Species-area and network-area relationships in host–helminth interactions

Tad A. Dallas1 and Pedro Jordano2

1Department of Biological Science, Louisiana State University, Baton Rouge, LA, USA
2Integrative Ecology Group, Estación Biológica de Doñana (EBD–CSIC), Avda. Americo Vespucio, Isla de La Cartuja, 41092 Sevilla, Spain

The scaling relationship observed between species richness and the geographical area sampled (i.e. the species-area relationship (SAR)) is a widely recognized macroecological relationship. Recently, this theory has been extended to trophic interactions, suggesting that geographical area may influence the structure of species interaction networks (i.e. network-area relationships (NARs)). Here, we use a global dataset of host–helminth parasite interactions to test existing predictions from macroecological theory. Scaling between single locations to the global host–helminth network by sequentially adding networks together, we find support that geographical area influences species richness and the number of species interactions in host–helminth networks. However, species-area slopes were larger for host species relative to their helminth parasites, counter to theoretical predictions. Lastly, host–helminth network modularity—capturing the tendency of the network to form into separate subcommunities—decreased with increasing area, also counter to theoretical predictions. Reconciling this disconnect between existing theory and observed SAR and NAR will provide insight into the spatial structuring of ecological networks, and help to refine theory to highlight the effects of network type, species distributional overlap, and the specificity of trophic interactions on NARs.

1. Introduction

The scaling relationship observed between the number of species and the geographical area sampled (i.e. the species-area relationship (SAR)) is a widely recognized ecological relationship [1,2]. Two forms of the relationship exist. The first considers geographically independent localities (often islands), and examines how species richness scales with geographical area. The second considers a contiguous geographical area, and subsamples the contiguous area scale. We focus on this second relationship.

There are several putative mechanisms underlying SARs, and specifically what species traits or environmental conditions influence the slope of the relationship [1,3,4]. From an island biogeographic perspective, SARs may be a function of the relationship between area and extinction rate. That is, larger areas may support larger populations, resulting in lower extinction rates in larger geographical areas [5,6]. Relatedly, available niche space should increase with geographical area, as larger geographical areas are more likely to contain a broader range of environmental conditions [7,8]. Regardless of the underlying mechanism, SARs are fundamental to community ecology [9–11], and are important to biological conservation efforts in fragmented systems [12,13].

SARs are typically examined at single trophic levels. However, several efforts to understand trophic consistency of SARs have been made [3,14,15], owing in part to the development of the trophic theory of island biogeography [16,17]. The explicit consideration of SARs for different trophic levels suggested that higher trophic levels would also have larger SAR slopes [3]. The
dependence of species of higher trophic levels on lower level species is one potential explanation for this difference in SARs. The scaling of species richness with area suggests that the number of species interactions may also scale with area, as the number of species interactions should depend on the number of potentially interacting species. One way to incorporate species interactions into the study of SARs is to consider how species interactions networks may change across spatial gradients, or in this case, with area. This approach recognizes that biodiversity is not simply the sum of unique species in an area, but is maintained by intraspecific and interspecific interactions among species which promote long-term co-occurrence.

According to the constant connectance hypothesis, the fraction of realized interactions should scale with network size, such that the number of interactions increases in a predictable fashion, but the fraction of all possible interactions remains constant. This assumes that the mean number of links per species does not saturate, instead that new species are added to the network in a manner that maintains connectance. While the constant connectance hypothesis sets a general upper bound on the relationship, the lower bound on the expected scaling of connectance is the link-species scaling hypothesis, which suggests that the number of links per species is constant (i.e. all species have the same mean number of links). Currently, both of these hypotheses have received quite mixed support in empirical data [18–20].

Apart from the effects of area and species richness on the number of ecological interactions between two trophic groups, it is important to consider the distribution of those interactions. How does the structure of interactions among species change with area and species richness? Considering the interactions between two groups (e.g. host and parasite) as a network of interactions between the groups (i.e. a bipartite network), measures of network structure can be examined for potential scaling relationships with geographical area. These network-area relationships (NARs) represent a multi-trophic extension of SARs, and may be useful in estimating the influence of area on network structure and stability [19]. This goes beyond simply capturing species number, but uses relationships between topological properties of ecological networks—which may be related to geographical area—to network dynamic measures related to stability and persistence [21,22]. Recent theoretical development of NARs has suggested that while species number and the number of interactions per species scales with area, some higher-level structural properties of the network do not change considerably (e.g. modularity; [19]). From the extension of community ecology theory to ecological networks, Galiana et al. [19] make several testable predictions generated from putative scaling relationships of food web network structure with area, which may apply to other ecological networks, such as host–parasite networks. We specifically examine host–helminth networks, owing to their diversity of species, transmission modes, and life histories. Further, data aggregation and curation efforts by the London Natural History Museum have produced an extensive dataset on host–helminth associations [23], allowing these questions to be tested from a pragmatic perspective.

(i) Host and helminth species richness should scale positively with increasing area, and the SAR slopes should be higher for the higher trophic level (i.e. the helminth parasites).

(ii) The average number of links per species should increase with geographical area in a saturating relationship.

(iii) Modularity should not be strongly influenced by geographical area, except when the landscape is heterogeneous, in which case we would expect an increase in modularity with increasing area.

The first prediction is a result of the SAR, an incredibly well-documented phenomenon [24], though the differences in slopes as a function of trophic level is a prediction stemming from previous work on food webs [3] based on island biogeography theory [16]. That is, higher trophic groups which rely on a subset of the resources available in a given area will have different richness scaling relationships as the resource community. The second prediction stems from the trophic sampling model of [19], which predicts that the mean number of links per species saturates with increasing area, as species are gained faster than novel links are gained. However, this relationship may not saturate if species interact with a different subset of species within increasing area, as species already gained will establish novel links with increasing area, leading to no clear saturation in the links-area relationship. Finally, network structure, and specifically modularity, is not expected to be influenced by geographical area generally. However, if adding geographical areas results in the addition of novel habitats or novel host and parasite species, increases in area may actually lead to increased modularity. Consider an island occupied by a set of dispersal-limited host and parasite species. Increasing from this network to include a nearby network with different, also dispersal-limited, host and parasite species will lead to a pronounced increase in modularity. There is some empirical evidence of this for ant-plant mutualistic networks at certain spatial scales [25].

Here, we use a global dataset of interactions between host species and helminth parasites [23] to examine the scaling in host–parasite network structure with increasing area. Apart from the clear implications to human and livestock health [26], host–parasite interactions offer an interesting system to test SAR and NAR for a number of reasons. First, helminth parasite species—especially terrestrial helminths—are unlikely to actively disperse, instead being vectored by their host species to new areas. This dispersal dependency may lead to differential scaling between host and parasite richness with increasing area, as well as influencing the scaling of parasite species richness with area. Further, host–parasite relationships may be quite species-specific, and require clear distributional overlap to occur. This creates a situation where host–parasite networks tend to be differently structured than other ecological networks [27], and these differences could strongly influence species-area and network-area scaling relationships.

In order to address the spatial scaling of the richness and interactions between host and helminth parasite species, we sequentially combined host–helminth networks constructed for geopolitical locations based on spatial proximity. We find significant SARs for both host and helminth parasite species, but detected larger SAR slopes for host species compared to helminth parasites, counter to theoretical predictions [19]. Second, we find an increase in the mean number of links per species with increasing area, though host and helminth parasites differed in the slope of this relationship. Finally, we find that network structure, specifically the tendency of species interactions to cluster into modules, decreased with
increasing area. Together, these results suggest that empirical host–parasite networks largely support existing species-area scaling theory, but highlight unexpected spatial scaling relationships in the slope of the relationship for host and parasite species. Reconciling this with existing theory will provide insight into the spatial structuring of ecological networks, and help to refine theory to highlight the effects of network type, species distributional overlap, and the specificity of trophic interactions on NARs.

2. Methods

(a) Host–helminth interaction data

Records of helminth parasite occurrence on host species were obtained from the parasite database of the London Natural History Museum [23], and accessed programmatically using the helmintR package [28]. These data currently represent one of the largest sources of host–parasite interaction data [29], despite being restricted to helminth parasites, including Platyhelminthes (trematodes and cestodes), Acanthocephalans, and Nematodes [23]. Host–helminth interactions are georeferenced to over 400 terrestrial and aquatic locations, largely determined by geopolitical boundaries (e.g. 'Spain'). Locations that were too vague or that other locations were nested within (e.g. 'western Europe') were removed, as were marine locations, terrestrial locations where area estimates were not available, and locations containing too few species for modularity calculation, resulting in a total of 281 locations (figure 1). The global network, the end point when sequentially adding location-specific sub-networks together, consisted of 10 680 host species, 16 929 helminth parasite species and the 55 159 associations between them (see the electronic supplementary material for further data description).

(b) Species-area and network-area relationships

To explore the scaling of species number and network structure with increasing geographical area, we selected 50 geographical locations randomly as starting points. At each starting point, we took the union of the initial network and the network which was geographically closest to the initial network, based on Haversine distance from the centre of the country. This process was repeated until the network incorporated host–helminth interactions for all countries (i.e. it was the global host–helminth network). With the addition of each country into the network, we calculated the number of unique host and helminth parasite species and the mean number of links per host and parasite species, as well as several measures which estimate aspects of network structure.

Specifically, we calculated connectance, or the fraction of realized links out of the maximum number of potential links ($H \times P$), the average number of links per species of each trophic level (i.e. hosts and parasites), and modularity using Barber’s $Q$ statistic [30]. Modularity was estimated using the random walk approach [31] in the R package igraph [32]. We standardized modularity based on the maximum modularity that we could obtain while maintaining the overall number of host and helminth parasite species and the overall number of interactions between species. This was performed by sequentially shuffling interactions to promote more modular structures, achieved by first creating a network with a fixed number of modules across a gradient of potential module sizes. Module sizes ranged from a minimum of 1 to a maximum where each interaction would be considered as a separate module. Networks of different module sizes were constructed ($n = 5000$). Within module connections were thinned randomly until the empirical network and the constructed network contained the same number of links. We then calculated modularity for the 5000 constructed networks, and took the maximum modularity value as the putative maximum modularity. By calculating the maximum modularity attainable for each module number, and taking the maximum value, we are able to approximate the maximum modularity the network could achieve given the existing empirical network constraints. The same relationship with area was found for standardized and unstandardized modularity (see the electronic supplementary material, figure S4 for unstandardized modularity).

The scaling of species richness and network properties with increasing cumulative area was examined using linear mixed effects models. Here, we examined species richness, links per species and modularity as a function of area and trophic level (i.e. host or helminth), while treating the identity of the iteration as a random effect, as spatial effects might produce differences in the shape of SARs or NARs owing to the initial starting location or how countries of different areas were combined. Models were fitted using the R package lme4 [33]. R code and data to reproduce the analyses are available as part of the electronic supplementary material.
3. Results

(a) Species-area relationships

Both host and helminth parasite species richness were positively related to cumulative area (table 1) after controlling for the effect of geographical starting position. Further, host and helminth species richness scaled with area differently, as helminth parasite species richness increased at a slightly slower rate than host species richness with increasing cumulative area (table 1 and figure 2), counter to theoretical predictions [19]. However, slope parameters from the SARs for host and helminth parasite species were positively related across the 50 different geographical starting points (electronic supplementary material, figure S2).

Further, we observe a positive scaling relationship with the average number of links per species as a function of cumulative area (table 1), after controlling for the effect of geographical starting position. This relationship differed between host and helminth parasite species with the number of links per host species increasing at a faster rate with area than the number of links per helminth parasite species (table 1). Theory predicts that this scaling relationship should saturate at large cumulative area values. However, we find limited evidence for a saturating relationship (figure 2), suggesting the possibility that novel links are added to existing species in the network, leading to the number of links per species consistently increasing with increasing area.

(b) Network-area relationships

Aside from the scaling of species richness or the average number of links per species, we might expect the structure of the network to change with increasing cumulative area. For instance, previous theoretical development suggests that the fraction of potential interactions (i.e. connectance) between host and helminth parasite species should remain constant with increasing cumulative area, as was demonstrated in food webs [34]. We find that connectance is large and variable when the networks are small, but increasing cumulative area reduces the variation, leading to a negative relationship between connectance and cumulative area (figure 3).

Lastly, it is expected that network structure—quantified here using modularity—would not differ as a function of cumulative area [19], barring a highly heterogeneous landscape that would promote a positive relationship between cumulative area and modularity. However, using a standardized measure of modularity which accounts for effects of the number of interacting species [35], we find that modularity decreases with increasing area (figure 4).

4. Discussion

We found clear support for species-area and network-area scaling relationships in host and helminth parasite species interactions. Specifically, species richness scaled positively

---

Table 1. Linear mixed effects model examining the scaling relationship between cumulative area and diversity of host and helminth parasite species. (Trophic level was important to model fit, suggesting that the number of helminth parasite species and the number of links per species increased at a slower rate for helminth parasites. (s.e. = standard error; d.f. = degrees of freedom.).)

<table>
<thead>
<tr>
<th>response</th>
<th>variable</th>
<th>estimate</th>
<th>s.e.</th>
<th>d.f.</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>species richness</td>
<td>log(area)</td>
<td>0.76</td>
<td>0.001</td>
<td>28094.51</td>
<td>512.47</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>trophic level</td>
<td>0.49</td>
<td>0.002</td>
<td>27816.96</td>
<td>206.31</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>links per species</td>
<td>log(area)</td>
<td>0.11</td>
<td>0.001</td>
<td>28057.41</td>
<td>155.51</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>trophic level</td>
<td>−0.49</td>
<td>0.001</td>
<td>27816.74</td>
<td>−434.23</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>modularity</td>
<td>log(area)</td>
<td>−0.02</td>
<td>0.0002</td>
<td>14047.70</td>
<td>−66.30</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Figure 2. The relationship between increasing log-transformed cumulative area and the number of unique host (blue) and helminth parasite (grey) species (a), and the average number of links per host and helminth species (b), suggesting that increasing area increases both the number of species and the average number of links per species. Dashed lines correspond to one standard deviation from the mean based on the 50 different geographical starting locations. (Online version in colour.)
and parasite species are novel additions to the growing species continue increasing, as geographically limited host increasing area makes the average number of links per positional turnover then may create a situation where difference between host and parasite spatial ranges and com-

new (or not) to the growing network are introduced. The number of links with increasing area, as parasites that are host already present in the network can gain an increasing each geographical location. This creates a situation where a parasites turnover quickly across spatial scales [36], and host parasites could be added from these isolated locations. The observed decrease in modularity with increasing area in host–helminth interaction networks suggests that the structure of species interactions—not simply the number of species or number of interactions—may adhere to a scaling rule similar to the SAR. It is important to note that the decrease in modularity with area was relatively small, and variation existed as a function of the initial geographical location. However, exploring how network structure is related to area for other types of networks (i.e. mutualistic networks) may provide some interesting insights into the role of interaction type on network-area scaling relationships. For instance, much of the current understanding of network-area scaling comes from food webs [19,37], which may operate in a fundamentally different manner relative to host–parasite systems. Previous work in an ant-plant mutualistic network demonstrated that spatial scale did influence resulting network structure, increasing modularity at regional scales [25]. Understanding how the structure of different ecological network types varies with increasing area [38] or across spatial gradients [39] is a clearly important research need. For instance, plant-pollinator network structure has previously been suggested to follow a latitudinal gradient [38], while similar structures have not been detected in host–parasite networks [40]. Further data collection and curation efforts will enable the thorough testing of the spatial scaling of ecological network structure.

Empirical examinations of scaling relationships in species richness, species interactions, and network structure with increasing geographical area are inherently subject to sampling, detection and reporting biases. Starting the simulations at different geographical locations and building up the network through the addition of the nearest neighbouring
location may help get around some of the biases. However, some of the existing theory assumes that if two trophically linked species occur in the same geographical area, they are presumed to interact [37]. This is not strictly true of host–parasite interactions, and is often not found in the empirical host–helminth data. Some part of this may be owing to detection biases, meaning that the co-occurring species do interact, but this interaction was simply not recorded. While this is possible, the host–helminth database examined here is presently one of the largest host–parasite databases in existence [29]. However, there are undoubtedly many unsampled interactions between host and helminth species in the data, and continued collection and curation efforts are vital to investigating macroecological patterns in only partially observed data.

The global scale of the host–helminth database is an incredible resource, but also is fundamentally different from previous explorations of NARs which largely focus on smaller spatial scales [19,37]. Recent attempts to explore link-species scaling patterns in global data argue that relationships hold across scales [20], though this examination was more related to the island SAR—treating each network independently—than the mainland SAR, which combines networks over some geographical area, building up to the full network. Large-scale spatial and climatic gradients present a fundamental limit to the distribution of species, resulting in potentially unique species subsets at large spatial scales. The degree to which spatial scale may modify theoretical predictions of link-area and NARs is presently largely unexplored.

Network size and fill can influence higher-level network properties (e.g. modularity), an issue that is not solved by using standardized z-scores [35]. However, standardizing by the maximal possible value of the network measure, we attempted to standardize measures of network modularity. Using standardized measures of network structure and acknowledging the influence of spatial scaling relationships on network structure is important when comparing ecological networks in different locations [39,41,42]. Future examinations at smaller spatial scales, and across a truly contiguous area would help address the influence of spatial sampling grain and the coarsity of considering host–parasite interaction networks at the geopolitical location on the resulting SAR and NAR. Lastly, tailoring theoretical expectations to the nature of the ecological interaction may help provide conceptual theory capable of explaining the differences between mutualistic and antagonistic networks. Even within antagonistic networks, host–parasite networks probably follow different scaling relationships with increasing area relative to predator–prey networks (i.e. food webs) given the specialized relationships between hosts and parasites [43].

Understanding the spatial variation in the number and structure of species–interactions—particularly of host–parasite interactions with implications for human health—is an important research frontier. Extending existing theory from community ecology into biogeographic space has led to the formation of macroecology. Considering how host–parasite interactions are structured along spatial and environmental gradients allows the ability to test and extend existing biogeographical and macroecological theory to interaction networks [39,44,45]. NARs, and other similar macroecological relationships of ecological networks, may elucidate spatial or environmental gradients of network structure and stability. Differences in SAR and NAR as a function of species interaction type (e.g. mutualism versus antagonism), phylogenetic scope, or spatial scale may provide insight into specific taxonomic groups or species interaction types that tend to be structured along environmental or spatial gradients.

Data accessibility. R code is available as part of the electronic supplementary material.

Authors’ contributions. All authors contributed equally.

Competing interests. We declare we have no competing interests.

Funding. This work has been supported by funding to the Viral Emergence Research Initiative (VERENA) consortium (NSF-B1-2021909) and to T.D. (NSF-DEB-2017826).

Acknowledgements. The work has been performed under the Project HPC-EUROPA3(INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Programme; in particular, the authors gratefully acknowledge the support of the Barcelona Supercomputing Centre. We thank Colin Carlson for helpful discussions, and several anonymous reviewers for their helpful comments.

References
