Missing links: testing the completeness of host-parasite checklists

ROBERT POULIN[1]*, ANNE A. BESSON[1], MATHIEU B. MORIN[1] *and* HASEEB S. RANDHAWA[2]

[1] *Department of Zoology, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand*
[2] *Ecology Degree Programme, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand*

(*Received 23 September 2015; revised 11 October 2015; accepted 12 October 2015; first published online 9 November 2015*)

SUMMARY

Host-parasite checklists are essential resources in ecological parasitology, and are regularly used as sources of data in comparative studies of parasite species richness across host species, or of host specificity among parasite species. However, checklists are only useful datasets if they are relatively complete, that is, close to capturing all host–parasite associations occurring in a particular region. Here, we use three approaches to assess the completeness of 25 checklists of metazoan parasites in vertebrate hosts from various geographic regions. First, treating checklists as interaction networks between a set of parasite species and a set of host species, we identify networks with a greater connectance (proportion of realized host–parasite associations) than expected for their size. Second, assuming that the cumulative rise over time in the number of known host–parasite associations in a region tends toward an asymptote as their discovery progresses, we attempt to extrapolate the estimated total number of existing associations. Third, we test for a positive correlation between the number of published reports mentioning an association and the time since its first record, which is expected because observing and reporting host–parasite associations are frequency-dependent processes. Overall, no checklist fared well in all three tests, and only three of 25 passed two of the tests. These results suggest that most checklists, despite being useful syntheses of regional host–parasite associations, cannot be used as reliable sources of data for comparative analyses.

Key words: checklists, connectance, interaction networks, helminths, species accumulation curves, vertebrates.

INTRODUCTION

Host-parasite checklists are fundamental tools in wildlife parasitology. Typically, such a checklist consists of the inventory of all parasite species found in each species of host belonging to a higher taxon, such as fish or birds, in a given geographical region; usually, this is accompanied by the reciprocal inventory, for each parasite species, of host species known to harbour them. Checklists are compiled from all existing surveys or other observational reports of parasites in wild hosts available from the literature at the time of compilation. For example, Moravec (2001) produced a checklist of parasites of freshwater fishes from the Czech and Slovak Republics, and Paredes-León *et al.* (2008) assembled a checklist of the parasites of amphibians and reptiles from Mexico. Host-parasite checklists are major contributions to parasitology for at least two reasons. First, they are the go-to resource for anyone contemplating a study of parasitism in a wild host population, as a guide to the kinds of parasites one might expect to find, or for anyone that comes across parasites as part of other types of studies, to provide clues regarding parasite identification. Second, they are often used as sources of data in comparative studies of the variation in parasite species richness across host species (e.g. Price and Clancy, 1983; Gregory *et al.* 1991; Sasal *et al.* 1997; Hughes and Page, 2007; Quiroz-Martinez and Salgado-Maldonado, 2013), or variation in host specificity among parasite species (e.g. Poulin, 1992; Edwards and Vidrine, 2006; Palm and Caira, 2008).

However, the usefulness of checklists depends on their completeness, that is, on how reliably they capture the full spectrum of host–parasite associations occurring in a particular region. All existing checklists are no doubt far from complete, but some are probably much further from completion than others. We can think of a host-parasite checklist as a bipartite interaction network, in which a set of P parasite species interact with a set of H host species (Poulin, 2010). Each possible link between a parasite species and a host species represents a potential host–parasite association. For each network, we have a number of known links, L, representing the total number of host–parasite associations that have been observed, with the maximum number of possible links equal to $P \times H$. The true number of existing links lies somewhere between L and $P \times H$, and it is its distance from L that determines how incomplete a particular checklist is.

There are at least three ways of determining whether the host-parasite checklist of a particular region is near completion. First, across well-sampled

---

* Corresponding author: Department of Zoology, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand. E-mail: robert.poulin@otago.ac.nz.

Parasitology (2016), **143**, 114–122. © Cambridge University Press 2015
doi:10.1017/S0031182015001559

local networks, the proportion of possible links that are actually realized, that is the connectance of the network or $L/(P \times H)$, decreases exponentially with increasing numbers of host or parasite species per network (Poulin, 2007; Mouillot *et al*. 2008). This pattern seems to apply universally to all kinds of ecological networks (Martinez, 1993; Stouffer *et al*. 2005). A large fraction of possible links is usually realized in species-poor networks, whereas this fraction becomes very small in species-rich networks. By applying the connectance *vs* species richness relationship to a large set of host-parasite checklists, it should be possible to identify checklists with greater connectance than expected for their species richness, and those with lower connectance. The former should represent relatively more complete checklists than the latter, as they include more host–parasite associations than one would expect for a typical network consisting of that number of species.

Second, the host-parasite networks that are presented as checklists are assembled piece by piece over time. After the first host–parasite association is reported in the literature for a given region, the number of known associations will grow over time until, in theory, the total number of existing host–parasite associations have been found. The rise in the cumulative number of known host–parasite associations as a function of time should follow a sigmoidal function and reach an asymptote, with the rate at which new ones are discovered decreasing as we are left with increasingly difficult ones to find (i.e. highly host-specific parasites occurring at very low prevalences). This is exactly the same situation as seen in species discovery: finding and reporting new host-parasite links follows the same temporal dynamics as finding and reporting new species. Thus, the methods long used to extrapolate species diversity within a large taxon or an area from species accumulation curves (Soberón and Llorente, 1993; Dolphin and Quicke, 2001; Cabrero-Sañudo and Lobo, 2003; Bebber *et al*. 2007) can be applied to host-parasite networks, to extrapolate the asymptote in the rising number of known host–parasite associations. Extrapolation is only possible, however, if the accumulation curve has passed its inflexion point, that is, if the rate at which new host-parasite links are reported has passed its maximum and started to slow down (Bebber *et al*. 2007). Checklists for which this is the case are not only closer to completion than those for which it is not, but we may also estimate the total number of existing (known and unknown) host-parasite links in those checklists.

Third, after a particular host–parasite association has been reported in the literature for the first time, it is likely to be observed and reported again in subsequent years. Assuming that observing and reporting host–parasite associations are frequency-dependent processes, determined mostly by the respective abundances of hosts and prevalences of parasites in a region, a positive relationship should develop between how often a given association has been reported and the time since its original discovery. The most common host–parasite associations should usually be among the first ones to be found, and should also be reported frequently thereafter. Checklists in which this inevitable relationship is either weak or non-existent are most likely based on very incomplete data, whereas those where this relationship emerges clearly are likely based on data from a region where the target host taxa have been well and repeatedly sampled for parasites.

Here, we use the above three approaches to test the completeness of multiple host-parasite checklists, involving metazoan parasites and vertebrate hosts. We show that most checklists perform poorly in all three tests of completeness; very few pass two of the three tests, and none pass all three tests. Our analysis provides solid grounds to discard most host-parasite checklists as potential sources of comparative data, because they are far from presenting a true and complete picture of natural host–parasite interactions.

MATERIALS AND METHODS

Data extraction from checklists

An internet search, using the keywords 'host parasite checklist' and the search engine Google, was conducted to find relevant checklists. Additional searches were conducted by replacing 'host' with 'fish', 'amphibian', 'reptile', 'bird', or 'mammal', and by replacing 'parasite' with 'helminth'. Of all relevant checklists found by the searches, only those that could be downloaded as PDF files or those that were available through the University of Otago library were used in the present study. We focused only on checklists covering broad host groups, such as fish or mammals, and not those restricted to host families. Although the final set of checklists does not include all published ones, it is nevertheless a large and representative sample of available ones, ranging from short (probably far from completion) to very long (possibly much closer to completion) lists of host–parasite associations.

When a checklist was updated by a new one published later on the same host higher taxon from the same region, the two were combined into a single network, with care taken to account for synonymies and changes in species nomenclature, as well as what only appeared as distinct species due to obvious spelling mistakes. If a checklist combined data for different higher taxa of hosts (e.g. amphibians and reptiles), or in the case of fish hosts combined those from freshwater and marine (including brackish waters) habitats, then these subsets were separated into different networks as they represent different host-parasite faunas. We only considered metazoan parasites, though some checklists focused on helminth parasites only; protozoan parasites were covered by some checklists but are not included in our analyses. We excluded all host-parasite combinations (i) observed only through experimental infection, or (ii) in which one of the two participants was not identified by a full Latin binomial name, as species identified only to genus or family level were possibly the same as those fully identified in an association with the same antagonist. We included the host–parasite associations for which the compilers of some checklists commented that misidentification of the parasite by the original authors was a possibility, since these amounted to only a small fraction of the total number of associations reported. When the authors of a checklist treated host subspecies or hybrids between host species as distinct hosts, so did we. Finally, in the case of helminths with complex life cycles, we treated different life stages of the same species, in the very rare cases in which they interacted with the same host species, as different parasite species if transition from one stage to the other required a transmission event (i.e. metacercaria and adult trematodes linked by trophic transmission) but not if that transition only required within-host growth (L3, L4 and adult stages in nematodes).

For each network, we recorded the number of parasite species, P, the number of host species, H, the number of observed host–parasite associations or links, L, and the connectance or $L/(P \times H)$. For each host–parasite association in each network, we recorded the year in which it was first recorded, and the number of published reports of that association.

Data analysis

First, we tested whether connectance, $L/(P \times H)$, decreases exponentially with increasing numbers of host and parasite species ($P + H$) per network. A better fit was achieved in log-log space, where the relationship becomes linear and negative. We therefore regressed log $[L/(P \times H)]$ against log $(P + H)$ and used the residuals of this regression to identify networks with greater (positive residuals) or lower (negative residuals) connectance than expected for their species richness. Initially, we performed separate regressions for networks with four different host types: freshwater fish, marine fish, herptiles (reptiles or amphibians) and endotherms (birds or mammals). However, as the slopes of the different regressions all fell within each other's 95% confidence intervals, in the end a single regression across all networks was used.

Second, to estimate the total number of host-parasite links in each network, we adopted the approach of Bebber *et al.* (2007), who use a generalized linear model with quasipoisson error to fit the model $E(L_t) = k(L_{tot} - L_{t-1})$, where $E(L_t)$ is the expected number of links discovered in year t, L_{tot} is the total number of existing links, L_{t-1} is the cumulative number of links described by year $t-1$, and k is a parameter to be estimated. We also considered a non-linear form for k, where $k = b + c\ L_{t-1}$ (see Bebber *et al.* 2007). The total number of existing host-parasite links or the bounds of its 95% confidence interval could not be estimated reliably in cases where the cumulative increase in the known number of records was erratic over time. In addition, we considered an estimated total invalid when its confidence interval was nonsensical, i.e. when its lower bound was lower than the currently known number of links and/or its upper bound was infinity.

Third, in each network, the relationship between the number of published reports mentioning a host–parasite association and the time elapsed, in years, between the year in which it was first recorded and the year of publication of the checklist, was tested across all host–parasite associations using Spearman rank correlations. A nonparametric test was necessary because data on the number of reports per host–parasite association are highly skewed and cannot be normalized by any type of transformation. The magnitude of the correlation coefficient, r_s, provided a measure of the strength of the relationship.

RESULTS

We obtained information from 18 (some with updates published after the original) host-parasite checklists from all over the world, published between 1972 and 2014, from which we derived 25 separate host-parasite networks (Table 1). These involved mostly fish hosts, and consisted of between 13 and 385 host species, 24 and 711 parasite species, and 27 and 1711 host-parasite links (Table 1). Across the networks, the number of host species correlated strongly with the number of parasite species per network (both log-transformed: $R^2 = 0{\cdot}728$, $N = 25$, $P < 0{\cdot}001$). Data on the year in which host–parasite associations were first recorded and/or the number of published reports of each association were unavailable, unclear, or not given in a way that was clearly linked to individual records, for 7 of these networks.

Across networks, connectance clearly decreased with increasing numbers of host and parasite species per network ($R^2 = 0{\cdot}594$, $N = 25$, $P < 0{\cdot}001$). The more host plus parasite species in a network, the smaller the proportion of possible host-parasite links that have been recorded (Fig. 1). The 4 networks with the clearly largest residual values of connectance, i.e. with the most observed links relative to what is expected from their species richness, were those for the freshwater fishes of the Czech and

Table 1. Summary data on the 25 host-parasite networks, derived from published checklists, included in the present analysis

Number	Hosts[a]	Parasites	Geographical region	No. host species, H	No. parasite species, P	No. host-parasite links, L (no. based on single record)	Connectance	Residual connectance[b]	Estimated total no. links (95% CI)[c]	Elapsed time *vs* no. records (Spearman r_s)	References
1	Fish (FW)	Metazoans	Ireland	23	56	199 (130)	0·155	0·331	235 (163 to NA)	0·2872 ($P < 0{\cdot}0001$)	Holland and Kennedy (1997)
2	Fish (M)	Metazoans	Germany	62	176	681 (355)	0·062	0·303	2047 (834 to NA)	0·3718 ($P < 0{\cdot}0001$)	Palm *et al*. (1999)
3	Fish (FW)	Metazoans	Latvia	54	188	783 (205)	0·077	0·400	540 (421 to NA)	0·6302 ($P < 0{\cdot}0001$)	Kirjusina and Vismanis (2007)
4	Fish (M)	Metazoans	Latvia	29	27	91 (10)	0·116	0·094	81 (54 to NA)	0·0072 ($P = 0{\cdot}9457$)	Kirjusina and Vismanis (2007)
5	Fish (FW)	Metazoans	Czech and Slovak Republics	81	332	1711 (545)	0·064	0·493	1666 (1592 to 1772)	0·5046 ($P < 0{\cdot}0001$)	Moravec (2001)
6	Fish (FW)	Metazoans	Turkey	75	136	369 (279)	0·036	0·026	525 (369 to NA)	0·2589 ($P < 0{\cdot}0001$)	Oktener (2003, 2014)
7	Fish (M)	Helminths	Turkey	56	95	210 (—)	0·039	−0·047	310 (179 to NA)	—	Oktener (2005)
8	Fish (FW)	Metazoans	Bangladesh	40	104	216 (102)	0·052	0·056	NA (25 to NA)	0·4759 ($P < 0{\cdot}0001$)	Arthur and Ahmed (2002)
9	Fish (M)	Metazoans	Bangladesh	29	41	46 (24)	0·039	−0·310	NA (30 to NA)	0·0896 ($P = 0{\cdot}5539$)	Arthur and Ahmed (2002)
10	Fish (FW)	Metazoans	Philippines	33	43	111 (62)	0·078	0·023	NA (91 to NA)	0·0806 ($P = 0{\cdot}4006$)	Arthur and Lumanlan-Mayo (1997)
11	Fish (M)	Metazoans	Philippines	116	110	207 (160)	0·016	−0·299	216 (196 to 257)	0·1024 ($P = 0{\cdot}1422$)	Arthur and Lumanlan-Mayo (1997)
12	Fish (FW)	Metazoans	New Zealand	27	35	117 (67)	0·124	0·155	123 (101 to NA)	0·2899 ($P = 0{\cdot}0015$)	Hewitt and Hine (1972); Hine *et al*. (2000)
13	Fish (M)	Metazoans	New Zealand	123	302	509 (393)	0·014	−0·164	541 (485 to 682)	0·1608 ($P = 0{\cdot}0003$)	Hewitt and Hine (1972); Hine *et al*. (2000)
14	Fish (FW)	Helminths	Mexico	179	183	1142 (677)	0·035	0·188	1376 (1156 to 2059)	0·5323 ($P < 0{\cdot}0001$)	Salgado-Maldonado (2006)
15	Fish (FW)	Metazoans	Brazil	385	711	1413 (—)	0·005	−0·274	—	—	Eiras *et al*. (2010)
16	Amphibians	Helminths	South America	175	212	682 (—)	0·018	−0·067	—	—	Campiao *et al*. (2014)
17	Amphibians	Metazoans	Mexico	61	113	283 (221)	0·041	0·017	1726 (423 to NA)	0·3295 ($P < 0{\cdot}0001$)	Paredes-León *et al*. (2008)
18	Lizards	Helminths	South America	110	125	317 (—)	0·023	−0·134	—	—	Avila and Silva (2010)
19	Reptiles	Helminths	Costa Rica	13	24	27 (—)	0·087	−0·171	NA (NA to NA)	—	Rodriguez-Ortiz *et al*. (2004*a*, *b*)
20	Reptiles	Metazoans	Mexico	205	236	693 (567)	0·014	−0·132	NA (1156 to NA)	0·1286 ($P = 0{\cdot}0007$)	Paredes-León *et al*. (2008)

Table 1. (*Cont.*)

Number	Hosts[a]	Parasites	Geographical region	No. host species, H	No. parasite species, P	No. host-parasite links, L (no. based on single record)	Connectance	Residual connectance[b]	Estimated total no. links (95% CI)[c]	Elapsed time *vs* no. records (Spearman r_s)	References
21	Reptiles	Nematodes	South Africa	41	59	98 (72)	0·041	−0·172	104 (86 to NA)	0·4106 ($P < 0·0001$)	Hering-Hagenbeck and Boomker (2000)
22	Birds	Helminths	Costa Rica	34	41	54 (—)	0·039	−0·287	NA (13 to NA)	—	Rodriguez-Ortiz *et al.* (2004*a*, *b*)
23	Birds	Metazoans	New Zealand	177	226	603 (—)	0·015	−0·140	—	—	Bishop and Heath (1998); McKenna (1998, 2010)
24	Mammals (Carnivora)	Helminths	Brazil	20	93	162 (118)	0·087	0·201	NA (NA to NA)	0·5777 ($P < 0·0001$)	Vieira *et al.* (2008)
25	Mammals	Helminths	Costa Rica	34	102	135 (119)	0·039	−0·087	NA (66 to NA)	0·0150 ($P = 0·8628$)	Rodriguez-Ortiz *et al.* (2004*a*, *b*)

[a] FW = freshwater, M = marine.

[b] From the linear regression of log connectance against log total number of host and parasite species per network.

[c] NA = calculation of the estimate or the bounds of its CI was not possible.

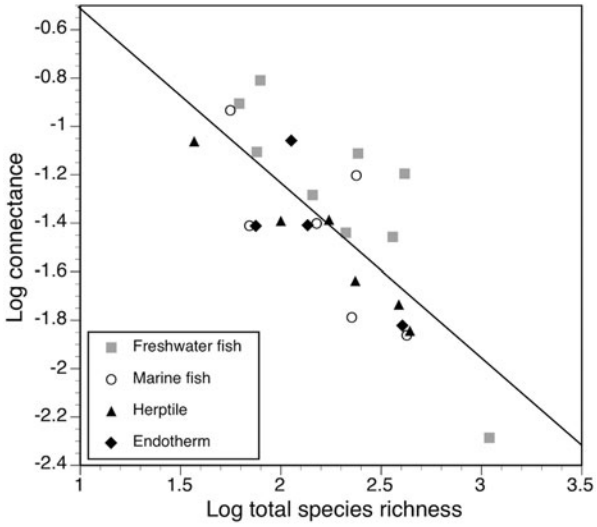

Fig. 1. Relationship between connectance and the total number of host and parasite species per network, in log-log space, among the 25 host-parasite networks. Different symbols represent networks with different host types: freshwater fish, marine fish, herptiles (reptiles or amphibians) and endotherms (birds or mammals). The line represents the following regression equation: log connectance = −0·762 (log total richness) + 0·304.

Slovak Republics, Latvia and Ireland, and the marine fishes of Germany (Table 1).

The cumulative increase over time in the known number of host–parasite associations showed little signs of slowing down in most cases. Also, it was often erratic, with significant monographs often causing a sharp spike in the year they were published, followed by few or no new findings for several years (see Fig. 2 for examples). As a consequence, the asymptote of the curve, corresponding to the total number of existing associations, could only be calculated with credible 95% confidence intervals for a single network out of the 21 for which data allowed this to be attempted (Table 1). This network was the one involving associations between helminths and freshwater fishes of Mexico, in which 1142 host–parasite associations are currently known and the total is estimated at 1376 (95% CI = 1156 to 2059).

In the majority of networks, more than half of host–parasite associations have only ever been recorded once (Table 1). Across all networks pooled, about 55% of host-–parasite associations have been recorded a single time, and most of the others fewer than 5 times (Fig. 3). The relationship between the number of published reports mentioning a host–parasite association and the time elapsed between the year of its first record and the year of publication of the checklist was positive and significant in 13 of the 18 networks in which it could be tested (Table 1). In all cases, there were associations first observed long ago that were never reported again, though most single records tended to be recent ones

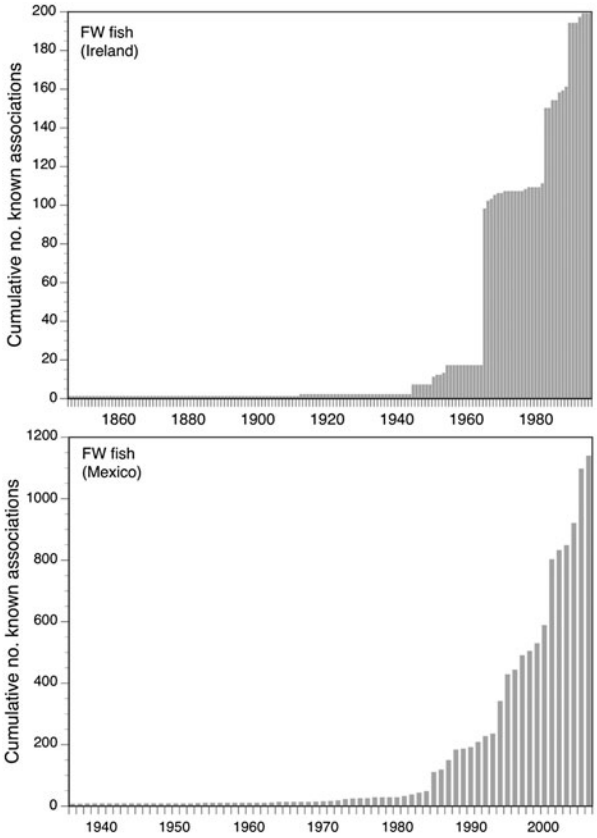

Fig. 2. Cumulative number of known host–parasite associations as a function of time, for metazoan parasites of Irish freshwater fish (period 1844–1997), and helminth parasites of Mexican freshwater fish (period 1936–2006).

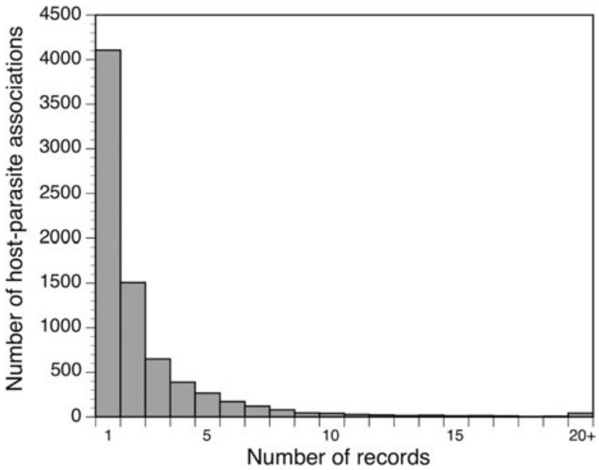

Fig. 3. Frequency distribution of the number of published records per host–parasite association, across all 7553 associations pooled from 18 host-parasite networks (see Table 1 for list of networks).

(see Fig. 4 for examples). However, the relationship between elapsed time and the number of published records was only strong ($r_s > 0{\cdot}5$) in 4 networks. These were the networks for the freshwater fishes of the Czech and Slovak Republics, Latvia and Mexico and the mammals of Brazil (Table 1).

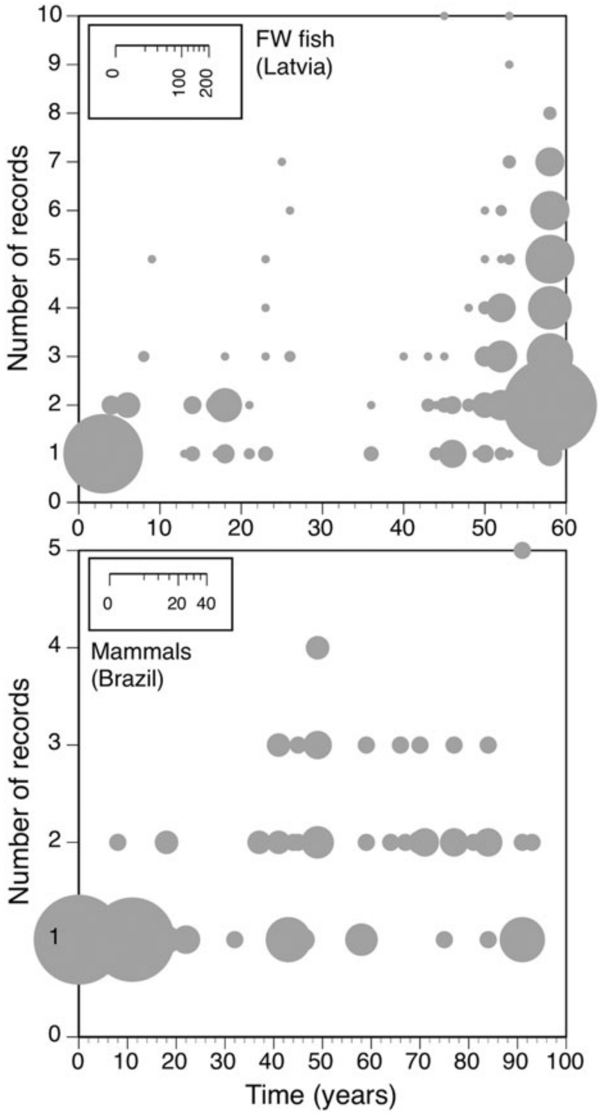

Fig. 4. Number of published records mentioning a host–parasite association as a function of the time elapsed between its first report and the year of publication of the checklist, across all known host–parasite associations for freshwater fish of Latvia and mammals of Brazil. The diameter of each point is proportional to the number of associations with those values (see scales).

DISCUSSION

Host-parasite checklists are essential resources that provide some expectations of the parasites one may find in a particular host species during a parasitological investigation. Checklists have also been used frequently as sources of data in comparative studies of either the variation in parasite species richness across host species, or variation in host specificity among parasite species (e.g. Price and Clancy, 1983; Gregory *et al*. 1991; Poulin, 1992; Sasal *et al*. 1997; Edwards and Vidrine, 2006; Hughes and Page, 2007; Palm and Caira, 2008; Quiroz-Martinez and Salgado-Maldonado, 2013). At the time of their publication, they typically provide an exhaustive list of all host–parasite associations known to date from a given region. However, that does not mean a complete list of all existing host–parasite associations in that region. Here, we used three approaches to estimate how close to complete were the 25 checklists for which we had data. None of the 25 networks passed all three tests; however, those involving the parasites of freshwater fishes in the Czech and Slovak Republics, Latvia and Mexico passed two of the tests.

The first approach consisted in treating host-parasite checklists as interaction networks, and determining whether the proportion of possible links that are actually realized, i.e. the connectance of the network, decreased with increasing total species richness. Local communities of hosts and their associated parasites are increasingly depicted and analysed as bipartite networks of interacting species to provide new insights about their structure (e.g. Vázquez *et al*. 2005; Lima *et al*. 2012; Bellay *et al*. 2015). We found that host-parasite checklists, although covering larger spatial scales and combining species that may not occur in actual sympatry, display some of the same properties as local interaction networks. For instance, there was a strong positive correlation between the number of host species and the number of parasite species per network, a pattern commonly observed among local host and parasite communities (Kamiya *et al*. 2014). Network connectance also decreased with increasing total host and parasite species richness, another pattern found across networks based on local communities (Poulin, 2007; Mouillot *et al*. 2008). In this relationship, observed connectance values that are higher than predicted ones indicate networks in which a greater proportion of host-parasite links has been observed than in a typical network consisting of that many host and parasite species. These are therefore more likely to approach a complete inventory of existing host–parasite associations. The networks for the freshwater fishes of the Czech and Slovak Republics, Latvia and Ireland and the marine fishes of Germany had distinctly higher positive residuals in the connectance-*vs*-richness regression than other networks. Although this suggests that a large proportion of existing host-parasite links have been discovered in those networks, it may also be that they consist of parasite species with low host specificity or host species compatible with many parasites, leading to unusually well connected networks for their size. Nevertheless, high positive residuals in the connectance-*vs*-richness regression are unlikely to indicate poorly-known networks.

The second approach we used assumed that the discovery of host–parasite associations in a region follows the same temporal pattern as that of species, with the rising cumulative known number eventually slowing down toward an asymptote as the total is approached. This should apply to the temporal discovery of any finite set of entities with varying probabilities of being observed, assuming a roughly constant research effort with occasional spikes (i.e. publication of major monographs). Using a method to extrapolate this asymptote based on the cumulative rise in known associations over time (Bebber *et al*. 2007), a credible estimate of the total number of existing host–parasite associations, with a proper confidence interval, could only be obtained for the network involving helminths of Mexican freshwater fishes. This may be because the discovery of parasite species in Mexican freshwater fish is itself at a very advanced stage (Pérez-Ponce de León and Choudhury, 2010). The analysis suggests that well over half, possibly more than 90%, of host–parasite associations have already been documented in this network, providing compelling evidence that this checklist is the only one considered here that comes close to presenting a relatively complete picture of host–parasite interactions in nature. Of course, this applies only to the host species covered by the checklist, and not to other fish species in the Mexican fauna that have not yet been investigated for parasites; their associations with parasites remain unknown. Also, it applies only to the drainage basins that have been sampled to date. The relative completeness of the network of associations among the currently known species does not mean that there are no more parasite species to be discovered within the Mexican freshwater fish fauna. Indeed, new parasite species continue to be found (see updates in Salgado-Maldonado and Quiroz-Martinez, 2013). However, among sampled hosts and parasite species known to date, the web of existing associations is well-known.

The final method we used made the plausible assumption that the most common host–parasite associations, i.e. those involving abundant host species and prevalent parasites, should tend to be among the first ones to be found, and should thereafter also be reported frequently. From this, we expected to find positive correlations between the time elapsed since an association was first observed and the total number of records of that association published since. We indeed found positive correlations in most networks, though these were only strong in four networks: those for the freshwater fishes of the Czech and Slovak Republics, Latvia and Mexico and the mammals of Brazil. These are the networks conforming most closely to the frequency-dependent pattern of temporal reporting expected a priori from situations where a regional group of hosts have been well-studied. Interestingly, our analysis also revealed that more than half of known host–parasite associations across all networks studied here have only been reported once; this fraction is even higher in several networks, surpassing two-thirds in those involving non-fish hosts. Thus, most host–parasite associations are reported once, and never again documented in the literature.

The last two methods we used to test for completeness could not be applied to several host-parasite checklists because these did not clearly identify all published reports documenting the occurrence of each host–parasite association. This made it impossible to obtain the year in which an association was first reported, or the total number of times it was reported. We urge future compilers of checklists to include this information clearly; it is not only valuable in its own right, but can also be used to ascertain how far from completion the current inventory is. The other piece of information lacking from almost all checklists is the sampling effort involved in the discovery of host–parasite associations. The original records almost invariably specify the number of host individuals examined for parasites; adding this information to each entry during the compilation of a checklist would allow other ways of assessing its completeness.

Admittedly, our set of checklists was biased toward those involving fish hosts (15 out of 25 networks considered here). It is possible that some of our findings would have been slightly different had we obtained more checklists of parasites from birds or mammals. Also, our analysis focused on checklists compiled for large areas (country or continent) and for broad taxonomic groups, such as vertebrate classes or all metazoans or helminths. It is very likely that checklists restricted to smaller spatial or taxonomic scales attain greater levels of completeness (e.g. Martinez-Aquino *et al*. 2014). Nevertheless, completeness should be tested, and not assumed.

There is one obvious reason why no current checklist can be complete: the recent application of molecular genetics to species identification is revealing a substantial cryptic diversity of parasites (Pérez-Ponce de León and Nadler, 2010; Poulin, 2011). Cryptic parasite species are morphologically indistinguishable from each other, and can only be separated using molecular markers. Most of the older records on which checklists are based have relied exclusively on morphology for species identification. Nevertheless, if we accept parasite morphospecies as true ‘species’, it is clear that most existing checklists do not present a summary of host–parasite interactions that is remotely close to what really exists in nature. In our analysis, no checklist out of 25 passed all three of our tests for completeness, and only three (those for the parasites of freshwater fishes in the Czech and Slovak Republics, Latvia and Mexico) passed two of the tests. By virtue of synthesizing information from multiple available records into a single list all host-parasite checklists are extremely useful. However, their usefulness has limits: they should not be used as reliable sources of data on parasite species richness per host species, or data on host specificity of different parasite species, unless it can be shown that the information they capture is relatively close to a full representation of natural host-parasite communities. Based on the present analysis, we suspect most existing checklists cannot be used as data sources for such analyses.

ACKNOWLEDGEMENTS

We thank Gerardo Pérez-Ponce de León for comments on an earlier version of the manuscript. Some of the ideas in the present study have benefitted from useful discussion within the Macroecology of Infectious Disease Research Coordination Network (funded by NSF/NIH/USDA DEB 131223), and the Otago Ecological Parasitology Group.

FINANCIAL SUPPORT

This study was funded by an internal grant from Otago University’s Department of Zoology to R. P.

REFERENCES

Arthur, J. R. and Ahmed, A. T. A. (2002). *Checklist of the Parasites of Fishes of Bangladesh*. Technical Paper no. 369/1, FAO Fisheries, Rome.

Arthur, J. R. and Lumanlan-Mayo, S. (1997). *Checklist of the Parasites of Fishes of the Philippines*. Technical Paper no. 369, FAO Fisheries, Rome.

Avila, R. W. and Silva, R. J. (2010). Checklist of helminths from lizards and amphisbaenians (Reptila, Squamata) of South America. *Journal of Venomous Animals and Toxins Including Tropical Diseases* **16**, 543–572.

Bebber, D. P., Marriott, F. H. C., Gaston, K. J., Harris, S. A. and Scotland, R. W. (2007). Predicting unknown species numbers using discovery curves. *Proceedings of the Royal Society B* **274**, 1651–1658.

Bellay, S., de Oliveira, E. F., Almeida-Neto, M., Mello, M. A. R., Takemoto, R. M. and Luque, J. L. (2015). Ectoparasites and endoparasites of fish form networks with different structures. *Parasitology* **142**, 901–909.

Bishop, D. M. and Heath, A. C. G. (1998). Parasites of birds in New Zealand: checklist of ectoparasites. *Surveillance* **25**, 15–26.

Cabrero-Sañudo, F. J. and Lobo, J. M. (2003). Estimating the number of species not yet described and their characteristics: the case of Western Palaearctic dung beetle species (Coleoptera, Scarabaeoidea). *Biodiversity and Conservation* **12**, 147–166.

Campiao, K. M., Morais, D. H., Dias, O. T., Aguiar, A., Toledo, G. M., Tavares, L. E. R. and da Silva, R. J. (2014). Checklist of helminth parasites of amphibians from South America. *Zootaxa* **3843**, 1–93.

Dolphin, K. and Quicke, D. L. J. (2001). Estimating the global species richness of an incompletely described taxon: an example using parasitoid wasps (Hymenoptera: Braconidae). *Biological Journal of the Linnean Society* **73**, 279–286.

Edwards, D. D. and Vidrine, M. F. (2006). Host specificity among *Unionicola* spp. (Acari: Unionicolidae) parasitizing freshwater mussels. *Journal of Parasitology* **92**, 977–983.

Eiras, J. C., Takemoto, R. M. and Pavanelli, G. C. (2010). *Diversidade dos Parasitas de Peixes de Agua Doce do Brasil*. Clichetec, Maringa, Brazil.

Gregory, R. D., Keymer, A. E. and Harvey, P. H. (1991). Life history, ecology and parasite community structure in Soviet birds. *Biological Journal of the Linnean Society* **43**, 249–262.

Hering-Hagenbeck, S. F. B. N. and Boomker, J. (2000). A check-list of the nematode parasites of South African Serpentes (snakes) and Sauria (lizards). *Onderstepoort Journal of Veterinary Research* **67**, 1–13.

Hewitt, G. C. and Hine, P. M. (1972). Checklist of parasites of New Zealand fishes and of their hosts. *New Zealand Journal of Marine and Freshwater Research* **6**, 69–114.

Hine, P. M., Jones, J. B. and Diggles, B. K. (2000). *A Checklist of the Parasites of New Zealand Fishes, Including Previously Unpublished Records*. Technical Report no. 75, National Institute of Water and Atmospheric Research, Wellington.

Holland, C. V. and Kennedy, C. R. (1997). A checklist of parasitic helminth and crustacean species recorded in freshwater fish from Ireland. *Proceedings of the Royal Irish Academy* **97B**, 225–243.

Hughes, J. and Page, R. D. M. (2007). Comparative tests of ectoparasite species richness in seabirds. *BMC Evolutionary Biology* **7**, 227.

Kamiya, T., O'Dwyer, K., Nakagawa, S. and Poulin, R. (2014). Host diversity drives parasite diversity: meta-analytical insights into patterns and causal mechanisms. *Ecography* **37**, 689–697.

Kirjusina, M. and Vismanis, K. (2007). *Checklist of the Parasites of Fishes of Latvia*. Technical Paper no. 369/3, FAO Fisheries, Rome.

Lima, D. P., Giacomini, H. C., Takemoto, R. M., Agostinho, A. A. and Bini, L. M. (2012). Patterns of interactions of a large fish-parasite network in a tropical floodplain. *Journal of Animal Ecology* **81**, 905–913.

Martinez, N. D. (1993). Effect of scale on food web structure. *Science* **260**, 242–243.

Martinez-Aquino, A., Mendoza-Palmero, C. A., Aguilar-Aguilar, R. and Pérez-Ponce de León, G. (2014). Checklist of helminth parasites of Goodeinae (Osteichthyes: Cyprinodontiformes: Goodeidae), an endemic subfamily of freshwater fishes from Mexico. *Zootaxa* **3856**, 151–191.

McKenna, P. (1998). Parasites of birds in New Zealand: checklist of helminth and protozoan parasites. *Surveillance* **25**, 4–12.

McKenna, P. (2010). An updated checklist of helminth and protozoan parasites of birds in New Zealand. *Webmed Central Parasitology* **1**, WMC00705.

Moravec, F. (2001). *Checklist of the Metazoan Parasites of Fishes of the Czech Republic and the Slovak Republic 1873–2000*. Academia, Prague.

Mouillot, D., Krasnov, B. R., Shenbrot, G. I. and Poulin, R. (2008). Connectance and parasite diet breadth in flea-mammal webs. *Ecography* **31**, 16–20.

Oktener, A. (2003). A checklist of metazoan parasites recorded in freshwater fish from Turkey. *Zootaxa* **394**, 1–28.

Oktener, A. (2005). A checklist of parasitic helminths reported from sixty-five species of marine fish from Turkey including two new records of monogeneans. *Zootaxa* **1063**, 33–52.

Oktener, A. (2014). Revision of parasitic helminths reported in freshwater fish from Turkey with new records. *Transylvanian Review of Systematical and Ecological Research* **16**, 1–56.

Palm, H. W. and Caira, J. N. (2008). Host specificity of adult versus larval cestodes of the elasmobranch tapeworm order Trypanorhyncha. *International Journal for Parasitology* **38**, 381–388.

Palm, H. W., Klimpel, S. and Bucher, C. (1999). *Checklist of Metazoan Fish Parasites of German Coastal Waters*. Report no. 307, Institute for Marine Science, Christian-Albrechts University, Kiel.

Paredes-León, R., Garcia-Prieto, L., Guzmán-Cornejo, C., León-Règagnon, V. and Pérez, T. M. (2008). Metazoan parasites of Mexican amphibians and reptiles. *Zootaxa* **1904**, 1–166.

Pérez-Ponce de León, G. and Choudhury, A. (2010). Parasite inventories and DNA-based taxonomy: lessons from helminths of freshwater fishes in a megadiverse country. *Journal of Parasitology* **96**, 236–244.

Pérez-Ponce de León, G. and Nadler, S. A. (2010). What we don't recognize can hurt us: a plea for awareness about cryptic species. *Journal of Parasitology* **96**, 453–464.

Poulin, R. (1992). Determinants of host-specificity in parasites of freshwater fishes. *International Journal for Parasitology* **22**, 753–758.

Poulin, R. (2007). Are there general laws in parasite ecology? *Parasitology* **134**, 763–776.

Poulin, R. (2010). Network analysis shining light on parasite ecology and diversity. *Trends in Parasitology* **26**, 492–498.

Poulin, R. (2011). Uneven distribution of cryptic diversity among higher taxa of parasitic worms. *Biology Letters* **7**, 241–244.

Price, P. W. and Clancy, K. M. (1983). Patterns in number of helminth parasite species in freshwater fishes. *Journal of Parasitology* **69**, 449–454.

Quiroz-Martinez, B. and Salgado-Maldonado, G. (2013). Taxonomic distinctness and richness of helminth parasite assemblages of freshwater fishes in Mexican hydrological basins. *PLoS ONE* **8**, e74419.

Rodriguez-Ortiz, B., Garcia-Prieto, L. and Pérez-Ponce de León, G. (2004*a*). Checklist of the helminth parasites of vertebrates in Costa Rica. *Revista de Biologia Tropical* **52**, 313–354.

Rodriguez-Ortiz, B., Garcia-Prieto, L. and Pérez-Ponce de León, G. (2004*b*). Addendum to the checklist of the helminth parasites of vertebrates in Costa Rica. *Revista de Biologia Tropical* **52**, 355–362.

Salgado-Maldonado, G. (2006). Checklist of helminth parasites of freshwater fishes from Mexico. *Zootaxa* **1324**, 1–357.

Salgado-Maldonado, G. and Quiroz-Martinez, B. (2013). Taxonomic composition and endemism of the helminth fauna of freshwater fishes of Mexico. *Parasitology Research* **112**, 1–18.

Sasal, P., Morand, S. and Guégan, J.-F. (1997). Determinants of parasite species richness in Mediterranean marine fish. *Marine Ecology Progress Series* **149**, 61–71.

Soberón, J. and Llorente, J. (1993). The use of species accumulation functions for the prediction of species richness. *Conservation Biology* **7**, 480–488.

Stouffer, D. B., Camacho, J., Guimerà, R., Ng, C. A. and Amaral, L. A. N. (2005). Quantitative patterns in the structure of model and empirical food webs. *Ecology* **86**, 1301–1311.

Vázquez, D. P., Poulin, R., Krasnov, B. R. and Shenbrot, G. I. (2005). Species abundance and the distribution of specialization in host-parasite interaction networks. *Journal of Animal Ecology* **74**, 946–955.

Vieira, F. M., Luque, J. L. and Muniz-Pereira, L. C. (2008). Checklist of helminth parasites in wild carnivore mammals from Brazil. *Zootaxa* **1721**, 1–23.