

# Host specificity and the probability of discovering species of helminth parasites

R. POULIN<sup>1\*</sup> and D. MOUILLOT<sup>2</sup>

<sup>1</sup> Department of Zoology, University of Otago, P.O. Box 56, Dunedin, New Zealand

<sup>2</sup> UMR CNRS-UMII 5119 Ecosystèmes Lagunaires, Université Montpellier II, CC 093, 34095 Montpellier cedex 5, France

(Received 4 November 2004; revised 1 December 2004; accepted 1 December 2004)

## SUMMARY

Different animal species have different probabilities of being discovered and described by scientists, and these probabilities are determined to a large extent by the biological characteristics of these species. For instance, species with broader geographical ranges are more likely to be encountered by collectors than species with restricted distributions; indeed, the size of the geographical range is often the best predictor of a species' date of description. For parasitic organisms, host specificity may be similarly linked to the probability of a species being found. Here, using data on 170 helminth species parasitic in freshwater fishes, we show that host specificity is associated with the year in which the helminths were described. Helminths that exploit more host species, and to a lesser degree those that exploit a broader taxonomic range of host species, tend to be discovered earlier than the more host-specific helminths. This pattern was observed across all helminth species, as well as within the different helminth taxa (trematodes, cestodes, nematodes and acanthocephalans). Our results demonstrate that the parasite species known at any given point in time are not a random subset of existing species, but rather a biased subset with respect to the parasites' biological properties.

Key words: freshwater fish, host specificity, taxonomic distinctness, year of description.

## INTRODUCTION

The list of living species known to science gets longer every day, with still a huge number of species remaining to be found and described. It is becoming increasingly clear that, although it involves an element of chance, species discovery is not random, i.e. the species in a given higher taxon that are found and described first are not a random subset of all living species in that taxon. Various biological characteristics of animal species apparently influence their probability of being discovered, such that all species in a higher taxon are not equally likely to be found rapidly. For instance, in many taxa ranging from insects to vertebrates, larger species are described earlier than smaller species (Gaston, 1991; Gaston & Blackburn, 1994; Gaston, Blackburn & Loder, 1995; Reed & Boback, 2002). This is easy to explain: larger animals are easier to detect and are thus found and formally described before the smaller species are discovered. Another important species attribute associated with the probability of discovery is geographical range (Blackburn & Gaston, 1995; Gaston *et al.* 1995; Allsop, 1997; Collen, Purvis & Gittleman, 2004). Species with broader geographical ranges are more likely to be encountered by collectors than species with restricted distributions. In fact,

among mammals, which probably represent the best-known animal taxon, the size of the geographical range is by far the best predictor of a species' date of description (Collen *et al.* 2004). Clearly, therefore, at any point in time, our view of biodiversity must be distorted, because the species that we know are not truly a random and representative sample of the species that exist; over time, as more species are described, the bias should become smaller and smaller.

The same phenomenon must apply to parasite biodiversity. There must exist biological determinants of the probability of discovering a parasite species (Poulin & Morand, 2004). For instance, in the case of ectoparasites of fish, with the exception of some specific field collections and recent studies involving microscopy, most new species described prior to the mid-19th century have been found as a secondary outcome of fisheries surveys. Thus, we might expect that the largest species, which are more obvious to the naked eye, would be discovered first. Indeed, body size of ectoparasite species correlates negatively with their year of description, among both copepods (Poulin, 1996) and monogeneans (Poulin, 2002) parasitic on fish. In contrast, internal parasites are normally found following dissections specifically aimed at parasite recovery. Thus, if investigators are purposely looking for internal helminths, we might expect that they would find them more or less independently of their size. Accordingly, there is no relationship between year of description and body size in trematodes parasitic in fish (Poulin, 1996).

\* Corresponding author: Department of Zoology, University of Otago, P.O. Box 56, Dunedin, New Zealand. Tel: +64 3 479 7983. Fax: +64 3 479 7584. E-mail: robert.poulin@stonebow.otago.ac.nz

Table 1. Results of linear regressions between three measures of host specificity and year of description in four groups of helminths parasitic in Canadian freshwater fishes

Helminth taxon	Measure of specificity	<i>N</i>	<i>r</i>	<i>P</i>
Trematodes				
	Log number of host species	63	-0.373	0.0026
	$S_{TD}$	39	-0.302	0.0619
	$VarS_{TD}$	39	-0.140	0.3951
Cestodes				
	Log number of host species	50	-0.422	0.0023
	$S_{TD}$	32	-0.380	0.0320
	$VarS_{TD}$	32	-0.363	0.0412
Nematodes				
	Log number of host species	39	-0.338	0.0355
	$S_{TD}$	22	-0.006	0.9782
	$VarS_{TD}$	22	-0.343	0.1185
Acanthocephalans				
	Log number of host species	18	-0.734	0.0005
	$S_{TD}$	15	-0.335	0.2217
	$VarS_{TD}$	15	0.021	0.9408

Thus, for endoparasite species, biological properties other than body size may be important determinants of the probability of discovery.

Since geographical range affects the likelihood that a species is encountered for free-living animals (e.g. Allsop, 1997; Collen *et al.* 2004), it may also be important for parasites. The geographical range size of an animal is essentially equal to the number of localities in which it occurs. For parasites, this is roughly equivalent to the number of host species in which a parasite occurs. Parasite species that occur in many host species are probably more likely to be discovered and described earlier than highly host-specific species, because the former have a higher probability than the latter of being found in a sample consisting of a random assortment of host species. Therefore, we should expect a negative relationship between the number of host species used by a parasite species and its date of description. It is not just the number of host species that matter, but also their identity. Consider two parasite species, A and B, that both exploit the same number of host species; parasite A exploits host species belonging to a wide range of taxa, however, whereas parasite B exploits a narrow taxonomic range of hosts, for instance hosts that all belong to the same genus. Surely parasite A has a higher probability of being discovered first than parasite B. The diverse host species used by A are likely to occupy a wider range of habitats, or display a wider range of body sizes, and are therefore likely to be susceptible to a wider range of sampling methods, than those of parasite B. Thus, we should expect the year in which parasite species are described to relate negatively not only with the number of host species

they exploit, but also with the taxonomic diversity of those hosts.

Here, we test both these predictions using data on helminth endoparasites of freshwater fishes. In addition to the number of host species exploited by a parasite, we used an index of host specificity that takes into account the taxonomic affinities of these hosts (Poulin & Mouillot, 2003). Our analysis demonstrates that host specificity is associated with the probability of finding parasite species, and suggests that the parasite species not yet discovered are most likely more host-specific, on average, than the ones we already know.

#### METHODS

We obtained data on the host fish species used by species of trematodes, cestodes, nematodes and acanthocephalans parasitizing Canadian freshwater fishes. The data were obtained from the checklists of Margolis & Arthur (1979) and McDonald & Margolis (1995). Data were gathered only for parasite species that use fish as definitive hosts, i.e. that occur in fish as adult worms.

The year in which each parasite species was first described was recorded. In cases of synonymy or subsequent revisions of the name or taxonomic position of a species, we took the original date of description, which corresponds to the time at which the species was first discovered.

Three different measures of host specificity were computed for these parasite species: (1) the number of known host species, (2) an index of host taxonomic distinctness,  $S_{TD}$ , and (3) the variance associated with this index,  $VarS_{TD}$ . The latter two measures were obtained only for parasite species with more than 1 known host species.

The specificity index  $S_{TD}$  measures the average taxonomic distinctness of all host species used by a parasite species (see Poulin & Mouillot, 2003). When these host species are placed within a taxonomic hierarchy, based on the Linnean classification, the average taxonomic distinctness is simply the mean number of steps up the hierarchy that must be taken to reach a taxon common to two host species, computed across all possible pairs of host species. Thus, if two host species are congeners, one step (species-to-genus) is necessary to reach a common node in the taxonomic tree; if the two species belong to different genera but the same family, 2 steps will be necessary (species-to-genus, and genus-to-family); and so on, with these numbers of steps averaged across all host species pairs. Equal step lengths of 1 unit are postulated between each level in the taxonomic hierarchy, in the absence of actual genetic distances. The greater the taxonomic distinctness between host species, the higher the number of steps needed, and the higher the value of the index  $S_{TD}$ : thus it is inversely proportional to specificity. A high index

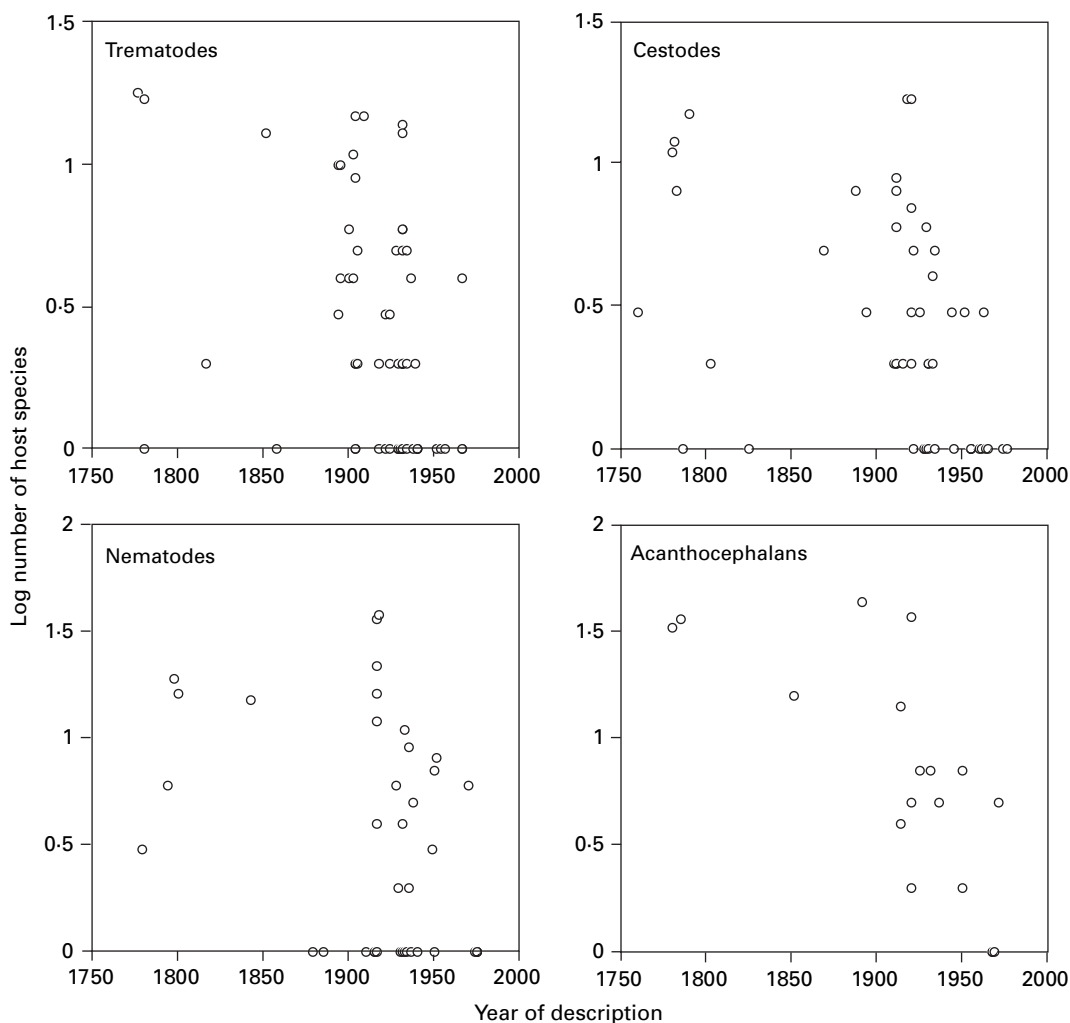


Fig. 1. Host specificity (number of known host species) of parasites as a function of the year in which they were described, presented separately for four taxa of helminths using Canadian freshwater fishes as definitive hosts.

value means that on average the hosts of a parasite species are not closely related. Using the standard 5 taxonomic levels above species, i.e. genus, family, order, class and phylum, the maximum value that the index  $S_{TD}$  can take (when all host species belong to different classes) is 5, and its lowest value (when all host species are congeners) is 1. The fish taxonomy used for the calculations was that proposed by Nelson (1994), because his comprehensive taxonomic hierarchy is based on a phylogenetic scheme. Asymmetries in the taxonomic distribution of host species across higher taxa can sometimes be missed by  $S_{TD}$ , which is only the average taxonomic distinctness; in these situations the variance in taxonomic distinctness,  $VarS_{TD}$ , can provide complementary information (see Poulin & Mouillot, 2003).

Data on number of known host species were log-transformed prior to all analyses. Analyses were computed both within each helminth group separately (trematodes, cestodes, nematodes, and acanthocephalans), and across all helminth species. Relationships between year of description and the three measures of host specificity were first assessed

separately using linear regressions. Then, the three measures were used as independent variables in multiple regressions, with year of description as the dependent variable.

## RESULTS

We used data for a total of 170 species of helminths parasitic in Canadian fishes. Together, these parasites occur in 147 fish species, representing approximately 81% of the entire Canadian freshwater fish fauna (Scott & Crossman, 1973). Of the 170 helminths, 108 species had at least 2 host species and could therefore be assigned both a  $S_{TD}$  and a  $VarS_{TD}$  value. The year of description for these helminths ranged between 1760 and 1976, with the majority described after 1900.

Within the different groups of helminths, the number of host species exploited by a parasite species was the only measure of host specificity consistently related to year of description (Table 1). As a rule, this negative relationship is explained by the fact that most species discovered before 1900 are known to

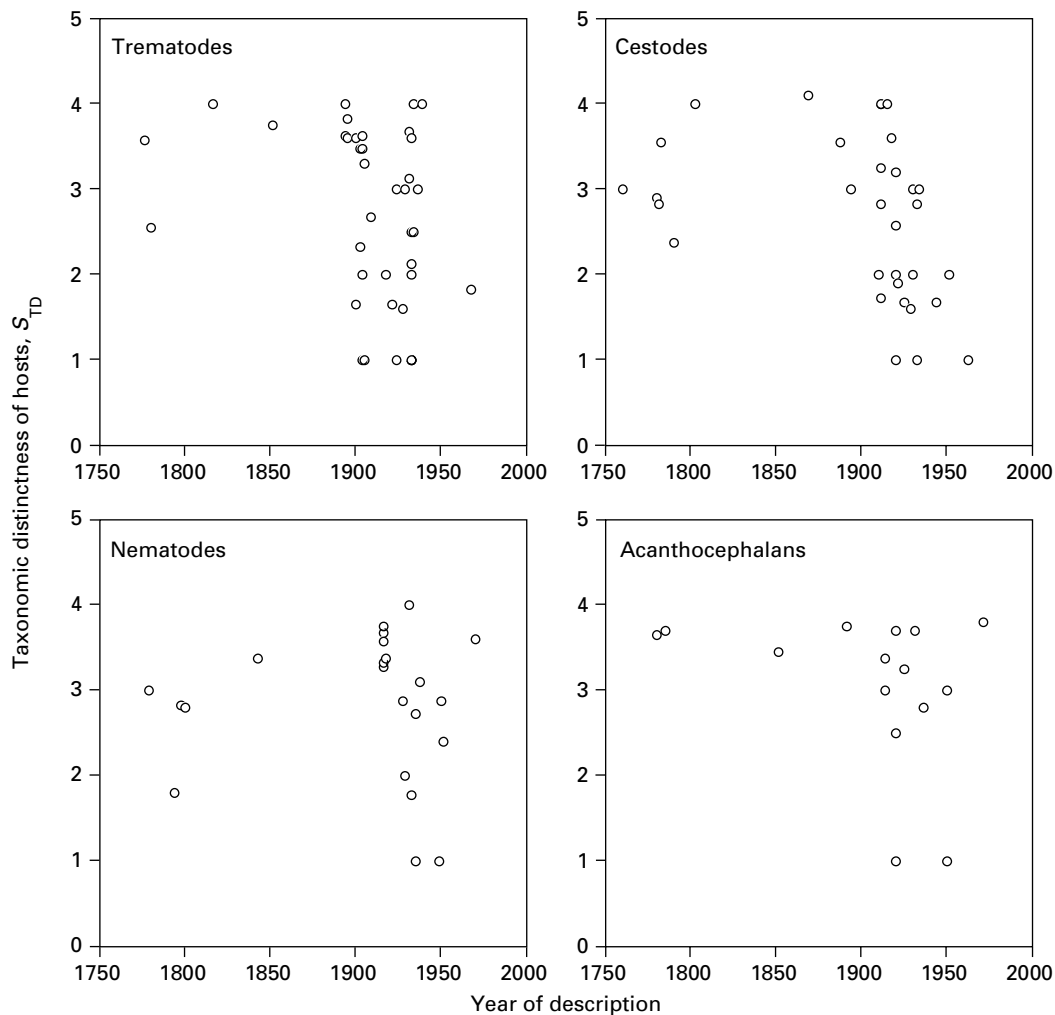


Fig. 2. Host specificity ( $S_{TD}$ ) of parasites as a function of the year in which they were described, presented separately for four taxa of helminths using Canadian freshwater fishes as definitive hosts.

exploit several host species, whereas most strictly host-specific helminths have been found after 1900 (Fig. 1). We also obtained weak negative relationships between year of description and either  $S_{TD}$  in trematodes and cestodes, or  $VarS_{TD}$  in cestodes (Table 1). Once again, this pattern can be explained by the fact that helminth species known before 1900 exploit taxonomically diverse host species, whereas those exploiting a narrow taxonomic range of hosts have mainly been found since 1900 (Fig. 2). When the three measures of host specificity are entered as predictors in a multiple regression with year of description as the dependent variable, only the number of host species exploited emerges as a significant predictor, and only for trematodes ( $P=0.0314$ ) and acanthocephalans ( $P=0.0098$ ). Still, these multiple regressions explain between 20% and 53% of the variation in year of description among helminth species, indicating that host specificity is associated with the probability of a parasite species being discovered and described.

As can be seen in Figs 1 and 2, many of the trends reported above are dependent on a few species

described prior to 1850. When these species are excluded from the analyses, relationships involving trematodes and nematodes become non-significant ( $P>0.05$ ).

When all helminth species are pooled, their year of description is negatively associated with all three measures of host specificity: the number of host species exploited ( $r=-0.401$ ,  $N=170$ ,  $P=0.0001$ ),  $S_{TD}$  ( $r=-0.257$ ,  $N=108$ ,  $P=0.0074$ ) and  $VarS_{TD}$  ( $r=-0.251$ ,  $N=108$ ,  $P=0.0089$ ). This is apparent when looking at how the average host specificity of all helminths known to date changes over time: the mean number of host species used by all helminths known to that date, and to a lesser extent their  $S_{TD}$ , has tended to decline over time (Fig. 3). The decrease happens because over time, the more host-specific helminths (with few host species, and a low  $S_{TD}$  value) are progressively found and added to the list. As in the analyses within the separate groups of helminths, the multiple regression including all three measures of host specificity returned only the number of host species exploited as a significant ( $P=0.0312$ ) predictor of the year of

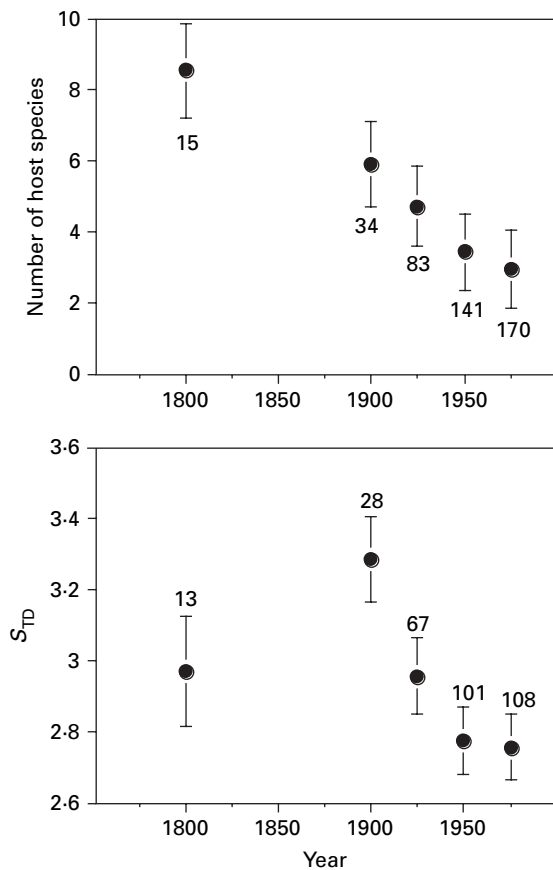


Fig. 3. Average ( $\pm$  S.E.) host specificity (number of known host species and  $S_{TD}$ ) of parasites known to date at five times over the past two centuries, for helminth parasites using Canadian freshwater fishes as definitive hosts. Number of known host species is presented as a geometric mean, i.e. as a back-transformed mean computed on log-transformed data. Sample sizes are indicated next to the symbols.

description across all helminth species. The multiple regression explained 15% of the variance in year of description.

#### DISCUSSION

Certain species are easier to detect than others, and this should relate directly to their probability of being discovered and described. Our results indicate that for internal helminths of fish, host specificity is an important biological correlate of year of description, and thus a key determinant of the probability of species discovery. Two separate aspects of host specificity, the number of host species used and their taxonomic distinctness, both play roles, although the former appears more important. Human efficiency at finding different types of parasites is therefore not independent from the biological features of these parasites.

Our results are based strictly on the helminth parasites of Canadian freshwater fishes. The Canadian freshwater fish fauna is not particularly

species-rich, given the size of the country (Scott & Crossman, 1973). It includes taxa found only in North America, such as the families Hiodontidae and Centrarchidae, as well as taxa also found on many other continents, such as the Cyprinidae, Salmonidae, Gasterosteidae or Percidae. At higher taxonomical levels, the helminth fauna of Canadian fish is similar to that of fish in other parts of the world. There are no reasons to believe that the results of the present study would not apply more broadly to all parasites of fish, in all geographical areas. Our findings are also consistent with those of the only other study to examine the relationship between host specificity and the probability of discovering a parasite species. Krasnov *et al.* (unpublished data) found that among fleas ectoparasitic on small mammals, host specificity was a better predictor of year of description than various host attributes. In addition, as in our study on fish parasites, the number of host species used by fleas appeared to be more important than the taxonomic distinctness of these hosts (Krasnov *et al.*, unpublished data). Therefore, the link between host specificity and the probability of finding parasite species may apply more generally to all parasite taxa.

Biodiversity surveys and species description are human endeavours, and we might expect a 'human factor' to influence the rate and timing of species discovery. Indeed, the variance in year of description left unexplained in our analyses may reflect the temporal changes in the activity of leading parasite taxonomists. For instance, in the early 1930s, J. F. Mueller and H. J. Van Cleave obtained a large sample of fish from Lake Oneida, Ontario, Canada. From their dissections of these fish, they recovered numerous parasite species, including many new ones that they proceeded to describe. As a result, 17 helminth species in our dataset, i.e. 10% of the total number of species, have been described by these two parasitologists over a couple of years in the early 1930s (Van Cleave & Mueller, 1932, 1934; Mueller, 1934*a, b*). These 17 species were described at that time mainly because of the concentrated efforts of these two researchers, and not because of the parasite's biological characteristics such as host specificity. The human element in the discovery of new species can produce irregularities in rates of species description. Other factors can also explain some of the variance in date of description. For instance, parasite body size may not matter as much as it does for ectoparasites (Poulin, 1996), but it can still determine whether a parasite will be found or not, especially in the case of small parasites occupying hidden locations (e.g. intestinal villi). In addition, the body size or general commercial importance of the fish hosts used by a parasite may also be important, as certain fish species are more likely than others to be surveyed for parasites. Finally, the prevalence and abundance of the different parasite species may also

affect their probability of discovery: the more common parasites should be found earlier. Epidemiological variables such as prevalence and abundance, however, vary greatly among host populations and species, and are also not readily available for a large-scale analysis such as the present one. All these factors merely add noise to the deeper trend we observed. Our results indicate that other influences are just superimposed over a larger pattern: as a rule, generalist parasite species are found earlier than specialist species.

An important finding of our study is that the estimate of average host specificity, based on all parasite species known at a given point in time, changes over the years. As more and more parasite species become known to science, their average host specificity increases, i.e. both the mean number of known host species per parasite species, and their mean  $S_{TD}$ , decrease over time. This is the direct result of the more host-specific parasite species being found later than the generalist species. As a consequence, the average number of host species exploited by parasites of fish described before 1800 was 8 host species; this value dropped to about 6 for all parasites known before 1900, and to 4 for all parasites known before 1950. Clearly, our perspective on the specificity of parasites may have changed over the years, because of our biased knowledge of the extant parasite fauna. By extrapolation, we can predict that the helminth parasite species not yet known are probably very host-specific, and that the true average host specificity of all extant parasites is lower than the one we can estimate at present.

An analysis like the one we performed obviously depends on the quality of the data. We have made significant efforts to account for all cases of synonymy and to include information from taxonomic revisions completed well after species discovery. It remains possible that the known host specificity of some parasite species is inflated, for instance in cases where a parasite species may in fact consist of a complex of cryptic species that can only be disentangled by using molecular genetic analyses. Still, there is no conceivable reason why such species complexes would be more common among recently described species than among species known for a century or more. Similarly, our measure of host taxonomic diversity is based on taxonomy and not on phylogeny. If they were known and available, using the true relationships among host species would provide a more accurate assessment of the influence of this aspect of host specificity on the probability of parasite discovery. Still, there is no reason to believe that the results we obtained using our 'rough'  $S_{TD}$  index do not reflect an underlying pattern. Therefore, minor errors or inaccuracies can generate noise but are extremely unlikely to generate the relationships we observed. The conclusion still stands: host specificity influences the probability that

a parasite species will be found and described. Our knowledge of parasite biodiversity is thus biased toward generalist species: there are probably more generalist species in our list of known parasites than among all extant parasites. The good news is that the bias is getting smaller over time.

## REFERENCES

- ALLSOP, P. G. (1997). Probability of describing an Australian scarab beetle: influence of body size and distribution. *Journal of Biogeography* **24**, 717–724.
- BLACKBURN, T. M. & GASTON, K. J. (1995). What determines the probability of discovering a species? A study of South American oscine passerine birds. *Journal of Biogeography* **22**, 7–14.
- COLLEN, B., PURVIS, A. & GITTLEMAN, J. L. (2004). Biological correlates of description date in carnivores and primates. *Global Ecology and Biogeography* **13**, 459–467.
- GASTON, K. J. (1991). Body size and probability of description: the beetle fauna of Britain. *Ecological Entomology* **16**, 505–508.
- GASTON, K. J. & BLACKBURN, T. M. (1994). Are newly described bird species small bodied? *Biodiversity Letters* **2**, 16–20.
- GASTON, K. J., BLACKBURN, T. M. & LODER, N. (1995). Which species are described first? The case of North American Butterflies. *Biodiversity and Conservation* **4**, 119–127.
- MARGOLIS, L. & ARTHUR, J. R. (1979). *Synopsis of the Parasites of Fishes of Canada*. Bulletin of the Fisheries Board of Canada **199**, Ottawa, Canada.
- MCDONALD, T. E. & MARGOLIS, L. (1995). *Synopsis of the Parasites of Fishes of Canada: Supplement (1978–1993)*. Special Publication of Fisheries and Aquatic Sciences **122**, National Research Council, Ottawa, Canada.
- MUELLER, J. F. (1934a). Two new trematodes from Oneida Lake fishes. *Transactions of the American Microscopical Society* **53**, 231–236.
- MUELLER, J. F. (1934b). Parasites of Oneida Lake fishes. Part IV. Additional notes on parasites of Oneida Lake fishes, including descriptions of new species. *Roosevelt Wild Life Annals* **3**, 335–373.
- NELSON, J. S. (1994). *Fishes of the World*, 3rd Edn. John Wiley & Sons, New York.
- POULIN, R. (1996). How many parasite species are there: are we close to answers? *International Journal for Parasitology* **26**, 1127–1129.
- POULIN, R. (2002). The evolution of monogenean diversity. *International Journal for Parasitology* **32**, 245–254.
- POULIN, R. & MORAND, S. (2004). *Parasite Biodiversity*. Smithsonian Books, Washington, D.C.
- POULIN, R. & MOUILLOT, D. (2003). Parasite specialization from a phylogenetic perspective: a new index of host specificity. *Parasitology* **126**, 473–480.
- REED, R. N. & BOBACK, S. M. (2002). Does body size predict dates of species description among North American and Australian reptiles and amphibians? *Global Ecology and Biogeography* **11**, 41–47.
- SCOTT, W. B. & CROSSMAN, E. J. (1973). *Freshwater Fishes of Canada*. Bulletin of the Fisheries Board of Canada **184**, Ottawa, Canada.

VAN CLEAVE, H. J. & MUELLER, J. F. (1932). Parasites of Oneida Lake fishes. Part I. Descriptions of new genera and new species. *Roosevelt Wild Life Annals* **3**, 1-72.

VAN CLEAVE, H. J. & MUELLER, J. F. (1934). Parasites of Oneida Lake fishes. Part III. A biological and ecological survey of the worm parasites. *Roosevelt Wild Life Annals* **3**, 161-334.