

## COMBINING PHYLOGENETIC AND ECOLOGICAL INFORMATION INTO A NEW INDEX OF HOST SPECIFICITY

R. Poulin and D. Mouillot\*

Department of Zoology, University of Otago, P.O. Box 56, Dunedin, New Zealand. e-mail: robert.poulin@stonebow.otago.ac.nz

**ABSTRACT:** Host specificity has 2 independent facets: the extent to which different host species are used by a parasite, and the phylogenetic distances among these hosts. Although the number of host species exploited by a parasite commonly is used as a measure of host specificity, it fails to capture ecological and phylogenetic differences among hosts. Here, a new index of host specificity,  $S_{TD}^*$ , is developed and illustrated. This index measures the average taxonomic distinctness among the host species used by a parasite, weighted for the parasite's prevalence in the different hosts. For a given number of host species, the index approaches its minimum value when a parasite achieves high prevalence in a few closely related host species, and the index approaches its highest value when a parasite reaches its highest prevalence values in distantly related host species. Simple hypothetical examples are used to demonstrate the index's computation and some of its properties. The new index is influenced independently both by the taxonomic (or phylogenetic) affinities of a set of host species and by the distribution of prevalence values among these hosts. A single value cannot truly capture all the nuances of a phenomenon as complex as host specificity; nevertheless, the proposed index incorporates the features of specificity that are most relevant to parasitologists and will be a useful tool for comparative studies.

The measurement of host specificity is of central importance to the study of parasite ecology and evolution (Adamson and Caira, 1994; Poulin, 1998). For example, the level of host specificity shown by parasites has profound implications for their population dynamics, their probability of local extinction, and the likelihood that they can switch to new host species (Bush and Kennedy, 1994; Paterson and Gray, 1997; Poulin, 1998). Thus, it is imperative that we use a proper measure of host specificity when comparing different parasite species if we aim to understand why they show different degrees of specialization. Traditionally, the number of host species used by a parasite species, or the host range (Lymbery, 1989), was used as a simple measure of host specificity. Although easy to compute, this rough index is inappropriate, because it involves the assumption that all host species used by a parasite are equal. In fact, the different host species exploited by a parasite differ on 2 important levels, and the mere number of host species used fails to capture these differences.

First, from an ecological perspective, some host species are used more intensely than others. That is, the prevalence, intensity, or abundance of infection by a particular parasite differs among its host species. Typically, the parasite will be highly prevalent in 1 host species but rare in others. It is possible to include ecological or epidemiological information in an index of host specificity. Rohde (1980, 1993) has developed indices of host specificity based on the relative intensity or prevalence of a parasite in its different host species; these indices provide a better idea of host specificity than the number of host species used provides, at least with respect to how evenly the parasite uses its host resources.

Second, from an evolutionary perspective, some of the host species used by a parasite are closely related phylogenetically, whereas others are distantly related. Taxonomic or phylogenetic data about the host species also carry information about host specificity. Consider 2 parasite species, each capable of infecting 4 host species. If the first parasite species infects 4 host species belonging to the same genus but the second parasite

species infects 4 host species belonging to different families, then surely we can say that the former is more host-specific than the latter, because it exploits a narrower taxonomic range of host species (Caira et al., 2003). Recently, following in the footsteps of Caira et al. (2003), Poulin and Mouillot (2003) proposed an index of host specificity that considers the taxonomic structure of the set of host species used by a parasite. Their index is derived from indices of biodiversity proposed in the context of conservation biology (e.g., Clarke and Warwick, 1998; Warwick and Clarke, 2001; Barker, 2002), and it measures the average taxonomic distance among all host species used by a parasite. This index provides new insight regarding host specificity and has great relevance for comparative studies focused on the evolutionary history of parasites (Poulin and Mouillot, 2003).

To date, and to our knowledge, no attempt has been made to integrate both ecological and phylogenetic information into a single, unified index of host specificity. Efforts to combine these different types of data into measures of biodiversity for use in conservation biology are progressing steadily (Ricotta, 2004), although the measures proposed to date are all inadequate for the determination of host specificity. Here, we present a new index of host specificity,  $S_{TD}^*$ , which achieves a simple, intuitive combination of information regarding the taxonomic or phylogenetic relationships among host species and the prevalence achieved by the parasite on these different hosts.

### THE NEW INDEX $S_{TD}^*$

The proposed specificity index,  $S_{TD}^*$ , measures the average taxonomic distinctness of all host species used by a parasite species, weighted by the prevalence of the parasite in these different hosts. The taxonomic side of the index is similar to an earlier index,  $S_{TD}$ , proposed by Poulin and Mouillot (2003). When the host species are placed within a taxonomic hierarchy based on the Linnean classification into phyla, classes, orders, families, genera, and species, the taxonomic distinctness between any 2 host species is the mean number of steps up the hierarchy that must be taken to reach a taxon common to both (Fig. 1). Thus, if 2 host species are congeners, then 1 step (species-to-genus) is necessary to reach a common node in the taxonomic tree; if the 2 species belong to different genera but

Received 13 May 2004; revised 22 August 2004; accepted 24 August 2004.

\* UMR CNRS-UMII 5119 Ecosystèmes Lagunaires, Université Montpellier II, CC 093, 34095 Montpellier cedex 5, France.

the same family, then 2 steps will be necessary (species-to-genus and genus-to-family); and so on. For any given host-species pair, the number of steps corresponds to half the path length connecting 2 species in the taxonomic tree, with equal step lengths (of 1) being postulated between each level in the taxonomic hierarchy.

Here, a brief mention of a previous index is necessary. The path length for all pairs of host species is simply a taxonomic distance, and we must point out that Rao (1982) already proposed a quadratic entropy index ( $Q$ ) expressing the average distance between 2 randomly selected individuals in a sample comprising  $S$  species:

$$Q = \sum_{i=1}^S \sum_{j=1}^S d_{ij} p_i p_j,$$

where  $d_{ij}$  is the distance between the 2 species  $i$  and  $j$  and where  $p_i$  and  $p_j$  are the proportions of individuals belonging to species  $i$  and  $j$ , respectively. However, prevalences in different host species cannot be compared to species proportions in a community, because the proportions must add up to 1 and the prevalences do not. In addition, Rao's (1982) index is dependent on species richness (the average distance between 2 randomly selected individuals increases with richness). Therefore, although we use the same kind of weighting method as Rao (1982), our new index must deal with prevalences and be independent from the number of host species used by a parasite to add orthogonal information.

We now return to the description of our index. Taxonomic distinctness is computed for all possible pairs of host species and then weighted by the product of the parasite's prevalence in each host species in a pair. Prevalence is defined here as the proportion of hosts harboring the parasite species. The weighting factor has a maximum value of 1 when the prevalence is 100% in both host species in a pair, and it converges toward 0 when the prevalence in the host species is very low. Thus, more weight is given to the taxonomic distance between 2 host species if the parasite achieves high prevalence in these hosts than is given if the parasite occurs infrequently in these hosts. This way, the average weighted taxonomic distinctness will reflect the taxonomic distances among the main hosts of the parasite more strongly. The index  $S_{TD}^*$  is the ratio of the sum of the weighted taxonomic distinctness values to the sum of the weighting factors. More formally, the index is computed as follows:

$$S_{TD}^* = \frac{\sum_{i < j} \omega_{ij}(p_i p_j)}{\sum_{i < j} (p_i p_j)},$$

where the double summation is over the set  $\{i = 1, \dots, s; j = 1, \dots, s, \text{ such that } i < j \text{ and } s \text{ is the number of host species used by the parasite}\}$ ;  $\omega_{ij}$  is the taxonomic distinctness between host species  $i$  and  $j$ , or the number of taxonomic steps required to reach a node common to both; and  $p_i$  and  $p_j$  are the prevalences of the parasite in host species  $i$  and  $j$ , respectively.

Figure 1 shows 4 hypothetical sets of host species for 4 different parasite species. In each case, the average (unweighted) taxonomic distance among host species is the same, and the prevalence values are the same. What differs is the taxonomic structure linking the host species, the distribution of prevalence

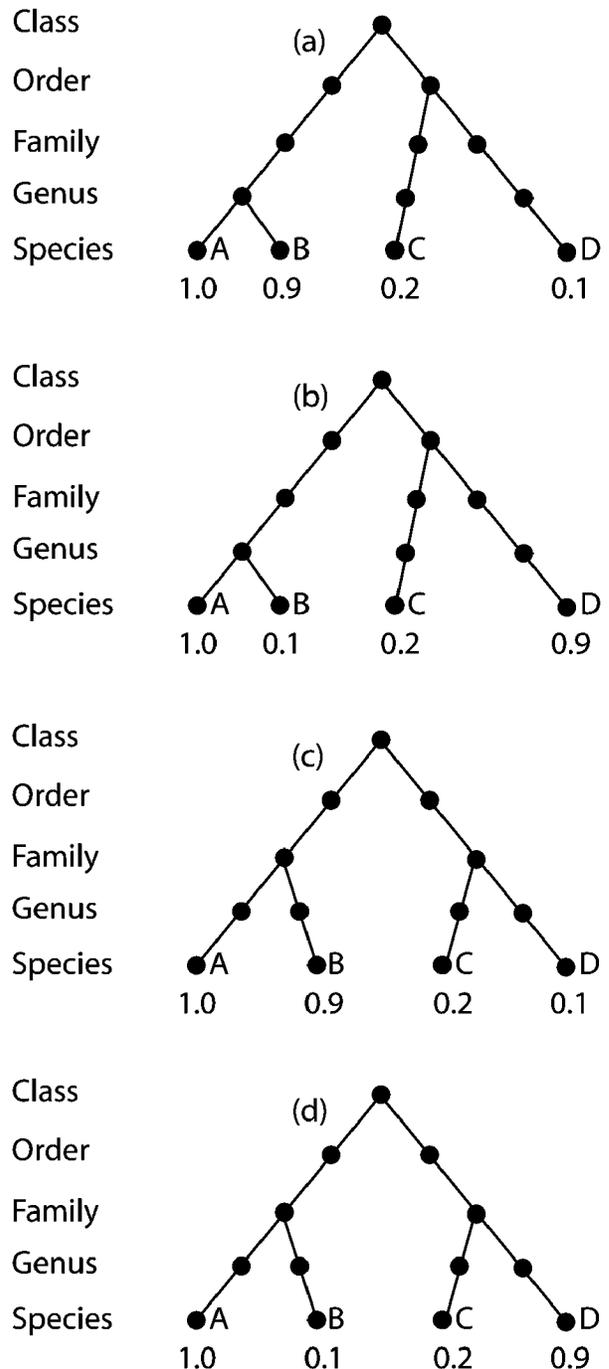


FIGURE 1. Taxonomic structure of the sets of hosts for 4 hypothetical parasites, with prevalence in each host indicated below. Four host species, A to D, appear in each example. Average (unweighted) taxonomic distinctness is the same (3.33) in all cases, as are the prevalence values. However, either the structure of the taxonomic tree or the distribution of prevalence values among host species changes from 1 example to the next. See Table I for computation of the index  $S_{TD}^*$ .

values among host species, or both (Fig. 1). A summary of the computations required to compute the index for each of these 4 hypothetical cases is shown in Table I. The value of  $S_{TD}^*$  increases as the taxonomic distinctness between the high-prevalence hosts increases. Therefore, the value of the index  $S_{TD}^*$

TABLE I. Summary of the computations required to obtain the index  $S_{TD}^*$  for the hypothetical examples in Figure 1.

Host pair	Unweighted distinctness [ $\omega_{ij}$ ]	Weighting factor [ $p_i p_j$ ]	Weighted distinctness [ $\omega_{ij} (p_i p_j)$ ]	$S_{TD}^*$
Example a				
A–B	1	0.90	0.90	2.17
A–C	4	0.20	0.80	
A–D	4	0.10	0.40	
B–C	4	0.18	0.72	
B–D	4	0.09	0.36	
C–D	3	0.02	0.06	
Total	20	1.49	3.24	
Example b				
A–B	1	0.10	0.10	3.68
A–C	4	0.20	0.80	
A–D	4	0.90	3.60	
B–C	4	0.02	0.08	
B–D	4	0.09	0.36	
C–D	3	0.18	0.54	
Total	20	1.49	5.48	
Example c				
A–B	2	0.90	1.80	2.77
A–C	4	0.20	0.80	
A–D	4	0.10	0.40	
B–C	4	0.18	0.72	
B–D	4	0.09	0.36	
C–D	2	0.02	0.04	
Total	20	1.49	4.12	
Example d				
A–B	2	0.10	0.20	3.62
A–C	4	0.20	0.80	
A–D	4	0.90	3.60	
B–C	4	0.02	0.08	
B–D	4	0.09	0.36	
C–D	2	0.18	0.36	
Total	20	1.49	5.40	

actually is inversely proportional to specificity. A high index value means that the host species most frequently used by a parasite are, on average, not closely related. Both ecological and phylogenetic data have separate and independent influences on the value of  $S_{TD}^*$ . For a given taxonomic structure among a set of host species, changing the distribution of prevalence values among them will affect  $S_{TD}^*$  (see examples a vs. b and c vs. d in Fig. 1 and Table I). Similarly, for a given distribution of prevalence values, altering the taxonomic structure of the set of host species also affects  $S_{TD}^*$  (compare examples a vs. c and b vs. d in Fig. 1 and Table I). In the examples used here, changing the taxonomic structure of the set of host species appears to have a lesser effect on  $S_{TD}^*$  compared to changing the distribution of prevalence values. This results only because of the nature of the examples chosen, however, and because the average (unweighted) taxonomic distance among host species was kept the same in all of them. More drastic changes to the taxonomic structure of the set of host species would have greater impact on  $S_{TD}^*$ .

We developed a computer program using Borland C++ Builder 6.0 to compute  $S_{TD}^*$  (available at [http://www.otago.ac.](http://www.otago.ac.nz/zoology/downloads/poulin/TaxoBiodiv2)

[nz/zoology/downloads/poulin/TaxoBiodiv2](http://www.otago.ac.nz/zoology/downloads/poulin/TaxoBiodiv2)). A guide accompanying the program explains how the data need to be formatted. All that is needed is a taxonomic hierarchy for the sets of host species used by each parasite species included and the prevalence of these parasites in each of their host species.

When prevalence values in all host species are the same, the index  $S_{TD}^*$  converges toward the earlier index  $S_{TD}$ , in which prevalence was not incorporated (Poulin and Mouillot, 2003). The range of possible values for  $S_{TD}^*$  also are the same as that for  $S_{TD}$ . Using the standard 5 taxonomic levels that occur 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 the minimum value (when all host species are congeners) is 1. Given that few (if any) parasite species at a given stage in their life cycle infect hosts belonging to different phyla, this range is sufficient for all practical purposes, and we recommend it as a standard procedure.

The only apparent weakness of  $S_{TD}^*$  is that it cannot be applied to parasite species infecting only a single host species. These cases have no pair of host species from which a taxonomic distance can be computed. Such highly host-specific parasites can be dealt with in 2 obvious ways. First, they can be excluded from any comparative analysis in which  $S_{TD}^*$  is computed for other, less specific parasite species. After all, they show maximum specificity, and the computation of an index will not tell us more about them. Second, they can be assigned a default  $S_{TD}^*$  value of 1 (or even 0), because by definition, “all” host species for such specialized parasites belong to the same genus.

## DISCUSSION

Advances in our understanding of parasite ecology and evolution require, among other things, that we properly measure the various parameters under scrutiny. Host specificity is a key property of parasites, and to elucidate how it has evolved, we first must be able to measure it. Most previous comparative studies of host specificity (e.g., Poulin, 1999; Barger and Esch, 2002) have used the number of host species exploited by a parasite as a measure of host specificity. Clearly, this simple number does not capture the differences that exist among host species with respect to how important they are to the parasite’s population dynamics or the taxonomic (phylogenetic) distances that the parasite had to “jump” to colonize all these hosts. A more recent study, which used a measure of host specificity that took into account the taxonomic affinities of host species but not the relative prevalence of the parasite among them, showed that using such an index leads to conclusions different from those based merely on the number of host species used (Poulin and Mouillot, 2004). Here, we propose a new, simple-to-compute index that encapsulates both the relevant ecological and phylogenetic information about the set of hosts used by a parasite. For a given number of host species, parasites that achieve high prevalence in few closely related host species are scored as more host specific (lower  $S_{TD}^*$  value) than those that achieve their highest prevalence values in distantly related host species.

Some features of the index require discussion. First, the index  $S_{TD}^*$  uses taxonomic information and not phylogenetic data, although the latter no doubt are preferable. The concept of taxonomic distinctness can easily be used in a full phylogenetic

context by replacing taxonomic trees, such as those in Figure 1, with proper phylogenetic trees having known branch lengths with only a slight modification to the computations involved (e.g., Clarke and Warwick, 2001). Indeed, one need only use actual branch lengths, based on molecular data, in our formula instead of the arbitrary taxonomic distances we postulate. However, given that we do not yet have fully resolved phylogenies for all host taxa and are unlikely to have them within the next several years, it probably is more practical to stay within a taxonomic framework, at least at present.

Second, the index  $S_{TD}^*$  uses data regarding prevalence rather than other measures of infection level, such as intensity or abundance. For some types of parasites, the latter 2 measures may be better descriptors of parasite epidemiology and infection level. Prevalence, however, usually is strongly correlated with intensity and abundance (e.g., Shaw and Dobson, 1995; Morand and Guégan, 2000). Using in our index either intensity or abundance as a weighting factor instead of prevalence would, therefore, make little difference. In addition, for many types of parasites, prevalence often is the only meaningful measure of infection level, or the only measure that can be obtained. This is true of many parasites that multiply rapidly within the host, such as protozoans in vertebrates or larval trematodes in snails, and of parasites that occur at such high intensities that counts of parasites per host are logistically impossible. Prevalence also is the only variable available when indirect evidence of infection is used to assess host specificity, such as when antibody titers or fecal egg counts are used instead of actual parasite numbers. Thus, the use of prevalence in our index not only is justified but also broadens the index's potential application.

Third, although the index  $S_{TD}^*$  combines data regarding patterns of host specificity on 2 scales (i.e., host use in ecological time and host colonization over evolutionary time), it does not incorporate information regarding the smallest of scales. Within the host, different parasite species may be specialized for different microhabitats, such as different parts of the gastrointestinal tract. It is difficult to imagine how a general index could capture this level of specificity and still remain useful for many kinds of parasites living in different organs that are not easily comparable.

Fourth, unlike the quadratic entropy proposed by Rao (1982), our new index does not have a clear combinatorial meaning, because the prevalence values are not normalized to a probability space (i.e., their sum is not always 1). Nevertheless, from a biological perspective, this does not affect the index's potential usefulness.

In summary, the new index  $S_{TD}^*$  integrates distinct ecological and phylogenetic information about the host species used by a parasite to capture, in a single number, the main components of host specificity. Clearly, a single number cannot express all the nuances associated with a phenomenon as complex as host specificity. Nevertheless, we are confident that the aspects of host specificity most relevant to parasitologists are incorporated

into our index, and we especially recommend its use for comparative purposes.

## ACKNOWLEDGMENTS

R.P. was supported by a James Cook Research Fellowship from the Royal Society of New Zealand during the preparation of this paper.

## LITERATURE CITED

- ADAMSON, M. L., AND J. N. CAIRA. 1994. Evolutionary factors influencing the nature of parasite specificity. *Parasitology* **109** (Suppl.): S85–S95.
- BARGER, M. A., AND G. W. ESCH. 2002. Host specificity and the distribution—abundance relationship in a community of parasites infecting fishes in streams of North Carolina. *Journal of Parasitology* **88**: 446–453.
- BARKER, G. M. 2002. Phylogenetic diversity: A quantitative framework for measurement of priority and achievement in biodiversity conservation. *Biological Journal of the Linnean Society* **76**: 165–194.
- BUSH, A. O., AND C. R. KENNEDY. 1994. Host fragmentation and helminth parasites: Hedging your bets against extinction. *International Journal for Parasitology* **24**: 1333–1343.
- CAIRA, J. N., K. JENSEN, AND K. E. HOLSINGER. 2003. On a new index of host specificity. In *Taxonomy, ecology and evolution of metazoan parasites*, Vol. 1, C. Combes and J. Jourdan (eds.). Presses Universitaires de Perpignan, Perpignan, France, p. 161–201.
- CLARKE, K. R., AND R. M. WARWICK. 1998. A taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology* **35**: 523–531.
- , AND ———. 2001. A further biodiversity index applicable to species lists: Variation in taxonomic distinctness. *Marine Ecology Progress Series* **216**: 265–278.
- LYMBERY, A. J. 1989. Host specificity, host range, and host preference. *Parasitology Today* **5**: 298.
- MORAND, S., AND J.-F. GUÉGAN. 2000. Distribution and abundance of parasite nematodes: Ecological specialization, phylogenetic constraint or simply epidemiology? *Oikos* **88**: 563–573.
- PATERSON, A. M., AND R. D. GRAY. 1997. Host-parasite cospeciation, host switching, and missing the boat. In *Host-parasite evolution: General principles and avian models*, D. H. Clayton and J. Moore (eds.). Oxford University Press, Oxford, U.K., p. 236–250.
- POULIN, R. 1998. *Evolutionary ecology of parasites: From individuals to communities*. Chapman and Hall, London, U.K., 212 p.
- . 1999. The intra- and interspecific relationships between abundance and distribution in helminth parasites of birds. *Journal of Animal Ecology* **68**: 719–725.
- , AND D. MOUILLOT. 2003. Parasite specialization from a phylogenetic perspective: A new index of host specificity. *Parasitology* **126**: 473–480.
- , AND ———. 2004. The relationship between specialization and local abundance: The case of helminth parasites of birds. *Oecologia* **140**: 372–378.
- RAO, C. R. 1982. Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology* **21**: 24–43.
- RICOTTA, C. 2004. A parametric diversity measure combining the relative abundances and taxonomic distinctiveness of species. *Diversity and Distributions* **10**: 143–146.
- ROHDE, K. 1980. Host specificity indices of parasites and their application. *Experientia* **36**: 1369–1371.
- . 1993. *Ecology of marine parasites*, 2nd ed. CAB International, Wallingford, U.K., 298 p.
- SHAW, D. J., AND A. P. DOBSON. 1995. Patterns of macroparasite abundance and aggregation in wildlife populations: A quantitative review. *Parasitology* **111** (Suppl.): S111–S133.
- WARWICK, R. M., AND K. R. CLARKE. 2001. Practical measures of marine biodiversity based on relatedness of species. *Oceanography and Marine Biology Annual Review* **39**: 207–231.