

# Parasite richness/Sampling effort/Host range: The Fancy Three-piece Jigsaw Puzzle

J.-F. Guégan and C.R. Kennedy

*In this article, Jean-François Guégan and Clive Kennedy propose an alternative explanation for the confounding effects of host geographical range and sampling effort on parasite species richness using pathway analysis procedure. They suggest that much of the species richness revealed by sampling effort is also a reflection of host range. Thus, the total contribution of host range logically incorporates a contribution from sampling effort. The implications of indirect effects of host range on richness estimates have not previously been discussed, and the authors here attempt to redress the balance. The contribution of host range to richness, as derived from control of sampling effort on richness estimates, therefore, is a mathematical expression that does not take into account the cause-and-effect nature of things.*

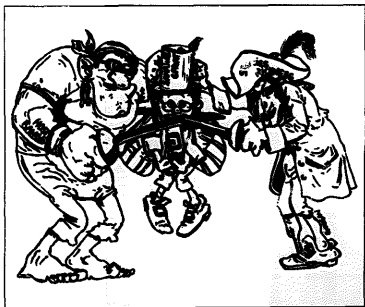
The question of parasite species richness among taxa being confounded by uneven sampling effort has recently been addressed<sup>1,2</sup>. Several contributions have assessed the relationship between regional parasite species number and host geographical range<sup>3-5</sup> (larger area sampled = more species found), but in reality very few have attempted to control for the perverse influence exerted by uneven sampling effort on richness estimates. Despite the fact that, without doubt, some sampling artefacts may considerably affect parasite richness estimates, we wish to discuss a different way of thinking about uneven sampling bias control in parasite community analyses. Usually, two possibilities have been considered to explain the correlation of host geographical area with parasite species richness<sup>1</sup>: (1) that there is a positive direct relationship between the two variables; and (2) that the relationship observed is a consequence of a co-varying factor, i.e. that parasite species richness may arise because larger areas have been more intensively sampled.

## Helminth species richness in British freshwater fish

The number of helminth parasite species found among species of British freshwater fish is highly correlated with the number of parasite surveys made per host species<sup>1,2</sup> (Fig. 1). When plotting the host geographical range versus parasite richness values<sup>1</sup> on the scatter diagram of the sampling effort versus richness relationship<sup>1</sup>, we observe a strong similarity between the two curves. High correlations observed between species richness, sampling effort and area must result from strong connections among these three variables. In addition, we have used Monte Carlo simulations<sup>6,7</sup> to estimate the probability of the observed parasite assemblage per host species under the null hypothesis of random assortment of parasites. They show that the species-area curve pattern does

not fit a nested species design (nestedness index (NI) = 1110, RANDOM1, 100 simulations,  $p = 0.670$ , RANDOM1, 1000 simulations,  $p = 0.625$ ). If specialists were excluded, then the distribution of parasites yielded similar results (NI = 756, RANDOM1, 100 simulations,  $p = 0.100$ , RANDOM1, 1000 simulations,  $p = 0.090$ ). Parasite composition among British freshwater fish shows a random distribution of species that differed from a nested species subset pattern (Fig. 2). The use of Wilcoxon's statistics<sup>9</sup> for testing the conformity of observed distribution to a perfect theoretical nestedness gave similar results ( $|z_{\text{obs}}| = 4.704$ ,  $p < 0.0001$  with all parasites, and  $|z_{\text{obs}}| = 4.625$ ,  $p < 0.0001$  when specialists were withdrawn) (NB  $z$  is a parameter of Wilcoxon's<sup>9</sup>; RANDOM1 is the Monte Carlo simulation program used to test for randomness). Clearly, this positive host geographic area-parasite species relationship<sup>1</sup> does not hold within the context of the island biogeographic theory<sup>10</sup> for which the equilibrium number is the result of a balance between the rate of immigration and the rate of extinction<sup>11</sup>. Thus, when regarding these fish host species as 'islands', we bias the results because each host species may be considered as a separate 'province' (slope coefficient of the relationship = 0.51, when species-area data are fitted to the power function) (see Refs 9, 10) with basic distinct parasite evolutionary histories and assemblages even if one can find some generalist parasites that inhabit more than one 'province'.

To date, contributions have considered independently that sampling effort may affect parasite richness estimates regardless of the methodology used, and that area may encompass the ranges of parasite species<sup>1,2</sup>. Thus, the possibility that sampling bias may confound the relationship between parasites and host range has been assessed using residual values derived from regressions. We have illustrated causal links between variables conventionally used (Fig. 3a). The subject of the three-way relationship between



Jean-François Guégan is at ORSTOM, BP 145, 97323 Cayenne cedex French Guiana. Clive Kennedy is at the Department of Biological Sciences, Prince of Wales Road, Exeter University, Exeter, UK EX4 4PS. Tel: +594 29 92 71, Fax: +594 31 98 55, e-mail: guegan@orstom.fr

area, sampling effort and species number has remained unexplored in parasite community ecology, and we propose a different model to establish the relationship among the three distinct variables. Interestingly, both host range and sampling effort contribute to the variation in parasite species number in a multiple regression procedure (species richness =  $-0.70 + 0.21 \times$  host range +  $0.55 \times$  number of surveys,  $R^2_{\text{adjusted}} = 0.77$ ,  $F$  test = 54.51,  $p < 0.0001$ , all terms logarithmically scaled; untransformed data yielded similar results) ( $R$  is the proportion of the species variation due to both the host range and the number of surveys done; for definition of  $F$  test, see Ref. 12).

### Caution

We must be cautious in interpreting results generated from multiple regression techniques due possibly to multi-collinearity among predictor variables, and because they provide an estimate of only the direct effect of correlations. We thus used path analysis procedures<sup>12</sup> to assess both direct and indirect effects of causal variables, assuming that variables measured early in the procedure may have causal effects on features characterized later; an assumption that cannot be handled by multiple regression techniques (see Fig. 3). The usefulness of the method in distinguishing the role of intercorrelated variables on a response variable can be seen in many contributions<sup>13-17</sup>. The proposed causal relationships among the three variables are illustrated in Fig. 3b.

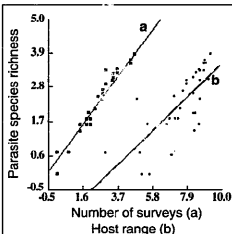


Fig. 1. Plots of helminth species richness in British freshwater fish as a function of number of surveys (a) and of host range (b) for 32 host species examined. Data show that the two curves are parallel ( $p < 0.05$ ). As host range (in km squares) increases, so proportionately does the number of surveys leading, in both cases, to higher species richness. The two axes are scaled logarithmically. No transformation of axes yields similar results. Data are from Ref. 4, and J.-F. Guégan, unpublished. Redrawn, with permission, from Ref. 4.

Path analysis reveals significant direct and indirect effects of both host range and sampling effort predictor variables on richness estimates (Table 1). In the two models we generated, the host range variable had little direct effect on richness, influencing instead the sampling effort variable, which subsequently affects richness estimates. Direct effects of host range were more than half the value of direct effects of sampling effort, but host range had a greater total (direct and indirect) effect. The implications of such indirect effects of host range, producing a larger total causal effect on richness estimates, have not previously been addressed.

Although parasite species richness varies strongly across different sampling efforts, we show here that host range exerts more than half (0.451 in Model 1, 0.584 in Model 2) (Table 1) its overall effect (0.762 in Model 1, 0.864 in Model 2) on richness through a powerful effect on sampling effort. We suggest that much of the contribution to parasite richness per host species made by sampling effort is therefore a reflection of host range, and the total contribution of host range logically incorporates a contribution from sampling effort.

The conventional approach and the one we propose here are conceptually different in that the former pre-judges the major importance of sampling effort on richness compared to any lesser predictor variables, whereas the latter incorporates sampling effort into a pathway model without any pre-assumption on its role and its importance on the response variable. In our study, host range and sampling effort correlate so closely, that it is only one source variable that is functionally useful: the effort to sample an area is simply a substitute.

### Open questions

Parasite richness among separate host species seems to arise not because larger host geographical ranges have been more intensively sampled, but just because some areas are larger and so require more surveys. Can path analysis procedure be applied to other measures of sampling effort, i.e. the time spent searching, the number or size of localities visited, the number of microhabitats, the number of collecting trips, the number of individuals examined? What can happen to the total effect of host range when the slope of the species richness-sampling effort relationship is different from that of the species richness-area relationship? Since more surveys are

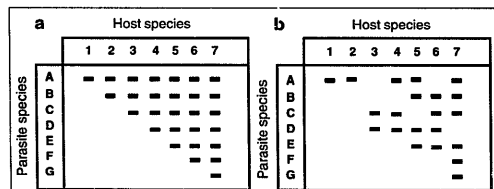


Fig. 2. Schematic representation of presence/absence of parasite species across host species. Generally, free-living island communities in an archipelago show a nested subset pattern (see Ref. 6). A completely parasite nested design (a) means that a parasite species found on one host with  $n$  parasites will be found on all host species with  $n + 1$  species (see Ref. 8 for a parasitological illustration). When such a hierarchical organization is not found, then the parasite composition deviates from a nested structure (b). The deviation is greater when such absences are higher. The Monte Carlo simulation RANDOM1 is based on the occurrence of species in the parasite pool, and thus it does not consider differences in probability distributions in hosts between specialists and generalists (see Ref. 7). The use of Wilcoxon's statistics which compare the observed distribution of parasite richness across host ranges (ranked from lower to higher values), to an hypothetical nested distribution produced similar results (see Ref. 9).

needed to census larger areas, more samples are taken to represent larger surveys. What can the introduction of a new source variable, i.e. number of individuals collected, into the path diagram produce on results? Does the uneven sampling bias control made on richness estimates correspond to an extra unnecessary control that serves first to control the sampling effort surrogate variable against parasite richness, and then to correlate richness residuals obtained from these regressions against the host range primary source variable? More generally, do we regrettably have to refute Island Biogeography theory and any other species-area curves, for the simple reason that sampling effort has probably biased all studies made on it? On the other hand, do we have to consider that surveys made on larger ranges implicitly need proportionally more hands for collecting species, but that this collaboration between many researchers of different institutions and countries over decades and centuries is absolutely useful to census all the biodiversity on Earth, and thus must be considered as a single meta-social work?

We open the debate, but we think that a discussion of the different conclusions from the various methods will be interesting and valuable. Undoubtedly, to solve this three-piece jigsaw puzzle, there is an absolute requirement for more published data!

#### Acknowledgements

We thank Claude Combes, Serge Morand, B. Walther and Bernard Huguency for helpful comments and discussions. JFG is supported by ORSTOM (Paris and Cayenne) and CRK by the University of Exeter.

#### References

- Gregory, R.D. (1990) Parasites and host geographic range as illustrated by waterfowl. *Funct. Ecol.* 4, 645-654
- Walther, B.A. et al. (1995) Sampling effort and parasite species richness. *Parasitol. Today* 11, 306-310
- Dritschilo, W. et al. (1975) Insular biogeography: of mice and mites. *Science* 190, 467-469
- Price, P.W. and Clancy, K.M. (1983) Patterns in number of helminth parasite species in freshwater fishes. *J. Parasitol.* 69, 449-454
- Price, P.W., Westoby, M. and Rice, B. (1988) Parasite-mediated competition: some predictions and tests. *Am. Nat.* 131, 544-555
- Patterson, B.D. and Atmar, W. (1986) Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biol. J. Linn. Soc.* 28, 65-82
- Andrén, H. (1994) Can one use nested subset pattern to reject the random sample hypothesis? Examples from boreal bird communities. *Oikos* 70, 489-491
- Guégan, J.F. and Huguency, B. (1994) A nested parasite species

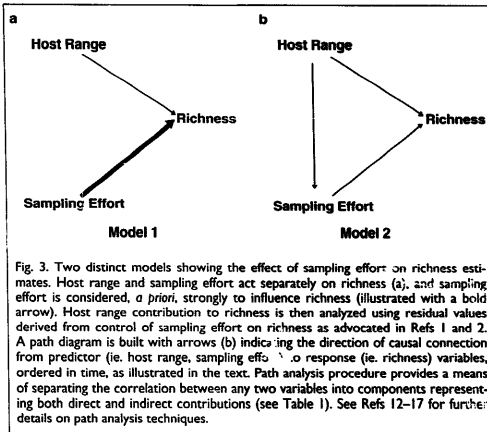


Fig. 3. Two distinct models showing the effect of sampling effort on richness estimates. Host range and sampling effort act separately on richness (a), and sampling effort is considered, *a priori*, strongly to influence richness (illustrated with a bold arrow). Host range contribution to richness is then analyzed using residual values derived from control of sampling effort on richness as advocated in Refs 1 and 2. A path diagram is built with arrows (b) indicating the direction of causal connection from predictor (i.e. host range, sampling effort) to response (i.e. richness) variables, ordered in time, as illustrated in the text. Path analysis procedure provides a means of separating the correlation between any two variables into components representing both direct and indirect contributions (see Table 1). See Refs 12-17 for further details on path analysis techniques.

Table 1. Path coefficients and effect coefficients for the path diagram illustrated in Fig. 3b

	Model 1	Model 2
<b>Path coefficients*:</b>		
HR effect on SE (direct)	$= a_1$	$= 0.702^{\dagger}$
SE effect on RSR (direct)	$= b_1$	$= 0.642^{\dagger}$
HR effect on RSR (direct)	$= b_2$	$= 0.311^{\ddagger}$
HR effect on RSR (indirect)	$= a_1 \times b_1$	$= 0.451$
<b>Effect coefficients:</b>		
HR effect on RSR (total)	$c_1 = b_2 + (a_1 \times b_1)$	$= 0.762$
SE effect on RSR (total)	$c_2 = b_1$	$= 0.642$
		$0.703$
		$0.280^{\ddagger}$
		$0.584$

\* All estimates are given in standardized form. Indirect causal effects on the response richness variable are calculated from the early pathway host range predictor variable via the sampling effort subsequent variable. Total causal influence sums all direct and indirect pathways. As path analysis assumes linear relations among variables, all three variables were log-transformed (Model 1) and kept untransformed (Model 2) for comparison, which yields similar results. HR, host species range; SE, sampling effort; RSR, regional helminth species richness.

<sup>†</sup> $p < 0.001$ .

<sup>‡</sup> $p < 0.001$ .

- subset pattern in tropical fish: host as major determinant of parasite infracommunity structure. *Oecologia* 100, 184-189
- Simberloff, D.S. and Martin, J.L. (1991) Nestedness of insular avifaunas, simple summary statistics masking complex species patterns. *Ornis. Fenn.* 68, 178-192
- MacArthur, R.H. and Wilson, E.O. (1967) *The Theory of Island Biogeography*. Princeton University Press
- Simberloff, D.S. (1974) Equilibrium theory of island biogeography and ecology. *Annu. Rev. Ecol. Syst.* 5, 161-182
- Sokal, R.R. and Rohlf, F.J. (1981) *Biometry*, W.H. Freeman & Co.
- Schemske, D.W. and Horvitz, C.C. (1988) Plant-animal interactions and fruit production in a neotropical herb: a path analysis. *Ecology* 69, 1128-1137
- Vitchell-Olds, T. and Bergelson, J. (1990) Statistical genetics of an annual plant, *Impatiens capensis*. II. Natural selection. *Genetics* 124, 417-421
- Kohn, D.D. and Walsh, D.M. (1994) Plant species richness - the effect of island size and habitat diversity. *J. Ecol.* 82, 367-377
- Kingsolver, J.G. and Schemske, D.W. (1991) Path analysis of selection. *Trends Ecol. Evol.* 6, 276-280
- Sheldon, A.L. and Meffe, G.K. (1995) Path analysis of collective properties and habitat relationships of fish assemblages in coastal plain streams. *Can. J. Fish. Aquat. Sci.* 52, 23-33