O P I O N
  P I   I O N
    I N I
    I   I
  P   O
O   N

Community nestedness and the proper way to assess statistical significance by Monte-Carlo tests: some comments on Worthen and Rohde's (1996) paper

Bernard Hugueny, ORSTOM, Laboratoire d'Ecologie des Eaux Douces et des Grands Fleuves, Université Lyon I, F-69622 Villeurbanne cedex, France (hugueny@biomserv.univ-lyon1.fr) and Jean-François Guégan, ORSTOM/ U.M.R. 5554-5 du C.N.R.S., Station Méditerranéenne de l'Environnement Littoral, Université de Montpellier II, 1 quai de la Daurade, F-34200 Sète, France.

In 1994, Guégan and Hugueny found that the infra-communities of parasites on the African cyprinid fish *Labeo coubie* exhibited significant nested-subset structure and were not random assemblages (Guégan and Hugueny 1994). In a recent analysis of our data, Worthen and Rohde (1996) reached the opposite conclusion. These authors suggested that we assessed the statistical significance of our Monte-Carlo test in an inappropriate way. However, in the light of recent reviews on the use of computer-intensive statistical methods in biology (Manly 1991, Crowley 1992, Potvin and Roff 1993), it seems to us that our approach is absolutely correct whereas the procedure advocated by Patterson and Atmar (1986) and used in most of the studies dealing with community nestedness (including that of Worthen and Rohde 1996) incorporates some drawbacks. Our note is thus mainly a discussion about the proper way to assess statistical significance by Monte-Carlo tests, a point we believe to be of concern for many community ecologists.

Community nestedness was first formulated by Patterson and Atmar (1986) to describe a particular non-random pattern of species occurrence along a species richness gradient of insular communities. A completely nested design occurs when species that compose a depauperate island community constitute a proper subset of those inhabiting richer islands. As a result, a set of such island communities arranged by species richness presents a nested series. Patterson and Atmar (1986) devised an index of nestedness (N) to quantify this pattern. This index equals zero when the presence/absence matrix is perfectly nested, and it grows larger as nestedness decreases. The statistical significance of the observed value of N is then tested by Monte-Carlo

simulations. Since pioneering works in the 1970s, Monte-Carlo methods have become common tools for community ecologists. The use of these methods has been facilitated by the availability of fast micro-computers because they need computer simulations to mimic the stochastic process assumed to have generated the data under a specific null hypothesis. This increasing importance of computation-intensive statistical methods (randomization tests, Monte-Carlo methods, bootstraping procedures) in biology has largely motivated recent reviews emphasizing the versatility of these methods due to their distribution-free nature (Manly 1991, Crowley 1992, Potvin and Roff 1993). However, the procedure used by Patterson and Atmar (1986) and subsequent workers, such as Worthen and Rohde (1996), does not take this non-parametric property into account. These authors developed computer programs to generate mean and variance of N by Monte-Carlo simulations under two different null hypotheses of species distribution within localities (RANDOM0 and RANDOM1). Under the RANDOM0 hypothesis, each species has the same probability to be drawn during the construction of a simulated community whereas under RANDOM1 each species has a probability to be drawn weighted by its observed occurrence. When the simulations are completed, the statistical significance of the observed value is evaluated by comparing the z-score (the difference between the observed value and the simulated mean value divided by the simulated standard deviation) with a standard normal distribution. Thus, it is assumed that N is normally distributed under the null hypothesis, which represents a conventional parametric approach. Empirical evidences suggest that, under the two currently used null hypotheses, N some-

times presents a skewed distribution (Wright and Reeves 1992, Worthen and Rohde 1996) and it is, at best, approximately normally distributed. These conditions lead to a biased estimation of p, the statistical significance, that does not converge toward the true value with increasing number of simulations. Moreover, the calculation of z-score is based on estimated mean and variance and, thus, it is exposed to an unknown amount of error even if the assumption of normality is met. As pointed out by Wright and Reeves (1992), N has a discontinuous distribution and a continuity correction should be integrated in the calculation of the z-score.

The procedure used by Worthen and Rohde (1996) following Patterson and Atmar (1986) differs actually from the correct way to assess statistical significance by Monte-Carlo tests or other computer intensive methods (see Manly 1991, for instance). If we want to test whether the observed N is lower than expected only by chance, the appropriate procedure is indeed very simple. At each simulation, we check if the simulated value is lower than or equal to the observed one. If it is, then the simulation is tallied. The p estimate is the proportion of simulated values lower than or equal to the observed value. No assumption has to be made on the statistical distribution (distribution-free test), and the estimated p-value is given unbiased with a direct known error value. A 95% confidence interval for p is $p \pm 1.96\sqrt{(p(1-p)/n)}$, where n is the number of simulations (Manly 1991). The statistical error is thus a decreasing function of the number of simulations (n). For testing nestedness, no continuity correction is needed because, in this case, Monte-Carlo simulations generate discrete distributions. We thus recommend the use of the procedure described above because it ensures unbiased p-values whatever the underlying distribution, and it permits adjustment of the number of simulations according to the precision fixed by the user.

It should be noted to the credit of Worthen and Rohde (1996) that they were aware of some of the problems raised by the use of z-scores. For instance, they noted, within their sample of 102 parasite communities, the occurrence of highly skewed distributions of simulation scores. This resulted in 95% confidence intervals for N in which zero was included. In this case, under the normality assumption, communities that have a perfect nestedness score ($N = 0$) cannot be distinguished from randomly constructed communities. Consequently, only 38 of the 102 communities have been included in the analyses made by Worthen and Rohde (1996). However it must be emphasized that all these 102 communities could have been analyzed with a distribution-free method. Studies on nestedness within parasite communities (Guégan and Hugueny 1994, Poulin 1996, Worthen and Rohde 1996) have opened the way for interesting comparisons between extinction- and colonization-dominated community. As pointed

out by Worthen and Rohde (1996), their study is one of the largest investigations of nestedness in a colonization-dominated system in which they conducted 38 statistical tests of parasite communities nestedness for fish host species from different habitats and with distinct life-histories. This made their study an ideal candidate for the use of methods allowing a quantitative and simultaneous analysis of several tests of the same hypothesis, such as meta-analysis (Arnqvist and Wooster 1995) or Fisher's combined probabilities (Sokal and Rohlf 1981). These methods are based on a combination of probabilities of individual tests. Consequently, the results must be treated with caution if systematically biased probabilities are included, as is possible if z-scores are used.

Spurred by Worthen and Rohde's (1996) re-analysis of our data set (Guégan and Hugueny 1994), we re-examined that data set ourselves and found a computational error. In our initial analysis, when computing the N-score, we erroneously included absences from communities with richness values greater than or equal to the most depauperate community in which the focal species was found; rather than only summing absences from communities with richness values greater than the most depauperate community containing that species. Including these 'tied' values resulted in a nestedness score which is not strictly comparable with Patterson and Atmar's N. The correct nestedness value for *Labeo coubie* is 53, not 85 as reported in our 1994 paper. The adequately revised Monte-Carlo test leads to a highly significant result (none of the 1000 simulated values, under the RANDOM1 hypothesis, was lesser than or equal to the observed one, $p < 0.001$) thus confirming our previous findings. The conclusions we made in our original study thus still hold, and we invite the readers to refer to our paper for further details. Nevertheless, we would add that a recent work (Andrén 1994) casts some doubts on our previous assertion that the observed nested pattern was not likely to result from random sampling because it differed from RANDOM1 simulations. The random sample hypothesis predicts, for statistical reasons, that only abundant species are to be found within small samples (those with a low total number of individuals), whereas rare species will mainly occur within large samples. Andrén (1994) pointed out that random samples of species with different abundances might produce a nested subset pattern which differs from RANDOM1 simulations. Thus, RANDOM1 does not seem to be a good mimic of random sampling. Ideally, the random sample hypothesis should be rejected before testing for community nestedness (see Worthen et al. 1996). However, in order to reject this hypothesis it is necessary to have information about species population sizes in the source pool. Unfortunately, these data are unavailable in most studies, including ours (Guégan and Hugueny 1994). Another shortcoming of the RANDOM1 algorithm is that

the incidence totals of matrices it generates do not approximate those of the observed matrix (Wright and Reeves 1992, Hugueny pers. obs.). Obviously, improvements of Monte Carlo algorithms are needed in this field but, whatever the algorithm used, our message is that the use of z-scores should be prohibited.

References

Andrén, H. 1994. Can one use nested subset pattern to reject the random sample hypothesis? Examples from boreal bird communities. – Oikos 70: 489–491.

Arnqvist, G. and Wooster, D. 1995. Meta-analysis: synthesizing research findings in ecology and evolution. – Trends Ecol. Evol. 10: 236–240.

Crowley, P. H. 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. – Annu. Rev. Ecol. Syst. 23: 405–447.

Guégan, J.-F. and Hugueny, B. 1994. A nested parasite species subset pattern in tropical fish: host as major determinant of parasite infracommunity structure. – Oecologia 100: 184–189.

Manly, B. F. J. 1991. Randomization and Monte Carlo methods in biology. – Chapman and Hall, London.

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.

Potvin, C. and Roff, D. A. 1993. Distribution-free and robust statistical methods: viable alternatives to parametric statistics? – Ecology 74: 1617–1628.

Poulin, R. 1996. Richness, nestedness, and randomness in parasite infracommunity structure. – Oecologia 105: 545–551.

Sokal, R. R. and Rohlf, F. J. 1981. Biometry, 2nd ed. – Freeman, San Francisco.

Worthen, W. B. and Rohde, K. 1996. Nested subset analyses of colonization-dominated communities: metazoan ectoparasites of marine fishes. – Oikos 75: 471–478.

– , Carswell, M. L. and Kelly, K. A. 1996. Nested subset structure of larval mycophagous fly assemblages: nestedness in a non-island system. – Oecologia 107: 257–264.

Wright, D. H. and Reeves, J. H. 1992. On the meaning and measurement of nestedness of species assemblages. – Oecologia 92: 416–428.