Ecology, 91(9), 2010, pp. 2568–2571

In the dragon's den: a response to the meta-analysis forum contributions

Robert J. Whittaker[1]

Biodiversity Research Group, Oxford University Centre for the Environment, South Parks Road, Oxford OX1 3QY United Kingdom and *Centre for Macroecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark*

Secondary analysis of previously published data has a long tradition in ecological science and is widely and successfully practiced as a means of efficiently addressing new questions and hypotheses. Meta-analysis is, in essence, the class of such analyses in which the findings of multiple primary studies are subject to further statistical analysis of emergent outcomes, and is a more recent practice within ecology. I recognize that this is a loose definition of meta-analysis (Ellison 2010, Gurevitch and Mengersen 2010) but continue to refer to the studies I critique using this common broader usage. Owing to the apparent power of such synthetic analyses, meta-analysis papers can be highly influential (Mittelbach 2010). This forum, together with other recent critical assessments (e.g., Englund et al. 1999, Gates 2002), demonstrates that there are good reasons to call for great care, improved rigor and transparency in the use of "meta-analysis" tools in ecology. However, in the article that initiated this forum exchange (Whittaker 2010), all the specific criticisms I made were restricted to recent meta-analyses of just one problem, which concerns the form of the species richness–productivity relationship (SRPR) in plants. In this brief response to the seven other contributions, I retain this focus while aiming to resolve several misconstructions of points made in my paper, and to comment on a few key points of disagreement regarding analyses of the SRPR.

Use of proxies.—First, there have been relatively few studies that have specifically set out to gather data to determine the form of the SRPR and so, in order to increase the power of analysis and refine the questions asked, those undertaking meta-analyses have sought other data sets that were initially gathered for different purposes. There are many published papers providing diversity data, but few that provide direct measurements of productivity, which is a difficult property to estimate accurately. Hence the reliance in Mittelbach et al. (2001), Pärtel et al. (2007), and Laanisto et al. (2008) on the use of proxies such as rainfall, vegetation height, biomass, etc., in order to generate surrogate productivity data for their analyses of the SRPR. Unfortunately, nonlinearities in relationships between actual productivity and the productivity proxies used in these analyses have the potential to result in misclassification of the form of the SRPR (see: Whittaker and Heegaard 2003, Gillman and Wright 2006, 2010, Huston and Wolverton 2009). By reference to details drawn from the original source papers and the wider literature, I have argued that this seriously undermines the analyses (Whittaker 2010: Appendix A). This was only one of a number of reasons leading me to stress the necessity of screening data sets for fitness-for-purpose prior to inclusion in analysis.

Criteria for selecting data sets.—Other contributors to the forum regard my criteria for including a data set in an SRPR meta-analysis as too limiting. For instance, Lajeunesse (2010) argues that "Erroneous elimination of a prohibitive number of studies is not a solution to handling variation due to study 'quality'..." Instead, we should "... gather all studies relevant to the conceptual topic under study, and then empirically test whether these differences ... actually influence research outcomes." However, the published SRPR meta-analyses demonstrate that different authors have adopted very different views of the "relevance" of an original study. Mittelbach et al. (2001) developed and reported a search strategy based on key words, e.g., a paper would have been screened if it had "species richness" in the key words but then rejected if it turned out there were no data *they* felt able to use as productivity proxies. They also eliminated studies of systems subject to severe anthropogenic disturbance, etc. By contrast, Pärtel et al. (2007) and Laanisto et al. (2008) did not reveal their criteria, and used many studies, that are not "relevant to the conceptual topic" and which in my view do not provide *suitable* data, free from confounding problems such as anthropogenic manipulation. It is thus of little practical help to say that we should use "all relevant studies" and then see if the (very many) factors identified as problematic have a statistical influence: the meta-analyst has first to decide and justify *which* are relevant (Gates 2002). Similarly, to provide a specific example, it

Manuscript received 10 September 2009; revised 19 November 2009; accepted 16 December 2009. Corresponding Editor: D. R Strong. For reprints of this Forum, see footnote 1, p. 2534.

[1] E-mail: robert.whittaker@ouce.ox.ac.uk

is no answer to the serious lack of standardization of sampling in Beadle's (1966) data set to say, as do Pärtel et al. (2010: Appendix A), that because Beadle saw fit to plot a regression line (actually it appears to be hand fitted) through a set of values, it is therefore safe to use for this new purpose. I therefore reiterate the view that a key initial step in meta-analysis should be to develop, articulate and apply a set of criteria for determining the studies that are to be included in the analysis.

I recognize that other ecologists may accept the need to have stated criteria while disagreeing with the particular set that I put forward for future use (e.g., see Hillebrand and Cardinale 2010). Here, I aim to clarify some of my choices regarding data set eligibility criteria. Criterion 1: I did not state that analysis of diversity indices are wrong, merely that alpha diversity indices provide different response variables, distinct from species richness, and that different response variables should be analyzed in separate (meta)analyses. Criterion 2: I may not have worded this clearly enough. My argument is that for a particular data set to be included in the meta-analysis, the plots reported *in that data set* should be of a fixed size (I suggested within $\pm 10\%$, but with very small plots within $\pm 5\%$). Holding plot size constant is necessary when sampling plant species richness because increasing the contiguous sample area from a small plot size to increasingly large plot sizes inevitably involves a stepped pattern of increased richness with area: failure to hold plot size constant within a data set means that area confounds analysis. I should emphasize that the criteria under discussion here are those I suggest for screening data for inclusion. In my article I also emphasized the need to organize the meta-analysis step with reference to scale of the study systems, but this is a separate step from screening individual studies for eligibility. Criterion 5 states that data sets involving other prominent confounding variables should be screened out, and I gave the examples of mowing, grazing, horticulture, or burning. An alternative to removing such studies is to examine whether the variable has explanatory power for the form of the SRPR. This approach has been adopted, for example, in examining the role of mesh size in meta-analyses of stream predation experiments (Englund et al. 1999). This may be tractable in systems in which there is a general consistency of approach and a limiting number of fairly obvious confounding factors. The difficulty presented in meta-analyses of the SRPR in plants, is that there appear to be so many potential confounding factors in the data sets gathered, that it becomes analytically intractable to deal with all of them at the formal meta-analysis step. Criterion 6 is the imposition of a minimum number of data points (within a particular study data set) for inclusion in these meta-analyses. Hillebrand and Cardinale (2010) argue that imposition of a 10-data-point minimum requirement is arbitrary and unnecessarily restrictive. Perhaps they are right, but the data sets in these analyses are noisy, productivity proxies are problematic, confounding variables are rarely entirely out of the equation, and inclusion of four- or five-point data sets in analyses testing between humped and linear fits seems risky in this context. This is why both Mittelbach et al. (2001) and Gillman and Wright (2006) used this criterion in their meta-analyses of the SRPR: I merely adopted their suggestion. Hillebrand and Cardinale sum up that my suitability criteria are "very arbitrary" but disappointingly do not provide their own alternative, less arbitrary set.

The logic of collapsing categories.—The first of the SRPR meta-analyses, by Mittelbach et al. (2001), set out to classify each SRPR as one of (1) positive linear, (2) humped, (3) negative linear, (4) U-shaped, and (5) unclassifiable, basing their decisions on standardized statistical procedures. In their analyses, Pärtel et al. (2007) collapse these categories. They assign U-shaped SRPR (which are very rare) to unclassifiable, on the grounds that they cannot see how to theorize a u-shaped relationship. They assign negative SRPR to humped SRPR on the basis that studies returning negative SRPR have probably not sampled environments of sufficiently low productivity to reveal the initial rising limb of what they theorize to be the real hump-shaped form. This sampling bias hypothesis is not a supportable generalization based on the source literature I have examined (Whittaker 2010: Appendix A). Moreover, if this logic is deemed acceptable, then why should we not convert positive linear relationships to humped relationships? Here the logic would be that systems showing positive linear relationships must merely have failed to sample high enough productivities to display the downwards part of the curve. This is as inherently plausible an argument as that concerning negative relationships, and, like that argument, may well apply in some cases (but in *which* and *how many* cases is unknowable). As these two arguments are logically equivalent, accepting one implies accepting the other, meaning that if statistical analysis reveals any one of forms 1, 2, or 3 (positive, humped, or negative), they should be (re-)classified as a humped SRPR; while other studies would be deemed to belong to the "no relationship" group. The humped SRPR then becomes general (the proposition Mittelbach et al. 2001 set out to test), but by proclamation rather than by statistical analysis. To pursue such arguments is to allow our beliefs about the likely true form of an unsampled portion of a relationship to hold sway over statistical analysis carried out within the empirical range of the study systems we have analyzed. This is unwarranted and, if undertaken, may easily be misunderstood by readers.

Agreements and disagreements on the detail.—As Gillman and Wright (2010) point out, we concur in most matters and there is a strong measure of agreement between our decisions on the form of the SRPR (but for differences, see Whittaker 2010: Appendix A case studies 10, 106/108, 131, 147, 151, 152, and 157). I thank them for pointing out my error in incorrectly transcribing

Mittelbach et al.'s classification of the study by Wheeler and Shaw (1991), although it remains the case that I differ from Gillman and Wright (2006) in regarding it as more likely a negative rather than U-shaped relationship. Bear in mind that my classification is based solely on reading the source paper and visual examination of the data set, not on new statistical analyses. In this instance, the data set includes a lot of scatter and has been variously regarded as negative by the original authors (and by me), humped (because negative is taken to mean humped) by Pärtel et al. (2007), and U-shaped by Mittelbach et al. (2001) and Gillman and Wright (2006). The limited consensus in this case is merely that the relationship is not positive.

Notwithstanding the concerns I have raised, Pärtel et al. (2010) repeat their claims to have demonstrated a tropical vs. temperate difference in the form of the productivity–diversity relationship, strongly implying that causation of this difference is related to species pool size. Here, indeed, be dragons. Their responses, especially as set out in their Appendix, provide revealing insights into the hitherto unstated criteria used by this team of authors and serve to illustrate that their data base cannot withstand forensic scrutiny. I find little scope for a more positive assessment of their findings in the light of this defense and recommend that any interested readers call up the source papers from online journal resources and archives, to evaluate how these data have been used in each meta-analysis.

Whither meta-analyses of the SRPR.—Those reviewing the first draft of this manuscript questioned whether it was productive to continue debating perceived flaws in the treatment of the SRPR. I sympathize with this perspective, but have invested in doing so because meta-analyses tend to carry influence and to become highly cited: they shape understanding and opinion disproportionately. So, for example, Oberle et al. (2009:6–7) comment that "... Recent work has shown that in herbaceous plant communities, clonal species may dominate high-productivity environments, increasing the prevalence of hump-shaped SRPRs, while this trait is less common among woody growth forms, resulting in more monotonic SRPRs ..." In support of this statement they cite the paper by Laanisto et al. (2008), which itself is a reworking and extension of the Pärtel et al. (2007) data base. I submit that while the foregoing statement could be correct at some scale of analysis, no such inference can reliably be based upon this particular source (see Oberle et al. [2009] for further discussion).

In some respects, I think we are seeing a classic trade-off here. In regular empirical papers, the reader gets to see the details of the sampling regime, study site, and key assumptions and can readily assess the strength of the inferences drawn. Such studies have value, but on their own provide singular cases that may not be representative. Meta-analyses (including quantitative data synthesis papers that are not technically meta-analyses), undoubtedly have greater agency (influence) than most primary data papers, but the properties of the underlying data are less easy for the reader to detect and scrutinize. This means, as other forum contributors argue, that it is doubly important that all key assumptions and analytical steps are clearly stated and that the meta-data are treated with great care by the meta-analysts (Gates 2002). The challenges involved for those involved in meta-analysis preparation and review are thus—like the influence such papers may have—disproportionate.

The scale issue.—I concur with a great deal of Mittelbach's (2010) thoughtful essay, although we continue to differ in our perspectives regarding scale, wherein I place relatively greater emphasis on the focal scale of analysis as an organizing principal in meta-analysis. On this issue, Mittelbach (2010) cites a specific study based on two data sets, demonstrating scale-invariance in the shape of the SRPR over a focal scale range of 10 m^2 to 200 m^2. However, as he recognizes, we cannot know if that scale-invariance would continue outside this empirical range, or for other systems. Other studies discussed by Whittaker (2010) do show (focal-)scale dependency. Notwithstanding our differences of perspective, I concur with Mittelbach's comments on Chase and Leibold (2002) as, in this instance, change in the form of the SRPR did not arise from changing plot sizes but rather from aggregating sites. This indicates that such changes in form can arise in studies of varying data structure, a common component being that as focal scale changes different diversity components are implicated.

Notwithstanding the concerns I have expressed about many of the case studies (Whittaker 2010: Appendix A), there appear to be sufficient recent empirical studies of robust design, to allow us to conclude that for a particular place and study system extent, the form of the SRPR can and frequently does change from linear to unimodal or vice versa with changing focal scale of analysis (Whittaker 2010). This means, I suggest, that we cannot view *any* study based on *a single focal scale* of analysis as adequately characterizing the general form of the relationship for that *place*, *system*, or *extent*. At finer or coarser focal scales it is quite likely that the system will have a different form of SRPR. Thus, all other issues aside, we cannot yet make any claim as to (e.g.) geographical differences in the form of the SRPR, without controlling in analysis for focal scale used. I suspect that this is a problem that has a wider relevance than yet realized in the quest for understanding geographical patterns of diversity.

Finally, I would like to make two points of clarification of the section *A few tasters* in Whittaker (2010), arising from correspondence following acceptance. First, I am grateful to L. N. Gillman and S. D. Wright for pointing out that Mittelbach et al. (2001) in fact classified the Wheeler and Shaw (1991) data set as U-shaped; hence, the reported relationships should read Wheeler and Shaw negative; GW2006 and M2001 U-shaped; P2007 humped. The inclusion of a humped relationship for M2001 in Appendix A (and Table A1) of Whittaker (2010) is thus in error. Second, regarding the Wardle et al. (1997) study, the comment about island area variation warrants further exposition; although the Shannon-Weiner diversity index data were collected from standard-sized plots, these plots are derived from islands of strongly contrasting size, and within the source paper it is demonstrated that island area is a strong determinant of environmental and ecosystem (including long-term successional) dynamics, thus confounding the interpretation of causal relationships.

Literature Cited

Beadle, N. C. W. 1966. Soil phosphate and its role in molding segments of the Australian flora and vegetation, with special reference to xeromorphy and sclerophylly. Ecology 47:992–1007.

Chase, J. M., and M. A. Leibold. 2002. Spatial scale dictates the productivity–biodiversity relationship. Nature 416:427–430.

Ellison, A. M. 2010. Repeatability and transparency in ecological research. Ecology 91:2536–2539.

Englund, G., O. Sarnell, and S. D. Cooper. 1999. The importance of data-selection criteria: meta-analyses of stream predation experiments. Ecology 80:1132–1141.

Gates, S. 2002. Review of methodology of quantitative reviews using meta-analysis in ecology. Journal of Animal Ecology 71:547–557.

Gillman, L. N., and S. D. Wright. 2006. The influence of productivity on the species richness of plants: a critical assessment. Ecology 87:1234–1243.

Gillman, L. N., and S. D. Wright. 2010. Mega mistakes in meta-analyses: devil in the detail. Ecology 91:2550–2552.

Gurevitch, J., and K. Mengersen. 2010. A statistical view of synthesizing patterns of species richness along productivity gradients: devils, forests, and trees. Ecology 91:2553–2560.

Hillebrand, H., and B. J. Cardinale. 2010. A critique for meta-analyses and the productivity–diversity relationship. Ecology 91:2545–2549.

Huston, M. A., and S. Wolverton. 2009. The global distribution of net primary production: resolving the paradox. Ecological Monographs 79:343–377.

Laanisto, L., P. Urbas, and M. Pärtel. 2008. Why does the unimodal species richness–productivity relationship not apply to woody species: a lack of clonality or a legacy of tropical evolutionary history? Global Ecology and Biogeography 17:320–326.

Lajeunesse, M. J. 2010. Achieving synthesis with meta-analysis by combining and comparing all available studies. Ecology 91:2561–2564.

Mittelbach, G. G. 2010. Understanding species richness–productivity relationships: the importance of meta-analyses. Ecology 91:2540–2544.

Mittelbach, G. G., C. F. Steiner, S. M. Scheiner, K. L. Gross, H. L. Reynolds, R. B. Waide, M. R. Willig, S. I. Dodson, and L. Gough. 2001. What is the observed relationship between species richness and productivity? Ecology 82:2381–2396.

Oberle, B., J. B. Grace, and J. M. Chase. 2009. Beneath the veil: plant growth form influences the strength of species richness–productivity relationships in forests. Global Ecology and Biogeography 18:416–425.

Pärtel, M., L. Laanisto, and M. Zobel. 2007. Contrasting plant productivity–diversity relationships across latitude: the role of evolutionary history. Ecology 88:1091–1097.

Pärtel, M., K. Zobel, L. Laanisto, R. Szava-Kovats, and M. Zobel. 2010. The productivity–diversity relationship: varying aims and approaches. Ecology 91:2565–2567.

Wheeler, B. D., and S. C. Shaw. 1991. Above-ground crop mass and species richness of the principal types of herbaceous rich-fen vegetation of lowland England and Wales. Journal of Ecology 79:285–301.

Whittaker, R. J. 2010. Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness–productivity relationship. Ecology 91:2522–2533.

Whittaker, R. J., and E. Heegaard. 2003. What is the observed relationship between species richness and productivity? Comment. Ecology 84:3384–3390.