CONCEPTS & SYNTHESIS
EMPHASIZING NEW IDEAS TO STIMULATE RESEARCH IN ECOLOGY

Ecology, 91(9), 2010, pp. 2522–2533

Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness–productivity relationship

ROBERT J. WHITTAKER[1]

Biodiversity Research Group, Oxford University Centre for the Environment, South Parks Road, Oxford OX1 3QY United Kingdom

Abstract. The form of the species richness–productivity relationship (SRPR) is both theoretically important and contentious. In an effort to distill general patterns, ecologists have undertaken meta-analyses, within which each SRPR data set is first classified into one of five alternative forms: positive, humped (unimodal), negative, U-shaped (unimodal), and no relationship. Herein, I first provide a critique of this approach, based on 68 plant data sets/studies used in three meta-analyses published in *Ecology*. The meta-analyses are shown to have resulted in highly divergent outcomes, inconsistent and often highly inappropriate classification of data sets, and the introduction and multiplication of errors from one meta-analysis to the next. I therefore call on the ecological community at large to adopt a far more rigorous and critical attitude to the use of meta-analysis. Second, I develop the argument that the literature on the SRPR continues to be bedeviled by a common failing to appreciate the fundamental importance of the scale of analysis, beginning with the confusion evident between concepts of grain, focus, and extent. I postulate that variation in the form of the SRPR at fine scales of analysis owes much to artifacts of the sampling regime adopted. An improved understanding may emerge from combining sampling theory with an understanding of the factors controlling the form of species abundance distributions and species accumulation curves.

Key words: confounding variables; diversity theory; meta-analysis; plants; primary productivity; scale.

INTRODUCTION

Species richness and productivity are two fundamental properties of (plant) ecological systems and the relationship between them has long been a subject of interest (e.g., Pianka 1966, Odum 1969). In experimental analyses using small plots the focus has sometimes been on how changing species richness changes system net primary productivity, but more usually the relationships is viewed, as herein, from the perspective of species richness as the dependent variable. The question that arose and which is at issue in the present paper is: what is the form of the species richness–productivity relationship (SRPR)? Is it (1) humped (unimodal), (2) U-shaped (negative unimodal), (3) positive monotonic, (4) negative monotonic, or is there (5) no relationship describable (i.e., neither linear nor unimodal)? The question is being asked because it is arguably fundamental to a mechanistic understanding of ecological diversity patterns (Whittaker et al. 2001) and because the relationship is poorly understood and contentious. The publication of a major meta-analysis of the SRPR including 121 plant data sets (90 of which are terrestrial systems, the rest aquatic) by Mittelbach et al. (2001) initially appeared to make an important contribution to understanding this problem, but closer examination revealed serious failings, leading Whittaker and Heegaard (2003) to call for the meta-analysis to be redone at consistent scales of analysis using more rigorous data-gathering and analytical protocols. I now realize that this call was a mistake on our part, because the data and protocols do not appear to exist to allow meaningful meta-analysis (cf. Slavin 1995). Three meta-analyses later, I now call for an end to meta-analyses of the SRPR, and a profound change in the criteria apparently being used by those undertaking, and reviewing submitted meta-analyses in ecology.

Subsequent to our critique and an accompanying defense by Mittelbach et al. (2003), Gillman and Wright (2006) responded to the challenge and reran a full meta-

---

Manuscript received 23 May 2008; revised 11 September 2008; accepted 8 December 2008. Corresponding Editor: D. R. Strong.

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

analysis for plants (terrestrial systems), adding a further 37 studies to those previously gathered by Mittelbach et al. (2001). Their analysis endorsed all the criticisms leveled by Whittaker and Heegaard (2003) and contrary to the claims of Mittelbach et al. (2003) that the original analysis was robust, obtained substantially different results. Gillman and Wright's (2006) paper is in substance a worthy and critical reanalysis, and it is thus with some regret that I note below errors of detail in their paper. Mittelbach et al. (2001) has turned out to be a significant paper, attracting over 300 citations thus far (ISI data), with remarkably few to date noting the existence of the two critical reanalyses (Table 1) or that the paper may be an unreliable analysis. Meanwhile, a third meta-analysis of the SRPR for plants, by Pärtel et al. (2007), has now been published, like each of the foregoing papers, in *Ecology*. Pärtel et al. (2007) claimed to build directly on the Mittelbach et al. (2001) data base, did not refer at all to Whittaker and Heegaard (2003) and side-stepped Gillman and Wright's (2006) damning reanalysis with a single "but see." As I show below, if you do take the trouble to "go and see," what you find is that none of the meta-analyses agree with one another on how to classify a large proportion of the data sets in their analyses, raising immediate concerns over the approach and doubts as to whether they constitute repeatable science.

The meta-analysis approach is supposed to provide an objective means of summing up the emergent outcome of numerous tests of the same thing (e.g., the effectiveness of a new medicine or treatment) by compiling the results of previously published analyses and objectively analyzing the distribution of the outcomes (Slavin 1995). Unfortunately, in many areas of ecology, sampling system and design properties are virtually unique from study to study, and potentially confounding factors abound. Moreover, the aims of the original studies have often been profoundly different from those of the meta-analyses, providing some form of data that may be scavenged and recycled, but not necessarily that are fit for purpose. Such problems affect other areas of science, including medicine (Slavin 1995), but I suspect may be particularly acute in ecology. In recent work on the SRPR, this has meant that some form of original analysis of the data sets (or something approximating the original data) has had to be undertaken case-by-case prior to assessing the emergent outcomes. The authors of the meta-analysis are not therefore objectively assessing objective tests of the SRPR made by *previous* authors: rather they are themselves undertaking extensive primary analyses in order first to classify each study before compiling the findings for meta-analysis.

Undertaking such analysis and interpreting the outcome requires careful exposition and discussion. Within the meta-analyses, perhaps as a result of journal restrictions on pagination, next to no space is given to the data properties of the source papers, appropriateness of the analyses, or contextualization of the end result (e.g., scrutinize Partël et al. 2007). It appears, moreover, that the "meta-" part of the analysis overwhelms the usual critical instincts of reviewers and readers who fail to dig into the underlying original case analyses. This has enabled, as I show here (see Appendix A) the passage of regrettable and often elementary sequences of errors, compounded from one meta-analysis to the next.

Table 1. Citations to the papers by Mittelbach et al. (2001), Whittaker and Heegaard (2003), and Gillman and Wright (2006), respectively M2001, WH2003, and GW2006, between 2003 and May 2008 inclusive, according to a search using ISI Web of Science on 9 September 2008.

Year	M2001	WH2003	GW2006
2003	50	1	NA
2004	45	2	NA
2005	53	14	NA
2006	57	10	1
2007	59	9	9
[2008]	[30]	[5]	[7]

Notes: Data for 2008 are given in square brackets as the year is incomplete. "NA" indicates not applicable.

All three of the meta-analyses contain error, although this is least apparent in the worthy attempt by Wright and Gillman (2006) to re-do using clear, stated criteria, the analysis for plants. The paper by Pärtel et al. (2007) provides no stated criteria or methods for the classification of studies and it attributes SRPR form inconsistently and inappropriately, often to fundamentally inadmissible data sets. These errors are repeated and compounded in a subsequent paper by the same team based on the same classification of studies (Laanisto et al. 2008), upon which I make little direct comment other than to regret its publication.

The primary purpose of the present paper is to ask that we draw a line under the whole approach. I anticipate that there may be some form of rejoinder published to this paper defending the meta-analytical approach, and so against that eventuality I ask that you, the reader, take the final call on whether the SRPR meta-analyses can be relied upon as repeatable ecological science. Colleagues, before you next cite their findings, please take the time to read this critical re-evaluation, and read at least some of the source papers (e.g., Beadle 1966, Flenley 1969, Wheeler and Giller 1982, Ehrman and Cocks 1990, Williams et al. 1996; and the following three papers of E. M. O'Brien's, which in fact present the same data set: O'Brien 1993, 1998, and O'Brien et al. 1998) in the light of the evidence and commentary I present in Appendix A. Having done so, why not set a class exercise for your students to read a few each, and then run an evaluation exercise with agreed criteria as to the attribution of SRPR form? See what *you* find. I predict that you will not wish to rely further upon the findings or meta-data presented in these meta-analyses and that it will lead into a wider discussion of the role of such analyses generally in ecology. My second goal is to develop the argument that the form of the SRPR is intrinsically scale-dependent and that much of the variation apparent at small focal scales of analysis constitutes an artifact of the use of inadequate plot sizes and protocols in the source literature.

CRITERIA FOR INCLUSION OF A STUDY IN A META-ANALYSIS OF THE SRPR (FOR PLANTS)

The start point for any meta-analysis has to be to establish a set of protocols for searching out case studies, criteria for including/excluding them, and adopting, a priori, a particular analytical strategy, statistical approach and probability level (Slavin 1995). Here I comment only on the criteria for including/excluding a data set (for discussion of the other issues see Whittaker and Heegaard 2003, Mittelbach et al. 2003, Gillman and Wright 2006). While Mittelbach et al. (2001) analyzed both plant and animal SRPR, the focus of the other meta-analyses and of this paper is entirely on plant data sets. I suggest that the following are reasonable and *necessary* criteria in order to include a data set in a meta-study of the SRPR for plants.

1) Data must be provided for plant species richness and must be complete and consistent within the source paper. (Other diversity metrics may be of interest to ecologists, but the response variable should be the same throughout, so papers reporting other alpha diversity indices should be placed in a separate analysis.)

2) Plot size (and sampling regime) must be held constant (I suggest within ±10%, but with very small plots within ±5%) to avoid sampling variation confounding the analysis.

3) An adequate measure or surrogate for productivity must be available, which does not hold the danger of distortion of the relationship at high or low values, where most cases of unimodality are detected.

4) The data distribution (and spatial structure of the sampling) should be consistent with the assumptions involved in the statistical tests employed (and—although this isn't a criterion of inclusion/exclusion of data—those tests should in turn be appropriate and robust).

5) The study design should not involve significant variation internally in potentially confounding variables of known or strongly suggested importance, and which have a strong likelihood of invalidating the analysis (e.g., including differential impacts of mowing, grazing, horticulture, or burning that are correlated with the "productivity" gradient).

6) As data sets consisting of a very small number of data points can be insufficient to capture the form of the SRPR reliably, a minimum qualifying number of plots should be set at the outset. Given that the goal is to discriminate linear from unimodal form, Gillman and Wright (2006) adopted a 10 data point minimum, which seems a reasonable (but admittedly arbitrary) minimum for present purposes, and which I endorse.

7) The same data points should not be included either wholly or in substance more than once. This may seem an obvious criterion, but it is one that needs careful checking given the habit in ecology of reanalyzing data sets to different ends in different papers.

(One reviewer commented about criterion 3 that it is surely necessary to standardize productivity measurements in regard to the division between aboveground and belowground productivity. This is a fair point. Most studies report only a measure of aboveground productivity, and while little is known about belowground productivity in many systems, there is reason to suspect that across some ecological clines, there can be strongly differential patterns of allocation switching between above- and belowground biomass [C. Girardin, *personal communication*]. While I have noted this point, I have not added it to the numbered list of criteria as I have not attempted to apply it herein.)

I recognize that these criteria are hard to meet, and that few studies are available that meet them (Mittelbach et al. 2003), but so be it. If the data aren't appropriate to meta-analysis, it is invalid to proceed with one. The solution is to read the literature, think about it, and do one of the following: (1) devise some critical experimental or other rigorous field study that will make a meaningful contribution to the question to hand, (2) undertake a narrative review, or (3) carry out what Slavin (1995) has termed "best evidence synthesis." For a description of what this final technique embodies, see Slavin (1995).

Of the three meta-analyses under consideration, Gillman and Wright (2006) have the most stringent and explicit criteria (with common elements to the seven I have listed), while Pärtel et al. (2007) have the least explicit and most liberal approach to inclusion of data sets. Unfortunately, all three meta-analyses include data sets that should have been excluded, in the case of Gillman and Wright (2006) and Pärtel et al. (2007) this even extends to accidentally including the same data set twice.

A CRITICAL AUDIT OF THE SRPR BASED ON 68 PLANT DATA SETS FROM THE META-ANALYSES

The method

I selected 68 studies previously classified in one or more of the meta-analyses for critical re-evaluation. First, I made use of the limited re-analysis and reclassification provided by Whittaker and Heegaard (2003; designated WH2003) of data sets for trees classed by Mittelbach et al. (2001) [designated M2001] as "regional" or "continental-global" in scale ($n = 12$). Second, I selected papers haphazardly from the Appendix of Pärtel et al. (2007; designated P2007), as this was the most recent of the meta-analyses. Initially, I focused on those SRPR classed by P2007 as unimodal (humped), as this was where the greatest problems were detected by Gillman and Wright (2006; designated GW2006). I extended the selection in order to ensure a reasonable representation of different SRPR forms as defined by P2007, subject to ease of retrieval of the article pdf. I continued collecting papers until I reached 68 studies,

Table 2. Summary of comparisons of the classification of the form of the species richness–productivity relationship (SRPR) across 68 data sets in the three published meta-analyses and in this paper.

Studies A vs. B	In study A only	In study B only	In A and B	Same result	Similar result	Different result	Percentage different
M vs. GW	0	5	35	5	5	25	71.4
M vs. P	6	30	30	18	0	12	40.0
M vs. RJW	0	30	34	5	1	28	82.4
GW vs. P	4	25	36	7	2	27	75.0
GW vs. RJW	0	23	39	29	3	7	18.0
P vs. RJW	0	8	57	10	4	43	75.4

Notes: Values in each column represent the number of studies, with the exception of the final column, which represents the percentage of those classified in both papers that are put into different classes. Note that each pair-wise comparison involves differing subsets of data sets and thus different total *n* values (column labeled "In A and B"). Key to studies: M, Mittelbach et al. (2001), consensus classification; GW, Gillman and Wright (2006); P, Pärtel et al. (2007); RJW, R. J. Whittaker (this paper). "Same result" indicates same classification in both papers; "Similar result" indicates broadly the same outcome, but e.g., described as uncertain, or being a relationship with biomass rather than productivity; "Differing result" indicates a differing classification; "Percentage different" is the percentage of those classified in both papers that are put into different classes.

which although fewer than compiled by M2001 (121 plant data sets), GW2006 (159), and P2007 (163), is sufficient to establish the consistency and reliability of the meta-analyses.

My approach to the re-evaluation took the form of two not entirely separable elements, first the application of the above criteria, and second, an evaluation based on the analyses and contextual information presented in the original source paper of the form of the SRPR. My method did not involve any statistical reanalysis, but took the form of scrutiny of the aims, methods, sampling strategy, results, and discussion of the original papers to determine the validity of the classification applied in each of the meta-analyses. I contend that as long as the evidential basis of this process is made transparent and explicit, this form of scrutiny of the internal consistency and ecological logic of each original analysis provides powerful evidence on which judgments can be taken. Accordingly, I provide as my evidence key details of the properties of each data set and how the SRPR was classified in the meta-analyses (see Appendix A). Of course, this is in essence just a first step. Note that many of the source papers either did not attempt to test the form of the SRPR or failed to carry out analyses that compared unimodal and linear models in a directly comparable fashion. To improve the power of my audit, it would be necessary to test model fits directly on the original data, using the stipulation adopted by M2001 that unimodal fits should only be accepted when the maximum (humped) or minimum (U-shaped) value in the fitted quadratic term falls within the empirical range of the observed data. There are other important issues (e.g., is the quadratic term in fact significant?) and both WH2003 and GW2006 have shown that the statistical procedure adopted for this evaluation is important. However, they have also shown that greater problems have arisen from failings of basic experimental design criteria, inappropriate treatment of surrogate productivity variables (which are often not fit for the purpose) and of ecological logic. If we can't be sure of the meaning and validity of the data being entered for statistical analysis, disputation over statistical protocols merely distracts attention from the really serious problems. In practice, it turns out that only a few of the data sets are fit for purpose (Appendix A).

The emergent outcome

Table 2 summarizes the outcome of the classification of the 68 data sets in the three meta-analyses and in my own audit. I should stress that it is tricky working out in some cases what particular data set is being referred to within P2007 and to a lesser degree in GW2006. Some data sets have been attributed to different source papers by different meta-analyses and in cases the same analysis has been included twice (below, Appendix A). Additionally, a few recent studies have provided analyses of the same system at multiple focal scales (e.g., Chase and Leibold 2002, Braschler et al. 2004, Chalcraft et al. 2004) and P2007 have been inconsistent in the number of "votes" assigned to these studies. Hence, as some papers provide multiple data sets, or multiple scales of analysis, and other data sets are included across the overall data base multiple times, the number of data sets could be deemed to be either more or less than 68.

The total number of data sets being directly compared varies between 30 (M2001 vs. P2007) and 57 (P2007 and RJW [this paper]) and the percentage of cases where the classification of the SRPR is different among the formal meta-analyses varies from 40% (M2001 vs. P2007) to 75% (GW2006 vs. P2007). Comparisons involving my own classification show that I largely concur with GW2006 (82% of cases), but I reject M2001's decisions in 82% of cases. In fact, in 17 of the 29 cases where GW2006 and RJW agree, we each classify the studies as inadmissible (i.e., invalid), meaning that we agree on only 12 meaningful classifications of SRPRs. From

TABLE 3. Summary of how analyses of the 68 data sets compare in their overall classification of studies, ignoring uncertain classifications, those deemed species richness–biomass relationships in Gillman and Wright (2006) and other such complexities (detailed in Appendix A).

Paper	Positive	Humped	Negative	U-shaped	Inadmissible
Original paper	8	12	3	0	NA
Mittelbach et al. (2001)	1	22	5	4	1
Gillman and Wright (2006)	6	5	0	2	21
Pärtel et al. (2007)	15	34	0	0	0
RJW (this paper)	5	7	3	0	35

Notes: Different subsets of the 68 data sets are included in each meta-analysis and my process of selecting studies may not have resulted in a representative subset of each meta-analysis. As a result, this table provides only a crude illustration of the way in which different approaches taken by each set of authors may have shaped the outcome of their analyses. Shape forms are as defined in *Introduction*; "Inadmissible" means failing the criteria outlined in the *Introduction* (i.e., invalid). NA = not applicable.

within 32 studies/data sets common to all three, the three meta-analyses concur in the classification of just three cases, make a very similar classification (i.e., basically opting for the same shape) in a further two cases, and disagree about the classification of 27 cases. So, at best, they agree on five studies, a meager 15% of the decisions. Of those five that have more-or-less agreed outcomes between the three meta-analyses, I dispute the classification of one more, meaning that across the four sets of authors, we have agreement on 4 studies (11%) out of 36 analyzed by each of us. You may as well classify the studies by random numbers. It is apparent that the meta-analyses of the SRPR provide no reproducible, objective basis for making any statement on emergent properties of the SRPR, how it varies with latitude (Pärtel et al. 2007), clonality of dominants (Laanisto et al. 2008), extent of study system (Mittelbach et al. 2001), and so on.

It is noteworthy that P2007 differ so much and analyze so many different data sets from M2001 because Pärtel et al. (2007) claim to have built their analyses largely on M2001 and because they provide no hint of how they classified the additional studies they included in their meta-analysis. Closer examination (Appendix A) suggests that their classification has also been influenced, to a limited degree, by GW2006 (and even by Whittaker and Heeagaard 2003). Part of the explanation for the difference in the classification of shared studies between P2007 and M2001 is that P2007 collapsed the initial five possibilities into three groups: (1) humped (including negative) SRPR, (2) positive SRPR, (3) no relationship (including U-shaped SRPR), of which more below. This resulted in four negative SRPR (M2001) being reclassified as humped SRPR (P2007). However, this collapsing of categories has not been carried out consistently. For instance, there are three cases where M2001 classified the SRPR as humped while P2007 didn't, and three of M2001's humps (each refuted by WH2003) were simply discarded from P2007's analysis. In addition, two of M2001's U-shaped relationships, instead of being reclassified by P2007 as "no relationship" were reclassified to a hump and a positive SRPR, respectively.

To begin to give some illustration of the breadth of the problems, taking the seven stated criteria listed above, in the Pärtel et al. (2007) paper: requirement 1 is broken in, e.g., case studies 9, 10, 11, 40, 129, 134, 135, 136, 137, 144, 145; requirement 2 in, e.g., cases 43, 66, 120, 129, 144; requirement 3 in, e.g., cases 8, 40, 46, 47, 62, 118, 120, 133; requirement 5 in, e.g., studies 8, 40, 51, 62, 91, 118, 133, 144, 146, 157; requirement 6 in, e.g., cases 62, 84, 91; requirement 7 in cases 106 and 108 (the exact same data set), and across the meta-analyses in other cases. Requirement 4 is also broken but is not formally demonstrated in my (nonstatistical) analysis other than by comparison across meta-analyses (see, e.g., study 147 in Appendix A). Appendix A demonstrates that there are more cases I could add to each list, those selected simply being "nice" examples. Unfortunately, very many of the above issues and examples apply also to the M2001 analysis, from which P2007 derived a large number of their classifications (Appendix A).

My sampling of the data sets used in the meta-analyses was not random in any formal sense but in a majority of cases I had not previously read the papers I selected for my reanalysis and so did not know in selecting them what I would find. In the earlier critique by Whittaker and Heegaard (2003) we provided a refutation of eight supposedly humped SRPR claimed by M2001, leading to the counter-charge that we were engaged in special pleading against humps (Mittelbach et al. 2003). So, I would like to emphasize that in this critique I do not simply dispute humped SRPR in this paper, and agree that they do occur (Table 3). However, it is apparent from close reading of the source material that both M2001 and P2007 are far too generous toward the notion of humped SRPR and far too liberal in assigning SRPR form without proper basis (Table 3, Appendix A). From the high level of erroneous and inconsistent treatments (both between and within meta-analyses) encountered for the 68 data sets examined, I anticipate that auditing of the remaining cases in the meta-analyses would reveal many additional errors and invalid classifications.

A few tasters

Limitations of space mean that I can provide just a few potted examples in the main text, as follows. Ehrman and Cocks (1990) provide data concerning the distribution of annual legumes in Syria, organized as a form of percentage incidence based on varying numbers of sites from 12 climate zones. The paper thus provides no proper species richness data, is focused only on a small taxonomic subset, lacks standardized sampling across the gradient, and includes in the study design confounding variables, but is classified using rainfall variation into humped SRPR by M2001 and P2007. It is clearly inadmissible. O'Brien's (1993) data set for southern African trees is included twice in both GW2006 and P2007. It is also wrongly classed as a humped SRPR by M2001, as shown by WH2003 and supported by GW2006 and P2007 (twice in each case!). Across the three meta-analyses the same data set is sourced to three original papers, and two different sets of meta-data are provided for this single richness vs. rainfall data set. Flenley's (1969) phytosociological study of the vegetation of the Wabag Hills (New Guinea highlands), including gardens, swamps, disturbed forest, and undisturbed forest is included (uniquely) in P2007 despite the obviously unsuitable nature of the "experimental design" of this study system and the absence of a meaningful productivity surrogate and inadequate size of the forest plots involved. Wardle et al. (1997) is classed as a humped SRPR by P2007, despite the fact that the *islands* concerned varied in area across two orders of magnitude, the source paper lacks species richness data and has no productivity data (stand biomass was used as surrogate). Wheeler and Shaw (1991) report a negative SRPR explaining 36% of the variation in a data set for herbaceous rich-fen vegetation from the United Kingdom. M2001 and P2007 regard it as a humped relationship, while GW2006 classify it as U-shaped, and claim incorrectly that M2001 did the same. But these descriptions really are just tasters. Please see the fuller accounts of all 68 studies in Appendix A, read the source papers, and judge for yourself.

Humps by proclamation?

Many of the studies included by Pärtel et al. (2007) were not conceived of as anything to do with the SRPR, and as several were not included in prior meta-analyses it is a mystery how the data were extracted, manipulated, analyzed and contextualized. As previously mentioned, however, Pärtel et al. are explicit that they started with five groups of relationships, which they then collapsed to three groups. First, "... the negative productivity–diversity relationship was merged with the unimodal relationship because most studies reporting a negative correlation focused on intermediate and high productivities" (Pärtel et al. 2007:1093). That is, they assume that all negative SRPR are merely incomplete humps in which the initial upward limb was (by design or accident) not sampled by the original authors. This is an extraordinary thing to do (1) as it invokes a complete reversal in the trend found in a data set by reference to no data at all, and (2) because the premise regarding productivity range is questionable (see, e.g., Appendix A, study 152: Wheeler and Shaw 1991). The merging of U-shaped relationships into the "no relationship" group on the grounds that U-shaped SRPR are theoretically implausible is also hard to justify (Gillman and Wright 2006), given that this is a mathematically equivalent form to their favored hump-shaped SRPRs. In fact, while they claim (Pärtel et al. 2007:1093) to have "used the earlier local and regional plant data from Mittelbach et al. (2001), but included additional studies (Appendix)," as shown above (and Appendix A) they have not simply incorporated the consensus decisions from M2001 but appear to have been influenced by decisions made both by M2001 and GW2006, while agreeing completely with neither. They have also, of course, added in further data sets scavenged from other source papers. Remarkably, I can find no trace in the paper or their appendix of the criteria and methods used to classify any of the SRPR in their meta-analysis. In addition, in at least two studies examined, it appears that an alpha diversity index was used instead of richness (Appendix A).

I fully accept that my own attempts to designate humps, U-shapes, and linear relationships in this article were based merely on visual examination of the source data and a reading of the source papers and whatever analyses they provide, but unlike M2001 and P2007, I explain the basis of my interpretation, I am explicit that the resulting designations are in cases highly uncertain, and I stress that they are not fit for summing for the purpose of further statistical analysis.

A final important point in attributing meaning to the SRPR is that having established, for instance, that a humped relationship is significant and explains more variation than a linear fit, if the overall variance explained is nonetheless very low, such that the majority of the variance in the data remains unexplained, this would suggests that something other than productivity is driving the system. Then the danger is that the apparent SRPR may in effect be an artifact of one or more other controlling factor(s): see Appendix A for discussion of a number of such cases.

WHY AND HOW FOCAL SCALE AND EXTENT ARE IMPORTANT ORGANIZING PRINCIPLES

It appears from the body of literature reviewed herein (i.e., the source papers as well as the recent meta-analyses) that the understanding of scale and its significance to the analysis of phenomena such as the SRPR is very uneven and incomplete among the ecological community. There are three relevant components: grain, focus, and extent (Whittaker et al. 2001, 2003). (1) The grain refers to the basic sampling unit (e.g., plot) used in collecting the data, which must be appropriate to the task. (2) The focal scale refers to the inference space used in analysis, either simply being to make use of the basic sampling unit (in which case focal scale and grain are identical), or it may refer to a coarser scale to which the basic data are aggregated prior to analysis. (3) The extent refers to the geographical area within which the entire data set is bounded. Grain and focal scale are true scale components, whereas extent is not: increasing extent is equivalent to unfolding a map sheet, gradually revealing more of the region at a consistent resolution.

Regardless of the underlying grain of the original data, it is the unit used in analysis (i.e., the focal scale) that must be the primary organizing principle when it comes to comparisons across (between) studies. This is because, first, the larger the space enclosed in a sample, the more individuals and the more species is it liable to contain. In relatively species-poor systems it is possible for the species accumulation curve (the "sampling curve") to level fairly quickly, indicating that a local community has been adequately sampled. But, with further expansion in plot area to incorporate differing habitat type(s) (beta diversity) or species pools from different source regions (~gamma diversity), species numbers rise again, producing either stepped or smoothly rising curves depending on the heterogeneity of the study system (e.g., see Cody 1975). Particularly if the study unit size (grain) corresponds with steep phases of the species accumulation curve, it is crucial to hold the sampling unit exactly constant in order to avoid sampling effects confounding the analysis, and consideration should be given to aggregating sets of nearby sites together into a consistent, but coarser focal scale to minimize the likelihood of noise or of systematic bias entering the analysis. Failure to hold focal scale constant within a particular data set fatally compromises analyses using species richness, perhaps the most scale-dependent of ecological response variables (Whittaker et al. 2001, 2003, Rahbek 2005).

As previously commented by Whittaker and Heegaard (2003) a key weakness of the meta-analytical design used by Mittelbach et al. (2001) was that having undertaken their initial classification of each SRPR, they then organized their analysis by grouping studies into extent classes, instead of by focal scale: an approach they subsequently defended. This is to mix up entirely dissimilar sets of relationships and, I argue, entirely confounds their analysis. This misconception of the scale problem is widespread in the SRPR literature. For instance, Schamp et al. (2003) implicitly accept this prioritization of extent over grain in their paper, describing their own study as a regional scale study. However, while the extent of the system is truly regional (spanning several hundred km across southern Ontario), the grain size and focal scale used in the analysis is 10 × 10 m plots. These are small plots for forest communities, which at best may capture the local diversity, or alpha diversity (sensu Whittaker 1977) of the stand. One consequence of the grouping of data sets by extent rather than grain, is that Mittelbach et al. (2001) and other authors following this rationale, are trying to find pattern across sites spanning several orders of magnitude of (focal) spatial scale. It is highly likely that the most general property of the SRPR is that its form will be found to change as the grain/focal scale of the study system is changed (Whittaker et al. 2001, Chase and Leibold 2002, Whittaker and Heegaard 2003), especially when dealing with small plots, as a difference between one square meter and a few tens of square meters will often be crucial to the form of the relationship while changing resolution from 10 000 km^2 to 25 000 km^2 may turn out to have trivial impact (cf. Gillman and Wright 2006).

Holding focal scale constant in analysis is also desirable because each data point in a SRPR is stable in both the dependent and independent variable. Imagine that we have 20 study sites each of 1 m^2 scattered across an area of 1 km^2. If the extent of the study system is increased to 10 km^2 to capture a greater range in environment, the original 20 data points will be afforced by additional data points but their productivity and richness values are unaltered. Altering extent while holding focal scale constant thus allows us to "fill in" the statistical distribution, and if we have indeed captured a greater range in environment, we may well add data points disproportionately at one or both "ends" of the distribution (i.e., very high or very low productivity), aiding in the discrimination of (and perhaps changing) the form of the SRPR but not altering in any way the values and structure of our initial 20 data points. Imagine instead that within our large study system extent of 10 km^2 we increase the size of each sample plot, beginning always with the same central location point, from 1 m^2 to 4 m^2 to 20 m^2 and so on, and what might happen? The richness of each plot either increases or remains constant with each increase in plot size (decrease being impossible given that each larger plot contains the previous smaller one), while the productivity value assigned to the site can increase, remain the same, or decrease. This is because, unlike richness, which is an additive variable in this context, values of productivity may be averaged across a site, and can be lower on average in a 20-m^2 area than within a particular 1-m^2 patch within that 20-m^2 space. In general, we should expect a reduction in range of values of productivity as we increase the focal scale of our 20 data points, providing of course that we do use a true average for estimating productivity and do not, for instance, simply rely upon the same clipped sample of aboveground biomass in one particular place within each site.

The instability of values of independent and dependent variables means that the form of the SRPR can change rapidly and profoundly with shifts in focal scale of analysis, particularly where starting with very small plots. The corollary of this is that where researchers have set out to study the SRPR and providing a sensible sampling strategy has been adopted, using a fixed-size analytical unit (focal scale) within a given study area (extent), I would predict that a robust and relatively stable form of SRPR can quite quickly be established, so that adding additional plots makes little impact on the relationship. Changes to the form of that relationship can be anticipated, however, if either the study system is expanded in extent to encompass higher or lower productivity areas outside the geographic bounds of the original data set, or if the sampling protocol is altered so that distinct and different habitat types are added to the data set within the same system (geographical) extent (cf. Nogués-Bravo et al. 2008).

The logical conclusions of this line of argument are that first, in order to establish how the SRPR changes with variation in the *range* of climate, or productivity, or between biogeographical regions and so forth, it is system sampling strategy and/or geographical extent that should be altered while focal scale must be held constant, and second, that whatever pattern is established in the analysis holds true only for the focal scale used in that analysis and cannot be generalized to different focal scales. Recent studies that have used data for the same system extents, but aggregated to different focal scales, have shown that this second conclusion, which is derivable from first principles of ecological science, is also empirically true: the form of the SRPR varies with focal scale (Chase and Leibold 2002, Braschler et al. 2004, Chalcraft et al. 2004). This finding opens a further challenge. My requirement 7 (data points should only be included once in the meta-analysis) is designed to avoid bias introduced by double-counting the same system. But, what should be done with data for the same system that have been analyzed at different focal scales? How should they be treated in a meta-analysis? If, for instance, the focal scale is changed trivially, and the form of the SRPR is humped at the two adjacent scales, should both counts be included? That would seem to constitute double counting. But what if a third much coarser focal scale is provided, and now the result is a positive SRPR. How should this system be entered into the meta-analysis? Should one scale or one form of SRPR have precedence over the other? And, if so, on what rationale? Those undertaking meta-analyses cannot simply ignore this question if they wish to claim objectivity and repeatability for their analysis. For just such a case, see studies 16 and 17 (Braschler et al. 2004), in Appendix A.

There is one further point to be made concerning system extent. While increasing the geographical extent of the study system can increase the range of productivity values within an analysis, when comparing across different studies we should expect no simple relationship between extent and the range of values. For instance, in the lower middle latitudes, orographic features can produce pronounced variation in climatic conditions (water regimes and temperature, but not day length), and thus in productivity, in the space of a few kilometers, as can rivers running through the world's more arid areas. On the other hand, some data sets used in meta-analyses of the SRPR include tropical rain forest sites sampled in different continents, spanning a vast geographical extent but only a limited range of climate space. While such a data set does contain huge variation in terms of the constituent species pools involved in the different regions, the range of variation in the independent variable, i.e., productivity, may be quite limited compared with the local dry–mesic scenario outlined above. Grouping studies for analysis of the SRPR by their geographical extent that comprise data sets varying in their focal scale across many orders of magnitude, as undertaken by Mittelbach et al. (2001), is to generate an analysis fundamentally confounded by (true) scale. The empirical analyses by Chase and Leibold (2002) and Braschler et al. (2004) show this to be so.

Pärtel et al. (2007), on the other hand, simply ignore focal scale and system extent altogether, which is even worse, as *both* parameters are fundamental to the emergent form of the SRPR. Again, examination of two case studies is instructive. Chase and Leibold (2002) analyzed the richness of aquatic macrophytes in 30 ponds of about 500 m^2. They report a unimodal SRPR at the pond scale, but a simple positive SRPR when the data were aggregated up to the catchment scale by combining approximately three ponds per catchment. As Pärtel et al. (2007) were interested in analyzing variation in form of the SRPR with latitude, and failed to structure the analysis by scale, this particular study system appears twice in their meta-analysis for the same geographical co-ordinates, once as a unimodal SRPR (study 25, pond scale) and once as a positive SRPR (study 26, catchment scale), i.e., two different votes for the same place. In a second case study, Braschler et al. (2004) report analyses at three spatial scales, in each case providing separate analyses for graminoids, forbs, and forbs with graminoids. If following Braschler et al. (2004: Fig. 2) we could score this study as providing two unimodal relationships, four negative relationships and three null relationships. Or, we could follow the rationale that including taxonomic subsets of the same data is a form of "double-dipping" and we could just include the combined data for forbs with graminoids ("all plants") at each of three reported scales, providing one unimodal, one null and one negative relationship. In this case, P2007 enter two unimodal records for this study system (their studies 16 and 17), i.e., two rather than three "votes" for this system. A third example of multi-scale analysis is the paper by Chalcraft et al. (2004), who provide two focal scales of analysis for two separate sites, providing potentially four "votes": recognized by GW2006, but not by P2007 who record two "votes" only for this system. Hence, multi-scale treatments have been handled in different ways within P2007 and across the different meta-analyses. In fact, while the Braschler et al. (2004) study has other important things to say, the key conclusion to emerge from each of these three source papers is that there is no single form of SRPR for the systems they have analyzed and crucially the outcome is dependent on the focal scale used in the study.

P2007 not only make no attempt to control scale effects, they do not even record the scale parameters of the study systems in their meta-analysis. Their approach is to contrast the form of SRPR between low and high latitudes. But, as we know that the form of the SRPR varies depending on the focal scale used in analysis of the same data sets, and as focal scale (and extent) varies across many orders of magnitude in the studies compiled in each of the meta-analyses, it is nonsensical to undertake such an analysis. So, even were their initial classifications of the form of each particular study system correct (which in the great majority of cases they are not), their meta-analysis would be fatally compromised by the variation in the distribution of focal scale and ecosystem-scale properties between low and high latitudes in their study.

I just referred to "ecosystem-scale properties," by which I had in mind another largely intractable problem in analyses of the SRPR (Marañón and García 1997, Gillman and Wright 2006). How should we handle systems in which there is a mix of extremes of vegetation types, e.g., low-herbaceous grasslands and woodland? There are several such studies in the meta-analyses (e.g., Weiher 2003). Trees have a modular unit size that is orders of magnitude larger than grass and forb ramets. To move across a gradient from open areas to oak woodland, as in Weiher's (2003) study, is to traverse a gradient in which the effective physical and resource space available to herbaceous species becomes vastly reduced (cf. Oksanen 1996). Overall system net primary productivity (NPP) is likely to be highest in the tree-dominated stands, so how should we treat such study systems? Should we record all plant diversity and all NPP, or should we restrict our measurements of both to the herbaceous layer? If we do the latter, how should we account for the reduction in physical space and especially resources in the woodland quadrats? Effectively, the incursion of trees into the stands means that sampling/resource space for herbaceous species has not been held constant even though plot dimensions have been (for an extreme example see Nilsson and Wilson 1991, who used 0.5 × 1.0 m quadrats despite the fact that their system included 5 m high stands of *Betula pubescens*; Appendix A, studies 103, 104). If we include the trees in both measurements, on the other hand, we have crossed an important boundary in ecosystem properties and seen a shift in the relative proportion of biomass contributed by many small plants (in treeless plots) in favor of a very few large plants: is this system going to provide a meaningful representation of the relationship between species richness and productivity? I regard this question as posing an unanswered theoretical challenge. For the record, Weiher's (2003) approach was to focus just on the herbaceous layer, but as his statistical analyses showed, the SRPR was in any case compromised by the active fire regime of the study system. Partël et al. (2007) classify it as a humped SRPR.

Plot Size Dictates the Form of the SRPR: A Theoretical Exposition

The question of what constitutes an acceptable minimum plot size is one that may depend in part on the purpose of the analysis, but it is surely self-evident that if your plot is too small to contain a single dominant individual, then it is too small to represent the local community (Gillman and Wright 2006). In a recent re-examination of appropriate plot sizes for phytosociological study of European vegetation, Milan and Zdenka (2003:563) come to the following conclusion: "... Based on our analysis, we suggest four plot sizes as possible standards. They are 4 m^2 for sampling aquatic vegetation and low-grown herbaceous vegetation, 16 m^2 for most grassland, heathland and other herbaceous or low-scrub vegetation types, 50 m^2 for scrub, and 200 m^2 for woodlands." Similar guideline plot sizes have in fact been around for decades, based largely on the wisdom that if the species accumulation curve for a vegetation type is beginning to approach an asymptote then a more-or-less stable representation of the local community may have been attained. Often, of course, plots need to be considerably larger than these sizes for stabilization of values to be reached (T. Stohlgren, *personal communication*). It is noteworthy that many of the studies used by Mittelbach et al. (2001) and by Pärtel et al. (2007) have plots significantly smaller than the least of these sizes (i.e., <4 m^2), including a number that were initially designed to analyze SRPR or species biomass–productivity relationships.

But, does this really matter? If the unit plot size is fixed, even if it is at a point on the species accumulation curve where richness is climbing steeply with increasing plot size, surely comparisons can be made? Yes, they can, but we should recognize that in such a case we are essentially working with point diversity (within community) rather than alpha diversity (richness representative of the local community) (sensu Whittaker 1977). This distinction may be important for interpretation of the SRPR (Oksanen 1996). Species accumulation curves typically rise very rapidly initially, and then flatten increasingly slowly until reaching an asymptote, rising again only when habitat boundaries are crossed to bring in genuine beta or gamma diversity (sensu Whittaker 1977) into the curve.

At very small plot sizes, beginning with perhaps 25-cm^2 grassland plots, physical competition for space, light, water, and nutrients is key in determining presence and richness of sub-patches within a sward. Using tiny plot sizes, we may therefore predict that analyses should typically return negative SRPR, as any increase in productivity will tend to be accompanied by a switch to larger ramets or clonal systems of one or two species (increased dominance, reduced equitability), reducing the likelihood of fitting in representatives of other species in these very small units of analysis. If there is an initial rising limb before the negative phase kicks in it will be apparent only briefly, with the negative phase starting at quite low productivity values. If we increase the plot size to and beyond the recommended Milan and Zdenka (2003) standards (i.e., a size where species accumulation curves are flattening) we can expect to see much more evidence of an initial rising limb as increased system productivity across a set of plots of varying productivity is matched by fitting in more ramets while retaining high equitability. Nonetheless, with further increases in productivity, we can again expect to find eventual decreases in species richness, particularly if our system includes artificially (or naturally) fertilized ("polluted") sites, within which those relatively few species in the local species pool that are best adjusted to exploiting high levels of nutrient are able to competitively out-grow other community members, expressing dominance (reducing equitability) and generating a reduction in richness. Thus, as we increase the focal scale, the position of the peak in richness should typically move from low in the productivity range toward higher values of productivity. And, as we escape the plot size at which local communities are defined and move to larger grain sizes (and different data types), and focal scales of analysis (up to and beyond 1000 km^2), we should expect to see increasing proportions of cases where species richness increases positively with productivity, either in a linear relationship or as an asymptotic curve with no downwards limb (Whittaker et al. 2001, Whittaker and Heegaaard 2003). This expectation is consistent with the overall findings of Gillman and Wright's (2006) meta-analysis, which I regard as the most rigorous of those reviewed herein.

To sum up, this theoretical exposition is linked to different conceptual realizations of diversity, invoking within-patch dynamics recorded at *point* scales of analysis, moving up to plot sizes more fully representing the local communities (i.e., to *alpha* scales of analysis), and eventually jumping to *gamma* scales of analyses, including whole landscapes or regions and in which climatic controls on species pools become apparent (in each case, point, alpha, and gamma are sensu Whittaker 1977). Hence, I posit that a lot of the variation reported in the literature on the form of the SRPR, in so far as it is based on adequate productivity data, and is meaningfully and accurately reported, essentially arises as an artifact or by-product of variation in the effective scale of sampling from point to alpha to gamma diversity. This argument is similar to but extends arguments made by Oksanen (1996). Variation in form at fine focal scales—and indeed what constitutes an appropriate scale of alpha analysis—will also depend on the range of physiognomic vegetation types incorporated (Marañón and García 1997, Chalcraft et al. 2004). As a crude generalization, however, negative SRPR should be expected to be most frequent for point scale data, with humped relationships more apparent at coarser alpha scales, and a gradual right shift of the hump, giving way to positive relationships within gamma scale analyses (shown schematically in Appendix B: Fig. B1). This somewhat speculative prediction could be tested by analyses using nested sampling based on plots of increasing grain size but fixed location across a fixed system extent.

CONCLUDING COMMENTS

In my former role as editor-in-chief of *Global Ecology and Biogeography*, it was my idea to introduce Meta-Analysis as an article type in that journal: then I was fired up with enthusiasm for the approach, now I wonder if we should not have labeled the section "Here be dragons," as might be found on some ancient maps to describe unknown and generally hazardous regions of the world from which even the bravest explorers have rarely returned unscathed.

On my first theme—failings of prior analyses—I conclude that much of the original research undertaken within what has become a paradigmatic framing of humped SRPRs has been poorly designed experimentally, has involved strongly confounding variables, inadequate plot sizes, and poor choices of incomplete surrogate variables. Several of these themes, notably the highly problematic nature of biomass as a productivity surrogate (Gillman and Wright 2006, Keeling and Phillips 2007), have scarcely been touched on in this critique, while others are detailed only in Appendix A. I hope, however, to have demonstrated that enough of these problems are important, to demand a reappraisal of thinking on the SRPR. The meta-analytical contribution to understanding the SRPR started with a transparent but flawed analysis (Mittelbach et al. 2001), which, however, succeeded in knocking down the notion that the SRPR has a general form (and that this general form is humped), progressed with a worthy (but imperfect) reanalysis (Gillman and Wright 2006), and has proceeded to the point where there no longer seems to be any stated or reproducible criteria or method involved (Pärtel et al. 2007, Laanisto et al. 2008). Despite efforts to correct failings in the original meta-analysis (Whittaker and Heegaard 2003, Wright and Gillman 2006), further meta-analysis papers have appeared that mutate outcomes from Mittelbach et al. (2001), compound many of the original failings, and add new ones, a process of multiplying small errors to the point of producing wholly unsound outcomes. All sorts of entirely inappropriate data sets have now been recycled to answer questions that are incompletely specified and essentially unanswerable. Meta-analysis has led, in short, to mega-mistakes.

I have written this article not with any desire to fall out with those whose work I have criticized but because I happen to think an understanding of the SRPR is of considerable importance within ecological and biogeographical theory and because I feel that ecology as a discipline would be ill served by letting these chronic failings multiply through the literature unchecked. These failings in the treatment of scale, sampling design, plot size, and so on, in fact extend well beyond the meta-analyses, but at least these weaknesses are readily detectable in the original case study papers. Colleagues, we have to do better than this when we undertake and review and read and cite meta-analyses. Perhaps we can, in time, address the lack of standardization of experimental design and the tendency to change our methods from one study to the next, and find more reliable ways of dealing with the inherent multivariate nature of ecological systems, but in the meantime, we should be wary of trying to crunch (analyze) chalk and cheese data sets together, and we should be circumspect in regard to the use of meta-analysis in ecology.

On the second theme of this article, my case is that analyses of the SRPR that are not placed in an explicit scale framework are essentially meaningless. And, while the geographical extent of the system can influence form of the SRPR, it is intrinsically less problematic to compare studies of different system extent than to attempt to meta-analyze systems of differing focal ("true") scale of data: in fact to do the latter is nonsensical in the same way that it would be nonsensical to compare the diversity of a 1-m^2 patch of grassland to a 1-km^2 area of grassland. We know this from first principles and we now know it from empirical proof of the relevance of focal scale to the form of the SRPR.

I think an understanding of the variation in form of the SRPR must involve an understanding of the different processes at work at different scales of analysis and of how these are likely to structure our data sets. At fine scales of analysis we need to combine sampling theory with an understanding of species abundance distributions and species accumulation curves (see, e.g., Oksanen 1996, Marañón and García 1997, Chalcraft et al. 2004), and at all scales we have to deal with the multivariate nature of ecological processes. My proposition in this article is speculative, and incomplete theoretically, focusing as it does on largely artifactual mechanisms, but for what it is worth, predicts a general switch in form from negative and unimodal to positive SRPR with increasing focal scale of analysis. While collecting together and "crunching" (i.e., analyzing) large collections of data sets has its place in ecology (I am not entirely averse to it myself), we may advance faster in our understanding of this particular relationship by framing innovative primary studies designed to test particular hypotheses than by paying attention to the misleadingly precise quantifications generated by the meta-analyses.

ACKNOWLEDGMENTS

For discussion of these issues and/or comments on a pre-submission draft I thank Bradford Hawkins, Kostas Triantis, Richard Ladle, Ole Vetaas, and David Nogués-Bravo; and for comments in review I thank Thomas Stohlgren, Gary Mittelbach (who kindly commented despite obvious differences of view), and two anonymous reviewers.

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.

Braschler, B., S. Zschokke, C. Dolt, G. H. Thommen, P. Oggier, and B. Baur. 2004. Grain-dependent relationships between plant productivity and invertebrate species richness and biomass in calcareous grasslands. Basic and Applied Ecology 5:15–24.

Chalcraft, D. R., J. W. Williams, M. D. Smith, and M. R. Willig. 2004. Scale dependence in the species-richness–productivity relationship: the role of species turnover. Ecology 85:2701–2708.

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

Cody, M. L. 1975. Towards a theory of continental species diversities: bird distributions over Mediterranean habitat gradients. Pages 214–257 *in* M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Harvard University Press, Cambridge, Massachusetts, USA.

Ehrman, T., and P. S. Cocks. 1990. Ecogeography of annual legumes in Syria: distribution patterns. Journal of Applied Ecology 27:578–591.

Flenley, J. R. 1969. The vegetation of the Wabag region, New Guinea highlands: a numerical study. Journal of Ecology 57: 465–490.

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.

Keeling, H. C., and O. L. Phillips. 2007. The global relationship between forest productivity and biomass. Global Ecology and Biogeography 16:618–631.

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.

Marañón, T., and L. V. García. 1997. The relationship between diversity and productivity in plant communities: facts and artefacts. Journal of Ecology 85:95–96.

Milan, C., and O. Zdenka. 2003. Plot sizes used for phytosociological sampling of European vegetation. Journal of Vegetation Science 14:563–570.

Mittelbach, G. G., S. M. Scheiner, and C. F. Steiner. 2003. What is the observed relationship between species richness and productivity? Reply. Ecology 84:3390–3395.

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.

Nilsson, C., and S. D. Wilson. 1991. Convergence in plant community structure along disparate gradients: are lakeshores inverted mountainsides? American Naturalist 137: 774–790.

Nogués-Bravo, D., M. B. Araújo, T. Romdal, and C. Rahbek. 2008. Scale effects and human impact on the elevational species richness gradients. Nature 453:216–219.

O'Brien, E. M. 1993. Climatic gradients in woody plant species richness: towards an explanation based on an analysis of Southern Africa's woody flora. Journal of Biogeography 20: 181–198.

O'Brien, E. M. 1998. Water–energy dynamics, climate, and prediction of woody plant species richness: an interim general model. Journal of Biogeography 25:379–398.

O'Brien, E. M., R. J. Whittaker, and R. Field. 1998. Climate and woody plant diversity in southern Africa: relationships at species, genus and family levels. Ecography 21:495–509.

Odum, E. P. 1969. The strategy of ecosystem development. Science 164:262–270.

Oksanen, J. 1996. Is the humped relationship between species richness and biomass an artefact due to plot size? Journal of Ecology 84:293–295.

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

Pianka, E. R. 1966. Latitudinal gradients in species diversity: a review of concepts. American Naturalist 100:33–46.

Rahbek, C. 2005. The role of spatial scale and the perception of large-scale species-richness patterns. Ecology Letters 8:224–239.

Schamp, B. S., L. W. Aarssen, and H. Lee. 2003. Local plant species richness increases with regional habitat commonness across a gradient of forest productivity. Folia Geobotanica Phytotaxonomica 38:273–280.

Slavin, R. E. 1995. Best evidence synthesis: an intelligent alternative to meta-analysis. Journal of Clinical Epidemiology 48:9–18.

Wardle, D. A., O. Zackrisson, G. Hörnberg, and C. Gallet. 1997. The influence of island area on ecosystem properties. Science 277:1296–1299.

Weiher, E. 2003. Species richness along multiple gradients: testing a general multivariate model in oak savannas. Oikos 101:311–316.

Wheeler, B. D., and K. E. Giller. 1982. Species richness of herbaceous fen vegetation in Broadland, Norfolk, in relation to the quantity of above-ground plant material. Journal of Ecology 70:179–200.

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. H. 1977. Evolution of species diversity in land communities. Pages 250–268 *in* M. K. Hecht, W. C. Steere, and B. Wallace, editors. Evolutionary biology. Volume 10. Plenum Press, New York, New York, USA.

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

Whittaker, R. J., K. J. Willis, and R. Field. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28:453–470.

Whittaker, R. J., K. J. Willis, and R. Field. 2003. Climatic–energetic explanations of diversity: a macroscopic perspective. Pages 107–129 *in* T. M. Blackburn and K. J. Gaston, editors. Macroecology: concepts and consequences. British Ecological Society Symposia Series. Blackwell Publishing, Oxford, UK.

Williams, R. J., G. A. Duff, D. M. J. S. Bowman, and G. D. Cook. 1996. Variation in the composition and structure of tropical savannas as a function of rainfall and soil texture along a large-scale climatic gradient in the Northern Territory, Australia. Journal of Biogeography 23:747–756.

APPENDIX A

Case-by-case evaluation of plant data sets used in three meta-analyses of the species richness–productivity relationship (*Ecological Archives* E091-185-A1).

APPENDIX B

Schematic diagram of how changing focal scale may influence the form of the species richness–productivity relationship (*Ecological Archives* E091-185-A2).