J Insect Conserv (2012) 16:87–92
DOI 10.1007/s10841-011-9395-6

ORIGINAL PAPER

Can higher taxa be used as a surrogate for species-level data in biodiversity surveys of litter/soil insects?

Neil Rosser · Paul Eggleton

Received: 7 October 2010 / Accepted: 5 March 2011

Abstract We investigate the usefulness of higher taxa as surrogates for species richness and diversity of litter/soil insects. We use data for Coleoptera (beetles) and Formicidae (ants) collected during biodiversity surveys of five tropical and two temperate countries, and use Pearson's product moment correlations to assess the surrogacy relationship. Our results suggest that genera would provide an adequate surrogate for species richness of Coleoptera, but not for Formicidae. We suggest that the usefulness of higher taxa as surrogates for richness is dependent on both taxonomy and scale. Higher taxa provided a poor surrogate for species diversity due to the added dimension of evenness of community structure, and we recommend that higher taxa should not be used as surrogates for species diversity.

Keywords Higher taxa · Surrogates · Insects · Richness · Diversity

Introduction

The assessment of biodiversity for conservation requires detailed knowledge of the spatial distribution of organisms, yet counting species richness and diversity is frequently impractical. In order to facilitate the measurement of biodiversity, at times it may be necessary to employ surrogates for species-level data (Margules and Pressey 2000). Possible surrogates include indicator groups, environmental variables, assemblages and higher taxa (Beccaloni and Gaston 1995; Gaston and Blackburn 1995; Gaston and Williams 1993; Margules et al. 2002). This study focuses on the usefulness of higher taxa as surrogates.

Higher taxa were first used as surrogates for species in palaeontology. For example, studies of mass extinctions generally use family-, order- or class-level data due to the problem of adequately sampling species in the fossil record (Raup 1979). In environmental impact studies of marine habitats the related idea that it may not be necessary to identify samples to species-level while still retaining sufficient information is known as "taxonomic sufficiency" (Ellis 1985; Warwick 1988). More recently, the possibility of applying the higher taxon approach to conservation biology has received considerable attention (e.g. Balmford et al. 1996a, b, 2000; Gaston and Williams 1993; Villaseñor et al. 2004; Williams and Gaston 1994).

There are a number of potential advantages of using higher taxa as surrogates, most notably a reduction in time and cost. Balmford et al. (1996b) estimated that for woody plants in Sri Lankan forests, surveying genera or families rather than species reduced costs by at least 60 and 85%, respectively. Furthermore, the use of family-level data in site-selection algorithms lost only 7–10% of species from the reserve network, and using genera lost a mere 1–2%. The approach is particularly appealing as the uneven "hollow curve" distribution of species among taxa (e.g. Roy et al. 1996) means that problematic, species-rich groups (e.g. Curculionidae or Staphylinidae) can be dealt with easily. These savings will be most useful in tropical areas where

N. Rosser · P. Eggleton
Department of Entomology, Natural History Museum,
Cromwell Road, London SW6 5BD, United Kingdom
e-mail: p.eggleton@nhm.ac.uk

Present Address:
N. Rosser (✉)
The Galton Laboratory, Department of Biology,
University College London, 4 Stephenson Way,
London NW1 2HE, United Kingdom
e-mail: neil.rosser@ucl.ac.uk

diversity is very high and many species have yet to be described. Another key advantage of using higher taxa rather than other surrogate methods is that some information on the identity of organisms and turnover between sites is retained, thus potentially enabling ecological studies and consideration of complementarity and character diversity (Balmford et al. 1996b; Gaston and Williams 1993; Pik et al. 1999).

The higher taxon approach suffers from certain limitations (Anderson 1995; Prance 1994). Most previous analyses of its usefulness have used sites dispersed across large scales, but it might be expected that at local scales higher taxa will be less likely to accurately predict species richness. This is because sites that are close together are likely to vary in richness considerably less than sites that are far apart, meaning that higher taxa may not reflect species patterns (Balmford et al. 1996a, but see Anderson 1995). Previous work has also been biased towards better known, less diverse groups such as vertebrates and vascular plants. This is unsurprising as few good quality, local-scale datasets for very species-rich groups currently exist (Balmford et al. 2000). The present study attempts to address these inequalities by testing the surrogacy relationship for two diverse insect groups, Coleoptera (beetles) and Formicidae (ants), at a local scale in five tropical and two temperate countries. Some of the data sets are from single locations with samples taken from within a habitat-type. The study therefore provides a particularly rigorous test of the scale to which the surrogacy relationship between species and higher taxa may be successfully applied.

A second related question which has received little attention is whether higher taxa can act as surrogates for species diversity and community structure as opposed to just richness (but see Kaesler et al. 1978). Species diversity indices combine species richness and equitability into a single number. Measures of equitability express the evenness of community structure (i.e. the distribution of abundances between species in an assemblage). Here we assess whether higher taxa are adequate surrogates for species using a parametric measure of diversity [the α parameter from Fisher's log-series model (Fisher et al. 1943), two commonly used non-parametric diversity indices (the Shannon index (Shannon and Weaver 1949) and the Simpson index (Simpson 1949)], and a pure evenness measure (the Berger-Parker index; Berger and Parker 1970).

Methods

Sources of data: Coleoptera

Abundance data from standard biodiversity surveys were obtained from databases of the Soil Biodiversity Group at the Natural History Museum. The survey datasets are summarised in Table 1. Coleopteran surveys were undertaken in the United Kingdom, Chile and Malaysian Borneo. Survey sample units comprised 100 m long transects sampled with 1 m^2 quadrats at 7 m intervals (thus fifteen quadrat sub-samples were pooled to make up a single sample; see Eggleton et al. (2009) for explanations of the pooling approach). Coleoptera were extracted by Winkler bag (Krell et al. 2005). Data from the United Kingdom consisted of 12 such 15-quadrat samples taken from a single wooded site over the course of a year (one sample per month). Data from Chile consisted of 14 samples collected from 4 different habitat types in the Aisén (two deciduous forests, temperate rain forest and steppe-edge forest). Data from Borneo were collected in Sabah over 2 years. During the first year five samples were collected from primary rainforest, small fragments of rain forest and an oil palm plantation. During the second year a further five samples were collected along a gradient from primary undisturbed rainforest to entirely deforested land. Specimens from Borneo had been separated into families and morpho-species, but only the first year data had been identified at genus-level. Thus second year data were combined with first year data and used only to test the utility of families as surrogates. When analysing the first year data some problems were encountered concerning unidentified genera. If a species had not been assigned to a genus it was assigned its own unique genus, distinct from all other species. If two or more species of undetermined genera had been noted to be similar to each other, they were assigned to the same (unknown) genus. Species that could not be assigned even to family were excluded from the analysis entirely. Similar problems were encountered with the Chilean data and the same rules applied as necessary (although in this case species-level data comprised a mixture of species and morpho-species).

Table 1 Summary of survey datasets

Taxa	Country	Number of samples	Sample units
Coleoptera	United Kingdom	12	100 m transect
Coleoptera	Chile	14	100 m transect
Coleoptera	Malaysia (Borneo)	10	100 m transect
Formicidae	Belize	79	1 m^2 quadrat
Formicidae	Cameroon	7	100 m transect
Formicidae	Gabon	39	1 m^2 quadrat
Formicidae	Ghana	28	100 m transect
Formicidae	Malaysia	138	1 m^2 quadrat

Sources of data: Formicidae

Formicidae surveys were conducted in Belize, Cameroon, Gabon, Ghana and peninsula Malaysia. Data from Cameroon and Ghana were collected using the sampling method described above. In Cameroon seven samples were taken from three undisturbed forests and two deforested areas. In Ghana 30 samples were taken from primary forest, degraded primary forest, secondary forest and coconut plantations. Data from Belize (79 samples), Gabon (39 samples), and Malaysia (138 samples) were collected from a single forested location each, with samples comprising randomly distributed 1 m^2 quadrats. Formicidae specimens had always been identified to genera- and to species-level (either named species or morphospecies). The higher taxonomic classification used for the Formicidae was in accordance with that described on AntWeb (http://www.antweb.org).

Statistical analysis

Statistical analyses, using the program R (version 2.5.0, R Core Development Team 2007), were applied to assess the usefulness of families and genera as surrogates for Coleopteran species, and sub-families and genera as surrogates for Formicidae species. We used the Shapiro–Wilk test to test the data for normality prior to parametric correlations. We then carried out Pearson's product moment correlations between richness and diversity of species and higher taxa, and calculated the coefficient of determination (r^2). To substantiate the significance of parametric correlations when assumptions were violated we also carried out a series of Spearman's rank correlations using all combinations of variables.

Results were compared using several different diversity indices: Fisher's α, the Shannon index, the Simpson index and the Berger-Parker index. Fisher's α is a parameter from Fisher's logarithmic series model which can be reliably used as a diversity index (Fisher et al. 1943; Magurran 2004). When calculating Fisher's α spurious values were occasionally produced for very small samples. In these cases the sample concerned was removed from the analysis. The Shannon index (Shannon and Weaver 1949) was calculated using the formula:

$$H' = -\sum p_i \ln p_i$$

where H' is the index of diversity and p is the proportion of the total sample belonging to the ith species. The Simpson index (Simpson 1949) was taken as the complement of D and calculated as:

$$D = \sum p_i^2$$

where D is Simpson's index of diversity and (1-D) it's complement, and p is the proportion of individuals in the ith species. The Berger-Parker index (Berger and Parker 1970) was used in its reciprocal form and calculated as:

$$d = N_{\max}/N$$

where d is the Berger-Parker index and (1/d) its reciprocal, N is the number of individuals in the sample and $N_{\max}$ is the number of individuals in the most abundant species.

Our results will not be confounded by sampling effort or area surveyed because these were held constant during data collection. However, a certain amount of spatial autocorrelation is to be expected due to closeness between data points within the datasets, particularly in cases where the same localities were sampled repeatedly. Analyses of the higher-taxon approach have statistical problems (Gaston 2000). Due to the statistical dependence between higher taxa and species, we consider primarily the strength and predictive power of correlations, rather than their statistical significance. For the Borneo and Chile datasets the strength of correlations between genera and species will have been slightly inflated by the method of dealing with unknown genera, but given the small proportion of unknown genera the effect should be small.

Results

Correlations with species richness

There were very strong correlations between richness of Coleopteran species and genera, with values for the coefficient of determination (r^2) ranging from 0.97 to 0.99 (Table 2). Richness of Formicidae species and genera were less well correlated, with r^2 values ranging from 0.61 to 0.91. Correlations between richness of species and families/sub-families were considerably weaker, with r^2 ranging from 0.09 to 0.88 for Coleoptera and 0.03–0.51 for Formicidae.

Correlations with diversity indices

The values of diversity indices combining richness and equitability (α, H', 1-D) were consistently well correlated between Coleopteran species and genera ($r^2 = 0.84$–1) but less so between species and families ($r^2 = 0.11$–1) (Table 2). Correlations between diversity of Formicidae species and higher taxa were weaker than those for Coleoptera (species versus genera $r^2 = 0.41$–0.96; species versus families $r^2 = 0$–0.35). Correlations between values of the Berger-Parker index of species and genera were very variable in strength, ranging from $r^2 = 0.2$–1 for Coleoptera and $r^2 = 0.38$–0.92 for Formicidae. Between species and families r^2 values ranged from 0.21 to 1 for Coleoptera and 0.04 to 0.38 for Formicidae.

The assumptions underlying product moment correlations were often not satisfied with transformation frequently

Table 2 Pearson's product moment correlations between richness, diversity and equitability of species and higher taxa

	Relationship		U.K. (beetles) (d.f. = 10)	Chile (beetles) (d.f. = 12)	Borneo (beetles) (d.f. = 3/8*)	Belize (ants) (d.f. = 77)	Cameroon (ants) (d.f. = 5)	Gabon (ants) (d.f. = 37)	Ghana (ants) (d.f. = 26)	Malaysia (ants) (d.f. = 136)
S	Genera versus species	r^2	0.97	0.97	0.99	0.87	0.61	0.64	0.81	0.91
		P	<0.001	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001
	Families versus species	r^2	0.09	0.55	0.88	0.51	0.31	0.03	0.38	0.41
		P	n.s.	<0.01	<0.001	<0.001	n.s.	n.s.	<0.001	<0.001
α	Genera versus species	r^2	0.95	0.97	0.94	0.49	0.68	0.81	0.62	0.68
		P	<0.001	<0.001	<0.01	<0.001	<0.05	<0.001	<0.001	<0.001
	Families versus species	r^2	0.11	0.83	0.54	0.1	0.24	0.05	0.24	0.17
		P	n.s.	<0.001	<0.05	<0.01	n.s.	n.s.	<0.01	<0.001
H′	Genera versus species	r^2	1	0.98	0.9	0.85	0.41	0.51	0.72	0.95
		P	<0.001	<0.001	<0.05	<0.001	n.s.	<0.001	<0.001	<0.001
	Families versus species	r^2	0.99	0.55	0.69	0.3	0.04	0	0	0.35
		P	<0.001	<0.01	<0.01	<0.001	n.s.	n.s.	n.s.	<0.001
1-*D*	Genera versus species	r^2	1	0.99	0.84	0.87	0.6	0.47	0.65	0.96
		P	<0.001	<0.001	<0.05	<0.001	<0.05	<0.001	<0.001	<0.001
	Families versus species	r^2	1	0.44	0.43	0.25	0.35	0.01	0.01	0.34
		P	<0.001	<0.01	<0.05	<0.001	n.s.	n.s.	n.s.	<0.001
1-*d*	Genera versus species	r^2	1	0.88	0.2	0.63	0.92	0.46	0.38	0.89
		P	<0.001	<0.001	n.s.	<0.001	<0.001	<0.001	<0.001	<0.001
	Families versus species	r^2	1	0.41	0.21	0.1	0.38	0.04	0.06	0.18
		P	<0.001	<0.05	n.s.	<0.01	n.s.	n.s.	n.s.	<0.001

* 3 d.f. for analyses with genera, 8 d.f. for analyses with families

d.f. degrees of freedom. *n.s.* not significant unable to rectify the problem. However, in such cases P values remained virtually the same whether a parametric or non-parametric (Spearman's rank) correlation was used.

Discussion

Higher taxa as surrogates for species richness

Our results suggest that the generic richness of Coleoptera would provide a good surrogate for species richness, a finding which is consistent with a previous study on Coleoptera from central Hungary (Báldi 2003). Coleoptera comprise a large proportion of global biodiversity, and in many areas their taxonomy is poorly known. While the use of species-level data will always be preferable, using genera could greatly facilitate the assessment of even very localised sites for conservation purposes. However, contrary to previous work (Báldi 2003), Coleopteran family richness did not provide a suitable surrogate for species richness. Richness of ant genera provided a poorer and less reliable surrogate for species richness than did Coleopteran genera, and richness of Formicidae sub-families would clearly be an unsuitable surrogate.

One would expect taxonomic groups with low species/higher taxon ratios to exhibit a stronger correlation between richness of species and higher taxa than those with a high species/higher taxon ratios (Anderson 1995; Prance 1994). Within our datasets the average number of species per higher taxon is always lower for Coleoptera than for Formicidae (Table 3), which may help to explain why genera provide a better surrogate for the former. This illustrates the unfortunate irony that higher taxa will be least effective as surrogates in the groups where their use would provide the greatest savings (Balmford et al. 1996a).

However, a second important factor affecting the ability of higher taxa to correctly predict species richness is the scale of sampling. It seems likely that higher taxa will successfully predict patterns in species richness down to a certain scale, after which variations in richness will be so low that patterns may not be reflected by higher taxa (Balmford et al. 1996a, but see Anderson 1995). At the scales explored in this study Coleopteran genera appear a suitable surrogate, but families did not. That the latter finding conflicts with previous work by Báldi (2003) may therefore be explained by a difference in scale; sample points in Báldi's study comprised protected areas distributed across central Hungary. Likewise, although higher taxa of Formicidae did not appear to provide a suitable surrogate at the scale addressed in this analysis, this is not to say they might not across a wider scale. For example, Negi and Gadgil (2002) assessed higher taxa of Formicidae as surrogates using data points dispersed across a 500 km^2 area which varied considerably in altitude. They found that sub-families correlated notably better with species richness ($r^2 = 0.72$) than was the case for any of the data sets considered here.

Higher taxa as surrogates for diversity

Kaesler et al. (1978) suggested that higher taxa could be used in place of species when calculating diversity indices with little loss of information. However, we found that correlations between diversity of higher taxa and species were generally weaker than those observed for richness, and also more variable in strength. As with species richness, taxonomy and scale are likely to strongly affect the strength of the surrogacy relationship. However, the consideration of evenness complicates the surrogacy relationship by adding a further dimension, leading to less consistently strong correlations than for richness alone. Furthermore, evenness of community structure probably fluctuates temporally and spatially to a greater extent than richness, further reducing confidence in the predictability of the surrogacy relationship. This means that the relationship may work very well at a certain point in space and time, yet be considerably less effective under only slightly different circumstances. This point is well illustrated in the UK dataset by the species *Acrotrichis* cf. *intermedia* (Gillmeister 1845), whose population explodes unpredictably and dominates the community in certain years (P. Eggleton, unpublished data). When *A. intermedia* was included in the analysis, very strong and highly significant relationships were noted between values for the Shannon, Simpson and Berger-Parker index calculated using families and those calculated using species. However, removing *A. intermedia* to create a dataset potentially more in keeping with a normal year resulted in the loss of these significant correlations.

Table 3 Average number of species per higher taxon

	Coleoptera U.K.	Coleoptera Chile	Coleoptera Borneo 1	Coleoptera Borneo 2	Formicidae Belize	Formicidae Cameroon	Formicidae Gabon	Formicidae Ghana	Formicidae Malaysia
Genus	1.45	1.15	1.93	–	2.9	3.36	3.33	4.07	2.85
Family	6.27	4.13	8.23	13.29	10.5	13.88	12.5	20.78	13.88

Conclusions

Our results show that genera of Coleoptera would provide a good surrogate for species richness in biodiversity surveys if resource limitations preclude the use of species. Because our study used fine scale data from three continents where taxonomic practises may vary considerably, this finding seems likely to be widely applicable. Formicidae genera, along with family-level data for both Coleoptera and Formicidae, seem too poorly correlated with species richness to provide suitable surrogates. As such, our results strongly highlight that whether a particular taxonomic rank can be successfully used as a surrogate requires careful consideration of both scale and the taxonomy of the group under consideration. Our data suggest that the use of higher taxa as surrogates for species diversity is inadvisable, because the addition of evenness of community structure to the surrogacy relationship severely reduces confidence in the predictability of the surrogacy relationship.

Acknowledgments We thank the following people who compiled the datasets used in this analysis; David Jones, Kelly Inward, Arthur Chung, Robert Belshaw, Peter Hammond, Daegan Inward, Emma Ross and Martin Brendell. Jeremy Field and Jonathan Green provided valuable feedback on an earlier version of the manuscript.

References

Anderson AN (1995) Measuring more of biodiversity: genus richness as a surrogate for species richness in Australian ant faunas. Biol Conserv 73:39–43

Báldi A (2003) Using higher taxa as surrogates of species richness: a study based on 3700 Coleoptera, Diptera, and Acari species in Central-Hungarian reserves. Basic Appl Ecol 4:589–593

Balmford A, Green MJB, Murray MG (1996a) Using higher-taxon richness as a surrogate for species richness: I. Regional tests. Proc R Soc Lond Ser B Biol Sci 263:1267–1274

Balmford A, Jayasuriya AHM, Green MJB (1996b) Using higher-taxon richness as a surrogate for species richness: II. Local applications. Proc R Soc Lond Ser B Biol Sci 263:1571–1575

Balmford A, Lyon AJE, Lang RM (2000) Testing the higher-taxon approach to conservation planning in a megadiverse group: the macrofungi. Biol Conserv 93:209–217

Beccaloni GW, Gaston KJ (1995) Predicting the species richness of Neotropical forest butterflies: Ithomiinae (Lepidoptera: Nymphalidae) as indicators. Biol Conserv 71:77–86

Berger WH, Parker FL (1970) Diversity of planktonic Foraminifera in deep sea sediments. Science 168:1345–1347

Eggleton P, Inward K, Smith J, Jones DT, Sherlock E (2009) A six year study of earthworm (Lumbricidae) populations in pasture woodland in southern England shows their responses to soil temperature and soil moisture. Soil Biol Biochem 41:1857–1865

Ellis D (1985) Taxonomic sufficiency in pollution assessment. Mar Pollut Bull 16:459

Fisher RA, Corbet AS, Williams CB (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. J Anim Ecol 12:42–58

Gaston KJ (2000) Biodiversity: higher taxon richness. Prog Phys Geogr 24:117–127

Gaston KJ, Blackburn TM (1995) Mapping biodiversity using surrogates for species richness: macro-scales and New World birds. Proc R Soc Lond Ser B Biol Sci 262:335–341

Gaston KJ, Williams PH (1993) Mapping the world's species–the higher taxon approach. Biodivers Lett 1:2–8

Kaesler RL, Herricks EE, Crossman JS (1978) Use of indices of diversity and hierachical diversity in stream surveys. In: Dickson KL, Cairns J, Livingston RJ (eds) Biological data in water pollution assessment: quantitative and statistical analyses. American Society for Testing and Materials, Philadelphia, pp 92–112

Krell FT, Chung AYC, DeBoise E, Eggleton P, Giusti A, Inward K, Krell-Westerwalbesloh S (2005) Quantitative extraction of macro-invertebrates from temperate and tropical leaf litter and soil: efficiency and time-dependent taxonomic biases of the Winkler extraction. Pedobiologia 49:175–186

Magurran AE (2004) Measuring biological diversity. Blackwell, Oxford

Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405:243–253

Margules CR, Pressey RL, Williams PH (2002) Representing biodiversity: data and procedures for identifying priority areas for conservation. J Biosci 27:309–326

Negi HR, Gadgil M (2002) Cross-taxon surrogacy of biodiversity in the Indian Garhwal Himalaya. Biol Conserv 105:143–155

Pik AJ, Oliver I, Beattie AJ (1999) Taxonomic sufficiency in ecological studies of terrestrial invertebrates. Aust J Ecol 24: 555–562

Prance GT (1994) A comparison of the efficacy of higher taxa and species numbers in the assessment of biodiversity in the neotropics. Philos Trans R Soc Lond Ser B Biol Sci 345:89–99

Raup DM (1979) Size of the Permo-Triassic bottleneck and its evolutionary implications. Science 206:217–218

Roy K, Jablonski D, Valentine JW (1996) Higher taxa in biodiversity studies: patterns from eastern Pacific marine molluscs. Philos Trans R Soc Lond Ser B Biol Sci 351:1605–1613

Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana, IL

Simpson EH (1949) Measurement of diversity. Nature 163:688

Villaseñor JL, Ibarra-Manriquez G, Meave JA (2004) Higher taxa as surrogates of plant biodiversity in a megadiverse country. Conserv Biol 19:232–238

Warwick RM (1988) The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. Mar Pollut Bull 19:259–268

Williams PH, Gaston KJ (1994) Measuring more of biodiversity: can higher-taxon richness predict wholesale species richness? Biol Conserv 67:211–217