Biodiversity: higher taxon richness
Kevin J. Gaston
Progress in Physical Geography 2000; 24; 117
DOI: 10.1177/030913330002400108

The online version of this article can be found at:
http://ppg.sagepub.com

Additional services and information for Progress in Physical Geography can be found at:

Email Alerts: http://ppg.sagepub.com/cgi/alerts
Subscriptions: http://ppg.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 11 articles hosted on the SAGE Journals Online and HighWire Press platforms):
http://ppg.sagepub.com/cgi/content/refs/24/1/117
I Introduction

Species richness has, arguably, become the common currency of studies of biodiversity (Gaston and Spicer, 1998). As is apparent from most of the multiple definitions of ‘biodiversity’ or ‘biological diversity’ (see DeLong, 1996; Gaston, 1996a), this is but one measure amongst many. However, it acts as a reasonable surrogate for some of the other dimensions of biodiversity, and is a convenient measure for numerous, if far from all, purposes (Gaston, 1996b; Williams and Humphries, 1996). None the less, species richness is often not readily measured. This is particularly so for taxonomic groups which are very speciose, and where time and resources are limiting (which in practice is almost always the case). For example, Lawton et al. (1998) calculated that their inventories of eight groups of animals (birds, butterflies, flying beetles, canopy beetles, canopy ants, leaf-litter ants, termites and soil nematodes) across study sites in the Mbalyamo Forest Reserve in Cameroon, required five scientist-years (more than 10 000 hours) to sample, sort and catalogue the approximately 2000 species. By extrapolation, they suggested that to produce an all-taxa biological inventory (ATBI) for a ‘representative hectare’ of forest might absorb the efforts of 10–20% of the estimated global workforce of systematists if it were to be completed in a reasonable time frame.

Against this background, there has been a search for short cuts to estimating the extant species richness of one or more higher taxa in an area (e.g., Cranston and Hillman, 1992; Pearson, 1994; Beccaloni and Gaston, 1995; Gaston, 1996b; 1996c; Oliver and Beattie, 1993; 1996a; 1996b; Duelli and Obrist, 1998; Howard et al., 1998; Lwanga et al., 1998; Williams and Gaston, 1998; Blair, 1999). One route is to find surrogates which are reasonably well correlated with species richness, but which are markedly easier to assess. One potential such surrogate is the numbers of higher taxa (e.g., genera, tribes, families) in the area. The roots of this approach are deep. Numbers of higher taxa have long been used as surrogates for numbers of species in palaeontological work, where it is common to conduct analyses particularly at the level of the family because of the low
likelihood that many individual species will have been preserved in recovered parts of the fossil record but the greater likelihood of preservation of at least one representative of any family (see Sepkoski, 1991; 1992; 1998). The usefulness or otherwise of this approach for assessing contemporary patterns of biodiversity, and the potential relative advantages and disadvantages of alternatives, have been much debated in recent years (e.g., Gaston and Williams, 1993; Beattie and Oliver, 1994; Prance, 1994; Williams and Gaston, 1994; Andersen, 1995a; 1995b; Balmford et al., 1996a; 1996b; Gaston, 1996b; Roy et al., 1996; Altaba, 1997; Crozier, 1997; Lee, 1997a; 1997b; Schoenly et al., 1998). In this article I selectively review these developments.

Throughout, I largely ignore the use of numbers of higher taxa as measures of biodiversity in their own right. However, such an approach has been employed, regardless of the presence or absence of a significant correlation with numbers of species (e.g., Eggleton, 1994; Gaston et al., 1995; Williams et al., 1997); numbers of higher taxa may be as valid a measure of biodiversity as many others. Some of the issues discussed here will also be relevant in that context.

II Rationale

The rationale behind using numbers of higher taxa as surrogates for numbers of species is simple. If the two are sufficiently well correlated across areas, then:

1) There are inevitably, and usually markedly, fewer higher taxa of a given rank to be counted in an area than there are species (with that number obviously declining for taxa of progressively higher rank). Thus, for example, even amongst the mammals, which have been particularly highly split at higher taxonomic levels (resulting in a disproportionate number of higher taxa), under one classification scheme there are 4629 extant and recently extinct species, but only 1135 genera and 136 families (Cole et al., 1994).

2) It is typically much easier to distinguish between higher taxa than it is amongst their constituent species.

3) Higher taxa are for some groups more clearly defined than are species.

4) Because most higher taxa contain relatively few species and a few contain very many (Willis, 1922; Williams, 1964; Anderson, 1974; Clayton, 1972; 1974; Bock and Farrand, 1980; Jablonski and Flessa, 1986; Dial and Marzluff, 1989; Hubbell, 1995; Roy et al., 1996), in enumerating higher taxa disproportionate effort is not expended on differentiating amongst species in the speciose taxa. This is particularly important because species of the highly speciose higher taxa are commonly morphologically very similar. Indeed, many of those higher taxa with large numbers of species (e.g., Asteraceae, Orchidaceae, Curculionidae, Pyralidae) are regarded as taxonomically more complex, albeit the reasons are not always well understood.

5) The numbers of higher taxa in many areas are already better enumerated than are the numbers of species (Gaston, 1996b).

An assessment of the potential savings in cost of only identifying higher taxa, and not their constituent species, has been made by Balmford et al. (1996b). Using data on the diversity of woody plants in Sri Lankan forests, they observe that targeting woody plant genera and families rather than species reduced survey costs by a minimum of
60% and 85%, respectively. If these findings generalize then, obviously, the possible benefits of simply enumerating higher taxa are substantial.

Two criticisms have, however, often been levelled at the principle (let alone the practice) of using higher taxa as surrogates for species richness (regardless of whether numbers of species and numbers of higher taxa are positively correlated). The first is that higher taxa are not comparable units, in that within or between groups (the latter being the more severe) higher taxa of the same rank (e.g., families) may constitute rather different entities. The second criticism is that higher taxa are not natural units, in that in many instances they may not be monophyletic, but may be paraphyletic or polyphyletic. These criticisms have been particularly vehement in the context of palaeontological studies. Without doubt, to varying degrees both observations are correct; taxonomic classificatory systems in parts reflect both the evolutionary process and the ways in which humans have chosen to organize information (Holman, 1985).

However, in the context of acting as surrogates for species richness, higher taxa function simply as packages of species diversity which are more amenable to analysis and hence make tractable some problems which otherwise would not be so (Lee, 1997a). As such, they need neither necessarily to be strictly comparable nor natural units. Sepkoski (1992) has illustrated this with reference to the strong positive correlation that exists between the size of the human population of a state in the USA and the number of municipalities that state contains. Despite variation in the size of municipalities of about four orders of magnitude and their arbitrary delineations, the number of municipalities in a state provides a good basis for estimating its relative population size. Sepkoski and Kendrick (1993) conducted Monte Carlo simulations to explore how accurately paraphyletic taxa captured underlying species patterns of diversity and extinction when compared with monophyletic groups. They found that the former perform adequately under a variety of conditions of diversification and mass extinction. In fact, of course, higher taxa are not arbitrary units, tending to comprise species that are on average far more closely related than would be expected by chance.

Arguably, on the rare occasions when detailed information is available regarding the relatedness of different groups then this information should be employed more explicitly in generating direct or surrogate measures of biodiversity, using some of the variety of measures of taxonomic/phylogenetic diversity which can be applied to such data (for a review and references, see Williams and Humphries, 1996).

III Empirical evidence

The widespread occurrence of positive correlations between the numbers of species and numbers of higher taxa across different areas might be expected because both have been found for a variety of groups to follow broadly similar patterns of geographic variation (e.g., Stehli et al., 1967; Wilson, 1974; Rabinovich and Rapoport, 1975; Taylor and Taylor, 1977; Wolseley et al., 1994; Roy et al., 1996), and to capture broadly similar patterns of community structure and its response to environmental factors (e.g., Kausler et al., 1978; Jones, 1993; Olsgard et al., 1997, 1998 and references therein). Indeed, a growing number of studies have reported significant positive relationships between numbers of species and numbers of higher taxa. These include studies of plants (Smith and Theberge, 1986; Gaston and Williams, 1993; Williams and Gaston, 1994; Williams et al., 1994, Balmford...
et al., 1996a; 1996b; O'Brien et al., 1998), amphipods (Myers, 1997), molluscs (Jablonski and Flessa, 1986; Roy et al., 1996), termites (Eggleton et al., 1994), butterflies (Williams, 1993; Williams and Gaston, 1994), tenebrionid beetles (Vilenken and Chikatunov, 1998), bumblebees (Williams, 1993), ants (Andersen, 1995a; 1995b), fish (McAllister et al., 1994), amphibians (Gaston and Williams, 1993), birds (Gotelli and Abele, 1982; Gaston and Williams, 1993; Williams and Gaston, 1994; Gaston and Blackburn, 1995; Balmford et al., 1996a) and mammals (Williams and Gaston, 1994; Balmford et al., 1996a).

Published examples where significant relationships do not exist are rather scarce, although there must be some concern that there is a strong selective bias against the publication of such findings (the ‘file drawer problem’; Rosenthal, 1979; Csada et al., 1996). Most published analyses of relationships between numbers of species and numbers of higher taxa have also been for multicellular organisms in temperate regions, and there is a need for a more diverse set of studies. In addition, there is a pressing need to move from a set predominantly concerned with establishing the occurrence of positive relationships towards a set which overall have been more rigorously conducted (see Analytical issues); this is not to deny the excellence of some of the work which has thus far been conducted.

Of course, the existence of significant positive correlations need not mean that the relationships that they describe are of any predictive value (which appears to be a source of confusion in some biodiversity studies). However, it has been argued that the correlations between species numbers and higher taxon richness are commonly sufficiently strong that this is indeed the case, and that any regional variations which they embody typically do not obscure general patterns. It is important to remember that frequently what is required of surrogates is the identification of relative levels of richness, and higher taxa may often perform reasonably well in this regard (but see Andersen, 1995a). Moreover, the use of higher taxa as surrogates of species numbers must be weighed against the predictive value of other approaches. As Balmford et al. (1996a) observe, they appear to perform noticeably better than, say, relationships between numbers of species across different taxonomic groups.

In perhaps the only published comparative analysis, Gaston and Blackburn (1995) contrasted the value of some environmental variables and of numbers of higher taxa for predicting numbers of species of birds on a coarse grid across the New World. Numbers of genera and families were found to have the strongest correlations with species richness. Moreover, they provided the best predictions of the numbers of species in areas of Central and South America on the basis of data for North America. Similar analyses, particularly at finer spatial resolutions, would be useful in clarifying the relative value of higher taxa as surrogates for species richness.

Typically, correlations between numbers of species and numbers of higher taxa become weaker towards higher (more inclusive) levels of the taxonomic hierarchy, in other words as the mean species richness of the higher taxa increases (e.g., from genera to families; Gaston and Williams, 1993; McAllister et al., 1994; Williams and Gaston, 1994; Balmford et al., 1996a; 1996b). Indeed, when the ratio of numbers of species to numbers of higher taxa becomes high, one would expect the latter to act as poor surrogates for the former; the total number of higher taxa may, under these circumstances, be recorded at relatively low species numbers (i.e., the number of higher taxa asymptotes rapidly). Thus, for example, Gaston et al. (1995) observe little systematic latitudinal variation in the numbers of families of beetles across the earth, at least when
recorded at coarse scales, although it is well established that their species numbers increase markedly towards tropical realms.

There will tend to be a trade-off between the taxonomic level that functions as the best predictor of species richness, and the costs of identifying those taxa. The closer the taxonomic level employed is to that of the species, the better a predictor of species richness it is likely to be, but the greater the associated cost (in time and resources).

The apparent broad generality of relationships between numbers of higher taxa and numbers of species in analyses of contemporary assemblages serves, it has been argued, to provide direct empirical support for the use of numbers of higher taxa as surrogates for numbers of species in palaeontological studies (Lee, 1997a). This is rather ironic given that such studies had been employing higher taxa as surrogates long before much consideration had been given to the possibility with regard to work on contemporary assemblages. Obviously, in many cases the veracity of the proposal depends fundamentally on the temporal consistency of the relationship between numbers of species and numbers of higher taxa; most studies of extant species have by definition drawn variation in numbers of taxa across space rather than time.

IV Analytical issues

The published literature documenting relationships between numbers of species and numbers of higher taxa should not be viewed uncritically. For many purposes the correlations may be inadequate. Moreover, there are some potential analytical problems with some or many published analyses for contemporary assemblages. Three are particularly notable. First, some studies are based on areas of differing size. That raises the possibility that numbers of species and numbers of higher taxa are positively correlated simply because the numbers of both increase with area, and that if area were to be held constant no such relationship would be found. Plainly area cannot explain all the correlations observed, because in some studies the sizes of different areas have been held constant (e.g., Gaston and Williams, 1993; Williams and Gaston, 1994; Williams et al., 1994). Balmford et al. (1996a) found that controlling for area made little difference to the relationships between numbers of species and numbers of genera, families or orders in protected areas in Indo-Malaya and the adjacent Pacific rim, for angiosperms, birds and mammals. Balmford et al. (1996b) found a similar outcome for woody plants from 35 forest reserves in the wet zone of Sri Lanka. This approach controls for area on the basis of the between-area species-area relationship, which may not reflect how the number of species would respond if each individual area was reduced in extent (see Holt, 1993). None the less, it is commonly the only practicable approach, as information on species-area relationships for each separate data point is seldom likely to be available.

Secondly, the correct null hypothesis will not strictly be the absence of a correlation between numbers of higher taxa and numbers of species. Because every higher taxon that occurs in an area must contain at least one species, there is a positive lower bound to the relationship. If data points were distributed at random above this lower bound then there would be a disproportionate probability that they would generate at least a weak positive relationship. In practice, in many analyses this is not likely to be a major problem, because the ratio of the numbers of higher taxa to numbers of species is very low (and thus the lower bound is very shallow). However, rather than simply plotting
the numbers of species against the numbers of higher taxa across different areas, it could readily be addressed by determining whether the number of species in excess of the number of higher taxa increased with the number of high taxa.

In fact, a null model of the relationship between numbers of species and numbers of higher taxa is provided by the hypergeometric draw, in which the number of species in an area is drawn at random (without replacement) from the overall regional pool of species, and the numbers of higher taxa this number comprises is determined (see Järvinen, 1982, for a historical review; Gotelli and Abele, 1982). This model has been found to fit most data sets reasonably well, although speciose areas tend to have more higher taxa than expected, and species poor areas tend to have fewer (Gotelli and Abele, 1982; Williams and Gaston, 1994; Roy et al., 1996). Of course, in situations where one might want to use numbers of higher taxa as a surrogate for numbers of species the detailed distribution of species amongst higher taxa will not be known.

Thirdly, probably all published analyses have ignored the effects of spatial autocorrelation on the statistics reported. Numbers of species and numbers of higher taxa are likely to be more similar in areas which are closer together than in areas which are far apart, which means that data points in a typical analysis are not statistically independent. Such problems afflict studies of many other issues in biodiversity and conservation biology. Several methods to address them have recently been discussed (e.g., Clifford and Richardson, 1985; Clifford et al., 1989; Carroll and Pearson, 1998; Pearson and Carroll, 1998).

Fourthly, probably all published analyses have ignored the phylogenetic nonindependence of data points, which results because some pairs of areas will share very similar taxonomic structure and most will have some taxa in common. Standard methods are not presently available to deal with this problem, but it should none the less be recognized as being such.

All these analytical issues deserve attention, as they will help determine both the true effectiveness with which numbers of higher taxa can be used as surrogates for numbers of species, and the best models by which this surrogacy can be achieved.

V Caveats

From the outset it is plain that there are a number of situations in which numbers of higher taxa are unlikely to prove to be good surrogates for species richness, no matter how good they may be in general:

1) Where numbers of higher taxa are very similar to numbers of species – the savings in time and effort will be limited, and one would do as well to count species directly.

2) Where numbers of higher taxa are extremely small relative to numbers of species – here it is likely that all the higher taxa will be recorded quite quickly with increasing numbers of species (see above).

3) When data are included from regions with very marked differences in taxonomic structure – insular biotas, particularly isolated islands and lakes, may be notably problematic, because here evolutionary radiations may lead to very high ratios of numbers of species to numbers of higher taxa (Gaston, 1996b; Altaba, 1997) and biotas with otherwise unusual taxonomic structures (Cronk, 1997).
Patterns of species richness are only one descriptor of spatial variation in the distribution of biodiversity. For some purposes they are not the most useful or important one. In particular, this is true in conservation planning where, especially at finer scales, an emphasis simply on those areas which represent peaks of species richness (the hotspots) fails to ensure coverage of many species and results in repeated representations of others (e.g., Williams et al., 1996). The principle of complementarity offers a solution to this problem, complementary sets of areas being those which maximize overall richness in the smallest number of areas (Vane-Wright et al., 1991; Pressey et al., 1993). With regard to the use of higher taxa as surrogates for species, the question arises as to whether sets of complementary areas are similar for species and for higher taxa.

Balmford et al. (1996b) examined this issue using their data on woody plants in forest reserves in Sri Lanka, and an algorithm which sought to represent each taxon at least once in the reserve network. They found that species representation was nearly as efficient when site-selection was based on genus-level information as when based on species-level data. Both approaches tended also to select the same sites as priorities. However, family-level data were less efficient in species representation and there was less overlap in the identities of the areas selected.

Van Jaarsveld et al. (1998) have also examined the similarity of complementary areas for species and higher taxa, using data for 9119 species from eight groups (vascular plants, butterflies, termites, antlions, buprestid beetles, scarabaeoid beetles, birds and mammals) in the Transvaal region of South Africa. For both genera and families, complementary sets of cells on a 25 × 25 km grid showed little overlap with complementary sets for species; the highest overlap between species and genus complementary sets was 38% (for plants) and between species and family complementary sets was 40% (termites). Moreover, the differences between the overlap in complementary sets between species and genera and between species and families were inconsistent between the different groups of organisms. However, it is not apparent from these analyses to what extent there were differences in species representation when areas were selected using higher taxa.

In a related vein, Vanderklift et al. (1998) examined the use of genus, family and class-level data in selecting marine protected areas, using data on the distributions of 977 taxa of fish, invertebrates and plants in Jervis Bay, Australia. At each level, taxa were classified into ‘assemblages’ using ordination, and an iterative selection algorithm was used to identify networks of priority areas. Using all the survey data, similar proportions of species were found to be included in selected areas, whether the selections were based on species, genus or family-level assemblages, although a lower proportion was included using class-level assemblages. Comparable results were found when fish, invertebrates and plants were considered separately, although here class-level assemblages performed much better. In sum, the findings were broadly supportive of the use at least of genus and family-level data in the selection of marine protected areas.

Their small number, and the lack of comparability in the findings of these analyses mean that others are required before any broader lessons can be drawn.
VII Conclusions

If numbers of higher taxa can be used reliably as surrogates for numbers of species to establish the basic spatial patterns of biodiversity, then the prospects of rapidly generating broad scale maps, and ultimately an atlas of biodiversity (Gaston, 1998), would be much improved. However, much remains to be done beyond simply the establishment of significant correlations between numbers of species and numbers of higher taxa in a biased set of areas and groups. The approach is promising, but it has yet to be revealed whether this promise is to be attained.

Acknowledgements

K.J.G. is a Royal Society University Research Fellow. I am grateful to Sian Gaston and Ana Rodrigues for comments on the manuscript.

References

Beccaloni, G.W. and Gaston, K.J. 1995: Predicting the species richness of neotropical forest butterflies: Ithomiinae (Lepidoptera: Nymphalidae) as indicators. Biological Conservation 71, 77-86.


Gaston, K.J. and Williams, P.H. 1993: Mapping the world’s species – the higher taxon approach. *Biodiversity Letters* 1, 2–8.


Stein, F.G., McAlester, A.L. and Helsley, C.E. 1967: Taxonomic diversity of recent bivalves,


