

Patterns in the global distribution of flowering plant genera

N.A. BRUMMITT

BRUMMITT, N.A. 2005. Patterns in the global distribution of flowering plant genera. *Biol. Skr.* **55**: 539-564. ISSN 0366-3612. ISBN 87-7304-304-4.

Distributions of all vascular plant genera have been compiled from herbarium specimens at the Royal Botanic Gardens, Kew, and this data has then been analysed for large-scale patterns in the diversity and distribution of flowering plants. A strong latitudinal gradient in diversity is apparent at family, genus and species levels, though while western South America is most diverse at species and genus levels, it is the SW Pacific which is most diverse at family level. However, the number of families and genera per region is very strongly correlated overall, irrespective of the region. There is a very strong relationship between both family and genus diversity and area, though not for numbers of endemic genera. For both families and genera, the range-size frequency distribution is highly skewed towards small range sizes (more so for genera than for families), which account for the majority of diversity. Analysing floristic similarity between different regions of the world reveals very strongly supported continental groups, since most genera are confined to particular continents, although the latitudinal difference between regions is a better predictor of floristic similarity than is simply distance between regions. Distribution patterns show strong regional clustering, with almost 40% of genera single-region endemics, and approximately 20% of world distribution patterns accounting for about 80% of total angiosperm genus diversity. Analysis of distribution patterns reveals a strong correlation between diversity and number of floristic elements, which intersect to form the diversity of a region. Since only a minority of genera are endemic to any one region, this floristic complexity forms an important component of the diversity within that region.

Neil Brummitt, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, United Kingdom. E-mail: n.brummitt@rbgkew.org.uk

Introduction

This paper presents results from several separate, yet inter-connected, analyses of patterns of diversity for different taxonomic scales, and also patterns of distribution, focusing at the level of genus, on a global scale for all angiosperms. By patterns of diversity, I mean taxonomic richness, or counts of all taxa (of the same rank) naturally occurring within an

area, compared between separate areas. Patterns of distribution therefore refer to the comparison of all native occurrences for collections of taxa, throughout different areas. Since most genera show distributions which extend outside any single one of the regions used for this study, the set of taxa occurring within a particular geographical region can be thought of as the intersection of many different patterns of distribution. Therefore, it is argued

that the diversity within that region depends not only on features unique to that region itself, but also on the wider distributions outwith the region of those taxa found there.

The Vascular Plant Families and Genera Database

At the Royal Botanic Gardens, Kew, a taxonomic database of all genera of vascular plants has been continuously maintained since its initial publication over a decade ago (Brummitt 1992). To this database have recently been added distributions of all taxonomically-accepted genera (14,724 genera). These distributions have been recorded as verbal text as well as scored by each of the 52 Level 2 geopolitical regions of the Taxonomic Databases Working Group (TDWG) world geographical scheme for recording plant distributions (Brummitt 2001; Table 1; Fig. 1). Distributions have been compiled primarily from records of herbarium specimens held at Kew, and these have been extensively supplemented with additional literature records from standard regional Floras (reviewed by Frodin 2001), from major taxonomic monographs, and from innumerable revisions of individual genera. Of 70,550 distribution records, 2,560 (3.63%) are from literature sources only, while very many individual records that were represented either by only a few or by doubtfully-determined specimens were further corroborated with literature records.

The Kew Herbarium is therefore judged to be comprehensive enough to give accurate distributions at this taxonomic and geographical scale, since a single reliably-identified specimen from a particular region is sufficient to stand as a distribution record. Inevitably there is some geographical bias in the numbers of specimens held (it is better represented in the former tropical British colonies and areas of more recent activity by Kew) – however, over

97% of all accepted genera are represented by at least one or some specimens, and over 96% of distribution records were supplied by specimen data. This database is updated on an almost-weekly basis, with new genera added and existing taxonomy revised in line with major new publications. Additional new generic names, which are published at the rate of about 110 per year, are entered either as accepted genera or as synonyms largely on the advice of herbarium staff at Kew. Distributions of genera are then modified as necessary in the light of these taxonomic changes, and compiled again from re-curated herbarium specimens and/or new publications as appropriate.

The database therefore contains a presence/absence data matrix of 52 regions x 14,724 genera. Collectively, the 52 regions are an amalgam of multiple small countries, or groups of islands, or multiple states within single countries, or single countries in their entirety (Fig. 1) – all are defined by hard political boundaries, and consequently none are of exactly equivalent size or shape. The geographic resolution is obviously extremely coarse, yet nevertheless this database represents a rich source of biogeographic data. Whilst many detailed local and regional Floras have been compiled, and many detailed distribution records compiled and biogeographic studies performed for particular taxa or areas, these have all been done individually and independently – there is no comparable comprehensive data source on a global scale. The main strength of this database is thus the comprehensive treatment of each genus in a comprehensive, standardised way.

Three obvious criticisms can be levelled at the design of the database and which may be thought detrimental. Firstly, the regions by which distributions have been scored are very large compared to the actual distributions of most taxa. Secondly, the regions have geo-political boundaries, which generally do not follow

any recognised biogeographical, ecological or climatic pattern. Thirdly, the regions vary greatly in size, from just over 200 square kilometres (Middle Atlantic Ocean – the islands of St. Helena and Ascension) to well over 12 million square kilometres (Antarctic Continent). However, even at this coarse scale, it took two years for one person fulltime to compile the whole database; the problems with compiling, and maintaining, more detailed geographical information for many poorly-known taxa would have been very much greater. Further work aims to capture genera with small distributions at finer geographical scales.

The second of these drawbacks may in fact be a strength of the database. The aim is to establish what are the distribution patterns – any biogeographic, ecological or climatic interpretations of these patterns must come secondarily, without this having a priori influenced the compilation of the data. If the data had instead been scored by pre-determined biomes or floristic regions, any subsequent analysis of the reality or authenticity of those regions would then be hopelessly circular since the existence of such areas would have been an assumption behind the compilation of the data in the first place. Geo-political boundaries, for all their arbitrariness, are much more sharply defined than are biomes or vegetation types or floristic regions: a plant either is in Brazil or it is not, and herbarium specimen labels which may well omit the vegetation type or the biogeographic region will almost certainly note the country. It was decided at the outset that distributions should be compiled from specimen records – literature records are simply not comprehensive enough to record distributions in a systematic way, whereas the Herbarium at Kew houses at least some material of more than 97% of accepted genera. This decision was borne out in compiling the data, when it was usually additional specimen records which were

found to augment distributions quoted in the literature, and not vice versa.

Notwithstanding the interest in the biogeography of pteridophytes and gymnosperms, in this paper all the analyses presented will refer solely to angiosperms (14,304 genera). Angiosperms represent a highly diverse crown group of established monophyly (Soltis *et al.* 1999; Savolainen *et al.* 2000) and increasing species number (Niklas 1988; Niklas & Tiffney 1994), rather than a formerly more diverse group now restricted in its size. Furthermore, many pteridophyte genera show highly scattered and irregular distributions when compared with angiosperm genera, presumably due to their spores being easily wind-dispersed, which makes the interpretation of the kind of analyses presented here considerably more difficult.

Patterns of genus richness

Accounting for area

The obvious differences in size within the set of TDWG regions (Fig. 1) complicate the initial comparison of their absolute taxonomic diversities. The first step in this study is therefore to examine the relationship between the size of an area and the number of species present in that area, which is so ubiquitous that it has been described as ‘one of the most important patterns in biogeography’ (Lomolino 1989). Although the exact form of the relationship is still not universally agreed upon (Connor & McCoy 1979), the relationship is conventionally expressed by the model

$$S = cA^z,$$

where S = number of species, A = area and c and z are fitted constants, usually interpreted as the intercept and the slope of the relationship on a double-logarithmic plot, respectively (Rosenzweig 1995).

Table 1. The 52 TDWG Level 2 regions, with counts of generic richness, size and relative generic richness after accounting for area given for each region; relative generic richness is represented on an arbitrary scale between 0 and 1, relative to the richness of the richest region, Western South America (83).

	TDWG Region	No. of genera	Area (km²)	Relative no. of genera
10	Northern Europe	578	1620215	0.24
11	Middle Europe	708	1080825	0.31
12	Southwestern Europe	1030	1158137	0.45
13	Southeastern Europe	1094	1056193	0.48
14	Eastern Europe	807	4690405	0.29
20	Northern Africa	972	5713373	0.34
21	Macaronesia	493	13998	0.40
22	West Tropical Africa	1574	6063089	0.75
23	West-Central Tropical Africa	1856	4120263	0.68
24	Northeast Tropical Africa	1582	5690458	0.56
25	East Tropical Africa	1861	1773068	0.77
26	South Tropical Africa	1807	3298998	0.68
27	Southern Africa	1874	2675954	0.72
28	Middle Atlantic Ocean	44	232	0.07
29	Western Indian Ocean	1404	602411	0.68
30	Siberia	610	9855164	0.20
31	Russian Far East	576	3063473	0.22
32	Central Asia	922	3974323	0.34
33	Caucasus	877	438212	0.45
34	Western Asia	1485	3845970	0.54
35	Arabian Peninsula	891	2789669	0.34
36	China	2438	9268483	0.79
37	Mongolia	480	1558842	0.20
38	Eastern Asia	1342	628999	0.63
40	Indian Subcontinent	2595	4423703	0.94
41	Indo-China	2096	1932370	0.85
42	Malesia	2174	2128984	0.86
43	Papuasia	1557	906147	0.70
50	Australia	1892	7704687	0.63
51	New Zealand	352	268760	0.19
60	Southwestern Pacific	910	57339	0.61
61	South-Central Pacific	265	4081	0.26
62	Northwestern Pacific	341	2640	0.35

	TDWG Region	No. of genera	Area (km ²)	Relative no. of genera
63	North-Central Pacific	242	16920	0.18
70	Subarctic America	336	7526750	0.11
71	Western Canada	521	2899891	0.21
72	Eastern Canada	521	3111341	0.20
73	Northwestern U.S.A.	715	1544658	0.30
74	North-Central U.S.A.	699	1842281	0.28
75	Northeastern U.S.A.	639	968634	0.28
76	Southwestern U.S.A.	1085	1207900	0.46
77	South-Central U.S.A.	1030	1000684	0.45
78	Southeastern U.S.A.	1023	1371180	0.43
79	Mexico	2236	1961910	0.90
80	Central America	1955	519581	0.95
81	Caribbean	1545	235202	0.85
82	Northern South America	2069	1355352	0.88
83	Western South America	2731	3783883	1.00
84	Brazil	2519	8506150	0.83
85	Southern South America	1714	4104401	0.62
90	Subantarctic Islands	90	24542	0.07
91	Antarctic Continent	2	12093000	0.00

A log-log graph of numbers of families, genera and species for each region against the area of that region is presented in Fig. 2; family- and genus-level data from the Vascular Plant Families and Genera database (absolute numbers of genera are given in Table 1) has here been augmented by species-level data extrapolated from figures provided by Rafaël Govaerts (cf. Govaerts 2001, 2003). At all taxonomic scales, the relationship between diversity and area is roughly linear in log-log space. The data and regression statistics presented in Fig. 2 exclude the data point for the Antarctic Continent, a recognised TDWG region of more than 12 million square kilometres but with only two native angiosperm genera (*Colobanthus* Bartl. in the Caryophyllaceae and the grass genus *Deschampsia* P.Beauv.). Including data for Antarctica

considerably reduces the slope of each regression line but without being informative of global diversity patterns.

Looking at Fig. 2, there is an obvious cluster of points towards the right hand edge. This reflects the efforts made in the TDWG Geographical Scheme to divide the world up into equal-sized political subdivisions, as far as is possible; most areas are within one order of magnitude, from 1 to 10 million square kilometres. At each taxonomic rank, the spread of points around the line is due to latitude. Tropical areas are spread along the top, in a gently arcing line, from small tropical island systems (Middle Atlantic Ocean [St. Helena and Ascension I.], 28) to large continental areas (Brazil, 84); the arcing is due to the geographically scattered (non-nested) relationships between



Fig. 1. Map of TDWG Level-2 regions following Brummitt (2001), as used in this study, with their 2-digit codes; regions are listed in Table 1.

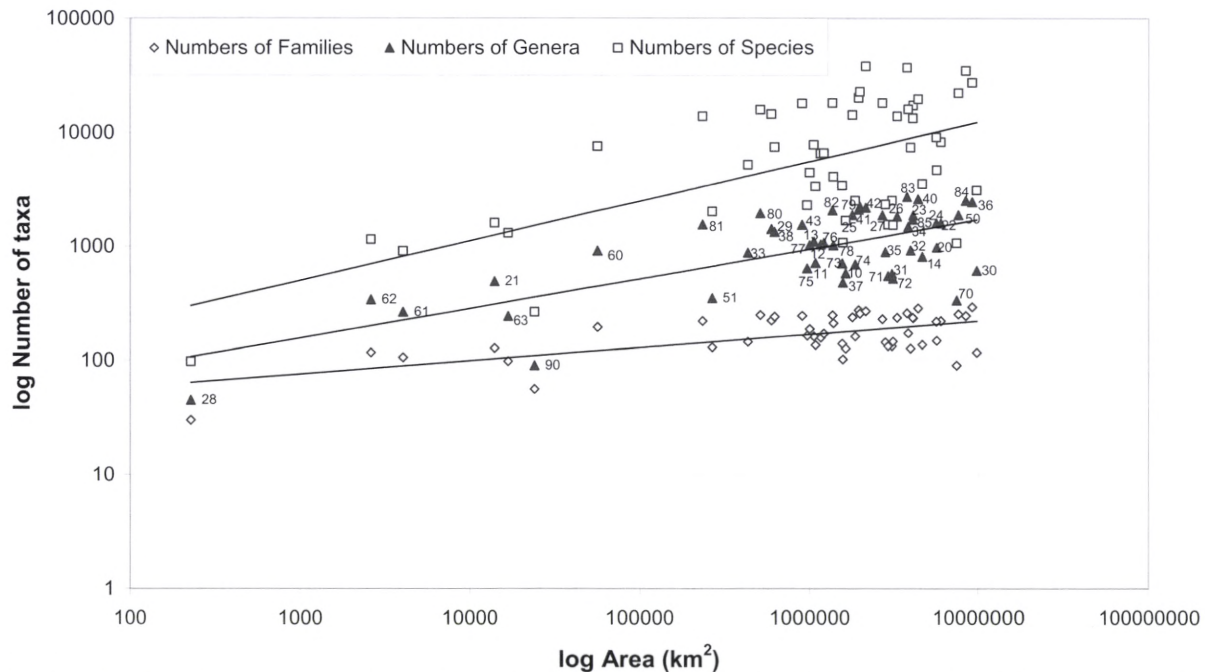


Fig. 2 Relationship between area and diversity for families (open triangles; $r^2 = 0.37$), genera (closed diamonds; $r^2 = 0.48$) and species (open squares; $r^2 = 0.36$) worldwide; regions are identified by their 2-digit code. Graph and regressions do not include data for the Antarctic Continent (91).

the regions (Rosenzweig 1995). Below that, there is a steady increase in latitude to the areas farthest from the equator (Siberia, 30; Subarctic America, 70), which are the least diverse continental areas. This can also be seen among the smaller island regions: the Subantarctic Islands (90) are much less diverse than comparably-sized tropical islands (North-Central Pacific [Hawaiian Islands], 63); the same is true for New Zealand (51) versus the Caribbean (81).

It is possible to detect geographic factors within the distribution of tropical regions. Neotropical areas are most diverse, with tropical Asian areas slightly below and African regions slightly below them (see also Fig. 3). Regions both directly north (34, 36, 38) and south (27, 50, 85) of the tropics have greater diversity than would be expected purely from their latitude; this is due to 'spill over' of essen-

tially tropical genera found in geographically adjacent regions. Below that, there is a band of 'warm-temperate' regions which is clustered in two distinct groups: southern Europe (12, 13) with southern U.S.A. (76, 77, 78), and Northern Africa (20), Middle Asia (32) and Arabian Peninsula (35) with Eastern Europe (14). In the case of Northern Africa and the Arabian Peninsula, this is presumably due to great aridity over most of these areas suppressing the diversity expected at that latitude (much lower than Eastern Europe and Middle Asia); however, it may be argued that it is the aridity itself which is expected at that latitude, and not high diversity (Archibold 1995). Below these clusters is a diffuse group of north temperate regions, both Old and New World (10, 11, 30, 31, 37, 70, 71, 72, 73, 74, 75).

Comparing the three taxonomic scales, patterns of relative diversity within a region are

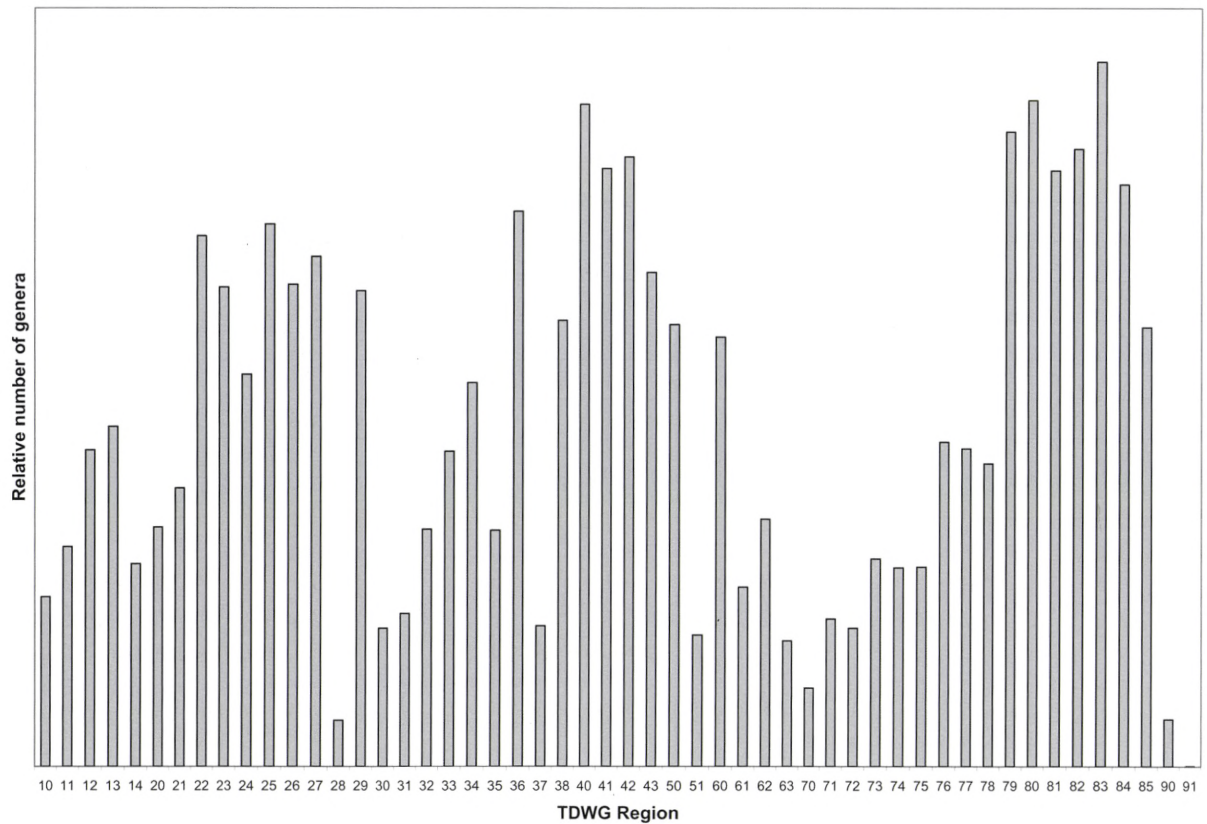


Fig. 3. Re-scaled generic richness for TDWG regions, independent of the relative sizes of regions. Generic richness is presented as an arbitrary scale on the y axis relative to that of the richest region (83); regions on the x axis are identified by their 2-digit code. Graph does not include data for the Antarctic Continent (91).

highly correlated, irrespective of the sizes of the regions (Spearman's r_s : between species richness and genus richness, 0.97; between genus richness and family richness, 0.96; between species richness and family richness, 0.93; $n = 52$, $p < 0.01$ in each case), confirming that higher taxonomic levels may indeed be a reasonable surrogate for species richness (Gaston 1996; La Ferla *et al.* 2002; but see also Prance 1994). The spread of points in Fig. 2 obviously increases from family- to genus- to species-level, with the exponent of the species area relationship likewise increasing from 0.12 to 0.26 to 0.35, respectively. The increasing

exponent values simply reflect the structure of the taxonomic hierarchy: there can be several species within one genus or genera within one family, but not vice versa; therefore the numbers of taxa will inevitably increase with decreasing taxonomic rank. However, the increasing spread of data points with decreasing taxonomic rank reflects the increase in the proportion of tropical taxa at lower ranks: there are not only more tropical species than temperate species, or tropical species than tropical genera, but there are more tropical species compared with temperate species than there are tropical genera compared with tem-

perate genera. That is to say, the strength of the latitudinal gradient of diversity increases with decreasing taxonomic rank.

Richness rescaled for area

Since differences in the absolute taxonomic diversity between regions are greatly influenced by differences in their size, a first task is therefore to first standardise the number of taxa within regions by the area of those regions; then relative diversity may be compared independently of the confounding influence of area. Simply dividing species number by area to give unscaled species-area ratios leads to highly spurious results (Connor & McCoy 1979; Brummitt & Nic Lughadha 2003) – they have the effect of over-estimating the diversity of small areas. In dividing number of species by area there lies an implicit assumption of a simple linear relationship between area and diversity, whereas ecologists have long known that such a linear relationship does not exist (Arrhenius 1921; Williams 1943; MacArthur & Wilson 1967). As already stated above, and shown in Fig. 2, the correct relationship is the power function model of $S = cA^z$. So if $S = cA^z$, then $c = S/A^z$, and thus the constant c is the ratio of diversity (S) to A^z (Rosenzweig 1995). So in order to obtain realistic scores of relative diversity, area needs to be scaled by a suitable exponent value (z), and relative values of c for areas of different size are then calculated with S/A^z . In this study an exponent (z) value of 0.14 was chosen as being within the range of empirical z values given for non-isolated island or mainland areas (0.12 – 0.17) by MacArthur and Wilson (1967), and the results from this analysis then compared with those from a comparable analysis with $z = 0.25$, a value both within the range of values for isolated mainland regions (0.20–0.35; MacArthur & Wilson 1967) and also predicted theoretically (Preston 1962; Brown *et al.* 2000). The results for $z = 0.14$ and for $z = 0.25$ were

highly correlated (Spearman's r_s 0.97; $n = 51$, $p < 0.01$); scores for relative generic richness when $z = 0.14$ are given in Table 1 and Fig. 3.

Several salient points can be inferred from Fig. 3. Firstly, three areas of tropical diversity can be clearly made out, for Africa (+ Madagascar), SE. Asia and the Neotropics, increasing in that order. The western Pacific is also markedly diverse, though this is only apparent from the rescaled data: the small areas of these regions mask their true diversity. Comparison between absolute and re-scaled rankings of China (36) and Indian Subcontinent (40) shows that very high genus richness scores for these regions have been inflated by their huge size. In addition, West Tropical Africa (22), West-Central Tropical Africa (23) and Northeast Tropical Africa (24) are all of considerably lower relative diversity than are other tropical regions; in fact comparable to southern Europe (Southwestern Europe, 12; Southeastern Europe, 13), to SW Asia (Caucasus, 33; Western Asia, 34) and to southern U.S.A. (Southwestern U.S.A., 76; South-Central U.S.A., 77; Southeastern U.S.A., 78). Again, this is masked in the absolute genus richness scores for these regions. This surprising result may be partly explained by the almost-barren Sahara Desert covering large expanses of West Tropical Africa (22) and Northeast Tropical Africa (24), giving lower generic diversities than would be expected for regions of that size.

For the Neotropical regions a similar story is revealed: in absolute terms it appears that the large regions Mexico (79) and Brazil (84) are very diverse in numbers of genera (Fig. 2); in fact, however, it is clear from Fig. 3 that Western South America (83) is the most diverse region of all for both absolute and relative numbers of genera, followed by Central America (80). The two smallest Neotropical regions, Central America (80) and Caribbean (81), which do not appear to be particularly diverse in terms of absolute numbers of genera, are

revealed as surprisingly diverse when numbers of genera are scaled by area, and four of the five most diverse regions are in the Neotropics, (Western South America, 83; Central America, 80; Mexico, 79; Northern South America, 82, in descending order of relative diversity) the exception being the Indian Subcontinent (40).

Latitudinal gradient of diversity

The latitudinal gradient of diversity has been described as “one of the most prominent features of the natural world” (Rohde 1992; Taylor & Gaines 1999), but also as (still) “the major, unexplained pattern in natural history” (R. E. Ricklefs, quoted in Lewin 1989). For, despite early descriptions of tremendous tropi-

cal diversity by such explorers as von Humboldt and Wallace (Whitmore 1990), and decades of subsequent ecological research, an adequate understanding of the cause(s) of the latitudinal gradient of diversity remains lacking. Greater numbers of species are found in the tropics for almost all groups of organisms and in almost every type of habitat, both terrestrial and aquatic, and so, if we are searching for a comprehensive understanding of the latitudinal gradient, the explanation must then apply equally well to nearly all organisms and environments – something which has not always been the case for different groups (Rohde 1992).

Fig. 4 plots genus richness of each TDWG region against minimum distance from the

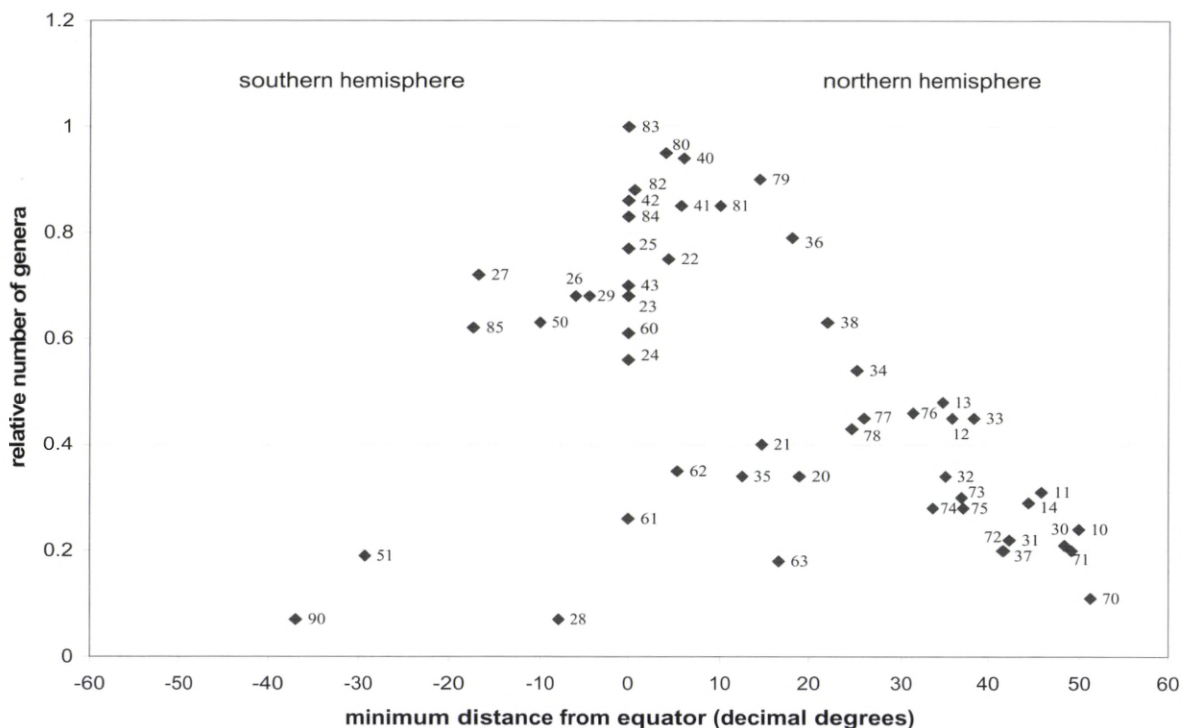


Fig. 4. The latitudinal gradient of generic diversity for TDWG regions, with the regions plotted by minimum distance from the equator. Generic richness is presented as an arbitrary scale on the y axis relative to that of the richest region (83), and independent of sizes of regions; regions are identified by their 2-digit code. Graph does not include data for the Antarctic Continent (91).

equator of that region (taken as minimum latitude of that region). Since any genus record will be included for that region if it barely crosses over the border, and diversity is known to decline away from the equator, this should have the effect of under-estimating the latitudinal gradient of diversity, if anything. The values plotted in Fig. 4 are re-scaled genus richness for regions against latitude; both northern and southern hemispheres are shown. Re-scaled genus richness values, calculated after correction by the species-area relationship (assuming $z = 0.14$), standardise for the effects of different sizes of regions. It is clear that there is a strong relationship between the latitude of a region and its genus richness. Thus the latitudinal gradient in diversity evident from Fig. 4 is independent of any considerations of land area.

The circumference of the scatter graph is roughly parabolic: diversity increases at the same rate in either hemisphere towards the equator. Regions falling below the apex of the distribution (that is, not as diverse as would be expected) can mostly be explained on a case-by-case basis. Northern Africa (20) and Arabian Peninsula (35) are both covered by extensive deserts; their low diversity is surely a product of their excessive aridity. Middle Atlantic Ocean (28), South-Central Pacific (61) and North-Central Pacific (63) are all isolated island systems. However, the low diversities of West Tropical Africa (22) and Northeast Tropical Africa (24) remain difficult to explain, excepting that the Sahara Desert also occupies large expanses of the northern portions of these regions.

Patterns of generic endemism

Patterns of generic diversity and patterns of generic endemism are less well correlated than is diversity at different taxonomic ranks (Spearman's r_s 0.80, $n = 46$, $p < 0.01$) Though all the

regions with highest genus richness have moderate degrees of generic endemism, those regions with the highest degree of endemism (Southern Africa, 27 [31%]; Western Indian Ocean, 29 [29%] and Australia, 50 [35%]) have themselves moderate genus richness. It is notable that amongst tropical regions, no tropical African region has generic endemism greater than 10%; indeed West Tropical Africa (22); East Tropical Africa (25); and South Tropical Africa (26), all have levels of generic endemism (2.2%, 2.8% and 1.9%, respectively) lower than that for Southwestern Europe (12; 3.0%). In SE Asia and the Neotropics, on the other hand, Malesia (42; 10.8%) and all of Mexico (79; 11.8%); Caribbean (81; 12.6%); Western South America (83; 10.8%); Brazil (84; 15.7%); and Southern South America (85; 13.6%) have values for generic endemism greater than 10%. Assuming still that patterns in the distribution of genera truly reflect underlying patterns in species distribution, this implies that levels of speciation have been far greater in extra-African tropical regions, and in the Neotropics in particular (Richardson *et al.* 2001), unless extinction of whole genera has been greater in Africa than in both SE Asia and the Neotropics.

Amongst temperate regions, some (Central Asia [32; 6.0%]; Western Asia [34; 6.6%]; Southwestern U.S.A. [76; 8.4%]) have levels of generic endemism greater than many tropical regions (all of tropical Africa; Indian Subcontinent [40; 6.1%]; Papuasiasia [43; 5.7%]; Central America [80; 3.7%]; and Northern South America [82; 4.7%]). Other, cold-temperate regions, however, have absolutely no endemic genera (Northern Europe, 10; Middle Europe, 11; Western Canada, 71; Eastern Canada, 72; North-Central U.S.A., 74; and, not surprisingly, the Antarctic Continent, 91) – and Eastern Europe (14) and Subarctic America (70), have only a single one apiece. Several isolated island regions (Middle Atlantic Ocean, 28; New

Zealand, 51; North-Central Pacific, 63) have moderate degrees of endemism but relatively low genus richness. In fact, Middle Atlantic Ocean (by far the smallest TDWG Region at only 232 km², and at least one order of magnitude smaller than the next smallest region), has a remarkable 11 endemic genera out of only 44 native angiosperm genera. However, conclusions from patterns of generic endemism remain tentative, since these genera are likely to be subject to future taxonomic change. Many endemic genera, particularly from isolated oceanic islands (*e.g.* St. Helena) or ecological islands (*e.g.* mountain tops) in temperate regions, contain only one or a few species marked by strong morphological adaptation to isolated, localised, often extreme, environments. The monophyletic status of many of these genera, the small temperate endemic genera in particular, has not yet been assessed; they may be shown to nest within other genera.

The potential number of possible distribution patterns

With 52 regions, and any individual genus able to occur or not occur in each of those regions, there will be a total of 2^{52} possible unique combinations of those regions making up the complete set of distribution patterns. The maximum distribution range is obviously a genus present in every region (although this is not actually shown by any genus); the minimum distribution is not those genera endemic to a single region, but a null distribution absent from every region, since in theory this is also a potential distribution. However, since recently-extinct genera are here treated as native in their former range, and intergeneric hybrids have been excluded from this analysis, in practice no genus actually shows this null distribution. Assuming therefore that we are not interested in the single empty (null) distribution

pattern containing no genera, then this still leaves $2^{52}-1$ potential generic distribution patterns – or more than 4.5×10^{12} possible distributions! Fortunately, for angiosperm genera only 2817 separate combinations of regions are actually found. This huge reduction in actual versus potential numbers of distribution patterns reveals that there must be a high degree of structure to the total set of genus distributions, and that some distribution patterns must be shown by many genera.

Of these 2817 distributions, c. 2200 are unicate (shown by only one genus); only c. 600 distributions are shown by more than one genus, but collectively these 600 distributions account for about 11000 genera. Respectively, therefore, only just over 20% of distribution patterns account for just over 80% of genera, while just under 80% of distribution patterns account for only about 20% of angiosperm genera. That is to say, the frequency distribution of genus distribution patterns is extremely right-skewed, with a modal (most common) value of only one genus per distribution pattern. This ‘hollow-curve’ frequency distribution (Willis 1922; Williams 1964) of distribution patterns is also shown by the frequency distribution of range-sizes, a pattern found in many groups (Colwell & Lees 2000; Gaston 2003). Some 38% of genera are endemic to a single region, but there is no genus found in every region, and only three genera (*Carex* L. and *Cyperus* L., both Cyperaceae; and *Plantago* L., Plantaginaceae) found in every region except Antarctica.

Floristic relationships

What are these repeating distribution patterns which account for such a high proportion of the total diversity of angiosperm genera? Since the total set of endemic distributions only numbers 46 separate regions (endemic genera are not found in all regions), the majority of

the 600 distribution patterns must therefore be shared between different regions. As a first step towards analysing the distribution patterns themselves, the strength of floristic relationships between regions was investigated using both clustering and ordination techniques. Firstly, all records from the Antarctic Continent (91) were again excluded, since the only two native genera are very widespread outside of Antarctica. Genera occurring in only one region were also excluded from the analysis, since they cannot be informative of relationships between regions but will increase dissimilarity between regions. Floristic relationships between regions were investigated using hierarchical cluster analysis by UPGMA and flexible beta clustering. Beals smoothing (Beals 1984) was applied to the data to draw out the structure within the matrix and reduce the associated noise for ordination analysis by non-metric multidimensional scaling (Kruskal 1964a, 1964b), again using the Sørensen similarity coefficient, to group regions in a non-hierarchical framework. All analyses of floristic relationships were carried out using PC-ORD Version 4.0 (McCune & Mefford 1997).

Clustering

Cluster analysis was undertaken with the UPGMA linking algorithm, using both Sørensen's and Jaccard's similarity coefficient (to assess the influence of selecting the coefficient); clustering was also undertaken with the flexible beta linking algorithm, using $\beta = -0.25$ (which emulates Ward's linking method for Euclidean distance measures) and $\beta = 0$. The dendrogram for UPGMA clustering with Sørensen's coefficient is given in Fig. 5 (lack of space precludes showing each of the other three dendrograms). There is strong agreement in the topologies of all four dendrograms; the differences between them are minor. Indeed, there is no difference at all in the topologies of the dendrograms between UPGMA clus-

tering with Sørensen's coefficient and UPGMA clustering with Jaccard's coefficient; the only difference is in the percentage information remaining, with Jaccard's coefficient consistently accounting for more information with each stage of the clustering process. This is because, whereas Sørensen's similarity coefficient is twice the shared abundance divided by total abundance, Jaccard's similarity coefficient is the shared abundance divided by the total non-shared abundance, creating greater distances between groups (McCune & Grace 2002).

Ordination

Ordination by non-metric multidimensional scaling, after transformation with the Beals smoothing function, was carried out under the 'Autopilot' mode in PC-ORD. Since non-metric multidimensional scaling is a re-iterative technique which converges on an optimal solution, the 'Autopilot' function sequentially calculates each successive stage of the analysis, stepping down in dimensionality before re-running a final ordination with the co-ordinates from the optimal number of dimensions as its starting co-ordinates (McCune & Grace 2002). The analysis was run with the following settings: Distance measure = SØRENSEN; Number of axes (max. = 6) = 6; Maximum number of iterations = 400; Starting coordinates (random or from file) = RANDOM; Reduction in dimensionality at each cycle = 1; Step length (rate of movement toward minimum stress) = 0.20; Random number seeds (use time vs. user-supplied) = USE TIME; Number of runs with real data = 40; Number of runs with randomized data = 50; Autopilot = YES; Stability criterion, standard deviations in stress over last 15 iterations = 0.00001; Speed vs. thoroughness = THOROUGH.

The final solution found by NMDS reached a low stress value of 2.47, with an instability value of 0.00001. These values are both remarkably low, and indicate a very strong ordination of highly structured data. Stress value, instability

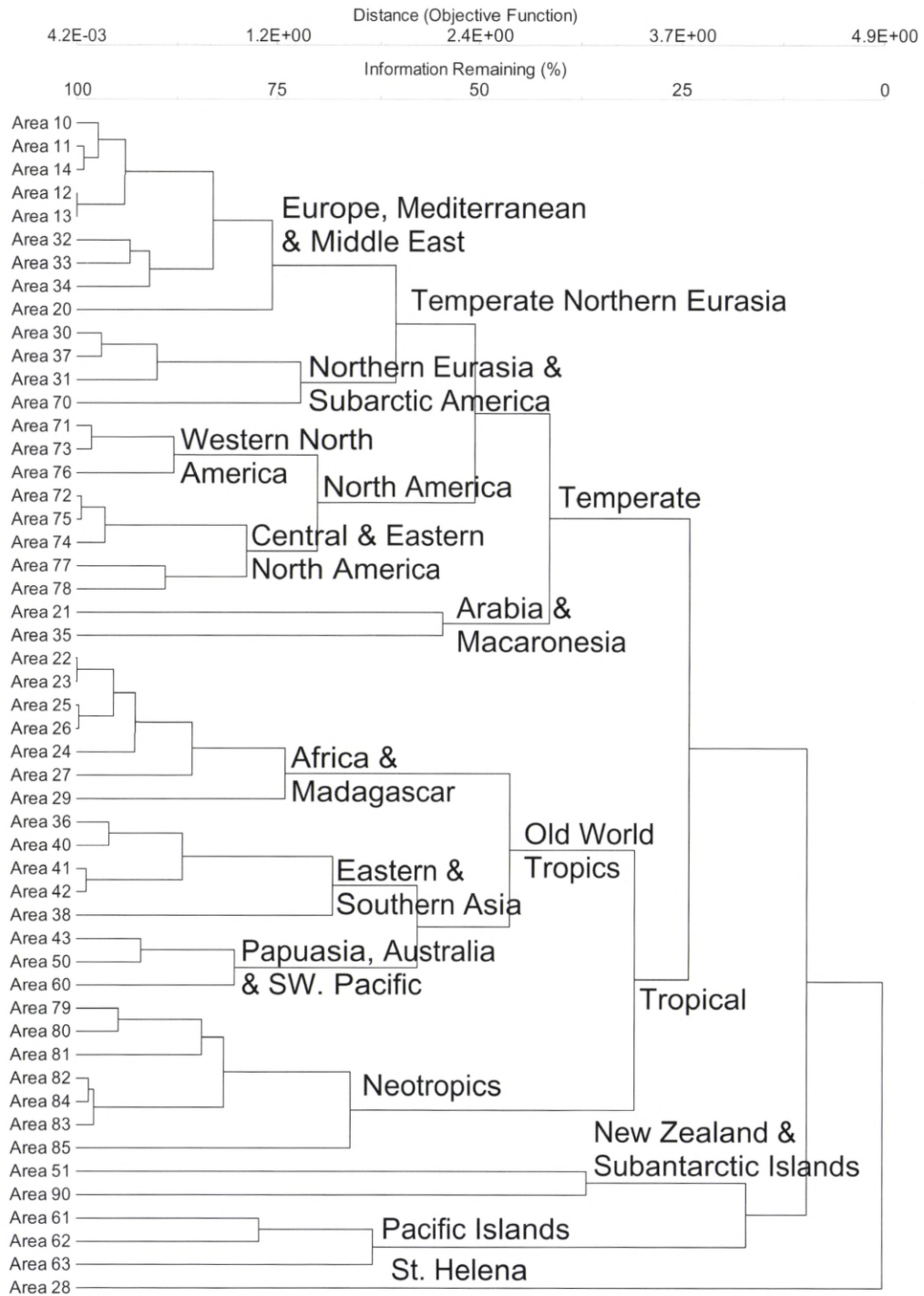


Fig. 5. Dendrogram of floristic relationships for TDWG regions; Sørensen similarity coefficient, UPGMA clustering, scaled by Wishart's objective function.

Table 2. Stress value, instability value, number of iterations and percentage of variance accounted for by each axis of the final best run of the non-metric multidimensional scaling ordination of floristic relationships between regions.

Stress value	Instability value	No. of iterations	Axis 1 (r^2)	Axis 2 (r^2)		Axis 3 (r^2)	
			increment	increment	cumulative	increment	cumulative
2.47	0.0001	80	0.51	0.36	0.87	0.11	0.98

value and the proportion of variation explained by each axis (r^2 value) are given in Table 2. Although the amount of variation explained by the first axis is very small (only 6%), 98% of variation has been accounted for after three axes have been constructed. The plot of iteration vs. stress (not shown here) demonstrates that as the number of iterations increases beyond 20, stress suddenly drops very sharply until levelling off at beyond about 40 iterations. By about 65 iterations there is no further decline in stress, and the analysis terminated at 80 iterations. Clarke (1993) recommended an ideal stress value of about 5, although McCune and Grace (2002) claim this is 'rarely achieved' with ecological data. The value of 2.47 obtained here, however, is almost equal to the indication of the 'excellent' quality of non-metric multidimensional scaling ordination proposed by Kruskal (1964a) of 2.5. Similarly, the instability value obtained here is below that recommended by McCune and Grace (2002) of 0.001.

From the NDMS ordination five large, well-defined groups are apparent (see Fig. 6), which are: an 'Africa and Madagascar' group, a distinct 'Neotropical' group, a diverse group of eastern and southern Asian regions, then a 'North American' group, and lastly there is a group of 'Temperate Northern Eurasian' regions. Several regions, however, do not fall neatly into any of these groups: Middle Atlantic

Ocean (28), Arabian Peninsula (35) and Subantarctic Islands (90). These strong continental clusters of regions, which are in broad agreement in both the clustering and ordination analyses, are a product of the highly-skewed frequency distribution of distribution patterns: the majority of generic distributions are within individual continents. The five groups recovered by non-metric multidimensional scaling correspond largely to the continental groups found by UPGMA cluster analysis at the level of 50% similarity or less. The 'Africa and Madagascar' group and the 'Neotropical' group are both identical in the two analyses. In the ordination, Subarctic America (70) shows greater similarity to other North American regions than it does in the cluster analysis, but otherwise this North American group is identical in composition also. Macaronesia (21), which grouped with northern temperate regions with cluster analysis, and Arabian Peninsula (35) group more strongly with the 'Temperate Northern Eurasia' group under ordination. The biggest single difference between the two analyses is with the position of the Pacific regions (61, 62, 63), which lie peripherally to the large group of eastern, tropical and Australasian regions by ordination analysis but as a distinct group with other 'island' regions by cluster analysis. However, some differences are to be expected, since the 3-dimensional, non-hierarchical

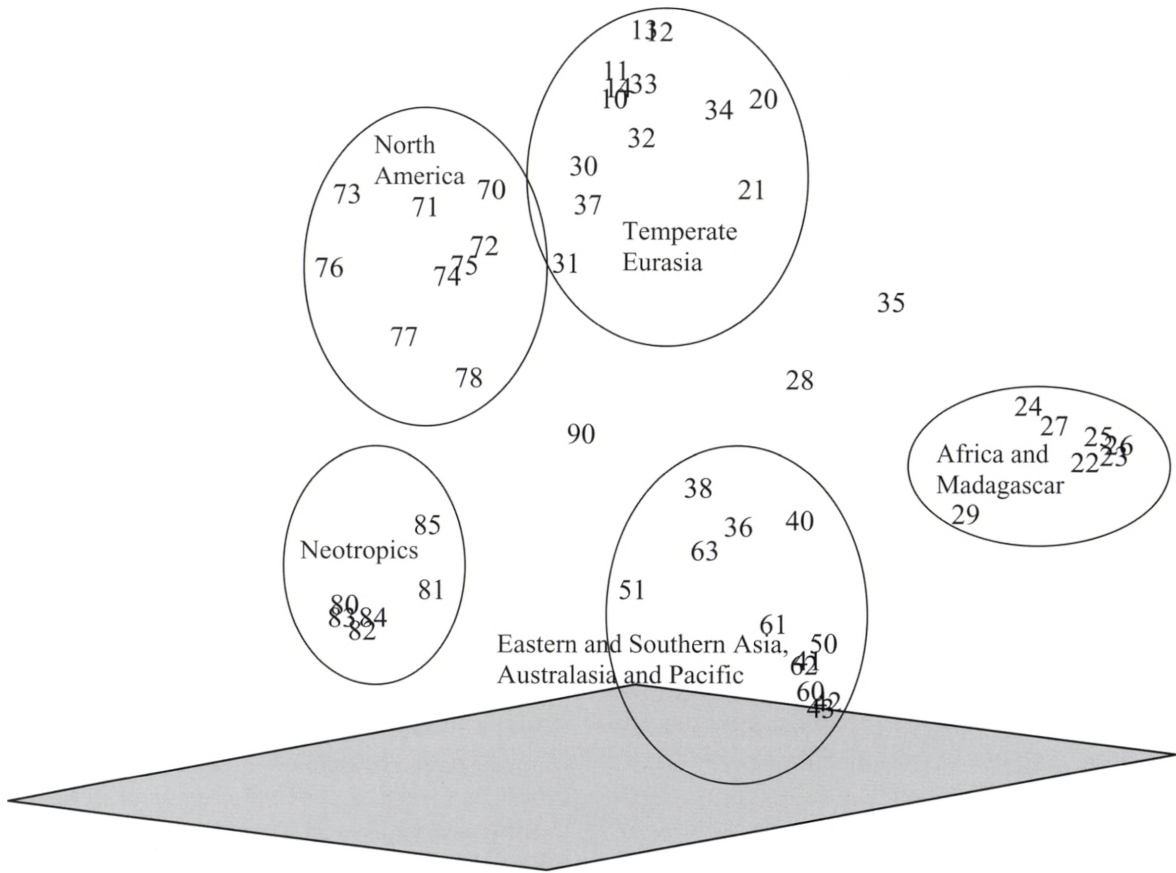


Fig. 6. Non-metric multidimensional scaling ordination diagram from analysis of floristic relationships; five broad groups are evident.

framework of the ordination reveals subsidiary floristic links between regions which are lost when the results are constrained onto a 2-dimensional dendrogram, which can only show the relationships of maximum similarity. The ordination results therefore more truly reflect the complexity of the underlying distribution patterns, but are correspondingly less easy to interpret.

Analysing distribution patterns

The strength of floristic relationships between regions of the same continent suggests that these common, repeating distribution patterns

are themselves confined to areas within single continents, an idea reinforced by the shape of the range-size frequency distribution. Returning therefore to the bigger question suggested by the frequency distribution of genus distribution patterns, we can ask: what actually are these distribution patterns, and how frequently are they found; and, furthermore, is there an optimal number of different distribution patterns for all angiosperm genera? However, defining just what is meant by 'different distribution patterns' proves to be no easy task. For a heterogeneous collection of overlapping distribution patterns, where should the discontinuities between them be drawn? For example,

if widespread distributions differ only in the presence or absence of a genus in a single extra region, then can these separate distribution patterns be treated as essentially 'the same' distribution pattern? If so, then what about a distribution pattern which differs from this one by only one region, and then where should the line be drawn between 'different' distributions when, for the totality of the data, there is so much overlap? If not, then are we left with having to deal with all 2817 unique distributions? If the latter, then we have made little progress with trying to analyse distribution patterns themselves, the actual data underlying all of the previous analyses, and questions of plant distribution will therefore remain essentially intractable.

In order to tackle this question, a non-hierarchical clustering approach was implemented using k-means partitioning (Lance & Williams 1968; Bailey & Gatrell 1995; Legendre & Gallagher 2001), a technique for dividing a set of values into a pre-determined number of groups based on the similarity between those values. The main advantage of using this technique was in the lack of structure imposed on the data: forcing overlapping distributions into a discrete hierarchical framework, such as by UPGMA clustering, had previously caused the arbitrary resolution of similar disjunct distributions into one or another, but seldom the same, group; this artefact was avoided within a non-hierarchical framework. Since the number of partitions being sought was unknown beforehand, the chosen partition was that with the maximum value of the Calinski-Harabasz pseudo-F-statistic (Calinski & Harabasz 1974). This statistic is in effect a ratio between the within-group homogeneity of all groups and the between-group heterogeneity of all groups; in simulation studies this is the optimality criterion which has been found to most closely return the optimum number of groups (Milligan & Cooper 1985).

However, since k-means partitioning can only be applied to values within a Euclidean geometric space, distributions first needed to be transformed from categorical data in a non-Euclidean geometric space in 52 dimensions (the raw data matrix) to continuous data in a lesser number of dimensions in Euclidean space, for example by ordination. Genera endemic to a single region were first excluded from the analysis, on the assumption that each endemic distribution would remain a distinct pattern from the others, while speeding computation time considerably. After then running Beals' smoothing, a non-metric multidimensional scaling ordination was used; ordination settings were the same as for ordination of regions, except that this time the thoroughness setting was reduced to 'MEDIUM'. The k-means partitioning analysis was then carried out (using a program downloaded from <http://www.fas.umontreal.ca/biol/legendre/indexEnglish.html>) on the taxon scores for the ordination axes, and group membership was outputted as simple ASCII text files, which could then be imported back into the Vascular Plant Families and Genera database to inspect the geographical distributions of each cluster.

Since the initial k-means partition was into only a small number of geographically-diverse groups, this stage was implemented recursively, saving the partition (number of groups) with the maximum value of the Calinski-Harabasz pseudo-F-statistic from each pass and subsequently running separate partitions on each of these groups. The initial assignment of objects to groups was at random, with 100 replicates of each random assignment for each partition; since the input data were the ordination scores, these were treated as unweighted and unstandardized, the ordination itself already having standardised the raw data. A non-parametric test of statistical significance (a multi-response permutation procedure) was applied to each pair of groups formed (Biondini *et al.*

1988), and the next round of k-means partitioning undertaken until further groups formed failed the test of statistical significance between group membership (*i.e.* they had become too similar to tell apart).

Ordination results

The final solution found by the non-metric multidimensional scaling was a 3-dimensional ordination; this gave the optimal reduction in stress – with additional dimensions there was little subsequent reduction. Although a decline in stress was seen until 30 iterations, the degree of stress showed no sign of stabilising before the maximum number (200) of iterations was reached. Ordination statistics for the final solution are given in Table 3: despite the continuing fluctuations in stress, the final solution

does seem to be stable, with a low instability value of 0.028. The stress value for the final solution is 9.76, still below 10 and so with ‘no real risk of drawing false inferences’ (Clarke 1993), and still below stress values of between 10 and 20 quoted by McCune and Grace (2002) as typical for ecological community data. Each axis represents roughly equal proportions of the variation, with a total cumulative variance of 91% explained. Results from a Monte Carlo test of significant difference between real and randomised data are given in Table 4, with stress values for each dimension significantly different from those for randomised data. It seems remarkable that this ordination technique can give such a robust and stable solution with such a large and complex data set.

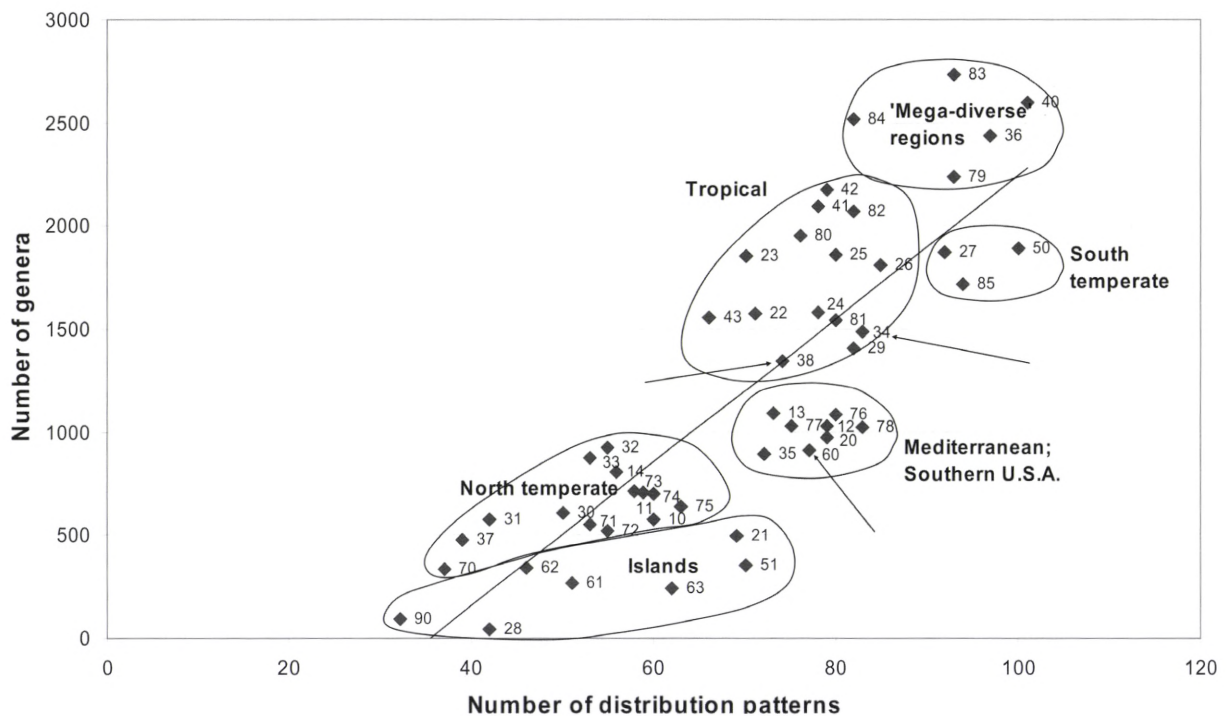


Fig. 7. Relationship between diversity of a region and number of distribution patterns shown by genera from that region ($r^2 = 0.68$); regions are identified by their 2-digit code. Graph and regression does not include data for the Antarctic Continent (91).

Table 3. Stress value, instability value, number of iterations and percentage of variance accounted for the final best run of non-metric multidimensional scaling ordination of generic distribution patterns.

Stress value	Instability value	No. of iterations	Axis 1 (r^2)	Axis 2 (r^2)		Axis 3 (r^2)	
			increment	increment	cumulative	increment	cumulative
9.76	0.028	200	0.27	0.42	0.87	0.11	0.98

Table 4. Results of Monte Carlo test for significant difference in stress between real and randomised data of non-metric multidimensional scaling ordination of generic distribution patterns. p = proportion of randomised runs with stress \leq observed stress, *i.e.* $p = (1 = \text{no. permutations} \leq \text{observed}) / (1 = \text{no. permutations})$.

No. axes	Stress in real data (15 runs)			Stress in random data (30 runs)			p
	minimum	mean	maximum	minimum	Mean	minimum	
1	31.70	41.40	46.92	56.46	56.52	56.56	0.0323
2	17.62	19.49	23.17	38.77	41.13	41.50	0.0323
3	9.76	13.82	22.32	29.59	30.01	32.40	0.0323
4	7.76	12.46	22.98	26.72	26.74	26.75	0.0323

Table 5. The 12 distribution patterns most frequently shown by angiosperm genera, as found by k -means partitioning of ordination scores from the non-metric multidimensional scaling ordination.

<i>Distribution pattern</i>	Number of genera
Tropical South America	683
Endemic to Australia	669
Endemic to Southern Africa	587
Throughout tropical Africa	501
Endemic to Brazil	414
Endemic to Madagascar or Mascarene Is.	403
Throughout the Neotropics	319
Central America and Caribbean and tropical South America	318
Endemic to Western South America	311
Endemic to Mexico	270
S. China and throughout tropical Asia	264
Endemic to Malesia	236

***k*-means partitioning**

The complete breakdown of k -means partitioning is too large to present in full here. However, Table 5 shows the 12 most frequent distribution patterns for angiosperm genera found.

Determining the optimum number of groups was to a certain extent arbitrary, by inspecting the geographical composition of the clusters and subjectively deciding to either continue or halt k -means partitioning depending on

whether or not they showed either too much or too little internal geographical variability. A non-parametric multi-response permutation procedure (MRPP) was also employed in a pair-wise fashion, testing for statistically significant between-group differences. With initial passes of the data, groups were, not surprisingly, highly significantly different. However, as partitioning progressed through successive passes, groups became progressively less statistically distinct until, on the 5th pass, a large proportion of groups no longer satisfied the test of statistical distinctness. These groups were then re-combined manually and k-means partitioning halted. Including endemic genera as separate distribution patterns 194 distribution patterns were found for *c.* 14304 angiosperm genera. Distributions throughout one tropical region or endemic to a single region remain the most common (see Table 5). Not only is tropical South America the most common distribution pattern found by this analysis, but two other widespread Neotropical patterns are within the top twelve. As both Western South America (83) and Brazil (84) are also in Table 3, each of these regions is represented by four distribution patterns out of the 12 most common.

Trying to summarise the results of such a large and complex analysis in simple terms is extremely difficult. Even after k-means partitioning has reduced the size of the data from 2817 unique distribution patterns to fewer than 200 distinct geographical clusters, however, the shape of the frequency distribution is still extremely skewed – the majority of diversity is still accounted for by only a few distribution patterns. Only 38 distribution patterns out of 194 are shared by 100 genera or more; together these account for 8785 genera, or two-thirds of the total number for all angiosperms. Furthermore, if we then pose the question ‘Is there a relationship between floristic richness and floristic complexity?’, the answer seems to

be: ‘yes’. Fig. 7 shows a strong correlation between the number of genera in a region, and the number of distribution patterns (as produced by k-means partitioning) shown by those genera – what we might refer to as the ‘complexity’ of the floristic composition of a region. Not surprisingly, there is considerable scatter in this graph. Indicating the identity of these regions and superimposing a crude geographical classification on top, however, as has been done for Fig. 7, reveals a striking uniformity within the evident clusters of the graph. Regions within each cluster are not necessarily geographically adjacent, or even geographically close, but do share important geographical characteristics. Some of these groups, furthermore, are essentially the same as those found in the analysis of floristic relationships presented earlier (see Fig. 5).

Moving up Fig. 7, at the bottom is a diffuse cluster of regions consisting only of islands: Macaronesia (21); Middle Atlantic Ocean (28); New Zealand (51); South-Central Pacific (61); Northwestern Pacific (62); North-Central Pacific (63); Subantarctic Islands (90). Island floras are not large (since most islands are small) and generally consist of a mixture of endemic elements and cosmopolitan elements, hence both the diversity and the complexity of these floras are low. This group of island regions was found repeatedly in the analysis of floristic relationships. Immediately above this is a large group of northern temperate regions; to the left of the group the most floristically depauperate of these: Russian Far East (31); Mongolia (37); Subarctic America (70) show considerably fewer distribution patterns than do the rest: Northern Europe (10); Middle Europe (11); Eastern Europe (14); Siberia (30); Middle Asia (32); Caucasus (33); Western Canada (71); Eastern Canada (72); Northwestern U.S.A. (73); North-Central U.S.A. (74); Northeastern U.S.A. (75).

Next come a group of more southerly but

still predominantly temperate regions which are here labelled 'Mediterranean; Southern U.S.A.' The North American component of this group is truly the southern U.S.A.: Southwestern U.S.A. (76); South-Central U.S.A. (77); Southeastern U.S.A. (78) while the 'Mediterranean' component (Southwestern Europe, 12; Southeastern Europe, 13; Northern Africa, 20) is missing one Mediterranean region (Western Asia, 34) but has gained another (Arabian Peninsula, 35). Though this group consists of two geographically separate entities, these two entities are very similar in their latitude. Collectively, they show many more distribution patterns (have a more 'complex' flora) than the regression line would predict for regions of such generic richness.

A large and diffuse group of tropical regions follows, representing all three of the tropical areas of Africa (West Tropical Africa, 22; West-Central Tropical Africa, 23; Northeast Tropical Africa, 24; East Tropical Africa, 25; South Tropical Africa, 26); SE. Asia (Indo-China, 41; Malasia, 42; Papuasia, 43); and the Neotropics (Central America, 80; Caribbean, 81; Northern South America, 82), with many more genera but with correspondingly more distribution patterns. To the right of these, with similar numbers of genera but with many more distribution patterns, is a group of three geographically disparate but biogeographically inter-related regions: Southern Africa (27); Australia (50); Southern South America (85). The existence of pan-Southern Hemisphere taxa has long been known (Hooker 1853), but in this study only now, with the analysis of distribution patterns, have these shared floristic elements been revealed. In the analysis of floristic relationships, the influence of greater tropical diversity meant that the greatest similarity of each of these regions was with adjacent tropical areas, whereas now, their more complex biogeographic relationships with distant areas are highlighted. Each of these regions is so large,

however, that at their northernmost limit they each extend into the tropics, so it is perhaps a moot point whether or not they should really be labelled as 'South temperate'.

Three regions, indicated by arrows in Fig. 7, do not 'fit' into this simple geographical classification: they have both more genera and more distribution patterns than they 'ought to' from their geographic location. Southwestern Pacific (60) falls within the 'Mediterranean; Southern U.S.A.' cluster, not with the group of 'Island' regions at the base of the graph; however, it did not group with these 'Island' regions in the analysis of floristic relationships either. The reasons for the anomalous position of the Southwestern Pacific region in Fig. 7 include: catching the tail end of the distributions of many SE. Asian genera which reach no further into the Pacific; New Caledonia within this region is a continental fragment derived from Gondwanaland, and shows not only an exceptional degree of generic endemism but also many floristic relationships with other regions (New Guinea, Australia, New Zealand) which are not shown by any other islands in the vicinity. Then within the tropical group are two predominantly temperate Asian areas: Western Asia (34) and Eastern Asia (38).

Finally, at the top of the graph is a group of what I have labelled 'mega-diversity' regions (China, 36; Indian Subcontinent, 40; Mexico, 79; Western South America, 83; Brazil, 84), which have much greater numbers of genera than do the other regions. These same regions were revealed as especially diverse both with respect to area and to latitude. One of them (Brazil, 84) may be regarded as exclusively tropical; three of the others (China, 36; Indian Subcontinent, 40; Mexico, 79) cross the important temperate-tropical boundary, so accumulating both exclusively temperate and exclusively tropical taxa; the final region (Western South America, 83) is the richest region at genus-level both in absolute and in relative

terms, but, as might Indian Subcontinent (40), it might be thought of as crossing the temperate/tropical divide in elevation rather than strictly by latitude, as these two regions are also by far the most mountainous, containing the bulk of, respectively, the great mountain chains of the Andes and the Himalaya.

Discussion

If genus-level patterns really are acceptable surrogates for species-level diversity (Gaston 1996; La Ferla *et al.* 2002), perhaps inferences drawn from these analyses may also be applicable at species-level. For example, the Neotropics in general, and in particular western South America and Central America, consistently emerge as the most diverse areas of the world for plants, at all spatial scales. At small scales, this was previously shown by Gentry (1988), in a global analysis of species-richness data from standard-sized 0.1-ha wet forest plots, where only one out of the five richest sites was not found in the Neotropics (Semengoh Forest in Sarawak [Borneo]); of the Neotropical sites, two were in Colombia and two in Peru. It has also been shown by a similar re-analysis of the raw data from Davis *et al.* (1994-1997) and from Mittermeier *et al.* (1999), again in each case re-scaling absolute species numbers by the species-area relationship where $z = 0.14$ (Brummitt & Nic Lughadha 2003). For the data from Davis *et al.* (1994-1997), the richest areas were found to be La Amistad (Costa Rica / Panama), the region of the upper Rio Negro (Brazil / Colombia / Venezuela) and Braulio Carillo – La Selva (Costa Rica). Similarly, re-analysing the ‘hotspots’ data of Mittermeier *et al.* (1999) reveals the Tropical Andes and Mesoamerica hotspots to be by far the most diverse for both total vascular plant species and endemic vascular plant species. Furthermore, in the pioneering study by Barthlott *et al.* (1996), three of the six areas where vascular

plant diversity was estimated to exceed 5,000 species/10,000 km² were found in the Neotropics: the Chocó-Costa Rica centre; the Tropical Eastern Andes centre; and the Atlantic Brazil centre. Notwithstanding that this result might prove to be dependent on the scale of analysis (Rahbek & Graves 2001), here again it is western South America which is the richest region of the world, this time for genera of flowering plants.

That Central America is among the most diverse was not expected, the small area of this region masking its apparent diversity. However, this may be due to Central America being both the southern limit of distribution for many northern hemisphere taxa, and the northern limit of distribution for many tropical and southern hemisphere taxa, given that degree of generic endemism is not especially high for Central America (3.7%). That is to say, Central America straddles different biogeographic regions, in addition to there also being many pantropical taxa or taxa from throughout the Neotropics there. This same argument applies to Mexico just to the north, although the flora of Mexico shows a more pronounced northern bias (many North American genera do not reach south beyond Mexico, while many South American genera do not reach north beyond Central America). Geologically, the mountains today forming the backbone of the Central American isthmus have their origin in the Andean orogeny caused by the meeting of the North American and South American plates, with a recent period of uplift under the north-western Andes from approximately 5 million years ago (mya), and a growing series of volcanic islands between the two continents then coalescing to form the Isthmus of Panama in the Pliocene, approximately 3.5 mya (Marshall *et al.* 1980; Coates & Obando 1996).

Gentry (1982) proposed that the exceptionally high floristic diversity of the northwestern Andean and southern Central American

region was principally due to rapid, sympatric in situ speciation caused by the uplift of the Andes mountains, on top of an already-rich tropical flora, and he estimated that this explosive speciation might account for approximately half of the total number of Neotropical species. Gentry (1982) distinguished three broad floristic elements within the Neotropics as a whole (notwithstanding the tremendous floristic complexity of such a large area at more detailed scales): Laurasian taxa, which were originally absent from the isolated South American continent but which migrated south from North America as the Isthmus of Panama closed; Gondwanan, Amazonian-centred trees and lianas with almost half of their species complement in Amazonia; and Gondwanan, Andean-centred taxa with only a minority of species present in Amazonia. Plant groups with a northern-Andean-centred distribution tend to be the abundant guilds of epiphytes, understory shrubs and palmettos that are less prominent in the Palaeotropics, which are characterised by numerous local endemic species in small habitat patches subject to frequent disturbance, with rapid generation times and specific pollinator relationships, all factors promoting rapid speciation (Gentry 1982). This hypothesis remained untested until borne out in a recent study of the species-rich genus *Inga* (Leguminosae), where molecular estimates of the mean divergence time of all species from the most-recent common ancestor of the genus from two separate gene regions were between 3.5 and 5.9 mya, respectively (after application of non-parametric rate smoothing; Richardson *et al.* 2001).

Gentry (1982) further hypothesized considerable floristic interchange between Northern and Southern America following the closing of the Panamanian isthmus, though he suggests that in this case the interchange occurred without significant amounts of speciation (at least for woody taxa) in either the northern or

southern elements. Also, there seems to be some ecological differentiation between the two elements, with many Laurasian (northern) taxa today confined to the montane Andes and absent from lowland South America, and Gondwanan, Amazonian-centred (southern) taxa dominant in lowland forest types (Gentry 1982). Furthermore, the woody Laurasian taxa show a higher proportion of wind-pollination, longer generation times and low numbers of species. Gondwanan, Amazonian-centred trees and lianas, however, show restricted allopatric distributions within a widespread lowland rain forest habitat, with species of unrelated groups found in the same small areas which have been proposed as Pleistocene forest refuges within a (then) more widespread savannah (Prance 1973; Gentry 1982) or seasonally-deciduous forest habitat (Pennington *et al.* 2000).

This scenario of both northern and southern migrations into Mexico, Central America and western South America may also apply to Region 81, the Caribbean. The proto-Antilles was created at the junction of the Northern and Southern American plates, and was then isolated by the pinching of the Caribbean plate from the easternmost section of the East Pacific plate when North and South America eventually joined, forming the Greater Antilles we know today (Gentry 1982; Hedges 2001). Gentry (1982) emphasised the dual northern and southern affinities of the Caribbean flora, which Rosen (1975) had suggested was due to the original vicariant stocking of a proto-Antillean island arc which was located between separate North and South American continents. The age of the Greater Antilles (Cuba, Hispaniola, Puerto Rico and Jamaica) is estimated at about 60 – 45 mya and for the Lesser Antilles at about 10 mya (Rosen 1975, 1985; MacPhee & Grimaldi 1996; Hedges 2001). Rosen (1975) identified two major terrestrial elements of the Caribbean biota, between each of North America and South America, and the Caribbean,

respectively, which 'represent extensions of the original biotas into the Caribbean region', as well as two marine elements between both the eastern Atlantic and the eastern Pacific, and the Caribbean. In this vicariance model, the biotic interchange occurred primarily through the proto-Antilles archipelago, not the modern-day Greater and Lesser Antilles, as it later became. However, a series of recent studies on zoological taxa (reviewed in Hedges 2001) calculate divergence times for Caribbean lineages which are too young to have been the product of vicariance and which imply later (Cenozoic) over-water dispersal from modern-day continents to various Caribbean islands.

The Caribbean also shows relatively higher generic endemism (12.6%) than do either western South America (10.8%) or Central America (3.7%), and in this case one might think the fragmentary nature of the region is contributing to increased generic diversity. In fact, however, this is not so; some two-thirds of Caribbean genera (134 of 201) are confined to either or both of only two islands: Cuba and Hispaniola. It is the isolation of this region from the continental Americas that is important, not the fragmented nature of the region itself, and the predominance of endemic genera on only the oldest islands of the Caribbean archipelago furthermore might perhaps give more weight to a vicariant rather than a dispersalist origin for these endemics, in contrast to zoological taxa, where the distribution of modern-day higher-level taxa is often widespread, but diversity of both modern-day and fossil higher-level taxa is low (Hedges 2001).

A single tectonic episode, the fusion of the two American continents, would therefore have allowed floristic interchange between two isolated floras across what became the most-diverse Neotropical regions, 79 (Mexico), 80 (Central America), 81 (Caribbean), and 83 (Western South America). However, this interchange would have occurred at different times:

in the late Mesozoic / early Cenozoic for the Caribbean (Rosen 1975 1985; Hedges 2001); during the Pliocene for Central America and western South America (Gentry 1982; Richardson *et al.* 2001). In the former case, the scenario is one of initial floristic interchange followed by isolation and then additional later dispersal to the Caribbean, resulting in high endemism. In the latter case, this initial diversity from floristic interchange is then thought to have been coupled with high speciation rates following isolation of populations by the successive phases of the Andean Orogeny in the Central American / northwest South American region. In each case, however, since the endemic genera are only ever a small component of the total generic flora, what is underpinning the diversity within that region is the number and strength of floristic relationships between that region and the regions beyond it. The greater the number of the floristic elements within that region, the greater is the diversity of that region.

Acknowledgements

The work presented here forms part of a PhD thesis carried out at the Royal Botanic Gardens, Kew, under the auspices of the University of Edinburgh. It was funded by the Bernard Sunley Charitable Trust, to whom I remain extremely grateful. I would also like to thank my principal supervisors, Professor Peter Furlley of the University of Edinburgh and Eimear Nic Lughadha of the Royal Botanic Gardens, Kew, for their help. Rafaël Govaerts kindly provided the figures from which I extrapolated species counts. In addition I have also benefited from discussions with William Mackness of the University of Edinburgh and from discussions with, and practical assistance from, Simon Mayo, David Simpson and Justin Moat of the Royal Botanic Gardens, Kew. Lastly, I would like to thank Henrik Balslev and Ib Friis

for asking me to speak at this symposium, and for their forbearance during the preparation of this manuscript.

Literature cited

- Archibold, O.W. 1995. *Ecology of World Vegetation*. London, U.K.
- Arrhenius, O. 1921. Species and area. *J. Ecol.* **9**: 95-99.
- Bailey, T.C. & Gatrell, A.C. 1995. *Interactive Spatial Data Analysis*. Harlow, U.K.
- Barthlott, W., Lauer, W. & Placke, A. 1996. Global distribution of species diversity in vascular plants. *Erdkunde* **50**: 317-327.
- Beals, E.W. 1984. Bray-Curtis ordination: and effective strategy for analysis of multivariate ecological data. *Advances in Ecological Research* **14**: 1-55.
- Biondini M.E., Mielke, Jr., P.W. & Berry, K.J. 1988. Data-dependent permutation techniques for the analysis of ecological data. *Vegetatio* **75**: 161-168.
- Brown, J.H., West, G.B. & Enquist, B.J. 2000. Scaling in Biology: patterns and processes, causes and consequences. In Brown, J.H. & West, G.B. (eds.), *Scaling in biology*. Oxford, U.K. Pp. 1-24
- Brummitt, N.A. & Nic Lughadha, E. 2003. Biodiversity – where's hot and where's not. *Conservation Biol.* **17**: 1442-1448.
- Brummitt, R.K. (comp.) 1992. *Vascular Plant Families and Genera*. Kew, U.K.
- Brummitt, R.K. 2001. *World Geographical Scheme for Recording Plant Distributions, 2nd edition*. Pittsburgh, U.S.A.
- Calinski, T. & Harabasz, J. 1974. A dendrite method for cluster analysis. *Communications in Statistics* **3**: 1-27.
- Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. *Austral. J. Ecol.* **18**: 117-143.
- Coates, A.G. & Obando, J.G. 1996. The geologic evolution of the Central American isthmus. In Jackson, J.B.C., Budd, A.F. & Coates, A.G. (eds.), *Evolution and Environment in Tropical America*. Chicago, U.S.A. Pp. 21-56
- Colwell, R.K. & Lees, D.C. 2000. The mid-domain effect: geometric constraints on the geography of species richness. *Trends Ecol. Evol.* **15**: 70-76.
- Connor, E.F. & McCoy, E.D. 1979. The biology and statistics of the species-area relationship. *Amer. Naturalist* **113**: 791-833.
- Davis, S.D., Heywood, V.H., Herrera-MacBryde, O., Villa-Lobos, J. & Hamilton, A.C. 1994-1997. *Centres of Plant Diversity: a guide and strategy for their conservation*. Vol. 1-3. Cambridge, United Kingdom.
- Frodin, D.G. 2001. *Guide to Standard Floras of the World, 2nd edition*. Cambridge, U.K.
- Gaston, K.J. 1996. Species richness: measure and measurement. In Gaston, K.J. (ed.), *Biodiversity: a biology of numbers and difference*. Oxford, U.K. Pp. 77-113.
- Gaston, K.J. 2003. *The Structure and Dynamics of Geographic Ranges*. Oxford, U.K.
- Gentry, A.H. 1982. Neotropical floristic diversity: phyto-geographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Ann. Missouri Bot. Gard.* **69**: 557-593.
- Gentry, A.H. 1988. Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Missouri Bot. Gard.* **75**: 1-34.
- Govaerts, R. 2001. How many species of seed plants are there? *Taxon* **50**: 1085-1090.
- Govaerts, R. 2003. How many species of seed plants are there? – a response. *Taxon* **52**: 583-584.
- Hedges, S.B. 2001. Biogeography of the West Indies: an overview. In Woods, C.A. & Sergile, F.E. (eds.), *Biogeography of the West Indies: patterns and perspectives, 2nd edition*. Boca Raton, U.S.A. Pp. 15-33.
- Hooker, J.D. 1853. *The botany of the Antarctic voyage of H.M. discovery ships Erebus and Terror in the years 1839-1843. Vol. II, Flora Novae-Zelandiae, Part I, Flowering Plants*. London, U.K.
- Kruskal, J.B. 1964a. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika* **29**: 1-27.
- Kruskal, J.B. 1964b. Non-metric multidimensional scaling: a numerical method. *Psychometrika* **29**: 115-129.
- La Ferla, B., Taplin, J., Ockwell, D. & Lovett, J.C. 2002. Continental scale patterns of biodiversity: can higher taxa accurately predict African plant distributions? *Bot. J. Linn. Soc.* **138**: 225-235.
- Lance, G.N. & Williams, W.T. 1968. A general theory of classification sorting strategies. II: Clustering systems. *Computer Journal* **10**: 271-277.
- Legendre, P. & Gallagher, E. 2002. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**: 271-280.
- Lewin, R. 1989. Biologists disagree over bold signature of nature. *Science* **244**: 527-528
- Lomolino, M.V. 1989. Interpretations and comparisons of constants in the species-area relationship: an additional caution. *Amer. Naturalist* **133**: 277-280.
- MacArthur, R.H. & Wilson, E.O. 1967. *The Theory of Island Biogeography*. Princeton, New Jersey.
- MacPhee, R.D.E. & Grimaldi D.A. 1996. Mammal bones in Dominican amber. *Nature* **380**: 489-490.

- Marshall, L.G., Webb, S.D., Sepkoski, J.J. & Raup, D.M. 1980. Mammalian evolution and the great American interchange. *Science* **215**: 1351-1357.
- McCune, B. & Grace, J.B. 2002. *Analysis of Ecological Communities*. Glenden Beach, Oregon, U.S.A.
- McCune, B. & Mefford, M.J. 1999. *PC-ORD: Multivariate Analysis of Ecological Data, version 4.0*. Glenden Beach, Oregon, U.S.A.
- Milligan, G.W. & Cooper, M.C. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**: 159-179.
- Mittermeier, R.A., Myers, N., Gil, P.R. & Mittermeier, C.G. 1999. *Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions*. Sierra Madre, Mexico.
- Niklas, K.J. 1988. Patterns of vascular plant diversification in the fossil record – proof and conjecture. *Ann. Missouri Bot. Gard.* **75**: 35-54.
- Niklas, K.J. & Tiffney, B.H. 1994. The quantification of plant biodiversity through time. *Philos. Trans., Ser. B* **345**: 35-44.
- Pennington R.T., Prado, D.E. & Pendry, C.A. 2000. Neotropical seasonally dry forests and Quaternary vegetation changes. *J. Biogeogr.* **27**: 261-273.
- Prance, G.T. 1973. Phylogeographic support for the theory of Pleistocene forest refuges in the Amazon Basin based on evidence from distribution patterns in Caryocaryaceae, Chrysobalanaceae, Dichapetalaceae and Lecythidaceae. *Acta Amazonica* **3**: 5-28.
- Prance, G.T. 1994. A comparison of the efficacy of higher taxa and species numbers in the assessment of biodiversity in the neotropics. *Proc. Roy. Soc. London, Ser. B, Biol. Sci.* **345**: 89-99.
- Preston, F.W. 1962. The canonical distribution of commonness and rarity. *Ecology* **43**: 185-215.
- Rahbek, C. & G.R. Graves 2001. Multiscale assessment of patterns of avian species richness. *Proc. Natl. Acad. U.S.A.* **98**: 4534-4539.
- Richardson, J.E., Pennington, R.T., Pennington, V & Hollingsworth, P.M. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* **293**: 2242-2245.
- Rohde, K. 1992. Latitudinal gradients in species diversity – the search for the primary cause. *Oikos* **65**: 514-527.
- Rosen, D.E. 1975. A vicariance model of Caribbean biogeography. *Syst. Zool.* **24**: 431-461.
- Rosen, D.E. 1985. Geological hierarchies and biogeographic congruence in the Caribbean. *Ann. Missouri Bot. Gard.* **72**: 636-659.
- Rosenzweig, M.L. 1995. *Species Diversity in Space and Time*. Cambridge, U.K.
- Savolainen, V., Chase, M.W., Hoot, S.B., Morton, C.M., Soltis, D.E., Bayer, C., Fay, M.F., de Bruijn, A.Y., Sullivan, S. & Qiu, Y.-L. 2000. Phylogenetics of flowering plants based on combined analyses of plastid atpB and rbcL gene sequences. *Syst. Biol.* **49**: 306-362.
- Soltis, P.S., Soltis, D.E. & Chase, M.W. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**: 402-404.
- Taylor, P.H. & Gaines, S.D. 1999. Can Rapoport's rule be rescued? Modelling causes of the latitudinal gradient in species richness. *Ecology* **80**: 2474-2482.
- Whitmore, T.C. 1990. *An Introduction to Tropical Rain Forests*. Oxford, U.K.
- Williams, C.B. 1943. Area and the number of species. *Nature* **152**: 264-267.
- Williams, C.B. 1964. *Patterns in the Balance of Nature*. London, U.K. and New York, U.S.A.
- Willis, J.C. 1922. *Age and Area: a study in the geographical distribution and origin of species*. Cambridge, U.K.