

## Lichen biogeography at the largest scales

Linda in ARCADIA

**Abstract:** A quantitative cluster analysis of lichen distribution data demonstrates that the main biogeographical subdivision in the world's lichen biota is into a Gondwanan and a Laurasian element. Patterns at smaller scales mainly reflect local climate. Wallace's line is not a significant boundary for lichens. The Gondwana / Laurasia split also applies to lichenicolous fungi. To a considerable extent, it applies to many of the larger families and orders of lichens too, though at these ranks the affinities of the lichen biota of eastern Asia and temperate North America are sometimes ambiguous.

**Key words:** cluster analysis, Gondwana, Laurasia, lichenicolous fungi, Wallace's Line

*Accepted for publication 21 February 2013*

---

### Introduction

This paper discusses the distribution of lichens at the largest scales, that is, their global biogeography. For a good introduction to all aspects of the subject of biogeography as a whole, see Lomolino *et al.* (2010).

The distribution of lichens is obviously not random. This non-randomness could be described in many ways, but three obvious ones include the following.

1) Divide the planet's land surface into regions, so that each point belongs to no more than one region. Carry out a cluster analysis to yield a dendrogram in which regions with similar lichen biotas are close together in some small subtree, but regions with very dissimilar lichen biotas are widely separated. Large subtrees correspond to major geographical regions that are, in some sense, meaningful units for lichen distribution. This is the method used in this paper, but it is worth contrasting briefly with other possible methods.

2) Consider, for example, *Tuckneraria pseudocomplicata*, which occurs along the Pacific coasts of NE Asia and NW North America (map in Randlane & Saag 2004: 373) in a well-defined pattern. One could

seek patterns, like this amphi-Beringian pattern, that are shared by many species. However, this approach requires distribution data defined on a much finer scale (e.g. latitude and longitude of numerous sites, rather than national or regional checklists) than is available for most lichens.

3) Biogeographers of flowering plants have used taxa that are endemic and autochthonous to define biogeographical regions. However, without a good fossil record, this cannot be done for lichens.

For an excellent overview of lichen biogeography treated in a qualitative way, with numerous references to earlier literature, see Galloway (2008). The only previous *quantitative* analysis of global lichen distributions is that of Feuerer & Hawksworth (2006). It does not attempt to determine meaningful regions, but instead takes 35 floristic regions already defined, by Takhtajan, and analyzes the relationships between them. Takhtajan's floristic regions were defined with vascular plants mainly in mind, and there is no reason to suppose *a priori* that they are meaningful for lichens. Feuerer & Hawksworth's algorithm was not fully specified, but appears to have been unsatisfactory as it concluded that the primary subdivision of the world's lichen biota is into those of St. Helena and Ascension Island, versus everything else – an implausible result. I have found it essential to

---

L. in Arcadia: Kastri, 22013, Arkadias, Greece. Email: linda\_in\_arcadia@cantab.net

construct a *robust* algorithm, and to monitor its workings.

Because little has been published about cluster analysis on large datasets of lichen distribution, and because I have found that there are many pitfalls, I explain why I adopted the method used (not merely what that method was), and the problems that arise in practice, as an aid to future workers. For more on the large subject of cluster analysis see, for example, Romesburg (2004).

## Materials and Methods

### Coefficient of similarity

Given two regions, P and Q, and a list of the lichen species present in each, we need a measure of the similarity of their lichen biotas. It will range from zero if P and Q have no species in common, to 1.0 if P and Q have identical biotas. Many coefficients of similarity have been suggested (Romesburg alone lists 12, and is not exhaustive), but we need only discuss three.

Suppose that  $p$  species are recorded for region P,  $q$  for Q, and  $r$  for both P and Q. The number recorded for P and Q combined is  $p + q - r$ . Denote the smaller of  $p$  and  $q$  by  $\min(p, q)$ . Possible coefficients include: i)  $r / (p + q - r)$ ; this is Jaccard's coefficient; ii)  $2r / (p + q)$ ; this is Sørensen's coefficient; and iii)  $r / \min(p, q)$ ; this is Simpson's coefficient. Feuerer & Hawksworth (2006) used Sørensen's coefficient, but as it leads to the same order relations, and thus the same dendrogram, as Jaccard's coefficient it need not be considered further.

Jaccard's coefficient has an obvious weakness. If P and Q are neighbouring regions with similar climatic and other factors, we expect their lichen biotas to be similar and a similarity coefficient to be close to 1.0. However, if P has been well studied and 1000 species are recorded, but Q is poorly studied and only 100 species are known, Jaccard's coefficient cannot exceed 0.1 and is misleading. Because our knowledge of lichen biotas is far from uniform, this is a pertinent and serious problem. Simpson's coefficient does not suffer from this problem.

I experimented extensively with Jaccard and Simpson coefficients. As expected, the latter gave much better results, that is, far fewer instances of a region appearing in the dendrogram at a place where it obviously did not belong. However, Simpson sometimes suffers from a different problem. If P is poorly studied, the lichens known from P may be mostly conspicuous common species. Those same species may also be present in other regions whose lichen biota overall does not have much in common with P, so Simpson may sometimes overestimate the similarity between P and those regions. With the algorithm described below, this problem, if it occurs, results in too many nodes of the dendrogram having extremely anisotomic splits and some branches being unreasonably long.

We need a coefficient that retains the good property of Simpson, its ability to cope with differences in intensity of study, but which reduces its tendency to be misled by biased sampling in poorly studied regions. Biased sampling is not a problem for Jaccard (it does not add to the problems that Jaccard already has) so adding a small mix of Jaccard into the coefficient used might be expected to help. In other words, we could use the linear combination:  $\lambda * \text{Jaccard} + (1 - \lambda) * \text{Simpson}$ . Extensive investigations with such combinations showed best results for lichen distribution data when  $\lambda$  is small. All results reported below used this linear combination with  $\lambda = 0.1$ ; this greatly reduced the kinds of problems experienced when either Jaccard or Simpson were used alone.

Although the choice of similarity coefficient greatly influenced the amount of 'noise' in the final dendrogram, the principal biogeographical subdivisions were generally recognizable whatever coefficient was used. To that extent, the reported results are robust.

### The algorithm

We start with  $n$  regions and a lichen checklist for each, and we wish to produce a dendrogram of regions. The best method will depend on the dataset and, as little work of this sort has been reported for lichens, I considered it inadvisable to choose a method *a priori*. Instead, I investigated many algorithms and developed one that worked well for the dataset under study. This necessarily involved writing some code, rather than depending on any 'black box' software.

There are two basic strategies for cluster analysis: bottom up (or agglomerative) and top down (or divisive). In the former, one calculates the similarity coefficient for every pair of regions, and merges into a single 'group of regions' the pair whose biota is the most similar (i.e. for which the similarity coefficient is greatest). After that merge, there are now only  $n - 1$  regions or groups of regions to work with, one fewer than before, and the algorithm continues in the obvious way. This method is computationally efficient, but it only looks locally at the data: it only compares two regions at a time. Also, having merged them there is no way to 'unmerge' them (i.e. the algorithm is 'greedy'). In extensive tests, a pure bottom up strategy gave results that, although not unreasonable overall, were noisy: too many regions appeared in the dendrogram in places where they clearly did not belong.

The top down method seeks the optimum way of splitting the  $n$  regions into two subsets. For any particular split one subset will have  $m$  regions and the other  $n - m$  regions. Merge the lists of species to obtain a list for the  $m$  regions combined and one for the  $n - m$  regions combined. Calculate a similarity coefficient from the two resulting lists. Repeat for all possible ways of partitioning the original  $n$  regions into two subsets. Choose the partition for which the similarity coefficient is smallest, that is the split which gives the two subsets whose lichen biota is as dissimilar as possible. This defines the first split of the data (i.e. it defines what happens at the root node of the dendrogram). Repeat recursively. The top down strategy is the ideal one, as it always takes a global view of the

data, but it is computationally unfeasible unless  $n$  is small.

After much investigation, I adopted a hybrid approach to retain advantages of both methods. Basically it has two steps, but for reasons of computational efficiency it was formulated as a 3-step algorithm. Choose some positive integer  $x$ . Use the bottom up method until only  $x$  (groups of) regions remain (first step). Then use the top down method to partition those  $x$  (groups of) regions into two subsets (second step). This defines the split at the root node of the dendrogram. Repeat recursively.

This method usually works well, but sometimes at the end of the bottom up phase we are left with  $x - 1$  (groups of) regions, each containing only a single region or a very few regions, with everything else placed into the other region group. (This problem appears to be related to the weakness of the Simpson coefficient discussed above. The problem is inherited, though to a much smaller degree, by the combined coefficient.) As a result, the top down stage cannot look sufficiently deeply into the data, and the resulting split may be far from optimum as far as the  $n$ -region dataset is concerned. The solution adopted was as follows. If, at the end of the bottom up phase, any region group has more than 30% of the original  $n$  regions, then (third step) perform a single bottom-up-followed-by top-down calculation (step 1, step 2, and if necessary step 3, which is implemented recursively) on that region group to split it into two, so that the main top-down algorithm will work with  $x + 1$ , not just  $x$ , (groups of) regions. If necessary, repeat step 3 until no region group has more than 30% of the original  $n$  regions when the top down algorithm acts. This third step is basically an optimization. The same, or similar, results could be obtained by dispensing with this third step and instead using a much larger value of  $x$  everywhere, but that would increase CPU requirements greatly.

For the dataset used,  $x = 14$  proved to be a suitable choice. It gave good results, and use of larger values for  $x$  did not noticeably increase the quality of the resulting dendrograms. The maximum value to which  $x$  was ever increased by the third step was 20. In many cases it was not increased at all, or was increased by only a small amount.

The algorithm was implemented using code written by myself in C++, compiled by GCC version 4.1.2 with optimization enabled, and run under Linux on an ordinary PC.

This algorithm is robust. Although the dataset included some parts of the world at very coarse resolution (e.g. much of South America) and some parts at very fine resolution (e.g. parts of south and SE Europe), and although the number of species per region varied (in the main run) from 50 to 2043, the algorithm gave generally sensible results.

### The database

Distribution data was taken from the author's personal database. This has been assembled, over many years, from information in 4689 publications, and contains distribution and other data on most described species of

lichens and lichenicolous fungi, organized under 82 590 names (most of which are, of course, synonyms). It includes 283 613 items of distribution data, each with a reference. For present purposes the distribution data does include some duplication, especially where a modern checklist has rendered data in older publications superfluous. For most regions of the world the distribution data is close to complete; for almost all others it is well representative of the lichen biota.

Only taxa normally treated at species rank were included in the analysis, though data reported under infra-specific synonyms of those names was included. The reason for ignoring infra-specific taxa is that many have been described in the past, but the status of most of them is unclear. Probably only a few represent good taxa.

Only lichenized and lichenicolous ascomycetes were included in the analysis. There are too few species of basidiolichens or of lichenicolous basidiomycetes to permit meaningful analysis of them separately from ascomycetes, and it seemed inadvisable to analyze them together.

The classification of organisms as 'lichens' or 'lichenicolous fungi' was made at the rank of genus. This means that a few lichenicolous or non-lichenized species in predominantly lichenized genera are misclassified as lichens. To do otherwise would have involved excessive labour.

Species in the database that I consider to be doubtful taxa were excluded from the analysis. There were 2027 such lichen species and 54 for lichenicolous fungi. For a further 1791 species of lichens and 74 of lichenicolous fungi I have no distribution data at all, so those species were also excluded. I expect that most of them would prove, on fuller investigation, to be doubtful taxa. Species known only from a single region (i.e. endemic species) were also excluded. The species excluded as endemic varied from one case to another, depending on how regions were defined for that case. The main reason for excluding them is that a significant proportion are probably not good species and would merely introduce noise into the analysis. Because the similarity coefficient used is based largely on Simpson's coefficient, excluding them has far less effect than it would if other kinds of coefficient were used. As a check, some analyses were made with endemic species included, but the overall conclusions were little changed.

After defining regions for analysis and removing duplicated data, the main analysis of lichens used 140 425 distinct (species, region) pairs, of which 91.4% are from formal publications, 7.2% are from miscellaneous online sources, and 1.4% are from the online checklists maintained at the University of Hamburg; the latter were used only to supplement my own dataset for China. Had I been able to make fuller use of the Hamburg checklists, the dataset could have been about 8% larger, but regrettably permission was not given for this. The main analysis for lichenicolous fungi used 4361 distinct (species, region) pairs, 99.1% of which are from formal publications and the remainder from miscellaneous online sources; the Hamburg datasets were not used for lichenicolous fungi. The analyses for individual families,

orders, etc., used smaller datasets, because fewer species are involved and because it was necessary to use larger, and thus fewer, regions.

### Geographical regions used

The regions used for the analysis, and shown on the Figures, are mostly individual countries, well-known subdivisions of a country, or groups of such countries or subdivisions. Subdivisions of Russia follow Urbanavichus & Andreev (2009).

The following 231 regions were used for the main analysis, that of global lichen distribution. The number of species included for each region is shown. Those numbers may differ from those in the standard checklists, since species which I consider doubtful taxa or for which I have no distribution data were removed, species endemic to a region are also excluded, and in some cases my database is incomplete. I may also have taken different views on synonymy than the authors of some checklists.

ACT (Australian Capital Territory) 372; Aegean Is (=Cyclades, Dodecanese, NE Aegean Islands) 626; Africa SE (=Mozambique, Zimbabwe) 115; Alaska 995; Albania 193; Alberta 469; Algeria 462; Andaman & Nicobar Is 298; Angola 138; Antarctic Peninsula 243; Arabia (=all the Arabian Peninsula, but excludes Socotra) 122; Argentina 959; Armenia + Azerbaijan + Georgia 130; Arunchal Pradesh 450; Ascension Is 118; Auckland Is 146; Austria 2043; Azores 553; Belgium 937; Belize + El Salvador + Honduras 106; Bhutan 291; Bjørnøya 188; Bolivia 470; Bosnia (includes Herzegovina) 628; Bouvet Is 50; Brazil 1847; British Columbia 799; Bulgaria 901; California 1309; Cameroon 112; Campbell Is 173; Canada E (=Labrador, Miquelon, Newfoundland) 718; Canada N (Mackenzie, Nunavut, Yukon) 345; Canada SE (=New Brunswick, Nova Scotia, PEI) 930; Canary Is 1068; Cape Verde Is 255; Channel Is (British Is, not California) 89; Chatham Is 60; Chile 747; China E (=Anhui, Henan, Hubei, Hunan, Jiangsu, Shandong, Zhejiang) 324; China N & W (=Gansu, Inner Mongolia, Ningxia, Qinghai, Shansi, Xinjiang) 418; China NE (=Heilungkiang, Jilin, Liaoning) 243; China S (=Fukian, Guangdong, Hainan, Hong Kong) 511; China SW (=Guizhou, Kwangsi, Sichuan, Yunnan) 706; Christmas Is 91; Colombia 1289; Congo Region (both Congos) 172; Continental Antarctica 78; Cook Is 62; Corfu 263; Corsica 291; Costa Rica 873; Crete 608; Croatia 967; Cuba 380; Cyprus 364; Czech Republic 1393; Dakotas + Nebraska 514; Denmark 921; Dominican Republic 115; Ecuador 539; Egypt 124; England 1414; Estonia 891; Ethiopia 320; European Russia Arctic 771; European Russia C 792; European Russia N 1435; European Russia S 360; FYROM (Former Yugoslav Republic of Macedonia) + Montenegro + Serbia 843; Faeroe Is 358; Falkland Is 182; Fiji 205; Finland 1544; France (excludes Corsica) 1836; French Guiana 322; Gabon Region (=Equatorial Guinea, Fernando Po, Gabon, Sao Tome) 88; Galapagos Is 244; Germany 1965; Greece N (=Epiros, Macedonia, Thrace) 639; Greece S (=Attica, Peloponnese) 434; Greece mid (=Evia, Sterea Ellada, Thessaly) 513; Greenland 1036;

Guatemala 453; Guyana 371; Haiti 70; Hawaii (the entire Hawaiian group) 641; Hungary 373; Iceland 662; Idaho 232; India E (=Madhya Pradesh, Orissa, Bihar) 228; India mid (=Andhra Pradesh, Maharashtra, Goa) 217; India NE (=Assam, Meghalaya, Nagaland, Manipur, W Bengal) 825; India NW (=Himachal Pradesh, Jammu & Kashmir) 485; Iran 452; Ireland (all 32 counties) 1057; Italy Central N (=Emilia Romagna, Tuscany) 1005; Italy Central S (=Abruzzi, Lazio, Marche, Umbria) 657; Italy NE (=Friuli, S Tirol, Veneto, Venezia Giulia) 1471; Italy NW (=Liguria, Lombardy, Piedmont, Valle Aosta) 1385; Italy S (=Basilicata, Calabria, Campania, Puglia) 845; Jamaica 165; Jan Mayen 145; Japan 1258; Java 367; Karnataka 343; Kazakhstan 517; Kenya 610; Kerala + Tamil Nadu 817; Kerguelen 78; Korea (both) 183; LHI (Lord Howe Is) 120; Latvia 489; Leeward Is 419; Lesotho + S Africa 694; Liberia Region (=Guinea, Ivory Coast, Liberia, Senegal, Sierra Leone) 215; Lithuania 563; Luxembourg 689; Macquarie Is 76; Madagascar 275; Madeira 656; Malaysia + Singapore 549; Manitoba + Ontario 483; Mauritius + Rodrigues Is 122; Mexico 1745; Michigan 816; Minnesota 751; Mongolia 203; Morocco 894; NSW (New South Wales) 1345; NZ (New Zealand) North Is 959; NZ South Is 1157; Namibia 111; Nepal 565; Netherlands 763; New Caledonia 575; New Guinea 787; New York 820; Nicaragua 252; Nigeria Region (=Nigeria, Benin, Ghana, Togo) 98; Norfolk Is 137; Northern Territory (Australia) 295; Norway 1873; Oregon 568; Pacific NW (=Caroline Is, Mariana Is) 98; Pacific SE (=Gambier Is, Henderson Is, Marquesas, Oeno Is, Pitcairn Is, Tuamotu Is, Tubai Is) 99; Pakistan 63; Palestine (all administrations) 310; Panama 347; Paraguay 443; Peru 574; Philippines 547; Poland 1534; Porto Rico 472; Portugal 922; Prince Edward Is (Antarctica, not Canada) 55; Quebec 941; Queensland 1465; Reunion Is 180; Romania 372; Russia Caucasus 1058; Russian Far East Arctic 724; Russian Far East N 879; Russian Far East S 1009; Rwanda Region (=Burundi, Rwanda) 195; S Georgia 193; S Orkney Is 210; S Shetland Is 219; Samoa (both) 158; Sardinia 1049; Saskatchewan 486; Scotland 1620; Siberia Arctic 798; Siberia E 820; Siberia S 1576; Siberia W 557; Sicily 673; Sikkim 486; Slovakia 1388; Slovenia 820; Society Is 183; Solomon Is 158; South Australia 442; Spain 1561; Spitzbergen 494; Sri Lanka 452; St Helena Group (=St Helena, Gough Is, Tristan Da Cunha) 248; Stewart Is 284; Sumatra 90; Sweden 2033; Switzerland 1723; Syria 353; Tadjzhikistan 473; Taiwan 957; Tanzania 646; Tasmania 944; Thailand 1061; Tibet 343; Tonga 62; Trinidad 97; Tunisia 403; Turkey (Asia) 1054; Turkey (Europe) 204; USA Appalachians (=Kentucky, Ohio, Tennessee, W Virginia) 608; USA Central (S) (=Arkansas, Oklahoma, Texas) 726; USA Central (mid) (=Illinois, Indiana, Iowa, Kansas, Missouri) 555; USA E Coast (N) (=Delaware, Maryland, New Jersey, Pennsylvania) 738; USA E Coast (S) (=Georgia, S Carolina) 936; USA E Coast (mid) (=N Carolina, Virginia) 898; USA NE (excl. NY) (=Connecticut, Maine, Massachusetts, New Hampshire, Rhode Is, Vermont) 870; USA Rockies (=Colorado, Montana, Wyoming) 1069; USA S (=Alabama, Florida, Louisiana, Mississippi) 857; USA SW (=Nevada, Utah)

364; USA far SW (=Arizona, New Mexico) 1050; Uganda 388; Ukraine 1276; Urals N 929; Urals S 733; Uruguay 452; Uttarakhand 530; Vanuatu 109; Venezuela 999; Victoria 862; Vietnam 99; Virgin Is (both) 97; Wales 1223; Washington 985; Western Australia 594; Windward Is 227; Wisconsin 722; Zambia Region (=Malawi, Zambia) 53.

In exploratory work, it was found that use of regions with fewer than *c.* 50 species in the main analysis was unsatisfactory. Such regions often became placed in parts of the dendrogram where they clearly do not belong. Countries or regions with fewer than 50 species were grouped with others for the analysis, or omitted entirely if there seemed no reasonable way to group them with others.

Other runs, those of Figs 2–5 and those for individual families etc., used different regions in some cases.

## Results

Figures 1–5 show the results. At each node in the dendrograms the corresponding similarity coefficient is also shown, as a percentage rounded to the nearest integer. The right hand side of Fig. 1 also indicates some groups of regions that seem, on an informal basis, to be natural groupings or that merit discussion.

The left to right direction in the dendrograms does not correspond to time (they are not cladograms). A subtree and its sister subtree are not to be interpreted as (necessarily) derived from some common precursor state.

All branches are drawn with the same length. In cladistics, branch lengths are sometimes used to indicate statistical significance, but it is not meaningful to do that here. There are fundamental difficulties in calculating any good measure of statistical significance for a cluster analysis of the present sort (the main, though not the only, one is that we do not know and cannot estimate objectively, the true number of species actually present in any region).

### Results for lichens as a whole

Figure 1 shows the dendrogram for the main analysis of lichen distribution. The primary split is between a southern region, groups 1–8, and a northern region, groups 9–15. This is a striking result. It demonstrates that the pattern of lichen distribution

on the largest scales is determined by geological history, not by climate. If the pattern were determined by climate, we would expect the main division to be between tropical lichens on the one hand and those of temperate and cold regions (of both hemispheres combined) on the other. That is not what we observe. Moreover, the result is robust: it is observed unambiguously whatever variant of the algorithm is employed. It was also observed with earlier, less complete, versions of the dataset.

It seems natural, therefore, to refer to the southern and northern elements as Gondwanan and Laurasian respectively, and I will do so. The Gondwanan lichen region comprises South America, the Caribbean, Central America (excluding Mexico), sub-Saharan Africa, the southern part of the Indian subcontinent, SE Asia, Malesia, Australasia, Antarctica, and the Pacific. This corresponds fairly well with Gondwana as defined geologically, the discrepancies being only at the edges, as one would expect.

Within the Gondwana region the primary split is between the predominantly tropical flora of the New World and most of sub-Saharan Africa (groups 6–8), versus the cold to temperate parts of Gondwana plus the tropical element provided by the India to Malesia region (groups 1–5).

Groups 1–3 could be described as cold and temperate Gondwana. Group 1 is Australasia, excluding the humid tropical part of Australia. Within group 1, temperate Australia is clearly distinct from New Zealand and Tasmania. The sister to group 1 is groups 2 (Antarctica) and 3 (southernmost Africa). The mid Atlantic islands of Ascension and St. Helena group here with southernmost Africa, though in earlier investigations their position in the dendrogram was found to be rather sensitive to the algorithm and dataset used. In this run, Argentina and Chile do not group with cold and temperate Gondwana, but in earlier investigations they sometimes did.

Groups 4 and 5 represent the tropical region of Gondwana, extending from the southern part of the Indian subcontinent through Malesia to humid tropical Australia. The

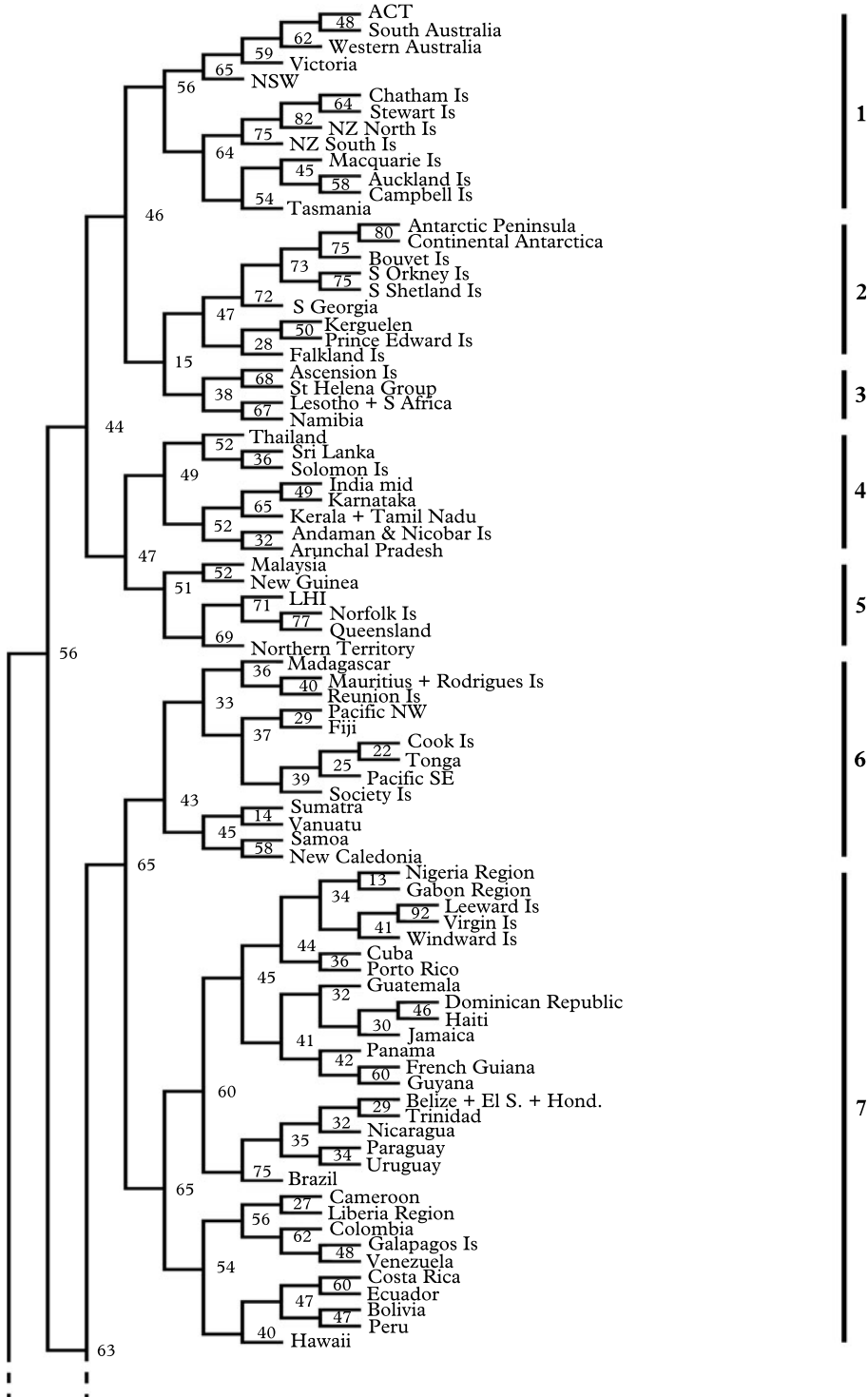


Fig. 1. Cluster analysis of world lichen distribution data.

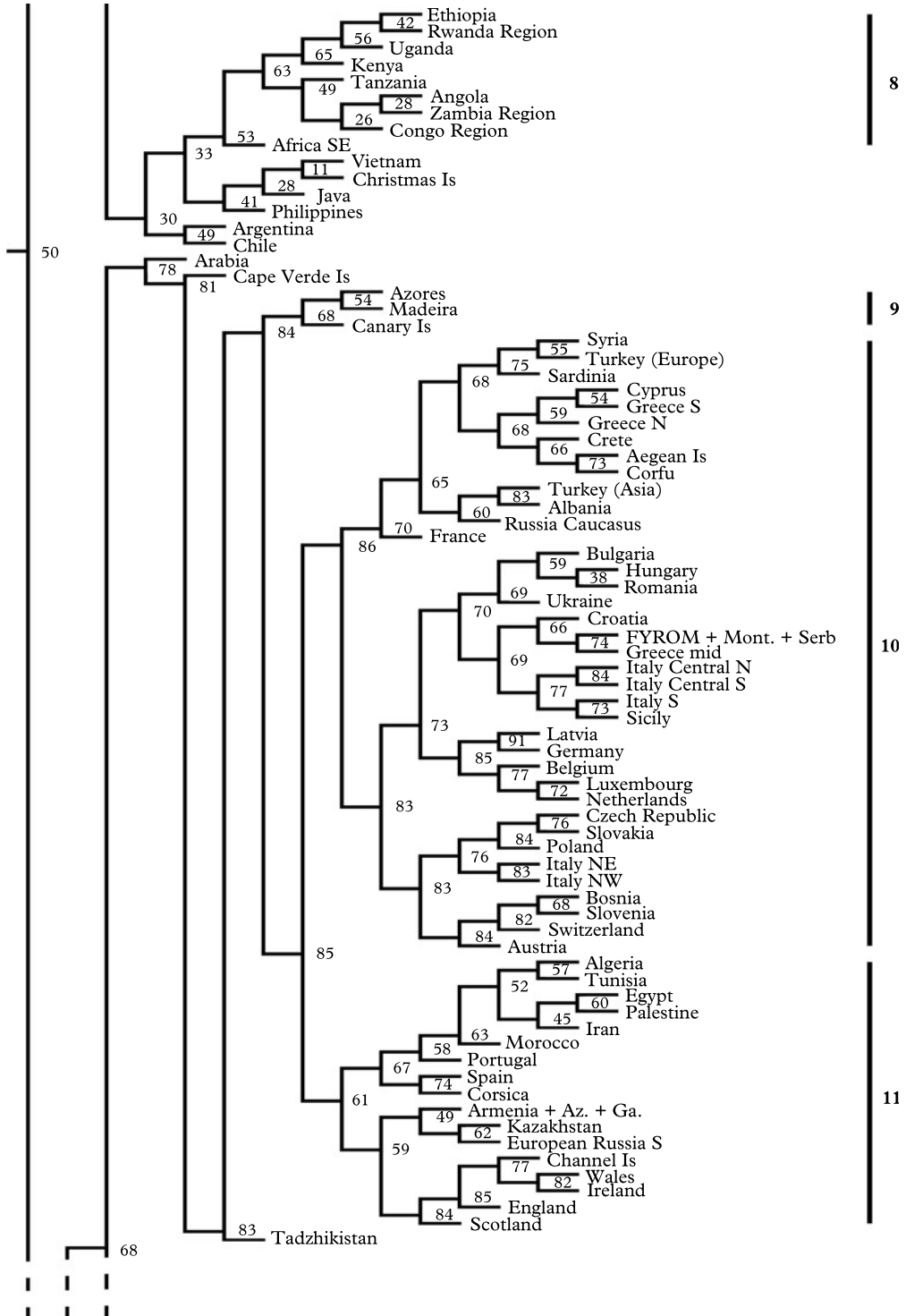


FIG. 1. Continued

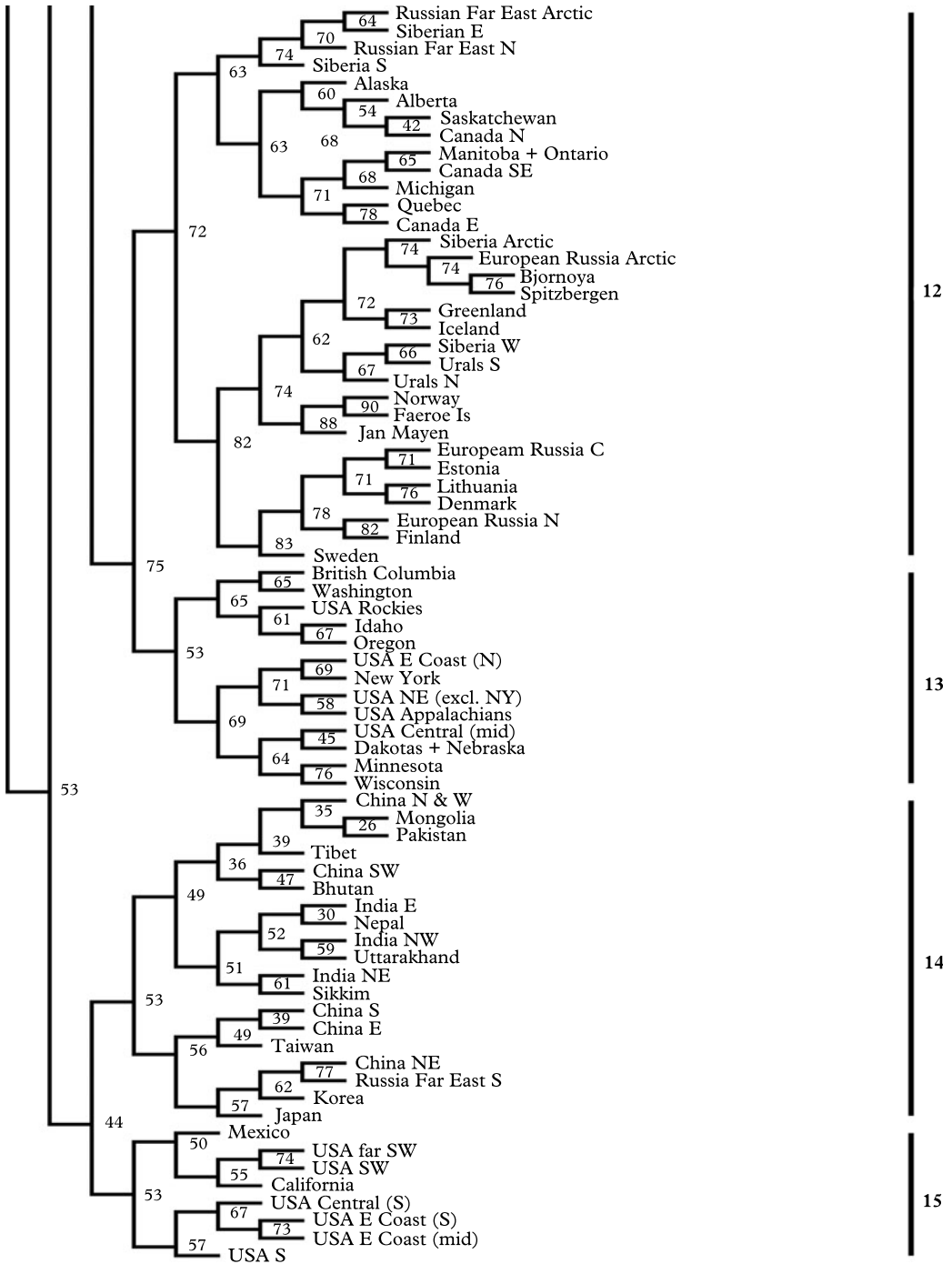


FIG. 1. *Continued*



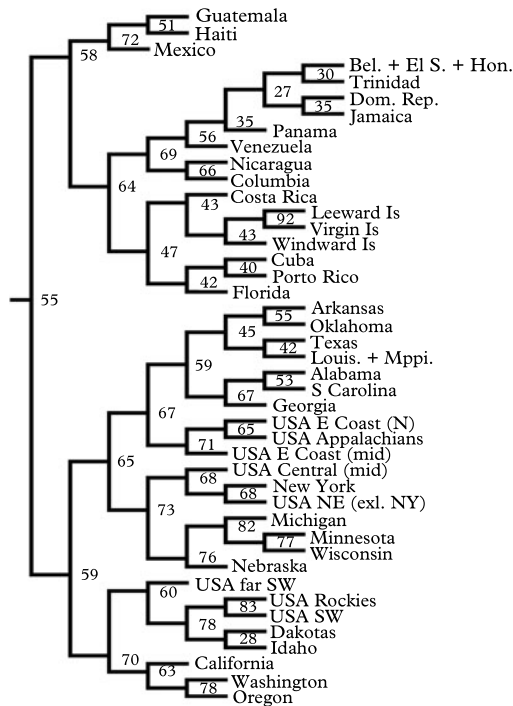


FIG. 2. Cluster analysis of lichen distribution: the Gondwana / Laurasia boundary in the Americas.

internal subdivisions do seem to be meaningful.

The rather heterogeneous group 6 includes most of the Pacific, though it also contains Madagascar and the nearby islands of the Indian Ocean. Possibly what these islands have in common is a warm, very oceanic climate.

Group 7 could be described as the tropical New World. It consists of South and Central America and the Caribbean, excluding only the southern, temperate end (Argentina, Chile). It does not include Mexico. It also includes a few small countries of tropical western Africa, but their lichen flora is very poorly known and it is unclear whether they really belong here. It also includes Hawaii. In all the analyses that I have made, including those with different algorithms and with much earlier versions of the dataset, the tropical New World has always been an obvious group. However, the subdivisions within group 7 are not so easy to interpret.

Group 8 contains most of sub-Saharan Africa, except for the southernmost part. It is sister to a rather heterogeneous assemblage of regions in SE Asia, whose lichen flora is not well known, and its grouping close to them may be an artefact. These two groups combined are sister to Argentina and Chile, which may also be an artefact.

Within Laurasia, the primary split is between cold and cool regions plus Europe and adjacent regions (groups 9–13), versus eastern Asia and warm temperate North America (groups 14 and 15).

Groups 9–11 comprise temperate Europe (group 10) and some peripheral regions. Macaronesia (group 9) is well defined. Group 11 is heterogeneous, and may be an artefact caused by the algorithm grouping together regions that have in common only that their biota differs from that of the core of temperate Europe. Three subdivisions are apparent within group 11: a Mediterranean region, a ‘Caucasus and surroundings’ region and a strongly maritime region (the British Isles). Whether these three subdivisions really have anything in common might merit further investigation. Group 12 is cold Laurasia; it can be subdivided into a predominantly North American and a Eurasian element. Group 13 is temperate North America.

Group 14 is cool and temperate eastern Asia and group 15 is warm-temperate North America. This link between eastern Asia and warm-temperate North America has been apparent from the start in the investigations that I have made, and it certainly merits further investigation.

### The Gondwana / Laurasia boundary regions

It is of interest to look more closely at some of the boundary regions. This can be done with more resolution than in Fig. 1, because when studying a small part of the world it proved possible to use regions with fewer species, and thus to use more and smaller regions, yet still obtain sensible results.

Figure 2 shows the boundary in the Americas. The Caribbean and Central America (including Mexico) group with Gondwana,

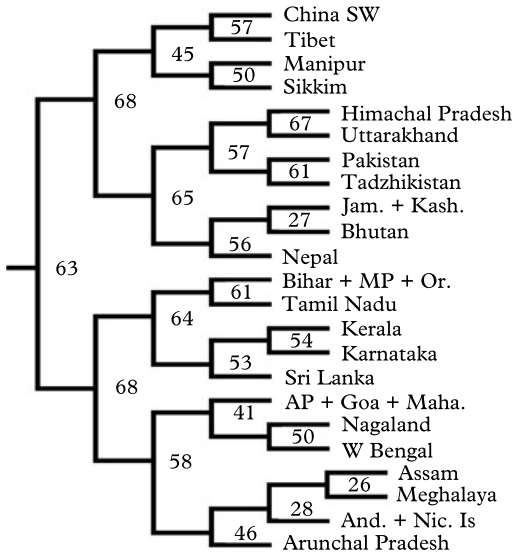


FIG. 3. Cluster analysis of lichen distribution data: the Gondwana / Laurasia boundary near India.

as does Florida. However, the allegiance of Florida and Mexico to Gondwana is marginal; with variant forms of the algorithm and/or earlier versions of the dataset they sometimes grouped with Laurasia, as indeed they do in Fig. 1.

The geological boundary between Laurasia and Gondwana lies at the isthmus of Panama. The lichen boundary lies well inside geological Laurasia, probably somewhere within Mexico. In other words, in the Americas the lichens of Gondwana were more successful at colonizing Laurasia than vice versa. This may be explicable. The northern part of South America had a tropical climate and must have been occupied by lichens adapted to that climate. These lichens would have found it comparatively easy to colonize the Caribbean and much of Central America, which also had warm, humid climates. On the other hand, southern North America had only small areas that were (sub)tropical, and its lichens would thus have found it more difficult to colonize southwards.

North American mammals were more successful in invading South America than the other way round. The explanation often given is that mammals had experienced more intense competition among themselves

in Laurasia than in Gondwana, essentially because of the larger area of the former, and were thus better adapted than, and so able to out-compete, the native Gondwanan fauna. This explanation sounds plausible, but it ought to apply to lichens too. The fact that it does not suggests that it may not be the whole story even for mammals.

The boundary in Africa is at the Sahara. This is clear from Fig. 1, and a more detailed analysis confirms it. The lichens of North Africa and the Iberian Peninsula, both of which were geologically part of Gondwana, group with those of Eurasia. This is unsurprising, as North Africa and Iberia have been in close proximity to Europe for a long time.

The Indian subcontinent is geologically part of Gondwana. As Fig. 3 shows, the lichens of most of India belong to Gondwana. Only the northern and NW provinces of India, those adjacent to the Himalayas, group with Laurasia. However, Assam and other parts of NE India group with Gondwana. That observation must be considered together with what is probably the largest single puzzle in the pattern of global lichen distribution, the affinities of the lichens of tropical SE Asia. Although there are less data for this region than one might wish, it seems unambiguously to group with Gondwana, even though geologically it is part of Laurasia. I am unable to offer a convincing explanation. One possibility, however, is that SE Asia is a fairly small region and its indigenous (i.e. Laurasian) lichen flora might have been swamped by invasion from India and Australasia. It might be of interest to analyze the lichen biota of this region genus by genus, to try to determine the geographical origins of the various elements.

Wallace's Line is a major discontinuity between Australasian (i.e. Gondwanan) and Malesian (Eurasian) faunas. It lies just east of Java. Figure 4 shows a clear distinction between the lichens of Australia and Malesia/SE Asia, but the boundary is not at Wallace's Line. (If it were, New Guinea would group with Australia.) This sort of situation is known for other groups of organisms too (Lomolino *et al.* 2010: 379–381), probably because Wallace's Line marks what has always been a significant extent of water, even

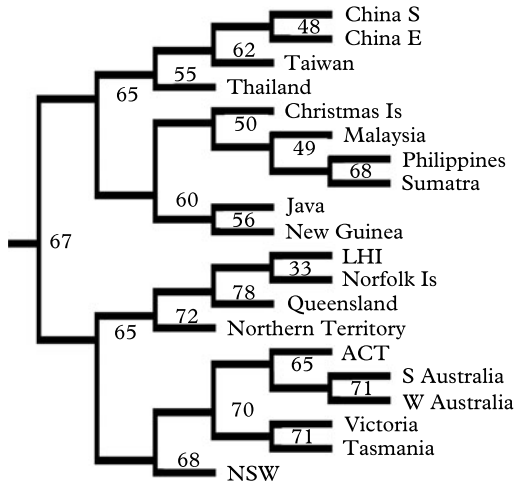


FIG. 4. Cluster analysis of lichen distribution: the Gondwana / Laurasia boundary in Malasia.

at times of low sea level, and thus a barrier to dispersal for the non-volant terrestrial fauna, but less of a barrier to other groups (including lichens which are commonly dispersed by aerial spores or vegetative propagules). However, the lichens of the boundary region are not well known, and further discussion may be premature.

### Lichenicolous fungi

Although there are less data for lichenicolous fungi than for lichens, there is enough to analyze provided that some countries are grouped together. The results in Fig. 5 are based on 60 regions and 4361 (species, region) pairs. The most striking conclusion is that the primary division is into the same Laurasian and Gondwanan elements that were observed for lichens (the only discrepancy is that East Africa groups with Laurasia, perhaps because the available data is scanty). This was entirely unexpected, and strongly reinforces the conclusion that these two subdivisions are real and meaningful.

### Results for individual families and orders

There are enough data to study some individual families and orders, provided that some

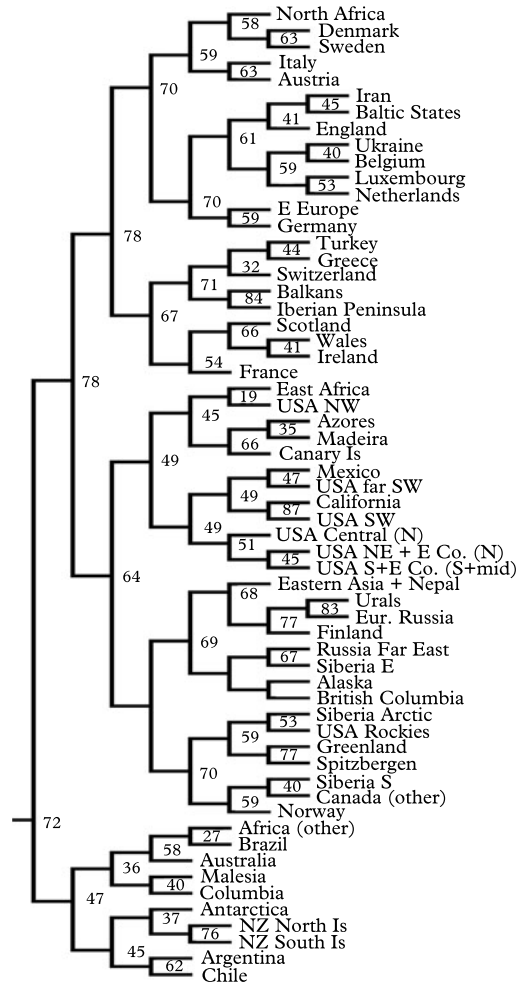


FIG. 5. Cluster analysis of distribution data for lichenicolous fungi.

of the regions used in the global analysis are combined. The choice of taxonomic groups to analyze was made in advance, and all cases are reported here. The taxonomic scheme mostly follows Lumbsch & Huhndorf (2009), with a few updates. The number following the name of the taxon is the number of (species/region) pairs used.

*Acarosporales*: (1281). There are too little data from the Gondwana countries for a satisfactory analysis, even when they are all grouped together as a single region. First split, not very well defined, is into Europe,

western and northern Asia and North Africa, versus everything else. More data from the Gondwana countries, and perhaps improved taxonomy, are needed.

*Agyriales*: (749). First split is into Laurasia and Gondwana (though there is insufficient data to subdivide the Indian subcontinent, which here groups with Laurasia).

*Arthoniomycetes*: (5098). First split is into Laurasia and Gondwana, though a few regions seem misplaced. That southern South America groups with Laurasia and Japan with Gondwana may just be a consequence of insufficient data. The same may be true for China grouping with Gondwana, or it might be the result of a Gondwanan element in warm, humid China.

*Catillariaceae*: (321). Not enough data. There are hints of a Gondwana group but it does not correspond to the top-level split.

*Cladoniaceae*: (5530). First split is into most of Gondwana versus everything else. However, in addition to Laurasia 'everything else' does include a block consisting of parts of Antarctica, southernmost South America and parts of the India to Malesia region. The Caribbean also groups with Laurasia.

*Collematineae*: (6174). First split is into most of Gondwana versus everything else. However, in addition to Laurasia 'everything else' does contain the mid-Atlantic islands of Ascension and St. Helena, most of sub-Saharan Africa (but not southernmost Africa), most of the India to Malesia region and the Pacific.

*Dothideomycetes*: (954). First split is into Europe and northern Asia versus everything else. However, it could also be interpreted as Laurasia versus Gondwana, with North America misplaced in Gondwana. The small dataset makes interpretation difficult.

*Eurotiomycetes*: (9057). First split is into a 'Gondwana with additions' group and an unambiguously Laurasian group. The 'additions' include parts of temperate eastern Asia as well as most of the USA; the Gondwanan and 'additions' elements are not always clearly separated. The historic tendency of lichenologists to assign names from well-studied regions, principally North America and Europe, to the lichens of less

well-studied regions, especially in taxonomically difficult groups, may here be blurring the distinction between Gondwanan and Laurasian biotas.

*Lecanoraceae*: (7092). Gondwana does not appear as the primary split, but most of it appears, as a fairly well defined group (though also containing a few apparently misplaced elements), at the next level down. Its sister group is basically North America. The other half of the primary split is essentially Eurasia, though with a few apparently misplaced elements, including Antarctica.

*Lecanorales*: (45 875). First split is into Gondwana (but minus Antarctica and plus the Himalayan region) versus 'everything else'. Antarctica splits from the rest of 'everything else' at the next step. The apparent misplacement of Antarctica may merely be a result of taxonomic difficulties, and the application of Northern Hemisphere names to what are in fact distinct Southern Hemisphere species.

*Lecideaceae*: (3216). First split is a clear Laurasia / Gondwana division. (Gondwana here excludes East Africa and southern India, because shortage of data made it necessary to merge both of those with surrounding regions.)

*Lichinomycetes*: (1443). Gondwana is not clearly distinguished, though all its parts do lie in only one of the primary subtrees, along with North America, Arabia and Macaronesia. Eastern Asia and the Indian Region, neither of which could be divided further as there was insufficient data, also fall into this subtree.

*Ostropales*: (8795). First split is into a group containing Gondwana, temperate North America and parts of eastern Asia, versus what is basically the Eurasian part of Laurasia.

*Parmeliaceae*: (1693). First split is into Gondwana (but excluding the part from southern India to Malesia and the adjacent Pacific), versus everything else.

*Peltigerales*: (10 772). First split is into Gondwana (but excluding parts of southern India) versus everything else.

*Peltigerineae*: (3958). First split gives American, Australasian and Antarctic parts

of Gondwana versus everything else. Within ‘everything else’, the Africa and India to Malesia parts of Gondwana separate out at the next split.

*Physciaceae*: (11 535). First split is basically into Gondwana and Laurasia, but the Gondwana branch does contain parts of China and northern India, and the Laurasia branch does contain some parts of Central and South America.

*Pilocarpaceae*: (1233). Most of the data for this family is from the Gondwana regions, so a global analysis is not easy to interpret. First split is into the African and American parts of Gondwana, versus everything else. Within ‘everything else’, Laurasia is sister to the remaining (Australasia, Pacific and India to Malesia) parts of Gondwana.

*Pyrenulales*: (1467). There are some patterns, but results are ‘noisy’ and difficult to interpret.

*Ramalinaceae*: (5725). First split is into Gondwana plus eastern Asia and North America, versus everything else. Within the former, Gondwana separates out at the next step.

*Teloschistales*: (16 657). Results are noisy. The main subdivisions of Gondwana are recognizable, but they do not form a coherent group. The overall pattern is hard to interpret. Taxonomic confusion in this group may be a problem.

*Verrucariales*: (5561). First split is into Gondwana and temperate North America versus everything else. Within the former, Gondwana separates fairly cleanly at the next step.

### Discussion

The main regions that Fig. 1 suggests are biogeographically meaningful are more likely to be upheld by future work than refuted. In at least some cases where lichenologists have applied familiar, especially Northern Hemisphere, names to unfamiliar, especially Southern Hemisphere, lichens, we can expect future research to show that the unfamiliar lichen is actually distinct. This will tend to make the biotas of widely-separated regions look even more dissimilar than at

present. The same will also occur if widespread species are recognized to be complexes of cryptic species. The main conclusion of the present work, that there is a strong and easily demonstrable Gondwanan and Laurasian imprint on the world’s lichen biota, is likely to be robust. A secondary conclusion that can be drawn from Fig. 1, that patterns of lichen distribution at smaller scales mainly follow climatic influences, was entirely expected and is also likely to be robust.

This work has also identified one region in particular where the simple Gondwana/Laurasia dichotomy may be an oversimplification. This is tropical SE Asia. To a lesser extent, much of temperate eastern Asia also has an ambiguous character, at least for some taxonomic groups of lichens, and this seems sometimes to be true of temperate North America too. Further work could usefully look in greater detail at these regions.

Most of the tropical areas of the world lie in the Gondwana region (and the one tropical region that does not, SE Asia, behaves anomalously). One might argue that this alone could be enough to separate much of Gondwana from Laurasia in analyses like the present one. However, that argument seems inadequate since the cool and cold regions of the Southern Hemisphere *almost never* group with the cool and cold regions of the Northern Hemisphere.

The Gondwana/Laurasia imprint is easy to understand. The two supercontinents were far enough apart for long enough that evolution must have proceeded largely independently in each. Long distance dispersal was evidently not frequent enough to homogenize the lichen biotas of the two regions. In fact, it has not homogenized even the lichen biota of India, which has been in contact with Laurasia for a long time. By building on these observations it might be possible to determine, quantitatively, how much long range dispersal actually occurs in lichens over any specified timescale. However, that is beyond the scope of this paper.

Most major families of lichens show the same Gondwana/Laurasia imprint, and also have many species within the present area of each supercontinent. This must mean that

those families were in existence before the breakup of Pangaea. If a family had arisen in, say, Gondwana, we would expect today to see many species in the Gondwana regions but rather few in the Laurasian ones, and we would probably not expect to see a clear Laurasia/Gondwana division. This may be the case for *Pilocarpaceae*, which seems predominantly Gondwanan, and possibly for *Acarosporaceae* which seems predominantly Laurasian, though other explanations are possible. The origin of a taxonomic group entirely within Gondwana or Laurasia seems more likely to have occurred only at the level of genera. For example, it seems plausible that *Pseudocyphellaria* may have originated in Gondwana.

The breakup of Pangaea occurred about 150 million years ago. If it is true that nearly all (and perhaps all) the major lichen families were in existence then, one can combine this with cladograms obtained in the usual way

to obtain a calibration point for molecular clocks.

#### REFERENCES

- Feuerer, T. & Hawksworth, D. L. (2007) Biodiversity of lichens, including a world-wide analysis of checklist data based on Takhtajan's floristic regions. *Biodiversity and Conservation* **16**: 85–98.
- Galloway, D. J. (2008) Lichen biogeography. In *Lichen Biology, 2nd Edition* (T. H. Nash, ed.): 315–335. Cambridge: Cambridge University Press.
- Lomolino, M. V., Riddle, B. R., Whittaker, R. J. & Brown, J. H. (2010) *Biogeography. Fourth edition*. Sunderland, Massachusetts: Sinauer Associates.
- Lumbsch, H. T. & Huhndorf, S. (2009) Outline of Ascomycota – 2009. *Myconet* **14**: 1–247.
- Randlane, T. & Saag, A. (2004) Distribution patterns of some primary and secondary cetrarioid species. *Acta Universitatis Upsaliensis. Symbolae Botanicae Upsalienses* **34**(1): 359–376.
- Romesburg, H. C. (2004) *Cluster Analysis for Researchers*. North Carolina: Lulu Press.
- Urbanavichus, G. & Andreev, M. (2010) *A Checklist of the Lichen Flora of Russia*. St. Petersburg: The Russian Academy of Sciences.