Journal of Biogeography (*J. Biogeogr.*) (2007) **34**, 1400–1411

SPECIAL ISSUE

Modelling the known and unknown plant biodiversity of the Amazon Basin

Michael J. G. Hopkins*[†]

Universidade Federal Rural da Amazonia, Belém, Pará, Brazil

ABSTRACT

Aims The overall aim of this study is to provide the data needed for Amazonian conservation and the sustainable management of the region. To this end I model the hypothetical distribution of plant species richness across the Amazon Basin, the distribution of the proportion of this species richness that can be accounted for by described species, and hence the distribution of the biodiversity which remains unknown.

Location Amazonia, Neotropics.

Methods Species richness across the Amazon Basin is estimated by comparing the occurrences of 1584 species of Magnoliophyta whose taxonomy and geographical distributions are relatively well known. These data are used to collate checklists for squares of 1° latitude by 1° longitude. Comparison of the checklists allows estimation of the relative expected diversity in the vicinity of each degree square. Summing the distributions of the hypothetical real ranges gives the proportion of the biodiversity that can be accounted for by described species. Subtraction of the second distribution from the first gives a distribution of the contribution to the overall biodiversity that the model predicts, potentially, results from as yet undescribed species.

Results Collections documented in recent botanical monographs show an extremely biased distribution with the best knowledge being found in a very few relatively well-collected areas. At the degree square level, this model predicts that gamma biodiversity in the Amazon Basin is uniformly high across most of the basin. The model predicts that four large areas of the basin are particularly poorly known, and that they should contain large numbers of uncollected species.

Main conclusions The model presented here highlights the difficulties of quantifying Amazonian plant diversity and its distribution. The low density of collections, and especially their extremely clumped distribution, undermines confidence in theories that seek to explain the apparent distribution of biodiversity. The model's prediction is substantially different from published predictions of the distribution of alpha diversity. Testing of this model in the areas identified as lacunae would require collecting programmes designed to collect fertile material of rare species. If the model's predictions are approximately accurate, the plant biodiversity of the Amazon Basin is considerably underestimated.

Keywords

Amazonia, biodiversity, Bignoniaceae, botanical monography, conservation biogeography, Chrysobalanaceae, *Inga*, Neotropics, Sapotaceae, species modelling.

*Correspondence: Michael J. G. Hopkins, Projeto PPBio, Instituto Nacional de Pesquisas da Amazônia, Caixa Postal 478 Manaus, Amazonas, 69011–970, Brazil.
E-mail: mikehopkins44@hotmail.com

[†] Present address: Instituto Nacional de Pesquisas da Amazônia, Av. André Araújo, 2936, Aleixo, 69060-001, Manaus, Amazonas, Brazil.

www.blackwellpublishing.com/jbi
doi:10.1111/j.1365-2699.2007.01737.x

INTRODUCTION

Mapping the biodiversity of Amazonia is severely compromised by the relatively low density of botanical collections. Furthermore, the density of collections is extremely clumped. Nelson *et al.* (1990) used herbarium data to show that for the genus *Inga* the number of collections was concentrated in a few areas, close to cities or research institutes. Schulman *et al.* (2007), again using herbarium data for the distribution of collecting localities, show that this situation has not changed over the past 15 years, and much of the Amazonian Basin shows no evidence of having been botanically explored. The Brazilian Amazonian herbaria probably contain *c.* 500,000 collections in total, giving an overall collecting density for the area of about one specimen per km^2.

As Schulman *et al.* (2007) point out (see references therein), species distribution maps based on collecting data are very unreliable, as they tend to map the distribution of collection activity rather than the actual distribution of species. Some authors have taken the known distributions of plants based on monographic data to support or refute theories of phytogeographical patterns. Some models (e.g. refugia theory; Prance, 1982) suggest that there are historical explanations for the uneven distribution of certain species, or groups of related species. Although Pitman *et al.* (1999) suggest that widely separated but similar types of Amazonian forest are dominated by a hierarchy of relatively few widely distributed species, there is evidence of local endemism at least for species that occur in specialized habitats (e.g. Gentry, 1986), and inspection of any of the recent botanical monographs for any Amazonian group will reveal many species with apparently very limited distributions.

Very little information is available on the distribution of gamma diversity of plants in Amazonia. Predictions of the distribution of plant diversity across Amazonia have principally used data from inventory plots (typically of trees > 10 cm diameter at breast height (DBH)). Modelling using these data suggests that forests on the upper Amazon are more diverse than those towards the east (ter Steege *et al.*, 2003). Data from this type of plot are of limited use in assessing the distribution of total diversity. The restriction to trees above a certain size limits the analysis to only a part of the total flora, and the typical plot size is inadequate to provide a comprehensive inventory of the locally occurring species. These limitations mean that it is necessary to employ extrapolations to estimate the actual number of species occurring in any locality. Furthermore, the identifications in plot studies are often made by non-specialists, often using only sterile plant collections, and by comparison with herbarium material that may be incorrectly identified. These shortcomings make it difficult to compare checklists between geographically distant areas.

However, for a very few areas relatively complete checklists of the flora have been published. In the Amazonian region the most complete are for the flora projects undertaken in the Ducke Reserve on the outskirts of Manaus, Brazil (Ribeiro *et al.*, 1999), the Iquitos area in Peru (Vásquez, 1997), and the Saül region of French Guiana (Mori *et al.*, 1997, 2002). In all of these studies it is clear that the many of the species occurring in any locality are extremely rare (e.g. Hopkins, 2005), and that undersampling rare species leads to a considerable underestimate of local biodiversity.

The Amazon Basin is poorly known floristically. Large geographical gaps in our knowledge and the small number of herbarium collections available for study for many species impede accurate mapping of plant distributions, mapping biodiversity and identifying regions of endemism. Consequently, it is difficult to plan for adequate conservation and sustainable use of the region's biota. That conservation planners usually use surrogate indicators rather than data based on the organisms themselves, runs the risk that conservation and planning measures may be inaccurate, and hence inadequate.

Herbarium data are intrinsically unreliable for mapping plant distributions. The identification of specimens in herbaria depends on the activity of resident identifiers, visitors to the herbaria and on the sending of duplicates to specialists for identification. The quality of identifications is a serious problem, and in practice many identifications are incorrect or out of date, and thus the amalgamation of species distribution data from several herbaria is likely to result in distributions biased by the eccentricity of the names used in different herbaria. For Amazonian plants, up to 40% of the specimens in the world's herbaria may bear incorrect identifications (Hopkins and Hamada, unpublished data). Furthermore, georeferencing of these specimens is either non-existent (in most collections made more than 15 years ago) or unreliable (when estimated from the available maps). Data in published monographs, in contrast, should be reliable. Monographed taxa have been examined by competent specialists, and there should be a consistency in the published identifications. Monographs also present maps of the known distribution of each species, based on studies by the monographer of all the specimens he or she has been able to examine. There may be errors in the localization of collecting localities, and there may be inconsistencies between monographers of different taxa, but these maps provide the most reliable source of available information.

This paper deals with an attempt to use the data from monographs to predict the distribution of gamma diversity across the Amazon Basin. By using the accumulated data for each degree square it is possible to construct a partial checklist for each and, by comparing the components of each checklist, estimate the local total biodiversity. This reasoning is used to produce maps of the hypothetical distribution of plant diversity for Amazonia. Furthermore, by modelling the hypothetical distributions of the monographed species, that part of the local diversity which can be explained by the distribution of described species can be subtracted from the overall biodiversity, thus giving an estimate of the biodiversity which may be the result of the distributions of species as yet undescribed.

METHODS

The data set used for distribution analysis

The most reliable data on Amazonian plant distribution are to be found in recently published monographs, notably Flora Neotropica. For the present study, data were taken directly from maps in treatments published since 1986. This date was chosen since a monograph published at this time would have had access to the material collected during Projeto Flora Amazonica, which increased the number of collections in Amazonian herbaria by about 50% (Prance *et al.*, 1984). There has been relatively little collecting of fertile material (at least in Brazil) since then (personal observation, from data sets of Amazonian herbaria). Only monographs whose authors plotted their data by degree squares (units of 1° longitude by 1° of longitude) were collated. Additional data came from an unpublished data set of georeferenced collections of the tribe Bignonieae compiled by Lúcia G. Lohmann. The sources and species numbers used are listed in Table 1.

Each of the simulations was performed for the entire data set, and was repeated for four partial subsets of the data (Bignonieae, Sapotaceae, Chrysobalanaceae and *Inga*). If the results of the modelling were similar between the subsets it would suggest that the conclusions are general and thus applicable to all Amazonian plants. However, if the results were substantially different then factors such as the phylogenetic history of the different taxonomic groups, or even the opinions of the responsible taxonomic specialist, might be important factors in the prediction of distribution of the region's biodiversity.

Distribution of species richness

Lists of the species recorded in each degree square between 15°N and 15°S were compiled. For each of these degree squares the contents of the lists of all the squares within a variable radius were tabulated. The radius of the circle enclosing the squares varied between one and nine degrees. Squares with empty lists were discarded from the analysis. These subsets of lists were used to calculate an index of the estimated number of species in the central (focal) degree square. Jack-knife first and second approximation and bootstrap were used for the calculations using the equations given in Colwell (2005).

Estimation of hypothetical distributions

The predicted actual distributions of plant species could be modelled by extrapolation from their known points of collection using topographic, environmental and climatic factors. Algorithms, such as the Mahalanobis distance (Fabron & Kadmon, 2003) and GARP (Stockwell & Peters, 1999), together with data sets of potentially relevant factors for Amazonia, could be used to generate hypothetical distributions to be used in this model. However, this approach was not used here specifically because of the small number of distribution points available for the majority of the species, and because the degree of accuracy of the georeferencing available (plus or minus 50 km) was not considered to be adequate to justify an attempt to correlate the observed distribution with environmental factors.

Using the distributional data set, each occurrence of each species was treated as being a proven occurrence of the species in the area, and hence $P_{\text{species present}}$(lat,long) was set at one in an array that included all possible degree squares between 15°N and 15°S. The probabilities of each species' occurrence in neighbouring degree squares were estimated by allocating a variable likelihood of occurrence relating to the distance from each degree square, with this effect being allowed to occur for variable distances. The effect was additive, so that the probability of a species occurring in a degree square in which there is currently no such record of its occurrence would be greater if several of the surrounding degree squares have confirmed occurrences. The probability of occurrence (P) in a

Table 1 The source of data used in the modelling. The number of species records those registered as occurring between 15°N and 15°S. The number of records is the sum of the number of degree squares registered for each species.

Taxon	Source	n_{species}	n_{records}
Chrysobalanaceae	Prance (1989)	365	4474
Sapotaceae	Pennington (1990)	263	2558
Inga	Pennington (1997)	225	3279
Bignonieae	Lohmann (unpublished data)	393	8538
Arecaceae	Henderson (1990)	11	238
Montagma	Hagburg (1990)	37	495
Parkia	Hopkins (1986)	17	247
Dimorphandra	Silva (1986)	25	146
Cecropiaceae	Berg *et al.* (1990)	68	768
Nectandra	Rohwer (1993)	88	978
Annonaceae	Maas & Westra (1992), Chatrou (1998), Chatrou & He (1999)	70	606
Anacardium	Mitchell & Mori (1987)	8	106
Erisma	Kawasaki (1998)	14	125
Total		1584	22,558

neighbouring square was allowed to vary between 0.1 and 0.9, and the distance of the effect (d) between 1 and 9 degree squares. Thus for example, the probability of occurrence in a square 3° distant from one with a confirmed occurrence would be 0.34 (P^d for $P = 0.7$ and $d = 3$) if the distance allowed was 3 or more, and would be 0.83 ($P^{d}{}_1 + P^{d}{}_2$ for $P = 0.7$) in the same case if there was a second confirmed occurrence 2 degree squares distant ($d_1 = 3$, $d_2 = 2$). In the analysis of the hypothetical distributions generated by this method, an eventual value of $P \geq 0.5$ was used to indicate that that species probably occurs in that degree square.

Examples of distributions generated for two species of *Inga* are shown in Fig. 1. Further examples using other values for distance and probability are illustrated in Appendix S1 in Supplementary Material.

In both these estimates of numbers of species per degree square, the data were standardized to give a measure of the percentage of the maximum recorded in any one degree

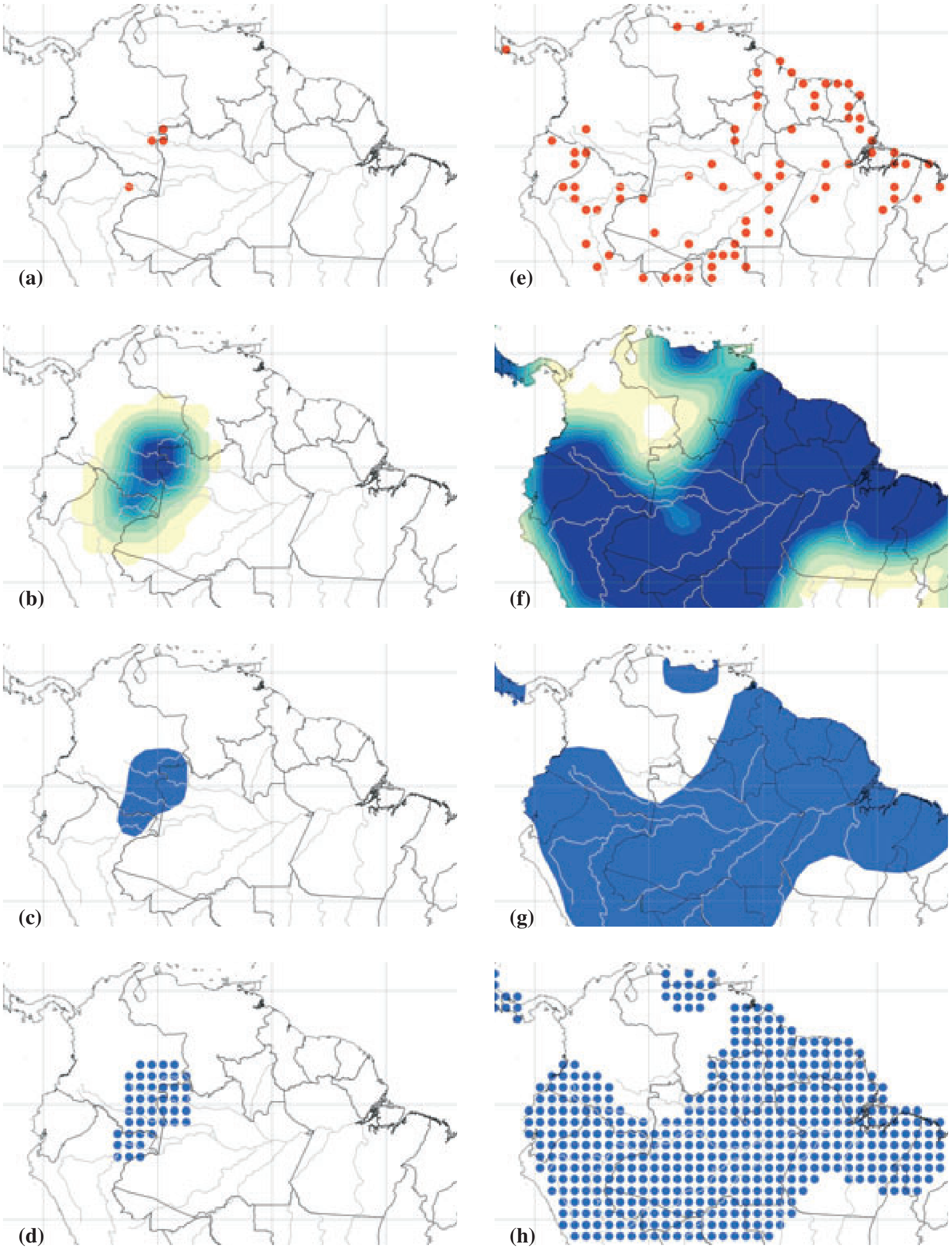

Figure 1 Examples of the steps in producing a hypothetical species distribution for a species with a restricted distribution (a–d, *Inga plumifera*) and a widespread distribution (e-h, *Inga capitata*). (a, e) The degree squares with confirmed occurrences (data from Pennington, 2000). (b, f) The contours of the predicted probability of occurrence, using a probability of occurrence in adjacent degree squares of 0.5 and allowing this effect to accumulate for 5 degree squares. (c, g) The hypothetical distribution deduced by accepting a probability of occurrence of greater than 0.5 in any degree square. (d, h) The degree squares for each species which contribute to the estimate of the total number of species hypothetically occurring in any one degree square.

square. Thus each measure could vary between 0 (no species recorded/expected) and 100 (the maximum recorded/expected across the area). Contours of colour shading of 10% intervals are used in the maps of the results (Figs 2–6) to illustrate the predicted distribution of the elements of the biodiversity.

The programming for the calculations was performed in programs written in Microsoft FoxPro 8.0 and the maps of these distributions were generated using Manifold System® 6.00. The contour maps were generated within Manifold® by generating contours from surfaces imported from tables of results generated in FoxPro.

Choice of values of the variables used

The results of the model depend on the values of the variables selected. See Appendix S1 for the results of 81 simulations using different values for the two variables used by the model. The distance from the focal square in the generation of both hypothetical species richness and hypothetical species distributions were varied from between 1 and 9 (equivalent to circles with approximately 200–1800 km diameter.

It would seem reasonable to assume that many of the apparent gaps in a species' currently known distribution are the result of a lack of collection data, rather than absence. Most species probably have more or less continuous distributions, at least on the degree-square scale being considered here. Thus a model that predicts a species' distribution with few gaps is probably more accurate than one which predicts numerous small lacunae. On the other hand a model that predicts that a species occupies much larger areas than current information supports would be unrealistic. The model should fill in the gaps without exaggerating the likely real distribution. Supplementary Appendices S1a and S1b show the effects of varying the values of the probability of occurrence in adjacent squares (P) from between 0.1 and 0.9 and the distance of the effect (d) from between 1 and 9 degree squares for two species of *Inga*. Lower values of P and d (simulations towards the top and left) tend to fail to fill in the gaps, while higher values (simulations towards the bottom and right) tend to fill in all the available area. For the purposes of this model, intermediate values of P and d of 0.5 and 5, respectively, were used, since they produce hypothetical simulations of distributions which are deemed to be realistic. A similar argument applies to the simulation of the bootstrap model for estimation of the distribution of gamma diversity, and again an intermediate value of 5 degree squares was used. The effects of varying the values of these parameters may be seen in Supplementary Appendices S2, S3 and S4.

RESULTS

The maps presented (Figs 2–6) show graphically the distribution of the parameters under discussion. In general, darker blue indicates a high level of the indicator and light yellow a lower level. The contours are colour-coded for intensity, each variation in shade being a 10% difference from the maximum value recorded.

The distribution of information about species distributions

Figure 2a shows a contour map of the level of knowledge of plant distributions across the region. The areas within the darker blue contours are those where the checklists are relatively long, that is to say, more species have been recorded in those areas than in the less blue, or yellow, areas.

The map shows a highly clumped distribution of information. Peaks of information density (longer checklists) are located in Manaus in Brazil, near Iquitos in Peru, in lowland Ecuador and in French Guiana. Other areas with moderately long checklists are seen close to regional centres, such as state capitals and in regions with relatively easy access such as highways and major rivers. Very large areas appear in yellow, which might indicate areas where plant diversity is actually low, or where no collections have been made.

The distribution of species richness

Figure 2b shows the distribution of species richness estimated by the bootstrap method using a 'medium' distance for inclusion of checklists (radius of 5 degree squares). The plots of the distribution made using other estimators (jack-knife first and second approximation) differed only in detail and the overall pattern was very similar (Supplementary Appendix S2). The principal result is that the central area of the basin, as delimited by the 80–90% of maximum contour, is predicted to have a relatively uniform distribution of high diversity across a very wide area. There is an indication that the distribution of the highest levels of biodiversity may be bimodal, being a little north of the Equator in the east, and a little south of the Equator in the west.

The distribution of the diversity of known species

Figure 2c shows the effect of modelling the distributions of described species by supposing that their true distribution is that estimated by a 'medium' estimate of range extension, 5° of distance with a probability of occurrence of 0.5 in adjacent squares (and accepting their occurrence when the probability in any degree square is ≥ 0.5). This distribution can be considered to be an enhanced version of Fig. 2a, in that the deduced distribution of biodiversity remains with a strong bias for areas that have been better collected. Consequently, well-collected areas would be expected to contain species already known locally, while others, more distant from these well-known areas, can be expected to contain relatively more undescribed species. Further results of the model using different values of the parameters are presented in Supplementary Appendix S3.

The distribution of diversity which remains unknown

Figure 2d shows the result of subtracting the distribution in Fig. 2c from that in Fig. 2b, resulting in a surface with peaks where it is least possible to account for the species expected to

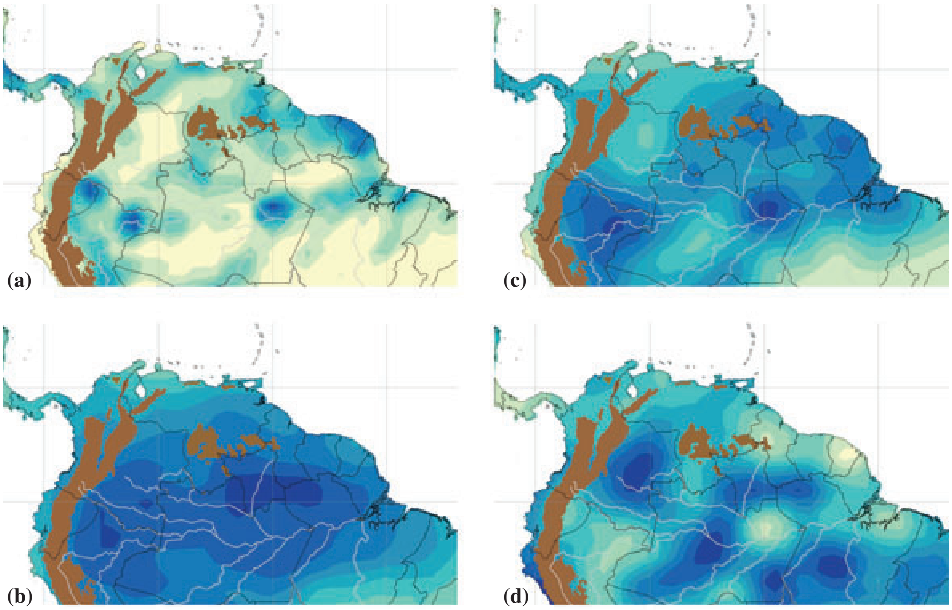

Figure 2 Maps of the results of the steps in the modelling of plant species diversity followed, using the same values of the variables as in Fig. 1. (a) The distribution of information of species occurrences using data from recent taxonomic studies. Deeper shades of blue indicate relatively well known areas, i.e. those where a relatively large number of species are recorded from the local degree squares. The lightest shade of yellow indicates areas where the checklists for the local degree squares have between 0 and 10% of the number of species of the best-known degree square. (b) The distribution of the expected diversity as predicted by a bootstrap model that compares the contents of the checklists within a circle with a radius of five degree squares of the focal square. Deeper shades of blue indicate areas predicted to have the highest numbers of species per degree square. Progressively lighter shades are predicted to have proportionately lower numbers of species per degree square. (c) The distribution of the diversity that can be explained by modelling the distributions of the 1584 species treated in recent monographs as predicted by assuming that each has a likelihood of occurrence of 50% in degree squares adjacent to those where they are already known to occur, and this additive effect extends within a radius of five degree squares. Deeper shades of blue indicate areas predicted to have higher diversities of described species. Lighter shades indicate areas distant from the known occurrences of described species. (d) The distribution of the difference between the predicted distribution of biodiversity (b) and the distribution of diversity of known species (c), and hence the distribution of the lack of taxonomic knowledge of the local biodiversity. Darker blue shades indicate areas where little of the predicted diversity can be explained by our present knowledge, and hence where the largest numbers of undescribed species may be expected to occur. Lighter shades indicate where our knowledge is relatively complete. The brown areas indicate land with an elevation of more than 1000 m. International borders and Brazilian state boundaries are shown in black. Major rivers of the Amazon Basin are shown in grey. Ten shades, from light yellow to dark blue, indicate intervals of 10% with reference to the highest occurring value (light yellow 0–10%, pale green 11–20%…dark blue 90–100%).

make up the local biodiversity. This distribution is in some ways the inverse of Fig. 2c, but it does not simply indicate those areas that have not been collected, as the estimate also takes account of expected local diversity. Thus an area with a short species list will appear in yellow if the local expected biodiversity is also low, but will appear in blue if the local biodiversity is expected to be high. Further results of the model using different values of the parameters are presented in Supplementary Appendix S4.

The pattern seen shows four main regions where botanical knowledge is particularly low, but where biodiversity is expected to be high, and hence these are regions likely to contain relatively more as yet undescribed species. These areas are:

1. lowland Colombia, centred in the area comprising parts of the departments of Vichada, Meta, Guainia, Guaviare and Vaupes;

2. western Amazonian Brazil, within the state of Amazonas, approximately between the cities of Tefé and Envira, comprising the interfluvial region between the Rio Purus and the Rio Juruá, and extending north of the Amazon River as far as the Jaú National Park;

3. northern Amazonian Brazil, extending from north-east Amazonas State across southern Roraima and the portion of Pará State about 300 km north of the Amazon River, and including the southern extremity of Guyana;

4. south-eastern Amazonian Brazil, extending from the south-east corner of Amazonas State (headwaters of the Rio Sucunduri and middle course of the Rio Aripuanã) and southern Pará State, especially the upper reaches of the Rio Irirí and Rio Curuá.

Comparisons between taxa

The four principal groups that contributed to the 'all species' model were analysed separately and the resultant maps shown in Figs 3–6, in the same order as for the combined data. Whilst the details vary between taxa, the overall patterns are similar.

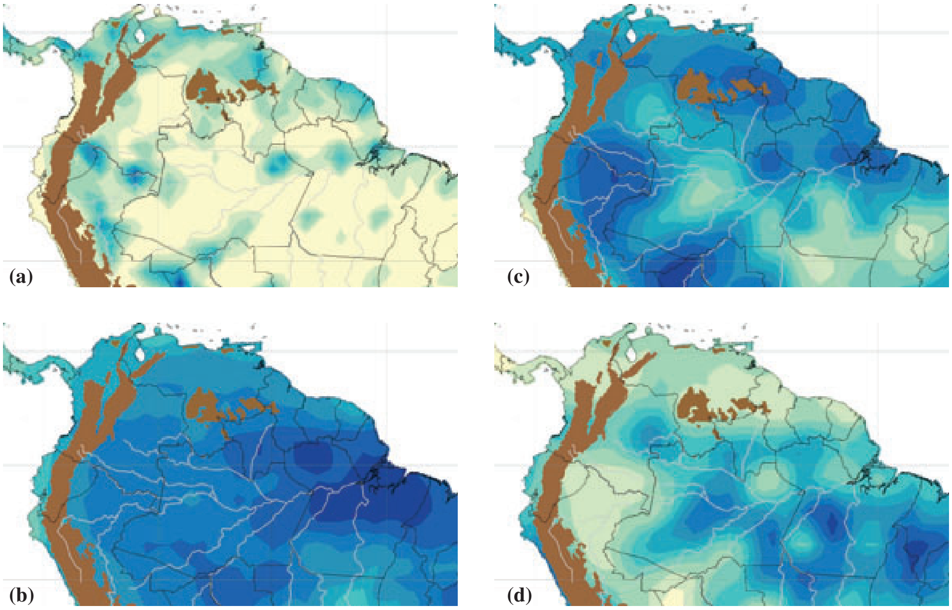

Figure 3 Distributions of diversity measures for species of Bignonieae: (a) distribution of taxonomic knowledge; (b) bootstrap model of predicted diversity; (c) diversity explicable by modelling the distribution of monographed species; (d) predicted distribution of lacunae in taxonomic knowledge. The coding of the colours of the contours is the same as for Fig. 2.

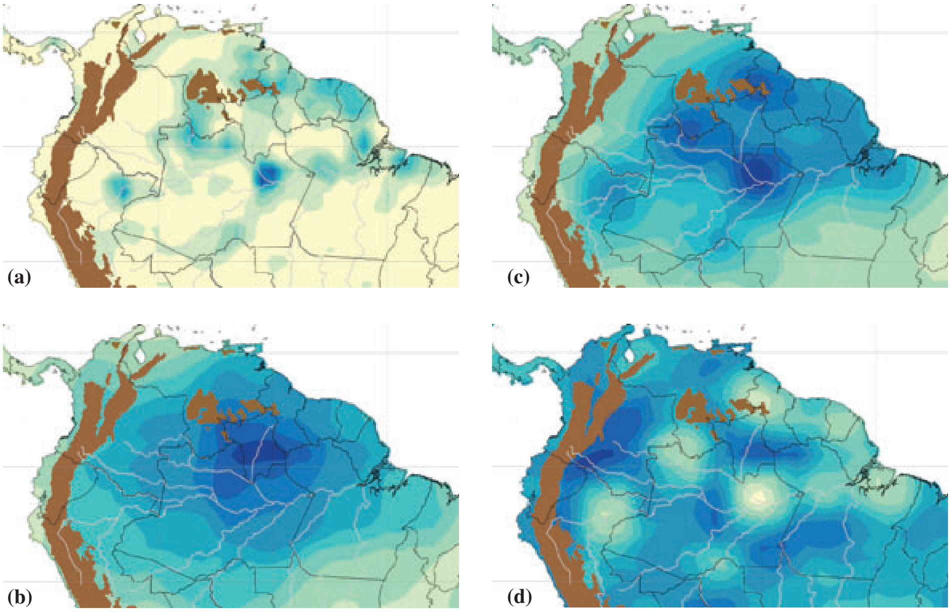

Figure 4 Distributions of diversity measures for species of Chrysobalanaceae: (a) distribution of taxonomic knowledge; (b) bootstrap model of predicted diversity; (c) diversity explicable by modelling the distribution of monographed species; (d) predicted distribution of lacunae in taxonomic knowledge. The coding of the colours of the contours is the same as for Fig. 2.

The possible reasons for the differences in magnitude of the contours are discussed below.

DISCUSSION

Gaps in knowledge

The mapping of the distributional data taken from monographs (Fig. 2a) shows a similar pattern to that shown by measures of collection density (e.g. Nelson *et al.*, 1990) and distribution of density of collection localities (Schulman *et al.*, 2007). The peaks in knowledge of plant species occurrences are highly biased towards relatively few geographical areas. The most pronounced of these are the areas where flora projects have been undertaken or are under way: lowland Ecuador (Jørgensen & León-Yánez, 1999); Iquitos in Peru (Vásquez, 1997); Manaus in Brazil (Ribeiro *et al.*, 1999) and Saül in French Guiana (Mori *et al.*, 1997, 2002). It is arguable whether

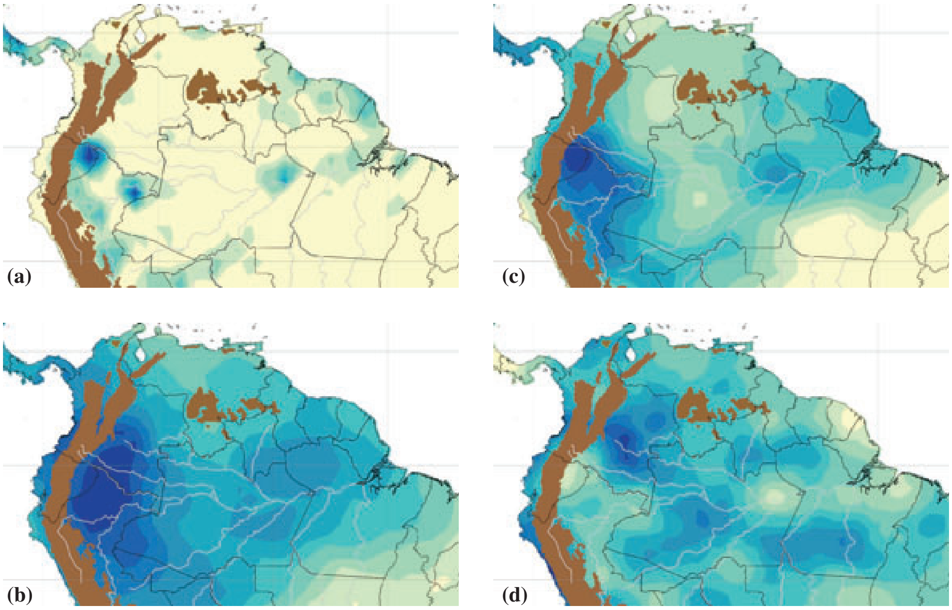

Figure 5 Distributions of diversity measures for species of *Inga*: (a) distribution of taxonomic knowledge; (b) bootstrap model of predicted diversity; (c) diversity explicable by modelling the distribution of monographed species; (d) predicted distribution of lacunae in taxonomic knowledge. The coding of the colours of the contours is the same as for Fig. 2.

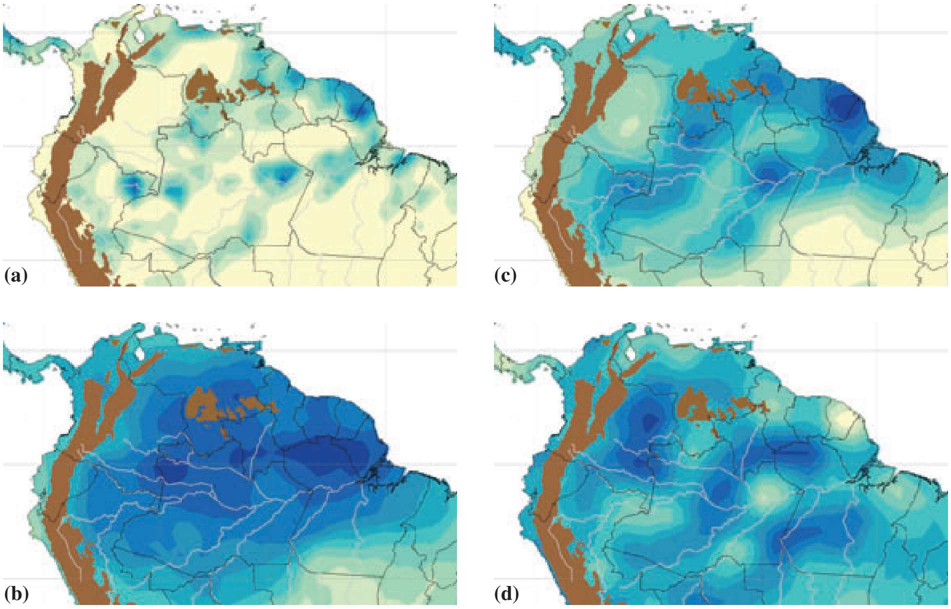

Figure 6 Distributions of diversity measures for species of Sapotaceae: (a) distribution of taxonomic knowledge; (b) bootstrap model of predicted diversity; (c) diversity explicable by modelling the distribution of monographed species; (d) predicted distribution of lacunae in taxonomic knowledge. The coding of the colours of the contours is the same as for Fig. 2.

better knowledge of the local flora facilitates efficient flora projects, or whether flora projects stimulate more collections and hence better knowledge. In Central America, at this scale of degree squares, Panama and Costa Rica are seen to be relatively well collected. Apart from floras, higher levels of knowledge are seen close to cities (for example Belém, Macapá, Porto Velho and Boa Vista in Brazil) though the level of knowledge appears to be much lower than in areas with floras. Major rivers are also sometimes associated with higher information content, such as the Tocantins and Tapajós in Pará State, and the Rio Negro and Rio Madeira in Amazonas State.

The 'white areas' on these maps are very extensive. In Brazil, about 50% of the land area (at the degree-square scale) is shown as light yellow, indicating that these areas are recorded as having between 0 and 10% of the number of species known from the peak areas.

Distribution of diversity

The map of predicted diversity (Fig. 2b) produced by a bootstrap analysis of the distributional data shows a pattern that seems to be fundamentally different from other published maps of Amazonian plant biodiversity.

Lleras *et al.* (1992; cited in Nelson & Oliveira, 2001) modelled the distributions of *c.* 3000 species, using data from Flora Neotropica monographs and data from herbaria. The resulting map (reproduced with modifications in Nelson & de Oliveira, 2001) indicated a peak in species diversity centred on Manaus, and Manaus was also shown as the main centre of endemism in the region. Lleras *et al.* (1992) concluded that these results are artificial, being strongly influenced by collecting activity.

Williams *et al.* (1996), using a smaller data set of 729 distributions of species belonging to five families monographed in Flora Neotropica, presented an analysis similar to the model discussed here. Their maps (presented as colour intensity on a degree-square basis), again show highest species diversity for the Manaus degree square, though the area of highest diversity stretched from north Amazonian Peru across the basin as far as Belém, and the Guianas also showed elevated diversity. However, the area which I estimate should have the highest regional biodiversity (area 3, northern Amazonian Brazil) was not indicated to have a high biodiversity. Williams *et al.* (1996) also presented a map of the distribution of endemism, which shows the lowest levels of endemism in those areas that had been least collected, including the four areas highlighted in the present study.

The maps of alpha diversity produced by ter Steege *et al.* (2003) show a high diversity on the upper reaches of the Amazon River, which gradually declines towards the east. In the current model there is no such decline; in contrast there is possibly a slight increase towards the east, and especially north of the River Amazon.

Various explanations are possible for the contrast among these models. The distribution presented here is theoretically the distribution of gamma diversity (species of all life-forms in all locally occurring habitats), whilst that of ter Steege *et al.* is of alpha diversity of trees (≥10 cm DBH) in single habitat plots in non-inundated forest. If the habitats (at a degree square level) were more homogeneous in the west of Amazonia, the increase in gamma diversity might be explained by a higher level of habitat diversity in the east. Possibly, in areas further from the main river, where there is more altitudinal variation, and a greater diversity of habitats per 100 km^2, there might be more species per degree square, even if each habitat contained fewer species. The same argument would apply to other heterogeneous environmental factors. For example, Fine *et al.* (2005) show that the local distribution of species of Burseraceae is closely associated with soil type. Higher diversity of soil types on a degree-square scale would result in higher gamma diversity.

A second explanation might be that the area to the north of the Amazon represents a zone where species of different phytogeographical biomes meet. If, for example, the ranges of species typically occurring in Amazonian lowland forest and Guianan highland forest interdigitate in the band to the north of the Amazon River, a degree square in this region would be likely to include species from both regions, while degree squares away from this zone of convergence would be likely to include only one set of species. Whilst the definition of phytogeographical regions is also limited by poor knowledge of species distributions, their distribution could also explain gamma diversity distribution at this scale. Thus, Prance's (1977) classification of Amazonian phytogeography, which shows confluence of four regions in north Amazonian Brazil, could be used to explain the hypothetical high diversity of the area, as could the confluence of the Guianan and Amazonian phytochoria of Prance (1989).

A third explanation is methodological. The model used here uses data of occurrence of species in degree squares distant from the focal square. This model could have the effect of overestimating local species diversity as some of the lists being used come from different phytogeographical regions. Hence the bootstrap methodology, which compares the differences between checklists, will predict that many species are missing when there is a lower concordance between checklists. However, if this were a factor, the same argument would apply to the region of interface between the Cerrado and Amazonian forest, which is not apparent in the maps.

Hypothetical distribution of known species

The map (Fig. 2c) of the summed distributions of hypothetical distributions of known (monographed) species also shows a biased pattern, with relatively high levels in northern Peru (and Ecuador) and central Brazilian Amazonia. The uniformity of the tones is also relatively constant throughout Central America and through the Guianas to Belém, suggesting that the species diversity in these areas is relatively easy to explain by the methodology of artificial range extension.

The distribution of missing information

The map (Fig. 2d) of the deduced distribution of the quantity of missing biodiversity information shows four very large areas which this model would predict would be likely to contain large numbers of as yet undescribed (and indeed as yet uncollected) species. This map predicts that intensive collecting in certain areas, notably French Guiana and the immediate area of Manaus and north and central Guyana will result in relatively few undescribed species being collected, while collecting in any of the deep blue areas is likely to result in many new species being discovered.

Comparison between species groups

The four subsets of data for different taxonomic groups show similarities and differences.

Bignonieae (Fig. 3)

Collecting activity has been higher in the west, whilst species diversity is predicted to be higher in the east. As a result the area of lack of knowledge in southern Pará is relatively deeper. This probably reflects the activities of plant collectors, in this case especially Alwyn H. Gentry, who collected extensively in the Andean countries, but very little in Brazil.

Chrysobalanaceae (Fig. 4)

Both collecting density and expected diversity are highest in central Amazonia, and hence the deepest area of poor knowledge is in the most distant of the principal lacunae, in Colombia. The higher level of collection density is largely attributable to the collecting activities of the specialist in the family, Ghillean T. Prance, who resided and worked in Manaus for many years.

Inga (Fig. 5)

In this case, both collecting activity and predicted biodiversity are highest in the west, especially in Ecuador and northern Peru. However, the deepest area of poor knowledge is in nearby Colombia. This probably reflects the large number of species with very few collections from Ecuador, and hence many rare species are also predicted for Colombia, but there has been insufficient collection activity to find them.

Sapotaceae (Fig. 6)

All the maps for Sapotaceae are very similar to those for all the species together. This would seem to indicate that Sapotaceae are a good indicator of the state of taxonomic knowledge across the region.

Testing the hypothesis

If this model reflects the real distribution of plant biodiversity, known and unknown, across Amazonia then there is a great deal of basic research needed before we can state that the Neotropical flora is reasonably well known. The model could be tested by using a standard methodology in areas with different levels of predicted unexplained biodiversity. For example, a flora project 100 km north of Manaus should record many fewer undescribed species than one based in lowland Colombia, in southern Roraima, on the Pará–Surinam border, on the Transamazon Highway in Pará, or on the upper Rio Tefé in Amazonas. However, the methodology of testing would have to be designed to capture the rare species. Short-term expeditions and inventory plots tend not to encounter the rarest species, and especially not encounter them in a fertile state adequate for taxonomic recognition and description. What is needed are intensive flora projects, of several years' duration, designed to ensure that rare species will be collected in flower and fruit, for example following the model used in the compilation of the Reserva Ducke flora (Hopkins, 2005).

ACKNOWLEDGEMENTS

I thank the GEOMA network (Rede Temática de Pesquisa em Modelagem da Amazonia) for financial support during 2003 while the ideas presented in this publication were being developed. I thank the directors of the Museu Paraense Emílio Goeldi, the Instituto Nacional de Pesquisas da Amazônia and Embrapa Amazonia Oriental for hosting me during my activities in Brazil. I thank Lúcia Lohmann for permission to use her unpublished data in these analyses. I especially thank Paulo de Marco, Bruce Nelson and Lyn Loveless for their encouragement.

REFERENCES

Berg, C.C., Akkermans, W.A.P. & van Heusen, E.C.H. (1990) Cecropiaceae: *Coussapoa* and *Pourouma*, with an introduction to the family. *Flora Neotropica Monograph*, **51**.

Chatrou, L.W. (1998) *Changing genera: systematic studies in Neotropical and West African Annonaceae*. PhD Thesis, Utrecht University.

Chatrou, L.W. & He, P. (1999) Studies in Annonaceae XXXIII. A revision of *Fusaea* (Baill.) Saff. *Brittonia*, **51,** 181–203.

Colwell, R.K. (2005) http://viceroy.eeb.uconn.edu/estimates.

Fabron, O. & Kadmon, R. (2003) Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalonobis distance. *Ecological Modelling*, **160,** 115–130.

Fine, P.V.A., Daly, D.C., Muñoz, G.V., Mesones, I. & Cameroni, K.M. (2005) The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the western Amazon. *Evolution*, **59,** 1464–1478.

Gentry, A.H. (1986) Endemism in tropical versus temperate plant communities. *Conservation biology* (ed. by M.E. Soulé), pp. 153–181. Sinauer Associates, Sunderland, MA.

Hagburg, M. (1990) *The genus Monotagma (Marantaceae)*. Dissertation, Department of Systematic Botany, University of Göttingen.

Henderson, A.J. (1990) Arecaceae. *Flora Neotropica Monograph*, **53**.

Hopkins, H.C.F. (1986) *Parkia*. *Flora Neotropica Monograph*, **43**.

Hopkins, M.J.G. (2005) Flora da Reserva Ducke, Amazonas, Brasil. *Rodriguésia*, **56,** 9–25.

Jørgensen, P.M. & León-Yánez, S. (1999) *Catalogue of the vascular plants of Ecuador*. Missouri Botanical Garden, St Louis.

Kawasaki, M.L. (1998) Systematics of *Erisma* (Vochysiaceae). *Memoirs of the New York Botanical Garden*, **81,** 1–120.

Maas, P.J.M. & Westra, L.Y.T. (1992) *Rollinia*. *Flora Neotropica Monograph*, **57**.

Mitchell, J.D. & Mori, S.A. (1987) The cashew and its relatives (*Anacardium*: Anacardiaceae). *Memoirs of the New York Botanical Garden*, **42,** 1–76.

Mori, S.A., Cremers, G., Gracie, C., de Granville, J.-J., Hoff, M. & Mitchell, J.D. (1997) Guide to the vascular plants of central French Guiana. Part 1. Pteridophytes, gymnosperms, and monocotyledons. *Memoirs of the New York Botanical Garden*, **76,** 1–422.

Mori, S.A., Cremers, G., Gracie, C., de Granville, J.-J., Hoff, M., Heald, S.V. & Mitchell, J.D. (2002) Guide to the vascular plants of central French Guiana. Part 2. Dicotyledons. *Memoirs of the New York Botanical Garden*, **76,** 1–776.

Nelson, B.W. & de Oliveira, A.A. (2001) Área Botânica. *Biodiversidade na Amazônia brasileira* (ed. by J.P.R. Copobianco, A. Veríssimo, A. Moreira, D. Sawyer, I. dos Santos and L.P. Pinto), pp. 132–176. Instituto Sociambiental, São Paulo.

Nelson, B.W., Fereira, A.C., da Silva, M.F. & Kawasaki, M.L. (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature*, **345,** 714–716.

Pennington, T.D. (1990) Sapotaceae. *Flora Neotropica Monograph*, **52**.

Pennington, T.D. (1997) *The genus Inga – botany*. Royal Botanical Gardens, Kew.

Pitman, N.C.A., Terborgh, J., Silman, M.R. & Nuñez, V.P. (1999) Tree species distributions in an upper Amazonian Forest. *Ecology*, **80,** 2651–2661.

Prance, G.T. (1977) The phytogeographic subdivisions of Amazonia and their influence on the selection of biological reserves. *Extinction is forever* (ed. by G.T. Prance and T.S. Elias), pp. 195–213. New York Botanical Garden, Bronx.

Prance, G.T. (1982) Forest refuges: evidence from woody angiosperms. *Biological diversification in the tropics: Proceedings of the Fifth International Symposium of the Association for Tropical Biology* (ed. by G.T. Prance), pp. 137–158. Colombia University Press, New York.

Prance, G.T. (1989) Chrysobalanaceae. *Flora Neotropica Monograph*, **9S**.

Prance, G.T., Nelson, B.W., Silva, M.F. & Daly, D.C. (1984) Projeto Flora Amazônica: eight years of binational botanical expeditions. *Acta Amazonica*, **14,** 5–29.

Ribeiro, J.L. da S., Hopkins, M.J.G., Vicentini, A., Sothers, C.A., Costa, M.A. da S., de Brito, J.M., Souza, M.A.D., da Martins, L.H.P., Lohmann, L.G., Assunção, P.A.C.L., Pereira, E.C, Silva, C.F., Mesquita, M.R. & Procópio, L.C. (1999) *Flora da Reserva Ducke*. INPA/DFID, Manaus.

Rohwer, J.G. (1993) Lauraceae: *Nectandra*. *Flora Neotropica Monograph*, **60**.

Schulman, L., Toivonen, T. & Ruokolainen, K (2007) Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. *Journal of Biogeography*, doi: 10.1111/j.1365-2699.2007.01716.x.

Silva, M.F. da (1986) *Dimorphandra* Caesalpiniaceae. *Flora Neotropica Monograph*, **44**.

ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., van der Hout, P., Daly, D.C., Silveira, M., Phillips, O., Vasquez, R., van Ande, T., Duivenvoorden, J., de Oliveira, A.A., Renske Ek, R., Lilwah, R., Thomas, R., van Essen, J., Baider, B., Maas, P., Mori, S., Terborgh, J., Nuñez Vargas, P., Mogollón, H. & Morawetz, W. (2003) A spatial model of tree α-diversity and tree density for the Amazon. *Biological Conservation*, **12,** 2255–2277.

Stockwell, D.R.B. & Peters, D.P. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Systems*, **13,** 143–158.

Vásquez, R. (1997) *Florula de las Reservas Biologicas de Iquitos, Peru*. Missouri Botanical Garden, St Louis.

Williams, P.H., Prance, G.T., Humphries, C.J. & Edwards, K.S. (1996) Promise and problems in applying quantitative complementary areas for representing the diversity of some Neotropical plants (families Dichapetalaceae, Lecythidaceae, Caryocaraceae, Chrysobalanaceae and Proteaceae). *Biological Journal of the Linnean Society*, **58,** 125–157.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article online:

Appendix S1 Examples of the generation of hypothetical species distributions using varying parameters.

Appendix S2 Examples of the results of using varying parameters on the distribution of total plant species biodiversity.

Appendix S3 Examples of the results of using varying parameters on the distribution of biodiversity of monographed species.

Appendix S4 Examples of the results of using varying parameters of mapping the distribution of the lack of knowledge.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2699.2007.01737.x

BIOSKETCH

Mike Hopkins has worked in Brazil for the past 14 years producing identification guides to Amazonian plants, conducting field courses in ecology and plant identification, and working on computerization of herbarium data. He has been associated with the National Amazonian Research Institute (INPA), the Museu Goeldi, with Embrapa Amazonia Oriental and most recently the Federal Rural Amazonian University. The research presented here derives from frustration over the difficulties of working with plants in Amazonia caused by the lack of basic knowledge and multiple difficulties in undertaking research in the area.

Editor: Mark Bush

Special Issue: The inspiration for this special issue on Amazonian biogeography was a meeting of the Association of Tropical Biology held in Uberlândia, Brazil, 24–28 July 2005. The papers comprise several commissioned as an outcome of this meeting, alongside other unsolicited submissions to the journal.