LETTERS

Continental-scale patterns of canopy tree composition and function across Amazonia

Hans ter Steege[1], Nigel C. A. Pitman[2], Oliver L. Phillips[3], Jerome Chave[4], Daniel Sabatier[5], Alvaro Duque[6], Jean-François Molino[5], Marie-Françoise Prévost[7], Rodolphe Spichiger[8], Hernán Castellanos[9], Patricio von Hildebrand[10] & Rodolfo Vásquez[11]

The world's greatest terrestrial stores of biodiversity and carbon are found in the forests of northern South America, where large-scale biogeographic patterns and processes have recently begun to be described[1–4]. Seven of the nine countries with territory in the Amazon basin and the Guiana shield have carried out large-scale forest inventories, but such massive data sets have been little exploited by tropical plant ecologists[5–8]. Although forest inventories often lack the species-level identifications favoured by tropical plant ecologists, their consistency of measurement and vast spatial coverage make them ideally suited for numerical analyses at large scales, and a valuable resource to describe the still poorly understood spatial variation of biomass, diversity, community composition and forest functioning across the South American tropics[9]. Here we show, by using the seven forest inventories complemented with trait and inventory data collected elsewhere, two dominant gradients in tree composition and function across the Amazon, one paralleling a major gradient in soil fertility and the other paralleling a gradient in dry season length. The data set also indicates that the dominance of Fabaceae in the Guiana shield is not necessarily the result of root adaptations to poor soils (nodulation or ectomycorrhizal associations) but perhaps also the result of their remarkably high seed mass there as a potential adaptation to low rates of disturbance.

The trees in our combined database number more than a quarter of a million (277,069). Nearly 40% of these are from the 1968–78 RADAMBRASIL surveys of the Brazilian Amazon[10]. The data set currently includes 89 families and 513 genera. The ten most common families in order of abundance are Fabaceae (legumes), Sapotaceae, Lecythidaceae, Moraceae, Burseraceae, Chrysobalanaceae, Malvaceae, Euphorbiaceae, Lauraceae and Myristicaceae; these account for 74% of all trees. One-quarter of all large trees in the data set belong to the Fabaceae. The ten most common genera in order of abundance are *Eschweilera* (5.6%), *Pouteria*, *Licania*, *Tetragastris*, *Eperua*, *Inga*, *Protium*, *Swartzia* and *Virola*.

Mapping the results of a multivariate ordination of the genus-level compositional data reveals two primary and independent geographical gradients in tree community composition, which together explain 24% of all variance. One gradient stretches from the Guiana shield to southwestern Amazonia (Fig. 1a and Supplementary Fig. S1) and the other from Colombia to southeastern Amazonia (Fig. 1b). The first gradient reflects the fact that none of the ten most abundant genera in the Guiana shield area (*Carapa*, *Lecythis*, *Aldina*, *Pentaclethra*, *Alexa*, *Dicorynia*, *Eperua*, *Catostemma*, *Mora* and *Dicymbe*) are among the ten most abundant genera (or at all) in western Amazonia (*Iriartea*, *Attalea*, *Otoba*, *Oenocarpus*, *Pseudolmedia*, *Ficus*, *Clarisia*, *Sapium*, *Spondias* and *Cecropia*), as well as the fact that seven of the ten top genera of the Guiana shield, but none of the top ten western Amazonian genera, are legumes. This first gradient in large-canopy trees agrees with earlier findings based on smaller trees and analysed at the family level[1,2]. The second gradient, northwest to southeast, is characterized by an increasing abundance of Burseraceae (*Protium* and *Tetragastris*), Bignoniaceae (*Jacaranda* and *Tabebuia*) and Rutaceae (*Fagara*) in drier forests at the southeastern margin of the Amazon basin.

The first gradient is congruent with parallel gradients in community-averaged wood density and seed mass (Figs 1c, d and 2a, b), and regional soil fertility (Fig. 2c). At the Guiana shield end of the gradient, soils are poorer, wood is denser, and seeds are larger. The latter life-history characteristics are important functional traits that can predict species behaviour under various disturbance regimes. Seed mass can be regarded as an indicator of the type of establishment[9,11]: species with small seeds germinate and grow fast in large gaps, whereas species with large seeds mainly survive and grow (slowly) in smaller gaps or the shaded understory. In other words, we consider average community seed mass to be an indicator of the level of landscape level disturbance[9]. Wood density can be regarded as an indicator of adult tree growth and survival[9,11]: models show that species with heavy wood are less competitive in more dynamic forests[9,11]. Because average wood density in a community is also expected to covary with long-term forest disturbance[9], our data indicate that long-term disturbance regimes might be lower on the geologically older and poorer Guiana shield than on the recently deposited sediments of western Amazonia. It has indeed been shown that higher soil fertility leads to higher productivity and, consequently, higher turnover of individuals[3,4]. The correspondence of these gradients to the gradient of tree diversity in the region (Fig. 1h; R^2 axis 1 score with Fisher's $\alpha = 0.35$) thus provides support for the intermediate disturbance hypothesis[12], which predicts that low disturbance rates lead to competitive exclusion and depress diversity. Indeed, mono-dominance is also strongly related to the first compositional gradient ($R^2 = 0.32$; data not shown).

The second major gradient is strongly related to dry season length

[1]Institute of Environmental Biology, Section Plant Ecology and Biodiversity, and the National Herbarium of the Netherlands NHN, Utrecht University branch, Sorbonnelaan 14–16, 3584 CA Utrecht, The Netherlands. [2]Amazon Conservation Association, 1731 Connecticut Avenue NW 3rd floor, Washington DC 20009, USA. [3]Earth and Biosphere Institute, School of Geography, University of Leeds, Leeds LS2 9JT, UK. [4]CNRS/UPS, Évolution et Diversité Biologique, UMR 5174, UPS Toulouse III, Bâtiment IVR3, F-31062 Toulouse, France. [5]Institut de Recherche pour le Développement (IRD), UMR AMAP, TA40/PS2, 34398 Montpellier cedex 5, France. [6]Universidad Nacional de Colombia, Departamento de Ciencias Forestales, Calle 64 x Carrera 65, A.A. 1027, Medellín, Colombia. [7]Institut de Recherche pour le Développement (IRD), UMR AMAP, BP 165, 97323 Cayenne cedex, French Guiana. [8]Conservatoire et Jardin Botaniques de la Ville de Genève, CP 60, CH-1292 Chambésy, Genève, Switzerland. [9]UNEG, Calle Chile, Urbaniz Chilemex, Puerto Ordaz, Edo. Bolívar, Venezuela. [10]Fundacion Puerto Rastrojo, Cra. no. 24-76 of. 1201, Bogotá DC A.A. 241438, Colombia. [11]Proyecto Flora del Perú, Jardín Botánico de Missouri, Jaen, Cajamarca, Peru.

(DSL; Fig. 2d). DSL is known to limit maximum tree α-diversity across the South American tropics[13] (in the current data set, DSL also affects Fisher's α in this manner). The forest inventory data confirm predictions that it also has a major effect on geographic variation in tree community composition across the Amazon basin.

The dominant floristic gradient probably predates the Quaternary period, because the underlying fertility gradient reflects a deep-seated contrast between slow weathering of the quartz-rich Precambrian Guyanan and Brazilian shields in the east, and rapid weathering and deposition of new materials from the Andes in the west, for more than 20 Myr (ref. 14). By contrast, the second gradient may be more dynamic, to the extent that the relative performance of species tracks the constantly shifting Cenozoic climate[15]. Thus, the decreased precipitation and increased temperature predicted for the Amazon basin in some climate change models[16] could result in significant shifts in tree composition along this gradient within human timescales.

Because the first gradient is so strongly associated with the abundance of legumes (Fabaceae; Fig. 1f), we investigated the geographical distribution of two characteristics commonly associated with the family: nitrogen-fixing nodulation and ectomycorrhizal associations, both assumed adaptations to low-fertility soils[17]. Ectomycorrhizal association is apparently insignificant in northern South America (Fig. 1e), apart from small areas in Guyana (where the ectomycorrhizal genus *Dicymbe* is common) and the upper Rio Negro (where the ectomycorrhizal genus *Aldina* is common). Not surprisingly, individuals of nodulating genera are more common on poor soils (Figs 1g and 2e). However, individuals of nodulating genera represent a smaller fraction of all Fabaceae on poor soils than they do on rich soils (Fig. 2f), and nodulation within Fabaceae is strongly negatively correlated with dominance of Fabaceae (Fig. 2g), indicating that some other factor might be responsible for the dominance of legumes in South American forests on poor soils.

One such factor may be seed mass. As shown in Fig. 1c, average community seed mass increases by more than an order of magnitude along the main compositional gradient from the southwest Amazonia to central Guyana; in addition, wood density increases along this gradient. The increase of community seed mass is driven by an increase in seed mass of both legumes ($R^2 = 0.81$) and non-legumes ($R^2 = 0.66$). However, the average seed mass of legume trees

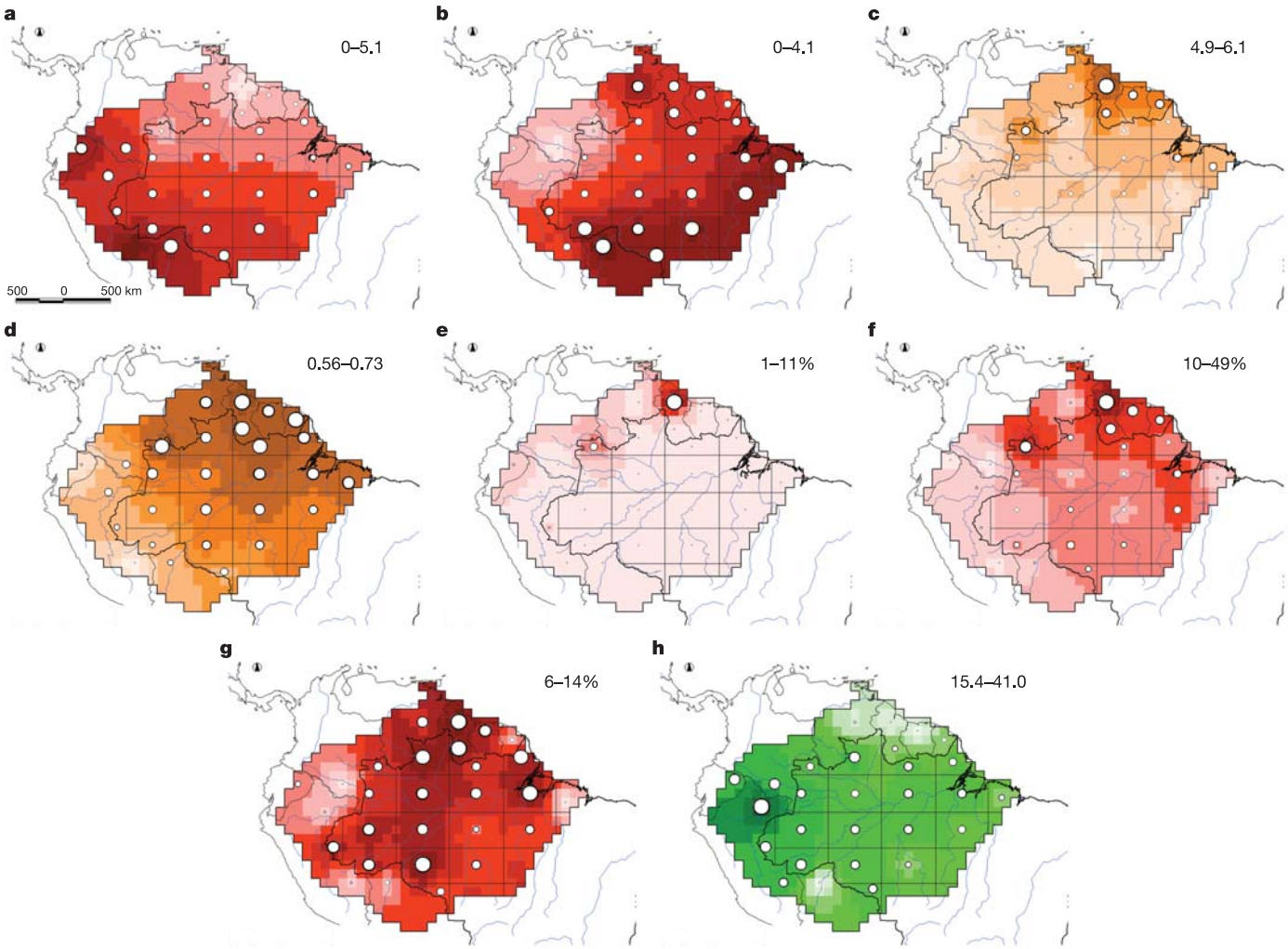

Figure 1 | Geographic variation in community characteristics of South American tree communities. Values in each region are illustrated by the sizes of the open circles and are based on all individuals in that region. Numbers in the top right corners indicate minimum and maximum values. Colours represent an interpolation of the same data by inverse distance weighting on a 1° grid; darker colours indicate higher values. The distance scale in **a** applies to all panels. **a**, Scores on the first axis of the gradient analysis (detrended correspondence analysis) of genus-level community composition. **b**, Scores on the second axis of the same analysis. **c**, Community-weighted seed mass in logarithmic classes (see Methods). **d**, Community-weighted wood density (oven-dried weight divided by green volume). **e**, Proportion of all trees belonging to ectomycorrhizal genera. **f**, Proportion of all trees belonging to Fabaceae. **g**, Proportion of all trees belonging to nodulating genera. **h**, Large tree diversity, calculated with Fisher's α index from the number of individuals and number of genera in each region.

increases more than that of non-legumes (Fig. 2h). The higher wood density in eastern Amazonia and the Guiana shield is also explained by an increase in wood density within Fabaceae ($R^2 = 0.85$) and non-Fabaceae ($R^2 = 0.27$) but their relative increase is identical (Fig. 2h) when plotted against Fabaceae abundance. Thus, on the richer soils of western Amazonia the average seed masses of Fabaceae and non-Fabaceae are similar, whereas on the poor soils of the Guiana shield Fabaceae have a 20% higher average seed mass. This indicates that Fabaceae might be successful in low-dynamics environments because Fabaceae species are better at producing large seeds and, subsequently, more shade-tolerant seedlings, than other families. Janzen[18] has linked the high seed mass of legumes to their relative success on poor soils and correctly predicted that their seeds should be very poisonous to protect them from excessive predation[19].

METHODS

Data gathering. Data were gathered from large-scale national resource inventories, often initiated by the United Nations Food and Agriculture Organization's Forestry Programme. The minimum diameter of the trees recorded was 30 cm, so our analysis is effectively confined to canopy trees. In cases where resource inventories were not available, incomplete, or too small to be representative, they were replaced or supplemented by the large trees of 1-ha inventories (see Supplementary Information).

Harmonizing names and taxonomy between the inventories. Vernacular names in the forest inventory data sets were converted to genus-level scientific names with the lists included in the references used (Supplementary Information). We conducted the analysis at the genus level for three reasons. First, vernacular names are often unreliable at the species level but accurate at the genus level. We estimate that more than 95% of individuals are correctly identified to genus[6,20]. Even so, we quantified the effect of potential 'translation' errors from the vernacular to the genus level by also performing an analysis at the family level (Supplementary Information). Second, because of incomplete sampling and because some species have small range sizes, many species will have been reported from only one inventory location, so that noise in a species-level analysis would overwhelm any underlying regional signal in floristic variation. Third, most species within a genus have similar values for the life-history characteristics we studied[21,22]. Analysis at the generic level thus reduces error and noise in compositional patterns, while not unduly sacrificing the subtlety of signal detectable in life-history characteristics. At the time of the analysis all trees in the database had been classified to family, and more than 98% classified to genus.

All genus names were checked with a variety of sources, notably w^{3}Tropicos (http://mobot.mobot.org/W3T/Search/vast.html), The International Plant Names Index (http://www.ipni.org/index.html) and the Legume Web of the International Legume Database and Information Service (http://www.ildis.org/LegumeWeb/). Taxonomy was standardized to family and genus level as described by the Angiosperm Phylogeny Group[23], because this system offers the best predictability for life-history characteristics[23].

Data analysis. To map compositional patterns (Fig. 1a, b), we first divided the study area into regions, based on the Radambrasil blocks of $4° \times 6°$, and combined all inventories within each region (see Supplementary Information). We converted the number of individuals of each genus in each region to a percentage, to correct for unequal sample sizes. Ordination of all regions was performed with detrended correspondence analysis (DCA; MVSP 3.11; Kovach Computing Services). We used DCA because this method models species occurrence as a unimodal, gaussian, response curve over gradients, just as we expect species to behave. (An ordination with nonmetric multi-dimensional scaling, which does not make this assumption, gave almost identical results; see Supplementary Information).

Life-history characteristics were scored by genus, because most variation is expected above this level (see above). For seed mass (in logarithmic classes: 0, more than 0.00001 to 0.0001 g; 1, more than 0.0001 to 0.001 g; and so on to 7, more than 100 to 1,000 g) we used data from ref. 9. If seed mass data were not available for a given genus in that reference, seed size was taken from refs 24 and

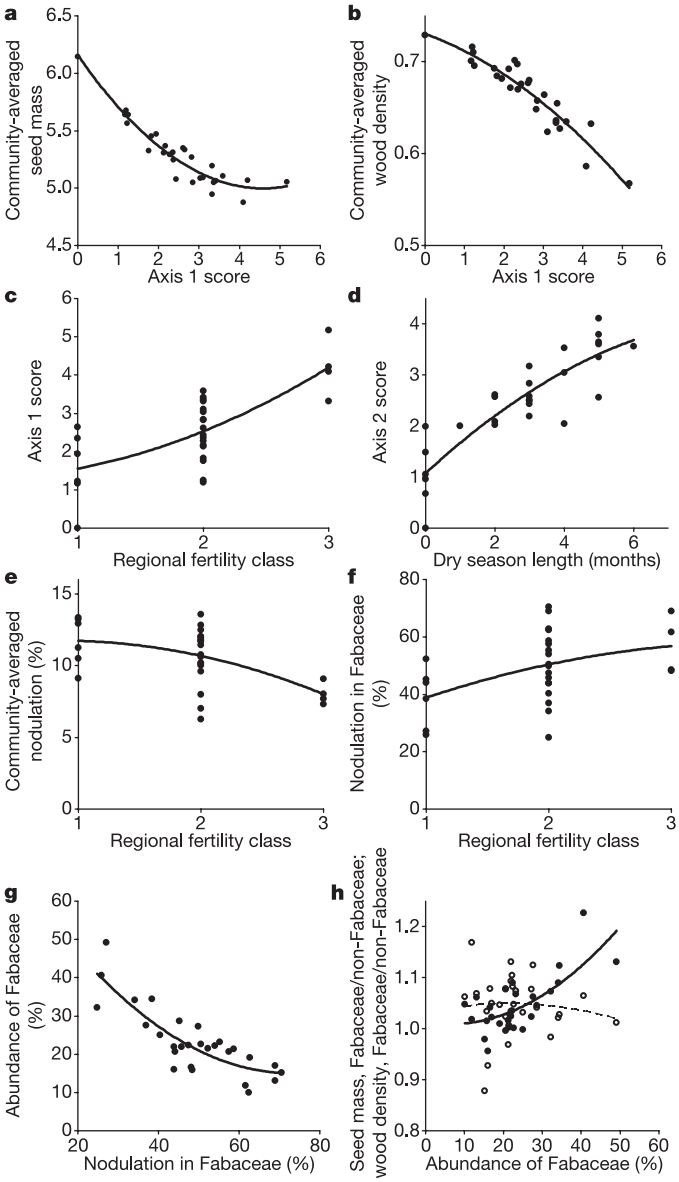

Figure 2 | Relationships between community characteristics of South American tree communities and community characteristics with abiotic factors. a, Community-weighted seed mass is strongly correlated with the primary compositional gradient in the study area. Plots on the left with high average seed mass are found in the Guiana shield, whereas plots on the right with low average seed mass are found in western Amazonia. $R^2 = 0.91$. **b**, Community-weighted wood density (oven-dried weight divided by green volume) is strongly correlated with the primary compositional gradient in the study area. Plots on the left with high wood density are found in the Guiana shield, whereas plots on the right with low wood density are found in western Amazonia. $R^2 = 0.88$. **c**, Regional fertility may be one of the factors influencing the primary compositional gradient in the study area. $R^2 = 0.52$. **d**, DSL is correlated with the secondary compositional gradient in the study area. $R^2 = 0.77$. **e**, The proportion of all trees belonging to nodulating genera is higher where soils are poorer. $R^2 = 0.29$. **f**, The proportion of Fabaceae trees belonging to nodulating genera is lower where soils are poorer. $R^2 = 0.21$. **g**, Abundance of Fabaceae (proportion of individuals of the community that belong to Fabaceae) is negatively related to the proportion of individuals of Fabaceae (in the community) with nodulation. $R^2 = 0.71$. **h**, The ratio of seed mass of Fabaceae individuals to non-Fabaceae individuals in Amazonian tree inventories (black circles and solid line) increases with the abundance of Fabaceae in the sample ($R^2 = 0.53$). Where Fabaceae account for less than 20% of trees, the average seed mass of Fabaceae individuals equals that of non-Fabaceae individuals; where Fabaceae account for more than 20% of trees, seed mass of Fabaceae is higher than that of non-Fabaceae. Where the abundance of Fabaceae is high, wood density of both Fabaceae ($R^2 = 0.68$) and non-Fabaceae ($R^2 = 0.40$) is higher but their relative increase is identical (open circles and dotted line; $R^2 = 0.01$).

25, and from local floras and converted to seed mass in accordance with relationships established in ref. 26. Wood density (oven-dried weight divided by green volume[22]) was taken from refs 9, 20 and 22. These sources provided seed mass data for 97.1% and wood density data for 97.5% of all individuals in the data set.

Data for nodulation among legumes were taken from ref. 17, which provided information for all except three rare genera. Information for ectomycorrhizal infection was compiled from a variety of references. Only five neotropical tree genera in our database were found to be ectomycorrhizal (*Dicymbe*, *Aldina*, *Neea*, *Pisonia* and *Coccoloba*).

When calculating site scores for the life-history characteristics, all individuals were counted. Thus, the values are not averages based on the number of species but real community averages weighted for all individuals in that community.

DSL (number of months per year averaging less than 100 mm of rainfall) was taken from ref. 13 (based on ref. 27) and averaged over the sub-areas of the 28 forest inventory regions.

Following refs 3 and 4, soil quality was classified into classes based on our field data where available, or else inferred from the landform and geographical context available[4]. The spatial unit used in our floristic analysis was large (more than 25,000 km^2), so we used three broad distinctive categories: 1, predominantly poor soil areas (Guianas, Venezuela and upper Rio Negro) with highly oligotrophic spodosols (white sands) and oxisols derived from highly weathered Precambrian sources; 2, edaphically young areas of western Amazonia (Bolivia, Peru, Ecuador and Colombia) with relatively eutrophic inceptisols and ultisols, predominantly derived from Quaternary and Holocene Andean sediment, or Miocene estuarine and marine deposits; and 3, all other areas, with moderately oligotrophic soils, predominantly oxisols derived from Precambrian substrates or weathered Tertiary sediments.

Received 21 June; accepted 3 August 2006.

1. Terborgh, J. & Andresen, E. The composition of Amazonian forests: patterns at local and regional scale. *J. Trop. Ecol.* **14,** 645–664 (1998).
2. ter Steege, H. *et al.* An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. *J. Trop. Ecol.* **16,** 801–828 (2000).
3. Phillips, O. L. *et al.* Pattern and process in Amazon tree turnover, 1976–2001. *Phil. Trans. R. Soc. Lond. B* **359,** 381–407 (2004).
4. Malhi, Y. *et al.* The above-ground coarse wood productivity of 104 Neotropical forest plots. *Glob. Change Biol.* **10,** 563–591 (2004).
5. van Rompaey, R. S. A. R. *Forest Gradients in West Africa*. PhD thesis, Wageningen Univ. (1993).
6. ter Steege, H. The use of forest inventory data for a National Protected Area Strategy in Guyana. *Biodivers. Conserv.* **7,** 1457–1483 (1998).
7. ter Steege, H. & Zondervan, G. in *Plant Diversity in Guyana with Recommendations for a National Protected Area Strategy* (ed. ter Steege, H.) 35–54 (Tropenbos Series 18, Tropenbos Foundation, Wageningen, 2000).
8. Bongers, F., Poorter, L. & Hawthorne, W. D. in *Biodiversity of West African Forests. An Ecological Atlas of Woody Plant Species* (eds Poorter, L., Bongers, F., Kouame, F. N. & Hawthorne, W.) 41–52 (CABI Publishing, Wallingford, 2004).
9. ter Steege, H. & Hammond, D. S. Character convergence, diversity, and disturbance in tropical rain forest in Guyana. *Ecology* **82,** 3197–3212 (2001).
10. RADAMBRASIL. *Levantamento de Recursos Naturais* (Ministério de Minas e Energia, Departamento Nacional de Produção Mineral, Rio de Janeiro, 1968–1978).
11. ter Steege, H. *Long-term Changes in Tropical Tree Diversity: Studies from the Guiana Shield, Africa, Borneo and Melanesia* (Tropenbos Series 22, Tropenbos International, Wageningen, 2003).
12. Connell, J. H. Diversity in tropical rain forests and coral reefs. *Science* **199,** 1302–1309 (1978).
13. ter Steege, H. *et al.* A spatial model of tree α-diversity and -density for the Amazon Region. *Biodivers. Conserv.* **12,** 2255–2276 (2003).
14. Gregory-Wodzicki, K. Uplift history of the Central and Northern Andes: a review. *Geol. Soc. Am. Bull.* **112,** 1091–1105 (2000).
15. Bush, M. B., Silman, M. R. & Urrego, D. H. 48,000 years of climate and forest change in a biodiversity hot spot. *Science* **303,** 827–829 (2004).
16. Houghton, J. T. *et al. Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
17. Sprent, J. I. *Nodulation in Legumes* (Royal Botanic Gardens, Kew, 2001).
18. Janzen, D. H. Tropical blackwater rivers, animals, and mast fruiting by the dipterocarpaceae. *Biotropica* **6,** 69–103 (1974).
19. Rankin, J. M. *The Influence of Seed Predation and Plant Competition on Three Species Abundances in Two Adjacent Tropical Rain Forest Communities in Trinidad, West Indies*. PhD thesis, Univ. of Michigan (1978).
20. Fearnside, P. M. Wood density for estimating forest biomass in Brazilian Amazonia. *For. Ecol. Manage.* **90,** 59–87 (1997).
21. Casper, B. B., Heard, S. B. & Apanius, V. Ecological correlates of single-seededness in a woody tropical flora. *Oecologia* **90,** 212–217 (1992).
22. Chave, J. *et al.* Regional and phylogenetic variation of wood density across 2456 neotropical tree species. *Ecol. Appl.* (in the press).
23. Stevens, P. F. Angiosperm Phylogeny Website, Version 6, May 2005. ⟨http://www.mobot.org/MOBOT/research/APweb/⟩ (2005).
24. van Roosmalen, M. G. M. *Fruits of the Guianan Flora* (Institute of Systematic Botany, Utrecht Univ., Utrecht, 1985).
25. Seed Information Database. *SID release 6.0* ⟨http://www.rbgkew.org.uk/data/sid/⟩ (2004).
26. Hammond, D. S. & Brown, V. K. Seed weight of woody plants in relation to disturbance, dispersal, soil type in wet Neotropical forests. *Ecology* **76,** 2544–2561 (1995).
27. Sombroek, W. Spatial and temporal patterns of Amazon rainfall. Consequences for the planning of agricultural occupation and protection of primary forests. *Ambio* **30,** 388–396 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Guyana Forestry Commission, the Center for Agricultural Research in Suriname (CELOS), Cirad-Forêt, Office National des Forêts (ONF) and Aernout Weeda of Zonisig for making data available for this research; P. Haripersaud for typing in practically all of the RADAMBRASIL data; and P.-M. Forget, M. Werger, H. During and M. Pascal for discussions and comments on the manuscript. This work would not have been possible without decades of dedicated fieldwork by hundreds of colleagues and forestry workers across South America.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.t.S. (h.tersteege@bio.uu.nl).