

# Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas

L. J. Gorenflo<sup>a,1</sup>, Suzanne Romaine<sup>b</sup>, Russell A. Mittermeier<sup>c</sup>, and Kristen Walker-Painemilla<sup>d</sup>

<sup>a</sup>Department of Landscape Architecture, Pennsylvania State University, University Park, PA 16802; <sup>b</sup>Merton College, Oxford University, Oxford OX1 4JD, England; and <sup>c</sup>Office of the President and <sup>d</sup>Center for Conservation and Governance, Conservation International, Arlington, VA 22202

Edited by B. L. Turner, Arizona State University, Tempe, AZ, and approved April 6, 2012 (received for review October 28, 2011)

As the world grows less biologically diverse, it is becoming less linguistically and culturally diverse as well. Biologists estimate annual loss of species at 1,000 times or more greater than historic rates, and linguists predict that 50–90% of the world's languages will disappear by the end of this century. Prior studies indicate similarities in the geographic arrangement of biological and linguistic diversity, although conclusions have often been constrained by use of data with limited spatial precision. Here we use greatly improved datasets to explore the co-occurrence of linguistic and biological diversity in regions containing many of the Earth's remaining species: biodiversity hotspots and high biodiversity wilderness areas. Results indicate that these regions often contain considerable linguistic diversity, accounting for 70% of all languages on Earth. Moreover, the languages involved are frequently unique (endemic) to particular regions, with many facing extinction. Likely reasons for co-occurrence of linguistic and biological diversity are complex and appear to vary among localities, although strong geographic concordance between biological and linguistic diversity in many areas argues for some form of functional connection. Languages in high biodiversity regions also often co-occur with one or more specific conservation priorities, here defined as endangered species and protected areas, marking particular localities important for maintaining both forms of diversity. The results reported in this article provide a starting point for focused research exploring the relationship between biological and linguistic-cultural diversity, and for developing integrated strategies designed to conserve species and languages in regions rich in both.

human dimensions of biodiversity | Biodiversity conservation priorities | endangered languages

Global biodiversity in the early 21st century is experiencing an extinction crisis, with annual losses of plant and animal species estimated to be at least 1,000 times greater than historic background rates (1, 2). Linguistic diversity is experiencing a similar crisis. Language loss in some areas, such as the Americas, has reached 60% over the last 35 y (3), and some linguists predict the disappearance of 50–90% of the world's languages by the end of this century (4). Prior studies have noted that biological and linguistic diversity often occur in the same places. Research conducted at continental and regional scales identified patterns of co-occurrence of linguistic and biological diversity in broad regions, such as West Africa, Melanesia, and Mesoamerica, and in mountainous regions, especially New Guinea (5–8). Previous inquiries noted that nations containing high biological diversity also tend to contain high linguistic and cultural diversity (4, 9–11). Research using geographic information system technology and examining locations of languages as geographic points concluded that ecoregions essential for conserving our planet's habitat types, ecosystems, and representative species often also contain large numbers of languages (12). Such studies have given rise to the notion of biocultural diversity, the tendency for biological, linguistic, and cultural diversity to co-occur (13, 14).

The availability of improved data on geographic distributions of languages and biodiversity enables closer examination of the co-occurrence of linguistic and biological diversity. Here we use

detailed global data showing the geographic extent of more than 6,900 languages, recently compiled by Global Mapping International (15), to analyze linguistic diversity in regions containing much of Earth's biological diversity, biodiversity hotspots and high biodiversity wilderness areas (*SI Text* and *Tables S1–S4*). We focus exclusively on indigenous and nonmigrant languages to identify those corresponding to particular cultural groups, as opposed to languages that have diffused throughout much of the world (such as English and Spanish). Our study covers all languages as well as those found only in individual regions (endemic to those regions), paying particular attention to languages in danger of extinction because of small numbers of speakers. Analyses consider multiple geographic scales, ranging from entire biodiversity regions to locations of protected areas (such as national parks) and sites where individual species occur.

We begin our analysis with a focus on regional conservation priorities defined by biodiversity hotspots and high biodiversity wilderness areas (Fig. 1*A*) (16). Hotspots are regions characterized by exceptionally high occurrences of endemic species and by loss of at least 70% of natural habitat (17). Totalling only about 2.3% of the earth's terrestrial surface, the remaining habitat in 35 hotspots contains more than 50% of the world's vascular plant species and at least 43% of terrestrial vertebrate species as endemics (18, 19). High biodiversity wilderness areas, also rich in endemic species, are large regions (minimally 10,000 km<sup>2</sup>) with relatively little human impact, having lost 30% or less of their natural habitat (20). Remaining habitat in the five high biodiversity wilderness areas, covering about 6.1% of the earth's terrestrial surface, contains roughly 17% of the world's vascular plant species and 8% of terrestrial vertebrate species as endemics.

The geographic distribution of languages indicates concentrations in regions of high biodiversity (Fig. 1*B*). A total of 3,202 languages, nearly half of those on Earth, currently are found in the 35 biodiversity hotspots (Fig. 2*A*). Hotspots with particularly high linguistic diversity include the East Melanesian Islands, Guinean Forests of West Africa, Indo-Burma, Mesoamerica, and Wallacea, each with more than 250 indigenous languages. In contrast, the Chilean Forests, Cape Floristic Region, New Zealand, Southwest Australia, and Succulent Karoo hotspots all contain three languages or fewer. Some 1,622 different languages occur in the five high biodiversity wilderness areas. The linguistic diversity of these regions is dominated by the New Guinea Wilderness Area, with 976 languages. We show detailed results of these and other analyses in *Tables S1–S3* and the *SI Text*.

A total of 2,166 of the languages in the biodiversity hotspots are endemic to individual regions (Fig. 2*A*). The Indo-Burma, East

Author contributions: L.J.G. and S.R. designed research; L.J.G. performed research; L.J.G. analyzed data; and L.J.G., S.R., R.A.M., and K.W.-P. wrote the paper.

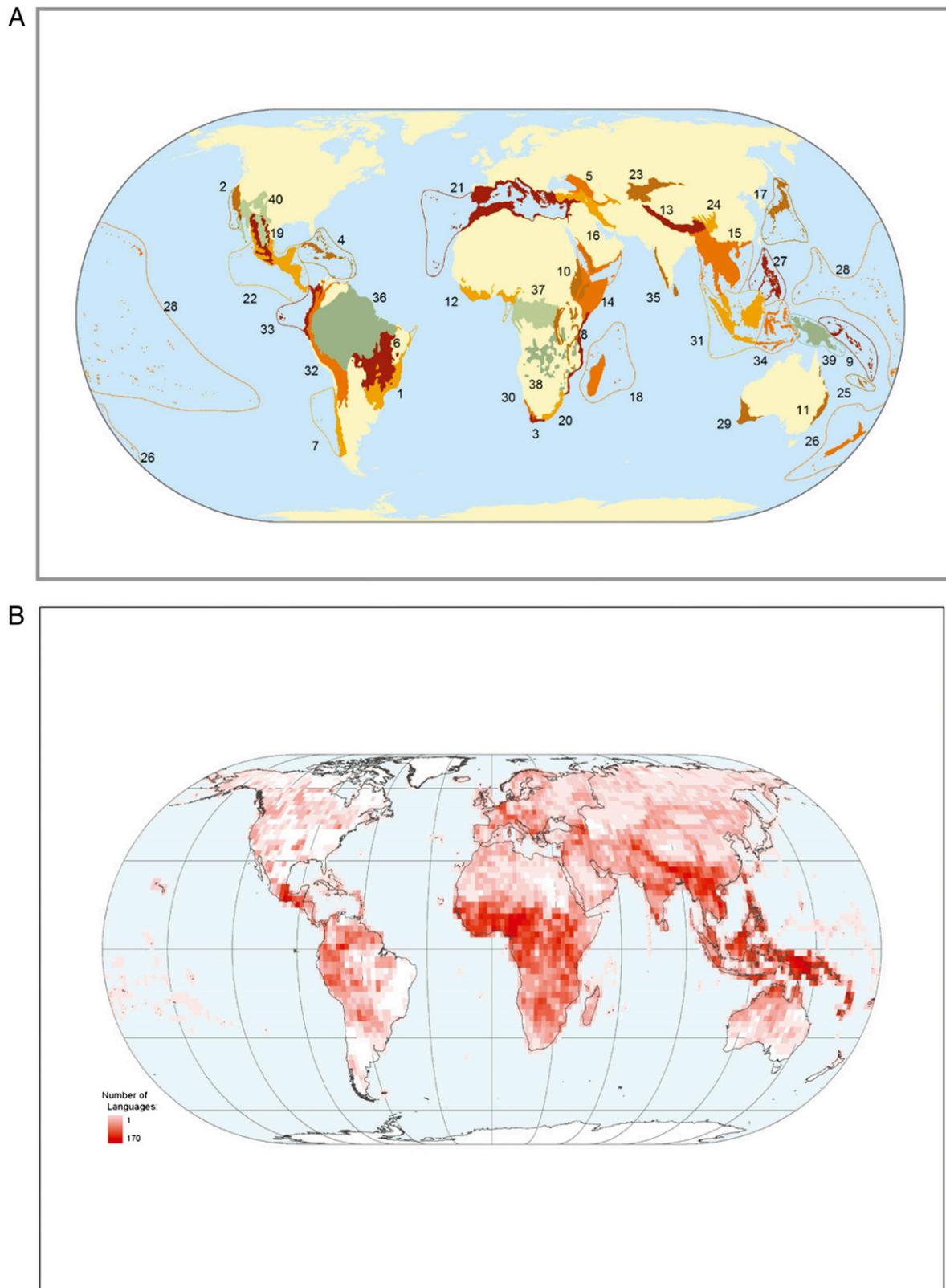
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

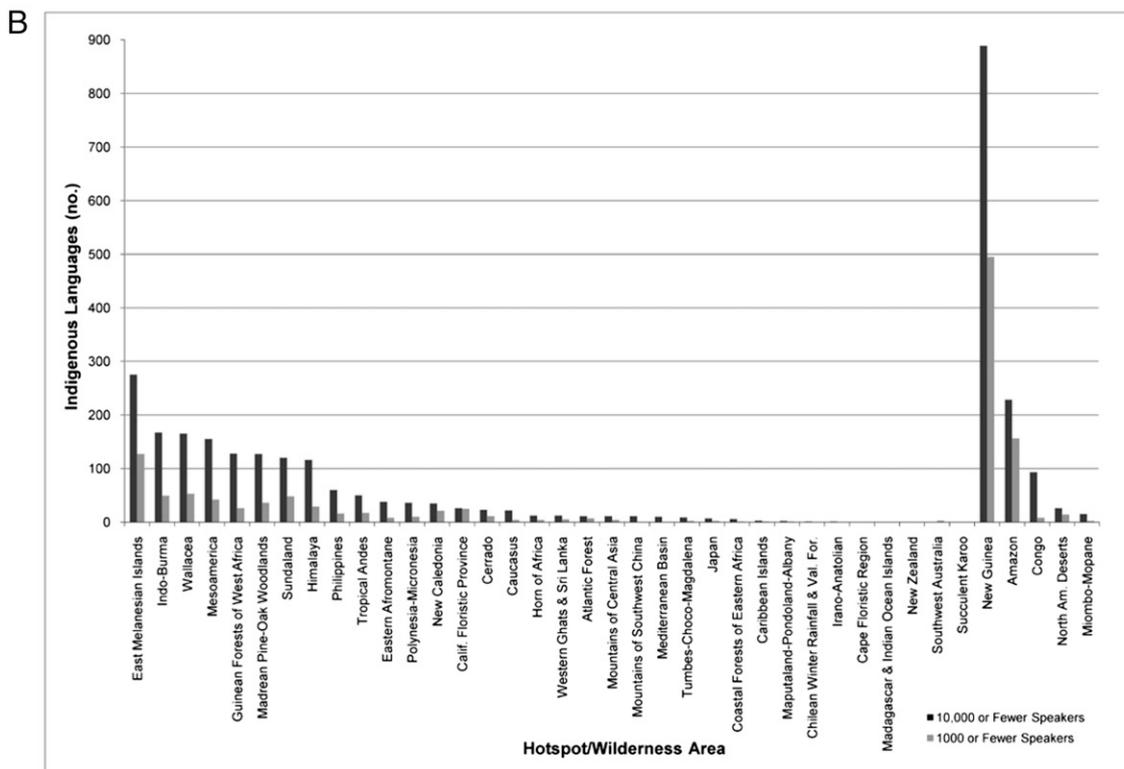
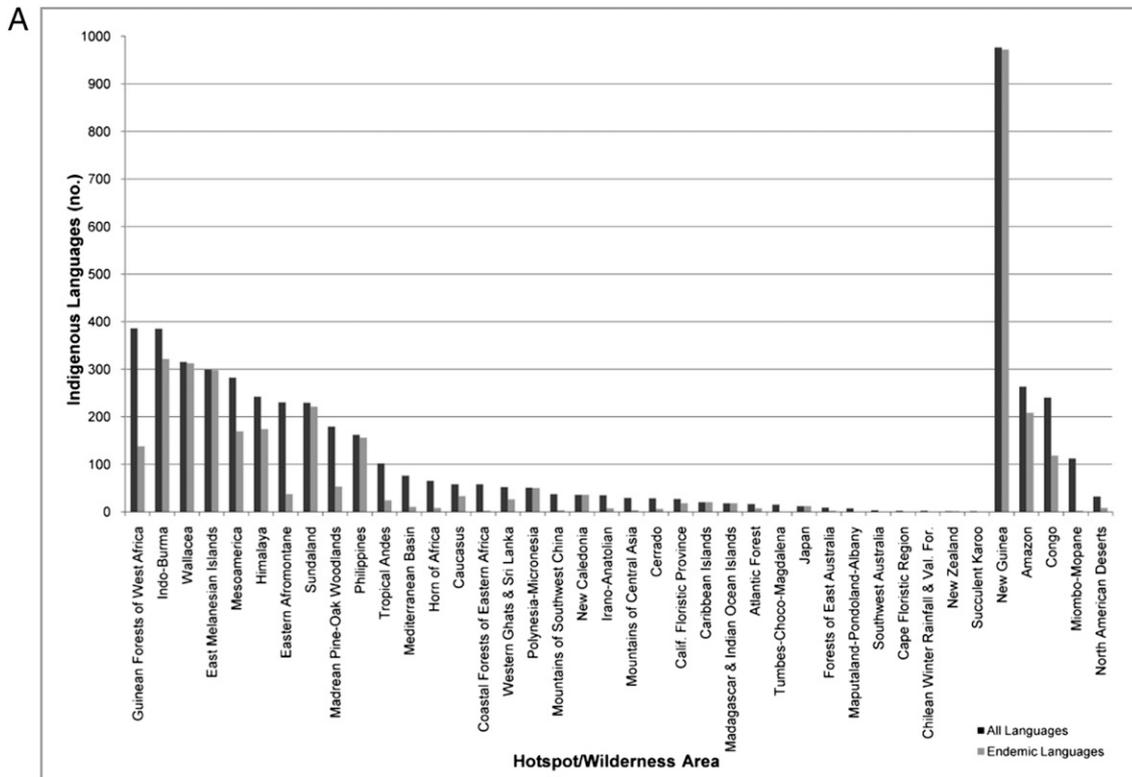
Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. E-mail: ljg11@psu.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1117511109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1117511109/-DCSupplemental).



**Fig. 1.** (A) Biodiversity hotspots (regions 1–35) and high biodiversity wilderness areas (regions 36–40). 1: Atlantic Forest; 2: California Floristic Province; 3: Cape Floristic Region; 4: Caribbean Islands; 5: Caucasus; 6: Cerrado; 7: Chilean Winter Rainfall-Valdivian Forests; 8: Coastal Forests of Eastern Africa; 9: East Melanesian Islands; 10: Eastern Afrotropical; 11: Forests of East Australia; 12: Guinean Forests of West Africa; 13: Himalaya; 14: Horn of Africa; 15: Indo-Burma; 16: Irano-Anatolian; 17: Japan; 18: Madagascar and the Indian Ocean Islands; 19: Madrean Pine-Oak Woodlands; 20: Maputaland-Pondoland-Albany; 21: Mediterranean Basin; 22: Mesoamerica; 23: Mountains of Central Asia; 24: Mountains of Southwest China; 25: New Caledonia; 26: New Zealand; 27: Philippines; 28: Polynesia-Micronesia; 29: Southwest Australia; 30: Succulent Karoo; 31: Sundaland; 32: Tropical Andes; 33: Tumbes-Chocó-Magdalena; 34: Wallacea; 35: Western Ghats and Sri Lanka; 36: Amazonia; 37: Congo Forests; 38: Miombo-Mopane Woodlands and Savannas; 39: New Guinea; 40: North American Deserts. (B) Geographic distribution of indigenous and nonmigrant languages in 2009.



**Fig. 2.** Bar charts showing the occurrence of the following by biodiversity hotspot and high biodiversity wilderness area. (A) All languages and languages endemic to individual regions. (B) Languages spoken by 10,000 or fewer and 1,000 or fewer people.

Melanesian Islands, Sundaland, and Wallacea hotspots have particularly high linguistic endemism, each with 220 or more languages unique to those respective regions. In contrast, 6 hotspots contain no endemic languages, and 10 more contain 10 or fewer endemic

languages. Several hotspots with considerable linguistic diversity have much less linguistic endemism, most notably the Guinean Forests of West Africa, Eastern Afrotropiane, and Mesoamerica. Some 1,308 languages are endemic to the high biodiversity

wilderness areas. The New Guinea Wilderness Area again contains the greatest number of endemic languages, totaling 972.

Many of the languages occurring in biodiversity hotspots and high biodiversity wilderness areas are spoken by relatively few people. Although various factors affect language vitality (in particular, the extent of intergenerational transmission), size may be the best generally available proxy for risk of language loss. Languages spoken by small numbers of people can disappear much faster than languages spoken by larger numbers of people because the vulnerability of small groups to external pressures in a rapidly changing world. Here we consider two thresholds to identify potentially endangered languages: those with 10,000 or fewer speakers and those with 1,000 or fewer speakers. Of the 3,202 languages in the hotspots, 1,553 are spoken by 10,000 or fewer people (Fig. 2B). Some 544 of those languages are spoken by 1,000 or fewer people. In the high biodiversity wilderness areas, of the 1,622 total languages 1,251 are spoken by 10,000 or fewer people, and 675 are spoken by 1,000 or fewer. As shown in Fig. 2B, the number of potentially endangered languages varies greatly by biodiversity region and depends greatly on the threshold used.

Other researchers have argued for a positive relationship between linguistic and biological diversity (4, 10, 12, 21–23). Using results from the current study and comparing the number of languages per region against total vascular plant species per region—a defining criterion for hotspots and high biodiversity wilderness areas—suggests a positive relationship; linear regression indicates a weak although significant ( $P < 0.05$ ) relationship with a Pearson's  $r$  value of 0.33, and calculating a Spearman's coefficient indicates a significant relationship ( $P < 0.02$ ) with an  $r_s$  value of 0.40. Comparing the number of endemic languages per region against total endemic vascular plant species similarly suggests a positive relationship, albeit weak, yielding a Pearson's  $r$  of 0.28 ( $P < 0.10$ ) and a Spearman's  $r_s$  of 0.30 ( $P < 0.10$ ). Examining the relationship between the number of languages and species in other taxa per region provides similar evidence of a weak positive association between linguistic and biological diversity (see *SI Text*). We discuss the possible explanations for these correlations below.

To refine our analysis geographically, within high biodiversity regions we examined the co-occurrence of languages with more precise definitions of conservation priorities, representing individual species, combinations of species, and key localities. We identified conservation priorities as endangered amphibians and existing protected areas, in both cases supported by datasets that include precise geographic localities beyond their presence in a particular biodiversity region (as is the case for vascular plants). We based our analysis of amphibians on data from the World Conservation Union's (IUCN's) Global Amphibian Assessment, a compilation by more than 500 experts of data that defined geographic range and population status of 5,743 described species of amphibians (24); the present study used data updated in 2006 to include such information on 5,816 species. We used the IUCN Red List to identify the level of threat associated with each species, focusing on species classified as “endangered” and “critically endangered” (25). Our analysis of protected areas used the 2010 World Database on Protected Areas (26), a dataset containing the boundaries of more than 17,000 protected areas (such as national parks) in the hotspots and high biodiversity wilderness areas.

The co-occurrence of languages with particular biodiversity conservation priorities reinforces the tendency found at a regional scale for linguistic and biological diversity to share geographic space. Fig. 3 summarizes the geographic intersection of indigenous and nonmigrant languages with endangered amphibians and protected areas within each biodiversity hotspot and high biodiversity wilderness area. The majority of each conservation priority in the hotspots shares at least part of its geographic location with indigenous and nonmigrant languages, with the amount of co-occurrence varying widely among hotspots and conservation priorities. Large percentages of these conservation priorities in the high biodiversity wilderness areas also overlap with indigenous and nonmigrant languages, again showing considerable variability

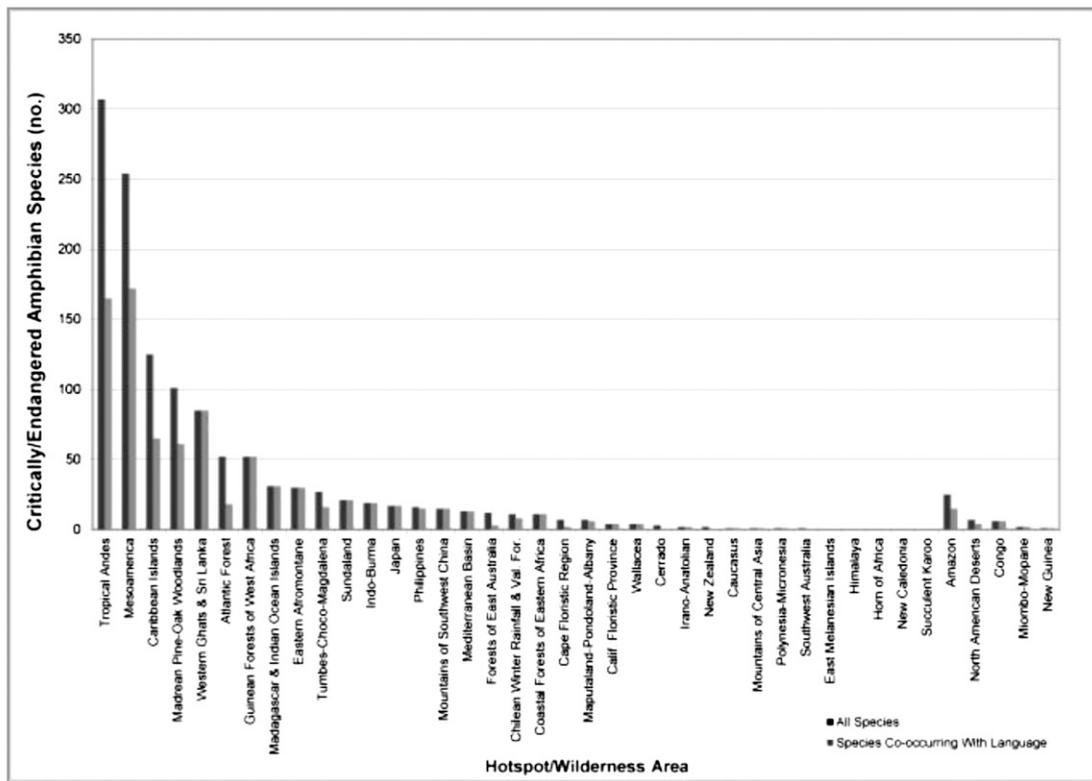
among wilderness areas and priorities. Details of these results, co-occurrences with other conservation priorities, and co-occurrences between conservation priorities and languages spoken by 10,000 or fewer and 1,000 or fewer, appear in the *SI Text*.

Of the more than 6,900 languages currently spoken on Earth, more than 4,800 occur in regions containing high biodiversity. As both hotspots and high biodiversity wilderness areas are defined by biological criteria and amount of natural habitat loss, there is no obvious reason why either would host large numbers of languages. Moreover, the geographic concentration is marked: nearly 70% of the world's languages are spoken in roughly 24% of the earth's terrestrial surface (26% if we exclude Antarctica), where only one-third of the planet's population lives (18). Although the total languages in biodiversity regions encompasses considerable variability for individual areas—the number of indigenous or nonmigrant languages spoken per region ranging from 1 to more than 970—in general, regions containing high biological diversity tend to have high linguistic diversity as well. The total linguistic diversity is greatly influenced by New Guinea, long known as the most linguistically complex and diverse area in the world. Although removing the New Guinea Wilderness Area from the analysis reduces the total languages in regions of high biodiversity to about 56% of the global total, the results still indicate remarkable linguistic diversity on the roughly one-fourth of the earth's surface comprising the remaining regions of high biodiversity. Focusing on other priorities for biodiversity conservation (16), such as certain nontropical biomes and selected rare species, the exclusion of which has led to criticism of the hotspot approach (27), would yield alternative patterns of co-occurrence with linguistic diversity, although for regions with lower concentrations of biological diversity than those used in this study.

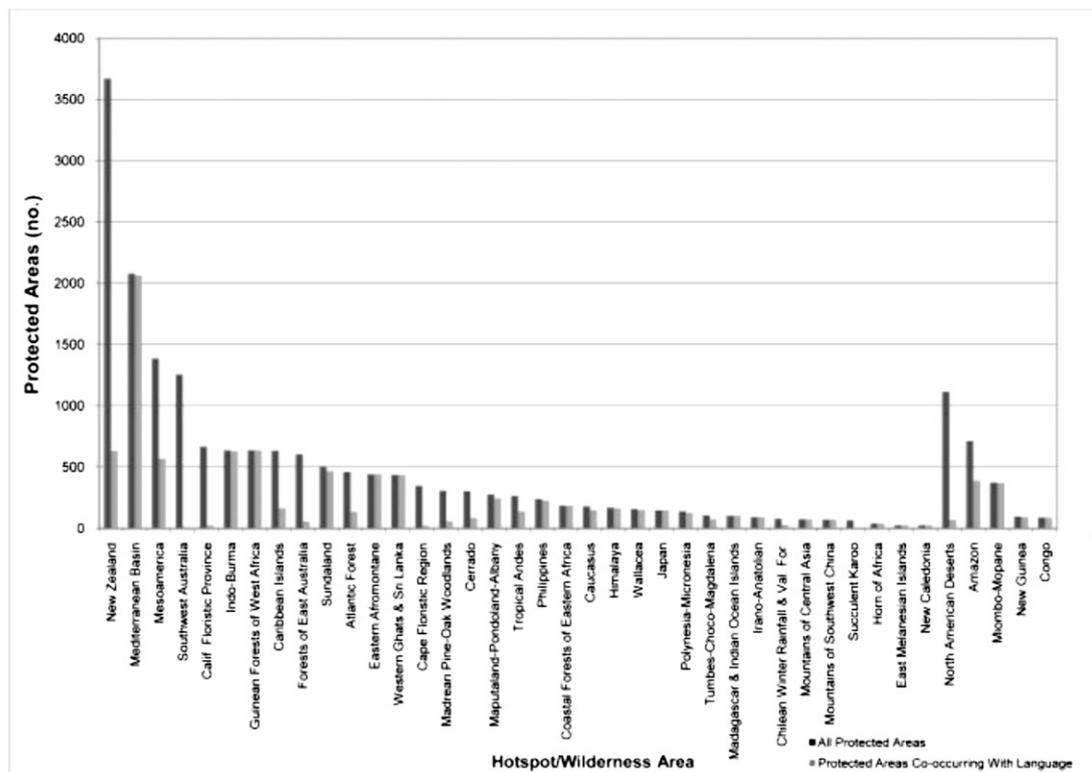
A variety of reasons may account for the co-occurrence of linguistic and biological diversity described in this study. Some prior research has proposed that the ecology of human societies has led to the emergence of high linguistic diversity in high biological diversity areas, in certain instances arguing that competition for larger numbers of resources generates greater linguistic diversity among people adapting to these more complex environments (28), and in others that more plentiful, diverse resources (lower ecological risk) enable greater linguistic diversity by reducing the likelihood of having to communicate and share resources with other groups in times of need (12, 21, 29). Other research has found that high linguistic diversity tends to occur in areas with high biological diversity, but suggests that the processes underlying co-occurrences vary and require separate examinations to identify underlying reasons (22, 30). The considerable variability in linguistic diversity with respect to biological diversity found across the regions examined in this study suggests that underlying reasons are complicated and may well differ from one area to another. This proposition is not surprising, given the complexity of the natural environments where large numbers of languages occur. For example, wetlands contain fewer species than other ecosystems but often provide resources valuable to humans, a pattern borne out in the Caucasus, where the distribution of small, highly localized languages correlates closely with locations near streams (31).

Other reasons, not based on functional connections between ecological and linguistic diversity, also offer possible explanations for similarities (and differences) in the two. At a global scale, the European biological expansion of people, crops, diseases, and languages served to reduce cultural and linguistic diversity in many localities on our planet (4). This expansion emphasized temperate areas more similar to Europe (32), its impacts therefore much less in the tropics, where much of the high biological and linguistic diversity often occurs. At the regional scale, particularities of human expansion have variously affected biological or linguistic diversity. For example, Madagascar, a large island with extremely high endemic biodiversity, hosts a small number of languages, the former because of geographic separation millions of years ago that allowed the evolution of unique species, the latter because of human colonization from a single region roughly 2,000 y ago that provided no such opportunity for linguistic evolution (33). Even amid overall

A



B



**Fig. 3.** Bar charts showing the occurrence of the following by biodiversity hotspot and high biodiversity wilderness area. (A) All languages and endangered and critically endangered amphibians. (B) All languages and protected areas.

high linguistic and biological diversity, New Guinea shows significant regional variation, with the isolated and rugged highlands featuring high biological diversity, but less linguistic diversity than the northeastern coast of that large island. Topographic barriers to

biological dispersion help account for the former; comparatively lower incidence of malaria in the interior allowed the emergence of large polities and the diffusion of associated language groups that did not occur on the more linguistically diverse northeastern coast,

where topography is less rugged but the incidence of malaria higher (34).

Regardless of the functional connection between linguistic and biological diversity, or the role that human movement and associated impacts have (or have not) had on one or both measures of diversity, the tendency for both to be high in particular regions suggests that certain cultural systems and practices, represented by speakers of particular indigenous and nonmigrant languages, tend to be compatible with high biodiversity. Independent inquiries support the view that indigenous economies and management practices essentially enable high biological diversity to persist. For example, analysis of satellite data shows that indigenous lands occupying one-fifth of the Brazilian Amazon (five times the area under protection in parks) currently are the most important barrier to Amazon deforestation, a major cause of biodiversity loss in the area (35). The inhibitory effect of indigenous lands on deforestation is not correlated with indigenous population density. Moreover, biodiversity is equal to if not higher in areas with more indigenous presence than areas with less.

Given the capacity of humans to dominate, and in many cases eradicate, other species on our planet, the importance of the relationship between people and the natural environments they inhabit cannot be overstated for biodiversity conservation. Unfortunately, the opportunity to enlist speakers of particular languages in biodiversity conservation is rapidly disappearing as languages are lost at an alarming rate (4). Although linguists have attempted to identify languages in danger of disappearance (36, 37), no system of language ranking in terms of risk can claim the broad attention and authority enjoyed by the IUCN Red List, the main means of evaluating the condition of species. The presence of so many languages in regions of high biodiversity, often spoken only in those regions by relatively small numbers of speakers, suggests that the future of many of these languages and the biodiversity supported by their speech communities is in

question in the face of expansion by the cultures and languages of a relatively few dominant societies.

Although different processes may have given rise to the diversification of languages, cultures, and species in different areas, similar forces currently appear to be driving biological extinctions and cultural/linguistic homogenization. Broad changes in the form of habitat loss because of large-scale human impacts from an expanding industrialized global economy also represent potential risks to languages and their associated cultures, similar to the impacts of the European expansion mentioned above. Our analysis reveals that for many conservation priorities and languages, efforts to maintain particular biodiversity targets in particular locations could benefit one or more languages in the same place, and vice versa. The co-occurrence of linguistic (and, in many ways, cultural) and biological diversity identified in this study is fortuitous in that it provides the basis for bringing together organizations and researchers focusing on biodiversity conservation and those concerned with linguistic and cultural conservation in particular regions (38). Adopting a shared framework for integrating biological and linguistic conservation goals will facilitate monitoring the status of species and languages at the same time as it may lead to better understanding of how humans interact with ecosystems. Indeed, it may be impossible to achieve large-scale conservation of species and the ecosystems that contain them without incorporating resident languages and the cultures they represent into biodiversity conservation strategies.

**ACKNOWLEDGMENTS.** We thank K. Alger, T. Brooks, D. Erwin, and M. Hoffman for reading and commenting on an earlier version of this paper; the editor for suggestions that improved the presentation of several important points; and M. Denil for help in designing Fig. 1. This research was supported in part by the Gordon and Betty Moore Foundation, through its generous support to Conservation International. S.R. acknowledges the financial support of the Arts and Humanities Research Council, United Kingdom, for research leave in support of this study.

1. Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. *Science* 269:347–350.
2. Mace GM, et al. (2005) Millennium ecosystem assessment. Current state and trends: Findings of the Condition and Trends Working Group. *Ecosystems and Human Well-Being*, eds Hassan R, Scholes R, Ash N (Island Press, Washington, DC), Vol 1, pp 77–122.
3. Harmon D, Loh J (2010) The Index of Linguistic Diversity: A new quantitative measure of trends in the status of the world's languages. *Language Documentation and Conservation* 4:97–151.
4. Nettle D, Romaine S (2000) *Vanishing Voices* (Oxford Univ Press, New York).
5. Nettle D (1996) Language diversity in West Africa: An ecological approach. *J Anthropol Archaeol* 15:403–438.
6. Stepp JR, et al. (2004) Development of a GIS for global biocultural diversity. *Policy Matters* 13:267–270.
7. Stepp JR, Castaneda H, Cervone S (2005) Mountains and biocultural diversity. *Mt Res Dev* 25:223–227.
8. Toledo VM (1994) Biodiversity and cultural diversity in Mexico. *Different Drummer* 1(3):16–19.
9. Harmon D (1995) The status of the world's languages as reported in 'Ethnologue'. *Southwest Journal of Linguistics* 14:1–28.
10. Harmon D (1996) Losing species, losing languages: Connections between linguistic and biological diversity. *Southwest Journal of Linguistics* 15:89–108.
11. Loh J, Harmon D (2005) A global index of biocultural diversity. *Ecol Indic* 5:231–241.
12. Oviedo G, Maffi L, Larsen PB (2000) *Indigenous and Traditional Peoples of the World and Ecoregion Conservation: An Integrated Approach to Conserving the World's Biological and Cultural Diversity* (World Wildlife Foundation International and Terralingua, Gland, Switzerland).
13. Maffi L, ed (2001) *On Biocultural Diversity: Linking Language, Knowledge, and the Environment* (Smithsonian Institution, Washington, DC).
14. Maffi L (2005) Linguistic, cultural, and biological diversity. *Annu Rev Anthropol* 29: 599–617.
15. Global Mapping International (2010) World Language Mapping System (Global Mapping International, Colorado Springs), version 16.0.
16. Brooks TM, et al. (2006) Global biodiversity conservation priorities. *Science* 313:58–61.
17. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.
18. Mittermeier RA, et al., eds (2004) *Hotspots Revisited* (CEMEX, Mexico City).
19. Williams KJ, et al. (2011) *Biodiversity Hotspots: Evolution and Conservation*, eds Zachos F, Habel JC (Springer, Berlin), pp 295–310.
20. Mittermeier RA, et al. (2003) Wilderness and biodiversity conservation. *Proc Natl Acad Sci USA* 100:10309–10313.
21. Sutherland WJ (2003) Parallel extinction risk and global distribution of languages and species. *Nature* 423:276–279.
22. Moore JL, et al. (2002) The distribution of cultural and biological diversity in Africa. *Proc Biol Sci* 269:1645–1653.
23. Mühlhäusler P (1995) *The Politics of Multiculturalism in the Asia/Pacific*, ed Myers D (Northern Territory Univ Press, Darwin, Australia), pp 154–161.
24. Stuart SN, et al. (2004) Status and trends of amphibian declines and extinctions worldwide. *Science* 306:1783–1786.
25. Baillie JEM, Hilton-Taylor C, Stuart SN, eds. (2004) 2004 IUCN Redlist of Threatened Species. A Global Species Assessment (IUCN, Gland, Switzerland and Cambridge, UK).
26. World Database on Protected Areas Consortium (2010) *World Database on Protected Areas Annual Release 2010* (United Nations Environment Programme, Nairobi, Kenya, and World Conservation Monitoring Centre, Cambridge, UK).
27. Kareiva P, Marvier M (2003) Conserving biodiversity coldspots. *Am Sci* 91:344–351.
28. Pagel M, Mace R (2004) The cultural wealth of nations. *Nature* 428:275–278.
29. Nettle D (1999) *Linguistic Diversity* (Oxford Univ Press, New York).
30. Manne LL (2003) Nothing has yet lasted forever: Current and threatened levels of biological diversity. *Evol Ecol Res* 5:517–527.
31. Comrie BL (2008) Linguistic diversity in the Caucasus. *Annu Rev Anthropol* 37: 131–148.
32. Crosby AW (1993) *Ecological Imperialism: The Expansion of Europe 900–1900* (Cambridge Univ Press, New York).
33. Dewar RE (2003) *The Natural History of Madagascar*, eds Goodman SM, Bensted J (Univ of Chicago Press, Chicago), pp 119–122.
34. Romaine S (1992) *Language, Education and Development. Urban and Rural Tok Pisin in Papua New Guinea* (Oxford Univ Press, Oxford).
35. Nepstad D, et al. (2006) Inhibition of Amazon deforestation and fire by parks and indigenous lands. *Conserv Biol* 20:65–73.
36. United Nations Educational, Scientific, and Cultural Organization (2005) *Interactive Atlas of the World's Languages in Danger*. <http://www.unesco.org/culture/languages-atlas/en/atlasmap.html>, accessed June 20, 2011.
37. Wurm SA (1996/2001) *Atlas of the World's Languages in Danger of Disappearing* (United Nations Educational, Scientific, and Cultural Organization, Paris).
38. Maffi L, Woodley E (2010) *Biocultural Diversity Conservation: A Global Sourcebook* (Earthscan, London).