



Geographic Range Size and Determinants of Avian Species Richness

Walter Jetz and Carsten Rahbek

Science **297**, 1548 (2002);

DOI: 10.1126/science.1072779

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 28, 2012):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/297/5586/1548.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2002/08/29/297.5586.1548.DC1.html>

This article **cites 22 articles**, 3 of which can be accessed free:

<http://www.sciencemag.org/content/297/5586/1548.full.html#ref-list-1>

This article has been **cited by** 198 article(s) on the ISI Web of Science

This article has been **cited by** 33 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/297/5586/1548.full.html#related-urls>

This article appears in the following **subject collections**:

Ecology

<http://www.sciencemag.org/cgi/collection/ecology>

5. K. Rohde, *Oikos* **65**, 514 (1992).
6. J. Damuth, *Biol. J. Linn. Soc.* **31**, 193 (1987).
7. B. J. Enquist, J. H. Brown, G. B. West, *Nature* **395**, 163 (1998).
8. J. F. Gillooly, J. H. Brown, G. B. West, V. M. Savage, E. L. Charnov, *Science* **293**, 2248 (2001).
9. Results are based on reduced-major-axis regression analyses of metabolic rate data (8) and molecular evolution data in J. F. Gillooly, A. P. Allen, G. B. West, J. H. Brown, in preparation.
10. Estimates of average population density \bar{N} , and therefore \bar{B}_T , obtained from Damuth's compilation of the published literature (6) are not directly comparable to estimates calculated on the basis of number of individuals, J , and species, S , in a plot of size A ($\bar{N} = J/AS$). Damuth's data refer to the number of individuals present per unit of suitable habitat for a particular species. Plots of fixed size A may include a mixture of suitable and unsuitable habitats for different species that constitute the communities, resulting in lower estimates for \bar{N} .
11. B. J. Enquist, K. J. Niklas, *Nature* **410**, 655 (2001).
12. S. P. Hubbell, *A Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
13. J. H. Brown, M. V. Lomolino, *Biogeography* (Sinauer, Sunderland, MA, 1998).
14. M. E. Ritchie, H. Olff, *Nature* **400**, 557 (1999).
15. M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, UK, 1995).
16. R. K. Colwell, D. C. Lees, *Trends Ecol. Evol.* **15**, 70 (2000).
17. D. H. Wright, *Oikos* **41**, 496 (1983).
18. J. F. Gillooly, E. L. Charnov, G. B. West, V. M. Savage, J. H. Brown, *Nature* **417**, 70 (2002).
19. A. H. Gentry, *Ann. Mo. Bot. Gard.* **75**, 1 (1988).
20. B. J. Enquist, K. J. Niklas, *Science* **295**, 1517 (2002).
21. T. Isobe, E. D. Feigelson, M. G. Akritas, G. J. Babu, *Astrophys. J.* **364**, 104 (1990).
22. Mass-corrected population densities for tree species were calculated with the use of a database compiled by Gentry, which includes counts of all individuals ≥ 2.5 cm in diameter collected from over 200 0.1-ha plots around the world (19). Extensive intraspecific variation in plant size required that we use a different approach to calculate mass-corrected population densities for the trees than for the other taxa. We estimated the dry mass (in grams) of each individual, M_j , from its diameter, D_j (in centimeters), using the formula $M_j = 156D_j^{2.53}$ (11, 20). We then computed the average mass-corrected population density as $NM^{3/4} = (AS)^{-1} \sum M_j^{3/4}$, where S is the number of species in the plot of size A .
23. D. R. Legates, C. J. Wilmott, *Theor. Appl. Climatol.* **41**, 11 (1990).
24. In Figs. 1 and 2, variation in the dependent variable (abundances of individual species) greatly exceeds errors in the independent variable (temperature) because abundance is known to vary by orders of magnitude among species in a community (12). This justifies the use of ordinary least-squares regression (21). In Figs. 3 and 4, we used reduced-major-axis regression because errors in the dependent variables (richness) and independent variable (temperature) are likely to be of comparable magnitude.
25. D. Lieberman, M. Lieberman, R. Peralta, G. S. Hartshorn, *J. Ecol.* **84**, 137 (1996).
26. W. E. Duellman, *Ann. Mo. Bot. Gard.* **75**, 79 (1988).
27. K. Rohde, M. Heap, *Int. J. Parasitol.* **28**, 461 (1998).
28. We thank D. J. Currie, T. Oberdorff, and J. T. Kerr for providing data, and D. J. Currie, E. H. Decker, W. Jetz, A. J. Kerkhoff, B. T. Milne, H. Olff, and three anonymous reviewers for comments. We gratefully acknowledge the support of NSF grant DEB-0083422. A.P.A. further acknowledges the support of NSF grant DEB-9910123, and J.H.B. acknowledges the support of a Packard Interdisciplinary Science Grant.

29 March 2002; accepted 19 July 2002

Geographic Range Size and Determinants of Avian Species Richness

Walter Jetz^{1,2*} and Carsten Rahbek³

Geographic patterns in species richness are mainly based on wide-ranging species because their larger number of distribution records has a disproportionate contribution to the species richness counts. Here we demonstrate how this effect strongly influences our understanding of what determines species richness. Using both conventional and spatial regression models, we show that for sub-Saharan African birds, the apparent role of productivity diminishes with decreasing range size, whereas the significance of topographic heterogeneity increases. The relative importance of geometric constraints from the continental edge is moderate. Our findings highlight the failure of traditional species richness models to account for narrow-ranging species that frequently are also threatened.

Most analyses of determinants of geographic patterns in species richness have traditionally looked only at overall species richness patterns, but this does not give a representative picture for most taxa. Wide-ranging species contribute many more distribution records to a species richness pattern than do narrow-ranging species. Thus, although most species tend to have range sizes well below average (1–3), insights from conventional biogeographical analyses of overall species richness are in fact largely based on wide-ranging species. This may produce a profound bias in the presumed determinants of species richness in space that any general model should address.

Determinants of overall species richness singled out so far include measures of productivity (4, 5), habitat heterogeneity (6, 7), area (8, 9), regional and evolutionary history (10), synergism between climate and evolutionary history (11), and effects from geometric constraints imposed by distribution boundaries such as the continental edge (12–14). Here we address the potential effect of range size on the pattern of species richness and its presumed determinants (15), using a 1°-resolution database summarizing the distribution of the 1599 breeding bird species endemic to sub-Saharan Africa (Fig. 1B) (15).

We first examine potential factors one by one, using traditional general linear model (GLM) and spatial linear model (SLM) regression analyses of overall species richness (see table S1 for all single-predictor results). We focus on the latter approach, which avoids inflation of type I errors and invalid parameter estimates due to spatial autocorrelation (15–17). Our results broadly confirm

the important role of net primary productivity (NPP, including a quadratic term, NPP^2) and habitat heterogeneity (HabHet) on species richness asserted in previous studies (4–7). Both emerge as main predictors of overall species richness (single-predictor SLM: t (HabHet) = 22.03; two-predictor SLM ($NPP + NPP^2$): t (NPP) = 25.08, t (NPP^2) = –15.62; all $P < 0.001$). These three variables together explain around 66% of the variance (SLM: r^2 of fitted values against observed values, log likelihood = –12,440). Annual precipitation (Rain) also has high predictive value (single-predictor SLM: t = 26.14), but because it is strongly collinear with NPP [$r_s = 0.91$, $n = 1738$; two-predictor SLM ($NPP + Rain$): t (Rain) = 8.69, t (NPP) = 13.66], we did not examine it further.

Other core environmental predictors include topographic heterogeneity [TopHet, measured as altitudinal range; single-predictor SLM: t = 13.77 (table S1)] and mean maximum daily temperature (MaxTemp; single-predictor SLM: t = –17.74). In contrast to the suggested direct positive link between temperature as a measure of energy and species richness (18, 19) and some of the empirical findings from South America (11), the relationship with maximum temperature is strongly negative: High-temperature quadrats support fewer species than do quadrats with more moderate temperatures. This may suggest an envelope effect of temperature on homeotherms, with both cold and hot temperatures adversely affecting species richness in addition to and conjunction with other factors.

Distribution constraints imposed by continental boundaries, together with the tendency of ranges to be continuous at larger scales of analysis, can demonstrably influence the geography of species richness, with higher levels of species richness expected in the middle of a bounded domain (12–14, 20). We found that geometric constraints modeled in two dimen-

¹Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. ²Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA. ³Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen O, Denmark.

*To whom correspondence should be addressed. E-mail: walter.jetz@zoo.ox.ac.uk

sions have an independent effect on overall species richness [single-predictor SLM: $t = 11.77$, $P < 0.001$ (table S1)]. This pattern is upheld when other factors are controlled for (Table 1), but its strength compared to productivity and habitat heterogeneity is relatively low (Table 1). Geometric constraints and most other variables are largely independent in effect, experiencing only minor change when considered together in the model. Individual test results are not shown, but, for example, in moving from a single-predictor to a two-predictor model based on geometric constraints and habitat heterogeneity, t values change very little (for geometric constraints, t changes from 11.77 to 11.88, whereas for habitat heterogeneity t changes from 22.03 to 22.10). Our results suggest that, although there exists a signature of hard distribution boundaries on the species richness pattern of wide-ranging species, environmental and evolutionary ecological aspects of species richness exceed these effects of continental geometry for our data set. It casts doubt on the assertion that geometric constraints may provide a simple answer to the latitudinal gradient in species richness (21, 22).

Investigations have so far tended to examine potential factors in isolation, although they are likely to act in concert. Collinearity of variables (such as between habitat and topographic heterogeneity, $r_s = 0.59$, $n = 1738$) may well affect model results and render presumed relationships as artefacts. To address this caveat, we selected six “core” variables following the initial single- and two-predictor results and consensus from the literature and examined them in a multiple-predictor spatial regression model (Table 1). All six core variables, notably habitat heterogeneity and topographic heterogeneity, retain the directions and magnitudes of their effects. The combined spatial autoregressive model explains 67% of the variance, a value that increases to 94% when the estimates for the spatial signal are included. Inclusion of nine further environmental predictors (15) (table S2) yields a statistically better model (log-likelihood test on SLM model: $G = 100$, d.f. = 6, $P < 0.001$), but the change in model fit is very small.

Partitioning species into range size quartiles [RSQs (15)] of ~400 species each highlights the fact that records are spatially distributed very disparately (Fig. 2): The widest-ranging quartile of species accounts for 70.5% of 366,853 records overall, whereas the first, second, and third quartiles represent only 1.3, 7.2, and 21% of records, respectively. The correlation between quadrat species richness of narrow-ranging species with overall species richness is weak and increases with larger range size. The dominance of wide-ranging species in quadrat assemblages contrasts with the strongly right-skewed range size distribution of sub-Saharan African birds (skewness = 1.99, $n = 1599$ species)

also found in other groups and regions (23).

As expected (13, 20), the effect of geometric constraints decreases with decreasing range size and is not detectable for the narrow-ranging half of species (Table 1). More surprisingly, the effect of productivity, a very prominent predictor of overall and wide-ranging species richness, shows a stark decrease with decreasing range size (Fig. 3). Advocates of the species-energy hypothesis support the idea that species numbers in local assemblages generally increase with ambient available energy (9, 18, 19, 24), although the shape of the relationship (hump-shaped or monotonic) appears to be affected by scale (9). We find that relationships between NPP

and richness vary markedly among groups of different range sizes, with a clear hump-shaped pattern only for the most wide-ranging species (fourth RSQ), a monotonic increasing one for moderately wide-ranging species, and a poor relationship for the narrow-ranging half of species (25).

Conversely, topographic heterogeneity, although of only marginal importance for overall species richness and wide-ranging species, increases in importance with decreasing range size and is the most important predictor of narrow-ranging species' richness (Fig. 4). Topographic heterogeneity measured as altitudinal range has often been used as a proxy for habitat heterogeneity (4, 26).

Fig. 1. Species richness pattern of (A) birds of continental Africa ($n = 1901$ species) and (B) birds endemic to Africa ($n = 1599$ species). Both patterns are highly correlated ($r_s = 0.99$, $n = 1738$ 1° quadrats). Equal-interval classification is shown, with color ramps indicating minimum (light, top of legend) to maximum (dark red, bottom of legend) species richness. (C and D) Geographic pattern of model residuals. For (C), the residuals come from the traditional (GLM) model and for (D) from the spatial autoregressive six-predictors model (SLM), as presented in Table 1. Standard deviation classification ranges from dark cyan (< -3 SD) to dark red ($> +3$ SD).

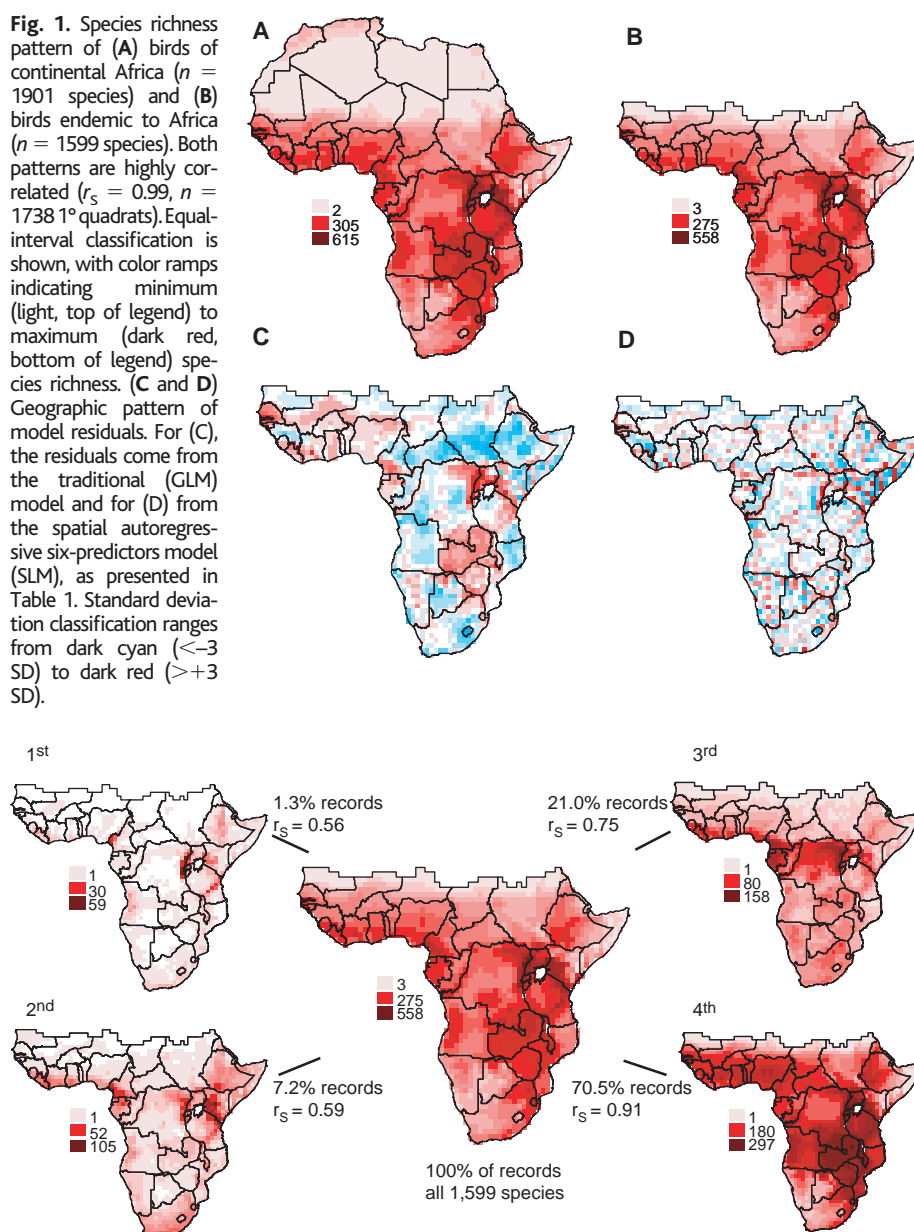


Fig. 2. Geographic pattern of sub-Saharan avian species richness by range size quartiles. Quartiles range from first (narrowest ranging 25%) to fourth (widest ranging 25%) of 1599 species overall. Equal-interval classification is shown, with color ramps indicating minimum (light, top of legend) to maximum (dark red, bottom of legend) species richness.

Fig. 3. Relationship between NPP (in tons of carbon per hectare per year) and avian species richness for species belonging to different range size quartiles, from first to fourth. Each point represents one 1° quadrat.

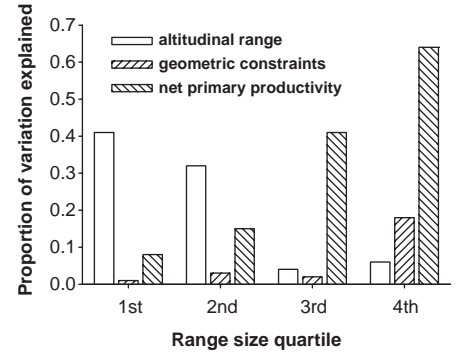
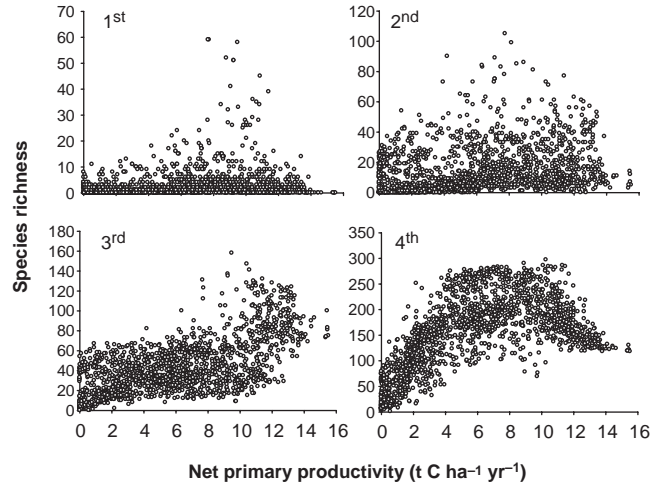


Fig. 4. Proportion of range size quartile species richness from first quartile to fourth quartile, explained by various predictors. r^2 values of fitted values from spatial autoregressive one-predictor regressions are shown.

Table 1. Regression results for selected predictors of all species richness of birds endemic to Africa (All species) and richness of species partitioned into range size quartiles [from first (narrowest-ranging 25% of species) to fourth (widest-ranging 25% of species)]. We performed traditional GLM and SLM analyses. We fitted a six-predictor model of core variables, including predictions from a two-dimensional geometric constraints model (GeoConst), the number of different vegetation classes (HabHet), net primary productivity (NPP), topographic heterogeneity/altitudinal range (TopHet), and mean daily maximum temperature (MaxTemp). We subsequently added nine further environmental variables in order to see whether they added further information left unexplained by the selected core predictor variables (for detailed results, see table S2). The high spatial autocorrelation of errors in the GLM analysis (all Moran's I , $P < 0.001$) confirms the expected violation of the non-independence assumption. Thus, for SLM only, we tested t values for their significance (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). P values were adjusted for error rate per variable $P_{adj} = P/6$.

| Model | Variable | | First quartile | Second quartile | Third quartile | Fourth quartile | All species |
|---------------------|--------------------|-------|----------------|-----------------|----------------|-----------------|-------------|
| <i>GLM</i> | | | | | | | |
| Six-predictor model | GeoConst | t | -4.80 | -5.75 | 8.22 | 9.08 | 6.28 |
| | HabHet | t | 6.97 | 15.78 | 4.46 | 6.76 | 9.61 |
| | NPP | t | 1.25 | -3.44 | -6.29 | 38.60 | 25.27 |
| | NPP ² | t | 0.76 | 6.77 | 14.31 | -35.00 | -18.91 |
| | TopHet | t | 20.18 | 10.94 | 0.82 | 0.70 | 5.21 |
| | MaxTemp | t | -6.97 | -18.10 | -21.57 | -7.13 | -14.75 |
| | Model | F | 290.82 | 450.82 | 434.32 | 701.14 | 758.87 |
| | Model | r^2 | 0.50 | 0.61 | 0.60 | 0.71 | 0.73 |
| 15-predictor model | Errors Moran's I | | 0.55*** | 0.70*** | 0.73*** | 0.79*** | 0.74*** |
| | Model | F | 133.84 | 218.30 | 252.41 | 337.17 | 364.78 |
| 15-predictor model | Model | r^2 | 0.54 | 0.66 | 0.69 | 0.75 | 0.76 |
| <i>SLM</i> | | | | | | | |
| Six-predictor model | GeoConst | t | -1.62 | 0.15 | 2.30 | 6.43*** | 6.42*** |
| | HabHet | t | 4.82*** | 9.66*** | 6.30*** | 10.44*** | 10.77*** |
| | NPP | t | 4.59*** | 5.31*** | 4.22*** | 23.03*** | 17.22*** |
| | NPP ² | t | -3.08* | -1.00 | 2.06 | -16.38*** | -10.12*** |
| | TopHet | t | 13.92*** | 8.45*** | 6.33*** | 3.68*** | 8.88*** |
| | MaxTemp | t | -7.41*** | -11.99*** | -6.76*** | -2.54 | -6.16*** |
| | Log likelihood | | -5,312 | -5,784 | -10,440 | -11,610 | -12,340 |
| | Fitted values | r^2 | 0.46 | 0.52 | 0.53 | 0.62 | 0.67 |
| 15-predictor model | Full model | r^2 | 0.82 | 0.92 | 0.93 | 0.96 | 0.94 |
| | Errors Moran's I | | -0.02 | -0.02 | 0.08*** | 0.03* | 0.01 |
| | Log likelihood | | -5,281 | -5,724 | -10,340 | -11,580 | -12,290 |
| 15-predictor model | Fitted values | r^2 | 0.48 | 0.52 | 0.60 | 0.62 | 0.66 |

However, we find that both habitat heterogeneity and topographic heterogeneity retain high individual contributions in a multiple predictor model. In addition to topographic heterogeneity, altitudinal range measures the degree to which some parts of a quadrat are topographically separate from others. Because of their relatively isolated geographic

position, habitat shifts on African mountains following temporally fluctuating changes in climate (27) may have given rise to many opportunities for speciation (11, 28). In Africa, mountain regions near the equator are known for their unusually high number of species and local endemics (29, 30). Our results support the potential historical signal

of altitudinal range per se (11), highlighting its specific effect on narrow-ranging species.

In the search for the determinants of species richness, it has been tacitly assumed that the pattern and determinants of overall species richness may be representative for the majority of species. However, the usual analyses of overall species richness elicit correlates that may be representative geographically and for species richness per se, but not necessarily taxonomically, for individual species or even for a great proportion of species. It follows that conservation strategies based simply on hotspots of species richness or the distribution and environmental correlates of overall species richness may be of limited use for threatened species, which tend to be restricted in range (31). Likewise, biogeographic models depending on such an approach may have to be rethought. Additional analyses will help to clarify to what extent the observed dissimilarity between widespread and localized species is due to differences in rates of speciation and extinction, dispersal and colonization potential, habitat availability and site suitability, community-wide competition at large spatial scales, or a yet unidentified principle. Progress in meeting this challenge is likely to come from high-quality distributional and ecological databases and from biologically realistic null models.

References and Notes

1. M. D. Pagel, R. M. May, A. R. Collie, *Am. Nat.* **137**, 791 (1991).
2. J. H. Brown, *Macroecology* (Univ. of Chicago Press, Chicago, 1995).
3. K. J. Gaston, *Rarity* (Chapman and Hall, London, 1994).
4. D. J. Currie, *Am. Nat.* **137**, 27 (1991).
5. M. L. Rosenzweig, Z. Abramsky, in *Species Diversity in Ecological Communities*, R. E. Ricklefs, D. Schluter, Eds. (Univ. of Chicago Press, London, 1993), pp. 52–65.
6. L. Hansson, L. Fahrig, G. Merriam, *Mosaic Landscapes and Ecological Processes*, (Chapman and Hall, London, 1995).
7. J. F. Guegan, S. Lek, T. Oberdorff, *Nature* **391**, 382 (1998).
8. E. F. Connor, E. D. McCoy, *Am. Nat.* **113**, 791 (1979).

9. M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
10. R. E. Ricklefs, D. Schluter, in *Species Diversity in Ecological Communities*, R. E. Ricklefs, D. Schluter, Eds. (Chicago Univ. Press, Chicago, 1993), pp. 350–363.
11. C. Rahbek, G. R. Graves, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4534 (2001).
12. R. K. Colwell, D. C. Lees, *Trends Ecol. Evol.* **15**, 70 (2000).
13. D. C. Lees, C. Kremen, L. Andriampianina, *Biol. J. Linn. Soc.* **67**, 529 (1999).
14. W. Jetz, C. Rahbek, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5661 (2001).
15. Materials and methods are available as supporting material on Science Online.
16. N. A. C. Cressie, *Statistics for Spatial Data* (Wiley, New York, 1993).
17. J. J. Lennon, *Ecography* **23**, 101 (2000).
18. J. R. G. Turner, J. J. Lennon, J. A. Lawrenson, *Nature* **335**, 539 (1988).
19. J. J. Lennon, J. J. D. Greenwood, J. R. G. Turner, *J. Anim. Ecol.* **69**, 581 (2000).
20. R. K. Colwell, G. C. Hurtt, *Am. Nat.* **144**, 570 (1994).
21. M. R. Willig, S. K. Lyons, *Oikos* **81**, 93 (1998).
22. P. Koleff, K. J. Gaston, *Ecography* **24**, 341 (2001).
23. S. Anderson, *Am. Mus. Novit.* **2833**, 1 (1985).
24. D. J. Currie, V. Paquin, *Nature* **329**, 326 (1987).
25. A. Balmford et al., *Science* **291**, 2616 (2001).
26. J. T. Kerr, L. Packer, *Nature* **385**, 252 (1997).
27. H. Elenga et al., *J. Biogeogr.* **27**, 621 (2000).
28. G. R. Graves, *Auk* **102**, 556 (1985).
29. D. Pomeroy, *Conserv. Biol.* **7**, 901 (1993).
30. T. M. Crowe, A. A. Crowe, *J. Zool.* **198**, 417 (1982).
31. A. J. Stattersfield, D. R. Capper, G. C. L. Dutton, *Threatened Birds of the World* (BirdLife International, International Union for Conservation of Nature and Natural Resources; Lynx Edicions, Barcelona, 2000).
32. We thank D. Rogers, B. Bakker, W. Wint, and S. Hay for Geographic Information Systems advice and discussions; I. Woodward for providing productivity data; L. A. Hansen for assistance with access to the African bird database on behalf of the Zoological Museum University of Copenhagen (ZMUC); and E. Baker, N. Baker, F.

Dowsett-Lemaire, R. Dowsett, J. Fjeldsø, M. E. Gartshore, H. M. de Klerk, M. Languy, R. B. Payn, COC/BirdLife Cameroon, and BirdLife International for providing data for the ZMUC database. The manuscript benefited tremendously from discussions with R. K. Colwell and comments from J. Brown, J. Fjeldsø, R. Freckleton, K. Gaston, N. J. Gotelli, G. R. Graves, P. Harvey, R. Lande, O. Lewis, R. May, I. Owens, and D. Rogers. W.J. was supported by Natural Environment Research Council and German Scholarship Foundation studentships; C.R. was supported by the Danish National Science Foundation (grant J. nr. 21-01-0547).

Supporting Online Material

www.sciencemag.org/cgi/content/full/297/5586/1548/DC1
Materials and Methods
Supporting Text
Tables S1 to S3

11 April 2002; accepted 25 July 2002

Hierarchical Organization of Modularity in Metabolic Networks

E. Ravasz,¹ A. L. Somera,² D. A. Mongru,² Z. N. Oltvai,^{2*}
A.-L. Barabási^{1*}

Spatially or chemically isolated functional modules composed of several cellular components and carrying discrete functions are considered fundamental building blocks of cellular organization, but their presence in highly integrated biochemical networks lacks quantitative support. Here, we show that the metabolic networks of 43 distinct organisms are organized into many small, highly connected topologic modules that combine in a hierarchical manner into larger, less cohesive units, with their number and degree of clustering following a power law. Within *Escherichia coli*, the uncovered hierarchical modularity closely overlaps with known metabolic functions. The identified network architecture may be generic to system-level cellular organization.

The identification and characterization of system-level features of biological organization is a key issue of postgenomic biology (1–3). The concept of modularity assumes that cellular functionality can be seamlessly partitioned into a collection of modules. Each module is a discrete entity of several elementary components and performs an identifiable task, separable from the functions of other modules (1, 4–8). Spatially and chemically isolated molecular machines or protein complexes (such as ribosomes and flagella) are prominent examples of such functional units, but more extended modules, such as those achieving their isolation through the initial binding of a signaling molecule (9), are also apparent.

Simultaneously, it is recognized that the

thousands of components of a living cell are dynamically interconnected, so that the cell's functional properties are ultimately encoded into a complex intracellular web of molecular interactions (2–6, 8). This is perhaps most evident with cellular metabolism, a fully connected biochemical network in which hundreds of metabolic substrates are densely integrated through biochemical reactions. Within this network, however, modular organization (i.e., clear boundaries between subnetworks) is not immediately apparent. Indeed, recent studies have demonstrated that the probability that a substrate can react with k other substrates [the degree distribution $P(k)$ of a metabolic network] decays as a power law $P(k) \sim k^{-\gamma}$ with $\gamma \cong 2.2$ in all organisms (10, 11), suggesting that metabolic networks have a scale-free topology (12). A distinguishing feature of such scale-free networks is the existence of a few highly connected nodes (e.g., pyruvate or coenzyme A), which participate in a very large number of metabolic reactions. With a large number of links, these hubs integrate all substrates into a single, integrated web in

which the existence of fully separated modules is prohibited by definition (Fig. 1A).

Yet, the dilemma of a modular versus a highly integrated module-free metabolic network organization remains. A number of approaches for analyzing the functional capabilities of metabolic networks indicate the existence of separable functional elements (13, 14). Also, from a purely topologic perspective, the metabolic network of *Escherichia coli* is known to possess a high clustering coefficient (11), a property that is suggestive of a modular organization. In itself, this implies that the metabolism of *E. coli* has a modular topology, potentially comprising several densely interconnected functional modules of varying sizes that are connected by few intermodule links (Fig. 1B). However, such clear-cut modularity imposes severe restrictions on the degree distribution, implying that most nodes have approximately the same number of links, which contrasts with the metabolic network's scale-free nature (10, 11).

To determine whether such a dichotomy is indeed a generic property of all metabolic networks, we first calculated the average clustering coefficient for 43 different organisms (10, 15, 16) as a function of the number of distinct substrates N present in their metabolism. The clustering coefficient, defined as $C_i = 2n_i(k_i - 1)$, where n_i denotes the number of direct links connecting the k_i nearest neighbors of node i (17), is equal to 1 for a node at the center of a fully interlinked cluster, and it is 0 for a metabolite that is part of a loosely connected group (Fig. 2A). Therefore, C_i averaged over all nodes i of a metabolic network is a measure of the network's potential modularity. We found that, for all 43 organisms, the average clustering coefficient is about an order of magnitude larger than that expected for a scale-free network of similar size (Fig. 2B), suggesting that metabolic networks in all organisms are characterized by a high intrinsic potential modularity. We also observed that, in contrast with

¹Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA. ²Department of Pathology, Northwestern University, Chicago, IL 60611, USA.

*To whom correspondence should be addressed. E-mail: zno008@northwestern.edu (Z.N.O.) and alb@nd.edu (A.-L.B.)