The Evolution and Distribution of Species Body Size
Aaron Clauset, *et al.*
Science **321**, 399 (2008);
DOI: 10.1126/science.1157534

Recent Research Studies" (Council on Library and Information Resources, Washington, DC, 2003).
7. A. Friedlander, "Dimensions and Use of the Scholarly Information Environment: Introduction to a Data Set Assembled by the Digital Library Federation and Outsell, Inc." (Council on Library and Information Resources, Washington, DC, 2002); www.clir.org/pubs/reports/pub110/contents.html.
8. P. Boyce, D. W. King, C. Montgomery, C. Tenopir, *Ser. Libr.* **46**, 121 (2004).
9. C. Tenopir, D. W. King, A. Bush, *J. Med. Libr. Assoc.* **92**, 233 (2004).
10. C. Shirky, "Ontology is Overrated: Categories, Links and Tags" (Clay Shirky's Writings About the Internet: Economics & Culture, Media & Community, Open Source, 2005); www.shirky.com/writings/ontology_overrated.html.

11. C. Manning, H. Schütz, *Foundations of Natural Language Processing* (MIT Press, Cambridge, MA, 1999).
12. J. Hausman, B. H. Hall, Z. Griliches, *Econometrica* **52**, 909 (1984).
13. R. P. Dellavalle *et al.*, *Science* **302**, 787 (2003).
14. A. L. Barabási, R. Albert, *Science* **286**, 509 (1999).
15. R. K. Merton, *Science* **159**, 56 (1968).
16. D. J. de Solla Price, *Science* **149**, 510 (1965).
17. H. A. Simon, *Biometrika* **42**, 425 (1955).
18. M. J. Salganik, P. S. Dodds, D. J. Watts, *Science* **311**, 854 (2006).
19. J. Berger, "Exploring ways to shorten the ascent to a Ph.D.," *New York Times*, 3 October 2007; www.nytimes.com/2007/10/03/education/03education.html.
20. I. Newton, *Principia* (Macmillan, New York, ed. 4, 1883) (first published in 1687).

21. C. Darwin, *The Origin of Species* (D. Appleton, New York, 1867) (first published in 1859).
22. I gratefully acknowledge research support from NSF grant 0242971, Science Citation Index data from Thompson Scientific, Inc., and *Fulltext Sources Online* data from Information Today, Inc. I also thank J. Reimer for helpful discussion and insight.

The Evolution and Distribution of Species Body Size

Aaron Clauset[1]* and Douglas H. Erwin[1,2]

The distribution of species body size within taxonomic groups exhibits a heavy right tail extending over many orders of magnitude, where most species are much larger than the smallest species. We provide a simple model of cladogenetic diffusion over evolutionary time that omits explicit mechanisms for interspecific competition and other microevolutionary processes, yet fully explains the shape of this distribution. We estimate the model's parameters from fossil data and find that it robustly reproduces the distribution of 4002 mammal species from the late Quaternary. The observed fit suggests that the asymmetric distribution arises from a fundamental trade-off between the short-term selective advantages (Cope's rule) and long-term selective risks of increased species body size in the presence of a taxon-specific lower limit on body size.

Most taxonomic groups show a common distribution of species body size (*1–3*), with a single prominent mode relatively near but not at the smallest species size (*4*) and a smooth but heavy right tail (often described as a right skew on a log-size scale) extending for several orders of magnitude (e.g., Fig. 1). This distribution is naturally related to a wide variety of other species characteristics with which body size correlates, including habitat, life history, life span (*5*), metabolism (*6*), and extinction risk (*7*). A greater understanding of the underlying constraints on, and long-term trends in, body size evolution may provide information for conservation efforts (*8*) and insight about interactions between ecological and macroevolutionary processes (*9*).

Studies of body-size distributions have suggested that the prominent mode may be indicative of a taxon-specific energetically optimal body size (*10, 11*), which is supported by microevolutionary studies of insular species (*12*). However, evidence for Cope's rule (*1, 13, 14*)—the observation that species tend to be larger than their ancestors—and the fact that most species are not close to their group's predicted optimal size [among other reasons (*15*)] suggest that this

theory may be flawed. Alternatively, species body sizes may diffuse over evolutionary time. If so, Cope's rule alone could cause size distributions to exhibit heavy right tails (*1*), although size-dependent speciation or extinction rates (*2, 9, 16*) or size-neutral diffusion near a taxon-specific lower limit on body size (*17*) could also produce a similar shape. Furthermore, different mechanisms may drive body-size evolution on spatial and temporal scales (*3*), and the importance of interspecific competition to the macroevolutionary dynamics of species body size is not known.

We developed a generalized diffusion model of species body-size evolution, in which the size distribution is the product of three macroevolutionary processes (Fig. 1). We combine these processes, each of which has been independently studied (*1, 2, 17, 18*), in a single quantitative framework, estimate its parameters from fossil data on extinct terrestrial mammals from before the late Quaternary (*19, 20*), and determine whether this model, or simpler variants, can reproduce the sizes of the 4002 known extant and extinct terrestrial mammal species from the late Quaternary (Recent species) (*21, 22*).

This model makes three assumptions: (i) Species size varies over evolutionary time as a cladogenetic multiplicative diffusion process (*1, 17*); the size of a descendant species x_D is the product of a stochastic growth factor λ and its ancestor's size x_A, that is, $x_D = \lambda x_A$. For each speciation event, a new λ is drawn from the dis-

tribution $F(\lambda)$, which models the total influence on species size changes from all directions. A bias toward larger sizes (Cope's rule) appears as a positive average log-change to size $\langle \log \lambda \rangle > 0$, and may depend on the ancestor's size. (ii) Species body size is restricted by a taxon-specific lower limit x_{min} (*6, 23*), which we model by requiring that $F(\lambda < x_{min}/x_A) = 0$, that is, the largest possible decrease in size for a particular speciation event is $\lambda = x_{min}/x_A$. In our computer simulations, time proceeds in discrete steps. At each step, exactly one new species is produced, which is the descendant of a randomly selected species. (iii) Every species independently becomes extinct with probability $p_e(x)$, which increases monotonically with size. A schematic of the model is shown in Fig. 2A. [For technical details, see (*24*).]

To make this model appropriately realistic, we estimated the form of each process from fossil data. The lower limit on mammalian body size is near 2 g, close to the size of both the Etruscan shrew (*Suncus etruscus*) and the bumblebee bat (*Craseonycteris thonglongyai*). Fossil evidence suggests that this limit has existed since at least the Cretaceous-Tertiary boundary (*19, 20, 25*). Further, a limit in this vicinity is supported by both
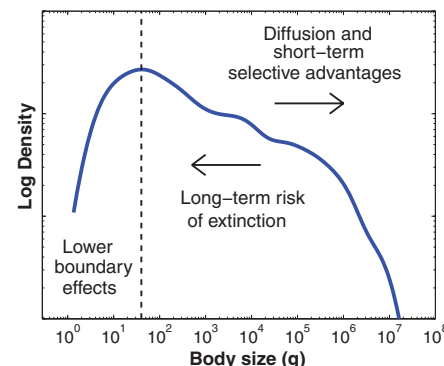
Fig. 1. Smoothed species body-size distribution of 4002 Recent terrestrial mammals [data from (*21*)], showing the three macroevolutionary processes that shape the relative abundances of different sizes. The left tail of the distribution is created by diffusion in the vicinity of a taxon-specific lower limit near 2 g, whereas the long right tail is produced by the interaction of diffusion over evolutionary time (including trends like Cope's rule) and the long-term risk of extinction from increased body size.

[1]Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA. [2]Department of Paleobiology, MRC-121, National Museum of Natural History, Post Office Box 37012, Washington, DC 20013–012, USA.

*To whom correspondence should be addressed. E-mail: aaronc@santafe.edu

experimental (*23*) and theoretical work (*6*) on mammalian metabolism.

Away from this limit, mammalian body-size evolution is governed mainly by diffusion with a bias (Cope's rule) (*14, 26*), whereas its evolution near the lower limit is likely constrained by the need for relatively specialized morphological structures (*1*). We expect this latter effect to appear in fossil data as a systematic intensification of Cope's rule for very small-bodied species, that is, increased $\langle \log \lambda \rangle$ as $x_A \to x_{min}$. From ancestor-descendant size data for 1106 extinct North American terrestrial mammals (*20*), we estimated and compared three models of the distribution $F(\lambda)$ as a function of ancestor size, including the model suggested by Alroy (*14*), which predicts a moderately bimodal distribution in body sizes. Of these, a piecewise model (Fig. 2B), with no effective optimal body size, has the best empirical support [model selection by likelihood ratio test and Bayesian information criterion (*24*)]. This model includes both a strengthening of Cope's rule for small-bodied species ($x \lesssim 32$ g) and a small but uniformly positive bias for larger species, resulting in an average body-size growth of $4.1 \pm 1.0\%$ between ancestors and their descendants ($\langle \log \lambda \rangle = 0.04 \pm 0.01$).

This result supports the existence of short-term selective advantages for increased species body size—for example, better tolerance of resource fluctuations, better thermoregulation, and better predator avoidance (*5*)—but also implies a more nuanced view: Small-bodied species exhibit even greater selective advantages from increased size, for example, because of greater morphological flexibility.

Empirical estimates of extinction rates (or, equivalently, speciation rates) as functions of body size are uncertain (*27*) because of the bias and incompleteness of the fossil record. We partly control for this uncertainty by using a simplistic model of extinction risk $p_e(x)$, largely estimated from the data, where extinction occurs independently, with a probability calculated only from the species' size. We specified a basal extinction rate β by assuming that the number of Recent terrestrial mammal species is close to a putative carrying capacity. We then let extinction risk per unit time increase logarithmically with body size (*24, 28*). This model leaves only the rate ρ by which risk increases with size as a free parameter, which was chosen by minimizing the statistical distance between the simulated and empirical distributions (*24*).

Inserting these three processes, as estimated above, into our computer model, we found that the model accurately predicted the distribution of Recent terrestrial mammal sizes over its seven orders of magnitude (Fig. 3A) and was particularly accurate for small-bodied species ($x < 80$ g). Our sensitivity analysis further indicated that this prediction was highly robust to variations in most of the estimated parameters, but highly sensitive to the location of the lower limit on body size. The estimated value of $x_{min} \approx 2$ g, however, is the most strongly supported of all model parameters. Thus, even large revisions to the other parameter estimates are unlikely to change our general conclusions (*24*). Also, although a range of ρ values produced size distributions that were statistically close to the empirical distribution, the model predicts a particular extinction risk curve (fig. S4B) that could be tested with appropriate empirical data.

To further discriminate among alternative explanations for the species size distribution, we tested simpler diffusion models, each with parameters estimated from fossil data (*24*), including (i) unbiased diffusion with a lower boundary, (ii) Cope's rule with size-dependent extinction, (iii) Cope's rule alone, (iv) size-dependent extinction alone, and (v) a version of the full model that omits the increased bias for small-bodied species ($x \lesssim 32$ g). We found that these models all predicted size distributions that differed, sometimes dramatically so, from the empirical distribution (Fig. 3, B and C, and figs. S9 and S10). Additionally, we found that a positive bias $\langle \log \lambda \rangle > 0$ for large-bodied species is not necessary if the extinction risk increases less quickly (*24*). These results support the inclusion of a fundamental lower limit, the diffusion of species size, and an increasing risk of extinction with size, as well as an increased bias toward larger sizes for small-bodied species ($x \lesssim 32$ g).

Thus, the shape of a body-size distribution can be interpreted in the context of these three macroevolutionary processes. An intermediate lo-
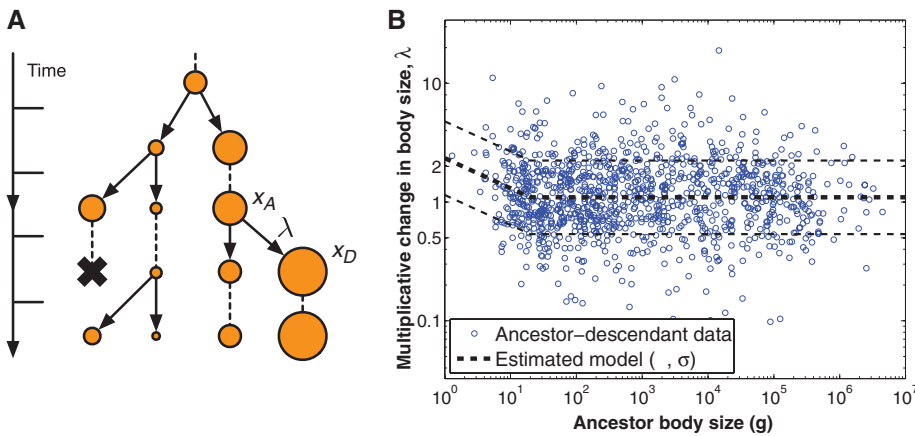
Fig. 2. (**A**) A schematic illustrating a simple cladogenetic diffusion model of species body-size evolution, where the size of a descendant species x_D is related to its ancestor's size x_A by a multiplicative factor λ. (**B**) Empirical data on 1106 changes in North American mammalian body size [data from (*20*)], as a function of ancestor size, overlaid with the estimated model of within-lineage changes, where the average log-change $\langle \log \lambda \rangle$ varies piecewise as a function of body size (*24*).
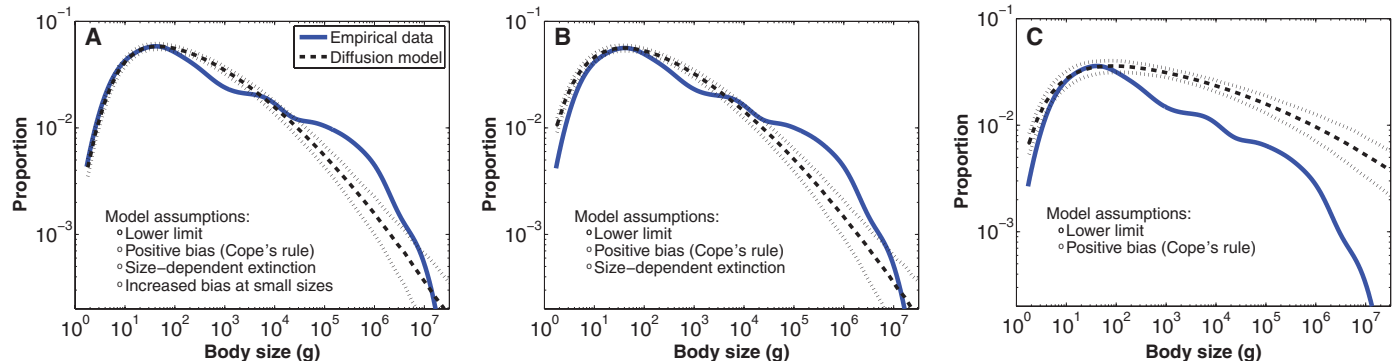
Fig. 3. Simulated distributions of species body size (central tendency ± 95% confidence intervals from 1000 repetitions; all model parameters estimated as described in the text) and the empirical distribution of Recent terrestrial mammals. (**A**) The model described in the text. (**B**) The same model as in (A) but with a bias $\langle \log \lambda \rangle$ that is independent of size. (**C**) The same model as in (B) but with an extinction risk that is independent of size. [For details and additional results, see (*24*).]

cation for the distribution's mode (40 g for terrestrial mammals) is mainly caused by diffusion in the vicinity of the physiological lower limit on body size, which prevents the smallest species from being the most abundant. A heavy right tail is then caused primarily by diffusion in the presence of extinction risks that increase weakly with size ($\rho > 0$). For mammals, the within-lineage tendency toward increased size (Cope's rule, $\langle \log \lambda \rangle > 0$) shifts the mode toward slightly larger sizes and slightly increases the heaviness of the right tail.

Under different conditions, these processes produce markedly different body size distributions. For instance, a long left tail extending toward small-bodied species would indicate that the risk of extinction decreases with larger size ($\rho < 0$). Similarly, a more symmetric distribution would indicate both that extinction rates are relatively size-independent ($\rho \approx 0$) and that changes to body size convey few selective advantages ($\langle \log \lambda \rangle \approx 0$). Although a suitable body-size distribution is not currently available for dinosaurs [but see (29)], evidence suggests that it may be more symmetric than for mammals. The right-skewed distribution's ubiquity, such as for insects and birds (1, 2), suggests that such circumstances are rare and that the mammalian distribution represents the norm.

This model omits explicit mechanisms for many canonical ecological and microevolutionary processes, including the impact of interspecific competition, geography, predation, population dynamics, and size variation between speciation events (anagenetic evolution), which suggests that their contributions to the systematic or large-scale character of species body-size distributions can be compactly summarized by the values of certain model parameters, for example, the strength of Cope's rule $\langle \log \lambda \rangle$ or the manner in which extinction risk increases with body size ρ. Some aspects of the body-size distribution, however, are not explained by this model, such as the slight overabundance of terrestrial mammal species around 300 kg and the slight underabundance around 1 kg (Fig. 3A). Whether such deviations can be attributed to phylogenetically correlated speciation and extinction events is an open question. A more thorough examination of these macroevolutionary processes may explain their particular form and origin, as well as why body size is weakly correlated with increased extinction rates (or decreased speciation rates), why physiological lower limits on body size exist and are conserved within taxonomic groups, and why some groups exhibit macroevolutionary trends but others do not.

References and Notes
1. S. M. Stanley, *Evol. Int. J. Org. Evol.* **27**, 1 (1973).
2. J. Kozlowski, A. T. Gawelczyk, *Funct. Ecol.* **16**, 419 (2002).
3. C. R. Allen *et al.*, *Ecol. Lett.* **9**, 630 (2006).
4. K. P. Dial, J. M. Marzluff, *Ecology* **69**, 1620 (1988).
5. J. H. Brown, *Macroecology* (Univ. of Chicago Press, Chicago, 1995).
6. G. B. West, W. H. Woodruff, J. H. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2473 (2002).
7. M. Cardillo *et al.*, *Science* **309**, 1239 (2005).
8. D. O. Fisher, I. P. F. Owens, *Trends Ecol. Evol.* **19**, 391 (2004).
9. S. M. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 646 (1975).
10. K. P. Sebens, *Annu. Rev. Ecol. Syst.* **18**, 371 (1987).
11. J. H. Brown, P. A. Marquet, M. L. Taper, *Am. Nat.* **147**, 1092 (1996).
12. M. Lomolino, *Am. Nat.* **125**, 310 (1985).
13. C. Deperet, *The Transformations of the Animal World* (D. Appleton, New York, 1909).
14. J. Alroy, *Science* **280**, 731 (1998).
15. J. Kozlowski, *Funct. Ecol.* **16**, 540 (2002).
16. L. Van Valen, *Evol. Int. J. Org. Evol.* **29**, 87 (1973).
17. D. W. McShea, *Evol. Int. J. Org. Evol.* **48**, 1747 (1994).
18. M. L. McKinney, in *Evolutionary Trends*, K. J. McNamara, ed. (Univ. of Arizona Press, Tucson, AZ, 1990), pp. 75–118.
19. M. Fortelius, Neogene of the Old World Database of Fossil Mammals (NOW) (2003). University of Helsinki, NOW public release 030717, www.helsinki.fi/science/now.
20. J. Alroy, North American Fossil Mammal Systematics Database (2008). Paleobiology Database Online Systematics Archive 3, http://paleodb.org.
21. F. A. Smith *et al.*, *Ecology* **84**, 3403 (2003).
22. F. A. Smith *et al.*, Macroecological Database of Mammalian Body Mass, MOM v 3.6.1 (2007).
23. O. P. Pearson, *Science* **108**, 44 (1948).
24. Materials and methods are available as supporting material on *Science* Online.
25. F. A. Smith *et al.*, *Am. Nat.* **163**, 672 (2004).
26. B. Van Valkenburgh, X. Wang, J. Damuth, *Science* **306**, 101 (2004).
27. D. Ludwig, *Ecol. Appl.* **6**, 1067 (1996).
28. L. H. Liow *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6097 (2008).
29. M. T. Carrano, *Amniote Paleobiology*, M. T. Carrano, T. J. Gaudin, R. W. Blob, J. R. Wible, Eds. (Univ. of Chicago Press, Chicago, 2006), pp. 225–268.
30. A.C. is grateful to A. Boyer, J. Dunne, J. Ladau, B. Olding, C. Shalizi, and J. Wilkins for helpful conversations. We thank J. Alroy, A. Boyer, and F. Smith for sharing data. Supported in part by the Santa Fe Institute and the Computer Science Department at the University of New Mexico.

Four-jointed Is a Golgi Kinase That Phosphorylates a Subset of Cadherin Domains

Hiroyuki O. Ishikawa,[1] Hideyuki Takeuchi,[2] Robert S. Haltiwanger,[2] Kenneth D. Irvine[1]*

The atypical cadherin Fat acts as a receptor for a signaling pathway that regulates growth, gene expression, and planar cell polarity. Genetic studies in *Drosophila* identified the *four-jointed* gene as a regulator of Fat signaling. We show that *four-jointed* encodes a protein kinase that phosphorylates serine or threonine residues within extracellular cadherin domains of Fat and its transmembrane ligand, Dachsous. Four-jointed functions in the Golgi and is the first molecularly defined kinase that phosphorylates protein domains destined to be extracellular. An acidic sequence motif (Asp-Asn-Glu) within Four-jointed was essential for its kinase activity in vitro and for its biological activity in vivo. Our results indicate that Four-jointed regulates Fat signaling by phosphorylating cadherin domains of Fat and Dachsous as they transit through the Golgi.

The Fat and Hippo signaling pathways intersect at multiple points and influence growth and gene expression through regulation of the transcriptional coactivator Yorkie (1–8). Fat signaling also influences planar cell polarity (PCP) (9). Fat acts as a transmembrane receptor, and is a large (5147 amino acids) atypical cadherin protein, with 34 extracellular cadherin domains (Fig. 1A) (10). Dachsous (Ds) is also a large (3503 amino acids) transmembrane protein with multiple cadherin domains (Fig. 1A) (11) and is a candidate Fat ligand because it appears to bind Fat in a cultured cell assay (12), acts non–cell autonomously to influence Fat pathway gene expression (2, 13,), and acts genetically upstream of *fat* in the regulation of PCP (14). A second protein, Four-jointed (Fj), also acts non–cell autonomously to influence Fat pathway gene expression and acts genetically upstream of *fat* in the regulation of PCP (2, 13–15). However, Fj is a type II transmembrane protein that functions in the Golgi (16, 17). Thus, Fj might influence Fat signaling by posttranslationally modifying a component of the Fat pathway.

To investigate the possibility of modification of Fat or Ds, we coexpressed FLAG epitope–tagged fragments of their extracellular domains together with Fj in cultured *Drosophila* S2 cells.

[1]Howard Hughes Medical Institute, Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854, USA. [2]Department of Biochemistry and Cell Biology, Institute for Cell and Developmental Biology, Stony Brook University, Stony Brook, NY 11794, USA.

*To whom correspondence should be addressed. E-mail: irvine@waksman.rutgers.edu