LETTERS

Curvature in metabolic scaling

Tom Kolokotrones[1], Van Savage[2], Eric J. Deeds[1] & Walter Fontana[1]

For more than three-quarters of a century it has been assumed[1] that basal metabolic rate increases as body mass raised to some power *p*. However, there is no broad consensus regarding the value of *p*: whereas many studies have asserted that *p* is 3/4 (refs 1–4; 'Kleiber's law'), some have argued that it is 2/3 (refs 5–7), and others have found that it varies depending on factors like environment and taxonomy[6,8–16]. Here we show that the relationship between mass and metabolic rate has convex curvature on a logarithmic scale, and is therefore not a pure power law, even after accounting for body temperature. This finding has several consequences. First, it provides an explanation for the puzzling variability in estimates of *p*, settling a long-standing debate. Second, it constitutes a stringent test for theories of metabolic scaling. A widely debated model[17] based on vascular system architecture fails this test, and we suggest modifications that could bring it into compliance with the observed curvature. Third, it raises the intriguing question of whether the scaling relation limits body size.

In 1932, Max Kleiber found that basal metabolic rate (B)—the power produced by a fasting, inactive organism—scales with body mass (M) across animal species[1]. Based on 13 data points, Kleiber concluded that this relationship was well described by a 3/4-power law:

$$B = B_0 M^{3/4} \quad (1)$$

This apparently simple relationship underlies and constrains an extensive web of scaling relationships, ranging from growth rates to lifespans to trophic dynamics[18–20].

Since 'Kleiber's law' was first proposed, significant amounts of data have been collected and analysed[4,7,8,13,15], fuelling debate about the value of the exponent[19,21–23], a quantity that is crucial for understanding the physical origins of metabolic scaling. An exponent of 2/3 has often been suggested[5–7,15,24] based on a simple surface-to-volume argument. In contrast, a 3/4 exponent emerges from a theory proposed by West, Brown and Enquist based on the properties of optimized resource distribution networks, such as the cardiovascular system[17]. Additionally, some investigators have noted deficiencies in the overall fit of the power law and suggested that the exponent itself might vary with factors such as taxonomic group or environment[6,8–16].

We show that the widely held assumption of a scale-free power law is incorrect. In our analysis, we utilize McNab's recently compiled data set[8] of measurements made reliably under basal conditions (inactive, thermoneutral, post-absorptive adults). It contains measurements of mean metabolic rate from 637 species of mammals spanning 6 orders of magnitude, making it one of the largest such collections yet assembled. To estimate the effect of body temperature on metabolic rate, we extracted temperature measurements from the original papers used in McNab's compilation. The resulting data set of 447 species spans 5 orders of magnitude in mass (Supplementary Information) and was used for those fits that take into account temperature effects. We excluded the orca because its large size has the potential to disproportionately influence the fit, though we found that this is not the case (Supplementary Information). We repeated our analysis using data from Savage[4] and Sieg[16]. Both data sets give essentially the same results as the analysis presented below (Supplementary Information). In all regressions, we use units of grams for mass, watts for basal metabolic rate, and kelvin for temperature.

On a logarithmic scale, a power law, like equation (1), but with an arbitrary scaling exponent β_1, becomes:

$$\log_{10}B = \beta_0 + \beta_1 \log_{10}M + \varepsilon \quad (2)$$

where β_0 is the logarithm of B_0 in equation (1), and ε is the error term. A fit to equation (2) accounts for a significant amount of the trend, but poorly describes the data for both small and large mammals (Fig. 1a, Supplementary Information). This suggests considering a nonlinear model (on the logarithmic scale). As every analytic function can be expanded as a power series, the natural next candidate is a quadratic model:

$$\log_{10}B = \beta_0 + \beta_1 \log_{10}M + \beta_2 (\log_{10}M)^2 + \varepsilon \quad (3)$$

This model results in a visibly better fit for mammals with $M > 50$ g (Fig. 1a, Supplementary Information), which is confirmed by the extremely small P value for β_2 of 9.0×10^{-10} (Table 1). Although the quadratic term explains only an additional 0.3% (96.1% versus 95.8%) of the total variation (7% of the unexplained variation), its impact is clearly seen in both residual and partial residual plots (Supplementary Information). The quadratic term is also necessary to correctly predict the metabolic rate of megafauna such as the orca and elephant (Fig. 1a). Importantly, the addition of higher-order terms beyond the quadratic does not significantly improve the fit (Supplementary Information), suggesting that the scaling relationship for the mammals in this data set is well approximated by a quadratic function of $\log_{10}M$.

Despite the improved fit, there is still considerable residual variation in the data (Supplementary Information). Several studies have demonstrated that temperature affects metabolic rate[7,14,25,26]. We attempt to capture this effect by including a Boltzmann–Arrhenius factor, that is, $B = f(M)\exp(-E/RT)$, where R is the gas constant and T is body temperature in kelvin. When f is a pure power law, equation (2), this new model fits significantly better, but still poorly describes the data for small and large mammals (Supplementary Information). However, when f is given by equation (3), the resulting temperature-corrected quadratic model:

$$\log_{10}B = \beta_0 + \beta_1 \log_{10}M + \beta_2 (\log_{10}M)^2 + \frac{\beta_T}{T} + \varepsilon \quad (4)$$

shows dramatically improved fit over the entire range of the data (Supplementary Information). A plot of the residuals (Supplementary Information) shows that the fit for mammals of intermediate size (between 25 g and 10 kg) is extremely good and that the deviation in the upper tail is small, though still increasing. All of the terms in the

[1]Harvard Medical School, Boston, Massachusetts 02115, USA. [2]David Geffen School of Medicine at the University of California at Los Angeles, Los Angeles, California 90024, USA.

©2010 Macmillan Publishers Limited. All rights reserved

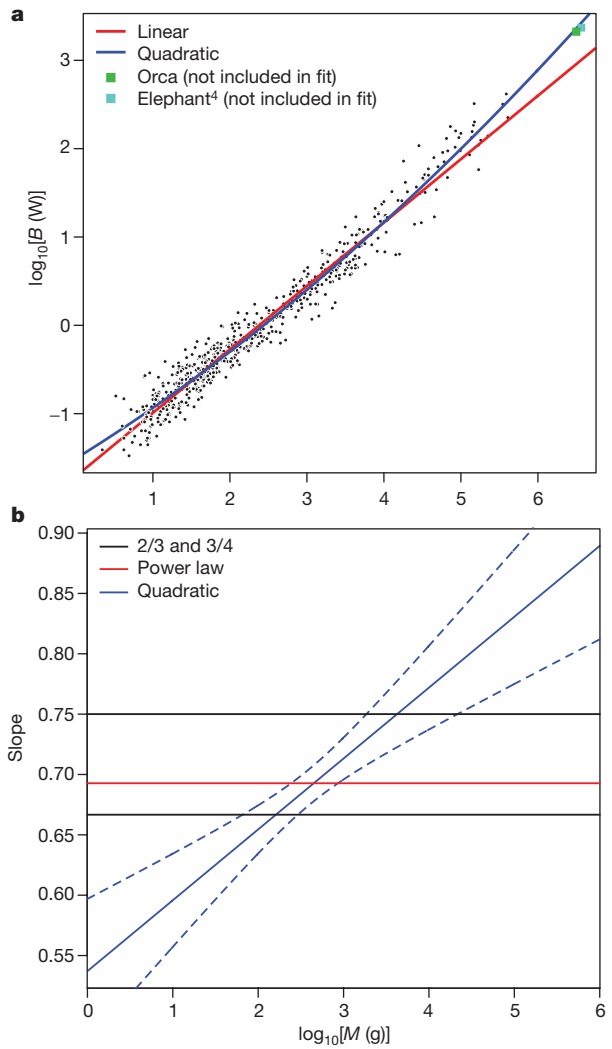


Figure 1 | Curvature in metabolic scaling. a, Linear (red) and quadratic (blue) fits (not including temperature) of $\log_{10}B$ versus $\log_{10}M$. The orca (green square) and Asian elephant (ref. 4; turquoise square at larger mass) are not included in the fit, but are predicted well. Differences in the quality of fit are best seen in terms of the conditional mean of the error, estimated by the lowess (locally-weighted scatterplot smoothing) fit of the residuals (Supplementary Information). See Table 1 for the values of the coefficients obtained from the fit. **b**, Slope of the quadratic fit (including temperature) with pointwise 95% confidence intervals (blue). The slope of the power-law fit (red) and models with fixed 2/3 and 3/4 exponents (black) are included for comparison. This panel suggests that exponents estimated by assuming a power law will be highly sensitive to the mass range of the data set used, as shown in Fig. 2.

Table 1 | Regression coefficients without and with temperature correction

Regression coefficient	Estimate	Standard error	*P* value
Without temperature correction*			
β_0	−1.5078	0.0377	$<2\times10^{-16}$
β_1	0.5400	0.0295	$<2\times10^{-16}$
β_2	0.0322	0.0053	8.9560×10^{-10}
With temperature correction†			
β_0	14.0149	1.1826	$<2\times10^{-16}$
β_1	0.5371	0.0305	$<2\times10^{-16}$
β_2	0.0294	0.0057	2.5680×10^{-7}
β_T	−4,799.0	362.22	$<2\times10^{-16}$

Regression coefficients, standard errors, and *P* values for quadratic models without and with temperature correction (for mass in grams, basal metabolic rate in watts, and temperature in kelvin). The former use the full McNab data set (minus the orca) of 636 species; the latter use a subset of 447 species for which we obtained temperature data. All coefficients are highly significant.

* $\log_{10}B = \beta_0 + \beta_1\log_{10}M + \beta_2(\log_{10}M)^2 + \varepsilon$.

† $\log_{10}B = \beta_0 + \beta_1\log_{10}M + \beta_2(\log_{10}M)^2 + \beta_T/\mathrm{T} + \varepsilon$.

regression are extremely significant ($P<3\times10^{-7}$ or better), suggesting that both the temperature and quadratic terms are important predictors of metabolic rate. From the value of β_T (the coefficient of the inverse temperature term) obtained from the quadratic fit, we calculate an effective activation energy of 21.9 ± 3.2 kcal mol^{-1} or 0.95 ± 0.14 eV (95% confidence intervals). This value is less than the free energy of the full hydrolysis of ATP to AMP under standard cellular conditions (26 kcal mol^{-1} or 1.13 eV; ref. 27), indicating that the model produces a biologically realistic coefficient.

In addition to temperature, previous studies have attempted to control for other factors that may affect metabolic rate, such as shared evolutionary history[16,28], habitat, climate and food type[8]. To account for these potential effects, we analyse the data using phylogenetic generalized least squares regression[29] and by conditioning on categorical variables (Supplementary Information). For both analyses, we find that the quadratic and temperature terms remain significant, with some changes in the magnitude of the coefficients (Supplementary Information). We also find that no single study or group of points is responsible for the curvature in the data, and that the quadratic and temperature terms remain significant across a variety of subsets of the data (Supplementary Information). These results suggest that the nonlinearity of the relationship between basal metabolic rate and mass on a logarithmic scale is highly robust.

The local scaling exponent, defined as the derivative of the scaling relationship (equation (4)) with respect to $\log_{10}M$, increases significantly—from 0.57 to 0.87—over the range of the fitted data (Fig. 1b). This stands in sharp contrast to the constant exponent of a pure power law, and indicates that the relationship between metabolic rate and mass is quite different for large and small animals. This finding explains the long-standing disagreement regarding the value of the scaling exponent, because assuming a power law at the outset results in linear fits to curved data. Carrying out such fits yields scaling exponents similar to the slopes of tangent lines at the mean of the $\log_{10}M$ distribution of the underlying data sets (Supplementary Information). Indeed, performing linear fits over partial mass ranges confirms this increasing trend and reveals different regions of the data that are consistent with either 2/3 or 3/4 (Fig. 2). Using the values of β_1 and β_2 from the fit of the full model (equation (4)), we can predict the scaling exponents obtained in previous studies using only the first three moments of their $\log_{10}M$ distributions (Fig. 2d, Supplementary Information). In general, we find that data sets with fewer large mammals[7,14] tend to exhibit smaller exponents than ones weighting large mammals more heavily[1,4]. Together, these results indicate that curvature in the data is a major factor underlying the historical variation in estimates of the scaling exponent (Supplementary Information).

Our findings have critical implications for theories of metabolic scaling. The West, Brown and Enquist (WBE) model[17] derives equation (1) as a consequence of the relationship between the volume of a vascular network (which is proportional to mass) and the number of capillaries (which is proportional to metabolic rate). However, it predicts pure 3/4-power scaling only as an asymptotic law in the limit of infinite body mass. For animals of finite size, the model instead yields an (implicit) scaling relation that exhibits curvature on a logarithmic scale[30]:

$$M = c_0B + c_1B^{4/3} \qquad (5)$$

Under the assumptions of West *et al.*, both coefficients in the extended model (equation (5)) are positive, predicting concave curvature—not the convex curvature found in the data—and resulting in a relatively poor fit (Fig. 3a and Supplementary Information). This raises the question of whether the theory can be adapted to agree with the data.

The WBE model posits that evolution resulted in a hierarchical vascular system that minimizes energy loss in the transport of blood. This assumption appears as an energy minimization criterion that

©2010 Macmillan Publishers Limited. All rights reserved

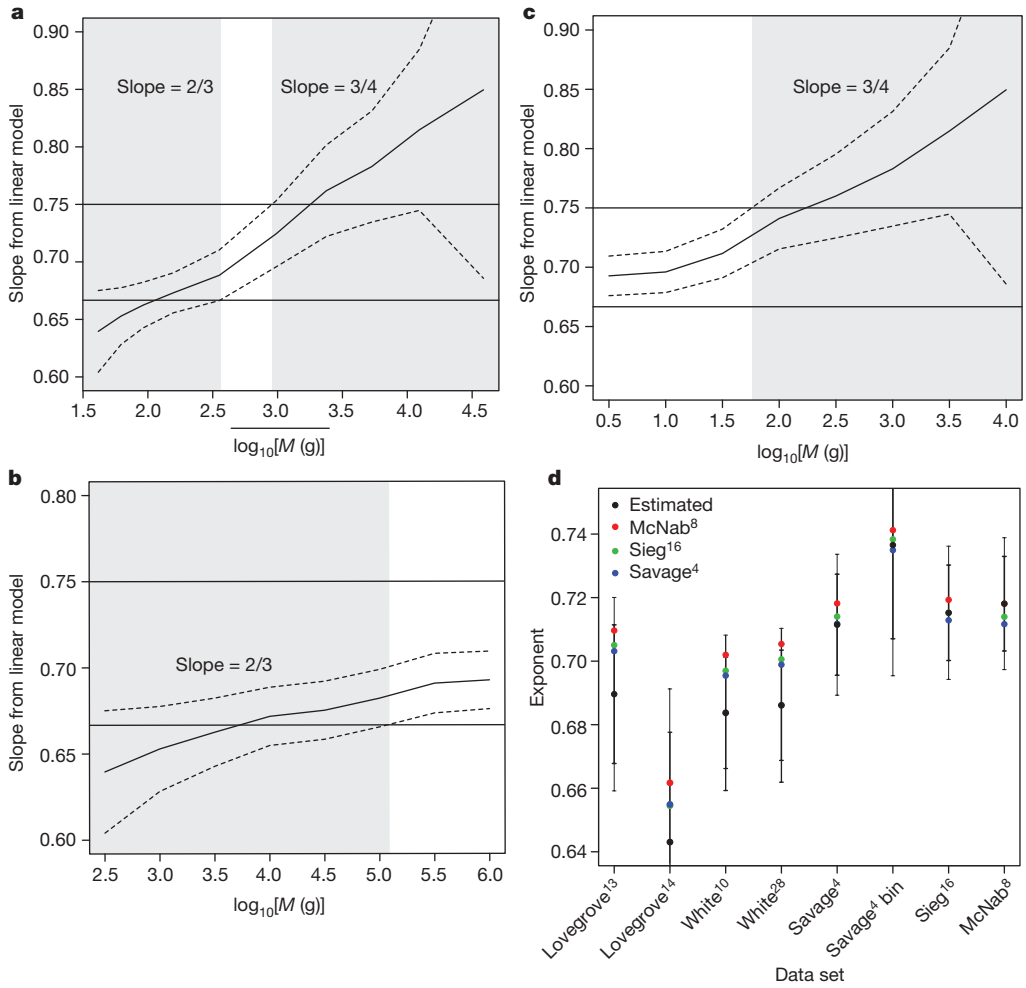


Figure 2 | Scaling exponent depends on mass range. **a**, Slope estimated by linear regression within a three log-unit mass range (smaller near the boundaries). Values on the abscissa denote mean $\log_{10}M$ within the range. When the 95% confidence regions (dashed lines) include the 2/3 or 3/4 lines, the local slope is consistent with a 2/3 or 3/4 exponent, respectively. These cases are indicated by the shaded regions (2/3 on the left and 3/4 on the right). **b**, Slope estimated by using all data points with $M < x$. The shaded region is consistent with 2/3 slope estimates. **c**, Slope estimated by using all data points with $M > x$. The shaded region is consistent with 3/4 slope estimates. **d**, Exponents estimated for eight historical data sets using linear regression (black filled circles): Lovegrove[13], Lovegrove[14], White[10], White[28], Sieg[16], McNab[8], and Savage[4] using species average data ('Savage[4]') and binned data ('Savage[4] bin'). Exponents predicted using coefficients from quadratic fits to McNab's (red), Sieg's (green), or Savage's (blue) data and the first three moments of $\log_{10}M$ (Supplementary Information). Thick lines represent uncorrected 95% confidence intervals. Thin lines are multiplicity corrected intervals.

fixes the vessel geometry (Supplementary Information). In the model, the vascular system is composed of two parts: large vessels with pulsatile blood flow and small vessels with smooth blood flow. The transition between these regions happens abruptly a constant number of levels from the capillaries. Together, these assumptions yield equation (5) (Supplementary Information). However, the calculation neglects physical effects, such as the attenuation of pulses as they travel away from the heart, which may affect the behaviour of large vessels and the position and nature of the transition between vessel types. This suggests several modifications to the model (Supplementary Information).

We first relax the assumptions about vessel geometry (model RG, 'relaxed geometry') in the pulsatile regime, resulting in a version of equation (5) in which the asymptotic exponent is no longer 3/4, but c_0 and c_1 are still positive, thus failing to produce convex curvature. Next, we modify the location of the transition between flow regimes. In one possibility, the transition occurs a constant number of levels from the heart (model FH, 'from heart'), rather than from the capillaries. In another possibility, the transition occurs a constant fraction of levels from the heart (model PT, 'proportional transition'). Both modifications lead to models that predict convex curvature, as detected in the data (Fig. 3a and b). However, the fit of the FH model is almost as poor as the original WBE model (Fig. 3a, Supplementary Information). In contrast, the PT model fits nearly as well as the quadratic model, suggesting that it merits further investigation. These modifications demonstrate that the WBE model can, in principle, be brought into agreement with the observed curvature, while still preserving core assumptions, such as the primacy of resource distribution networks. A more detailed energy minimization calculation should help to determine if these adaptations represent physically realistic cases or suggest alternative corrections.

The WBE model and its variants necessarily predict an asymptotic scaling exponent, suggesting that metabolic rate does not limit animal size without additional assumptions, such as the existence of a minimal cellular metabolic rate. On the other hand, the quadratic model with temperature (equation (4)), which provides the best fit to the data, predicts that the slope of the scaling function increases without bound (though this apparent behaviour may be due to the paucity of data for large animals). If this is correct, the metabolic scaling relationship may directly determine maximum animal size. This limit might occur at the mass at which the slope equals 1. Beyond this point, bigger is no longer better, meaning that an x% increase in

©2010 Macmillan Publishers Limited. All rights reserved

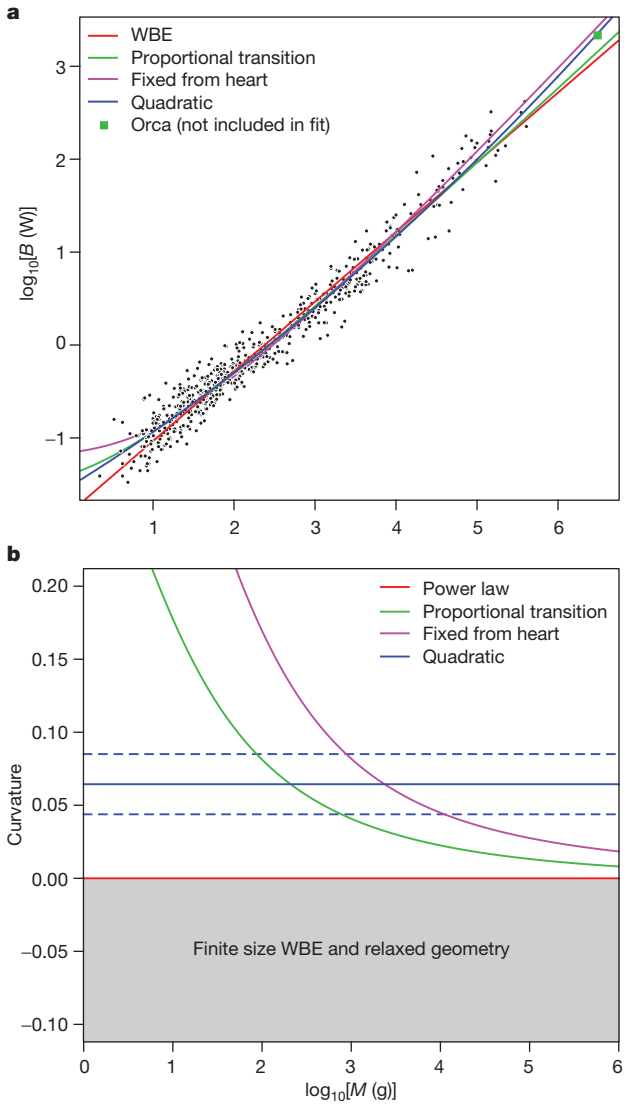


Figure 3 | Modified WBE models. a, Fits of the 'proportional transition' (PT, green) and 'from heart' (FH, magenta) models. Fits of the quadratic (blue) and WBE (red) models are included for comparison. The FH model posits that the transition from large to small vessels occurs a fixed number of levels from the heart. This is in contrast to the WBE model, which assumes that this transition occurs a fixed number of levels from the capillaries. The PT model represents another possibility, in which the transition occurs a fixed fraction of levels from the heart. **b**, Curvatures, as measured by the second derivative, achievable by the models considered in the main text. The quadratic model (blue) provides an estimate of the empirical curvature of the data itself. Dashed lines represent 95% confidence intervals. Pure power-law models (red) have no curvature, which is inconsistent with the data. The finite-size corrected WBE and 'relaxed geometry' (RG) models (the latter is the variant with relaxed geometry for large vessels) exhibit negative or concave curvature (not shown in **a**), which is also inconsistent with the data. The PT (green) and FH (magenta) variants of WBE have mass-dependent positive or convex curvature, consistent with the data, and asymptotically have no curvature (meaning that they become pure power laws for very large animals).

body mass requires a greater than x% increase in metabolic rate. Our fit suggests that this point occurs around 10^8 g (100 t): intriguingly, this is about the size of the blue whale, which is believed to be the largest animal that has ever lived.

Received 1 November 2009; accepted 12 February 2010.

1. Kleiber, M. Body size and metabolism. *Hilgardia* **6,** 315–353 (1932).
2. Bartels, H. Metabolic rate of mammals equals the 0.75 power of their body weight. *Exp. Biol. Med.* **7,** 1–11 (1982).
3. Feldman, H. A. & McMahon, T. A. The 3/4 mass exponent for energy metabolism is not a statistical artifact. *Respir. Physiol.* **52,** 149–163 (1983).
4. Savage, V. M. *et al.* The predominance of quarter-power scaling in biology. *Funct. Ecol.* **18,** 257–282 (2004).
5. Heusner, A. A. Energy metabolism and body size. I. Is the 0.75 mass exponent of Kleiber's equation a statistical artifact? *Respir. Physiol.* **48,** 1–12 (1982).
6. Dodds, P. S., Rothman, D. H. & Weitz, J. S. Re-examination of the "3/4-law" of metabolism. *J. Theor. Biol.* **209,** 9–27 (2001).
7. White, C. R. & Seymour, R. S. Mammalian basal metabolic rate is proportional to body mass$^{2/3}$. *Proc. Natl Acad. Sci. USA* **100,** 4046–4049 (2003).
8. McNab, B. K. An analysis of the factors that influence the level and scaling of mammalian BMR. *Comp. Biochem. Physiol. A* **151,** 5–28 (2008).
9. Glazier, D. S. A unifying explanation for diverse metabolic scaling in animals and plants. *Biol. Rev. Camb. Phil. Soc.* **85,** 111–138 (2009).
10. White, C. R., Phillips, N. F. & Seymour, R. S. The scaling and temperature dependence of vertebrate metabolism. *Biol. Lett.* **2,** 125–127 (2006).
11. McNab, B. K. Complications inherent in scaling the basal rate of metabolism in mammals. *Q. Rev. Biol.* **63,** 25–54 (1988).
12. Bokma, F. Evidence against universal metabolic allometry. *Funct. Ecol.* **18,** 184–187 (2004).
13. Lovegrove, B. G. The zoogeography of mammalian basal metabolic rate. *Am. Nat.* **156,** 201–219 (2000).
14. Lovegrove, B. G. The influence of climate on the metabolic rate of small mammals: a slow-fast metabolic continuum. *J. Comp. Physiol. B* **173,** 87–112 (2003).
15. Heusner, A. A. Size and power in mammals. *J. Exp. Biol.* **160,** 25–54 (1991).
16. Sieg, A. E. *et al.* Mammalian metabolic allometry: do intraspecific variation, phylogeny, and regression models matter? *Am. Nat.* **174,** 720–733 (2009).
17. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276,** 122–126 (1997).
18. Lindstedt, S. L. & Calder, W. A. Body size, physiological time, and the longevity of homeothermic mammals. *Q. Rev. Biol.* **56,** 1–16 (1981).
19. Schmidt-Nielsen, K. *Scaling: Why Is Animal Size So Important?* (Cambridge Univ. Press, 1983).
20. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85,** 1771–1789 (2004).
21. McMahon, T. A. & Bonner, J. T. *On Size and Life* (Scientific American Library, 1983).
22. Peters, R. H. *The Ecological Implications of Body Size* (Cambridge Univ. Press, 1983).
23. Calder, W. A. *Size, Function, and Life History* (Dover, 1996).
24. Heusner, A. A. Energy metabolism and body size. II. Dimensional analysis and energetic non-similarity. *Respir. Physiol.* **48,** 13–25 (1982).
25. Robinson, W. R., Peters, R. H. & Zimmermann, J. The effects of body size and temperature on metabolic rate of organisms. *Can. J. Zool.* **61,** 281–288 (1983).
26. Gillooly, J. F. & Allen, A. P. Changes in body temperature influence the scaling of $V_{O2,max}$ and aerobic scope in mammals. *Biol. Lett.* **3,** 99–102 (2007).
27. Alberts, B. *et al. Molecular Biology of the Cell* 4th edn (Garland Science, 2002).
28. White, C. R., Blackburn, T. M. & Seymour, R. S. Phylogenetically informed analysis of the allometry of mammalian basal metabolic rate supports neither geometric nor quarter-power scaling. *Evolution* **63,** 2658–2667 (2009).
29. Grafen, A. The phylogenetic regression. *Phil. Trans. R. Soc. Lond. B* **326,** 119–157 (1989).
30. Savage, V. M., Deeds, E. J. & Fontana, W. Sizing up allometric scaling theory. *PLoS Comp. Biol.* **4,** e1000171 (2008).

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

Acknowledgements We thank the staff of the library of the Museum of Comparative Zoology at Harvard University for their assistance and access to their resources. A. Duncan assisted in indexing and copying original printed materials.

Author Contributions T.K. carried out the statistical analysis, performed the analytical calculations, and annotated the data of ref. 8; T.K., V.S. and E.J.D. extended and analysed vascular scaling models; T.K., V.S., E.J.D. and W.F. interpreted the results and wrote the paper.

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 W.F. (walter@hms.harvard.edu).

©2010 Macmillan Publishers Limited. All rights reserved

SUPPLEMENTARY INFORMATION

Contents

1 Regression Coefficients for Various Data Sets

In place of standard linear regression, we use generalized estimating equations with an independent working correlation [1], considering each data point as its own cluster. This method gives identical coefficient estimates to those found by standard linear regression, but uses a robust estimator for the variance that attempts to correct for heteroscedasticity. Generally, this method gives larger estimates for the variance of the coefficients (and thus larger confidence intervals) than standard linear regression, reflecting the fact that the data contains less information than if it were homoscedastic. All statistical analysis utilized R [2].

When analyzing the subset of the metabolic data for which temperature measurements are available, we also report coefficients for (linear and quadratic) fits that do not include the temperature term ($1/T$), for comparison.

1.1 Data from McNab (2008)

The data set is described in [3]. Since we exclude the orca, 636 of the 637 species are used in all fits that do not include temperature. However, inclusion of the orca does not disproportionately influence the fit, as shown in Section 1.1.3. For 447 species, we also extracted temperature measurements from the original papers. This smaller data set is used for those fits that include temperature effects. For a phylogenetic generalized least squares (PGLS) analysis of the McNab data set, see Section 2.2.4.

1.1.1 Without temperature (636 species)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	-1.7092	0.0179	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.7181	0.0076	$< 2 \cdot 10^{-16}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	-1.5078	0.0377	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5400	0.0295	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0322	0.0053	$8.956 \cdot 10^{-10}$

1.1.2 With cubic term (636 species)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \beta_3(\log_{10}(Mass))^3 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.6144	0.0811	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.6863	0.0985	$3.170 \cdot 10^{-12}$
$\hat{\beta_2}$	-0.0250	0.0363	$4.922 \cdot 10^{-1}$
$\hat{\beta_3}$	0.0066	0.0041	$1.083 \cdot 10^{-1}$

1.1.3 With orca (637 species)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.7133	0.0179	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.7201	0.0076	$< 2 \cdot 10^{-16}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.5094	0.0355	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.5415	0.0269	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0319	0.0046	$6.278 \cdot 10^{-12}$

1.1.4 With temperature (447 species)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.7003	0.0226	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.7057	0.0099	$< 2 \cdot 10^{-16}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.5434	0.0467	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.5632	0.0371	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0269	0.0068	$8.470 \cdot 10^{-5}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \frac{\beta_3}{T} + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	13.6265	1.1963	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.6927	0.0085	$< 2 \cdot 10^{-16}$
$\hat{\beta}_3$	-4731.87	367.96	$< 2 \cdot 10^{-16}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \frac{\beta_3}{T} + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	14.0149	1.1826	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5371	0.0305	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0294	0.0057	$2.568 \cdot 10^{-7}$
$\hat{\beta}_3$	-4798.95	362.22	$< 2 \cdot 10^{-16}$

1.2 Data from Sieg (2009)

The data set is described in [4]. For a PGLS analysis of the Sieg data set, see Section 2.2.4.

1.2.1 Without temperature (695 species)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	0.5852	0.0187	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.7152	0.0077	$< 2 \cdot 10^{-16}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	0.8010	0.0391	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5262	0.0295	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0340	0.0050	$1.724 \cdot 10^{-11}$

1.2.2 With temperature (535 species)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	0.6291	0.0218	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.6861	0.0092	$< 2 \cdot 10^{-16}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	0.7528	0.0488	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.5707	0.0402	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0226	0.0076	$3.104 \cdot 10^{-3}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \frac{\beta_3}{T} + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	15.3303	1.0671	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.6741	0.0077	$< 2 \cdot 10^{-16}$
$\hat{\beta_3}$	-4540.90	328.52	$< 2 \cdot 10^{-16}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \frac{\beta_3}{T} + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	15.8313	1.0792	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.5311	0.0304	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0279	0.0057	$8.750 \cdot 10^{-7}$
$\hat{\beta_3}$	-4648.40	331.38	$< 2 \cdot 10^{-16}$

1.3 Data from Savage (2004)

The data set is described in [5]. We exclude the Asian Elephant from most analyses involving this data set, leaving 625 species. However, including the elephant leads to very similar results, as shown in Section 1.3.2.

1.3.1 Without elephant (625 species)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.6769	0.0192	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.7091	0.0080	$< 2 \cdot 10^{-16}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.4789	0.0421	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.5324	0.0323	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0324	0.0057	$1.019 \cdot 10^{-8}$

1.3.2 With elephant (626 species)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.6816	0.0193	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.7115	0.0081	$< 2 \cdot 10^{-16}$

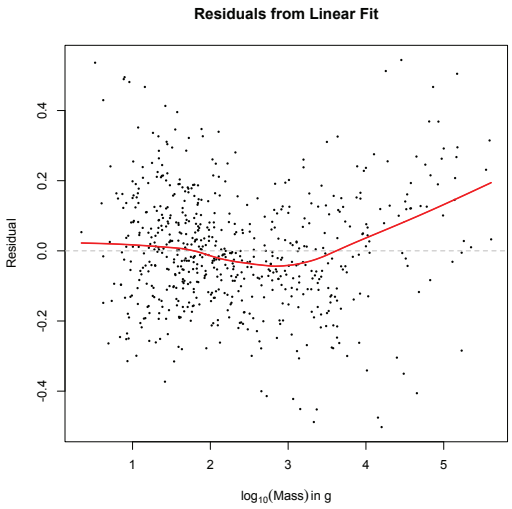
$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.4812	0.0393	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.5347	0.0291	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0320	0.0049	$8.014 \cdot 10^{-11}$

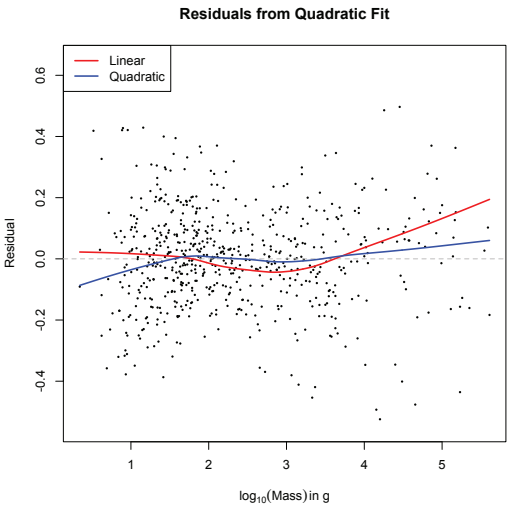
2 Regression Diagnostics

2.1 Analysis of Residuals

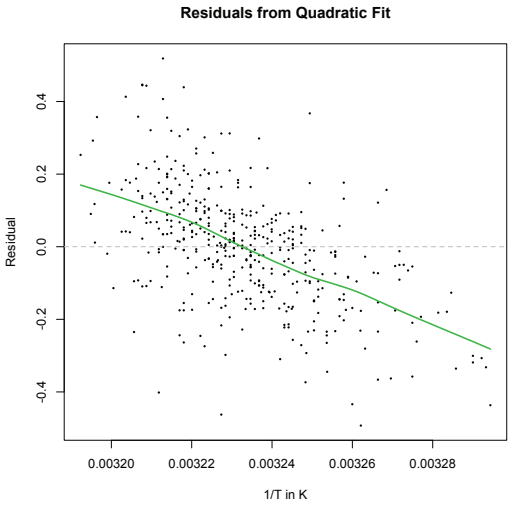
The analysis of residuals plays a major role in analyzing the quality of fit of a regression model. As noted in the main text, the linear model fits the data quite badly, a fact that is evident from a plot of the residuals vs. $\log_{10}(Mass)$ (Figure 1a). In a correctly specified model, the conditional mean of the error, estimated by the lowess fit of the residuals [6], is close to zero. The conditional mean here deviates significantly from zero, indicating that the pure power-law model is incorrect. Adding a quadratic term significantly improves the fit, but the conditional mean still diverges substantially from zero (Figure 1b). This suggests that we determine whether the behavior of the


Residuals from Linear Fit
Residual
log10(Mass) in g

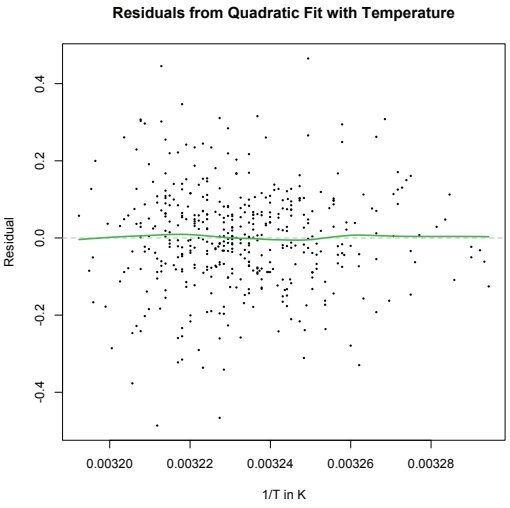
a


Residuals from Quadratic Fit
Linear
Quadratic
Residual
log10(Mass) in g

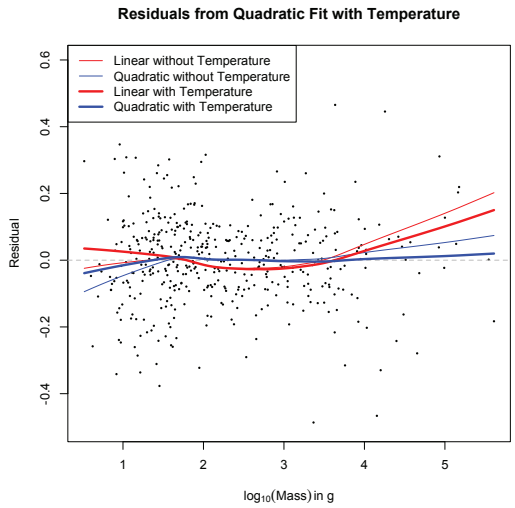
b


Residuals from Quadratic Fit
Residual
1/T in K

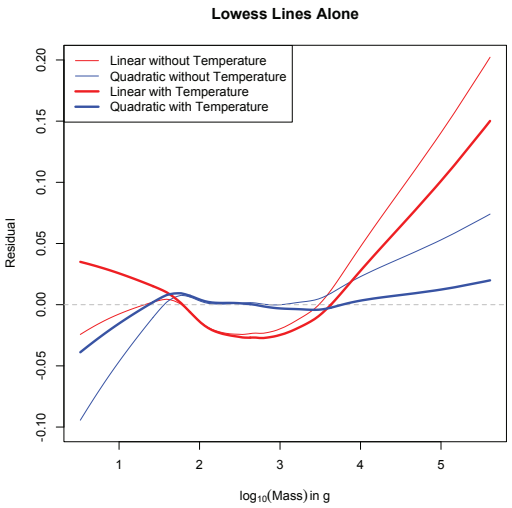
c


Residuals from Quadratic Fit with Temperature
Residual
1/T in K

d


Residuals from Quadratic Fit with Temperature
Linear without Temperature
Quadratic without Temperature
Linear with Temperature
Quadratic with Temperature
Residual
log10(Mass) in g

e


Lowess Lines Alone
Linear without Temperature
Quadratic without Temperature
Linear with Temperature
Quadratic with Temperature
Residual
log10(Mass) in g

f

Figure 1: **a:** Residuals from the linear fit with lowess line (red). **b:** Residuals from the quadratic fit with lowess line (blue). The lowess line for the linear fit (red) is overlaid for comparison. **c:** Residuals from the quadratic fit plotted against inverse temperature with lowess line (green). **d:** Residuals from the full model (quadratic plus temperature) plotted against inverse temperature with lowess line (green). **e:** Residuals from the full model with lowess line (thick blue). Lowess lines for the linear (thin red), quadratic (thin blue), and linear with temperature (thick red) models are overlaid for comparison. **f:** Lowess lines from **e** presented alone to emphasize differences.

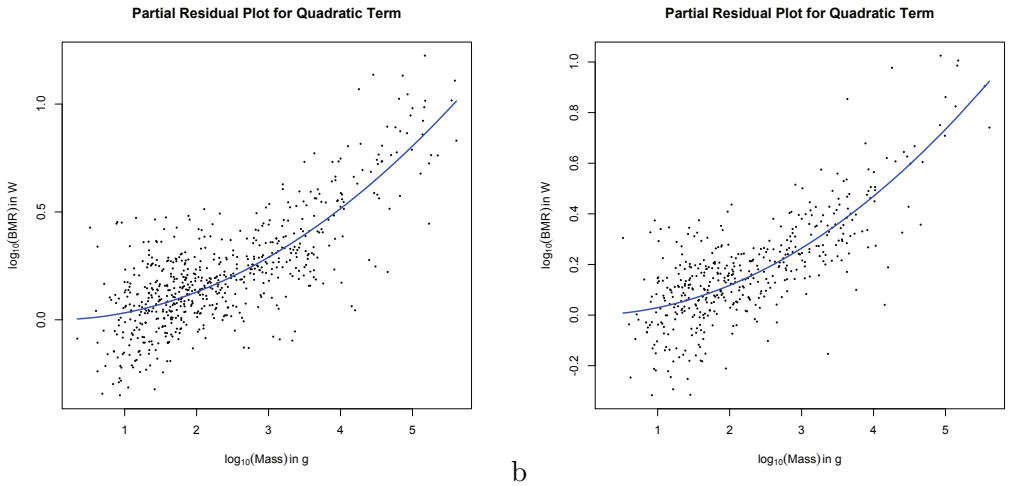
residuals corresponds to other known predictors. Temperature is one potential factor. Indeed, refitting a quadratic model using the subset of the data with associated temperature information and plotting the residuals vs. inverse Temperature shows a strong linear trend (Figure 1c). Adding Temperature (as a $1/T$ term) to the model shows a dramatic improvement, as can be seen when plotting the new residuals vs. $1/T$ (Figure 1d). From the residuals we see that the $1/T$ term appears to fully account for the effect of Temperature without the need for higher order terms. Figure 1e shows the residuals from the full model vs. $\log_{10}(Mass)$ and the lowess lines for the linear and quadratic models with and without temperature. As indicated by the thick blue lowess line, the full model provides a superior fit to any of the submodels, as discussed in the main text. Figure 1f emphasizes the difference in fits by focusing on the lowess lines alone. Figure 2 shows partial residual plots (also called residual plus component plots) for the quadratic term. These plots separate the effect of one term of a regression from the others. Figure 2a shows the partial residuals for the quadratic term of the quadratic model without temperature ($\log_{10}(BMR) - \beta_0 - \beta_1 \log_{10}(Mass)$). Figure 2b shows the partial residuals for the quadratic term of the full model ($\log_{10}(BMR) - \beta_0 - \beta_1 \log_{10}(Mass) - \frac{\beta_3}{T}$). These plots show that, with or without temperature, the curvature captured by the quadratic term occurs over approximately one order of magnitude. This is significantly less than the four orders of magnitude that are largely explained by the linear trend, but nevertheless constitutes an important property of the data.

2.2 Sensitivity Analysis (Robustness of Results)

In this section we check whether our findings are valid in general, or are just artifacts of the data. We do this by testing for and removing influential points, controlling for a collection of environmental and physiological factors, and accounting for phylogeny. All data points are from McNab [3], as in the main text.

2.2.1 Data Sources

If some or all of the data points share an unnoticed commonality, the estimates obtained from the data could be biased, since the data may not be reflective of mammals in general. One way in which this could occur would be if the data includes one or more large studies that systematically differ from the others. Fortunately, this is not the case for McNab's, Sieg's, or Savage's data. McNab's data comes from 321 primary sources of which the majority (212) contribute only a single data point. Only 18 sources contribute more than 5 data points. The largest four studies



a b

Figure 2: Partial residual plots for the quadratic term for models without (**a**) and with (**b**) temperature. These plots remove the effects of the other terms in the model and show the change in BMR that is directly attributable to the quadratic term. In both cases, the blue line is $\beta_2(\log_{10}(Mass))^2$. See text for details.

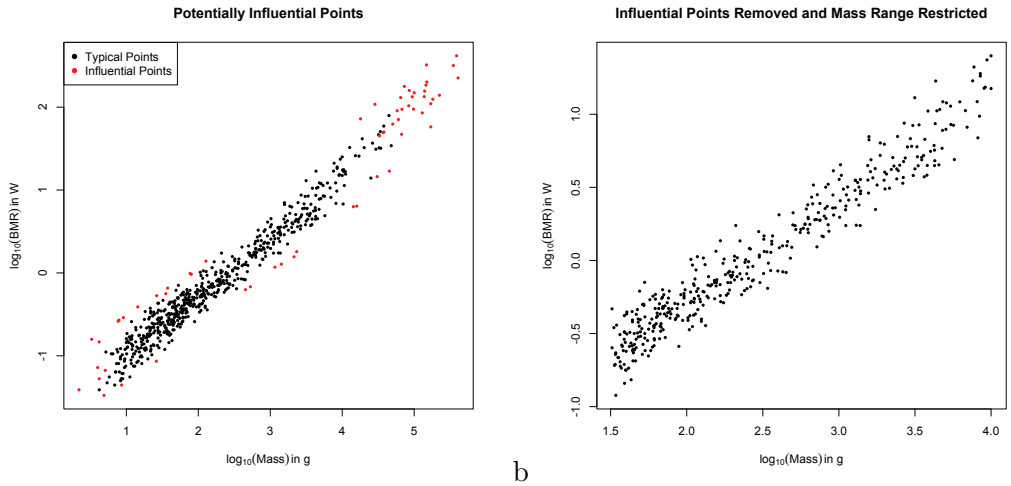
are authored by McNab and contain 20, 20, 15, and 11 animals. Perhaps not surprisingly, McNab's compilation is enriched for measurements taken from his own work. It is important to note that the data sets from Savage et al. [5] and Sieg et al. [4] are not as strongly enriched for McNab's data, but nonetheless yield essentially the same results (Sections 1.2 and 1.3). Additionally, since McNab has selected the points in his compilation, the other studies included should be the ones that are most consistent with his methods, potentially removing some confounding factors. Specifically, McNab has attempted to select only adult, post-absorptive, resting individuals within their thermoneutral zones, since these criteria are required for metabolic rate measurements to be basal.

2.2.2 Identifying Influential Points

We next consider whether any points appear to be unduly influential as measured by a collection of standard diagnostic statistics: DFBETAS, DFFITS, covariance ratio, Cook's Distance, and the magnitude of the diagonal entries of the hat matrix. These statistics detect whether a point influences individual coefficient values, the fit at that point, the covariance matrix, or all coefficient values or has high leverage, respectively. If any of these quantities suggest that a point might be influential, we consider it as such. Points so identified are shown in red in Figure 3a.

A total of 59 potentially influential points come from 45 different studies, of which only 8 contribute more than one animal and the largest contributor contains 5. Figure 3a shows that large animals may have a pronounced influence on the fit, which is to be expected since the effect of curvature will increase as mass increases. There is a second group of potentially influential points that seem to be separated from the main concentration of data, primarily at small masses. A third group is composed of points that simply appear to be isolated from the rest of the data.

In order to determine the effect of these points on the fit, we simply remove them and refit. The



a b

Figure 3: **a:** Plot of the data with potentially influential points marked in red. **b:** Data with potentially influential points removed and mass range restricted to 2.5 orders of magnitude.

data appears much more uniform with these points removed, yet the fit is largely unchanged. The quadratic term remains significant, though its magnitude has decreased somewhat, which is not surprising considering that we have eliminated many of the points that contribute most to the curvature. To determine how the significance of the quadratic term responds to further reductions in the data set, we remove points from both ends of the remaining data (.5 orders of magnitude at a time). Again, we see that the quadratic term remains significant even after reducing the range of the data to 2.5 orders of magnitude in mass (1.5 to 4). At this point the data appears very regular with no apparent outliers (Figure 3b), but the magnitude of the quadratic coefficient begins to fluctuate and its standard error starts to increase. Given that the quadratic term remains significant despite the removal of influential points and a considerable reduction of the mass range, we conclude that the significance of the quadratic term is due to an inherent property of the data rather than a particular set of influential points.

$$\text{Model: } \log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \epsilon$$

Complete McNab data set (636 species), without temperature

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.5078	0.0377	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.5400	0.0295	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0322	0.0053	$8.956 \cdot 10^{-10}$

Species remaining (577) after removal of influential points

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	-1.5871	0.0407	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.6051	0.0350	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0205	0.0070	$3.190 \cdot 10^{-3}$

Restriction to mass range $1 \leq \log_{10}(Mass) \leq 4.5$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	-1.5010	0.0508	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5402	0.0439	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0315	0.0088	$3.521 \cdot 10^{-4}$

Restriction to mass range $1.5 \leq \log_{10}(Mass) \leq 4$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	-1.4461	0.1013	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.4980	0.0818	$1.149 \cdot 10^{-9}$
$\hat{\beta}_2$	0.0390	0.0157	$1.280 \cdot 10^{-2}$

We repeat the analysis for the full model (quadratic with temperature). In this case, only 47 points from 37 studies are potentially influential. Of these, only 6 studies are represented more than once (the largest contribution consists of 5 data points). Proceeding as before, we first remove the potentially influential points and then incrementally restrict the mass range. The quadratic and temperature terms remain significant even after the mass range is reduced to 3 orders of magnitude (1 to 4).

$$\text{Model: } \log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \frac{\beta_3}{T} + \epsilon$$

McNab data set with temperature (447 species)

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	14.0149	1.1826	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5371	0.0305	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0294	0.0057	$2.568 \cdot 10^{-7}$
$\hat{\beta}_3$	-4798.95	362.22	$< 2 \cdot 10^{-16}$

Species remaining (400) after removal of influential points

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	13.0989	1.1865	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5858	0.0374	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0198	0.0077	$9.681 \cdot 10^{-3}$
$\hat{\beta}_3$	-4532.04	364.87	$< 2 \cdot 10^{-16}$

Restriction to mass range $1 \leq \log_{10}(Mass) \leq 4.5$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	12.9510	1.1701	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5377	0.0428	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0286	0.0086	$8.731 \cdot 10^{-4}$
$\hat{\beta}_3$	-4467.64	360.22	$< 2 \cdot 10^{-16}$

Restriction to mass range $1 \leq \log_{10}(Mass) \leq 4$

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	12.8797	1.1681	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5342	0.0455	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0294	0.0093	$1.533 \cdot 10^{-3}$
$\hat{\beta}_3$	-4444.50	358.92	$< 2 \cdot 10^{-16}$

2.2.3 Controlling for Environmental and Physiological Factors

There are likely to be factors besides temperature that impact metabolic rate. Brian McNab includes seven environmental and physiological factors in his data set. In order to test whether the quadratic and temperature terms remain significant, even after accounting for these factors, we employ conditional regression. This approach estimates the quadratic and temperature terms for all the data points while allowing the intercept or both the intercept and slope to vary between groups. Conditional regression requires a likelihood for its definition, so, for this analysis, we employ standard linear regression, since generalized estimating equations do not, in general, possess a true likelihood. For each conditioning variable, the first table includes conditioning only for the intercept while the second table includes the effects of conditioning for both the intercept and linear slope.

$$\text{Model: } \log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2 (\log_{10}(Mass))^2 + \frac{\beta_3}{T} + \epsilon$$

No Conditioning

	Estimate	Standard Error	p-value
$\hat{\beta}_0$	14.0149	1.0991	$< 2 \cdot 10^{-16}$
$\hat{\beta}_1$	0.5371	0.0275	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0294	0.0051	$1.209 \cdot 10^{-8}$
$\hat{\beta}_3$	-4798.95	338.85	$< 2 \cdot 10^{-16}$

Conditioning on Food Type

	Estimate	Standard Error	p-value
$\hat{\beta}_1$	0.4891	0.0296	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0364	0.0053	$2.594 \cdot 10^{-11}$
$\hat{\beta}_3$	-3974.54	384.48	$< 2 \cdot 10^{-16}$

	Estimate	Standard Error	p-value
$\hat{\beta}_2$	0.0360	0.0079	$7.918 \cdot 10^{-6}$
$\hat{\beta}_3$	-3654.50	408.08	$< 2 \cdot 10^{-16}$

Conditioning on Climate

	Estimate	Standard Error	p-value
$\hat{\beta}_1$	0.5424	0.0271	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0289	0.0050	$1.524 \cdot 10^{-8}$
$\hat{\beta}_3$	-4349.98	350.00	$< 2 \cdot 10^{-16}$

	Estimate	Standard Error	p-value
$\hat{\beta}_2$	0.0294	0.0053	$4.483 \cdot 10^{-8}$
$\hat{\beta}_3$	-4319.37	352.67	$< 2 \cdot 10^{-16}$

Conditioning on Habitat

	Estimate	Standard Error	p-value
$\hat{\beta}_1$	0.5563	0.0270	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0245	0.0050	$1.582 \cdot 10^{-6}$
$\hat{\beta}_3$	-4757.88	331.65	$< 2 \cdot 10^{-16}$

	Estimate	Standard Error	p-value
$\hat{\beta}_2$	0.0225	0.0052	$2.081 \cdot 10^{-5}$
$\hat{\beta}_3$	-4743.92	331.96	$< 2 \cdot 10^{-16}$

Conditioning on Substrate

	Estimate	Standard Error	p-value
$\hat{\beta}_1$	0.5549	0.0270	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0255	0.0049	$3.211 \cdot 10^{-7}$
$\hat{\beta}_3$	-4406.43	336.03	$< 2 \cdot 10^{-16}$

	Estimate	Standard Error	p-value
$\hat{\beta}_2$	0.0295	0.0053	$4.887 \cdot 10^{-8}$
$\hat{\beta}_3$	-4365.08	335.47	$< 2 \cdot 10^{-16}$

Conditioning on Torpor

	Estimate	Standard Error	p-value
$\hat{\beta}_1$	0.5050	0.0282	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0334	0.0051	$1.309 \cdot 10^{-10}$
$\hat{\beta}_3$	-4381.96	349.00	$< 2 \cdot 10^{-16}$

	Estimate	Standard Error	p-value
$\hat{\beta}_2$	0.0375	0.0057	$1.024 \cdot 10^{-10}$
$\hat{\beta}_3$	-4395.69	347.88	$< 2 \cdot 10^{-16}$

Conditioning on Exclusive Island Residence

	Estimate	Standard Error	p-value
$\hat{\beta}_1$	0.5350	0.0275	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0297	0.0050	$8.295 \cdot 10^{-9}$
$\hat{\beta}_3$	-4702.22	344.06	$< 2 \cdot 10^{-16}$

	Estimate	Standard Error	p-value
$\hat{\beta}_2$	0.0289	0.0051	$2.686 \cdot 10^{-8}$
$\hat{\beta}_3$	-4713.43	344.24	$< 2 \cdot 10^{-16}$

Conditioning on Exclusive Mountain Residence

	Estimate	Standard Error	p-value
$\hat{\beta}_1$	0.5403	0.0274	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0290	0.0050	$1.569 \cdot 10^{-8}$
$\hat{\beta}_3$	-4728.19	338.12	$< 2 \cdot 10^{-16}$

	Estimate	Standard Error	p-value
$\hat{\beta}_2$	0.0284	0.0050	$2.518 \cdot 10^{-8}$
$\hat{\beta}_3$	-4596.52	340.94	$< 2 \cdot 10^{-16}$

We next allow either the intercept or both the intercept and slope to vary as linear functions of all categorical variables.

	Estimate	Standard Error	p-value
$\hat{\beta}_1$	0.5128	0.0288	$< 2 \cdot 10^{-16}$
$\hat{\beta}_2$	0.0309	0.0051	$2.971 \cdot 10^{-9}$
$\hat{\beta}_3$	-2777.22	381.54	$1.839 \cdot 10^{-12}$

	Estimate	Standard Error	p-value
$\hat{\beta}_2$	0.0361	0.0083	$1.911 \cdot 10^{-5}$
$\hat{\beta}_3$	-2402.39	429.08	$4.374 \cdot 10^{-8}$

In all these regressions, the quadratic and temperature terms remain extremely significant. The quadratic term also remains within the 95% confidence interval from the regression of the full model. The same cannot be said for the temperature coefficient, which increases significantly when conditioning on food type to > -4000. In the case of the linear model for the intercept, or the intercept and slope, the coefficient increases to > -2800. This suggests that part of the effect of temperature on the regression may be as a surrogate for the influence of environmental or physiological factors.

2.2.4 Accounting for Phylogenetic Information

Standard linear regression does not account for the fact that data points for different species are not truly independent, since different animals share environments and resources, as well as evolutionary history. The previous section considered the effects of the first two factors; here we attempt to account for the third.

If the covariance structure for the data is known, regression methods can include it in the fitting process. However, one of the greatest difficulties with trying to account for covariance due to evolution is selecting this structure. Early attempts to do this assumed that the evolution of quantitative traits could be modeled as a diffusion process and that the theory of Brownian motion could be used to define the covariance structure. Such strong assumptions about the nature of evolution do not always fit the data well. One solution, originally proposed by Freckleton, Harvey and Pagel [7], is to weaken these assumptions by including an extra parameter, λ, in the covariance structure, thereby attenuating the correlation between species, while still modeling the variance for a particular species with pure Brownian motion. λ typically ranges from 0 to 1, where 0 corresponds to complete independence between points and 1 corresponds to the original diffusion model. The fitting process determines this parameter by attempting to find the optimal value for the data. Due to its increased flexibility, we use Pagel's covariance structure for the following phylogenetic generalized least squares (PGLS) regressions.

In order to use phylogenetic regression, we had to link McNab's data to a phylogenetic tree. We did this by first harmonizing his species names with the third edition of Wilson and Reeder [8]. Since the supertree we used [9] is based on the second edition of Wilson and Reeder we then matched changed spellings, species, and genus names to that edition. With this done, we attempted to associate any species that were separated from another species between the second and third editions with the original species, if this species was not already represented in McNab's data. After this procedure, we were able to link all but eight species to the supertree.

For fitting the models, we use REstricted Maximum Likelihood (REML) instead of standard Maximum Likelihood (ML), since it accounts for the loss of degrees of freedom in the estimation process. We fit the models using the APE and NLME packages for R [2, 10].

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	-1.7964	0.1078	$< 2 \cdot 10^{-16}$
$\hat{\beta_1}$	0.6614	0.0290	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0126	0.0050	$1.162 \cdot 10^{-2}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \frac{\beta_3}{T} + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	7.2170	1.3310	$9.750 \cdot 10^{-8}$
$\hat{\beta_1}$	0.6054	0.0327	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0218	0.0058	$1.759 \cdot 10^{-4}$
$\hat{\beta_3}$	-2748.12	406.85	$4.582 \cdot 10^{-11}$

Both the quadratic and temperature terms remain significant using PGLS. Interestingly, while the magnitude of the quadratic coefficient decreases significantly in the fit without temperature, its value is much more stable when temperature is included. As in the case of conditioning on categorical factors (Section 2.2.3), the value of the temperature coefficient increases dramatically, suggesting that part of the apparent effect of temperature is to act as a surrogate for the role of shared evolution. However, when conditioning on categorical factors, the magnitude of the quadratic term increased, while in this case it decreases. The meaning of this is unclear.

We can also allow the intercept to vary as a linear function of the categorical variables within the PGLS framework. Unfortunately, attempting to also treat the slope in this manner leads to singularities in the fitting process.

PGLS and categorical variables (intercept variation only)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_1}$	0.6261	0.0291	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0123	0.0048	$1.063 \cdot 10^{-2}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \frac{\beta_3}{T} + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_1}$	0.5694	0.0348	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0223	0.0059	$1.648 \cdot 10^{-4}$
$\hat{\beta_3}$	-1688.89	430.20	$1.023 \cdot 10^{-4}$

Interestingly, the inclusion of the categorical variables does not seem to affect the values or significances of the coefficients, with the exception of the temperature term, which increases further. This may be an indication that the effect of temperature is partially due to its role as a surrogate for evolutionary and environmental factors. The relative constancy of the other coefficients may indicate that the covariance between points in the PGLS model captures much of the information encoded by the categorical variables (at least when they are only allowed to affect the intercept).

Finally, we repeated our PGLS analysis on data from Sieg et al. [4], using the same supertree as above [9]. 685 and 529 data points were used for PGLS without and with temperature, respectively. As one can see from the tables below, Sieg's data yields similar results to McNab's.

PGLS for the data from Sieg (2009)

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	0.5577	0.1135	$1.121 \cdot 10^{-6}$
$\hat{\beta_1}$	0.6171	0.0320	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0199	0.0054	$2.314 \cdot 10^{-4}$

$$\log_{10}(BMR) = \beta_0 + \beta_1 \log_{10}(Mass) + \beta_2(\log_{10}(Mass))^2 + \frac{\beta_3}{T} + \epsilon$$

	Estimate	Standard Error	p-value
$\hat{\beta_0}$	10.2386	1.3540	$1.787 \cdot 10^{-13}$
$\hat{\beta_1}$	0.5879	0.0353	$< 2 \cdot 10^{-16}$
$\hat{\beta_2}$	0.0211	0.0065	$1.279 \cdot 10^{-3}$
$\hat{\beta_3}$	-2955.96	414.37	$3.260 \cdot 10^{-12}$

3 Linear Regression and the Slope of a Quadratic Function

3.1 Derivation

We show that a (misspecified) linear regression of a quadratic function estimates its tangent at the mean of the data.

Let $Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \epsilon$ be a correctly specified model for a data set (X, Y), such that the errors, ϵ, satisfy $E(\epsilon|X) = 0$ (i.e. their conditional mean is 0). Note that the ϵ terms in the derivation below correspond to the errors of the *quadratic* model, and not the errors for the linear model. If we consider fitting a linear model $E(Y|X) = \beta_0 + \beta_1 X$ for (X, Y), the estimated slope of the linear regression $\hat{\beta}_1$ will be given by:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i (Y_i - \overline{Y})(X_i - \overline{X})}{\sum_i (X_i - \overline{X})^2} = \frac{\sum_i (\alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \epsilon_i - (\alpha_0 + \alpha_1 \overline{X} + \alpha_2 \overline{X^2}))(X_i - \overline{X})}{\sum_i (X_i - \overline{X})^2} \\
&= \frac{\sum_i (\alpha_1 (X_i - \overline{X}) + \alpha_2 (X_i^2 - \overline{X^2}) + \epsilon_i)(X_i - \overline{X})}{\sum_i (X_i - \overline{X})^2} \\
&= \frac{\sum_i (\alpha_1 (X_i - \overline{X})^2 + \alpha_2 (X_i^2 - \overline{X^2})(X_i - \overline{X}) + \epsilon_i (X_i - \overline{X}))}{\sum_i (X_i - \overline{X})^2} \\
&= \alpha_1 \frac{\sum_i (X_i - \overline{X})^2}{\sum_i (X_i - \overline{X})^2} + \alpha_2 \frac{\sum_i (X_i^2 - \overline{X^2})(X_i - \overline{X})}{\sum_i (X_i - \overline{X})^2} + \frac{\sum_i \epsilon_i (X_i - \overline{X})}{\sum_i (X_i - \overline{X})^2} \\
&= \alpha_1 + \alpha_2 \frac{\sum_i (X_i^2 - \overline{X^2}) X_i}{\sum_i (X_i - \overline{X})^2} + \frac{\sum_i \epsilon_i (X_i - \overline{X})}{\sum_i (X_i - \overline{X})^2} \qquad (1)
\end{aligned}$$

The second term can be further simplified:

$$\begin{aligned}
\alpha_2 \frac{\sum_i (X_i^2 - \overline{X^2}) X_i}{\sum_i (X_i - \overline{X})^2} &= \alpha_2 \frac{\sum_i (X_i^3 - \overline{X^2} X_i)}{\sum_i (X_i - \overline{X})^2} = \alpha_2 \frac{\sum_i (X_i^3 - \overline{X^2}\,\overline{X})}{\sum_i (X_i - \overline{X})^2} = \alpha_2 \frac{\sum_i (X_i^3 - \overline{X} X_i^2)}{\sum_i (X_i - \overline{X})^2} \\
&= \alpha_2 \frac{\sum_i (X_i^3 - \overline{X} X_i^2 - 2\overline{X}(X_i - \overline{X})^2 + 2\overline{X}(X_i - \overline{X})^2)}{\sum_i (X_i - \overline{X})^2} \\
&= \alpha_2 \frac{\sum_i (X_i^3 - \overline{X} X_i^2 - 2\overline{X}(X_i^2 - 2\overline{X} X_i + \overline{X}^2))}{\sum_i (X_i - \overline{X})^2} + 2\alpha_2 \overline{X} \frac{\sum_i (X_i - \overline{X})^2}{\sum_i (X_i - \overline{X})^2} \\
&= 2\alpha_2 \overline{X} + \alpha_2 \frac{\sum_i (X_i^3 - 3\overline{X} X_i^2 + 4\overline{X}^2 X_i - 2\overline{X}^3)}{\sum_i (X_i - \overline{X})^2} \\
&= 2\alpha_2 \overline{X} + \alpha_2 \frac{\sum_i (X_i^3 - 3\overline{X} X_i^2 + 3\overline{X}^2 X_i + \overline{X}^2 X_i - 2\overline{X}^3)}{\sum_i (X_i - \overline{X})^2} \\
&= 2\alpha_2 \overline{X} + \alpha_2 \frac{\sum_i (X_i^3 - 3\overline{X} X_i^2 + 3\overline{X}^2 X_i + \overline{X}^3 - 2\overline{X}^3)}{\sum_i (X_i - \overline{X})^2} \\
&= 2\alpha_2 \overline{X} + \alpha_2 \frac{\sum_i (X_i^3 - 3\overline{X} X_i^2 + 3\overline{X}^2 X_i - \overline{X}^3)}{\sum_i (X_i - \overline{X})^2} = 2\alpha_2 \overline{X} + \alpha_2 \frac{\sum_i (X_i - \overline{X})^3}{\sum_i (X_i - \overline{X})^2}
\end{aligned}$$

Equation (1) thus becomes:

$$\hat{\beta}_1 = \alpha_1 + 2\alpha_2\overline{X} + \alpha_2\frac{\sum_i (X_i-\overline{X})^3}{\sum_i (X_i-\overline{X})^2} + \frac{\sum_i \epsilon_i(X_i-\overline{X})}{\sum_i (X_i-\overline{X})^2},$$

or

$$\hat{\beta}_1 - (\alpha_1 + 2\alpha_2\overline{X}) = \alpha_2\frac{\sum_i (X_i-\overline{X})^3}{\sum_i (X_i-\overline{X})^2} + \frac{\sum_i \epsilon_i(X_i-\overline{X})}{\sum_i (X_i-\overline{X})^2}. \quad (2)$$

Let $M \geq \sup|X_i - \overline{X}|$. Then:

$$|\sum_i (X_i-\overline{X})^3| \leq \sum_i |X_i-\overline{X}|^3 \leq \sum_i M|X_i-\overline{X}|^2 = M\sum_i (X_i-\overline{X})^2 \quad (3)$$

Taking the expectation of equation (2) conditional on X and using (3) yields:

$$\begin{aligned}
&|\mathrm{E}(\hat{\beta}_1|X) - (\alpha_1 + 2\alpha_2\overline{X})| = |\mathrm{E}(\hat{\beta}_1 - (\alpha_1 + 2\alpha_2\overline{X})|X)| \\
&\leq \left|\alpha_2\frac{\sum_i (X_i-\overline{X})^3}{\sum_i (X_i-\overline{X})^2} + \frac{\sum_i \mathrm{E}(\epsilon_i(X_i-\overline{X})|X)}{\sum_i (X_i-\overline{X})^2}\right| \leq |\alpha_2|\left|\frac{\sum_i (X_i-\overline{X})^3}{\sum_i (X_i-\overline{X})^2}\right| + \left|\frac{\sum_i \mathrm{E}(\epsilon_i(X_i-\overline{X})|X)}{\sum_i (X_i-\overline{X})^2}\right| \\
&\leq |\alpha_2|\frac{\sum_i M(X_i-\overline{X})^2}{\sum_i (X_i-\overline{X})^2} + \left|\frac{\sum_i \mathrm{E}(\epsilon_i|X)(X_i-\overline{X})}{\sum_i (X_i-\overline{X})^2}\right| = |\alpha_2|M\frac{\sum_i (X_i-\overline{X})^2}{\sum_i (X_i-\overline{X})^2} + \left|\frac{\sum_i \mathrm{E}(\epsilon_i|X)(X_i-\overline{X})}{\sum_i (X_i-\overline{X})^2}\right| \\
&= |\alpha_2|M + \left|\frac{\sum_i 0(X_i-\overline{X})}{\sum_i (X_i-\overline{X})^2}\right| = |\alpha_2|M
\end{aligned}$$

Thus,

$$|\mathrm{E}(\hat{\beta}_1|X) - (\alpha_1 + 2\alpha_2\overline{X})| \leq |\alpha_2|M \quad (4)$$

The absolute difference between the conditional expectation of $\hat{\beta}_1$ given X and the tangent of the quadratic curve at the mean of the data is thus less than or equal to $|\alpha_2|M$. This means that, as M gets larger (i.e., as the maximum difference between the X_i's and the mean of X increases), $\hat{\beta}_1$ will, in general, become a less accurate estimate for the slope of the tangent. The larger the range of the X_i's, the more sensitive the results become to the distribution of the data.

Under independent homoscedastic errors (for heteroscedastic ϵ with variance bounded by σ^2 the following result will be an upper bound) we obtain:

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_1|X) &= \mathrm{Var}\left(\alpha_1 + 2\alpha_2\overline{X} + \alpha_2\frac{\sum_i (X_i-\overline{X})^3}{\sum_i (X_i-\overline{X})^2} + \frac{\sum_i \epsilon_i(X_i-\overline{X})}{\sum_i (X_i-\overline{X})^2}|X\right) \\
&= \mathrm{Var}\left(\frac{\sum_i \epsilon_i(X_i-\overline{X})}{\sum_i (X_i-\overline{X})^2}|X\right) = \frac{\sum_{i,j}\mathrm{Cov}(\epsilon_i(X_i-\overline{X}),\epsilon_j(X_j-\overline{X})|X)}{\left(\sum_i (X_i-\overline{X})^2\right)^2} \\
&= \frac{\sum_{i,j}(X_i-\overline{X})(X_j-\overline{X})\mathrm{Cov}(\epsilon_i,\epsilon_j|X)}{\left(\sum_i (X_i-\overline{X})^2\right)^2} = \frac{\sum_{i,j}(X_i-\overline{X})(X_j-\overline{X})\sigma^2\delta_{i,j}}{\left(\sum_i (X_i-\overline{X})^2\right)^2} \\
&= \frac{\sigma^2\sum_i (X_i-\overline{X})^2}{\left(\sum_i (X_i-\overline{X})^2\right)^2}
\end{aligned}$$

Therefore:

$$\mathrm{Var}(\hat{\beta}_1|X) = \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2} \tag{5}$$

Together, equations (4) and (5) imply that $\hat{\beta}_1$ is a reasonable estimator of the slope of the quadratic function at the mean of the data, $\alpha_1 + 2\alpha_2\overline{X}$, and that this estimate can be made arbitrarily accurate by choosing a sufficiently small region (resulting in a small M) and sufficiently large n. Also note that $\sum_i (X_i - \overline{X})^3$ is a measure of the asymmetry of the sample and therefore, for symmetrically distributed X, the estimator may still be accurate, even over large regions.

3.2 Estimated and Predicted Exponents

Using the formulae derived above, we can attempt to predict the exponents that would be estimated for various data sets by using simple linear regression (as is frequently done). Taking the expectation of equation (2) conditional on X gives:

$$\mathrm{E}(\hat{\beta}_1|X) = \alpha_1 + 2\alpha_2\overline{X} + \alpha_2\frac{\sum_i (X_i - \overline{X})^3}{\sum_i (X_i - \overline{X})^2}. \tag{6}$$

Using equation (6) we predict the exponents for seven well known data sets: Lovegrove (2000) [11], Lovegrove (2003) [12], White (2006) [13], White (2009) [14], Sieg (2009) [4], McNab (2008) [3] and Savage (2004) [5] (using both species average data and binned data).

We compare these predictions to the exponents that would be estimated by simply applying a standard linear fit (in log-log space) to each data set. (Note that these estimates may not match the actual results reported in a particular paper, since the investigators may have used more complex methods.) The estimated exponents are shown in Figure 4a. Indeed, the exponents appear to increase with $\overline{\log_{10}(Mass)}$, as predicted by equation (6). Figure 4a indicates that this trend is not perfectly linear, presumably because some data sets have a skewed $\log_{10}(Mass)$ distribution. We therefore include the last term of equation (6) when calculating our predictions. Since we do not have true values for α_1 and α_2, we estimate them by fitting a quadratic model to McNab's, Sieg's, or Savage's data. Panels C and D of Figure 4 show that our predicted exponents using (6) are reasonably close to the estimated exponents, regardless of the data set we use when fitting the quadratic model. In all cases, the predicted exponents fall within the 95% multiplicity corrected confidence intervals of the estimated exponents, and in all but two cases (both using McNab's data), they fall within the uncorrected confidence intervals. Since both the estimated and predicted exponents are random variables, it is the confidence interval for the difference between the two exponents that matters. Figure 4b shows that, for all data sets considered, this difference is always included within the uncorrected 95% confidence interval, meaning that our predictions are consistent with the data, even without correcting for multiplicity.

We conclude that the quadratic model not only fits the data better than the linear model, but seems to be able to predict the exponent that is observed in a given study, using only information about the first three moments of the $\log_{10}(Mass)$ distribution. Although these predictions are quite good, and are consistent with the data, they are not perfect, and appear to be biased upwards. This is likely because, as we have shown, the quadratic model does not completely

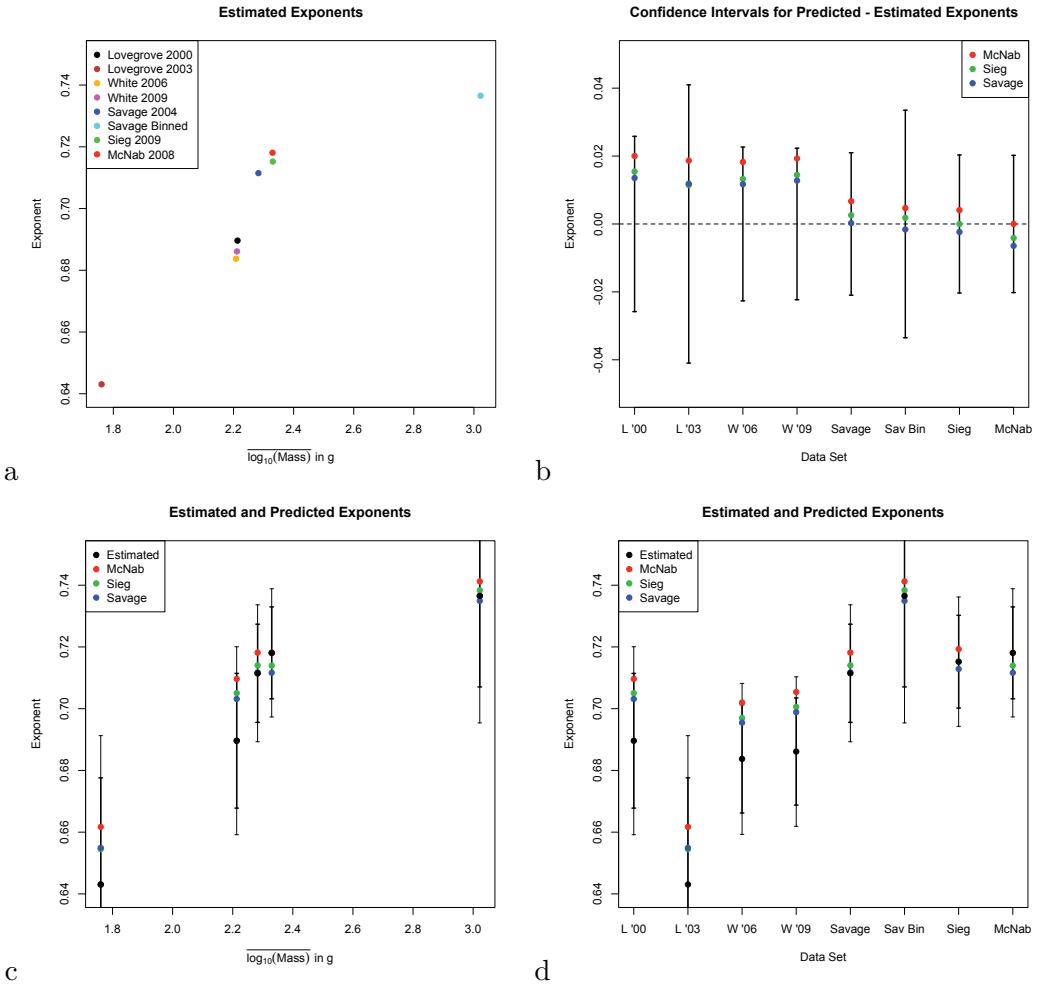


Figure 4: **a:** Exponents estimated for various data sets using simple linear regression. **b:** Exponents of the data sets mentioned in the text were predicted using equation (6) with α_1 and α_2 estimated by the coefficients obtained from quadratic fits to McNab's, Sieg's, or Savage's data. The panel shows the 95% uncorrected confidence intervals for the differences between the predicted exponents and exponents estimated using linear regression. All differences fall within the 95% confidence intervals. The displayed confidence intervals are computed using McNab's data, but are almost identical for Sieg's or Savage's data. **c:** Estimated and predicted exponents vs. $\overline{\log_{10}(Mass)}$. Thick lines represent uncorrected 95% confidence intervals. Thin lines are multiplicity corrected intervals. In this panel, only five of eight data sets are included (Lovegrove (2000 and 2003), Savage (unbinned and binned), and McNab (2008)), since the masses of the other three data sets are very similar to some of those plotted and are therefore obscured. **d:** As in **c**, but estimated and predicted exponents are now organized by data set. (This panel also appears in the main text as Figure 2d.)

explain BMR. Temperature is also an important predictor that must be included, and there will certainly be additional factors that influence BMR, like those considered by McNab. Despite this, the simple quadratic model predicts scaling exponents remarkably well. Much of the variance in observed exponents obtained in previous large-scale studies can therefore be attributed to the curvature in the data (which is captured by the coefficient of the quadratic term). As seen in equation (6), due to differences in body mass distributions between studies, the nonlinearity in the true scaling relationship can result in divergent estimates of "the scaling exponent."

4 Extending the West-Brown-Enquist model

For many symbolic computations we utilized Mathematica [15].

4.1 The finite-size WBE model

We provide a brief summary of the derivation of the West-Brown-Enquist (WBE) model [16, 17] to provide context for the modifications detailed in Section 4.2. These new models result in several functions that we fit in the main text, in order to identify WBE variants that can produce the convex dependence on body mass observed in the data. A full explanation of the model's assumptions and its derivation can be found in [17].

The central assumption of WBE is that the basal metabolic rate of an organism, B, is determined by the structure of the network it uses to distribute resources throughout its body. In the case of mammals, this network is the arterial vascular system, which the model assumes to be a hierarchical tree beginning with a single vessel at level 0 (the aorta). At each level, the vessels branch into n symmetric daughter vessels of identical radius and length. This process continues until the network reaches the capillaries, which are assumed to have the same size (radius and length) across species. To service every part of the organism, the network is assumed to be space filling. This means that, at each branch point, the sum of the cubed lengths of the daughter vessels will be equal to the cubed length of the parent vessel: for the branching of a vessel k levels from the aorta, $l_k^3 = \sum_{i=0}^{n} l_{k+1,i}^3 = n l_{k+1}^3$ (where i indexes daughter vessels and the second equality holds because the daughter vessels have identical lengths). Defining γ as the ratio of the length of a daughter vessel to its parent, gives $\gamma \equiv l_{k+1}/l_k = n^{-1/3}$. Given the length of the aorta (l_0), we can calculate the length of the vessels at level k as $l_k = l_0(l_1/l_0)(l_2/l_1)\dots(l_k/l_{k-1}) = l_0\gamma^k$.

The WBE model also assumes that vascular networks have evolved to minimize energy loss as blood flows through the network. Combining this assumption with the space filling constraint allows one to calculate how the *radii* of daughter vessels are related to the *radii* of parent vessels. The energy minimization calculation is quite involved and is discussed in detail elsewhere [17], but we provide a brief sketch. In order to simplify the calculation, WBE separates the network into two regions. The first region includes vessels near the heart, where blood flow is pulsatile due to the action of the heart. The second region includes small vessels near the capillaries, where blood flow is smooth (i.e. Poiseuille) and dominated by viscous forces. In the WBE model, the pulse is not damped in the large vessels, but instead abruptly ceases at level $\bar{k}$, where the flow becomes smooth.

Under these assumptions, energy loss in large vessels is minimized by eliminating the reflection of waves at branch points, which occurs if the impedance to flow in the parent vessel is exactly equal to the total impedance of the daughter vessels. This "impedance matching" leads to the constraint that $r_k^2 = \sum_{i=0}^{n} r_{k+1,i}^2 = n r_{k+1}^2$, so that $\beta_< \equiv r_{k+1}/r_k = n^{-1/2}$ for $k \leq \bar{k}$. For small vessels, energy loss is minimized by minimizing the viscous drag of the fluid against the vessel walls. Combined with other assumptions, this yields the constraint $r_k^3 = \sum_{i=0}^{n} r_{k+1,i}^3 = n r_{k+1}^3$, which is known as Murray's Law [18–20]. Therefore, $\beta_> \equiv r_{k+1}/r_k = n^{-1/3}$ for $k > \bar{k}$. Using the assumption that capillary radius is constant across species, these ratios can be used to directly compute the radius of any vessel in the network.

Using the ratios defined above we can determine the volume of a vascular network, which is equivalent to the volume of blood in the body (i.e. V_b, neglecting a factor of 2 that would arise from including the blood in the venous system). To do so, we calculate the volume of vessels at each level k (each of which is considered to be a cylinder, with $V_k = \pi r_k^2 l_k$), multiply this volume by the number of vessels at that level, and sum over all levels k while respecting the transition between $\beta_<$ and $\beta_>$ at $\bar{k}$. For an organism with N total levels of vasculature, this yields:

$$V_b = N_c V_c \left((n\gamma\beta_>^2)^{\bar{k}-N} \sum_{k=0}^{\bar{k}} (n\gamma\beta_<^2)^{k-\bar{k}} + \sum_{k=\bar{k}+1}^{N} (n\gamma\beta_>^2)^{k-N} \right) \tag{7}$$

where N_c is the number of capillaries in the network and V_c is the volume of a capillary. From the relationships above, we obtain $n\gamma\beta_<^2 = n n^{-1/3} n^{-1} = n^{-1/3}$ and $n\gamma\beta_>^2 = n n^{-1/3} n^{-2/3} = 1$. Substituting these results into equation (7), we obtain:

$$V_b = N_c V_c \left(\sum_{k=0}^{\bar{k}} (n^{-1/3})^{(k-\bar{k})} + \sum_{k=\bar{k}+1}^{N} 1 \right) = N_c V_c \left(\frac{n^{(\bar{k}+1)/3} - 1}{n^{1/3} - 1} + (N - \bar{k}) \right) \tag{8}$$

Evaluating equation (8) depends on the definition of $\bar{k}$. In WBE, the transition between the two regimes occurs at a constant radius, r_{trans} [16]. Since WBE assumes that the radius and length of capillaries do not vary significantly across organisms of different sizes, this transition must occur a constant number of levels before the capillaries. In this case, we can define a new constant, $\bar{N} \equiv N - \bar{k}$, that does not depend on the size of the organism in question. Noting that $N_c = n^N$, we have:

$$\begin{aligned} V_b &= N_c V_c \left(\frac{n^{(N-\bar{N}+1)/3} - 1}{n^{1/3} - 1} + \bar{N} \right) = N_c V_c \left(\frac{N_c^{1/3} n^{(1-\bar{N})/3} - 1}{n^{1/3} - 1} + \bar{N} \right) \\ &= N_c V_c \left(\left(\bar{N} - \frac{1}{n^{1/3} - 1} \right) + \frac{n^{(1-\bar{N})/3}}{n^{1/3} - 1} N_c^{1/3} \right) \end{aligned} \tag{9}$$

The final set of assumptions in the model is that the metabolic rate of an organism is proportional to its number of capillaries (i.e. $B = c_B N_C$) and that body mass is proportional to

blood volume ($M = c_M V_b$). This yields:

$$\begin{aligned} M &= \frac{c_M}{c_B} B V_c \left(\left(\bar{N} - \frac{1}{n^{1/3}-1} \right) + \frac{n^{(1-\bar{N})/3}}{n^{1/3}-1} \left(\frac{B}{c_B} \right)^{1/3} \right) \\ &= \left(\frac{c_M V_c}{c_B} \left(\bar{N} - \frac{1}{n^{1/3}-1} \right) \right) B + \left(\frac{c_M V_c}{c_B^{4/3}} \frac{n^{(1-\bar{N})/3}}{n^{1/3}-1} \right) B^{4/3} \\ &= c_0 B + c_1 B^{4/3} \end{aligned} \tag{10}$$

where c_0 and c_1 represent the two constant terms multiplying B and $B^{4/3}$, respectively. Equation (10) is equation (5) of the main text. In the limit of infinite B (or, equivalently, the limit of infinite mass), equation (10) reduces to the pure power-law relationship $B = (M/c_1)^{3/4}$, i.e. 3/4-power scaling.

The proportionality constants relating M to V_b and B to N_c are unknown, making c_0 and c_1 free parameters. Given $n \geq 2$ and that V_C, c_B and c_M are all positive, it is clear that $c_1 > 0$. Equation (10) implies that $c_0 > 0$ if and only if $\bar{N} > \frac{1}{n^{1/3}-1}$. The value of $\bar{N}$ can be estimated by using known parameters of the vascular system (such as the viscosity of the blood, the size of the capillaries, etc.) to calculate r_{trans}. For $n = 2$ (which is reasonable for the mammalian vasculature), one obtains $\bar{N} \approx 24$ [16, 17]. This is consistent with measurements of actual vessels in mammals, in which $\bar{N}$ is at least 7 [20]. Thus, for the WBE model as originally formulated, we have the additional constraint that $c_0 > 0$.

Since equation (10) is a quartic function in $B^{1/3}$ (i.e. $M = c_0(B^{1/3})^3 + c_1(B^{1/3})^4$), we can invert it to find a function f such that $B = f(M, c_0, c_1)$. We use standard non-linear least squares fitting procedures to fit this function to our data and estimate c_0 and c_1 under the constraint that both constants are positive. Given this restriction on the values of c_0 and c_1, it is clear that equation (10) has *negative* (concave) curvature, which is exactly the opposite of the *positive* (convex) curvature observed in the data. It is therefore not surprising that the constrained non-linear fit results in $\hat{c}_0 = 0$, producing a pure power law with a fixed exponent of 3/4. This fit is displayed in Figure 3a in the main text (labeled "WBE") and is considerably worse than a standard power-law fit with a free exponent (Figure 1a).

4.2 Modifications to WBE

The energy minimization calculation in WBE is quite involved and ignores a number of physical effects that could influence network geometry. Perhaps most prominent among these effects is the attenuation of pulse waves as blood travels away from the heart. In WBE, the wave number and angular frequency of blood pulses is constant from level 0 to level $\bar{k}$. At exactly that level, the blood flow abruptly changes from pulsatile to perfectly smooth. Such a sudden transition seems unphysical. Although one could, in principle, include attenuation effects to produce a more realistic model, such calculations are far beyond the scope of this work. We can, however, explore how relaxing various WBE constraints that are directly related to the energy minimization calculation (particularly in large vessels) might influence the overall scaling function obtained from equation (7).

4.2.1 Generic scale-free ratio of large-vessel radii (RG)

We first consider a simple case in which we set $\beta_< = n^{-r}$, preserving the scale-free ratio of $\beta_<$ without specifying the exact relationship between the radii of parent and daughter vessels in the large-vessel regime. Given that Murray's law has been observed empirically [18–20], it is reasonable to assume $n\gamma\beta_> = 1$. Following the derivation that led to equation (8), but using this modified $\beta_<$, we obtain:

$$\begin{aligned} M &= \frac{c_M}{c_B} B V_c \left(\left(\bar{N} - \frac{1}{n^q - 1} \right) + \frac{n^{(1-\bar{N})q}}{n^q - 1} \left(\frac{B}{c_B} \right)^q \right) \\ &= \left(\frac{c_M V_c}{c_B} \left(\bar{N} - \frac{1}{n^q - 1} \right) \right) B + \left(\frac{c_M V_c}{c_B^{q+1}} \frac{n^{(1-\bar{N})q}}{n^q - 1} \right) B^{q+1} \\ &= c_0 B + c_1 B^{q+1} \end{aligned} \tag{11}$$

where $q = 2r - 2/3$ is a new free parameter. If we have $c_0 > 0$ and $c_1 > 0$, the curvature of equation (11) is again concave, opposite to the curvature observed in the data. Since the above equation cannot be analytically inverted for arbitrary values of q, we determine its inverse using standard numerical root-finding techniques. The numerical inverse is then used to perform a non-linear fit of equation (11) to the data, providing us with estimates of c_0, c_1 and q. As in the case of the original WBE model, the mismatch in curvature between the model and the data forces $\hat{c}_0 = 0$, reducing equation (11) to a standard power-law fit. We conclude that simply relaxing the radius scaling relationship for large vessels does not result in a model that is consistent with the empirical data.

4.2.2 Flow-mode transition close to the capillaries

As mentioned previously, one aspect of the WBE model that appears to be worth exploring is the level at which the transition from pulsatile to smooth flow occurs. If it is sufficiently close to the capillaries (i.e. if $\bar{N}$ is sufficiently small), c_0 will become negative, resulting in a model with curvature equivalent to that observed in the data. As one might expect, performing fits for equations (10) or (11) without the constraint that $c_0 \geq 0$ results in negative estimates for this parameter. In the case of equation (11) the fit appears to be remarkably good and is superior to a pure power law, though it is not quite as good as the fit of the quadratic model discussed in the main text. (This fit is identical to the optimal fit of the Proportional Transition Model, seen in the main text, so Figure 3 may be used for comparison.) In order for c_0 to be negative, however, we must have $\bar{N} < \frac{1}{n^q - 1}$. Using the estimated value for q obtained from this fit, and a branching ratio of $n = 2$, implies that $\bar{N} \leq 6$ for this to be true. However, physiological measurements indicate that $\bar{N}$ is at least 7 [20], suggesting that this small-$\bar{N}$ scenario may not represent a biologically realistic case, no matter how well it fits the empirical data.

4.2.3 Flow-mode transition a constant number of levels from the heart (FH)

Here we consider the possibility that $\bar{k}$ might be constant: that is, the transition occurs a constant number of levels away from the aorta, rather than a constant number of levels from the capillaries. In this case, the first term in equation (7) becomes a constant, yielding:

$$V_b = N_c V_c \left(\frac{n^{(\bar{k}+1)q} - 1}{n^q - 1} - \bar{k} + N \right) \tag{12}$$

Since $N_c = n^N$, and therefore $N = \frac{\log N_c}{\log n}$, we have:

$$V_b = N_c V_c \left(\frac{n^{(\bar{k}+1)q} - 1}{n^q - 1} - \bar{k} + \frac{\log N_c}{\log n} \right) \tag{13}$$

Using $B = c_B N_c$, $M = c_M V_b$, as before, results in:

$$\begin{aligned} M &= \frac{c_M}{c_B} B V_c \left(\frac{n^{(\bar{k}+1)q} - 1}{n^q - 1} - \bar{k} + \frac{\log \frac{B}{c_B}}{\log n} \right) \\ &= \frac{c_M V_c}{c_B} \left(\left(\frac{n^{(\bar{k}+1)q} - 1}{n^q - 1} - \bar{k} - \frac{\log c_B}{\log n} \right) + \frac{\log B}{\log n} \right) B \\ &= \left(\frac{c_M V_c}{c_B} \left(\frac{n^{(\bar{k}+1)q} - 1}{n^q - 1} - \bar{k} - \frac{\log c_B}{\log n} \right) \right) B + \left(\frac{c_M V_c}{c_B \log n} \right) B \log B \\ &= c_0 B + c_1 B \log B \end{aligned} \tag{14}$$

where c_0 and c_1 are again free parameters in the model. In this scenario, it is clear that $c_1 > 0$, while the sign of c_0 will vary depending on the values of $\bar{k}$ and c_B, neither of which can be estimated independently at this time. Equation (14) can be inverted analytically to yield:

$$B = M \left(c_1 \mathrm{W} \left(\frac{e^{c_0/c_1} M}{c_1} \right) \right)^{-1} \tag{15}$$

where the function W is the inverse of xe^x and is known as the Lambert W function or the product log. In the limit of small mass, equation (15) has a slope of 0. The value of the slope increases up to a value of 1 in the limit of infinite mass (where the smooth-flow regime dominates). The FH model thus displays convex curvature, matching that of the data. Non-linear fits of this model, however, do not appear to fit the data well (Figure 3a). It is interesting to note that, despite the large difference in asymptotic exponents (1 for FH and 3/4 for the original WBE model [17]), both the FH and WBE models seem to fit the data similarly poorly (Figure 3a).

4.2.4 Flow-mode transition proportional to organism size (PT)

Finally, we consider a model in which the transition point in the network is some function of N (the number of total levels in the network hierarchy), with the simplest case being $\bar{k} = \alpha N$. Then,

from equation (8) we have:

$$V_b = N_c V_c \left(\frac{n^{(\alpha N+1)q} - 1}{n^q - 1} + (1-\alpha)N \right) = N_c V_c \left(\frac{n^{N\alpha q} n^q - 1}{n^q - 1} + (1-\alpha)N \right)$$
$$= N_c V_c \left(\frac{N_c^{\alpha q} n^q - 1}{n^q - 1} + (1-\alpha)\frac{\log N_c}{\log n} \right) \quad (16)$$

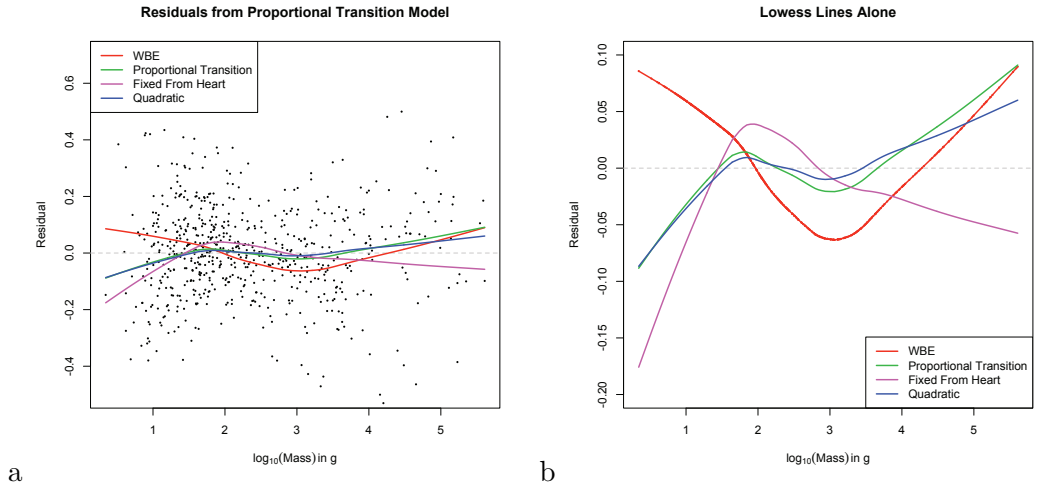
Rewriting the equation in terms of M and B, as before, yields:

$$M = \frac{c_M}{c_B} V_c \left(\frac{\left(\frac{B}{c_B}\right)^{\alpha q} n^q - 1}{n^q - 1} + (1-\alpha)\frac{\log \frac{B}{c_B}}{\log n} \right) B$$
$$= \frac{c_M V_c}{c_B} \left(\frac{n^q}{c_B^{\alpha q}(n^q-1)} B^{\alpha q} - \left(\frac{1}{n^q - 1} + \frac{(1-\alpha)\log c_B}{\log n} \right) + \frac{(1-\alpha)}{\log n} \log B \right) B$$
$$= \left(-\frac{c_M V_c}{c_B} \left(\frac{1}{n^q - 1} + \frac{(1-\alpha)\log c_B}{\log n} \right) \right) B \; + \left(\frac{(1-\alpha) c_M V_c}{c_B \log n} \right) B \log B$$
$$+ \left(\frac{c_M V_c}{c_B^{\alpha q+1}} \frac{n^q}{(n^q-1)} \right) B^{\alpha q+1}$$
$$= c_0 B + c_1 B \log B + c_2 B^p \quad (17)$$

where c_0, c_1, c_2 and $p = \alpha q + 1$ are all free parameters. This large number of parameters gives the model great flexibility. Since c_1 and c_2 are necessarily greater than 0, the scaling function can produce convex curvature, with an asymptotic exponent of p^{-1}. Since the expression has so many parameters, one cannot definitively estimate values for them, and indeed the model can result in very similar fits for different sets of parameters. Interestingly, nearly optimal fits of the scaling function (17), in particular the best fit (for which $\hat{c}_1 = 0$), tend to have shapes nearly identical to the small $\bar{N}$ case (Section 4.2.2). Equation (11) for the small $\bar{N}$ model is a special case of equation (17), which explains the performance of the small $\bar{N}$ model, despite its apparently unphysical nature. It is important to note here that, while equation (17) is capable of fitting the data well for large sets of parameters, it is unclear whether any of these parameters represent physically realistic cases. We leave further evaluation of the PT model and other modifications of WBE to future work.

4.3 Residuals of WBE Modifications

In order to evaluate the quality of fit of each of the above models, we examine their residuals. Figure 5a and 5b clearly demonstrate that both the WBE and FH models fit the data poorly, as claimed. The PT model appears to fit the data nearly as well as the quadratic model, though it requires one more parameter to do so.



a b

Figure 5: **a:** Plot of residuals from PT Model with lowess line (green). Lowess lines for the WBE (red), FH (magenta), and quadratic (blue) models are overlaid for comparison. **b:** Lowess lines from **a** presented alone in order to accentuate differences.

References

1. Liang, K. Y. and Zeger, S. L. *Biometrika* **73**(1), 13–22 (2004).
2. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, (2009). ISBN 3-900051-07-0.
3. McNab, B. *Comparative Biochemistry and Physiology A-Molecular & Integrative Physiology* **151**, 5–28 (2001).
4. Sieg, A. E., O'Conner, M. P., McNair, J. N., Grant, B. W., Agosta, S. J., and Dunham, A. E. *American Naturalist* **174**, 720–733 (2009).
5. Savage, V. M., Gillooly, J. F., Woodruff, W. H., West, G. B., Allen, A. P., Enquist, B. J., and Brown, J. H. *Func. Ecol.* **18**(2), 257–282 (2004).
6. Cleveland, W. A. *Journal of the American Statistical Association* **74**, 829–836 (1979).
7. Freckleton, R. P., Harvey, P., and Pagel, M. *American Naturalist* **160**, 712–726 (2002).
8. Wilson, D. E. and Reeder, D. M. *Mammal Species of the World. A Taxonomic and Geographic Reference (3rd ed)*. Johns Hopkins University Press, Baltimore, Maryland, (2005).
9. Bininda-Emonds, O. R., Cardillo, M., Jones, K. E., MacPhee, R. D., Beck, R. M., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L., and Purvis, A. *Nature* **446**(7135), 507–512 (2007).
10. Paradis, E., Claude, J., and Strimmer, K. *Bioinformatics* **20**, 289–290 (2004).
11. Lovegrove, B. *The American Naturalist* **156** (2000). 201–219.

12. Lovegrove, B. *Journal of Comparative Physiology B* **173** (2003). 87–112.

13. White, C. R., Phillips, N. F., and Seymour, R. S. *Biol Lett* **2**, 125–127 (2006).

14. White, C. R., Blackburn, T. M., and Seymour, R. S. *Evolution* **63**, 2658–2667 (2009).

15. Wolfram Research, Inc. *Mathematica Edition: Version 7.0.* Wolfram Research, Inc., Champaign, Illinois, (2008).

16. West, G. B., Brown, J. H., and Enquist, B. J. *Science* **276**, 122–126 (1997).

17. Savage, V. M., Deeds, E. J., and Fontana, W. *PLoS Comp. Biol.* **4**, e1000171 (2008).

18. Zamir, M. *The Physics of Coronary Blood Flow.* Springer, New York, (2005).

19. Murray, C. D. *Proc. Natl. Acad. Sci. USA* **12**(1), 207–214 (1926).

20. Sherman, T. F. *J. Gen. Physiol.* **78**, 431–453 (1981).