Identifying branching principles in biological networks using imaging, modeling, and machine learning

Alexander Byers Brummer¹,²,³, Panagiotis Lymeropoulos¹*, Jocelyn Shen⁴*, Elif Tekin¹,², Lisa P. Bentley⁵, Vanessa Buzzard⁶, Andrew Gray⁶, Imma Oliveras⁷, Brian J. Enquist⁶,⁸, & Van M. Savage¹,²,³,⁸

¹Institute for Quantitative and Computational Biology, University of California, Los Angeles
²Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles
³Department of Ecology and Evolutionary Biology, University of California, Los Angeles
⁴Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology
⁵Department of Biology, Sonoma State University
⁶Department of Ecology and Evolutionary Biology, University of Arizona
⁷Environmental Change Institute, University of Oxford
⁸Santa Fe Institute

Branching in vascular networks and in overall organismic form is one of the most common and ancient features of multicellular plants¹, fungi² and animals¹,³. These networks deliver resources and eliminate wastes from early development onward⁴, and even play a vital role in the growth, prognosis, and treatment of tumors and stroke recovery⁵–⁷. Because of these basic and applied reasons there is immense interest in identifying key fea-
tures of vascular branching and their connection to biological function\textsuperscript{4,8}. Here we classify diverse branching networks—mouse lung, human head and torso, angiosperm plants, and gymnosperm plants\textsuperscript{9–12}—by harnessing recent advances in medical imaging, algorithms and software for extracting vascular data\textsuperscript{11}, theory for resource-distribution networks\textsuperscript{12,13}, and machine-learning\textsuperscript{14,15}. Specifically, we apply standard machine-learning techniques to a variety of feature spaces. These spaces include the untransformed raw data for vessel radii and lengths as well as spaces predicted to be the most biologically informative\textsuperscript{13}. Our results show that our theoretically-informed feature spaces—especially those that determine blood flow rate—combined with Kernel Density Estimation (KDE)\textsuperscript{14,15} are best at distinguishing networks. Our categorization of networks enhances the mapping between biologic function—such as the dependence of metabolic rate on body mass—to vascular branching traits among organisms and organs. We accomplish this by analyzing how variation in metabolic scaling exponents—around the canonical value of 3/4\textsuperscript{16–18}—arises despite differences in vascular traits. Our results reveal how network categorization and variation in metabolic scaling are both heavily determined by scaling ratios of vessel radii—changes and asymmetries across branching generations—that strongly constrain rates of fluid flow. These linkages will improve understanding of evolutionary convergence across plants and animals while also potentially aiding prognosis and treatment of vascular pathologies and other diseased states.

It is a great challenge to decipher which features of biological branching networks are shared, which are different, and when these differences matter\textsuperscript{8,19}. For instance, branching in plant and animal networks exhibit strikingly similar features despite profound physiological and environmental
differences (e.g., carbon dioxide and sap versus oxygen and blood, mobile versus stationary organisms, heart and pulsatile flow versus smooth flow). Similarly, differences in loopiness and “noisiness” are well documented between vascular branching in tumors or stroke-damaged tissue versus healthy tissue. The shared branching features are argued to lead to functional convergence in plant and animal networks via biological rates despite the notable physiological differences just listed. Yet, the extent of shared versus distinct branching features has not been systematically and quantitatively analyzed across plants and animals in the same study. Consequently, there is a need to understand the forces that shape the full spectrum of form and function in branching networks (Figure 1a).

With recent advances in automated methods of image analysis developed by us and others, increasing amounts of data are becoming available to tackle these problems. The tools that are missing are efficient and accurate algorithms for categorizing branching across whole networks and different organisms. In this paper we apply machine-learning methods to theoretically-informed feature spaces to leverage all available information and technology to achieve these goals.

In this paper we analyze the largest-ever compilation of branching network data. We collected these data over the last decade for both mammalian cardiovascular systems and plant architecture in both angiosperms and gymnosperms. The mammalian networks are the major arterial branching junctions of the human head and torso (HHT) for 18 adult individuals (H. sapiens), and the full branching of one wild-type adult mouse lung (M. musculus). The plant networks consist of: 1. whole, above-ground, adult trees for one Balsa (O. pyramidale), one Piñon (P. edulis), and
Figure 1: Network diagrams and comparison of machine learning methods. a, Examples of mouse lung and angiosperm branching networks. b and c, Diagrams of positive and negative asymmetric branching. d and e, Results for the Kernel Density Estimation (KDE) and Logistic Regression (LR) methods, respectively, of classification of mammalian and plant networks. Both methods use the radius average and difference scale factors ($\bar{\beta} = (r_{c1} + r_{c2})/2r_p$, $\Delta \beta = (r_{c1} - r_{c2})/2r_p$). For each method, data are randomly split into training (75%) and testing (25%) groups. Following testing, classified points are binned based on predicted probability significance (or score), and comparison is made while varying the level of classification significance from high (left) to low (right). f, Receiving operator characteristic (ROC) curves comparing true positive versus false positive rates of classification for methods of Support Vector Machine (SVM), LR, and KDE for each significance bin. The group defined as the positive classifier varies from mammal (left), to plant (center), and finally mixed (right) where the plant networks are defined as the positive classifier for the KDE method and the mammal networks are defined as the positive classifier for the SVM and LR methods.
five Ponderosa pines (*P. ponderosa*)\(^9\), 2. an array of angiosperm root clusters belonging to Andean tropical montane cloud forests\(^10\), and 3. a collection of 50 cm long clippings of the terminal ends of canopy branches from three species each of angiosperms and gymnosperms comprised of Maple (*A. grandidentatum*), Scrub Oak (*Q. gambelii*), Robinia (*R. neomexicana*), White Fir (*A. concolor*), Douglas Fir (*P. menziesii*), and White Pine (*P. strobiformis*). Tree measurements are of the external branching structures (limbs), not the xylem that are directly responsible for water transport. Scaling relationships for the external limbs directly determine similar relationships for the internal xylem based on previous empirical studies\(^28,29\) and established branching theory\(^25,30\), thus enabling comparisons of plant and animal networks for the structure, flow, and function in the present study\(^9,27,31\).

To search for patterns, machine-learning is often applied to the full set of untransformed raw data. These raw data thus represent one feature space, yet there are always infinitely more choices of feature spaces based on specific combinations, subsets, mathematical operations (e.g., logarithms or ratios), or other transformations of the raw data. Informed choices of feature space hold the promise of greatly improving the convergence time, accuracy, and inference of machine-learning algorithms. Here we show how crucial this choice can be and the role theory plays in its selection.

The default choice for feature spaces for our networks would be the raw data—all vessel radii and lengths for branching networks. However, theory grounded in evolution, biology, and physics predicts that the parent-to-child ratios of radii and length—along with associated scaling
exponents throughout the networks\textsuperscript{24,25}—encapsulate the most biologically-informative properties because they are directly tied to organismic function. Specifically, numerous models tie these ratios to the ability of branching networks to efficiently fill space and to deliver resources\textsuperscript{13,22,24,25,30}.

We use recent theory developed by some of us (Brummer, et al. (ref. 10)) for the asymmetric branching patterns that are pervasive throughout our data. In this theory the two sibling vessels (labelled $c_1$ and $c_2$, Figure 1b-c) and the parent vessel (labelled $p$) are combined to give two radial scale factors $\beta_1 = r_{c1}/r_p$ and $\beta_2 = r_{c2}/r_p$ and two length scale factors $\gamma_1 = l_{c1}/l_p$ and $\gamma_2 = l_{c2}/l_p$. Thus the average radial and length scale factors are

$$\bar{\beta} = \frac{\beta_1 + \beta_2}{2}, \quad \bar{\gamma} = \frac{\gamma_1 + \gamma_2}{2}$$ (1)

To capture asymmetry the difference radial and length scale factors are

$$\Delta \beta = \frac{\beta_1 - \beta_2}{2}, \quad \Delta \gamma = \frac{\gamma_1 - \gamma_2}{2}$$ (2)

Corresponding constraint equations for area-preserving and space-filling branching—used in canonical models—are

$$(\bar{\beta} + \Delta \beta)^2 + (\bar{\beta} - \Delta \beta)^2 = 1$$ (3)

$$(\bar{\gamma} + \Delta \gamma)^3 + (\bar{\gamma} - \Delta \gamma)^3 = 1$$ (4)

If aspects of blood flow primarily determine differences in the formation and evolution of branching architecture, the scale factors involving vessel radius are expected to be most informative. This is because branching theory and empirical evidence show that changes in these radial
scale factors most strongly determine resistance to blood flow, speed of blood flow, and energy loss. In contrast, if the space-filling constraints and body plan of the organism primarily determine differences in branching architecture, the scale factors for vessel lengths should best discriminate. Moreover, variation in average scale factors versus difference scale factors might reflect specific types of selective pressures.

We generate distributions of our data for combinations of the raw radius and length measurements and of the scale factors \((β_1, β_2, γ_1, γ_2, \bar{β}, γ, Δβ, Δγ)\) for the combined mammal and plant networks. We then use an array of standard machine-learning techniques to categorize our network data\(^{32,33}\). We test support vector machine (SVM), logistic regression (LR), and kernel density estimation (KDE) machine-learning methods (Figure 1c-f), as well as principle components analysis (PCA) for feature space selection. See Supplemental Information for additional detail on testing protocols.

We find that the combination of the KDE method and the radial average and difference scale factors for radius \((\bar{β}, Δβ)\) are the most effective for classifying branching network data (Figure 1d and e). This finding strongly suggests that hydrodynamic principles are the primary drivers of branching patterns and overall network form. In particular, the KDE method exhibits greater sensitivity to distinguish plant networks, while the SVM and LR methods are more sensitive to distinguish mammal networks (Figure 1f). This is likely due to the fact that the KDE method excels at resolving multimodality\(^{15}\) that characterizes the radial scale factors for the plant dataset (Figures 1d and 2a). Since the distribution means are approximately equivalent, the SVM and
LR methods are strongly influenced by outliers and the higher moments comprising the mammal
dataset (Figures 1e and 2a) 33.

Interpreting the KDE results (Figure 2) we see that mammalian branching exhibits more area-
increasing branching than plants. Area-increasing branching is necessary to slow blood flow as it
travels from the heart to the capillaries and transitions from pulsing to smooth flow, a phenomenon
not present in plants. However, values of $\bar{\beta} \approx 1.0$ and $\Delta \beta \approx 0$ represent a deviation from the
theoretical predictions of $\Delta \beta = 0$ and $\bar{\beta} = 1/2^{1/3} \approx 0.794$ for the smooth flow expected in this
region. This marked increase in cross-sectional area is shared by both the HHT and ML networks
as indicated by the nearly null relative abundances of these two networks (Figure 2c) as well as
by the insignificant $p$-value score of 0.2 from the global-level implementation of the KDE method
(Extended Data Table 1). This suggests that transitions in blood flow type from pulsing to smooth
may occur across a greater range of branching generations, and begin nearer to the heart, than in
current theory17,22.

The majority of plant networks adhere to area-preservation while exhibiting a greater ten-
dency than mammals to branch asymmetrically (specifically the Balsa, Piñon, Ponderosas, and
GS Tips, Figure 2a and b). This asymmetry could arise from light-seeking behavior, self- and
wind-induced pruning, and gap-filling21,31.

We find that differentiation within mammals and plants is driven by the species level (Fig-
ure 2e and Extended Data Table 1) unrelated to plant categorization as angiosperm or gymnosperm.
This suggests the potential for high-resolution architectural-based categorization of plant taxonomy27.
Figure 2: Classification based on features (the radial scale factors $\bar{\beta}$ and $\Delta \beta$) that are related to fluid transport—blood or sap—via volume-flow rate and hydraulic resistance through networks and vessels. a, Joint and marginalized distributions for the mammal (left) and plant (right) radius scale factors using the KDE method. Mammals are divided into mouse lung and HHT, and plants are divided into the groups of Gymnosperms (GS), Angiosperms (AS), and Roots. Black contours represent lines of constant probability density, ranging from 0.5 to 0.05 in steps of 0.05. White dashed lines are graphs of the radius conservation equation for area-preservation. b, Regions of significantly ($p < 0.05$) greater joint probability density for the mammals (red) or plants (green). c, Representative diagrams of tree networks and bar plots of relative abundances of each group/species are presented for each region of significant classification in b (clockwise). Scale factor values for tree networks are determined by geometrically averaging over all classified data points within each significance region. Means and standard deviations for bar plots are determined by bootstrapping the KDE method 1000 times. Horizontal black dashed lines represent null expectations of relative abundances. These significance-region abundances are corroborated with global-level testing of all pairs of branching networks (Extended Data Table 1). The global-level test is a method that effectively integrates over the entire feature space to produce one singular $p$-value for the comparison.
For example, the Balsa is the only species present in two regions (I and III in Figure 2c). Thus, the Balsa consists of two unique branching motifs that distinguish it from GS Tips, Piñon, and Roots that each have large relative abundances in only one region. Such motif identification can be coupled to known mechanism. The branching motif of the roots in the contour for cross-sectional area decrease may be associated with (i) an abundance of fine absorbing roots that have yet to undergo any secondary growth, or (ii) the fact that the direction of the flow of water in roots is opposite to the above-ground networks34.

Connecting length-based categorization to mechanism—the space-filling constraint (Eq. 4)—remains a challenge. The combination of the KDE method and length scale factor feature space ($\bar{\gamma}, \Delta \gamma$) identified only one region of significance. In this region differentiation is driven by the plants, specifically the Piñon and Roots (Figure 3 and Extended Data Table 2). The inability of the length scale factors to inform classification between networks suggests either of two extremes—a universal architecture or completely random architecture that is being followed by both the mammals and plants12. This result is unlike the radial scaling that is strongly coupled to hydraulics. Current theory suggests that this architecture is guided by the principles of space-filling fractals13,22,24,25. However, the large deviations observed between the joint distributions of the length scale factors and the theoretical curves determined by the space-filling conservation equation (Figure 3a) support the need for including missing constraints, variables, and assumptions (e.g. branching angles, multi-fractal scaling, etc.), or alternative mathematical frameworks35–38.

To better understand the physiological and biological implications of these categorizations,
Figure 3: Classification based on features (the length scale factors $\bar{\gamma}$ and $\Delta \gamma$) that are related to costs of materials and construction for these networks as well as the extent to which they fill the space of the organisms that they are supplying with nutrients and resources. a-c, See caption for Figure 2 for description of subfigures.
we examine the influence of asymmetric branching on estimates of biological rates—specifically the metabolic scaling exponent $\theta$ that canonically relates metabolic rate $B$ to body mass $M$ as $B = M^\theta$. Previous studies spanning orders of magnitude in body mass have shown that $\theta$ converges on a value near $3/4$, yet exhibits variation specific to mammals or plants\textsuperscript{16–18}.

To probe this variation we use branching data to estimate metabolic scaling (Figure 4) by directly accounting for network geometry and size\textsuperscript{13,22,24},

$$\theta = \frac{\ln(2^N)}{\ln(2^N) + \ln(1 - \nu^{N+1}) - \ln(\nu^N(1 - \nu))}$$ (5)

where $N$ is the total number of branching generations in the network and $\nu$ represents volumetric scaling. Specification of $\nu$ allows estimation of $\theta$ under different model assumptions for symmetric ($\nu = 2/\beta_1^2\gamma_1$) or asymmetric ($\nu = 2/\beta_2^2\beta_2 + 4/\beta_2\Delta/\beta\Delta\gamma + 2/\gamma\Delta\beta^2$) branching. We also use a regression method between the number of terminal branches $N_{\text{TIPS}}$ and total volume $V_{\text{TOT}}$ distal to a given branch ($N_{\text{TIPS}} \propto V_{\text{TOT}}^{\theta}$) that does not depend directly on geometry (See Supplemental Information).

We find that asymmetric branching increases the predicted values of metabolic scaling exponents when compared to the symmetric- and regression-based methods (Figure 4a). This is due to all networks exhibiting some length asymmetry, and more importantly suggests that previous studies have underestimated metabolic scaling exponents by not accounting for such variation\textsuperscript{9,39}.

To understand which different scale factors are primarily responsible for observed variation in the predicted metabolic scaling exponents we focus on the asymmetric version of Eq. 5. Estimated metabolic scaling exponents are graphed for each individual organism in terms of the...
Figure 4: Variation in metabolic scaling exponents related to variation in branching geometry. 

a. Comparison of symmetric (red) and asymmetric (green) estimates of metabolic scaling exponents to regression (blue) based estimates. For groups with multiple species and/or multiple individuals (AS Tips, GS Tips, Ponderosa, and HHT), metabolic scaling exponents were calculated at the species/individual level when averaged. Error bars represent 95% confidence intervals. The horizontal dashed line represents a metabolic scaling exponent value of 3/4. Note that a also serves as a legend for the symbols in all other subfigures.

b. Empirically based estimates of metabolic scaling exponents are presented as functions of the geometrically averaged length and radial average scale factor values (left), and compared to theoretical predictions (right) reproduced from Brummer et al. (ref. 10).

c. Analogous results are presented as in b but instead for the length and radius difference scale factors. Solid black lines represent contours of constant metabolic scaling exponent values. Axis ranges differ between empirical data and theory-based predictions due to observed deviations from conservation equations.

d. Curvature of metabolic rate versus mass (log-log) as a function of volume scaling ($\nu_{ASYM}$) and the number of branching generations (N).
average scale factors \((\bar{\beta}, \bar{\gamma})\) in Figure 4b and difference scale factors \((\Delta\beta, \Delta\gamma)\) in Figure 4c. We compare these graphs against the corresponding theoretical predictions reproduced from Brummer et al. (ref. 10) where we have graphed the approximate form of Eq. 5,

\[
\theta \approx \frac{\ln(2)}{\ln(2) - \ln(2\bar{\beta}^2\bar{\gamma}) - \ln(1 + \frac{2\Delta\beta\Delta\gamma}{\beta\gamma} + \frac{\Delta\gamma^2}{\beta^2})}
\]  

(6)

assuming small volume scaling \((\nu < 1)\), generationally large networks \((N >> 1)\), and enforcing area-preserving and space-filling (Eqs. 3 and 4).

We observe a striking amount of grouping among the mammals and plants when graphing the metabolic scaling exponent \(\theta\) versus the average radial and length scale factors \(\bar{\beta}\) and \(\bar{\gamma}\) (Figure 4b). This indicates that the average scale factors \((\bar{\beta}, \text{and} \bar{\gamma})\) are the primary determinants of variation in the metabolic scaling exponent and thus organism function.

In contrast to previous theory and importantly for understanding how diverse branching architectures could lead to universal scaling exponents, we find near constancy of the metabolic scaling exponent despite large fluctuations in length scaling (Figure 4c). These shared exponents are likely driven by the little to no radial asymmetry observed in mammalian networks and suggests that variation in length asymmetry \((\Delta\gamma)\) in vascular networks has little influence on whole organism metabolic function in the presence of symmetric radial branching \((\Delta\beta = 0)\).

Figures 4b-c demonstrate marked deviation in the observed grouping (or lack thereof) between the empirically based estimates of Eq. 5 and the theoretical predictions of Eq. 6 for metabolic scaling. To explore this we calculate curvature between metabolic rate and mass in log-log space.
When branching networks are strictly assumed to be very large, $(N \gg 1)$ and decreasing in volume across generations ($\nu < 1$ (Eq. 6)), we predict zero curvature. When accounting for variation in network size and volume scaling (Eq. 5) we predict positive (concave up) curvature (Figure 4d). These predictions are both in agreement with respiration-based studies of mammals$^{17}$. Furthermore, we predict that curvature decreases to zero with increasing network size, or generation $N$ in agreement with respiration-based studies of plants$^{18}$. These results can be informative for future studies connecting branching patterns and vascular data to ontogenetic- and size-based shifts in organism function. Such shifts are observed in growth and reproduction curves for tumors$^5$, plants$^{18}$, mammals$^{40}$, and fish$^{41}$.

We now demonstrate the importance of choosing theoretically-informed feature spaces over raw data to classify vascular organisms relating form to function. Classification using only raw data (branch radii and lengths) results only in size-based categorization, an approach that can distinguish between a mouse lung and a Balsa tree, but is not easily applicable to medical diagnostics between healthy and diseased tissues of near-equal size. Once networks are normalized for size, distributions of the raw data are greatly overlapped$^{9,11}$ (see Supplemental Information) and machine-learning methods applied to the raw data cannot distinguish the networks. Incorporating topological features—connectivity, loops, and branching angles—could enhance categorization methods because these features provide structural integrity and redundancy to damage in plant leaves and in capillaries$^{42,43}$. Furthermore, utilizing more robust applications of machine-learning methods and model complexity might help improve classification based on raw data but should improve classification using feature spaces based on theory as well. We thus conclude that
our theoretically-informed feature spaces are objectively superior at categorizing branching networks. In addition, these theoretically-informed feature spaces facilitate much easier translation into known biological principles and constraints on biologic function related to blood flow and metabolic rate.

In Extended Data Table 1 we show that the scaling of branch radius—$\bar{\beta}$ and $\Delta \beta$—are the strongest discriminants of vascular branching within and between mammal and plant networks. This finding suggests that hydrodynamic constraints and resource flow are the dominant drivers of biological branching networks. We also find that scaling of branch length—$\bar{\gamma}$ and $\Delta \gamma$—does little to differentiate vascular branching, possibly suggesting a broader architectural principle that is not specific to mammals or plants. Furthermore, we find variation in metabolic scaling exponents is primarily determined by variation in average scale factors ($\bar{\beta}$ and $\bar{\gamma}$), symmetric radial branching, and relative network size.

These results inform our understanding of the evolutionary pressures that determine convergence in organismal form and function. Furthermore, they provide a proof-of-principle that a mechanistically-based automatic classification and detection scheme for vascular networks could be used for medical diagnostics (e.g., tumor growth and stroke recovery). Such application would serve as a new dimension in radiomics and the emerging personalized healthcare paradigm, where classification based on vascular branching is wholly absent.
1. Brown, J. H. & West, G. B. (eds.) *Scaling in Biology* (Oxford University Press on Demand, 2000).

2. Roper, M., Simonin, A., Hickey, P. C., Leeder, A. & Glass, N. L. Nuclear dynamics in a fungal chimera. *Proceedings of the National Academy of Sciences* **110**, 12875–12880 (2013). URL https://www.pnas.org/content/110/32/12875. https://www.pnas.org/content/110/32/12875.full.pdf.

3. Fung, Y. C. *Biomechanics: Circulation* (Springer Science & Business Media, 1997).

4. Metzger, R. J., Klein, O. D., Martin, G. R. & Krasnow, M. A. The branching programme of mouse lung development. *Nature* **453**, 745–750 (2008).

5. Herman, A. B., Savage, V. M. & West, G. B. A quantitative theory of solid tumor growth, metabolic rate and vascularization. *PLoS One* **6**, e22973 (2011).

6. Prakash, R. *et al.* Vascularization pattern after ischemic stroke is different in control versus diabetic rats. *Stroke* (2013).

7. Bazmara, H. *et al.* The vital role of blood flow-induced proliferation and migration in capillary network formation in a multiscale model of angiogenesis. *PLOS ONE* **10**, 1–31 (2015). URL https://doi.org/10.1371/journal.pone.0128878.

8. Lambin, P. *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology* **14**, 749 (2017).
9. Bentley, L. P. *et al.* An empirical assessment of tree branching networks and implications for plant allometric scaling models. *Ecology Letters* **16**, 1069–1078 (2013).

10. Oliveras, I. *et al.* Changes in forest structure and composition after fire in tropical montane cloud forests near the andean treeline. *Plant Ecology & Diversity* **7**, 329–340 (2014).

11. Newberry, M. G., Ennis, D. B. & Savage, V. M. Testing foundations of biological scaling theory using automated measurements of vascular networks. *PLoS Computational Biology* **11**, e1004455 (2015).

12. Tekin, E., Hunt, D., Newberry, M. G. & Savage, V. M. Do vascular networks branch optimally or randomly across spatial scales? *PLoS Computational Biology* **12**, 1–28 (2016).

13. Brummer, A. B., Savage, V. M. & Enquist, B. J. A general model for metabolic scaling in self-similar asymmetric networks. *PLoS Computational Biology* **13**, 1–25 (2017).

14. Duong, T., Goud, B. & Schauer, K. Closed-form density-based framework for automatic detection of cellular morphology changes. *Proceedings of the National Academy of Sciences* **109**, 8382–8387 (2012).

15. Duong, T. Local significant differences from nonparametric two-sample tests. *Journal of Nonparametric Statistics* **25**, 635–645 (2013).

16. Savage, V. M. *et al.* The predominance of quarter-power scaling in biology. *Functional Ecology* **18**, 257–282 (2004).
17. Kolokotrones, T., Savage, V., Deeds, E. J. & Fontana, W. Curvature in metabolic scaling. *Nature* **464**, 753–6 (2010).

18. Mori, S. *et al.* Mixed-power scaling of whole-plant respiration from seedlings to giant trees. *Proceedings of the National Academy of Sciences* **107**, 1447–1451 (2010).

19. Price, C. A. *et al.* Testing the metabolic theory of ecology. *Ecology letters* **15**, 1465–1474 (2012).

20. Conn, A., Pedmale, U. V., Chory, J. & Navlakha, S. High-resolution laser scanning reveals plant architectures that reflect universal network design principles. *Cell Systems* **5**, 53 – 62.e3 (2017).

21. Smith, D. D. *et al.* Deviation from symmetrically self-similar branching in trees predicts altered hydraulics, mechanics, light interception and metabolic scaling. *New Phytologist* **201**, 217–229 (2014).

22. Savage, V. M., Deeds, E. J. & Fontana, W. Sizing up allometric scaling theory. *PLoS Computational Biology* **4**, e1000171 (2008).

23. Ronellenfitsch, H. & Katifori, E. Global optimization, local adaptation, and the role of growth in distribution networks. *Physical Review Letters* **117**, 138301 (2016).

24. 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).
25. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the structure and allometry of plant vascular systems. *Nature* **400**, 664–667 (1999).

26. Jain, R. K. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* **307**, 58–62 (2005).

27. Lau, A. *et al.* Quantifying branch architecture of tropical trees using terrestrial lidar and 3d modelling. *Trees* **32**, 1219–1231 (2018).

28. Olson, M. E. *et al.* Universal hydraulics of the flowering plants: vessel diameter scales with stem length across angiosperm lineages, habits and climates. *Ecology Letters* **17**, 988–997 (2014).

29. Reich, P. B. The world-wide ‘fast–slow’ plant economics spectrum: a traits manifesto. *Journal of Ecology* **102**, 275–301 (2014).

30. Savage, V. M. *et al.* Hydraulic trade-offs and space filling enable better predictions of vascular structure and function in plants. *Proceedings of the National Academy of Sciences* **107**, 22722–7 (2010).

31. Eloy, C., Fournier, M., Lacointe, A. & Moulia, B. Wind loads and competition for light sculpt trees into self-similar structures. *Nature Communications* **8**, 1014 (2017).

32. Moses, A. M. *Statistical Modeling and Machine Learning for Molecular Biology* (CRC Press, 2017).
33. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2016), 2nd edn.

34. Bouda, M., Brodersen, C. & Saiers, J. Whole root system water conductance responds to both axial and radial traits and network topology over natural range of trait variation. *Journal of Theoretical Biology* (2018).

35. Banavar, J. R. *et al.* A general basis for quarter-power scaling in animals. *Proceedings of the National Academy of Sciences* **107**, 15816–15820 (2010).

36. Dodds, P. S. Optimal form of branching supply and collection networks. *Physical Review Letters* **104**, 048702 (2010).

37. Hunt, D. & Savage, V. M. Asymmetries arising from the space-filling nature of vascular networks. *Physical Review E* **93**, 062305 (2016).

38. Barnsley, M. F. *Fractals Everywhere* (Academic press, 2012).

39. Huo, Y. & Kassab, G. S. Intraspecific scaling laws of vascular trees. *Journal of The Royal Society Interface* **9**, 190–200 (2012).

40. West, G. B., Brown, J. H. & Enquist, B. J. A general model for ontogenetic growth. *Nature* **413**, 628 (2001).

41. Barneche, D. R., Robertson, D. R., White, C. R. & Marshall, D. J. Fish reproductive-energy output increases disproportionately with body size. *Science* **360**, 642–645 (2018).
42. Mileyko, Y., Edelsbrunner, H., Price, C. A. & Weitz, J. S. Hierarchical ordering of reticular networks. *PLoS ONE* **7**, e36715 (2012).

43. Katifori, E., Szőllősi, G. J. & Magnasco, M. O. Damage and fluctuations induce loops in optimal transport networks. *Physical Review Letters* **104**, 048704 (2010).

**Acknowledgements**  *These authors contributed equally to this work.*

**Author Contributions**  A.B.B, B.J.E., and V.M.S. conceived the analysis, E.T., L.P.B., V.B., A.G., and I.O. acquired the data. A.B.B., P.L., and J.S., analyzed the data. A.B.B. and V.M.S. developed theory. A.B.B., B.J.E., and V.M.S. wrote the manuscript. All authors edited the manuscript.

**Competing Interests**  The authors declare that they have no competing financial interests.

**Correspondence**  Correspondence and requests for materials should be addressed to A.B.B. (email: abrummer@ucla.edu).

**Data Availability**  The data that support the findings of this study are available from the corresponding author upon reasonable request.
Extended Data Table 1 | Results of pairwise classification of plants and animal groups using scale factors that are related to hydraulics, and fluid transport for the cases of strictly symmetric branching ($\beta, \gamma$) or asymmetric branching ($\bar{\beta}, \Delta\beta$)

| SYM   | HHT* | Mouse Lung | Balsa   | Piñon   | Ponderosa* | GS Tips* | AS Tips* | Roots |
|-------|------|------------|---------|---------|------------|----------|----------|-------|
|       |      |            |         |         |            |          |          |       |
| HHT*  | -    | 2 $\times$ 10^{-1} | 6 $\times$ 10^{-13} | 1 $\times$ 10^{-4} | 3 $\times$ 10^{-2} | 9 $\times$ 10^{-9} | 1 $\times$ 10^{-1} | 2 $\times$ 10^{-2} |
| Mouse Lung | 6 $\times$ 10^{-2} | - | 5 $\times$ 10^{-18} | 7 $\times$ 10^{-3} | 7 $\times$ 10^{-5} | 2 $\times$ 10^{-19} | 2 $\times$ 10^{-2} | 2 $\times$ 10^{-6} |
| Balsa  | 1 $\times$ 10^{-1} | 4 $\times$ 10^{-4} | - | 1 $\times$ 10^{-16} | 6 $\times$ 10^{-2} | 1 $\times$ 10^{-6} | 6 $\times$ 10^{-2} | 6 $\times$ 10^{-13} |
| Piñon  | 3 $\times$ 10^{-7} | 3 $\times$ 10^{-6} | 2 $\times$ 10^{-7} | - | 4 $\times$ 10^{-13} | 3 $\times$ 10^{-56} | 2 $\times$ 10^{-5} | 2 $\times$ 10^{-13} |
| Ponderosa* | 2 $\times$ 10^{-1} | 7 $\times$ 10^{-2} | 2 $\times$ 10^{-3} | 4 $\times$ 10^{-4} | - | 9 $\times$ 10^{-5} | 3 $\times$ 10^{-1} | 1 $\times$ 10^{-4} |
| GS Tips* | 9 $\times$ 10^{-3} | 3 $\times$ 10^{-2} | 8 $\times$ 10^{-4} | 1 $\times$ 10^{-3} | 1 $\times$ 10^{-1} | - | 7 $\times$ 10^{-12} | 2 $\times$ 10^{-11} |
| AS Tips* | 2 $\times$ 10^{-1} | 1 $\times$ 10^{-1} | 7 $\times$ 10^{-3} | 2 $\times$ 10^{-2} | 4 $\times$ 10^{-1} | 3 $\times$ 10^{-2} | - | 5 $\times$ 10^{-3} |
| Roots  | 1 $\times$ 10^{-1} | 6 $\times$ 10^{-3} | 2 $\times$ 10^{-3} | 2 $\times$ 10^{-5} | 1 $\times$ 10^{-1} | 3 $\times$ 10^{-2} | 3 $\times$ 10^{-1} | - |

Table of $p$-values for pairwise KDE testing on plant and animal groups comparing different feature spaces. The lower diagonal corresponds to scale factors associated with strictly symmetric branching ($\beta, \gamma$), while the upper diagonal corresponds to scale factors associated with hydraulics and asymmetric radial branching ($\bar{\beta}, \Delta\beta$). Significant classification is defined as occurring when $p < 0.05$. Datasets with multiple individuals that were aggregated are indicated with asterisks (*). Of note is the universally enhanced ability of the radial scale factor feature space to differentiate between groups when compared to the symmetric scale factor feature space.
Extended Data Table 2 | Results of pairwise classification of plant and animal groups using scale factors that are related to both hydraulics and space-filling for the cases of asymmetric branching ($\bar{\gamma}, \Delta \gamma$) or the combined asymmetric feature space of ($\bar{\beta}, \Delta \beta, \bar{\gamma}, \Delta \gamma$).

| LENGTH | HHT* | Mouse Lung | Balsa | Piñon | Ponderosa* | GS Tips* | AS Tips* | Roots |
|--------|------|------------|-------|-------|------------|----------|----------|-------|
|        | –    | 7 x 10^{-2} | 8 x 10^{-1} | 6 x 10^{-3} | 1 x 10^{-1} | 1 x 10^{-2} | 3 x 10^{-2} | 8 x 10^{-3} |
| HHT*   | 1 x 10^{-1} | – | 1 x 10^{-1} | 2 x 10^{-3} | 4 x 10^{-1} | 4 x 10^{-1} | 9 x 10^{-2} | 9 x 10^{-6} |
| Mouse Lung | 9 x 10^{-5} | 2 x 10^{-5} | – | 2 x 10^{-2} | 1 x 10^{-2} | 1 x 10^{-3} | 1 x 10^{-2} | 3 x 10^{-2} |
| Piñon  | 4 x 10^{-11} | 4 x 10^{-8} | 1 x 10^{-8} | – | 3 x 10^{-1} | 6 x 10^{-2} | 5 x 10^{-1} | 3 x 10^{-1} |
| Ponderosa* | 8 x 10^{-3} | 9 x 10^{-3} | 7 x 10^{-3} | 6 x 10^{-5} | – | 3 x 10^{-1} | 5 x 10^{-1} | 1 x 10^{-1} |
| GS Tips* | 2 x 10^{-4} | 6 x 10^{-4} | 1 x 10^{-3} | 1 x 10^{-5} | 2 x 10^{-1} | – | 3 x 10^{-1} | 9 x 10^{-4} |
| AS Tips* | 5 x 10^{-2} | 2 x 10^{-2} | 1 x 10^{-1} | 1 x 10^{-2} | 4 x 10^{-1} | 1 x 10^{-1} | – | 3 x 10^{-1} |
| Roots  | 3 x 10^{-4} | 1 x 10^{-4} | 8 x 10^{-8} | 1 x 10^{-3} | 2 x 10^{-2} | 1 x 10^{-3} | 4 x 10^{-2} | – |

Table of $p$-values for pairwise KDE testing on plant and animal groups comparing different features. The lower diagonal corresponds to scale factors associated with asymmetric branching related to space-filling ($\bar{\gamma}, \Delta \gamma$) while the upper diagonal corresponds to scale factors associated with the combined asymmetric feature spaces related to both hydraulics and space-filling ($\bar{\beta}, \Delta \beta, \bar{\gamma}, \Delta \gamma$). Significant classification is defined as occurring when $p < 0.05$. Datasets with multiple individuals that were aggregated are indicated with asterisks (*). In this scenario we see that nearly all groups (excluding the AS Tips) are more differentiable when using all scale factors for the feature space when compared to the length scale factor feature space.
Extended Data Table 3 | Results of pairwise classification of plant and animal groups using scale factors that are related to hydraulics and space filling and that are associated with the central means of asymmetric branching ($\bar{\beta}$, $\bar{\gamma}$) or scale factors associated with variation in asymmetric branching ($\Delta\beta$, $\Delta\gamma$).

| AVE   | DIFF     | HHT* | Mouse Lung | Balsa | Piñon | Ponderosa* | GS Tips* | AS Tips* | Roots |
|-------|----------|------|------------|-------|-------|------------|----------|----------|-------|
| HHT*  |          | 5 × 10^{-3} | 1 × 10^{-1} | 3 × 10^{-1} | 4 × 10^{-2} | 3 × 10^{-1} | 4 × 10^{-2} | 2 × 10^{-1} | 4 × 10^{-2} |
| Mouse Lung | 2 × 10^{-1} | 4 × 10^{-3} | 3 × 10^{-2} | 1 × 10^{-2} | 2 × 10^{-1} | 4 × 10^{-2} | 2 × 10^{-1} | 7 × 10^{-3} | 7 × 10^{-3} |
| Balsa  | 3 × 10^{-2} | 8 × 10^{-7} | 3 × 10^{-6} | 2 × 10^{-4} | 2 × 10^{-1} | 8 × 10^{-2} | 2 × 10^{-1} | 8 × 10^{-2} | 1 × 10^{-5} |
| Piñon  | 4 × 10^{-2} | 1 × 10^{-2} | 3 × 10^{-6} | 1 × 10^{-4} | 1 × 10^{-1} | 8 × 10^{-2} | 2 × 10^{-1} | 8 × 10^{-2} | 1 × 10^{-5} |
| Ponderosa* | 3 × 10^{-2} | 1 × 10^{-3} | 3 × 10^{-2} | 1 × 10^{-4} | 1 × 10^{-1} | 6 × 10^{-1} | 5 × 10^{-1} | 9 × 10^{-2} | 1 × 10^{-2} |
| GS Tips* | 3 × 10^{-6} | 1 × 10^{-11} | 3 × 10^{-7} | 2 × 10^{-12} | 9 × 10^{-2} | 3 × 10^{-3} | 5 × 10^{-1} | 3 × 10^{-2} | 1 × 10^{-1} |
| AS Tips* | 1 × 10^{-1} | 1 × 10^{-2} | 9 × 10^{-2} | 6 × 10^{-2} | 3 × 10^{-1} | 4 × 10^{-3} | 1 × 10^{-1} | 1 × 10^{-2} | 1 × 10^{-1} |
| Roots  | 1 × 10^{-1} | 4 × 10^{-3} | 2 × 10^{-2} | 1 × 10^{-1} | 9 × 10^{-2} | 1 × 10^{-3} | 4 × 10^{-1} | 1 × 10^{-1} | 1 × 10^{-1} |

Table of $p$-values for pairwise KDE testing on plant and animal groups comparing feature spaces. The lower diagonal corresponds to scale factors associated with the central means of asymmetric branching ($\bar{\beta}$, $\bar{\gamma}$) while the upper diagonal corresponds to scale factors associated with variation in asymmetric branching ($\Delta\beta$, $\Delta\gamma$). Significant classification is defined as occurring when $p < 0.05$. Datasets with multiple individuals that are aggregated are indicated with asterisks (*). Of note is the enhanced differentiation that occurs when using the average scale factors as the feature space— in particular for the Mouse lung, Balsa, Piñon, Ponderosas, and GS Tips.Interestingly, the Roots exhibit greater differentiation between all other groups using the difference scale factors as the feature space.
Extended Data Table 4 | Results of pairwise classification of Ponderosa individuals using scale factors that are related to hydraulics, and fluid transport ($\bar{\beta}, \Delta \beta$) and that are associated with strictly symmetric branching ($\beta, \gamma$) or asymmetric branching.

|        | ASYM   | POND03 | POND05 | POND06 | POND07 | POND16 |
|--------|--------|--------|--------|--------|--------|--------|
| SYM    |        |        |        |        |        |        |
| POND03 | -      | $5 \times 10^{-2}$ | $5 \times 10^{-1}$ | $7 \times 10^{-1}$ | $3 \times 10^{-1}$ |        |
| POND05 | $5 \times 10^{-1}$ | -      | $3 \times 10^{-2}$ | $8 \times 10^{-2}$ | $3 \times 10^{-1}$ |        |
| POND06 | $3 \times 10^{-1}$ | $4 \times 10^{-1}$ | -      | $4 \times 10^{-1}$ | $3 \times 10^{-1}$ |        |
| POND07 | $6 \times 10^{-1}$ | $4 \times 10^{-1}$ | $5 \times 10^{-1}$ | -      | $6 \times 10^{-1}$ |        |
| POND16 | $2 \times 10^{-1}$ | $2 \times 10^{-1}$ | $2 \times 10^{-1}$ | $2 \times 10^{-1}$ | -      |        |

Table of $p$-values for pairwise KDE testing on the Ponderosa individuals comparing different feature spaces. The lower diagonal corresponds to scale factors associated with strictly symmetric branching ($\beta, \gamma$), while the upper diagonal corresponds to scale factors associated with hydraulics and asymmetric radial branching ($\bar{\beta}, \Delta \beta$). Significant classification is defined as occurring when $p < 0.05$. Note that no single pairwise comparison results in differentiation between Ponderosa individuals as per the $p < 0.01$ threshold.
Extended Data Table 5 | Results of pairwise classification of the human head and torso individuals using scale factors that are related to hydraulics and fluid transport and that are associated with strictly symmetric branching ($\beta, \gamma$) or asymmetric branching ($\bar{\beta}, \Delta\beta$).

| SYM  | HHT01 | HHT02 | HHT03 | HHT04 | HHT05 | HHT06 | HHT07 | HHT08 | HHT09 | HHT10 | HHT11 | HHT12 | HHT13 | HHT14 | HHT15 | HHT16 | HHT17 | HHT18 |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| HHT01| 6 x 10^{-5} | 7 x 10^{-5} | 6 x 10^{-5} | 4 x 10^{-4} | 4 x 10^{-4} | 3 x 10^{-4} | 5 x 10^{-3} | 5 x 10^{-3} | 6 x 10^{-3} | 6 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 2 x 10^{-2} | 6 x 10^{-3} | 6 x 10^{-3} | 6 x 10^{-3} |
| HHT02| 3 x 10^{-4} | 7 x 10^{-5} | 7 x 10^{-5} | 7 x 10^{-4} | 3 x 10^{-4} | 6 x 10^{-3} | 7 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 2 x 10^{-2} | 2 x 10^{-3} | 6 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} |
| HHT03| 5 x 10^{-3} | 3 x 10^{-4} | 8 x 10^{-4} | 6 x 10^{-3} | 3 x 10^{-4} | 5 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 1 x 10^{-1} | 6 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} | 7 x 10^{-3} |
| HHT04| 6 x 10^{-4} | 4 x 10^{-3} | 4 x 10^{-3} | 3 x 10^{-3} | 4 x 10^{-3} | 3 x 10^{-3} | 7 x 10^{-3} | 6 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 5 x 10^{-3} | 1 x 10^{-1} | 6 x 10^{-3} | 6 x 10^{-3} | 6 x 10^{-3} | 6 x 10^{-3} | 6 x 10^{-3} |
| HHT05| 4 x 10^{-5} | 6 x 10^{-5} | 3 x 10^{-5} | 2 x 10^{-5} | 2 x 10^{-5} | 1 x 10^{-5} | 6 x 10^{-5} | 2 x 10^{-5} | 2 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} | 1 x 10^{-5} |
| HHT06| 7 x 10^{-6} | 3 x 10^{-5} | 3 x 10^{-5} | 2 x 10^{-5} | 2 x 10^{-5} | 5 x 10^{-5} | 3 x 10^{-5} | 4 x 10^{-5} | 4 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} | 3 x 10^{-5} |
| HHT07| 3 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} | 2 x 10^{-6} |
| HHT08| 6 x 10^{-6} | 4 x 10^{-6} | 4 x 10^{-6} | 4 x 10^{-6} | 4 x 10^{-6} | 4 x 10^{-6} | 4 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} | 3 x 10^{-6} |

Table of $p$-values for pairwise KDE testing on the human head and torso individuals comparing different feature spaces. The lower diagonal corresponds to scale factors that are related to hydraulics and space filling and that are associated with strictly symmetric branching ($\beta, \gamma$) while the upper diagonal corresponds to scale factors associated with just hydraulics and asymmetric radial branching ($\bar{\beta}, \Delta\beta$). Significant classification is defined as occurring when $p < 0.05$. Note that only pairwise comparisons between HHT individuals 13 and 7, 11 and 5, and 7 and 5 result in differentiation between HHT individuals as per the $p < 0.01$ threshold. All other pairwise comparisons are indistinguishable.
Extended Data Table 6 | Results of pairwise classification of tree tips species using scale factors associated with strictly symmetric branching ($\beta, \gamma$), or asymmetric branching, hydraulics, and fluid transport ($\tilde{\beta}, \Delta \beta$).

|       | ASYM | Maple | Oak   | Robinia | Whitefir | Dougfir | Whitepine |
|-------|------|-------|-------|---------|----------|---------|-----------|
| SYM   |      |       |       |         |          |         |           |
| Maple | –    | $3 \times 10^{-2}$ | $7 \times 10^{-1}$ | $4 \times 10^{-10}$ | $1 \times 10^{-2}$ | $8 \times 10^{-3}$ |
| Oak   | $4 \times 10^{-1}$ | –     | $3 \times 10^{-2}$ | $9 \times 10^{-6}$ | $4 \times 10^{-3}$ | $1 \times 10^{-1}$ |
| Robinia | $5 \times 10^{-1}$ | $2 \times 10^{-1}$ | –     | $3 \times 10^{-6}$ | $4 \times 10^{-2}$ | $3 \times 10^{-2}$ |
| Whitefir | $3 \times 10^{-1}$ | $1 \times 10^{-1}$ | $4 \times 10^{-1}$ | –     | $2 \times 10^{-1}$ | $3 \times 10^{-1}$ |
| Dougfir | $4 \times 10^{-1}$ | $2 \times 10^{-1}$ | $5 \times 10^{-1}$ | $5 \times 10^{-1}$ | –     | $1 \times 10^{-1}$ |
| Whitepine | $5 \times 10^{-1}$ | $3 \times 10^{-1}$ | $6 \times 10^{-1}$ | $6 \times 10^{-1}$ | $6 \times 10^{-1}$ | –     |

Table of $p$-values for pairwise KDE testing on tree tips across species comparing different feature spaces. The lower diagonal corresponds to scale factors that are related to hydraulics and space filling and that are associated with strictly symmetric branching ($\beta, \gamma$) while the upper diagonal corresponds to scale factors associated with just hydraulics and asymmetric radial branching ($\tilde{\beta}, \Delta \beta$). Significant classification is defined as occurring when $p < 0.05$. All species had multiple individuals that were aggregated due to low branch number per individual. Of note is the clustered differentiation between Angiosperm and Gymnosperm species using the radial asymmetric scale factors, specifically the Whitefir with all Angiosperms, the Dougfir with the Oak, and the Whitepine with the Maple.
Supplementary Information

Contents

A Classifier methods 2

B Feature space selection 3
   B.1 Raw radius and length measurements 3
   B.2 Scale factor feature spaces 4

C Normalizing results from classifier methods 5
   C.1 Kernel density estimation 6
   C.2 Logistic regression and support vector machine 7
   C.3 ROC comparison of methods 8
   C.4 Data grouping 10

D Derivation of exact metabolic scaling exponent formula 10

E Derivation of curvature in metabolic rate versus mass 14

List of Tables

| 1 | Global scores for machine learning methods and feature spaces.          | 6 |

List of Figures

| 1 | Mammal and plant diameters versus lengths, log-log scale              | 4 |
| 2 | Mammal and plant diameters versus lengths, standardized scales       | 5 |
| 3 | Principle component analysis of asymmetric scale factors            | 7 |
| 4 | Data classification versus sensitivity for KDE, LR, and SVM methods  | 8 |
| 5 | Receiver operating characteristic curve comparing KDE, LM, and SVM methods | 9 |
| 6 | Sample regression analysis on mouse lung data                       | 12 |
| 7 | Exact formula for metabolic scaling exponent versus volume scaling and network size | 13 |
A Classifier methods

To generate probability models using the support vector machine and logistic regression methods, approximately 75% of each combined plant and and combined mammal dataset was taken as the training sample and the remaining 25% used as the testing sample. This was done after randomization within respective datasets to minimize the chances of accidentally removing an entire individual or species. Then, using the radial scale factor feature space ($\bar{\beta}$ and $\Delta \beta$) the two groups of mammals and plants were compared using three methods: a support vector machine (SVM) model and a logistic regression (LR) model, and a non-parametric kernel density estimator (KDE) method.

The support vector machine method, a supervised machine learning algorithm for classification problems, plots data points in an n-dimensional space and draws a decision boundary by maximizing the margin between points from different classes. To generate a support vector machine model for our datasets, we used the SVC (support vector classification) function from the Python scikit-learn package. When running SVC, we used the polynomial kernel with a degree of 2.

The LR method generates probabilities of how likely certain data points are going to fall under a certain class, based on the logit function $\frac{1}{1 + e^{-x}}$. To produce a non-linear LR model for our two-dimensional datasets, we used the Python scikit-learn logistic regression function and added three nonlinearity columns, $x_1^2$, $x_2^2$, and $x_1 \times x_2$. All pairwise combinations were run on this LR model and the probabilities were recorded.

For both the SVM method and the LR method, a polynomial kernel was used. Through a series of tests between different kernels, including linear and radial basis function, the polynomial kernel yielded the highest accuracy score when classifying the testing data. The LR and SVM methods differ in their nonlinearities, where the LR method uses an added nonlinearity term and the SVM method uses a radial basis function kernel. The probabilities for the LR method are assigned using the logit function, and the probabilities assigned using the SVM method are based on the point distances from the decision boundary. For both of these methods, testing points are classified based on the value of their score. When using training data that is equally split between the two categories then points that receive scores of 0.5 or greater are classified as one group, and scores below 0.5 are classified as the other group.

The kernel density estimator (KDE) technique introduced by Duong et al. [1, 2] generates non-parametric, multi-dimensional probability distributions, $P_i(x)$, of vascular traits, $x$, for each testing group, $i = A, B$. These distributions are then compared against one another for their extent of uniqueness, or non-overlap, represented by the test statistic $T = \int [P_A(x) - P_B(x)]^2 dx$. This test can be thought of as a multi-dimensional generalization of the two-sample Kolmogorov-Smirnov (KS) test, where significance of classification is conventionally communicated through p-values. We use the nominal threshold of $p \leq 0.05$ as a threshold for significance when using the KDE method.

The KDE method can be applied both globally [1] and locally [2]. At the global level, the test statistic $T$ is calculated as described above (and in more detail in [1]), and converted to a p-value using standard tables. While $p = 0.05$ is used as the nominal threshold for significance, it should be noted that $p$ values ranging orders of magnitude in size from $p = 10 \times 10^{-2}$ to $p = 10 \times 10^{-20}$ are observed, and thus interpreted as relative levels of significance. The local application of the KDE method is effectively an inversion of the calculation for the test statistic $T$. Here, one sets the desired threshold for significance, or the value of the test statistic $T$, then calculates which regions in the vascular trait-space, $x$ correspond to the chosen test statistic.
B  Feature space selection

Here we describe our methods for selecting the feature space. We tested each machine learning method on a variety of vascular trait feature spaces. Specifically, we tested the feature spaces of: raw diameter and length measurements; the symmetric WBE diameter and length scale factors, $\beta$ and $\gamma$; and five combinations of the asymmetric scale factors. The five combinations of asymmetric scale factors were: the average scale factors $\bar{\beta}$ and $\bar{\gamma}$; the difference scale factors $\Delta \beta$ and $\Delta \gamma$; the radial scale factors $\bar{\beta}$ and $\Delta \beta$; the length scale factors $\bar{\gamma}$ and $\Delta \gamma$; and all four asymmetric scale factors $\bar{\beta}, \bar{\gamma}, \Delta \beta,$ and $\Delta \gamma$. These results are presented in Table S1. We also conducted a principle components analysis (PCA) on the asymmetric scale factors to identify which combinations of scale factors explain the most variance in the datasets. Performing a PCA on all eight vascular variables is non-trivial due to the non-random presence of empty-cells in the dataset [3].

It is an open question as to which combinations of the scale factors will work best as a feature space for discriminating vascular networks. If dynamics of blood flow dominate the formation and evolution of vascular architecture, the scale factors involving vessel radius would be expected to be most informative because vascular theory and empirical evidence show blood flow is most strongly determined by vessel radius. If the space-filling constraints and body plan of the organism primarily determine vascular architecture, the scale factors for vessel lengths should best discriminate. Moreover, average properties might be fixed across species and differences might reflect specific selective pressure, in which case the difference scale factors may be most effective at discriminating. Alternatively, if selective pressures are different enough, even the average or symmetric scale factors may exhibit explanatory differences in vascular networks. Finally, if length scale factors are more important, the fact that observed vessel lengths exhibit more asymmetry in vascular branching also suggests that difference scale factors may be more important. Conversely, if radius scale factors are more important, the high symmetry of observed vessel radii branching may mean that average scale factors are more important.

B.1  Raw radius and length measurements

Prior to transforming the radial and length measurements of our vascular datasets as motivated by scaling theory [4, 5, 6], we applied all three of the logistic regression (LR), support vector machine (SVM), and kernel density estimator (KDE) methods on the raw, untransformed data. This was done in two ways, first on the data after all measurements were converted to meters (Figure S1), then again after performing a standardized transformation by translating each species- or group-level distribution to be centered about zero, and then normalizing by the respective standard deviations (Figure S2).

Classifying vascular data based on metric size is both trivial and uninformative. Small networks (mouse lung) are clearly distinct from large networks (whole trees), and varying degrees of overlap will exist in the intermediate range of all other networks (Figure S1). All three methods yield significant global classification scores, as demonstrated by the example scores of LR: 82%, SVM: 88%, and KDE: $p \approx 0$ found in Table S1. One can certainly argue that classifying networks in this manner is possible, even though it is simply demonstrating that networks of differing size are distinguishable. However, this approach does not provide an obvious path toward understanding at a mechanistic level why certain patterns are observed beyond the simple size-based classification. Furthermore, when applying these methods for the purposes of disease detection one is oftentimes examining healthy and diseased tissues that are of similar size classes. Thus, the utility of size-based classification is rendered moot, and we must transform.

The most common transformation is to standardize the data such that the distribution means are centered about the origin and to normalize by the variance [7, 8] (Figure S2). While this approach has the benefit of removing the global size-based hierarchy between the networks, it fails to address the common pattern of
Figure S1: Graph of lengths versus diameters for all data analyzed, scaled logarithmically. Note the clear grouping present due to physical size differences between mammal and plant data. The cluster in the lower left corresponds to the mouse lung data, while the cluster in the upper right corresponds to the whole tree data of Balsa, Piñon, and Ponderosa. Some overlap occurs between the larger HHT vessels and the smaller plant branches.

local size-based hierarchy that is commonly found within networks [4, 5, 9]. Specifically, the abundance-size distribution of vessels in a vascular network is approximately exponential due to the fact that the vessels, on average, decrease in size at every bifurcation. Thus, classification becomes immediately obscured, as demonstrated by the noticeably decreased example scores of LR: 54%, SVM: 56%, and KDE: $p \approx 0.002$ in Table S1. The fact that the KDE method still yielded a significant score is a consequence of this method excelling at detecting multimodality between distributions. Yet, even though the KDE method does yield a significant score, we still have the original problem of how to interpret the results, only now it is further compounded after having performed the standardization transformation. It is possible to consider alternative transformations based on mechanical principles—flow-rates and pressures [10]—yet the problem of the hierarchy of sizes returns, only now with different physical units. As a consequence of these complications in analyzing raw data, we turn to classification based on scale factor variables as guided by the metabolic scaling theory literature.

B.2 Scale factor feature spaces

To investigate the scale factor feature spaces, all three machine learning methods were tested on different combinations of candidate feature spaces, and a principle component analysis (PCA) was conducted on the asymmetric scale factor feature space. Results from the application of different machine learning methods are presented in Table S1 and graphs from the PCA are presented in Figure S3. Table S1 shows that the top
two combinations of features for classifying animal and plant vascular networks are the set of all asymmetric scale factors \((\bar{\beta}, \Delta \bar{\beta}, \bar{\gamma}, \Delta \bar{\gamma})\) and the radial scale factors \((\bar{\beta}, \Delta \bar{\beta})\). That the set of all asymmetric scale factors performs the best is not surprising as this is the most comprehensive set of data for the vascular networks considered. However, it does not help to identify which subset of variables are primarily responsible for classification over other variables. To do that, we focus our attention to the PCA results.

A common feature of vascular network data is the large variation observed in the scaling of branch lengths [5, 4, 9], often times exhibited over many orders of magnitude. In Figure S3 we see that the first principle component explains 48.4% of the variance in the data, and is composed primarily of the length scale factors \((\bar{\gamma}, \Delta \bar{\gamma})\). Interestingly, the length scale factors are not powerful for classification purposes (see Table S1 and Figure 3 in the main text), even though they are responsible for much of the variance in the data. On the other hand, we find that correlations between the radial scale factors account for 26.8% and 18.8% of the variance through the second and third principle components. This result, in combination with the fact that the radial scale factors had the second best global classification scores, led to the selection of the radial average and difference scale factors as the final choice for the feature space.

C Normalizing results from classifier methods

While the three machine learning methods tested effectively perform the same task (e.g. classification of data), the manner in which this is done varies significantly, and thus requires careful consideration when trying to compare results. This difficulty is compounded by the inclusion of a variable classification sensitivity level. Here we describe the process used to standardize classification output from the three methods of Kernel
Table S1 | Global scores for machine learning methods and feature spaces for classification of mammal plant datasets.

|            | (Radius, Length) | (Radius, Length)* | (β, γ) | (β, γ) | (∆β, ∆γ) | (∆β, ∆γ) | (∆β, ∆γ) | (∆β, ∆γ, ∆β, ∆γ) |
|------------|------------------|-------------------|--------|--------|----------|----------|----------|-----------------|
| LogReg     | 0.82             | 0.54              | 0.52   | 0.59   | 0.55     | 0.59     | 0.53     | 0.58            |
| SVM        | 0.88             | 0.56              | 0.57   | 0.62   | 0.59     | 0.64     | 0.58     | 0.67            |
| KDE        | 0                | 0.002             | 0.018  | 0.024  | 0.088    | 0.00011  | 0.0049   | 0.0000058       |

The Logistic Regression and Support Vector Machine scores represent the ratio of correctly classified vessels/nodes for a given feature space, and are compared to a baseline of 0.52 (as determined by the starting ratio of mammal to plant data). The Kernel Density Estimation scores are $p$-values representing level of significance in differentiating mammal from plant networks, with $p = 0.05$ being defined as the nominal level of significance. * indicates the standardized radius and length distribution.

Density Estimation (KDE), Support Vector Machine (SVM), and Logistic Regression (LR). Only one variable feature space was used for the comparison between methods, and that was the radial scale factor feature space ($\bar{\beta}, \Delta \beta$). This feature space was chosen as it performed best for classification across all methods (Table S1, as well as it having explain up to 55% of the variance in all of the asymmetric scale factor variables (Figure S3). The testing groups used were that of mammal and plant. These were chosen to increase the performance of methods reliant on dataset size for training and testing. Finally, we present receiver operating characteristic (ROC) curves that are used to perform the final comparison once standardization has been achieved.

C.1 Kernel density estimation

The non-parametric kernel density estimation procedure put forth by Duong et al. [1, 2] tests for uniqueness and overlap between two different probability distributions generated from empirical data. Probability distributions, $P_i(x, X)$ are defined on a discretized feature space, where $x$ represents the discretized coordinates in the feature space, $X$ represents actual empirically measured values, and $i = A$ or $B$ represents our known classifier. When performing the local significant difference test, one must first set the $p$-value that denotes “significance”. Conventionally this value is set at either $p = 0.01$ (as done for the analysis presented in the main text) or $p = 0.05$. The KDE method then identifies regions of the feature space where the squared-difference between the probabilities, $P_i(x, X)$, is less than the selected significance level. Finally, by examining which probability, $P_i(x, X)$, is greatest within each region, one can identify which classifier is driving differentiation. Thus, one can measure the performance of the KDE method at a given significance level by counting the number of correctly and incorrectly classified points with respect to each region (Figure S4).

By varying the $p$-value one can vary the relative size of the classified regions of the test to examine the efficacy of classification (Figure S4). For the KDE method, the $p$-value was varied by integer orders of magnitude from $10^{-12}$ to $10^4$ to ensure that the limiting scenarios of zero points classified and all points classified were contained. Then, for each sensitivity level, correctly and incorrectly classified points were tallied for further analysis and comparison.
Figure S 3: Principle components analysis of asymmetric scale factors \((\bar{\beta}, \Delta\beta, \bar{\gamma}, \Delta\gamma)\). Arrows indicate loadings for the different scale factors with respect to the four principle components. Ellipses correspond to 95% concentration contours for the data.

C.2 Logistic regression and support vector machine

The logistic regression (LR) and support vector machine (SVM) methods represent two approaches at using supervised machine learning methods for classification [7, 11]. These methods differ from the KDE approach by using a training set of data to partition the feature space into two separable regions (separated by the decision boundary). Then, points from a testing set of data are assigned a classification score based on their positions in the feature space with respect to the decision boundary. In Figure S4e-h are graphs of results from the LR method, and in Figure S4i-l are graphs of results from the SVM method. In these examples, the decision boundary corresponds to a probability score of 0.5, and any testing point with a score between 0 and 0.5 is classified as a plant, while a score between 0.5 and 1 is classified as a mammal.

To compare to the output of the KDE approach, an analogue of significance to the KDE approach must be defined for the LR and SVM approaches. In the context of the LR and SVM approaches this was done by defining significance as the distance a probability score is from 0.5. Regions of equal significance were determined by binning along the probability score axis. Thus, points with probability scores in the ranges of \([0, 0.05]\) for mammals or \([0.95, 1]\) for plants would all be characterized as equally, highly significant predictions (or in terms of the KDE method, would correspond to the next lowest possible \(p\)-value). The bins were then successively enlarged to reflect a decrease in test significance. So, for mammals, the bins used were \([0, 0.05], [0, 0.1], [0, 0.15], [0, 0.2], [0, 0.25], [0, 0.3], [0, 0.35], [0, 0.4], [0, 0.45], [0, 0.5]\). For plants, the bins used were \([0.95, 1], [0.9, 1], [0.85, 1], [0.8, 1], [0.75, 1], [0.7, 1], [0.65, 1], [0.6, 1], [0.55, 1], [0.5, 1]\). In Figure S4 are graphs of results from the LR and SVM methods showing the successive binning approach. Finally, for each binned region corresponding to varying levels of significance, correctly and incorrectly classified points were identified for comparison.
C.3 ROC comparison of methods

To finally compare the three methods of kernel density estimation (KDE), logistic regression (LR), and support vector machine (SVM), we graphed receiver operating characteristic (ROC) curves of the results of the methods as sensitivities were varied [7, 11]. ROC curves are graphs of a methods true positive rate (TPR) versus its false positive rate (FPR). In calculating the TPR and FPR, one must first choose which category will represent a true positive. Should one choose the plant group for this role, then the TPR and FPR can be calculated as,
\[ TPR = \frac{N_{\text{plant}^+}}{N_{\text{plant}^+} + N_{\text{mammal}^-}} \]

\[ FPR = \frac{N_{\text{plant}^-}}{N_{\text{plant}^-} + N_{\text{mammal}^+}} \]

where \( N_{\text{plant}^\pm} \) corresponds to the number of data points correctly (or incorrectly) classified as plant as denoted by the + sign (− sign), and \( N_{\text{mammal}^\pm} \) corresponds to the number of data points correctly (or incorrectly) classified as mammal. A TPR and an FPR is then calculated for each level of significance (defined by a p-value for the KDE approach, or a probability bin for the LR or SVM approaches). Finally, the ROC curve can be graphed, as presented in Figure S5a for the case of identifying true positives as plants.

![Figure S5: Receiver Operating Characteristic curve comparing KDE, LM, and SVM methods. a is plant positive, b is mammal positive, and c is mixed.](image)

Conventionally, when comparing classification schemes using an ROC curve, the best method is identified as whichever method sits most in the upper-left corner of the graph, as this represents a maximal true positive rate and a minimal false positive rate. However, a seemingly peculiar feature is present in Figure S5a. The points corresponding to the highest significance levels (smallest diameter points) for the LR and SVM methods begin at the lower left corner of the graph. This is to be expected since the fewest classifications have been made for such strict values of significance, and thus the TPR and FPR are naturally low. Typically, as significance is decreased, or as more classifications occur, the TPR and FPR begin to both increase (although one hopes that the TPR increases at a rate greater than that of the FPR). However, the points corresponding to the highest significance of the KDE method are in the upper right corner. This is a result of having chosen plants as representing true positives, and combined with the fact that the KDE method does not classify any mammal data points at its starting significance values, but only plant data points. As can be seen by the structure of Eq. 1, if \( N_{\text{mammal}^\pm} \approx 0 \), then both the TPR and FPR are going to be very near to 1.

This effect can be validated by redefining true positives as mammals. In this case, Eq. 1 takes the form of,
TPR = \frac{N_{\text{mammal}+}}{N_{\text{mammal}+} + N_{\text{plant}−}} \\
FPR = \frac{N_{\text{mammal}−}}{N_{\text{mammal}−} + N_{\text{plant}+}} \tag{2}

and we can expect to have “corrected” the presentation of the KDE ROC curve, as presented in Figure S5b. However, we can see that not only as the KDE ROC curve been inverted about the $y = −x$ axis, but so too have the LR and SVM ROC curves. This latter effect suggests that the LR and SVM methods do not classify any plant data points at their starting significance values, but only mammal points. As before, we can verify this effect by examining the structure of Eq. 2, and seeing that if $N_{\text{plant}±} \approx 0$, then both the TPR and FPR are going to be very near to 1. What we learn from these tradeoffs is that the KDE method is better suited for classifying the plant networks, where as the LR and SVM methods are best for classifying the mammal networks. This result is due to the fact that the KDE method performs well at discerning internal distribution structure, whereas the LR and SVM methods are designed to slice the feature space in two, thus being driven more by outlier location in this instance.

To standardize the ROC curves for the different methods of KDE, LR, and SVM, we can use Eq. 1 for the LR and SVM methods, and Eq. 2 for the KDE method, resulting in Figure S5c. Here we can observe that overall the KDE method outperforms the LR and SVM methods as indicated by the multiple values located near TPR $\approx 0.85$ and FPR $\approx 0.15$.

C.4 Data grouping

Once selection of the classifier was made, the subgroups of datasets were prepared for higher resolution classification. Using the KDE method, Individual HHT and ponderosas were found indistinguishable, as were species within the GS Tips and AS Tips groups, hence their final merging into larger datasets (see Extended Data Tables 4-6. We obtained 8 different major species/groups: HHT, mouse lung, Balsa, Piñon, Ponderosa, GS Tips, AS Tips, and Roots, each with 6 recorded variables: $\bar{\beta}$, $\bar{\gamma}$, $\Delta\beta$, $\Delta\gamma$, and the merging of $\beta_1$ and $\beta_2$ into a single distribution as well as for $\gamma_1$ and $\gamma_2$. Specific groupings of the scale factor variables used were: $(\beta_1, \beta_2; \gamma_1, \gamma_2)$ as a two-dimensional distribution representative of the symmetric formalism; $(\bar{\beta}, \bar{\gamma}, \Delta\beta, \Delta\gamma)$ as a four-dimensional distribution representative of the asymmetric formalism; $(\bar{\beta}, \bar{\gamma})$ as a two-dimensional distribution for average scaling; $(\Delta\beta, \Delta\gamma)$ as a two-dimensional distribution for difference scaling; $(\bar{\beta}, \Delta\beta)$ as a two-dimensional distribution for radial scaling; and $(\bar{\gamma}, \Delta\gamma)$ as a two-dimensional distribution for length scaling. For the full list of comparison results using the KDE method, see Extended Data Tables 1-3.

D Derivation of exact metabolic scaling exponent formula

Here we present a derivation of the metabolic scaling exponent under the general assumptions of the West, Brown, Enquist model for vascular branching [12, 13, 6]. This approach deviates from conventional derivations in that zero approximations will be made regarding network size or the particulars of volume scaling outside of the bounds of self-similarity. Some discussion will be presented at the end regarding how network geometry can be included in terms of symmetric or asymmetric branching.

The typical starting point for modeling metabolic scaling is Kleiber’s Law [14], the empirically motivated, power-law relationship between organismal basal metabolic rate and mass, expressed as,
Here $B$ represents the measured metabolic rate, $M$ the mass, $\theta$ the metabolic scaling exponent observed to cluster around $3/4$ [14, 15], and $B_0$ and $M_0$ normalization constants. Treating metabolic rate as the combined sum of metabolism of every terminal branch in an organism, we can substitute $B = B_{\text{tip}} N_{\text{tips}}$, where $B_{\text{tip}}$ represents total metabolism per terminal branch and $N_{\text{tips}}$ the total number of metabolizing terminal branches. Doing so results in,

$$B_{\text{tip}} N_{\text{tips}} = B_0 \left( \frac{M}{M_0} \right)^\theta$$

Next, mass is substituted with total volume. The validity of this substitution is the result of having optimized the geometrical scaling of a hierarchically branching vascular network against the dual demands of hydrodynamic resistance to fluid flow and fractal space-filling [12, 13, 6]. Performing the substitution gives us,

$$N_{\text{tips}} = \left( \frac{V_{\text{tot}}}{V_0} \right)^\theta$$

where we have cancelled out $B_{\text{tip}}$ with $B_0$. At this point it is important to pause and recognize Eq. 5 as a method by which one can estimate the metabolic scaling exponent in a vascular organism free from explicitly imposing assumptions regarding network geometry via symmetric or asymmetric branching. One powerful aspect of Eq. 5 is that it can be applied recursively on any individual vascular branching network. Specifically, for every branch (or vessel) in a network, the total number of distal terminal branches can be represented by $N_{\text{tips}}$, and the total volume of all distal branches are represented by $V_{\text{tot}}$. Thus, a standard major axis regression analysis can be performed to determine the value of $\theta$ that corresponds to an individual organism as per this model. An example of such an analysis being performed on the mouse lung data set is presented in Figure S6.

It is at this point that we must specify in greater detail the functional form of the total volume of the vascular network, $V_{\text{tot}}$. Assuming that we are working with a strictly self-similar, hierarchically branching, pipe-like model, then the total volume of the network can be expressed as

$$V_{\text{tot}} = V_{N,\text{tot}} \sum_{j=0}^{N} \nu^{-j}$$

where $j$ and $N$ represent the $j^{th}$ and $N^{th}$ generations of the network, $V_{N,\text{tot}}$ is the total volume of the terminal ($N^{th}$) generation, and $\nu$ represents the ratio of total volume from sibling branches to their respective parent branch. For example, in a bifurcating symmetric network, $\nu = 2(\pi r_j^2 l_{j+1})/(\pi r_j^2 l_j) = 2\beta^2 \gamma$. Recognizing Eq. 6 as a geometric series, we can write it’s exact form as,

$$V_{\text{tot}} = V_{N,\text{tot}} \frac{1 - \nu^{-(N+1)}}{1 - \nu^{-1}}$$

which is valid for all values of $\nu$ except for $\nu \approx 1$. It is not uncommon to find individual organisms with $\nu \approx 1$. This scenario is problematic if using the above formula due to its asymptotic nature. However, this problem can be remedied by using L’Hôpital’s rule, producing the piecewise result,
Upon substituting Eq. 8 into Eq. 5, we arrive at the following exact, piecewise function for the metabolic scaling exponent,

$$\theta = \begin{cases} 
\frac{\ln(2^N)}{\ln(2^N) + \ln(1 - \nu^{N+1}) - \ln(\nu^N(1 - \nu))} & \text{for } \nu \neq 1 \\
\ln((N+1)2^N) - N\ln(\nu) & \text{for } \nu \approx 1 
\end{cases}$$

(9)

where we set $V_{N,\text{TOT}} = N_{\text{TIPS}}V_{\text{TIP}}$, $V_0$ was cancelled out with $V_{\text{TIP}}$, and we have restricted ourselves to considering strictly bifurcating networks such that $N_{\text{TIPS}} = 2^N$. One benefit of Eq. 9 is that we can easily examine the functional relationship between metabolic rate and the scaling of volume in a general sense, as presented in Figure S7. Examining the influence of symmetric or asymmetric branching can then be done separately.

Several important features are present in Figure S7. Most notably is the effect that network size, $N$, has on varying the sensitivity of the metabolic scaling exponent to the value of the volume scale factor, $\nu$. For all values of network size, the metabolic scaling exponent is bounded between zero and one. For the case of large networks ($N = 100$ in Figure S7), the metabolic scaling exponent nearly reaches the asymptotic...
value of $\theta = 1$ when $\nu = 1$, where as for smaller size networks (lesser values of $N$), the rate of increase in $\theta$ is markedly less. This result lends support to the argument that the approximate form of Eq. 9, namely $\theta \approx \frac{\ln(2)}{\ln(2) - \ln(2^{\beta_2})}$ for a symmetrically branching network, holds in the limit that $N \gg 1$ and $\nu < 1$. This approximate form of $\theta$ is presented in Figure S7 by the solid, blue line.

A second important result present in Figure S7 is the fact that all network sizes can take on the empirically observed value of $\theta = 3/4$ as long as $\nu$ is large enough. This result supports previous arguments regarding transitions in blood flow related to shifts in the scaling of radii. In these circumstances, the radial scaling transitions from cross-sectional area-preserving ($\beta^2 = 1/2$ for a symmetrically bifurcating network) to area-increasing ($\beta^3 = 1/2$). Due to the conservation fluid flow-rates across branching junctions, this latter scaling acts to slow the flow of blood. However, a simultaneous result of changing the scaling of radii in such a way is to increase the volume scale factor, assuming that there is no change in the scaling of lengths. Thus, this predicted trade-off between decreasing network size and increasing volume scaling would appear consistent with the biological demands associated with the cardiovascular system in mammals. As for plants, ontogenetic shifts in metabolic scaling (as measured through total respiration) have been reported between sapling and mature trees and shrubs, yet not in the context of the measurement of vascular branching traits [16]. Thus, it would be interesting to see if future studies corroborate trade-offs between metabolic scaling
exponents, vascular branching traits, and network size.

In using Eq. 9, one must specify (and thereby either measure or estimate) the total number of branching generations in a vascular network. Unfortunately this is not a particularly straightforward task. With the conventional WBE generational labeling scheme, the generation index is advanced once a branching occurs for all branches within the previous generation. For highly asymmetric networks this can result in having branches of wildly differing size be members of the same generation. Alternative labelling schemes have been proposed outside the context of the WBE framework, specifically that of Horton-Strahler [17, 18]. While capable of handling moderate levels of asymmetry, one can show that the Horton-Strahler labelling scheme is insufficient at handling networks resembling the fishbone branching pattern. For the purposes of this work, we adopt an approach based on the asymptotically symmetric expectation that for a network with \( N \) generations, there should be approximately \( 2^N \) terminal branches. Thus, from a count of the number of terminal branches, \( N_{\text{TIPS}} \), one can estimate the number of generations as \( N = \ln(N_{\text{TIPS}}) / \ln(2) \).

Lastly, in the main texts standard errors are reported for calculated values of the metabolic scaling exponent in Figure 4a. These errors were determined from standard methods, namely \( \sigma^2 \approx \| \theta \|_2^2 \sigma_\nu \), where \( \sigma_\nu \) is determined from the chosen symmetric or asymmetric scale factor variables.

### E Derivation of curvature in metabolic rate versus mass

Here we present a derivation of the curvature in the metabolic rate, \( B \), of an organism in terms versus its mass, in log-log space. Beginning again with Kleiber’s Law after having linearized the equation, we have,

\[
\ln(B) = \ln(B_0) + \theta \ln \left( \frac{M}{M_0} \right)
\]

(10)

As we are interested in examining the mass related curvature in metabolic rate, we must evaluate \( \partial^2 \ln(B) / \partial(\ln(M))^2 \). Since \( \theta \) is effectively a mass-dependent quantity through its dependence on network size via total generation number \( N \), we need to find an appropriate substitute for mass. Assuming that the total mass of a vascular organism can be expressed asymptotically as the sum total of all cells serviced by the vascular architecture, then \( M \approx M_{\text{TIP}} 2^N \). Thus, we can build a derivative operator as,

\[
\frac{\partial}{\partial \ln(M)} = \frac{1}{\ln(2)} \frac{\partial}{\partial N}
\]

(11)

Using this operator, curvature in metabolic rate can be expressed as follows,

\[
\frac{\partial^2 \ln(B)}{\partial(\ln(M))^2} = \frac{2}{\ln(2)} \frac{\partial \theta}{\partial N} + \frac{N}{\ln(2)} \frac{\partial^2 \theta}{\partial N^2}
\]

(12)

where we have maintained the explicit notation for mass, \( M \), on the left hand side as a reminder of what the quantity on the right represents. Now the problem is reduced to calculating derivatives of the metabolic scaling exponent \( \theta \) with respect to total number of generations \( N \). An important and immediate consequence of Eq. 12 is that the approximate (\( \nu < 1 \)) and asymptotic (\( N >> 1 \)) version of Eq. 9, \( \theta \approx \ln(2)/[\ln(2) - \ln(\nu)] \), results in zero curvature due to its invariance in relative network size. This prediction is consistent with alternative parameterizations of curvature in metabolic scaling [19], and could inform observed deviations from previous theoretical predictions of metabolic scaling exponents as discussed in the main text.

Focusing on the version of Eq. 9 for \( \nu \neq 1 \), we find that the curvature in \( B \) can be expressed finally in the relatively compact form of,
\[ \frac{\partial^2 \ln(B)}{\partial (\ln(M))^2} = \frac{1}{x^4} \left\{ 2x^3 + 2N^2x \left( \frac{\partial x}{\partial N} \right)^2 - 4Nx^2 \frac{\partial x}{\partial N} - N^2x^2 \frac{\partial^2 x}{\partial N^2} \right\} \]  

(13)

where \( x = \ln \left( \frac{V_{\text{tot}}}{V_{\text{tip}}} \right) \). Graphs of Eq. 13 are presented in the main text, where we find that curvature is predicted to always be positive, consistent with previous studies based on mammalian respiration [19], but inconsistent with studies based on plant respiration [16]. Noting the form of \( x \), we can identify that \( x \propto N \), thus curvature is proportional to \( 1/N \) and decreases with an increasing number of generations. This network size-based suppression of curvature has been reported in relation to plant respiration [16].

These results demonstrate a current knowledge gap within the field of metabolic scaling, and motivate the simultaneously conducting respiration-based measurements of metabolic rates and vascular imaging to better synchronize prediction with observation.

References

[1] Duong, T., Goud, B. & Schauer, K. Closed-form density-based framework for automatic detection of cellular morphology changes. *Proceedings of the National Academy of Sciences* **109**, 8382–8387 (2012).

[2] Duong, T. Local significant differences from nonparametric two-sample tests. *Journal of Nonparametric Statistics* **25**, 635–645 (2013).

[3] Dray, S. & Josse, J. Principal component analysis with missing values: a comparative survey of methods. *Plant Ecology* **216**, 657–667 (2015).

[4] Newberry, M. G., Ennis, D. B. & Savage, V. M. Testing foundations of biological scaling theory using automated measurements of vascular networks. *PLoS Computational Biology* **11**, e1004455 (2015).

[5] Bentley, L. P. *et al.* An empirical assessment of tree branching networks and implications for plant allometric scaling models. *Ecology Letters* **16**, 1069–1078 (2013).

[6] Brummer, A. B., Savage, V. M. & Enquist, B. J. A general model for metabolic scaling in self-similar asymmetric networks. *PLoS Computational Biology* **13**, 1–25 (2017).

[7] Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2016), 2nd edn.

[8] Moses, A. M. *Statistical Modeling and Machine Learning for Molecular Biology* (CRC Press, 2017).

[9] Tekin, E., Hunt, D., Newberry, M. G. & Savage, V. M. Do vascular networks branch optimally or randomly across spatial scales? *PLoS Computational Biology* **12**, 1–28 (2016).

[10] Yang, J., Yu, L., Rennie, M. Y., Sled, J. G. & Henkelman, R. M. Comparative structural and hemodynamic analysis of vascular trees. *American Journal of Physiology-Heart and Circulatory Physiology* **298**, H1249–H1259 (2010).

[11] Géron, A. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems* (O’Reilly Media, Inc., 2017).

[12] 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).
[13] Savage, V. M., Deeds, E. J. & Fontana, W. Sizing up allometric scaling theory. *PLoS Computational Biology* **4**, e1000171 (2008).

[14] Kleiber, M. Body size and metabolism. *Hilgardia* **6**, 315–353 (1932).

[15] Savage, V. M. *et al.* The predominance of quarter-power scaling in biology. *Functional Ecology* **18**, 257–282 (2004).

[16] Mori, S. *et al.* Mixed-power scaling of whole-plant respiration from seedlings to giant trees. *Proceedings of the National Academy of Sciences* **107**, 1447–1451 (2010).

[17] Horton, R. E. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological society of America bulletin* **56**, 275–370 (1945).

[18] Strahler, A. N. Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union* **38**, 913–920 (1957).

[19] Kolokotrones, T., Savage, V., Deeds, E. J. & Fontana, W. Curvature in metabolic scaling. *Nature* **464**, 753–6 (2010).