Title: MAT drives the variation in maximum potential species richness and frost organizes the residual variation

Ricardo A. Segovia[1,2]

1 School of GeoSciences, University of Edinburgh, United Kingdom.
2 Instituto de Ecología y Biodiversidad (www.ieb-chile.cl), Santiago, Chile.

* ricardo.segovia@ed.ac.uk

Abstract

The kinetic hypothesis of biodiversity proposes that Mean Annual Temperature (MAT) is the main driver of variation in species richness, given its exponential effect on rates of energy flux and thus, potentially, on rates of biological interaction and diversification. However, limited support for this hypothesis has been found to date. We tested the fit of this model on the variation of tree species richness across the Americas. We found that the kinetic hypothesis accurately predicts the upper bound of the relationship between the inverse of temperature (1,000/kK) and the natural logarithm of species richness. In addition, we found that the number of frost days organizes a substantial portion of the residual variation. Historically, attempts to explain large-scale variation of species richness has focused on gradients of independent variables, but explanatory power has been limited. More than a gradient, the fit of the upper bound of the exponential Boltzmann temperature model of variation of observed species richness can be seen as an upper limit on the species richness per unit of MAT. Likewise, the distribution of the residuals of the upper bound model in function of the number of days with freezing temperatures, shows the importance of environmental thresholds, rather than gradients driving species richness variation.

Main text

While correlations between species richness and environmental variables have been widely documented [1, 10, 11, 14, 15, 19, 21, 24, 27–30, 35, 37], a general law or even a single, coherent explanation for these correlations is still elusive [5, 16, 33]. The most formal and testable hypothesis to mechanistically relate species richness to environmental parameters is the kinetic model, which forms part of the framework of the "Metabolic Theory of Ecology" [5]. Under this hypothesis, Mean Annual Temperature (MAT) would be the main driver of variation in species richness, given its exponential effect on rates of energy flux and thus, potentially, on rates of biological interaction and diversification [2, 33]. As a result, a predictive model derived from the exponential Boltzmann temperature relationship has been proposed and tested across different taxa [2]. However, subsequent studies at global [24] and regional [19] scales show no support for two of the main predictions of the model: 1) a linear relationship between species richness and temperature, and 2) a slope of -0.65 between the inverse of temperature (1.000/kK) and the natural logarithm of species richness [2].

In addition to temperature, species richness has also been related to water availability, often measured as mean annual precipitation [37]. The effect of water availability could explain the failure of previous studies to support the kinetic model [12]. Currie (2007) [12] hypothesized that the model proposed by Allen *et al.* (2002) [2] only fits the upper bound of the relationship between species richness and MAT in plants, but cannot explain the species richness variation in general. Allen *et al.* (2007) [3] agree with this, and suggest that the variation below the proposed upper bound fit would be associated with variation in total tree abundance due, mainly, to water deficit. Indeed, statistical models show that forests can reach their maximum diversity when high annual energy input is combined with high water availability throughout the year [24]. However, the original model and the posterior empirical approaches have only tested the central tendency of the variation in species richness [19, 24], not the upper bound

of species richness. In order to explore the behavior of the kinetic model around the upper bound species richness, given observed MAT values, we analyzed the variation of species richness in 10,721 inventories of tree species across the Americas (between 49°N and 55°S). We tested the predictions of linearity and the negative slope (~ 0.65) using regression models with MAT as an explanatory variable for species richness on its upper quantiles. In addition, we explored the environmental drivers of the variation in the residuals of the mean annual temperature – maximum potential species richness relationship.

To test the linearity of the relationship between the inverse of temperature (1,000/kK) and the upper quantiles of the natural logarithm of species richness, we assessed if breakpoint models, that allow for a change in slope at some point over the range of the explanatory variable, provided better statistical fits than linear models. Thus, we compared linear versus breakpoint models above the mean of the relationship (55%, 65%, 75%, 85%, 95%). Although segmented models are more robust in all of the quantiles analyzed (see AIC, Table 1), the slopes in the segmented

Table 1 Slopes and linearity across the upper bound of the relationship between the inverse of temperature (1,000/kK) and the natural logarithm of species richness. Slope values for linear and segmented quantile regressions (slope 1 and slope 2) are provided. The p-value column means the results of the t tests comparing if both slopes in the segmented model are significantly different (p-values $<$ 0.05). Finally, the Akaike Information Criteria are provided for linear and segmented in each quantile regression towards the upper bound.

Quantile	Linear Slope	Segmented Slope 1	Segmented Slope 2	p-value	AIC (linear)	AIC (segmented)
55%	-0.781	0.297	-1.442	0.041	29117.46	27895.73
65%	-0.789	0.306	-1.458	0.039	29206.52	27986.76
75%	-0.823	0.147	-1.312	0.072	29599.82	28520.47
85%	-0.808	-0.194	-1.011	0.207	30390.64	29279.77
95%	-0.661	-0.236	-1.008	0.22	32444.54	30946.64

models towards the upper bound are not significantly different (Table 1, p-values greater to 0.05 to the test of the null hypothesis, slope 1 = slope 2). Therefore, we cannot reject a linear relationship towards the upper bound (Fig. 1). Actually, only towards the very upper bound (quantile 85% and 95%) are both slopes negatives, while in 55%, 65% and 75%, the first slope is positive. Furthermore, we found that the negative linear slope of ~ 0.65, proposed by Allen *et al.* (2002) [2] is reached around the 95% quantile (Table 1, Fig. 1, Fig. S1). Analyzing the residuals of the linear model at quantile 95%, we found a trend of water availability driving the residual variation below the upper bound (Fig. 2a). In addition, we found a pattern of change of state along the axis of the cumulative number of frost days. Those sites with more than $\sim 1,000$ frost days during the last 117 years tend to be significantly further from the model proposed in the quantile 95% than those with fewer frost days (Fig. 2b).

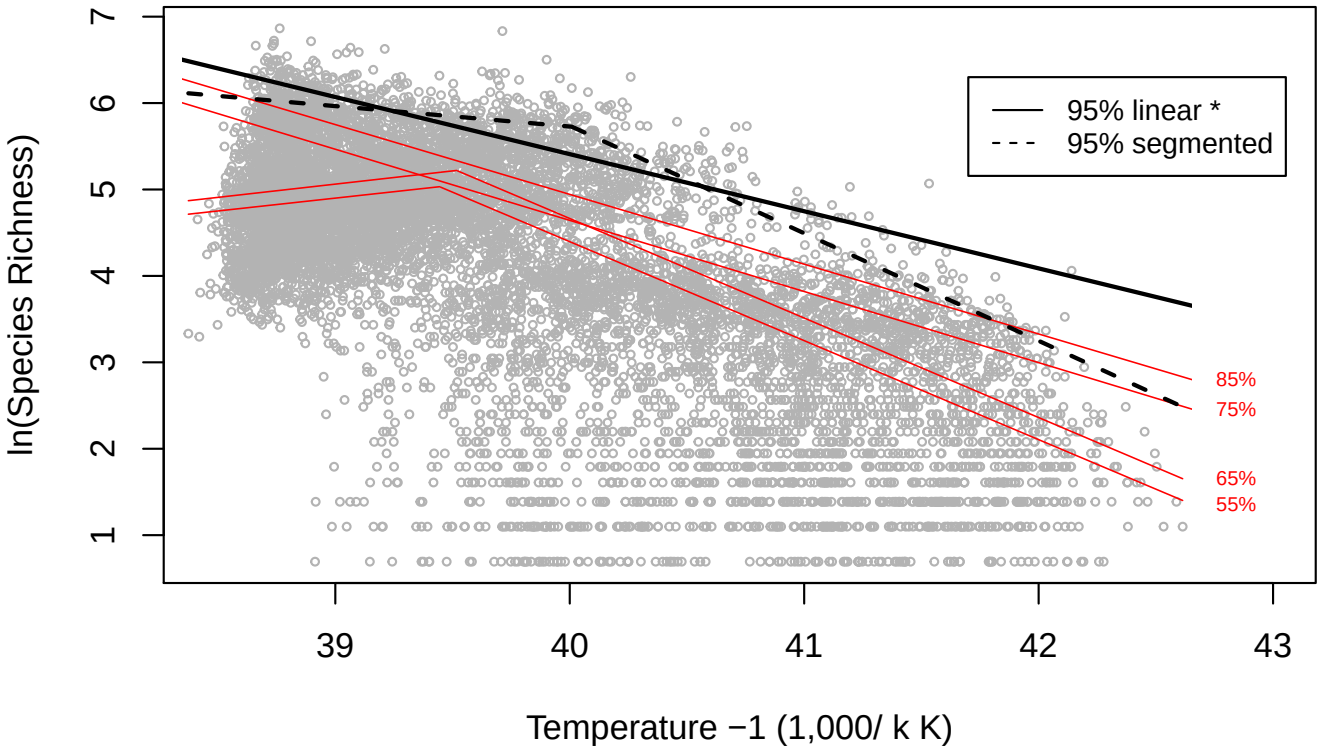


Fig.1 Effects of Mean Annual Temperature (MAT) on log-transformed species richness of trees in the Americas. Black lines are the model fit at 95% quantile for a linear regression (continuous) and for a one-breakpoint segmented regression (dashed). The red lines represents the stronger models in 55%, 65%, 75% and 85% quantiles.

Limits and thresholds controlling species richness

Our results point to the importance of environmental limits and thresholds controlling the variation in species richness. We show that the predictions of the kinetic model [2] are restricted to the upper bound of the relationship between species richness and Mean Annual Temperature (MAT) for tree assemblages from the Americas. That suggests that the effect of MAT on metabolic rates of individual trees drives an upper limit for the number of species, and that assemblages with species richness below their potential given MAT must be constrained by other factors. In concordance to the hypothesis of Allen *et al.* (2007) [3], we found that the residual variation below the upper bound is related to variation water deficit (Fig. 2a). However, we also show that an environmental threshold across the cumulative number of frost days axis ($\sim 1,000$ frost days between 1901 and 2018) triggers a change of state in the residual variation below the upper bound fit (Fig. 2b). Community assemblages subjected to more than $\sim 1,000$ frost days during the last 117 years have different behavior with respect to upper bound fit than assemblages that do not experience or experience few events of frost (less than $\sim 1,000$ days or ~ 10 frost days per year).

The variation in the observed species richness below the upper bound model is related to water deficit and freezing temperature. Therefore, our results are coherent with hypotheses suggesting that eco-evolutionary process associated to the conservative evolution of lineages to inhabit "harsh", non tropical (i.e. dry and/or freezing) environments drives the variation in species richness [10, 36]. We suggest that this evolutionary effect is complementary to the metabolic mechanism, inhibiting assemblages to reach the potential number of species expected by MAT. While water deficit may constrain the potential species richness by drop in total abundance [2, 13, 16], the cumulative frost days may constrain it by interrupting (freezing) the Boltzman relationship, and, therefore, slowing down the ecological and evolutionary rates. Indeed, the kinetic relationship

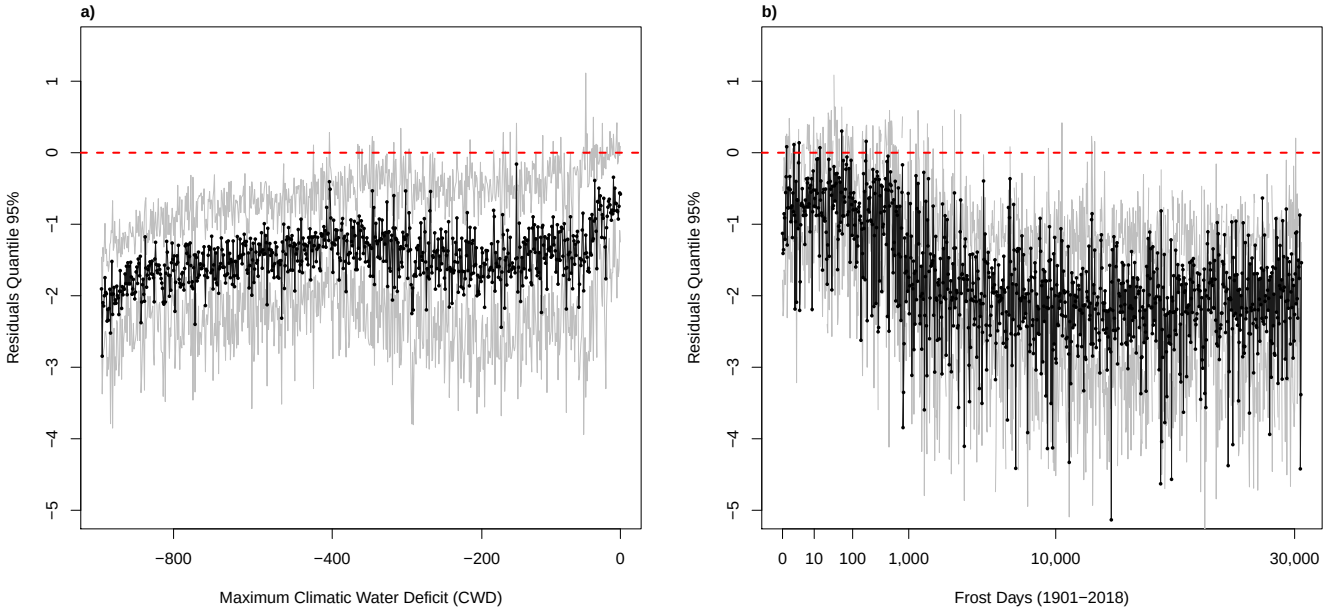


Fig.2 Distribution of the residuals according to the 95% quantile fit along two limiting species richness environmental variables: A) the Maximum Climatological Water Deficit, and B) the Number of Frost Days between 1901 and 2018.

between temperature and metabolic rates is expected to hold only over the temperature above 0°C [6].

Changing the focus from the central tendency towards the limits allows for clarifying that the metabolic effect of MAT works on the maximum species richness and, therefore, to conclude that the kinetic hypothesis described a potential model of global species richness variation. Instead, the observed variation of species richness is controlled by secondary environmental variables in different ways, suggesting that its effect can be spatial or geographically structured. In fact, at a global scale, measures of ambient energy (mean annual temperature or potential evapotranspiration) best predicted the species richness variation at high latitudes; and water-related variables (actual evapotranspiration and annual rainfall) best predicted richness variation in low-latitude, high-energy regions [10, 14, 19, 24]. Likewise, new structural patterns rise from a perspective of species richness being controlled by an environmental threshold, rather than gradients. For example, the structure of wet-tropics rich versus dry-tropics and extratropics poor seems more feasible than the

famous pattern of the latitudinal gradient of species richness.

Methods

Dataset

Our tree assemblage dataset was derived by combining the NeoTropTree (NTT; `http://neotroptree.info`) database [31] and the Forest Inventory and Analysis (FIA) Program of the U.S. Forest Service [7], accessed on March, 2019 via the BIEN package [25] for the R Statistical Environment [32]. We excluded from the FIA database any plot that had less than two species. The inventories in the NTT database are defined by a single vegetation type within a circular area with a 5 km radius and contains records of tree and tree-like species, i.e., freestanding plants with stems that can reach over 3m in height (see www.neotroptree.info and Silva de Miranda *et al.* (2018) [34] for details). Each FIA plot samples trees that are > 12.7 cm diameter at breast height (dbh) in four subplots (each being 168.3 m2) that are 36.6 m apart. We aggregated plots from the FIA dataset in a 10 km diameter space, to parallel the spatial structure of the NTT database. This procedure produced a total dataset of 10,721 tree assemblages distributed across major environmental and geographic gradients in the Americas. Taxonomic nomenclature was made consistent by querying species names against The Plant List (`http://www.theplantlist.org`). Mean annual temperature and minimum temperature of the coldest month were extracted from the Worldclim dataset [20] and Maximum Climatological Water Deficit (CWD) from Chave *et al.* 2014 [9] for the coordinates assigned to each tree assemblage.

Statistical Analyses

Kinetic hypothesis testing: In order to explore the fit of the relationship between the inverse of temperature (1,000/kK) and the natural logarithm

of species richness towards the upper bound, we used Quantile Regressions [23] above the mean linear regression (50%). As quantile regression is insensitive to heteroscedastic errors and dependent variable outliers [23], it is useful to explore the prediction of the kinetic hypothesis. In addition, quantile regressions have been previously described as particularly useful in exploratory and inferential analyses concerning limiting factors in ecology [8], which makes it useful to study the relationship of species richness and temperature towards the upper bound. Thus, we tested the two main predictions derived from the kinetic hypothesis: linearity and a negative slope of -0.65.

Linearity: We fitted linear Quantile Regression and one-breakpoint segmented Quantile Regression in 55%, 65%, 75%, 85% and 95% quantiles, with "quantreg" [22] and "segmented" [26] packages in R [32]. In order to define if a linear or a segmented fit is better in each quantile, we used two approaches. First, we compared the models by Akaike information criterion (AIC). Second, following Hawkins and Felizola Diniz-Filho (2004) [19], we tested if both slopes in the segmented model are different from each other using a t test. If the two slopes were not significantly different ($P > 0.05$), the relationship between rescaled temperature and ln-transformed richness was classified as being linear on the quantile, whereas data sets with significantly different slopes were classified as being nonlinear.

Slope: Brown *et al.* (2004) [6] and Allen *et al.* (2007) [3] argued that slopes of richness–temperature regressions should fall near ~ 0.65. In order to test this prediction towards the upper bound of the relationship, we simply compared the slopes of all the linear Quantile Regressions. Examining quantile regression functions across the entire range provides a more complete view of the relationship between variables than what is captured by least squares or individual quantile regression functions.

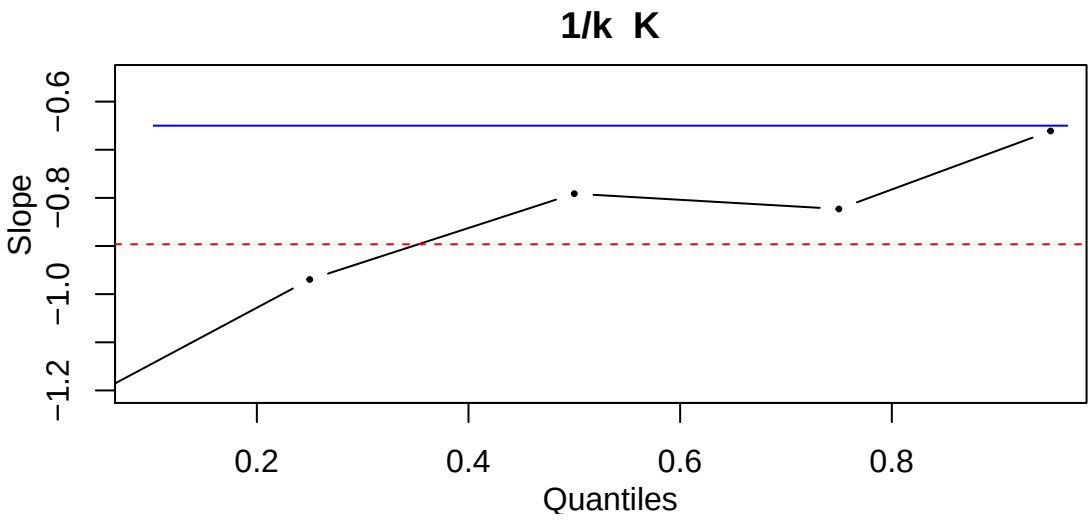
Residuals Variation: As we found the 95% quantile regression is the best model to fit the kinetic model for species richness, we explored the environmental determinants of the residual variation at this quantile. We

tested the variation of the residuals with the variation of Maximum Climatological Water Deficit (CWD) from Chave et al. (2014) [9] because Allen *et al.* (2007) [3] hypothesized that the residual variation on an upper bound fit should be associated with water availability. Alternatively, we evaluated the effect of cumulative frost days from 1901 until 2018, from the Climatic Research Unit's (CRU) [18] on the residual variation because the frost line has been proposed as an environmental threshold for taxonomic, functional, and evolutionary diversity on trees and, therefore, we hypothesized that it could also affect the relationship between species richness and mean annual temperature. In order to plot the distribution of residuals across the continuous environmental variables, we use the function discretize() in the "arules" package [17] with the "cluster" method. Thus, we calculated a mean and a standard deviation across the ordered categories in both of the variables.

Acknowledgments: CONICYT PIA APOYO CCTE AFB170008. R.A.S. is supported by a Newton International Fellowship from The Royal Society and by Conicyt PFCHA/Postdoctorado Becas Chile/2017 N° 3140189.

Supplemenntary material

Fig.S1 Slopes of the estimated linear quantile regression for the species richness - mean annual temperature relationship. The red line represents the mean slope. The blue line represents the slope predicted by Allen, Brown, and Gillooly (2002) [2], Allen *et al.* (2006) [4], Brown *et al.* (2004) [6].



References

1. JM Adams and FI Woodward. Patterns in tree species richness as a test of the glacial extinction hypothesis. *Nature*, 339(6227):699, 1989.
2. Andrew P Allen, James H Brown, and James F Gillooly. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, 297(5586):1545–1548, 2002.
3. Andrew P Allen, James F Gillooly, James H Brown, et al. Recasting the species-energy hypothesis: the different roles of kinetic and potential energy in regulating biodiversity. *Scaling biodiversity*, 1, 2007.
4. Andrew P Allen, James F Gillooly, Van M Savage, and James H Brown. Kinetic effects of temperature on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences*, 103(24):9130–9135, 2006.
5. James H. Brown. Why are there so many species in the tropics? *Journal of Biogeography*, 41(1):8–22, 2014.
6. James H Brown, James F Gillooly, Andrew P Allen, Van M Savage, and Geoffrey B West. Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789, 2004.
7. Elizabeth Burrill, Andrea Wilson, Jeffery Turner, Scott Pugh, James Menlove, Glenn Christiansen, Barbara Conkling, and Winnie David. *The Forest Inventory and Analysis Database: database description and user guide version 8.0 for Phase 2. U.S.* Department of Agriculture, Forest Service, Available at web address: www.fia.fs.fed.us/library/database-documentation, 2018.
8. Brian S Cade, James W Terrell, and Richard L Schroeder. Estimating effects of limiting factors with regression quantiles. *Ecology*, 80(1):311–323, 1999.
9. Jérôme Chave, Maxime Réjou-Méchain, Alberto Búrquez, Emmanuel Chidumayo, Matthew S Colgan, Welington BC Delitti, Alvaro Duque, Tron Eid, Philip M Fearnside, Rosa C Goodman, et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global change biology*, 20(10):3177–3190, 2014.
10. David J Currie, Gary G Mittelbach, Howard V Cornell, Richard Field, Jean-Francois Guégan, Bradford A Hawkins, Dawn M Kaufman, Jeremy T Kerr, Thierry Oberdorff, Eileen O'Brien, et al. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology letters*, 7(12):1121–1134, 2004.

11. David J Currie and Viviane Paquin. Large-scale biogeographical patterns of species richness of trees. *Nature*, 329(6137):326, 1987.

12. DJ Currie. Regional-to-global patterns of biodiversity, and what they have to say about mechanisms. *Scaling biodiversity*, pages 258–282, 2007.

13. Brian J Enquist and Karl J Niklas. Invariant scaling relations across tree-dominated communities. *Nature*, 410(6829):655, 2001.

14. Richard Field, Eileen M O'Brien, and Robert J Whittaker. Global models for predicting woody plant richness from climate: development and evaluation. *Ecology*, 86(9):2263–2277, 2005.

15. Anthony P Francis and David J Currie. A globally consistent richness-climate relationship for angiosperms. *The American Naturalist*, 161(4):523–536, 2003.

16. Kevin J Gaston. Global patterns in biodiversity. *Nature*, 405(6783):220, 2000.

17. Michael Hahsler, Christian Buchta, Bettina Gruen, and Kurt Hornik. *arules: Mining Association Rules and Frequent Itemsets*, 2019. R package version 1.6-3.

18. IPDJ Harris, Philip D Jones, Timothy J Osborn, and David H Lister. Updated high-resolution grids of monthly climatic observations–the cru ts3. 10 dataset. *International journal of climatology*, 34(3):623–642, 2014.

19. Bradford A Hawkins, Eric E Porter, and José Alexandre Felizola Diniz-Filho. Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology*, 84(6):1608–1623, 2003.

20. Robert J Hijmans, Susan E Cameron, Juan L Parra, Peter G Jones, and Andy Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 25(15):1965–1978, 2005.

21. Walter Jetz and Carsten Rahbek. Geographic range size and determinants of avian species richness. *Science*, 297(5586):1548–1551, 2002.

22. Roger Koenker. *quantreg: Quantile Regression*, 2018. R package version 5.38.

23. Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

24. Holger Kreft and Walter Jetz. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences*, 104(14):5925–5930, 2007.

25. Brian S Maitner, Brad Boyle, Nathan Casler, Rick Condit, John Donoghue, Sandra M Durán, Daniel Guaderrama, Cody E Hinchliff, Peter M Jørgensen, Nathan JB Kraft, et al. The bien r

package: A tool to access the botanical information and ecology network (bien) database. *Methods in Ecology and Evolution*, 9(2):373–379, 2018.

26. Vito M.R. Muggeo. segmented: an r package to fit regression models with broken-line relationships. *R News*, 8(1):20–25, 2008.
27. Jens Mutke and Wilhelm Barthlott. Patterns of vascular plant diversity at continental to global scales. *Biologiske skrifter*, 55(4):521–531, 2005.
28. Jens Mutke, Gerald Kier, Gerald Braun, Chr. Schultz, and Wilhelm Barthlott. Patterns of african vascular plant diversity: A gis based analysis. *Systematics and Geography of Plants*, pages 1125–1136, 2001.
29. Eileen O'Brien. Water-energy dynamics, climate, and prediction of woody plant species richness: an interim general model. *Journal of Biogeography*, 25(2):379–398, 1998.
30. Eileen M O'Brien. Climatic gradients in woody plant species richness: towards an explanation based on an analysis of southern africa's woody flora. *Journal of Biogeography*, pages 181–198, 1993.
31. AT Oliveira-Filho. Neotroptree, flora arbórea da região neotropical: Um banco de dados envolvendo biogeografia, diversidade e conservação. *Belo Horizonte: Universidade Federal de Minas Gerais*, 2014.
32. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
33. Klaus Rohde. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos*, pages 514–527, 1992.
34. Pedro Luiz Silva de Miranda, Ary T. Oliveira-Filho, R. Toby Pennington, Danilo M. Neves, Baker, Timothy R., and Kyle G. Dexter. Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America. *Global ecology and biogeography*, 27(8):899–912, AUG 2018.
35. Nathan L Stephenson. Climatic control of vegetation distribution: the role of the water balance. *The American Naturalist*, 135(5):649–670, 1990.
36. JJ Wiens and MJ Donoghue. Historical biogeography, ecology and species richness. *Trends in ecology & evolution*, 19(12):639–644, DEC 2004.
37. David Hamilton Wright. Species-energy theory: an extension of species-area theory. *Oikos*, pages 496–506, 1983.