Midpoint attractors and species richness: Modeling the interaction between environmental drivers and geometric constraints

APPENDICES

APPENDIX 1: SUPPLEMENTAL TABLES AND FIGURES

Tables S1–S3

Figures S1–S3

APPENDIX 2: SUPPLEMENTAL TEXT

Supplemental Introduction

Supplemental Materials and Methods

REFERENCES FOR THE APPENDICES

APPENDIX 1: SUPPLEMENTAL TABLES AND FIGURES

SUPPLEMENTAL TABLES S1–S3

Table S1. The datasets and their characteristics. *Sampling limits* represent the lowest and highest occurrence on a unit-line transect, after range adjustments described in the *Supplemental Materials and Methods* (Dataset Selection and Preparation). *Sampling scope* is the difference between the sampling limits.

Dataset	Locality	Transect coordinates	Sampling elevations	Sampling limits	Sampling scope	Domain limits (m asl)	Sampling method	Species	Environmental variables and their units	Data provider	Collection dates	Published references to the dataset
Costa Rica Datasets												
Ants	Barva Transect (Prov. Heredia)	10°08'N–10°26'N, 84°00'W–84°07'W	7	0.004, 0.705	0.701	0, 2900	Miniwinkler extractors	332	MAT (°C), Mean RH (%), MAP (mm), Area (% of total per 100m band)	John T. Longino	2001-2007	(Longino & Colwell 2011; Longino *et al.* 2014)
Arctiine moths	Barva Transect (Prov. Heredia)	10°08'N–10°26'N, 84°00'W–84°07'W	12	0.013, 0.940	0.927	0, 2900	Light traps, manual collection	222	MAT (°C), Mean RH (%), MAP (mm), Area (% of total per 100m band)	Gunnar Brehm	2003–2004	None
Geometrid moths	Barva Transect (Prov. Heredia)	10°08'N–10°26'N, 84°00'W–84°07'W	12	0.013, 0.940	0.927	0, 2900	Light traps, manual collection	739	MAT (°C), Mean RH (%), MAP (mm), Area (% of total per 100m band)	Gunnar Brehm	2003–2004	(Brehm *et al.* 2007)

Dataset	Locality	Transect coordinates	Sampling elevations	Sampling limits	Sampling scope	Domain limits (m asl)	Sampling method	Species	Environmental variables and their units	Data provider	Collection dates	Published references to the dataset
Ferns	Barva Transect (Prov. Heredia)	10°08'N–10°26'N, 84°00'W–84°07'W	29	0.011, 0.986	0.975	0, 2900	Plot-based (20x20m²)	434	MAT (°C), Mean RH (%), MAP (mm), Area (% of total per 100m band)	Jürgen Kluge	2002-2003	(Kluge *et al.* 2006)
Mammals	Tilarán Mt. Range	10°23'N–10°17'N, 84°47'W–84°26'W	18	0.000, 0.998	0.989	0, 1840	Live traps, kill traps, and pitfall traps	18	Average Temperature (°C), Annual Precipitation (mm) [worldclim], elevational area (km2 per 100m elevational band) [DEM, ArcGIS]	Christy McCain	2000-2002	(McCain 2004; McCain 2005)
Papua New Guinea Datasets												
Ants	Mt. Wilhelm Transect	5°44'S–5°47'S, 145°03'E–145°20'E	8	0.007, 0.822	0.815	0, 4509	Pitfall trapping and hand-collecting Yusah et al.(2012)	116	MAT (°C), Mean RH (%), MAP (mm), Area (% of total per 100m band)	Jimmy Moses, Tom M. Fayle, Petr Klimes	2012	(Moses 2015)
Butterflies	Mt. Wilhelm Transect	5°44'S–5°47'S, 145°03'E–145°20'E	8	0.022, 0.876	0.854	0, 4509	Modified Pollard transects (Caldas & Robbins 2003)	264	MAT (°C), Mean RH (%), MAP (mm), Area (% of total per 100m band)	Legi Sam	2009	(Sam 2011)

Dataset	Locality	Transect coordinates	Sampling elevations	Sampling limits	Sampling scope	Domain limits (m asl)	Sampling method	Species	Environmental variables and their units	Data provider	Collection dates	Published references to the dataset
Birds	Mt. Wilhelm Transect	5°44'S–5°47'S, 145°03'E–145°20'E	8	0.022, 0.876	0.854	0, 4509	Point-counts, mist-netting	245	MAT (°C), Mean RH (%), Mean Tree Height (m), Mean Tree Basal Area (cm2)	Katerina Sam	2010-2012	(Tvardikova 2013; Sam & Koane 2014)
Ferns	Mt. Wilhelm Transect	5°44'S–5°47vS, 145°03'E–145°20'E	8	0.022, 0.876	0.854	0, 4509	Plot-based (20x20m²)	359	MAT (°C), Mean RH (%), MAP (mm), Area (% of total per 100m band)	D.N. Karger, S. Noben, M. Lehnert, M.S. Sundue	2014	None
Australia Datasets												
Moths (macromoths + Pyraloidea)	Border Ranges (NSW)	28°24'S–28°22'S, 153°1'E–153°5'E	5	0.220, 0.959	0.739	0, 1100	Light traps	612	°C min, max median, average plant richness	Louise Ashton, Roger Kitching	2009-2010	(Ashton *et al.* 2015)
Leaf miners (Lepidoptera, Coleoptera, Diptera, Hymenoptera)	Border Ranges (NSW)	28°24'S–28°22'S, 153°1'E–153°5'E	5	0.183, 0.981	0.798	0, 1100	Hand collecting and rearing	34	Average Temperature (°C), Annual Precipitation (mm), Vegetation cover (log of cm's intercepted)	Sarah Maunsell	2011 - 2012	(Maunsell *et al.* 2016)

Dataset	**Locality**	**Transect coordinates**	**Sampling elevations**	**Sampling limits**	**Sampling scope**	**Domain limits (m asl)**	**Sampling method**	**Species**	**Environmental variables and their units**	**Data provider**	**Collection dates**	**Published references to the dataset**
Leaf miner parasitoids (Hymenoptera)	Border Ranges (NSW)	28°24'S–28°22'S, 153°1'E–153°5'E	5	0.183, 0.981	0.798	0, 1100	Hand collecting and rearing	14	Average Temperature (°C), Annual Precipitation (mm), Vegetation cover (log of cm's intercepted)	Sarah Maunsell	2011 - 2012	(Maunsell *et al.* 2015)
Borneo Datasets												
Geometrid moths	NE Borneo	1°28'N-6°16'N, 112°06'E-117°53'E	70	0.000, 0.958	0.958	0, 4095	Light traps	775	Average Temperature (°C), Annual Precipitation (mm) [worldclim], forest stratum, vegetation type [field descriptions]	Jan Beck, Jeremy Hollo-way, Chey Vun Khen	1965-2003	(Beck *et al.* 2012) (undis-turbed habitats only)
Sphingid moths	NE Borneo	0°05'S-6°18'N, 109°43'E-118°10'E	19	0.000, 0.958	0.958	0, 4095	Light traps	102	Average Temperature (°C), Annual Precipitation (mm) [worldclim], area [of 200m bands], vegetation type [globcov]	Jan Beck, Ian Kitching et al.	1965-2005	(Beck & Kitching 2009)

North America Datasets												
Butterflies	Califor-nia	38°34'N-39°20'N, 120°20'-121°25'W	6	0.001, 0.966	0.965	0, 2775	Pollard walk, presence/ absence	129	Average Max Daily Temperature (°C), Average Min Daily Temperature (°C), Annual Precipitation (mm)	Arthur Shapiro	1973-2014	(Forister *et al.* 2010).
Mammals	Yosemite NP (Califor-nia)	37°30'N–37°59'N, 118°56'–120°28'W	40	0.000, 0.990	0.990	0, 3997	Live traps, kill traps, hunting, visual observations	46	Average Temperature (°C), Annual Precipitation (mm) [worldclim], elevational area (km per 100m elevational band) [DEM, ArcGIS]	Joseph Grinnell & Tracy Storer	1914-1916, 1919.	(Grinnell & Storer 1924; McCain 2005)

Table S2. Midpoint attractors in relation to environmental variables and observed species richness, analyzed by multiple regression with AIC-based model selection. Values of $R^2 < 0.5$ are set in italics. Values of delta AIC > 3 are boldfaced. The corresponding scatterplots appear in Fig. S1. Because the attractor is a continuous function (a doubly-truncated Gaussian distribution) and the other variables are spatially autocorrelated, significance probabilities cannot be assigned to R^2 values, which are thus best viewed as comparative. Dark grey fill indicates datasets for which the same environmental predictor (or predictor set) best explains both the Attractor and Empirical richness, by a strict ΔAIC criterion (ΔAIC > 3). Light grey fill indicates datasets for which the same environmental predictor (or predictor set) best explains both the Attractor and Empirical richness, disregarding the ΔAIC criterion.

Dataset	**Response Variable**	**Predictor Variables**	***n***	R^2	**Condition Number**	**Delta AIC**
Costa Rica Datasets						
Ants	Empirical richness	Modeled richness	10	0.942	1	0
		Attractor	10	0.845	1	**9.563**
	Attractor	Temperature & Relative humidity	10	0.971	2.354	0
		Temperature & Area	10	0.971	4.625	1.332
		Temperature & Precipitation	10	0.972	2.548	1.857
	Empirical richness	Temperature	10	0.856	1	0
Arctiine moths	Empirical richness	Modeled richness	10	0.879	1	0
		Attractor	10	0.762	1	**6.77**
	Attractor	Precipitation & Area	10	0.852	1.616	0
		Relative humidity & Precipitation	10	0.833	1.04	1.217
		Temperature & Precipitation	10	0.825	2.548	1.712
		Precipitation	10	0.678	1	1.793
	Empirical richness	Temperature & Area	10	0.927	4.625	0
		Precipitation	10	0.839	1	1.896
Geometrid moths	Empirical richness	Modeled richness	10	0.898	1	0
		Attractor	10	0.869	1	2.448
	Attractor	Relative humidity	10	0.623	1	0

Dataset	**Response Variable**	**Predictor Variables**	***n***	R^2	**Condition Number**	**Delta AIC**
	Empirical	Area	10	0.647	1	0
		Temperature & Area	10	0.788	4.625	0.899
		Relative humidity	10	0.608	1	1.037
		Precipitation & Area	10	0.744	1.616	2.377
Ferns	Empirical richness	Modeled richness	10	0.898	1	0
		Attractor	10	0.883	1	1.312
	Attractor	Relative humidity	10	0.666	1	0
	Empirical richness	Precipitation & Area	10	0.785	1.616	0
		Temperature & Area	10	0.770	4.625	0.69
		Relative humidity	10	0.560	1	1.165
Mammals	Empirical richness	Attractor	10	0.630	1	0
		Modeled richness	10	0.611	1	0.538
	Attractor	Area	10	*0.483*	1	0
	Empirical richness	Area	10	*0.043*	1	0
		Precipitation	10	*0.020*	1	0.267
		Temperature	10	*0.013*	1	0.348
Papua New Guinea Datasets						
Ants	Empirical richness	Modeled richness	8	0.894	1	0
		Attractor	8	0.861	1	2.142
	Attractor	Temperature	8	0.586	1	0
		Tree height	8	0.539	1	0.859
	Empirical richness	Temperature	8	0.867	1	0
Butterflies	Empirical richness	Modeled richness	8	0.975	1	0
		Attractor	8	0.950	1	**5.461**
	Attractor	Temperature	8	0.925	1	0

Dataset	**Response Variable**	**Predictor Variables**	***n***	R^2	**Condition Number**	**Delta AIC**
		Temperature & Relative humidity	8	0.968	1.812	2.486
	Empirical richness	Temperature	8	0.842	1	0
Birds	Empirical richness	Modeled richness	8	0.935	1	0
		Attractor	8	0.876	1	**5.222**
	Attractor	Temperature	8	0.804	1	0
		Tree height	8	0.731	1	2.53
	Empirical richness	Temperature	8	0.958	1	0
		Temperature & Basal area	8	0.985	1.36	1.299
Ferns	Empirical richness	Modeled richness	8	0.813	1	0
		Attractor	8	0.810	1	0.137
	Attractor	Basal area	8	*0.447*	1	0
		Humidity	8	*0.272*	1	2.207
	Empirical richness	Basal area	8	*0.442*	1	0
		Humidity	8	*0.236*	1	2.518
Australia Datasets						
Moths*	Empirical richness	Attractor	10	0.926	1	0
		Modeled richness	10	0.907	1	1.966
	Attractor	Temperature-Precipitation PCA	10	*0.139*	1	0
		Tree Richness	10	*0.078*	1	0.625
	Empirical richness	Temperature-Precipitation PCA	10	*0.123*	1	0
		Tree Richness	10	*0.122*	1	0.007
Leaf-miners	Empirical richness	Modeled richness	10	*0.342*	1	1
		Attractor	10	*0.163*	1	2.162
	Attractor	Temperature-Precipitation PCA	10	0.560	1	0
	Empirical richness	Temperature-Precipitation PCA & Tree richness	10	0.704	1.357	0

Dataset	**Response Variable**	**Predictor Variables**	***n***	R^2	**Condition Number**	**Delta AIC**
		Tree Richness	10	*0.338*	1	0.032
		Temperature-Precipitation PCA	10	*0.164*	1	2.131
Leaf-miner parasitoids	Empirical richness	Attractor	10	0.939	1	0
		Modeled richness	10	0.770	1	**11.878**
	Attractor	Temperature-Precipitation PCA	10	*0.442*	0.939	0
	Empirical richness	Temperature-Precipitation PCA	10	*0.476*	1	0
Borneo Datasets						
Geometrid Moths	Empirical richness	Attractor	10	*0.469*	1	0
		Modeled richness	10	*0.461*	1	0.152
	Attractor	Temperature	10	0.680	1	0
	Empirical richness	Temperature	10	*0.188*	1	0
		Precipitation	10	*0.068*	1	1.337
Sphingid moths	Empirical richness	Modeled richness	10	0.994	1	0
		Attractor	10	0.713	1	**38.012**
	Attractor	Temperature	10	0.702	1	0
		Cover Classes	10	0.683	1	0.614
	Empirical richness	Temperature & Area	10	0.944	2.034	0
North American Datasets						
Butterflies	Empirical richness	Modeled richness	11	0.968	1	0
		Attractor	11	0.936	1	**7.506**
	Attractor	Precipitation	11	*0.404*	1	0
		Precipitation & Minimum temperature	11	0.624	2.324	0.17
		Precipitation & Maximum temperature	11	0.590	2.265	1.12
	Empirical richness	Precipitation	11	0.533	1	0

Dataset	**Response Variable**	**Predictor Variables**	***n***	R^2	**Condition Number**	**Delta AIC**
Mammals	Empirical richness	Modeled richness	10	0.725	1	0
		Attractor	10	0.697	1	**4.655**
	Attractor	Precipitation	10	*0.429*	1	0
	Empirical richness	Precipitation	10	*0.154*	1	0
		Area	10	*0.140*	1	0.163
		Temperature	10	*0.034*	1	1.327

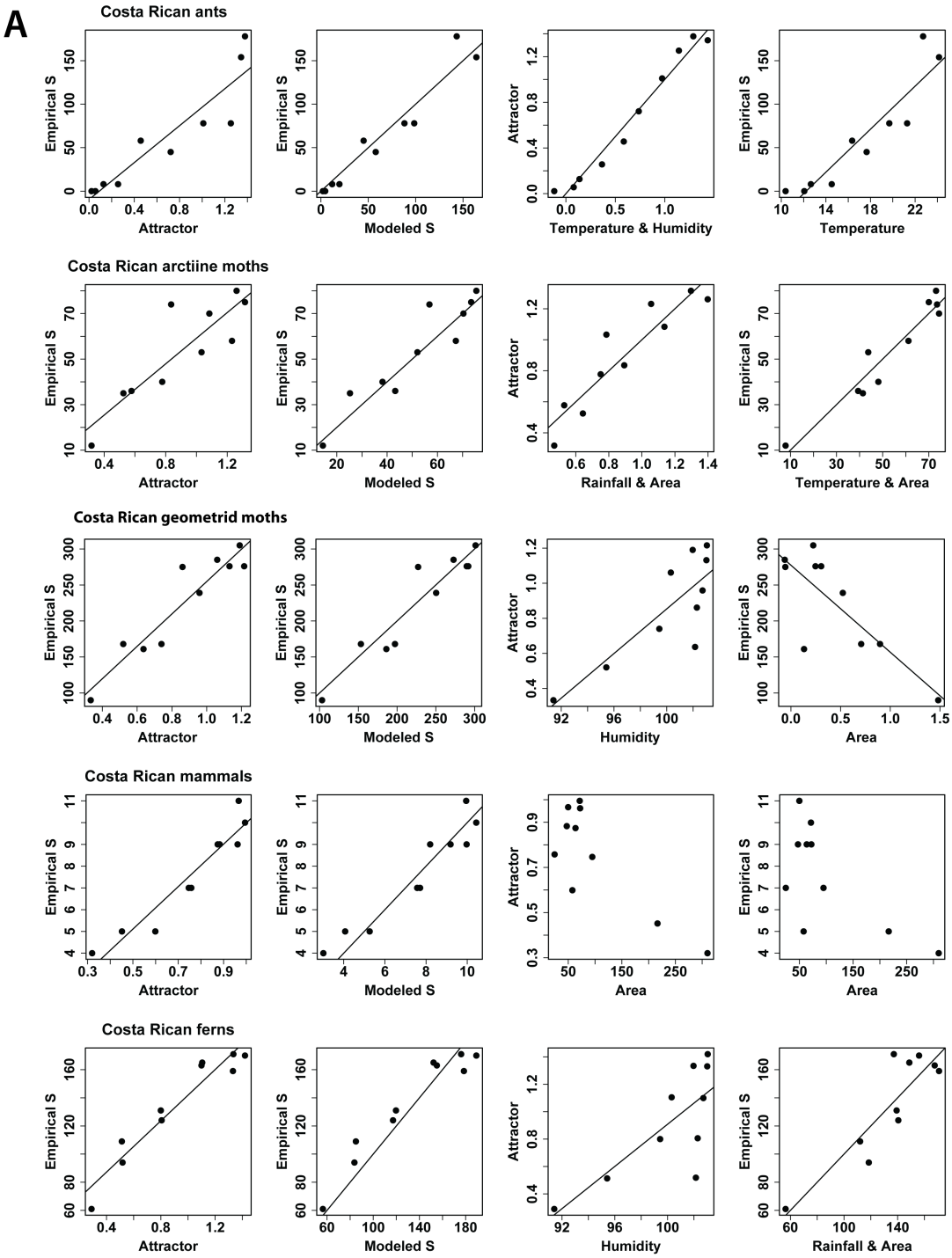
*Temperature and precipitation were highly (inversely) correlated for the Australian moths dataset (Condition Number = 21.696). PCA was extracted to reduce the effects of collinearity.

Table S3. Analysis of midpoint predictor models for range midpoint locations. Each row represents a different environmental variable used to model probabilities of midpoint occurrence along the domain. A plus sign (+) indicates $P < 0.05$, meaning that the results were improbable relative to a particular model ($P(data|model)$). Numerical entries indicate one-tailed P values, based on 1000 simulations, for which $P > 0.05$ indicates that the data were not improbable, given the model. See *Materials and Methods* in the main text for the algorithms of the two midpoint predictor models.

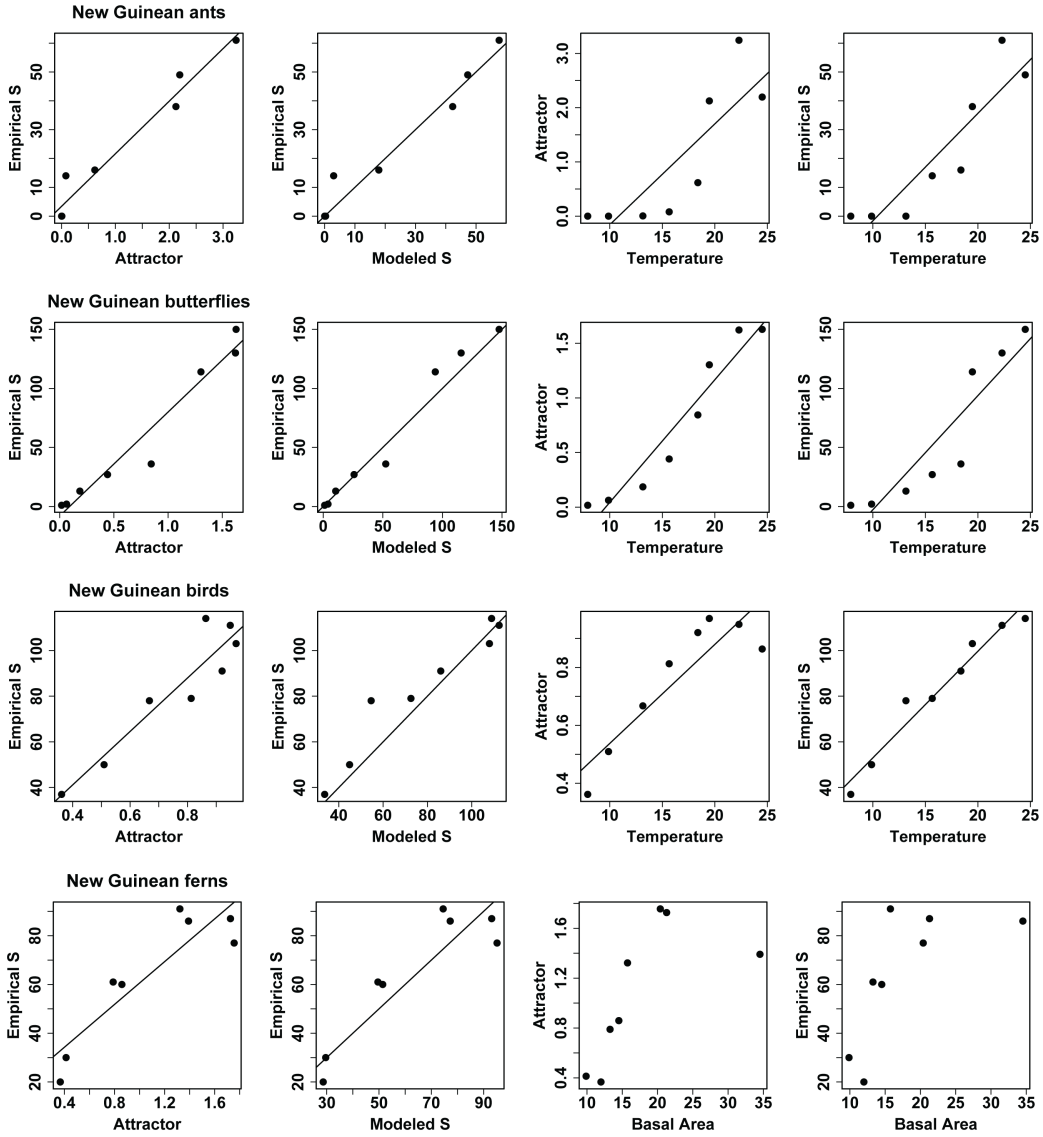
Dataset	**Environmental Variable**	**Model 1**	**Model 2**
Costa Rica Datasets			
Ants	Temperature	+	+
	Precipitation	+	+
	Relative humidity	+	+
	Area	+	+
Arctiine moths	Temperature	+	+
	Precipitation	+	+
	Relative humidity	+	+
	Area	+	+
Geometrid moths	Temperature	+	+
	Precipitation	+	+
	Relative humidity	+	+
	Area	+	+
Ferns	Temperature	+	+
	Precipitation	+	+
	Relative humidity	+	+
	Area	+	+
Mammals	Temperature	0.277	0.294
	Precipitation	0.300	0.305
	Area	+	+
Papua New Guinea Datasets			
Ants	Temperature	+	+
	Relative humidity	+	+
	Tree height	+	+
	Basal area	+	+
Butterflies	Temperature	+	+
	Relative humidity	+	+
	Tree height	+	+
	Basal area	+	+
Birds	Temperature	+	+
	Relative humidity	+	+
	Tree height	+	+

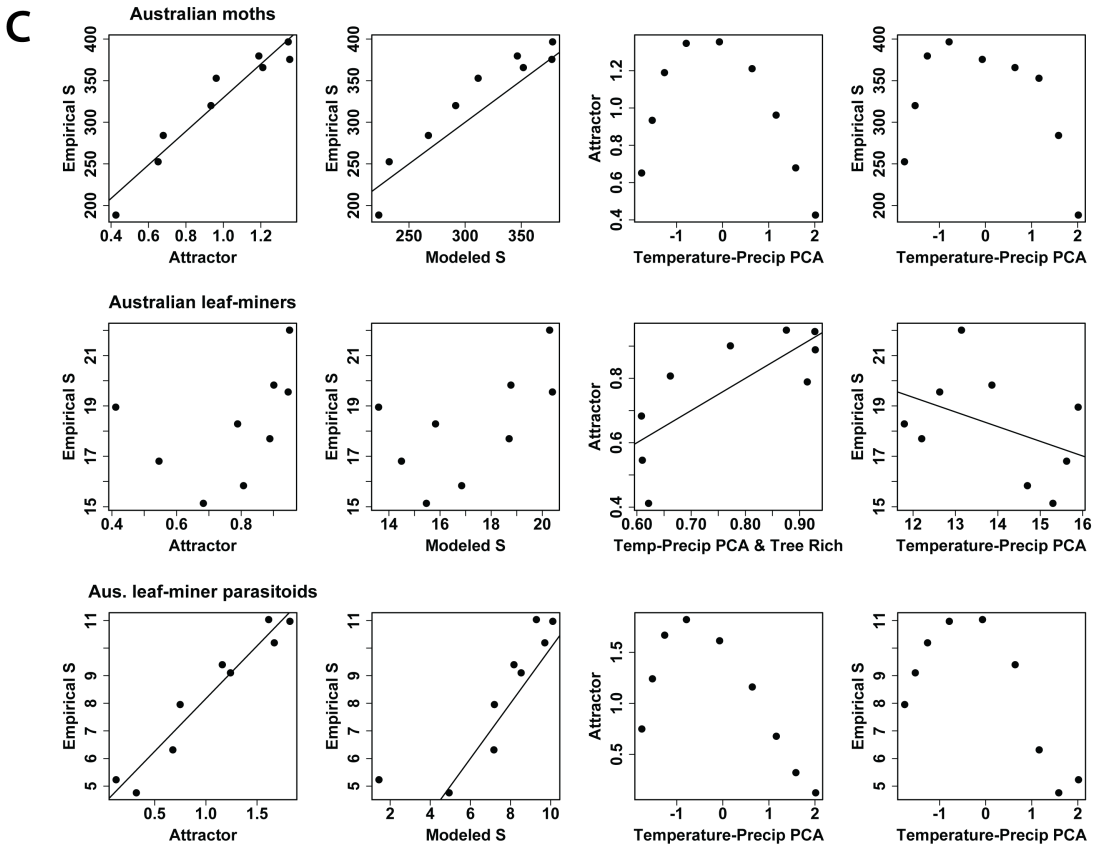
Dataset	Environmental Variable	Model 1	Model 2
	Basal area	+	+
Ferns	Temperature	+	+
	Relative humidity	+	+
	Tree height	+	+
	Basal area	+	+
Australia Datasets			
Moths	Temperature	+	+
	Precipitation	+	+
	Tree richness	+	+
Leaf-miners	Temperature	+	+
	Precipitation	0.132	0.083
	Tree richness	+	+
Leaf-miner parasitoids	Temperature	+	0.053
	Precipitation	+	0.056
	Tree richness	0.074	0.153
Borneo Datasets			
Geometrid moths	Temperature	+	+
	Precipitation	+	+
Sphingid moths	Temperature	+	+
	Precipitation	+	+
	Area	+	+
	Cover classes	+	+
North American Datasets			
Butterflies	Minimum temperature	+	+
	Maximum temperature	+	+
	Precipitation	+	+
Mammals	Temperature	0.266	0.069
	Precipitation	0.230	+
	Area	0.154	+

SUPPLEMENTAL FIGURES S1–S3



B


New Guinean ants
Empirical S
Attractor
Modeled S
Attractor
Temperature
Empirical S
Temperature
New Guinean butterflies
Empirical S
Attractor
Modeled S
Attractor
Temperature
Empirical S
Temperature
New Guinean birds
Empirical S
Attractor
Modeled S
Attractor
Temperature
Empirical S
Temperature
New Guinean ferns
Empirical S
Attractor
Modeled S
Attractor
Basal Area
Empirical S
Basal Area


C
Australian moths
Empirical S
Attractor
Modeled S
Attractor
Temperature-Precip PCA
Empirical S
Temperature-Precip PCA
Australian leaf-miners
Empirical S
Attractor
Modeled S
Attractor
Temp-Precip PCA & Tree Rich
Empirical S
Temperature-Precip PCA
Aus. leaf-miner parasitoids
Empirical S
Attractor
Modeled S
Attractor
Temperature-Precip PCA
Empirical S
Temperature-Precip PCA

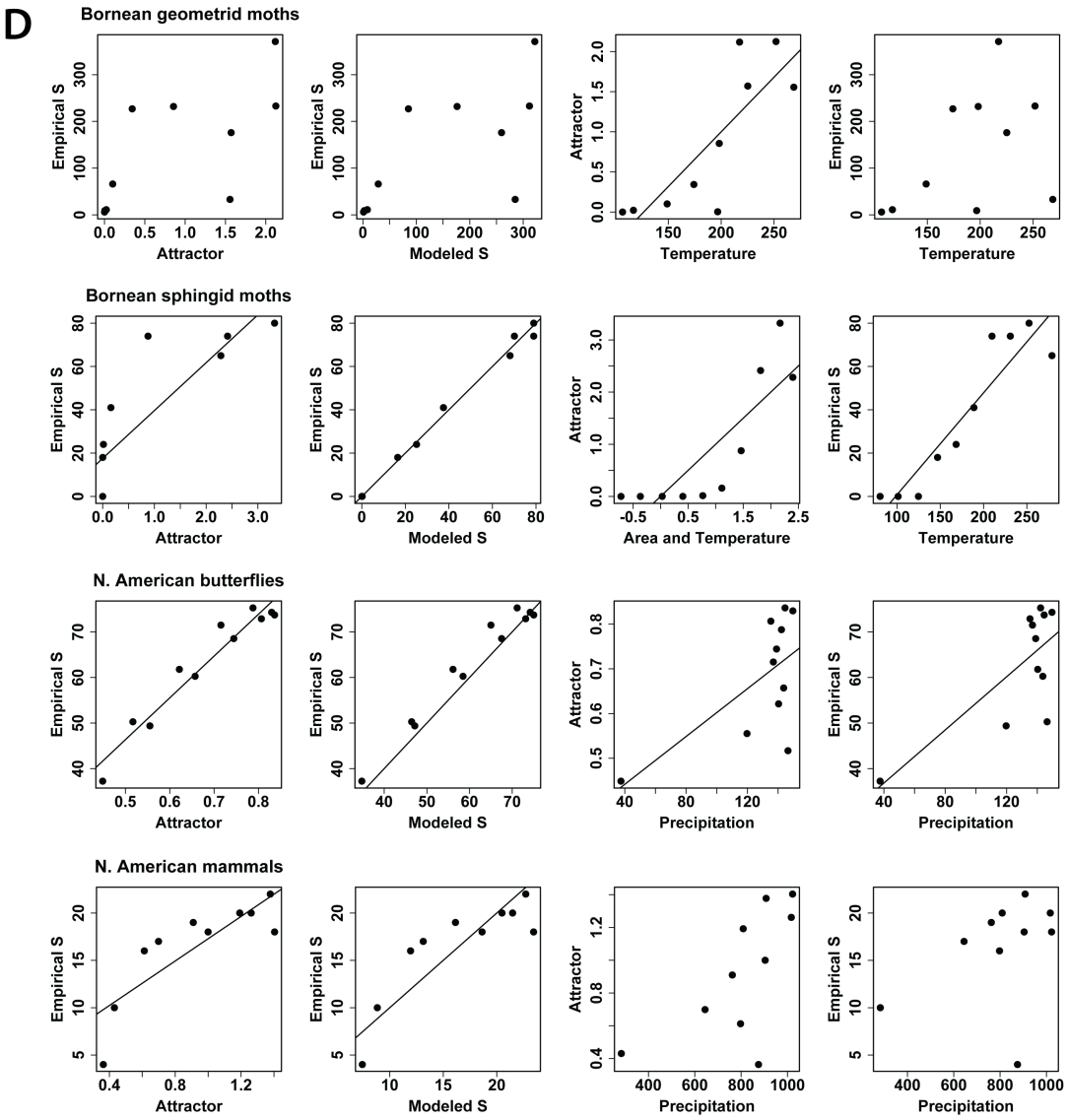


Fig. S1, A-D. Relationships between the modeled attractor, simulated species richness, empirical species richness, and measured environmental variables for each of the 16 datasets (in four geographical groups). Each dataset is represented by the four panels in a row. Within a panel, each point represents one of 9 or 10 elevations within the (rescaled) domain at which variables were evaluated. ***First panel:*** the regression of empirical richness *vs.* the magnitude of the modeled midpoint attractor function. ***Second panel:*** unity-line regression (*slope* = 1, Romdal *et al.* 2005) of modeled richness *vs.* empirical richness. ***Third panel*:** regression of the magnitude of the modeled midpoint attractor function *vs.* the best-fitting (by AIC) environmental variables. ***Fourth panel:*** the regression of empirical species richness *vs.* the best-fitting (by AIC) environmental variables. See Table S2 for statistical results. As explained in the caption for Table S1, the regressions plotted in this figure cannot be assessed for statistical significance, because the points are not independent. Nevertheless, R^2 is as an appropriate measure of

linear goodness-of-fit between variables, sensitive both to linearity and scatter. In the plots here, we have set an arbitrary lower threshold of $R^2 = 0.5$ for display of regression lines.

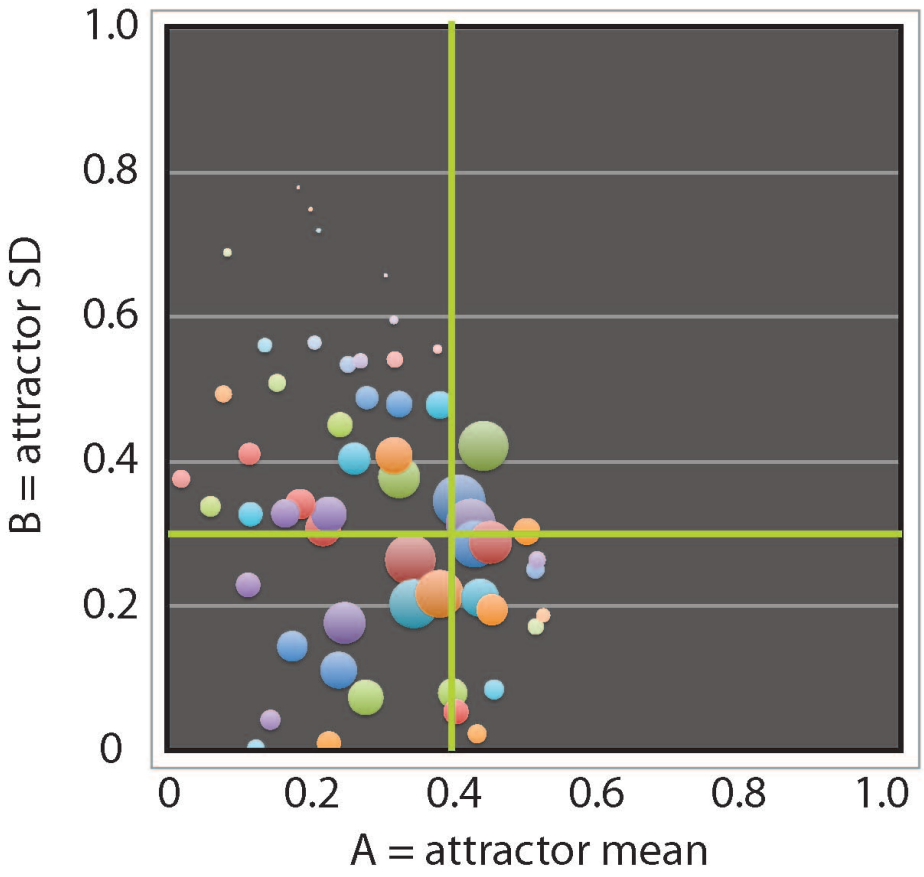


Fig. S2. Sampled (*A*, *B*) pairs of midpoint attractor parameters generated by the MCMC Gibbs sampler for the Costa Rican arctiine moth dataset. Point width is proportional to the coefficient of determination (R^2) between modeled and observed species richness across the elevational domain. Point color is arbitrary. The green lines indicate an optimized pair of parameter values ($A = 0.378$, $B = 0.294$), centered in the cluster of coordinate pairs with highest R^2, which was used to produce the model for the arctiine moth dataset in Fig. 3 (main text).

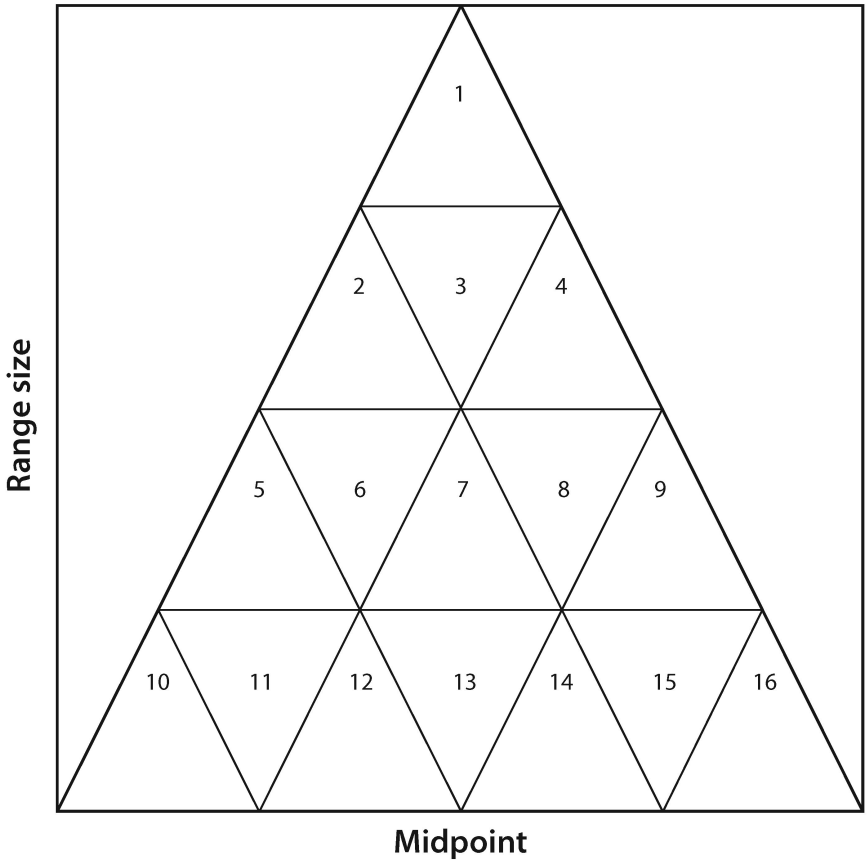


Fig. S3. The geometric constraint triangle, subdivided into 16 smaller, equal-sized triangles.

APPENDIX 2: SUPPLEMENTAL TEXT

SUPPLEMENTAL INTRODUCTION

Beginning with Lees *et al.* (1999) and Jetz and Rahbek (2001), many authors have taken a statistical approach to integrating geometric constraints with environmental variables, treating "pure" MDE model predictions as candidate predictor variables. In most of these studies, the observed range-size frequency distribution (RSDF) was sampled without replacement to generate the MDE model predictions of expected species richness at each location in the domain (Colwell *et al.* 2004, 2005). The MDE predictions and standard environmental variables were then used together in traditional correlative modeling of species richness patterns. Increasingly rigorous versions of this statistical approach have incorporated formal model selection, spatial statistics, and assessment of multicollinearity (Bellwood *et al.* 2005; Davies *et al.* 2007; Wu *et al.* 2012).

Several studies have integrated constraints and drivers directly, incorporating the interacting effects of geometric constraints and environmental drivers on species richness (Gotelli *et al.* 2009), using environmental variables to condition probabilities of range placement and expansion within a spatially bounded domain (Storch *et al.* 2006; Rahbek *et al.* 2007), thus relaxing the assumption of a pure MDE model that all parts of the domain are environmentally identical. These models were also conditioned on the empirical range size frequency distribution (RSFD). In contrast, Grytnes et al. (2008) modeled plant species richness on a bounded elevational gradient by drawing range sizes from theoretical distributions and range midpoints from a probability distribution fitted directly to the observed richness gradient.

Rangel and Diniz-Filho (2005) built a stochastic, mechanistic model that integrates speciation, range expansion, and extinction on a bounded, monotonic environmental "favorability" gradient, without reference to empirical data. The model is effectively a spatially explicit version of the neutral model (Hubbell 2001) in a one-dimensional bounded domain, but with an underlying environmental gradient. The Rangel and Diniz-Filho (2005) model generated off-center species richness peaks that emerged from the interaction between the gradient and the geometric constraints (Colwell & Rangel 2009). Without the environmental gradient—or with a very weak gradient—Rangel and Diniz-Filho's model generated a peak of species richness in the center of the domain that was qualitatively similar to the predictions of a simple MDE model.

Wang and Fang (2012) developed a third approach. They fitted a multiple regression model of species richness as a response to environmental variables, but they used only the subset of species with the smallest geographic ranges to parameterize the model. They reasoned that the placement of small-ranged species within a bounded domain is little affected by the location of range boundaries, so that, for this subset of taxa, correlations between species richness and environmental variables would not be distorted by geometric constraints. They then used the resulting model coefficients, together with the empirical RSFD, to simulate the placement of range midpoints of the larger-ranged species within the bounded domain. They showed that a single

environmental model, combined with strong geometric constraints, best explains the species richness of both small- and large-ranged plant species along elevational gradients in China.

SUPPLEMENTAL MATERIALS AND METHODS

Dataset selection and preparation

As a criterion for inclusion in this study, we applied the rule (McCain 2007; McCain 2009) that at least 70% of the physical gradient between sea level and mountaintop must have been sampled and at least four environmental variables had been reported for the gradient.

Each of the 16 datasets (Table S1) was prepared in the same way. Domain limits were defined as sea level and the highest elevation on the mountain massif upon which the gradient was located. This domain was converted to the unit line, and all empirical sampling elevations were proportionally scaled within this [0,1] domain. Environmental variables (Table S1) were resampled, as necessary, after smoothing with cubic spline interpolation, using the *splinefun* function in R, version 3.1.1 (R Core Team 2014).

If the highest elevation at which a species was recorded was not at the highest sampling location, the upper boundary for that species range was estimated to occur halfway between the highest elevation of recorded occurrence and the next higher sampling elevation. If the highest elevation at which a species was recorded at was the highest sampling elevation, the upper boundary of that species range was estimated to occur halfway between that sampling elevation and the upper limit of the domain. The lower boundary for each range was treated analogously, being extended halfway to the next lower sampling elevation or halfway to the lower domain limit (sea level), if a species was recorded at the lowest sampling elevation, but that sampling elevation was not the domain limit. The ranges of each species found at only one sampling elevation were treated similarly; otherwise, these point ranges would have had a zero range, and would have been lost from the model. We assumed that the occurrence of each species was continuous between its estimated upper and lower recorded range boundaries. These range-adjustment procedures and assumptions have been widely used in previous studies (e.g., Cardelús *et al.* 2006; Longino *et al.* 2014).

The protocol for range adjustment, described above, leaves most datasets without any empirical ranges that actually reach the domain boundaries, resulting in zero estimated empirical richness at one or both limits of the domain. A few zeroes are real (e.g., ants do not occur at very high elevations in the Costa Rica and New Guinea gradients), but most others are artifacts of the location of original sampling elevations and the range estimation protocol. Data providers (Table S1) were asked in each case whether such zeroes in their data sets were real or artifactual. If real, zero richness at the domain endpoint (and in some cases adjacent sampling points) was plotted and included in analyses; if artifactual, we proportionally adjusted all empirical range midpoints so that ranges nearest to the domain limit exactly reached it. The shifts needed to achieve this

adjustment, which effectively shifts the domain boundary slightly, were consistently very small (0.002 to 0.02 on the unit line).

To cope with the wide variation among datasets in number and spacing (often not uniform) of empirical sampling points, we took a mixed approach. For fitting the attractor (see below), we used a series of 11 evenly spaced sampling locations across the entire unit line (domain), including both ends of the domain (0 and 1), for all datasets except the New Guinea group. The New Guinea transect was sampled in the field at 8 evenly-spaced elevations, so with the domain ends added, we used 10 sampling points for fitting the attractor in those datasets. For plotting model results (main text Figs. 3, 4, and 5), we used the original sampling points for datasets with fewer than 11 original points (eight points for the four Papua New Guinea datasets, five for the three Australia datasets, and six for North American butterflies), and 11 points for all other datasets.

The Bayesian Midpoint Attractor model

The MCMC sampler and richness pattern simulation. We designed a simple MCMC Gibbs sampler (Gelman *et al.* 2013) to select (A, B) pairs for the mean (A) and standard deviation (B), the parameters of the Gaussian midpoint attractor, with the objective of simulating the richness pattern over the domain for a particular empirical dataset, using *only* the range-size frequency distribution (RSFD) as input. Empirical midpoints were completely ignored for the simulations. The goodness of fit between modeled and empirical richness was then assessed for each simulation, as detailed below.

Running the simulation. For each candidate (A, B) pair, each empirical range was placed stochastically on the domain, without replacement, using either Algorithm 1 or 2 (*Main text, Materials and Methods*). The modeled richness was recorded for L (10 or 11, see above) evenly spaced sampling locations across the domain, always including both ends of the domain (0 and 1). The process was repeated M (= 100) times, for the same (A, B) pair. The mean richness for each of the L sampling points on the domain was then computed, among the M runs, to estimate the expected richness pattern, given the (A, B) pair and the empirical RSFD.

Measuring goodness-of-fit. The next step in the MCMC procedure assessed the goodness-of-fit (GOF) between the empirical richness pattern and the mean modeled richness pattern, for a given candidate (A, B) pair, at the L sampling points. We applied three alternative GOF measures: (1) r, the Pearson product-moment correlation coefficient (but only when positive), squared; (2) the chi-squared statistic computed on standardized richness (the richness at each sampling point, divided by total richness at all L points), treating the empirical richness as "expected" and the modeled richness as "observed" (as is customary in Bayesian modeling); and (3) the two-sample Kolmogorov-Smirnov (K-S) statistic. Note that none of these measures can be used in this way to yield a probability test of significance; they are simply mathematically suitable measures of GOF for richness patterns. The protocol for choosing the best GOF for each dataset is described, in context, in the next section.

Sampling the parameter space. Using the procedure just described, the MCMC sampler tested a series of (A, B) pairs. At each step in this process, a candidate (A, B) pair was

proposed by drawing a new value for A and a new value for B from uniform distributions $[0 < A < 1]$ and $[0 < B < 1]$. In Bayesian terms, A was a flat prior, with the full [0,1] domain sampled uniformly for the location of the mean (A). For the standard deviation (B), we also set the upper limit at 1 because this value produces a spatial pattern of richness broader and flatter than any empirical richness pattern we have seen; thus the prior distribution of B incorporated this information. (An even higher limit for B could have been used, but the results would not have changed.)

The candidate (A, B) pair was evaluated by running the simulation (M times) and assessing goodness-of-fit (GOF) between the mean modeled richness (averaged among M runs) and empirical richness (as described above). If the GOF for the candidate (A, B) pair was better, or not much worse, than the GOF for the previous pair, the new pair was added to the chain and the process repeated. The criterion for "not much worse" is important. If only parameter sets (A, B pairs) that yield a better fit than the previous step are kept, the chain may become stuck on a local GOF "peak" in the parameter space, and fail to detect a higher peak nearby.

The criterion for accepting a candidate (A, B) pair in our model was the *threshold-for-acceptance ratio T*, between the GOF of the candidate (A, B) pair and the GOF of the previous (A, B) pair in the chain. The ratio T was compared to a uniform random number on the interval [0,1] (Gelman *et al.* 2013). If T was greater than this number, the candidate (A, B) pair was accepted and the chain continued; if T was smaller than this number, the candidate pair was rejected, and a new candidate pair was proposed. In this way, better pairs ($T > 1$) were always accepted, and some not-as-good pairs ($T < 1$) were also accepted, ensuring a better sampling of the parameter space.

For each dataset, $C = 200$ to 500 candidate pairs were tried, and the accepted (A, B) pairs (the chain) were tabulated, each with its GOF and step number in the chain. When the process was complete, the accepted (A, B) pairs were plotted (Fig. S2), and ranked by their GOF (largest to smallest for Pearson and Kolmogorov-Smirnov GOFs, smallest to largest for the chi-squared GOF).

For each dataset, when results differed substantially between the two stochastic range placement algorithms in the Bayesian attractor model (*Main text, Materials and Methods*), GOF measures were used to choose the better of the two algorithms. When results differed substantially among GOF measures for a given algorithm for a particular dataset, choice of GOF was based on minimizing overall deviation of empirical points from the 95% confidence intervals of the model. On the basis of this procedure, Pearson correlation emerged as the most successful GOF (13 of 16 datasets), with chi-squared providing a better result in two cases (Australian leaf-miners and Bornean geometrid moths), and Kolmogorov-Smirnov in one case (North American mammals).

Midpoint predictor models

For each of the two midpoint predictor models, we assessed the same set of environmental variables used to interpret modeled attractors in the Bayesian midpoint attractor model (Table S1), one variable at a time. To construct the probability density functions, the [0,1] domain was divided into 1000 bins. For each bin, the magnitude of the environmental variable was approximated by linear interpolation between measured values at sampling locations on the elevational gradients (Table S1). Next, probabilities

for each bin were assigned proportional to these measured values. Finally, a range midpoint representing each empirical species was placed stochastically in the domain in proportion to these values. For Model 1, no geometric constraints were enforced. In Model 2, range placement was constrained by the domain boundaries.

Midpoint predictor model evaluation. For each midpoint predictor model, we calculated the cumulative distribution function (cdf) of species range midpoints across the domain, averaged over 1000 simulations. Steeply rising sections of this cdf indicate elevations with a high concentration of species range midpoints, whereas flatter sections of the cdf indicate elevations where few or no species range midpoints occur. We refer to this averaged cdf as the *model reference cdf*.

We next constructed the cdf for the empirical midpoint data and calculated the maximum difference between this curve and the model reference cdf. This difference is the traditional Kolmogorov-Smirnov test statistic. To generate a null distribution and estimate the tail probability for the empirical data, we generated 1000 additional midpoint distributions with the midpoint predictor model, and for each of these we calculated the K-S test statistic between the cdf of the single simulated midpoint distribution and the model reference cdf.

We then compared the histogram of the 1000 simulated K-S differences with the observed K-S difference between the empirical data and the model reference cdf. A non-significant one-tailed value ($P > 0.05$) indicates an adequate fit with the data. In contrast, unusually large K-S values for the observed data would suggest that the midpoint predictor model does not successfully reproduce the pattern of midpoints in the data.

Software

The midpoint attractor simulator and the MCMC sampler were implemented in 4th Dimension, in an extension of the RangeModel application (Colwell 2008) that is available from the authors. The midpoint predictor models were programmed in R version 3.1.1 (R Core Team 2014), with base functions from the EcoSimR development package (https://github.com/GotelliLab/EcoSimR), which is available as an R package. R scripts for the midpoint predictor model analyses and for plotting the graphics for the midpoint attractor models (main text Figs. 3, 4, and 5) are available from the authors.

REFERENCES FOR THE APPENDICES

Ashton, L.A., Odell, E.H., Burwell, C.J., Maunsell, S.C., Nakamura, A., McDonald, W.J.F. *et al.* (2015). Altitudinal patterns of moth diversity in tropical and subtropical Australian rainforests. *Austral Ecol.*, 41, 197–208,.

Beck, J., Holloway, J.D., Khen, C.V. & Kitching, I.J. (2012). Diversity partitioning confirms the importance of beta components in tropical rainforest Lepidoptera. *Am. Nat.*, 180, E64-E74.

Beck, J. & Kitching, I.J. (2009). Drivers of moth species richness on tropical altitudinal gradients: a cross-regional comparison. *Global Ecol. Biogeogr.*, 18, 361-371.

Bellwood, D., Hughes, T., Connolly, S. & Tanner, J. (2005). Environmental and geometric constraints on Indo-Pacific coral reef biodiversity. *Ecol. Lett.*, 8, 643-651.

Brehm, G., Colwell, R.K. & Kluge, J. (2007). The role of environment and mid-domain effect on moth species richness along a tropical elevational gradient. *Global Ecology & Biogeography*, 16, 205-219.

Cardelús, C.L., Colwell, R.K. & Watkins, J.E. (2006). Vascular epiphyte distribution patterns: explaining the mid-elevation richness peak. *J. Ecology*, 94, 144-156.

Colwell, R.K. (2008). RangeModel: Tools for exploring and assessing geometric constraints on species richness (the mid-domain effect) along transects. *Ecography*, 31, 4-7.

Colwell, R.K., Rahbek, C. & Gotelli, N. (2004). The mid-domain effect and species richness patterns: what have we learned so far? *Am. Nat.*, 163, E1-E23.

Colwell, R.K., Rahbek, C. & Gotelli, N. (2005). The mid-domain effect: there's a baby in the bathwater. *Am. Nat.*, 166, E149–E154.

Colwell, R.K. & Rangel, T.F. (2009). Hutchinson's duality: the once and future niche. *PNAS* 106, 19651-19658.

Davies, R.G., Orme, C.D.L., Storch, D., Olson, V.A., Thomas, G.H., Ross, S.G. *et al.* (2007). Topography, energy and the global distribution of bird species richness. *Proc. R. Soc. B-Biol. Sci.*, 274, 1189-1197.

Forister, M.L., McCall, A.C., Sanders, N.J., Fordyce, J.A., Thorne, J.H., O'Brien, J. *et al.* (2010). Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. *Proc. Natl. Acad. Sci. U. S. A.*, 107, 2088-2092.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2013). *Bayesian data analysis*. CRC press.

Gotelli, N., Anderson, M.J., Arita, H.T., Chao, A., Colwell, R.K., Connolly, S.R. *et al.* (2009). Patterns and causes of species richness: a general simulation model for macroecology. *Ecol. Lett.*, 12, 873-886.

Grinnell, J. & Storer, T.I. (1924). *Animal life in the Yosemite: an account of the mammals, birds, reptiles, and amphibians in a cross-section of the Sierra Nevada*. University of California Press.

Grytnes, J.A., Beaman, J.H., Romdal, T.S. & Rahbek, C. (2008). The mid-domain effect matters: simulation analyses of range-size distribution data from Mount Kinabalu, Borneo. *J. Biogeogr.*, 35, 2138-2147.

Hubbell, S.P. (2001). *The unified theory of biodiversity and biogeography*. Princeton University Press, Princeton, N. J.

Jetz, W. & Rahbek, C. (2001). Geometric constraints explain much of the species richness pattern in African birds. *PNAS*, 98, 5661-5666.

Kluge, J., Kessler, M. & Dunn, R.R. (2006). What drives elevational patterns of diversity? A test of geometric constraints, climate and species pool effects for pteridophytes on an elevational gradient in Costa Rica. *Global Ecology & Biogeography*, 15, 358-371.

Lees, D.C., Kremen, C. & Andriamampianina, L. (1999). A null model for species richness gradients: bounded range overlap of butterflies and other rainforest endemics in Madagascar. *Biol. J. Linn. Soc.*, 67, 529-584.

Longino, J.T., Branstetter, M.G. & Colwell, R.K. (2014). How Ants Drop Out: Ant abundance on tropical mountains. *PLoS One*, 9, e104030.

Longino, J.T. & Colwell, R.K. (2011). Density compensation, species composition, and richness of ants on a neotropical elevational gradient. *Ecosphere*, 2(3):art29, doi:10.1890/ES10-00200.1.

Maunsell, S.C., Burwell, C.J., Morris, R.J., McDonald, W.J., Edwards, T., Oberprieler, R. *et al.* (2016). Elevational turnover in the composition of leaf miners and their interactions with host plants in Australian subtropical rainforest. *Austral Ecology*, 41, 238-247.

Maunsell, S.C., Kitching, R.L., Burwell, C.J. & Morris, R.J. (2015). Changes in host–parasitoid food web structure with elevation. *J. Anim. Ecol.*, 84, 353-363.

McCain, C. (2005). Elevational gradients in diversity of small mammals. *Ecology*, 86, 366-372.

McCain, C.M. (2004). The mid-domain effect applied to elevational gradients: species richness of small mammals in Costa Rica. *J. Biogeogr.*, 31, 19-31.

McCain, C.M. (2007). Could temperature and water availability drive elevational species richness patterns? A global case study for bats. *Global Ecology & Biogeography*, 16, 1-13.

McCain, C.M. (2009). Vertebrate range sizes indicate that mountains may be 'higher' in the tropics. *Ecol. Lett.*, 12, 550-560.

Moses, J. (2015). Tropical elevational gradient in ants (Hymenoptera: Formicidae): Diversity patterns, food preferences and nutrient redistribution rates on Mt Wilhelm, Papua New Guinea. MSc Thesis, University of Papua New Guinea, Port Moresby., p. 85 pp.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. , URL http://www.R-project.org/.

Rahbek, C., Gotelli, N., Colwell, R.K., Entsminger, G.L., Rangel, T.F.L.V.B. & Graves, G.R. (2007). Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proceedings of the Royal Society of London Series B*, 274, 165-174.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F. (2005). An evolutionary tolerance model explaining spatial patterns in species richness under environmental gradients and geometric constraints. *Ecography*, 28, 253-263.

Romdal, T.S., Colwell, R.K. & Rahbek, C. (2005). The influence of band sum area, domain extent, and range sizes on the latitudinal mid-domain effect. *Ecology*, 86, 235–244.

Sam, K. & Koane, B. (2014). New avian records along the elevational gradient of Mt. Wilhelm, Papua New Guinea. Bulletin of the British Ornithologists' Club 134, 116-133.

Sam, L. (2011). Responses of butterfly (Lepidoptera) communities along an altitudinal forest gradient in Papua New Guinea. MSc Thesis, University of Papua New Guinea, Port Moresby, p. 66 pp.

Storch, D., Davies, R.G., Zajicek, S., Orme, C.D.L., Olson, V., Thomas, G.H. *et al.* (2006). Energy, range dynamics and global species richness patterns: reconciling mid-domain effects and environmental determinants of avian diversity. *Ecol. Lett.*, 9, 1308-1320.

Tvardikova, K. (2013). Trophic relationships between insectivorous birds and insect in Papua New Guinea. Thesis Series, No. 9. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, p. 184 pp.

Wang, X. & Fang, J. (2012). Constraining null models with environmental gradients: a new method for evaluating the effects of environmental factors and geometric constraints on geographic diversity patterns. *Ecography*, 35, 1147–1159.

Wu, Y., Yang, Q., Wen, Z., Xia, L., Zhang, Q. & Zhou, H. (2012). What drives the species richness patterns of non-volant small mammals along a subtropical elevational gradient? *Ecography*, 36, 185-196.