Ecography

Supplementary material

Appendix 1: Supplemental tables and figures

Table A1. Site details.

Sites were sampled for 90 days during the summer, with 20 pairs of pitfall traps at each site, collected multiple times during the 90-day period. Each San Juan (SJ) transect contained one extremely disturbed site which was excluded from the analysis (italicized, gray text), while only one site in the Front Range (FR) was similarly disturbed.

Transect	Year Sampled	Elevation (m)	Latitude (°N)	Longitude (°W)	Ant worker count	Number of collections	Obs. species richness	Obs. genus richness	Obs. subfamily richness
SJ-1	2012	1796	37.30	-108.67	8,328	5	27	15	4
SJ-1	2012	1970	37.24	-108.30	4,034	5	43	16	4
SJ-1	2012	2349	37.47	-108.39	8,442	5	42	14	3
SJ-1	2012	2580	37.52	-108.31	8,545	5	32	9	3
SJ-1	2012	2887	37.54	-108.22	1,460	5	16	7	3
SJ-1	*2012*	*3241*	*37.55*	*-108.10*	*208*	*2*	*7*	*6*	*2*
SJ-1	2012	3386	37.72	-107.93	82	5	3	3	2
SJ-1	2012	3508	37.71	-107.91	1,321	4	6	4	2
SJ-2	2011	1493	37.33	-109.03	10,000	4	34	16	4
SJ-2	2011	1717	37.43	-109.02	9,477	4	27	14	5
SJ-2	2011	2240	37.55	-108.46	15,014	4	44	15	3
SJ-2	2011	2528	37.69	-108.33	5,783	3	37	15	3
SJ-2	2011	3038	37.80	-108.12	3,992	3	21	7	3
SJ-2	2011	3228	37.82	-108.13	4,142	3	15	6	3
SJ-2	2012	3390	37.84	-108.03	339	5	3	3	2
SJ-2	*2012*	*3517*	*37.84*	*-108.01*	*---*	*---*	*---*	*---*	*---*
FR-1	2010	1728	40.43	-105.23	3,646	3	43	15	4
FR-1	2011	1938	40.46	-105.28	8,006	4	46	14	3
FR-1	2010	2153	40.45	-105.30	5,690	3	37	10	3
FR-1	2011	2410	40.42	-105.50	3,363	5	27	8	3
FR-1	2011	2811	40.40	-105.64	619	2	13	7	3
FR-1	2011	3022	40.38	-105.66	6,590	5	16	7	3
FR-1	2012	3364	40.39	-105.67	6,273	6	15	5	2
FR-1	2011	3640	40.40	-105.70	186	4	2	1	1
FR-2	2010	1811	40.02	-105.30	441	2	17	12	3
FR-2	2010	1904	40.01	-105.33	886	2	16	11	3
FR-2	2010	2215	40.01	-105.37	4,853	1	35	10	3
FR-2	2010	2712	40.02	-105.48	6,199	2	34	7	3
FR-2	2010	2881	40.02	-105.53	1,968	2	24	6	3
FR-2	2010	3117	40.03	-105.55	728	2	12	5	2
FR-2	2010	3466	40.05	-105.58	4,424	3	3	3	2
FR-2	*2010*	*3659*	*40.05*	*-105.61*	*---*	*---*	*---*	*---*	*---*

Table A2: Parameter prior distributions.

Prior distributions for each parameter. See main text (Methods: Model structure: Phylogenetic effects) for the phylogenetic structuring of genus-level and species-level parameters. Priors were weakly informative, primarily just constraining the range of each parameter to realistic values. In the normal distributions, $\tau = 1/\sigma^2$, following the use of *precision* by JAGS.

Parameter	Prior Distribution
Aggregate slopes & intercepts	$\beta_x \sim \text{Normal}(\mu = 0,\ \tau = 0.01)$
Pagel's λ	$\lambda_x \sim \text{Uniform}(\text{min} = 0,\ \text{max} = 1)$
Intragenus standard deviation	$e_x \sim \text{Uniform}(\text{min} = 0,\ \text{max} = 4)$
Detection probability	$\delta_s \sim \text{Beta}(d_1,\ d_2)$ $d_x \sim \text{Gamma}(1,1)$
Site effect on trap	$\rho \sim \text{Normal}(\mu = 0,\ \tau = 0.01)$

Table A3: Aggregate posterior distribution summaries.

Aggregate slopes represent the overall response of the regional ant community to each variable. Posterior distribution means, medians, 95% highest posterior density intervals (HPDIs), and 90% HPDIs for all aggregate parameters are shown below for the full model (m_{full}) and the optimal model (m_{opt}). The models were parameterized using the two San Juan transects (*SJ-1*, *SJ-2*), and model selection was performed using the Front Range transects (*FR-1*, *FR-2*) as novel transects to compare the out-of-sample predictive ability of each model. Bolded intervals indicate posterior intervals that do not include 0, indicating a non-zero estimate of the slope with 95% confidence.

| Parameter	Model	Mean	Median	95% HPDI	90% HPDI
$\beta_{H'(LandCover)}$	m_{full}	0.08	0.08	-0.11 – 0.24	-0.08 – 0.22
	m_{opt}	---	---	---	---
$\beta_{log(Area)}$	m_{full}	-0.19	-0.19	-0.38 – 0.02	**-0.36 – -0.02**
	m_{opt}	---	---	---	---
$\beta_{H'(SoilType)}$	m_{full}	-0.27	-0.26	**-0.52 – -0.03**	**-0.48 – -0.07**
	m_{opt}	---	---	---	---
$\beta_{TempBreadth}$	m_{full}	-2.87	-2.86	**-3.66 – -2.09**	**-3.50 – -2.19**
	m_{opt}	-2.35	-2.35	**-3.16 – -1.54**	**-2.99 – -1.60**
β_{NPP}	m_{full}	0.59	0.58	**0.24 – 1.04**	**0.26 – 0.90**
	m_{opt}	---	---	---	---
$\beta_{DailyTempVar}$	m_{full}	0.37	0.36	**0.003 – 0.77**	**0.05 – 0.70**
	m_{opt}	---	---	---	---
$\beta_{CanopyCover}$	m_{full}	-0.31	-0.31	**-0.57 – -0.05**	**-0.52 – -0.09**
	m_{opt}	-0.36	-0.35	**-0.61 – -0.14**	**-0.56 – -0.16**
β_{nTrees}	m_{full}	-0.07	-0.07	-0.30 – 0.16	-0.26 – 0.13
	m_{opt}	---	---	---	---
$\beta_{VegBiomass}$	m_{full}	-0.23	-0.22	-0.48 – 0.02	**-0.44 – -0.03**
	m_{opt}	---	---	---	---
$\beta_{BareGround}$	m_{full}	-0.10	-0.10	-0.30 – 0.08	-0.26 – 0.06
	m_{opt}	---	---	---	---
ω	m_{full}	1.09	1.09	**0.89 – 1.27**	**0.94 – 1.27**
	m_{opt}	1.00	1.00	**0.78 – 1.19**	**0.82 – 1.16**
ρ	m_{full}	1.46	1.44	**1.06 – 1.96**	**1.12 – 1.88**
	m_{opt}	1.74	1.66	**1.17 – 2.68**	**1.23 – 2.42**

Table A4: Pagel's λ posterior distribution summaries.

Posterior means, medians, 95% highest posterior density intervals (HPDIs), and 90% HPDIs for all Pagel's λ parameters are shown below for the full model (m_{full}) and the optimal model (m_{opt}). The models were parameterized using the two San Juan transects (*SJ-1*, *SJ-2*), and model selection was performed using the Front Range transects (*FR-1*, *FR-2*) as novel transects to compare the out-of-sample predictive ability of each model. Values range from 0, indicating no phylogenetic covariance among genera, and 1, indicating strong covariance.

Parameter	Model	Mean	Median	95% HPDI	90% HPDI
$\lambda_{TempBreadth}$	m_{full}	0.41	0.35	0.01 – 0.94	0.04 – 0.86
	m_{opt}	0.53	0.53	0.09 – 1.00	0.51 – 0.97
λ_{NPP}	m_{full}	0.55	0.55	0.11 – 1.00	0.15 – 0.99
	m_{opt}	---	---	---	---
$\lambda_{DailyTempVar}$	m_{full}	0.37	0.29	0.00 – 0.92	0.00 – 0.85
	m_{opt}	---	---	---	---
$\lambda_{CanopyCover}$	m_{full}	0.50	0.49	0.02 – 0.95	0.08 – 0.95
	m_{opt}	0.47	0.45	0.04 – 0.95	0.04 – 0.90
λ_{nTrees}	m_{full}	0.58	0.60	0.15 – 1.00	0.22 – 1.00
	m_{opt}	---	---	---	---
$\lambda_{VegBiomass}$	m_{full}	0.51	0.51	0.09 – 0.99	0.12 – 0.98
	m_{opt}	---	---	---	---
$\lambda_{BareGround}$	m_{full}	0.36	0.29	0.03 – 0.89	0.03 – 0.78
	m_{opt}	---	---	---	---
λ_{ω}	m_{full}	0.38	0.31	0.03 – 0.90	0.04 – 0.82
	m_{opt}	0.51	0.51	0.08 – 1.00	0.07 – 0. 49

Figure A1: Genus-level phylogeny for ants in Colorado.

The genus-level phylogeny used to account for phylogenetic non-independence in species' responses. A Newick file of the phylogeny for genera in the family Formicidae (213 genera) was accessed from TimeTree <https://www.timetree.org> on 2018 October 17 and pruned in R with the *ape* package using the function `drop.tip()` to include only genera present in Colorado (Kumar et al. 2017). This pruned phylogeny is shown here. We then calculated the variance-covariance matrix using the `vcv.phylo()` function. The slope estimates for species (e.g., b_x) within each genus were distributed about genus-level means (e.g., B_x), which were then distributed about aggregate means (e.g., β_x) with an error structure allowing for phylogenetic relationships among genera based on the estimated values for Pagel's λ. A species-level phylogeny is not currently available for this species set.
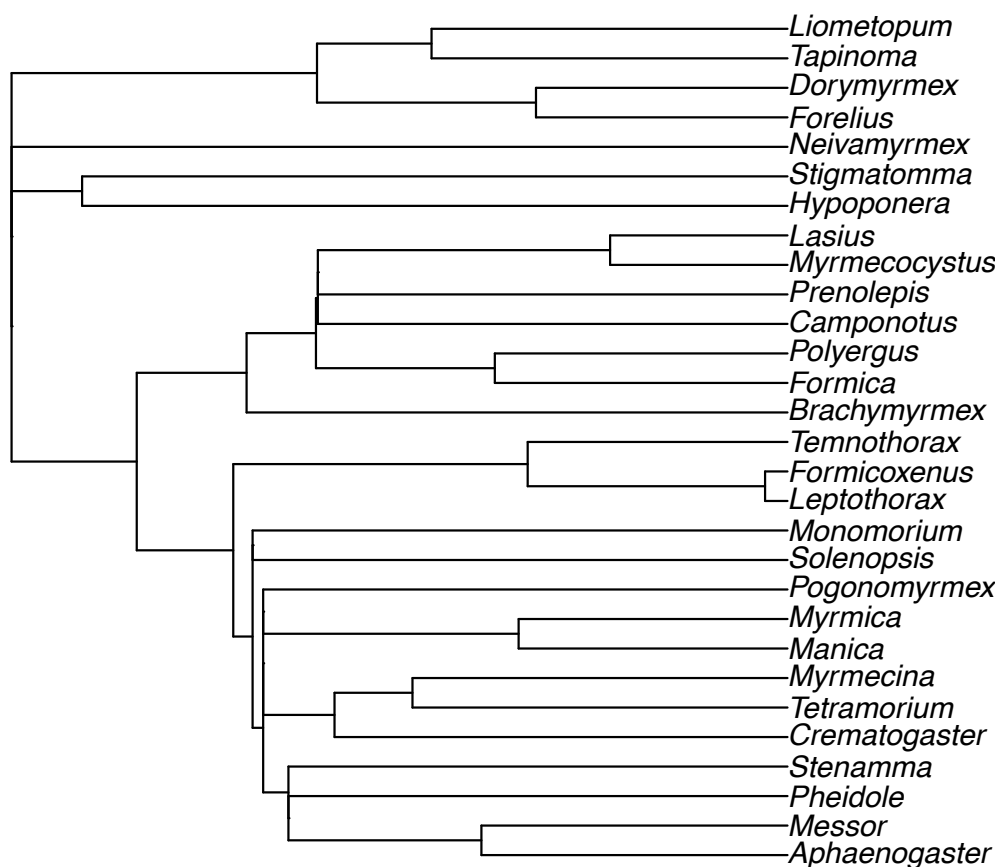
References:
Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol 34 (7): 1812-1819

Figure A2. Species accumulation and rarefaction-extrapolation curves.

In addition to accounting for detection error and comparing observed richness with Fisher's α and *ICE*, we generated sample-based species accumulation curves (Fig. A2a–A2d) and rarefaction-extrapolation curves (Fig. A2e) for each site to evaluate sampling effort. Species accumulation curves were calculated using the R package *vegan*, including the expected accumulation (i.e., the Mao Tau estimate) and the distribution based on 1000 permutations (Colwell et al. 2012). With the R package *iNEXT*, we calculated rarefaction and extrapolation curves using traps as samples. Extrapolated richness was estimated to 40 traps at each site, representing a doubling of the actual trap number. Site details can be found in Appendix A1: Table A1. Sites are ordered and numbered from low (Site 1) to high (Site 8) elevation. The three sites that were excluded due to disturbance are not shown.

References:

Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.Y., Mao, C.X., Chazdon, R.L. & Longino, J.T. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *J. Plant Ecol.* 5: 3–21.

Figure A2a. Species accumulation curves for transect *SJ-1*. Solid lines represent expected accumulation (i.e., the Mao Tau estimate) with the 95% confidence intervals shown in blue. Boxplots show the distribution based on 1000 random permutations of the traps. Note that the y-axis varies with each pane, and that Site 6 was excluded due to heavy disturbance.
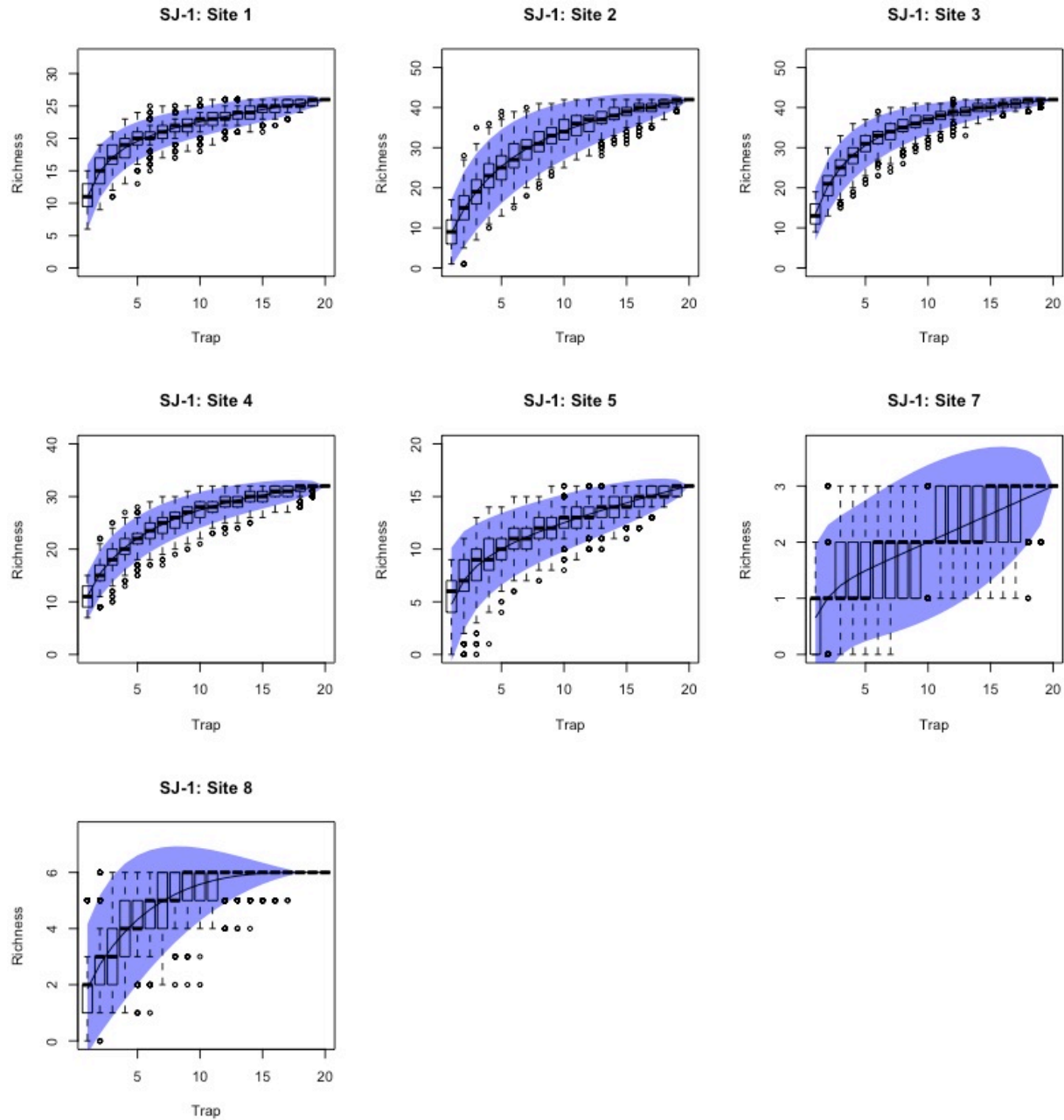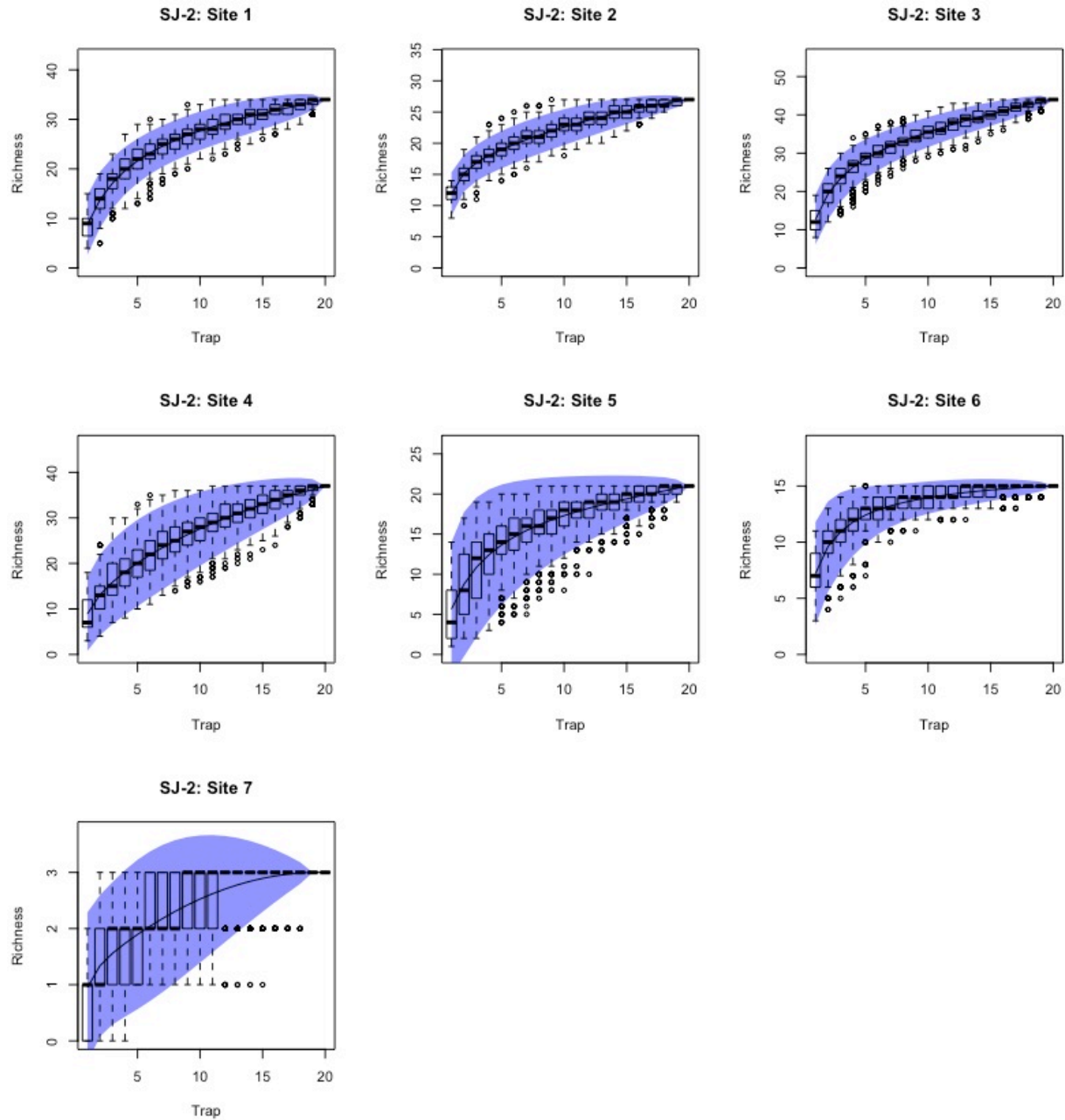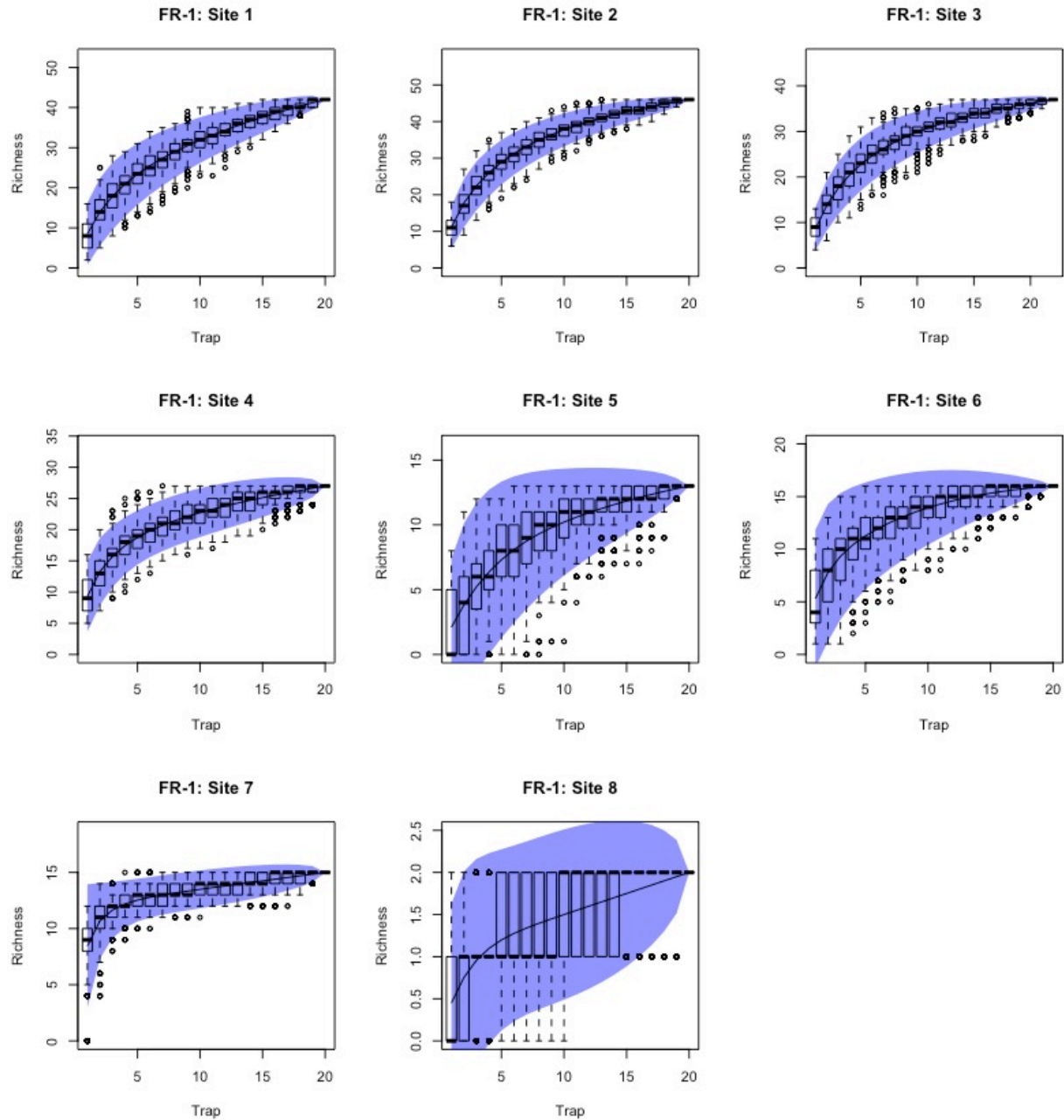
Figure A2b. Species accumulation curves for transect *SJ-2*. Solid lines represent expected accumulation (i.e., the Mao Tau estimate) with the 95% confidence intervals shown in blue. Boxplots show the distribution based on 1000 random permutations of the traps. Note that the y-axis varies with each pane, and that Site 8 was excluded due to heavy disturbance.

Figure A2c. Species accumulation curves for transect *FR-1*. Solid lines represent expected accumulation (i.e., the Mao Tau estimate) with the 95% confidence intervals shown in blue. Boxplots show the distribution based on 1000 random permutations of the traps. Note that the y-axis varies with each pane.

Figure A2d. Species accumulation curves for transect *FR-2*. Solid lines represent expected accumulation (i.e., the Mao Tau estimate) with the 95% confidence intervals shown in blue. Boxplots show the distribution based on 1000 random permutations of the traps. Note that the y-axis varies with each pane, and that Site 8 was excluded due to heavy disturbance.
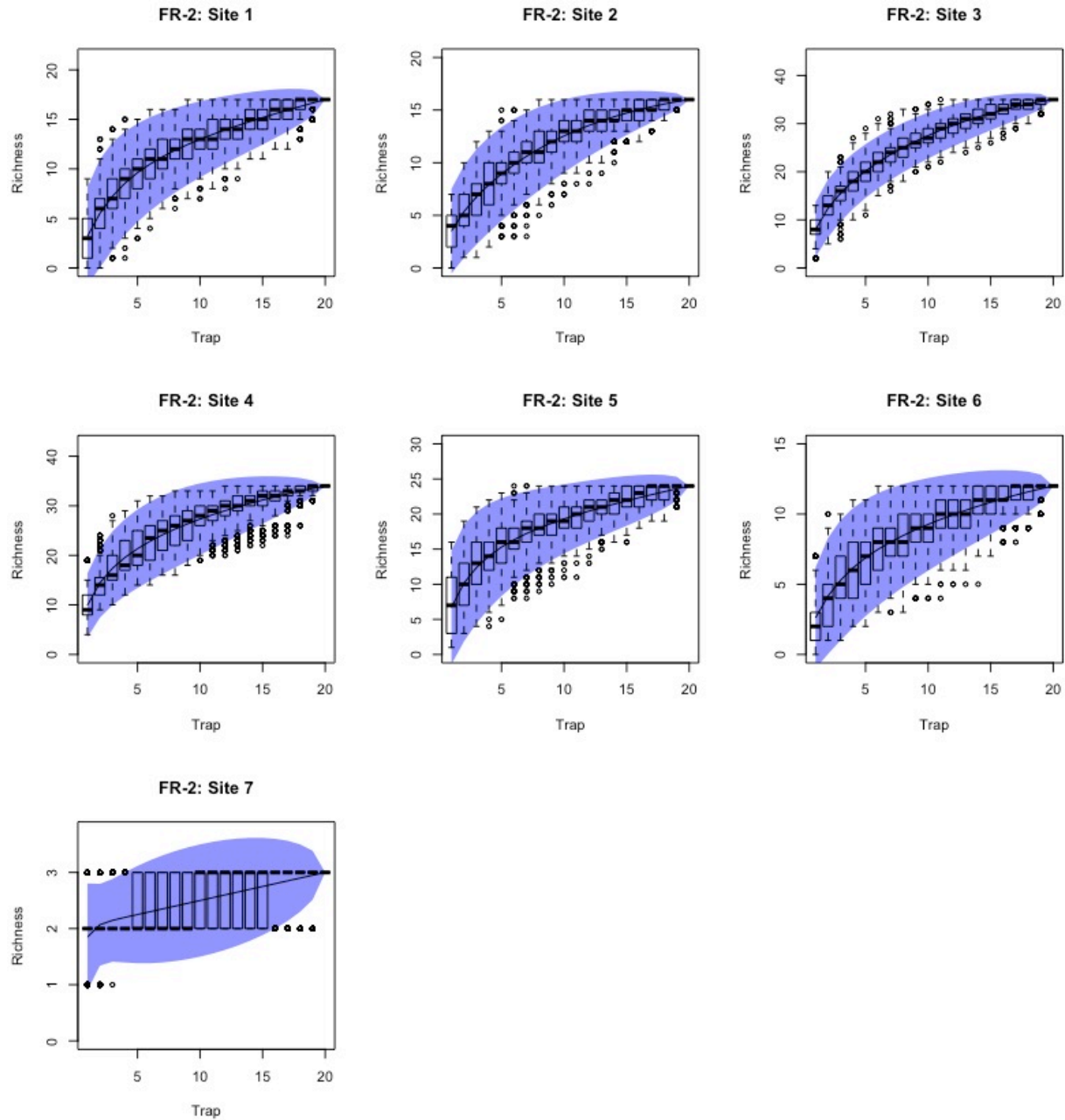
Figure A2e. Rarefaction and extrapolation curves. Lines represent expected interpolated (1-20 traps) and extrapolated (21-40 traps) richness, with observed richness shown as points at 20 traps (dotted vertical line). The shaded regions indicate 95% confidence intervals. Sites that were excluded due to heavy disturbance (*SJ-1* Site 6, *SJ-2* Site 8, and *FR-2* Site 8) are not shown.
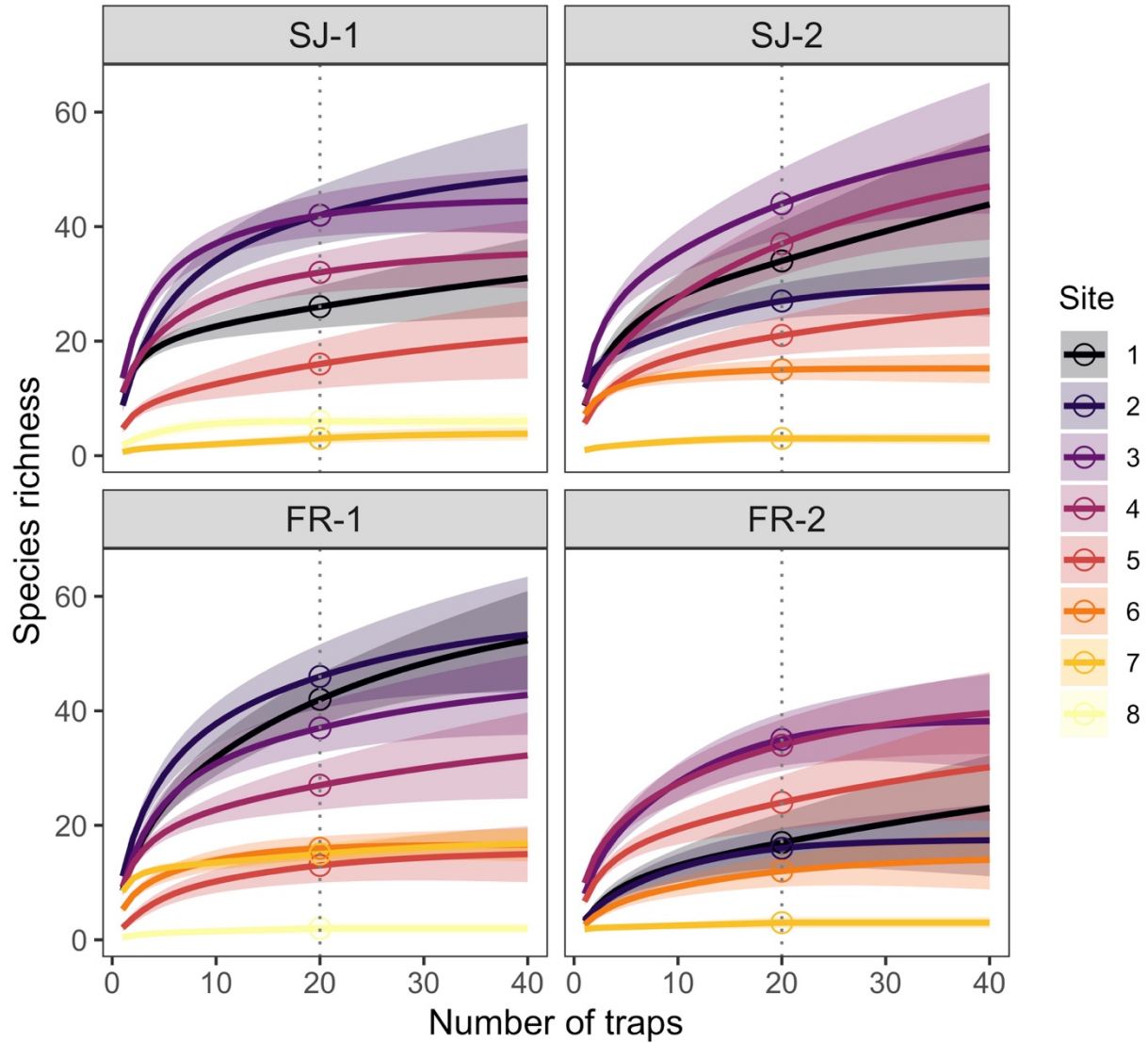
Figure A3: Species-specific posterior distribution medians.

For each variable below, species-specific responses were estimated in addition to genus and aggregate responses. Species within a genus tended to respond similarly. Each point (partly transparent) represents the median of the posterior distribution for each species' response to each variable. Note that species-specific responses were not estimated for variables expected to affect richness more directly (i.e., area, habitat diversity, and soil diversity). Results shown are from the full model, m_{full}, which included all possible covariates.
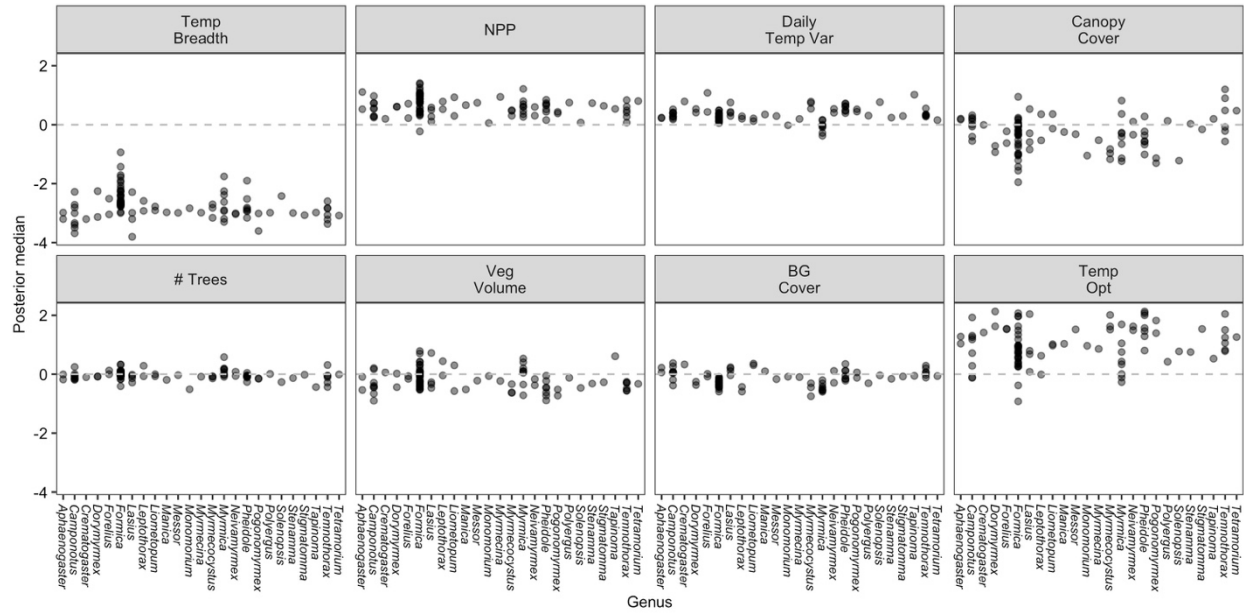
Figure A4. Posterior distributions of Pagel's λ and intragenus standard deviation for species-specific variables in m_{full} and m_{opt}.

(a) Pagel's λ ranges from 0, indicating no influence of the genus phylogeny in response to a covariate, to 1, indicating strong phylogenetic covariance. Generally, there was little phylogenetic covariance at the genus level, with medians almost always below 0.25 (thick bar: median; box: 50% highest posterior density interval (HPDI); capped line ranges: 95% HPDI) in the full model (m_{full}: green) and optimal model by prediction of the novel FR transects (m_{opt}: light blue). (b) Higher values of intragenus standard deviation, e, correspond to more variation among congeners in their response to a variable. The confidence in the estimate varied, as shown by the range of each distribution, though estimates tended to be less than 1 (thick bar: median; box: 50% HPDI; capped line ranges: 95% HPDI).
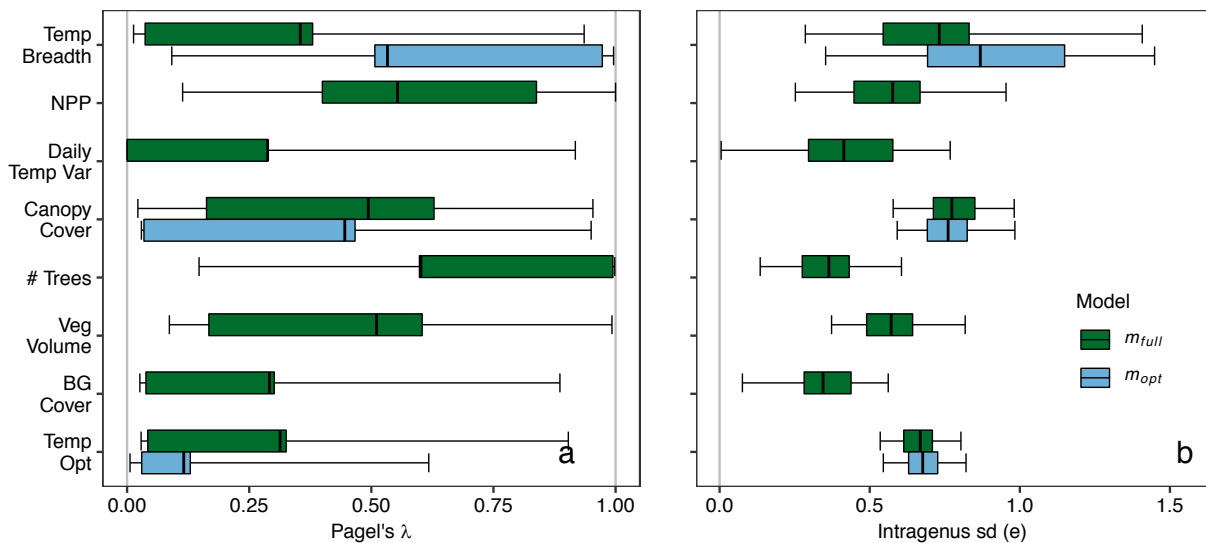
Figure A5: Detection probabilities

In addition to the species-specific detection probabilities, the model estimated the probability distribution from which those detection probabilities were drawn ($\delta_s \sim \text{Beta}(d_1, d_2)$). Each curve (partially transparent) shows the beta distribution defined by the d_1 and d_2 hyperparameter estimates in that iteration. Detection probabilities vary widely among ant species, with a tendency toward lower probabilities, as indicated by the skew. An ant with a detection probability of 0.4, or roughly the mode of this distribution, would be detected 2 out of every 5 pitfall trap collections.
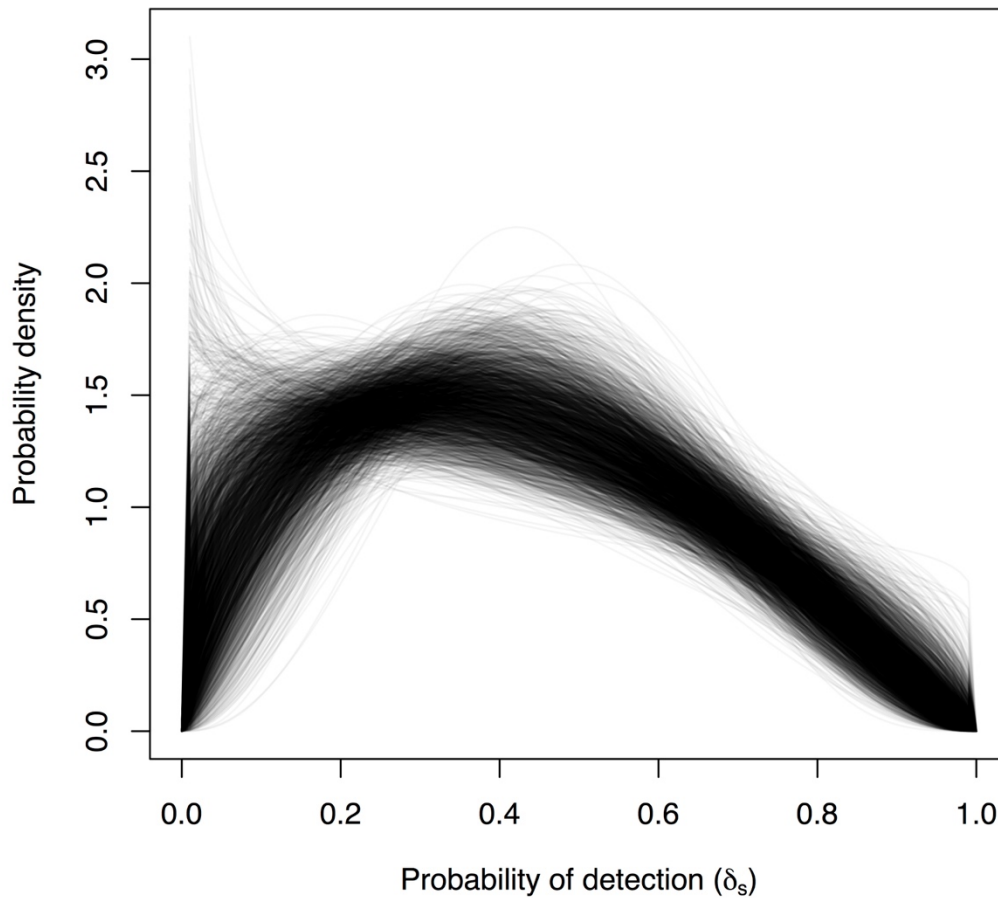
Figure A6: Posterior distribution estimates for all models.

Posterior estimates for aggregate responses (i.e., β) remained generally stable across all candidate models. The median of each posterior distribution is shown by each point, with the line ranges showing the 95% highest posterior density intervals (HPDIs). Models are indicated on the x-axis, with 'mod_full' as the full model (m_{full}: green). The best model based on out-of-sample predictions of the Front Range transects (m_{opt}: light blue) is 'mod_full_minus8'. Note that 'd[1]' and 'd[2]' are the hyperparameters that determine the distribution from which each detection probability, δ_s, is drawn (Appendix A1: Fig. A5). The parameter estimates were generally stable across models, as variables were sequentially removed.
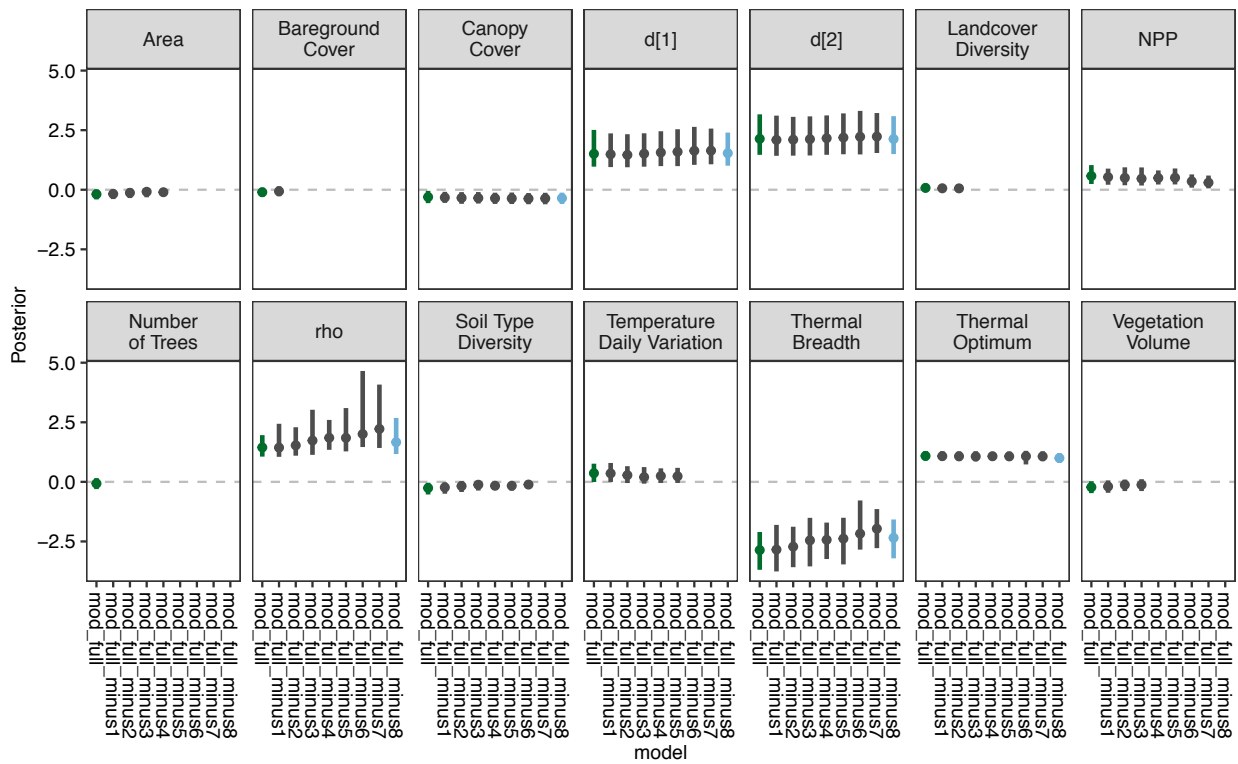
Appendix 2: Model code

This model was written to run in JAGS. The code is for the full model with all possible covariates included. To optimize for clarity in the text and for computational efficiency in the model, there are two naming differences compared to the text of the manuscript: the intercepts (text: a_{1s} and a_{2s}) and thermal optimum (text: ω_s) are treated as additional b_{xs} terms in the model. Where applicable, upper case (PSI, Z) indicates site-level, while lower case (psi, z) indicates trap-level. The following table describes each object in the model code, with those input as data listed in black, and those estimated by the model in blue.

Object Name	Quantity
i	Number of sites
j	Number of traps per site
k	Number of collections per trap
s	Number of possible species
g	Number of possible genera
nbetaScov	Number of site-scale β-only covariates
nbScov	Number of site-scale species-level covariates
nTcov	Number of trap-scale covariates
nCov	Total number of covariates
SiteX	Matrix [i,nbetaScov+nbScov] of site covariates
TrapX	Array [i,j,nTcov] of trap covariates
y	Array [s,i,j] of number of detections
genCov	Genus-level phylogenetic variance-covariance matrix [g,g]
Im	Identity matrix [g,g]
tax	Matrix [s,2] matching each species (1:s) with its genus (1:g)
beta, B, b	Aggregate, genus-level, and species-level slopes
lam	Pagel's λ for each species-level covariate
e	Intra-genus standard deviation for each species-level covariate
delta	Detection probability
d	Hyperparameters for gamma distribution on delta
rho	Effect of site occupancy probability on trap occupancy probability
PSI, psi	Site-scale [s,i] and trap-scale [s,i,j] occupancy probability
Z, z	Site-scale [s,i] and trap-scale [s,i,j] occupancy probability

```
model{
### Priors
  for(x in 1:(nCov+3)) {    # covariates + intercepts + thermal optimum
    beta[x] ~ dnorm(0, 0.01)
  }
# species-specific site, trap, opt, intercept parameters
  for(x in 1:(nbScov+nTcov+3)) {    # excludes 3 covariates with family-only slopes
    lam[x] ~ dunif(0, 1)
    e[x] ~ dunif(0, 4)
    B[1:g,x] ~ dmnorm(rep(beta[x+nbetaScov],g),
                       lam[x]*genCov + (1-lam[x])*Im)
    for(sp in 1:s) {
      b[sp,x] ~ dnorm(B[tax[sp,2],x], 1/(e[x]*e[x]))  # tax: matrix listing each species' genus
    }
  }
# detection
  for(sp in 1:s) {
    delta[sp] ~ dbeta(d[1], d[2])
  }
  d[1] ~ dgamma(1, 1)
  d[2] ~ dgamma(1, 1)
# site effect on trap
  rho ~ dnorm(0,0.01)
### Likelihood
  for(site in 1:i) {
    for(sp in 1:s) {
      logit(PSI[sp,site]) <- beta[1]*SiteX[site,1] +
                             beta[2]*SiteX[site,2] +
                             beta[3]*SiteX[site,3] +
                             b[sp,1]*abs(b[sp,8] - SiteX[site,4]) +
                             b[sp,2]*SiteX[site,5] +
                             b[sp,3]*SiteX[site,6] +
                             b[sp,9]
      Z[sp,site] ~ dbern(PSI[sp,site])
      for(trap in 1:j) {
        logit(psi[sp,site,trap]) <- rho*logit(PSI[sp,site]) +
                               b[sp,4]*TrapX[site,trap,1] +
                               b[sp,5]*TrapX[site,trap,2] +
                               b[sp,6]*TrapX[site,trap,3] +
                               b[sp,7]*TrapX[site,trap,4] +
                               b[sp,10]
        z[sp,site,trap] ~ dbern(psi[sp,site,trap]*Z[sp,site])
        y[sp,site,trap] ~ dbin(z[sp,site,trap]*delta[sp], k[site,trap])
      }
    }
  }
}
```