Title

Opposing macroevolutionary and trait-mediated patterns of threat and naturalization in flowering plants

Opposing patterns of threat and naturalization in flowering plants

Authors

Example: J. P. Schmidt,[1]* T. J. Davies,[2] M. J. Farrell[3]

Affiliations

[1]Odum School of Ecology, University of Georgia
Athens, Georgia 30602
jpschmidt270@gmail.com

[2]Departments of Botany, Forest & Conservation Sciences
Biodiversity Research Centre, University of British Columbia
Vancouver, BC V6T 1Z4, Canada

[3]Department of Ecology and Evolutionary Biology
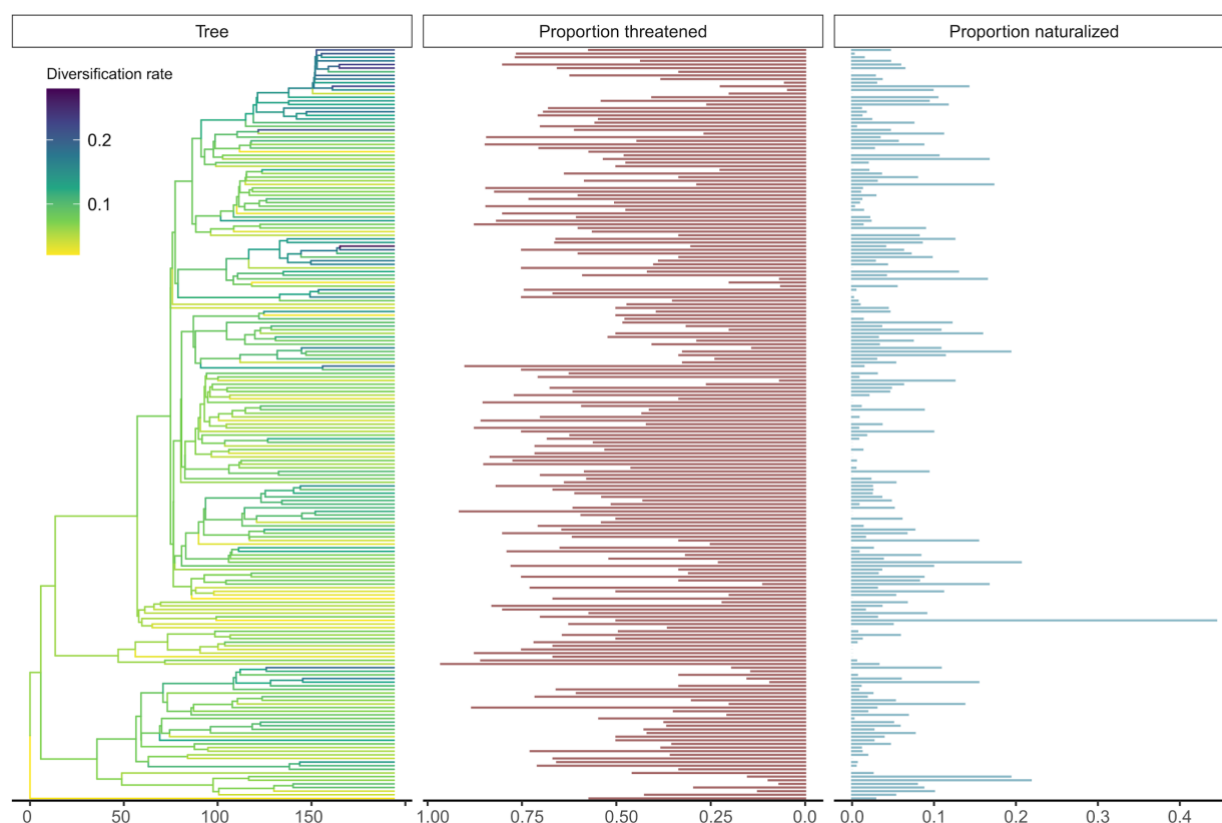University of Toronto
Toronto, ON M5S 3B2, Canada

Abstract

Due to expanding global trade and movement, new plant species are establishing in exotic ranges at increasing rates while the number of native species facing extinction from multiple threats grows. Yet, how species losses and gains globally may together be linked to traits and macroevolutionary processes is poorly understood. Here we show that, adjusting for diversification rate and age, the proportion of threatened species across flowering plant families is negatively related to the proportion of naturalized species. Moreover, naturalization is positively associated with climate variability, short generation time, autonomous seed production, and interspecific hybridization, but negatively with age and diversification; whereas threat is negatively associated with climate variability and hybridization, and positively with biotic pollination, age and diversification. Such a pronounced signature of naturalization and threat across plant families suggests that both trait syndromes have coexisted over deep evolutionary time and that neither strategy is necessarily superior to the other.

Introduction

Ongoing habitat loss, degradation, and fragmentation (Corlett 2016), disruption of historic disturbance regimes (Alstad *et al.* 2016), increased invasion success of alien species (Winter *et al.* 2009), and climate change (Willis *et al.* 2008) are driving plant extinctions at accelerating rates (Alstad *et al.* 2016, Yessoufou & Davies 2016). At the same time, global trade has increased the pace at which alien plants are introduced and become established outside their native ranges (Seebens *et al.* 2017). Yet, while a broad body of work suggests threatened and invasive species contrast sharply in traits and have distinct phylogenetic distributions (Davies *et al.* 2011), our understanding of global patterns of threat and naturalization as potentially interrelated macroecological and macroevolutionary phenomena remains lacking (Vellend *et al.* 2017).

48 Extinction risk at broad scales appears to be strongly structured on plant phylogeny (*e.g.*
49 Vamosi &Wilson 2008, Davies *et al.* 2011, Vamosi *et al.* 2018) with the frequency of
50 threatened angiosperms highest within speciose clades (Schwartz & Simberloff 2001) and
51 generally young, rapidly diversifying lineages (Davies *et al.* 2011). Critically, extinction
52 risk may be more related to evolutionary dynamics (Davies *et al.* 2011) than to traits, and
53 contingent on extinction drivers such as habitat loss, exploitation, etc., (Godefroid *et al.*
54 2014, Davies 2019). That extinction risk correlates strongly with clade age and richness
55 suggests that the defining characteristics of rarity – endemism, limited range, and small
56 population sizes (Rabinowitz *et al.*1986) – may simply follow from high rates of
57 speciation (Davies *et al.* 2011). However, Vamosi and Wilson (2008) found extinction
58 risk to be elevated in species-poor families, moreover, in some habitats, phylogenetically
59 distinct species may be more threatened (Daru *et al.* 2013). Therefore, in old and species-
60 poor families, perhaps remnants of formerly large and diverse clades, extinction risk may
61 be linked to life history features that are sensitive to extinction drivers (Yessoufou &
62 Davies 2016), while diversification dynamics dominate in young, species-rich families,
63 implying that the drivers of extinction differ between old vs young clades (Vamosi *et al.*
64 2018, Davies 2019).

65 Successful plant invasions are conditioned on context-specific factors that include use
66 (van Kleunen *et al.* 2020) and transport by humans (Kueffer 2017), degree of climate
67 matching (Thuiller *et al*. 2005), residence time (Wilson *et al*. 2007), propagule pressure
68 (Simberloff 2009), geography of habitat alteration and anthropogenic disturbance
69 (Lembrechts *et al.* 2016), and the invasibility of particular communities and biogeographic
70 regions (Richardson & Pyšek 2006). Nonetheless, successful invasions have been
71 correlated in comparative studies with a suite of traits – auto-fertility (Razanajatovo *et al.*
72 2016), self-compatibility (Hao *et al*. 2011), height (van Kleunen *et al*. 2007), small seeds
73 (Hamilton *et al*. 2005), high specific leaf area (Hamilton *et al*. 2005), large native range
74 size (Schmidt *et al*. 2012, van Kleunen *et al*. 2007), broad climate and habitat tolerances
75 (Schmidt *et al*. 2012, Bradshaw *et al*. 2008), competitive ability (Guo *et al*. 2018),
76 variability and perhaps plasticity in growth form and life history (Schmidt *et al*. 2012),
77 abiotic pollination (Hao *et al*. 2011), polyploidy (Schmidt *et al*. 2012) and hybridization
78 (Ellstrand & Schierenbeck 2000) – that appear consistently advantageous (Table 1). Thus,
79 while naturalized species may also be non-randomly distributed among angiosperm
80 families (Pyšek 2017, Pyšek 1998) and benefit from factors extrinsic to ecological
81 features, trait syndromes across clades appear to strongly influence invasion success.

82 Here, adjusting for phylogenetic relatedness and clade size, we ask 1) how the
83 representation of globally threatened species by flowering plant family is related to that of
84 naturalized species (Fig. 1); 2) whether the frequency of species in either category tends
85 toward the opposite extremes of the same trait axes; and 3) if the relative importance of
86 traits vs evolutionary history as explanatory variables differs for threat vs naturalization.
87 We evaluate these hypotheses in a hierarchical Bayesian regression framework – allowing
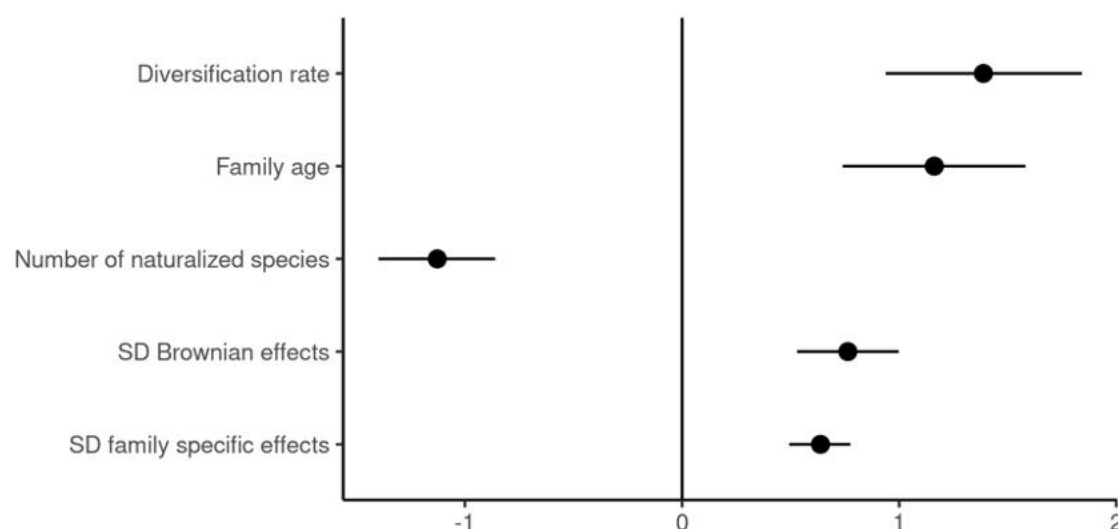88 us to quantify uncertainty in analyses, and to better compare effect sizes among predictors.

Fig. 1. Phylogeny of the 207 angiosperm families. Families included have at least one species vetted by the IUCN and less than 100% of species categorized as either threatened or naturalized. Internal branches (*left*) show family level diversification rate, log(species richness)/clade age (ancestral state reconstruction calculated using *fastAnc* in the phytools R package). Proportion IUCN vetted species classed as globally threatened in red (*middle*), and proportion of all species in the family naturalized in blue (*right*).
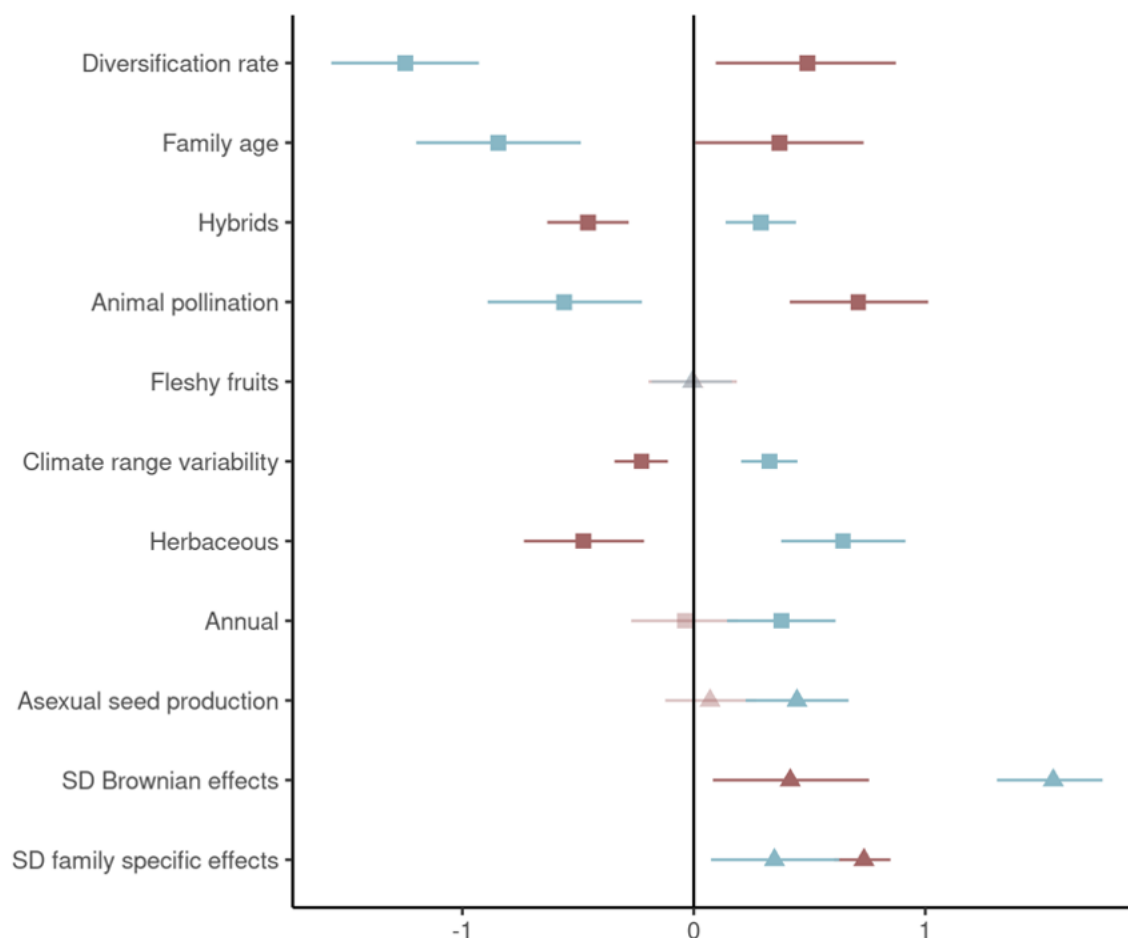
Results

Across angiosperm families, the proportion of vetted species classed as threatened was negatively related to the frequency of naturalized species (Fig. 2). In separate models predicting the proportion by family of threatened to vetted species or naturalized to total species as a function of traits and evolutionary history, the proportion threatened was positively related and the proportion naturalized negatively related – to family age and diversification rate (Fig. 3). Trait covariates also showed opposing patterns for threatened and naturalized models. Threat was lower in families with greater climate variability (number of climate zones occupied), and those characterized by many interspecific hybrids and predominantly herbaceous life forms, and higher in families that included animal pollinated species (Fig. 3). In contrast, proportion naturalized per family was positively related to herbaceous growth form, climate variability, number of interspecific hybrids, and the presence of species with asexual seed production, and annual life-histories, but negatively correlated with animal pollination (Fig. 3). Notably, the negative effect sizes of diversification rate and family age on proportion naturalized were greater in magnitude than the positive effects of these same factors on proportion threatened (Fig. 3).

Model fit (normalized root mean squared error: *NRMSE*) for threat (*NRMSE* = 0.20, *sd* = 0.03) was better than for naturalization (*NRMSE* = 0.54, *sd* = 0.08). And, contrary to expectations, the proportion of the variance in family level effects explained by the Brownian phylogenetic component (phylogenetic heritability, Lynch's λ), was higher for

116   naturalization ($\lambda = 0.93$, s$d = 0.08$) than threat ($\lambda = 0.25$, $sd = 0.22$, Fig. 4). Sensitivity
117   analyses using the total number of species per family rather than IUCN vetted species in
118   the binomial response ($NRMSE = 0.49$, $sd = 0.06$, Tables S3,S4) and excluding family
119   level effects ($NRMSE = 0.55$, $sd = 0.03$, S3.1) both reduced threat model fit,
120   demonstrating the importance of controlling for data deficiencies and family level
121   variation not accounted for by trait data. Sensitivity analyses investigating the proportion
122   of invasive rather than naturalized species per family also reduced model fit ($NRMSE =$
123   0.78, $sd = 0.11$) likely reflecting the smaller number of taxa classed as invasive, and
124   perhaps the more arbitrary application of the invasive vs naturalized label.
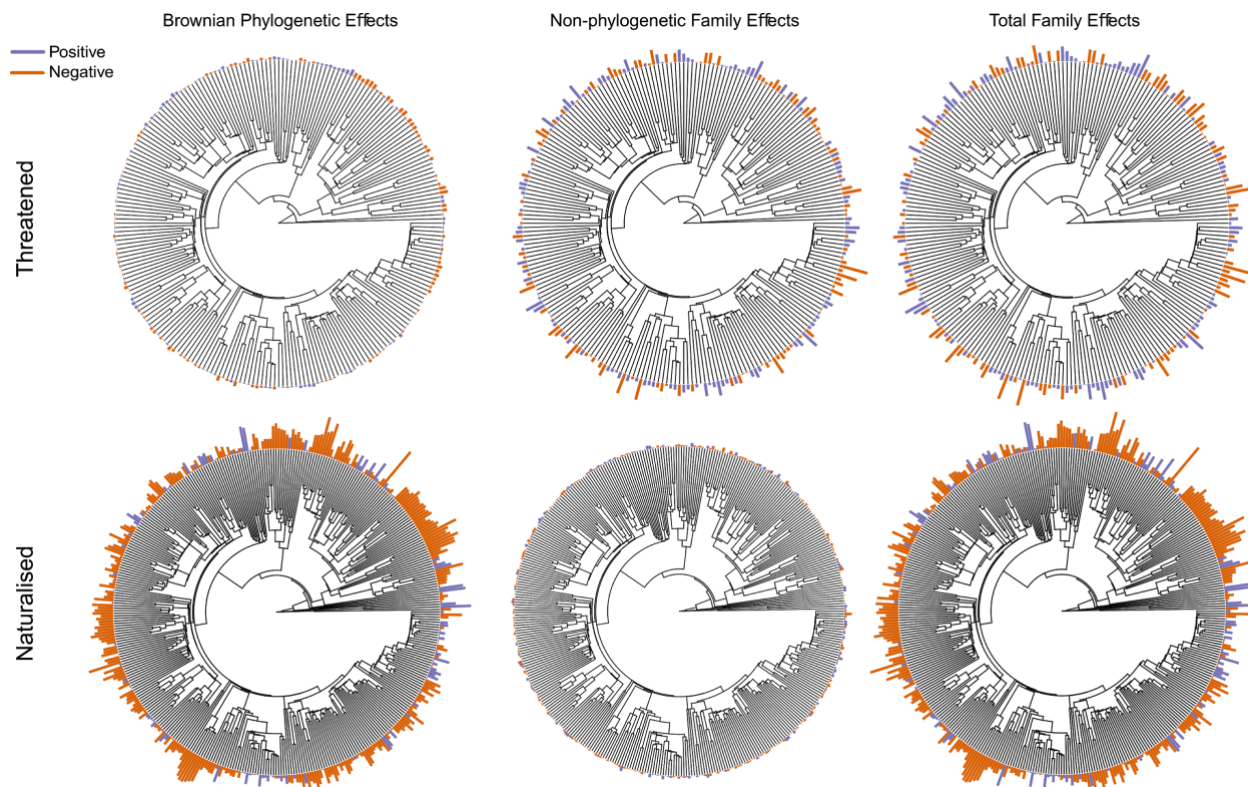
125

126   **Fig. 2. Estimated model parameters for predicting proportion threatened.** Regression
127   coefficients and 80% credible intervals for hierarchical effects by family among the 236
128   families with IUCN vetted species as a function of the number of naturalized species,
129   while adjusting for diversification rate (log(family size)/family age) and family-level
130   effects.

Fig. 3. Estimated model parameters for threatened and invasive. Regression coefficients and hierarchical standard deviations, with 80% credible intervals, by family (*red*) among the 267 families with IUCN vetted species and the proportion naturalized (*blue*) among all species per family for all 395 families included in the study. Squares (rather than triangles) indicate variables with opposite effects on naturalized vs threatened status, pale triangles credible intervals overlapping zero.

Fig. 4. Family-level hierarchical effects for the full model. Proportion IUCN vetted species threatened per family (*top row*), and the proportion naturalized (*bottom row*) among all species per family, separated into Brownian phylogenetic, non-phylogenetic, and total family effects (Brownian + non-phylogenetic). Purple bars indicate positive effects, orange bars indicate negative effects, and bar lengths indicate the relative magnitude of the mean estimated effect per family.

Discussion

At the global scale, we show that the drivers of threat and naturalization across angiosperms are opposing, and that this can be explained by divergent macro-evolutionary and ecological trait relationships. We note that models excluding family-specific effects or phylogeny were poorer fits, indicating that a large component of the variation among families still remains unexplained; and family-level estimates of age and diversification rate (not entirely separable) may not optimally capture the signature of macroevolutionary processes towards the present. Nonetheless, our results provide the first quantitative support at a global scale across angiosperms for the hypothesis that naturalization and threat represent "two sides of the same coin" (Schmidt *et al.* 2012, Bradshaw *et al.* 2008).

Consistent with previous studies (Davies *et al.* 2011), diversification rate was positively related to proportion threatened per family. In contrast, naturalization was negatively related to both diversification rate and family age. Fast diversifying clades are often associated with localized radiations that give rise to many endemics with narrow ranges that are restricted to particular habitat types (*e.g.* many plant lineages in the Fynbos, South Africa, have diversified rapidly, and are characterized by a high diversity of narrowly distributed species, frequently restricted to particular soil types). This high ecological specialization likely restricts geographical expansion and naturalization outside of the native range, resulting in a negative correlation between diversification and naturalization. More established lineages, and those in older families may have had more time to spread

165   and thus have had more opportunities to become naturalized outside of their native range.
166   However, as species age they may also contract in their geographic extent, especially if
167   the environmental conditions which favored their initial expansion change (c.f. taxon
168   cycle: Ricklefs & Bermingham 2002). Species experiencing range contraction are also less
169   likely to become naturalized elsewhere. As only old families can contain old species, this
170   could lead to a negative correlation between naturalization and family age. The magnitude
171   of the negative effects of diversification rate and family age on naturalization were also
172   somewhat greater than those of the positive effects on threat. Contrary to initial
173   expectations, after adjusting for macroevolutionary and life-history predictors, we also
174   found that in the naturalization model, a larger fraction of variance in family level effects
175   was attributable to Brownian phylogenetic effects, compared to the equivalent model for
176   threat. We suggest that one possible explanation for the weaker signature of
177   macroevolutionary process on threat is the mixing of threatened taxa found within both
178   young and old clades such that observed threat captures two independent processes.

179   The opposing relationship between threat and naturalization was also manifested in the
180   trait syndromes that characterize either status. Naturalization was positively related to
181   herbaceous growth form, climate variability, frequency of interspecific hybrids, asexual
182   seed production, and annual life history, and negatively related to animal pollination;
183   whereas threat was negatively related to herbaceous growth form, climate variability, and
184   frequency of interspecific hybrids, and positively to animal pollination and variation in
185   growth form. The correlations we detect among angiosperm families match closely to
186   expectations from theory and existing comparative studies across various scales (Table 1)
187   – with naturalization characterized by habitat/climatic generalism, short generation time,
188   asexuality and polyploidy, and threat by habitat specialism and endemism, dependence on
189   mutualists, and diploidy.

190   That we recover such strong associations at higher taxonomic levels and at a global scale
191   is notable and suggests that the trait signatures characterizing threat and naturalization in
192   the present day extend back over deep evolutionary time. Thus, while particular trait
193   syndromes appear to predispose some species to higher risk of extinction and others to
194   ecological expansion, this might not translate straightforwardly to long-term
195   macroevolutionary dynamics. For example, the traits characteristic of threatened species
196   today may be features that permit chronically rare species to persist (*e.g.* the directed
197   movement of pollen by an animal vector) and, by limiting outcrossing, allow adaptation to
198   spatially restricted environmental conditions (Vermeij & Grosberg 2018). An alternative
199   strategy – promoting colonization and expansion at range margins – relies on features such
200   as abiotic pollination, asexual seed production, and, to enable rapid niche shifts in the face
201   of novel climatic conditions or sudden environmental change, hybridization and
202   polyploidy (Baniaga *et al.* 2019). Both strategies thus represent successful macro-
203   evolutionary syndromes, but under contrasting selective regimes. Include a Discussion
204   that summarizes (but does not merely repeat) your conclusions and elaborates on their
205   implications. There should be a paragraph outlining the limitations of your results and
206   interpretation, as well as a discussion of the steps that need to be taken for the findings to
207   be applied. Please avoid claims of priority.

208 **Materials and Methods**
209   *Data*
210   To estimate species richness, we tabulated the number of species (including interspecific
211   hybrids) within each of 404 angiosperm families (per APG3, Angiosperm Phylogeny

212   Group 2009) from The Plant List (TPL, http://www.theplantlist.org/) – counting only the
213   301,639 species with 'accepted' names. To determine the number of species in each class,
214   we compiled a list of all angiosperms labeled as 1) naturalized (12,256 species) in any
215   region, globally, from the GloNAF database (van Kleunen 2019); 2) 'invasive' (4,540
216   species) from Randall (2017); and 3) as 'threatened' (12,894 species), if listed as globally
217   "Vulnerable" or more threatened by the International Union for the Conservation of
218   Nature (https://www.iucnredlist.org/, accessed March 11, 2019). We labeled species as
219   'not threatened' if listed as 'Least Concern', 'Near Threatened' or 'data deficient' by the
220   IUCN. To control for study effort, we summed the number of species per family that have
221   currently been assessed by the IUCN. In all cases (naturalized, invasive, threatened and
222   vetted), we filtered species through TPL to include only accepted names. Of the 404
223   families, 239 included a naturalized species, 196 included an invasive species, 237
224   included an IUCN-vetted species, and all vetted families included a threatened species.

225   For family level trait data, we retrieved binary data on climate zones (tropical, subtropical,
226   temperate, frigid zone) occupied, and presence of animal pollination and fleshy fruits –
227   from the Watson and Dallwitz (1992 onward) online key to angiosperm families. To
228   capture variability in climate tolerance, we summed the number of climate zones (1-4)
229   reported among species within each family. Animal pollination and fleshy fruits were
230   coded for each family as 1 if biotic pollination (346 families) or fleshy fruits (77 families)
231   was the sole mode, 0.5 if multiple traits (biotic and abiotic pollination (13 families) or
232   fleshy and non-fleshy fruits, 130 families), and 0 if only abiotic pollination (45 families)
233   or dry fruits (197 families) are known. From Watson and Dallwitz and data from the
234   Plants National Database (https://plants.sc.egov.usda.gov/), we labeled families as known
235   to include annual species (1, 99 families) or otherwise (0, 305 families). From Zanne *et al.*
236   (2014) and Hawkins *et al.* (2011) we classified families as predominantly herbaceous (1,
237   156 families), predominantly woody (0, 192 families), or mixed (0.5, 56 families). Lastly,
238   we used data from Hojsgaard *et al.* (2014) to identify families in which asexual seed
239   production (1, 173 families) is known to occur vs all other families (0, 231 families).

240   For estimates of age and diversification rates, we relied on stem family ages from the
241   Zanne *et al.* (2014) phylogeny of vascular plants. Zanne *et al.* first used a method of
242   congruification to make their tree ultrametric based on the phylogeny of Soltis *et al.*
243   (2011). Branch lengths were then time-scaled using 39 fossil calibration points, which at
244   the time represented the most reliable set of fossils spanning the angiosperm phylogeny.

245   *Statistical Analysis*
246   We used Bayesian binomial-logit multilevel regressions to model the proportion of species
247   threatened and naturalized plants per family. To assess the relationship between threat and
248   naturalization, we fit an initial model with, for each family, the proportion of all species
249   naturalized, and family age and diversification rate as continuous predictors, and with the
250   proportion of vetted species (those assessed by the IUCN) that have been classed as
251   threatened as the response. To aid in the comparison of effect sizes across continuous and
252   binary predictors, we log-transformed, centered and scaled to a standard deviation of 0.5
253   all continuous predictors prior to analyses. To identify traits associated with the proportion
254   of species threatened and naturalized per family, we fit two additional models that
255   included family-level traits, and family age and diversification rate. We calculated
256   diversification rate as $\log(N)$/clade age, which is the maximum likelihood estimate of
257   diversification rate assuming negligible extinction (eq. 3 in Magallon & Sanderson 2001).
258   A more complex formulation (eq. 6 in Magallon & Sanderson 2001) allows a constant

259 extinction, but empirically, family level diversification rates estimates are little influenced
260 (Jansson & Davies 2008). To account for phylogenetic non-independence and family-level
261 effects in all models, we included hierarchical effects by family separated into
262 phylogenetic (assuming a Brownian motion model of evolution) and non-phylogenetic
263 effects, following the additive quantitative genetic model adapted to interspecific data.
264 The correlation matrix for determining family-level Brownian phylogenetic effects was
265 calculated from the Zanne *et al.* (2014) phylogeny of vascular plants pruned to a single
266 representative species per family.

267 We fit all models in Stan version 2.18.0 (The Stan Development Team) accessed using the
268 R package *brms* version 2.7.0 (Buerkner 2017). We fit models with the *brms* default
269 priors, which use a uniform prior for the regression coefficients, and half-Student's *t*
270 distributions with three degrees of freedom and scale parameter of 10 for the variance
271 components of the family-level effects. As these represent extremely weak priors, for the
272 simplified and full threatened models we also explored the use of alternative custom priors
273 (normal (0,1.5) for regression coefficients and normal (0,1) for family effect standard
274 deviations), which provide more realistic expectations (Figs. S1, S2). We found a
275 negligible influence of each set of priors on posterior estimates (Figs. S4, S5, S7, S8, S10,
276 S11, S13, S14), and thus assumed the *brms* default priors for all other models. We ran
277 models across four chains, with 10000 iterations per chain. For each chain, the first 5000
278 iterations were used as burn-in and discarded. The remaining iterations were thinned to
279 retain every fourth iteration, resulting in a total of 5000 posterior draws. We diagnosed
280 convergence by visual inspection of traceplots and observation of Rhat values equal to 1.0
281 for all estimated parameters, and we assessed model fit using posterior predictive checks
282 (see Data & Code supplement) and root mean square error (*RMSE*). To compare model
283 fits across varying responses, we calculated the normalized *RMSE* (*NRMSE*) scaling by
284 the interquartile range of the observed data, and across all posterior samples. To calculate
285 phylogenetic heritability (Lynch 1991) we used the *hypothesis* function in *brms* with the
286 hypothesis $\sigma_{2\text{Brownian}}/(\sigma_{2\text{brownian}} + \sigma_{2\text{family-specfic}}) = 0$ across all posterior samples.

287 To explore the effect of including both Brownian and family-specific effects, we ran
288 sensitivity models which fit the full threatened model with a) no family specific effects, b)
289 only family specific effects, b) only Brownian effects. To determine the effect of using the
290 number of IUCN vetted species rather than the total number of species per family, we fit
291 the full threatened model with total species richness per family as *n* (the denominator) in
292 the binomial model. Finally, to assess the sensitivity of the naturalized model results to the
293 definition of an invasive species, we re-ran the naturalized model using the number of
294 invasive species, defined by Randall (2017) as naturalized species that have spread and
295 have significant ecological/economic impact.

References and Notes

297 Alstad, A.O. *et al.* The pace of plant community change is accelerating in remnant
298   prairies. *Science Advances* **2**, p.e1500975 (2016).

300 Baniaga, A.E. *et al.* Polyploid plants have faster rates of multivariate niche differentiation
301   than their diploid relatives. *Ecology Letters* **23**, 68-78 (2019).

303 Bradshaw, C.J.*et al.* Threat or invasive status in legumes is related to opposite extremes of
304   the same ecological and life-history attributes. *Journal of Ecology* **96**, 869-883
305   (2008).

Buerkner, P.C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* **80**, 1-28 (2017).

Corlett, R.T. Plant diversity in a changing world: status, trends, and conservation needs. *Plant Diversity* **38**, 10-16 (2016).

Daru, B.H. *et al.* A global trend towards the loss of evolutionarily unique species in mangrove ecosystems. *PloS One* **8**, p.e66686 (2013).

Davies, T.J. *et al.* Extinction risk and diversification are linked in a plant biodiversity hotspot. *PLoS Biology* **9**, p.e1000620 (2011).

Davies, T.J. The macroecology and macroevolution of plant species at risk. *New Phytologist* **222**, 708-713 (2019).

Ellstrand, N.C., Schierenbeck, K.A. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences* **97**, 7043–7050 (2000).

Godefroid, S. *et al.* Do plant reproductive traits influence species susceptibility to decline? *Plant Ecology and Evolution* **147**, 154-164 (2014).

Guo, W.-Y. *et al.* The role of adaptive strategies in plant naturalization. *Ecology Letters* **21**, 1380–1389 (2018).

Hamilton, M.A. *et al.* Life-history correlates of plant invasiveness at regional and continental scales. *Ecology Letters* **8**, 1066-1074 (2005).

Hao, J.H. *et al.* A test of Baker's law: breeding systems of invasive species of Asteraceae in China. *Biological Invasions* **13**, 571-580 (2011).

Hawkins, B.A. *et al.* Global angiosperm family richness revisited: linking ecology and evolution to climate. *Journal of Biogeography* **38**, 1253-1266 (2011).

Hojsgaard, D. *et al. Critical Reviews in Plant Sciences* **33**, 414–27 (2014).

Jansson, R., Davies, T.J. 2008. Global variation in diversification rates of flowering plants: energy vs. climate change. *Ecology Letters* **11**, 173-183.

Kueffer, C. Plant invasions in the Anthropocene. *Science* **358**, 724-725 (2017).

Lembrechts, J.J. *et al.* Disturbance is the key to plant invasions in cold environments. *Proceedings of the National Academy of Sciences* **113**, 14061-14066 (2016).

Lockwood, J.L. *et al.* The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution* **20**, 223-228 (2005).

Magallon, S., Sanderson, M.J. 2001. Absolute diversification rates in angiosperm clades. *Evolution* **55**, 1762-1780.

Pyšek, P. Is there a taxonomic pattern to plant invasions? *Oikos* 282-294 (1998).

Pyšek, P. *et al.* Naturalized alien flora of the world. *Preslia* **89**, 203-274 (2017).

Rabinowitz D. *et al.* Seven forms of rarity and their frequency in the flora of British Isles. In: Soule M. (ed.), Conservation Biology: The science of scarcity and diversity. Sinauer Associates, Sunderland, 1882–1204 (1986).

Randall, R.P. *A global compendium of weeds* (No. Ed. 3). R.P. Randall (2017).

Razanajatovo, M. *et al.* Autofertility and self-compatibility moderately benefit island colonization of plants. *Global Ecology and Biogeography* **28**, 341–352 (2019).

Richardson, D.M., Pyšek, P. Plant invasions: merging the concepts of species invasiveness and community invasibility. *Progress in Physical Geography* **30**, 409-431 (2006).

Ricklefs, R.E., Bermingham, E. The concept of the taxon cycle in biogeography. *Global Ecology and Biogeography* **11**, 353-361 (2002).

Schmidt, J. P. *et al.* Cost-sensitive risk assessment for invasive plants in the United States. *Ecosphere* **3**:art46 (2012).

Schwartz, Mark W., Simberloff, D. Taxon size predicts rates of rarity in vascular plants. *Ecology Letters* **4**, 464–69 (2001).

Seebens, H. *et al.* No saturation in the accumulation of alien species worldwide. *Nature Communications* **8**, 14435 (2017).

Simberloff, D. The role of propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics* **40**, 81-102 (2009).

Soltis, D. E. *et al.* Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* **98**, 704–730 (2011).

The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**, 105–121 (2009).

The Stan Development Team.Stan Modeling Language User's Guide and Reference Manual.

Thuiller, W. *et al.* Consequences of climate change on the tree of life in Europe. *Nature* **470**, 531–534 (2011).

Vamosi, J.C., Wilson, J.R. Nonrandom extinction leads to elevated loss of angiosperm evolutionary history. *Ecology Letters* **11**, 1047-1053 (2008).

Vamosi, J.C. *et al.* Macroevolutionary patterns of flowering plant speciation and extinction. *Annual Review of Plant Biology* **69**, 685-706 (2018).

406 van Kleunen, M. *et al.* Economic use of plants is key to their naturalization success.
407   *Nature Communications* **11**, 1-12 (2020).

409 van Kleunen, M. *et al.* The Global Naturalized Alien Flora (Glo NAF) database. *Ecology*
410   **100**, e02542 (2019).

412 Vermeij, G.J., Grosberg, R.K. Rarity and persistence. *Ecology Letters* **21**, 3-8 (2018).

414 Watson, L., Dallwitz, M.J. The families of flowering plants: descriptions, illustrations,
415   identification, and information retrieval. *World Wide Web: http://delta–intkey.*
416   *com/angio/, version*, *25 (1992 onwards).*

418 Willis, C.G. *et al.* Phylogenetic patterns of species loss in Thoreau's woods are driven by
419   climate change. *Proceedings of the National Academy of Sciences* **105**, 17029-
420   17033 (2008).

422 Wilson, J.R. *et al*. Residence time and potential range: crucial considerations in modelling
423   plant invasions. *Diversity and Distributions* **13**, 11-22 (2007).

425 Winter, M. et al. Plant extinctions and introductions lead to phylogenetic and taxonomic
426   homogenization of the European flora. *Proceedings of the National Academy of*
427   *Sciences* **106**, 21721-21725 (2009).

429 Yessoufou, K., Davies, T.J. Reconsidering the Loss of Evolutionary History: How Does
430   Non-random Extinction Prune the Tree-of-Life? In *Biodiversity conservation and*
431   *phylogenetic systematics* (pp. 57-80). Springer, Cham (2016).

433 Zanne, A.E. *et al.* Three keys to the radiation of angiosperms into freezing environments.
434   *Nature* **506**, 89–92 (2014).

Opposing macroevolutionary and trait-mediated patterns of threat and naturalization in flowering plants
Supplementary Materials

J.P. Schmidt[1,2*], Maxwell J. Farrell[2,3] & T. Jonathan Davies[4,5]

[3]Odum School of Ecology, University of Georgia

[2]Center for the Ecology of Infectious Diseases, University of Georgia

[3]Ecology & Evolutionary Biology Department, University of Toronto

[4]Botany, Forest & Conservation Sciences, University of British Columbia

[5]African Centre for DNA Barcoding, University of Johannesburg

[*]To whom correspondence should be addressed: e-mail: jpschmidt270@gmail.com

1   Exploration of priors

As a default for binomial-logit models, the *brms* package uses extremely flat priors for the regression coefficients and standard deviation parameters. These are improper unbounded uniform priors, and a half Student t distribtion with three degrees of freedom and a scale parameter of ten, respectively. To explore the influence of these priors, and their impact on model performance, we compare them to custom priors reflecting recommendations by Richard McElreath in his textbook *Statistical Rethinking* (a normal(0,1.5) and half-normal(0,1) respectively).
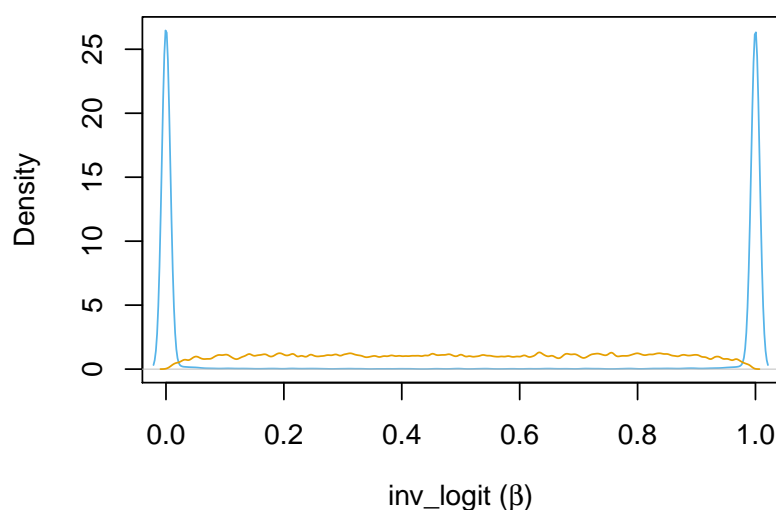
Figure SM 1: Priors on the regression coefficients for the *brms* default (blue), and custom (yellow) priors

The default improper uniform prior used by *brms* is flat on the log-odd scale, so when converted to the probability scale, puts excess mass near 0 and 1. A Normal(0, 1.5) prior is more reasonable, having a flatter distribution with a mode at 0.5 on the probability scale.
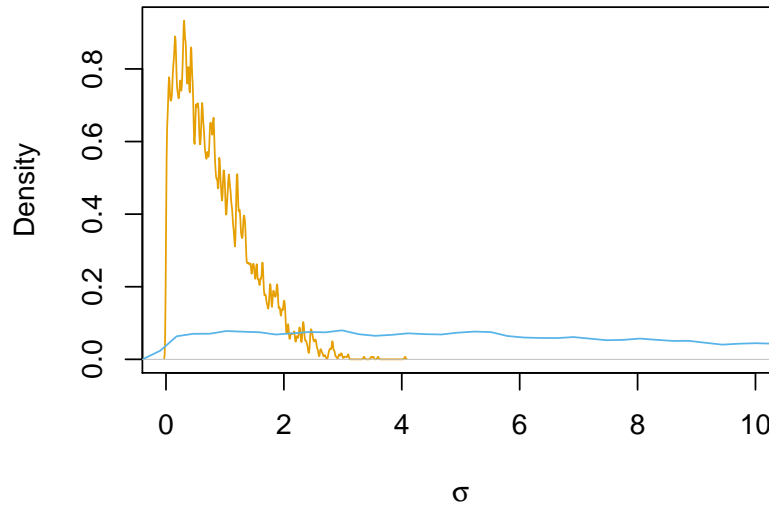
2

Figure SM 2: Priors on the regression coefficients for the *brms* default (blue), and custom (yellow) priors

The *brms* prior is extremely long tailed. Considering we are fitting a non-linear model in which changes in log-odds over over 4 have a diminishing influence due to the ceiling effect of the binomial distribution. With this in mind, a prior of Normal(0,1) constrains the posterior to a more realistic range, and is likely to improve sampling efficiency.

To explore the impact of each of these priors, we fit the simple and full threatened models using each set of priors, and visualize the relationships between the priors and associated posteriors for two representative parameters. Overall, the choice of prior was found to have no influence on the posterior parameter estimates in these two models, so we used the brms defaults for all other models.

3

2 Main models

2.1 Simple threatened model - *brms* default priors

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Number of naturalized species	-0.97	0.20	-1.37	-0.97	-0.56	4320.48	1.00
Diversification rate	1.67	0.35	1.00	1.66	2.36	4526.64	1.00
Family age	1.33	0.33	0.68	1.33	1.98	4473.99	1.00
SD brownian effects	0.98	0.20	0.59	0.98	1.37	991.53	1.00
SD family specific effects	0.60	0.14	0.28	0.61	0.83	813.83	1.00

Table SM 1: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=236; RMSE=5.95 (+/- 0.73 SD); NRMSE = 0.20 (+/- 0.03).
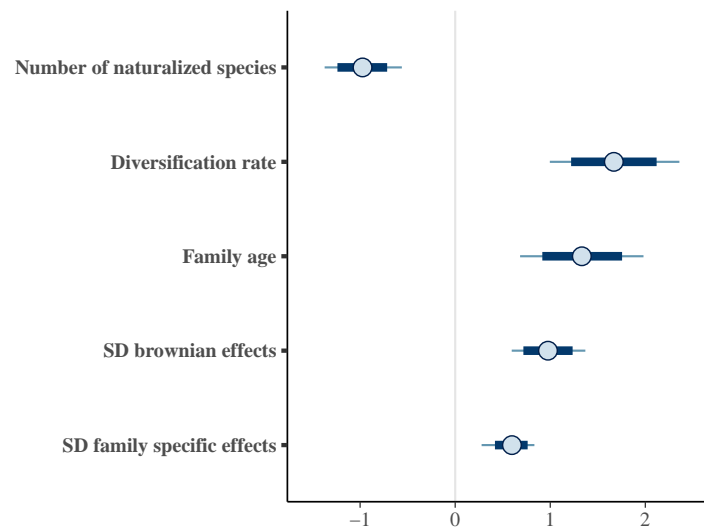

Figure SM 3: Forest plot for estimated β and σ parameters for the simple threatened model. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.
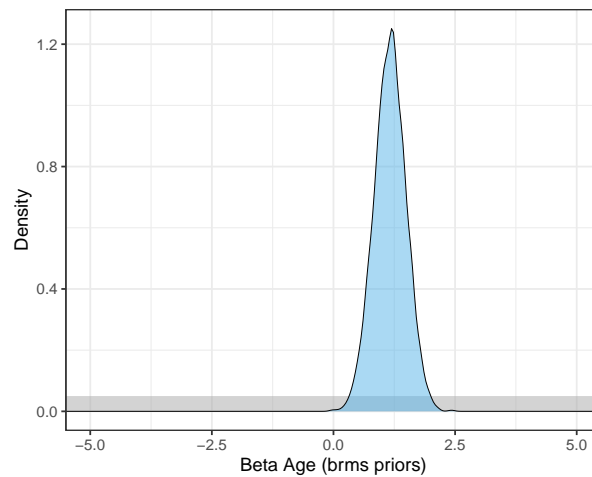
4

Figure SM 4: Prior (grey) and posterior (blue) distributions for the Age regression coefficient for the simple rarity model using *brms* priors. The improper uniform prior was restricted to (-10,10) to aid visualization.
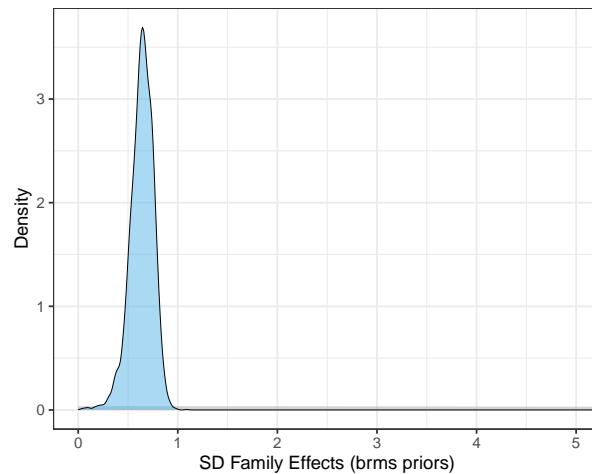

Figure SM 5: Prior (grey) and posterior (blue) distributions for the Family Effects standard deviation parameter for the simple rarity model using *brms* priors.

2.2   Simple threatened model - custom priors

Table SM 1: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=236; RMSE=5.93 (+/- 0.73 SD); NRMSE = 0.20 (+/- 0.03).

5

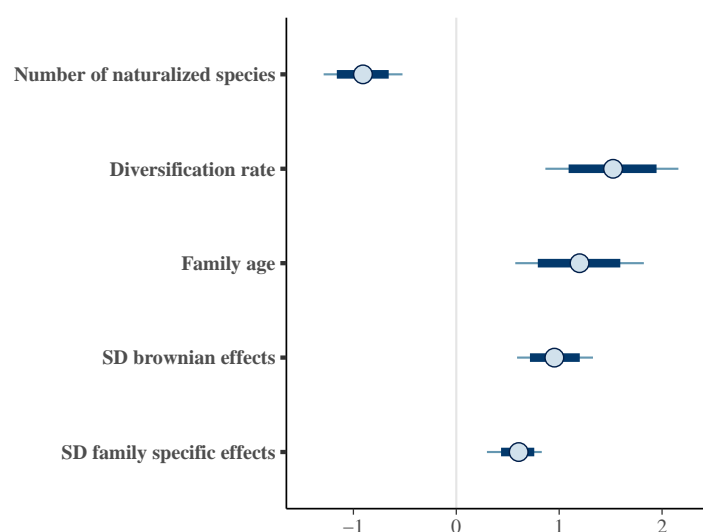|	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Number of naturalized species	-0.91	0.20	-1.29	-0.91	-0.52	4462.00	1.00
Diversification rate	1.52	0.33	0.87	1.52	2.16	4579.12	1.00
Family age	1.20	0.32	0.57	1.20	1.82	4506.43	1.00
SD brownian effects	0.95	0.19	0.59	0.95	1.33	1175.25	1.00
SD family specific effects	0.61	0.13	0.30	0.62	0.83	1044.40	1.00

Figure SM 6: Forest plot for estimated β and σ parameters for the simple threatened model with custom priors. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.
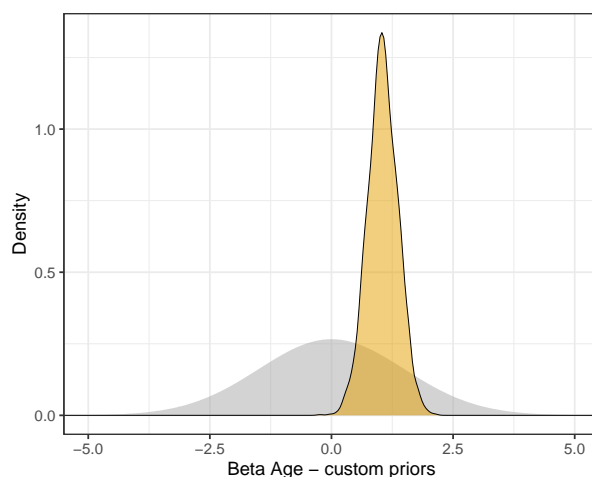
6

Figure SM 7: Prior (grey) and posterior (yellow) distributions for the Age regression coefficient for the simple rarity model using custom priors. The improper uniform prior was restricted to (-10,10) to aid visualization.
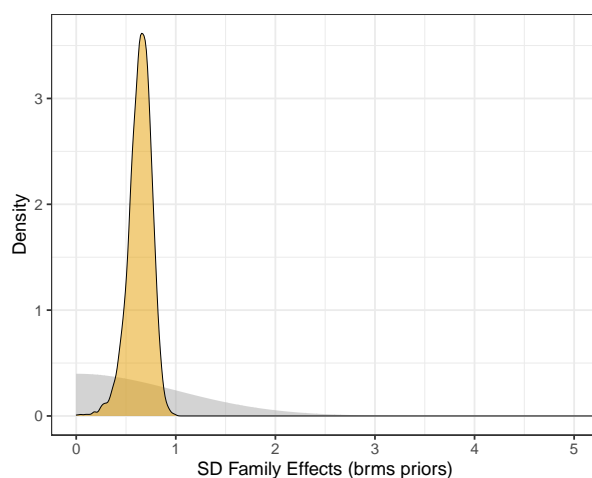

Figure SM 8: Prior (grey) and posterior (yellow) distributions for the Family Effects standard deviation parameter for the simple rarity model using custom priors.

2.3 Full threatened model - IUCN vetted species - *brms* default priors

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Diversification rate	0.47	0.33	-0.18	0.47	1.13	4411.32	1.00
Famly age	0.42	0.31	-0.17	0.42	1.01	4175.69	1.00
Woodiness	0.61	0.14	0.34	0.61	0.87	3874.33	1.00
Herbaceous	-1.05	0.23	-1.48	-1.05	-0.59	2332.31	1.00
Climate sum	-0.26	0.10	-0.45	-0.26	-0.07	4508.21	1.00
Number of species hybrids	-0.18	0.13	-0.43	-0.18	0.08	4658.36	1.00
Hermaphroditic	0.13	0.23	-0.30	0.12	0.58	4991.91	1.00
Dioecious	-0.18	0.15	-0.48	-0.18	0.13	4657.14	1.00
Monoecious	-0.14	0.16	-0.45	-0.15	0.17	4848.15	1.00
Asexual	0.05	0.16	-0.28	0.05	0.37	4352.38	1.00
Fleshy fruit	0.03	0.15	-0.25	0.03	0.32	4339.32	1.00
Animal pollinated	0.55	0.25	0.05	0.55	1.04	4139.00	1.00
SD brownian effects	0.52	0.28	0.04	0.53	1.06	472.22	1.01
SD family specific effects	0.73	0.12	0.47	0.74	0.92	795.47	1.01

Table SM 2: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=236; RMSE=5.94 (+/- 0.72 SD); NRMSE = 0.20 (+/- 0.03).
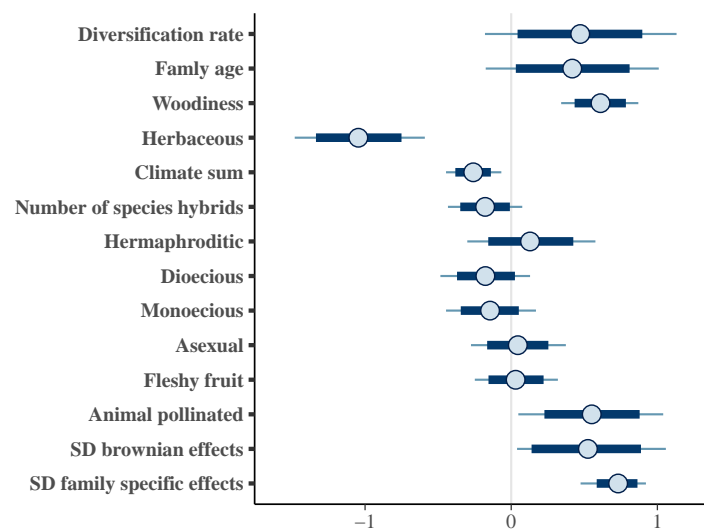
8

Figure SM 9: Forest plot for estimated β and σ parameters for the full threatened model with IUCN vetted species. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.
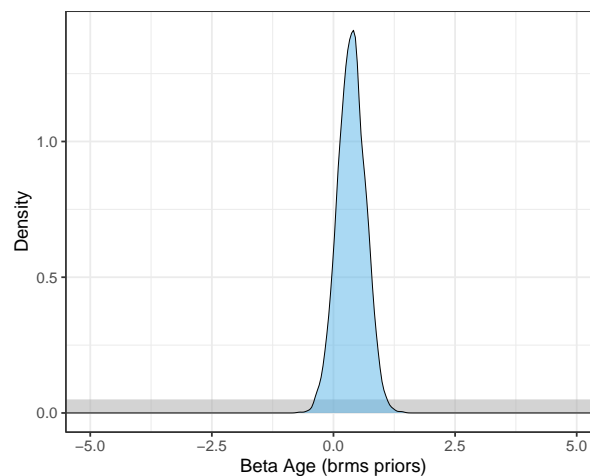

Figure SM 10: Prior (grey) and posterior (blue) distributions for the Age regression coefficient for the full rarity model using *brms* priors. The improper uniform prior was restricted to (-10,10) to aid visualization.
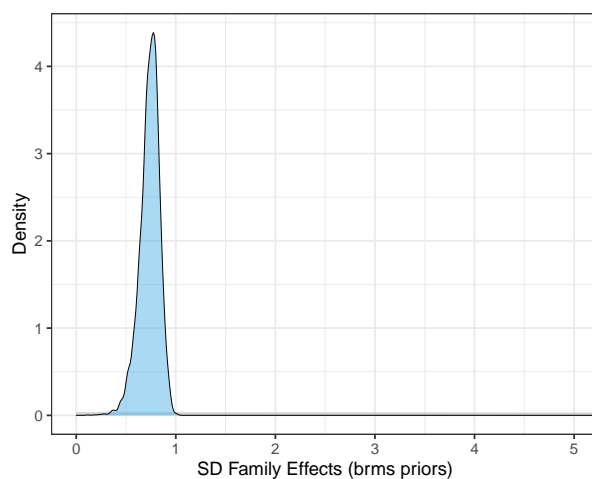
Figure SM 11: Prior (grey) and posterior (blue) distributions for the Family Effects standard deviation parameter for the full rarity model using *brms* priors.

2.4 Full threatened model - IUCN vetted species - custom priors

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Diversification rate	0.44	0.32	-0.20	0.44	1.07	4537.10	1.00
Famly age	0.40	0.29	-0.19	0.40	0.95	3913.99	1.00
Woodiness	0.61	0.13	0.35	0.61	0.87	4766.49	1.00
Herbaceous	-1.04	0.22	-1.46	-1.04	-0.59	3420.69	1.00
Climate sum	-0.26	0.09	-0.44	-0.26	-0.07	4853.71	1.00
Number of species hybrids	-0.18	0.13	-0.43	-0.18	0.08	5120.11	1.00
Hermaphroditic	0.14	0.23	-0.32	0.14	0.58	4905.70	1.00
Dioecious	-0.18	0.15	-0.46	-0.18	0.12	4982.72	1.00
Monoecious	-0.14	0.15	-0.44	-0.14	0.16	4151.91	1.00
Asexual	0.04	0.16	-0.27	0.04	0.37	4815.08	1.00
Fleshy fruit	0.03	0.15	-0.25	0.03	0.33	4621.55	1.00
Animal pollinated	0.54	0.25	0.04	0.54	1.02	4005.88	1.00
SD brownian effects	0.49	0.26	0.03	0.50	1.00	798.59	1.00
SD family specific effects	0.74	0.11	0.50	0.75	0.92	1078.48	1.00

Table SM 2: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=236; RMSE=5.93 (+/- 0.73 SD); NRMSE = 0.20 (+/- 0.03).
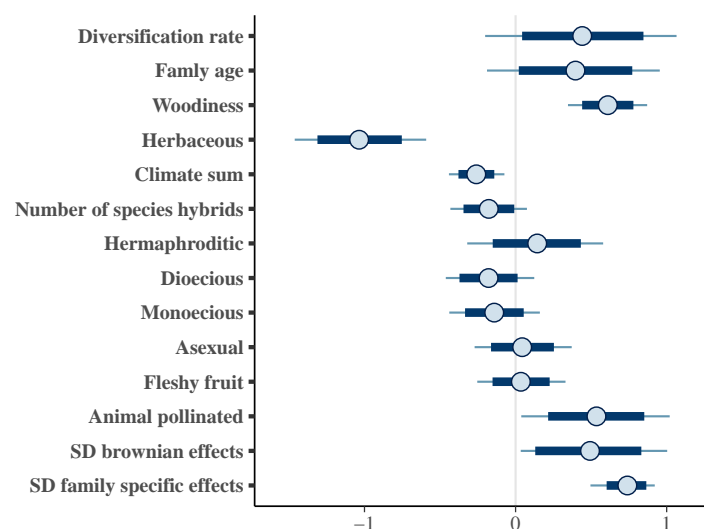
Figure SM 12: Forest plot for estimated β and σ parameters for the full threatened model with IUCN vetted species with custom priors. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.
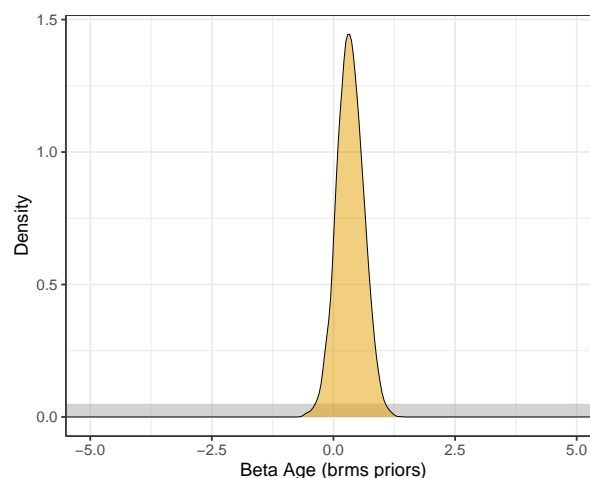

Figure SM 13: Prior (grey) and posterior (yellow) distributions for the Age regression coefficient for the full rarity model using custom priors. The improper uniform prior was restricted to (-10,10) to aid visualization.
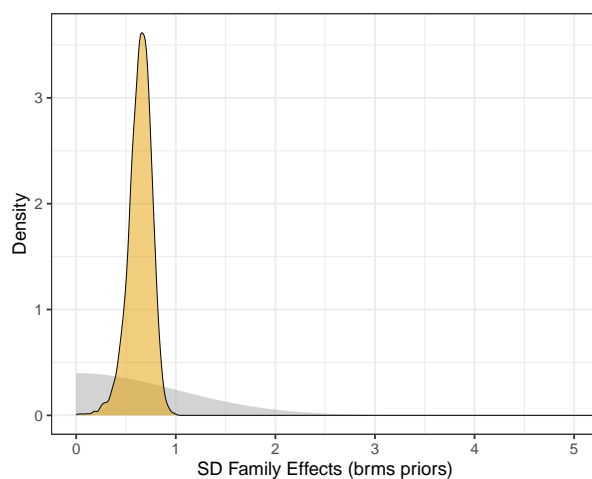
Figure SM 14: Prior (grey) and posterior (yellow) distributions for the Family Effects standard deviation parameter for the full rarity model using custom priors.

2.5   Naturalization model

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Diversification rate	-1.25	0.25	-1.72	-1.25	-0.76	4733.68	1.00
Family age	-0.84	0.28	-1.38	-0.84	-0.30	4829.02	1.00
Hybrids	0.29	0.12	0.05	0.29	0.53	4517.69	1.00
Animal pollinated	-0.56	0.26	-1.06	-0.56	-0.05	4491.14	1.00
Fleshy fruit	-0.01	0.14	-0.28	-0.00	0.26	4757.24	1.00
Climate sum	0.33	0.09	0.14	0.33	0.51	4612.59	1.00
Herbaceous	0.64	0.21	0.24	0.64	1.05	4822.37	1.00
Annual	0.38	0.18	0.02	0.38	0.74	4643.58	1.00
Axsexual	0.45	0.17	0.10	0.45	0.79	4600.48	1.00
SD brownian effects	1.55	0.18	1.16	1.57	1.87	639.44	1.01
SD family specific effects	0.35	0.20	0.02	0.34	0.74	419.10	1.01

Table SM 3: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=395; RMSE=7.57 (+/- 1.08 SD); NRMSE = 0.54 (+/- 0.08).
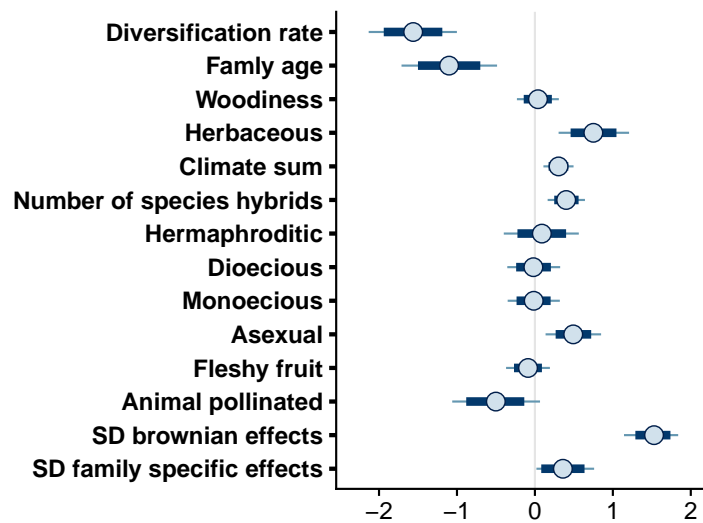

Figure SM 15: Forest plot for estimated β and σ parameters for the naturalization model. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.

14

3 Sensitivity analyses

3.1 Full threatened model: no family effects

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Diversification rate	0.18	0.09	-0.00	0.18	0.36	4842.48	1.00
Famly age	0.21	0.09	0.04	0.21	0.38	4644.13	1.00
Woodiness	0.50	0.03	0.44	0.50	0.56	5171.53	1.00
Herbaceous	-0.71	0.05	-0.81	-0.71	-0.61	5069.99	1.00
Climate sum	-0.37	0.02	-0.41	-0.37	-0.32	4836.18	1.00
Number of species hybrids	-0.15	0.04	-0.24	-0.15	-0.07	4870.95	1.00
Hermaphroditic	0.19	0.06	0.07	0.19	0.31	4867.92	1.00
Dioecious	0.12	0.03	0.05	0.12	0.19	4568.06	1.00
Monoecious	-0.26	0.04	-0.34	-0.26	-0.18	4373.63	1.00
Asexual	0.18	0.03	0.11	0.18	0.25	4663.13	1.00
Fleshy fruit	-0.24	0.02	-0.28	-0.24	-0.20	4868.31	1.00
Animal pollinated	0.92	0.06	0.80	0.92	1.03	5010.02	1.00

Table SM 4: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=236; RMSE=16.20 (+/- 0.98 SD); NRMSE = 0.55 (+/- 0.03).
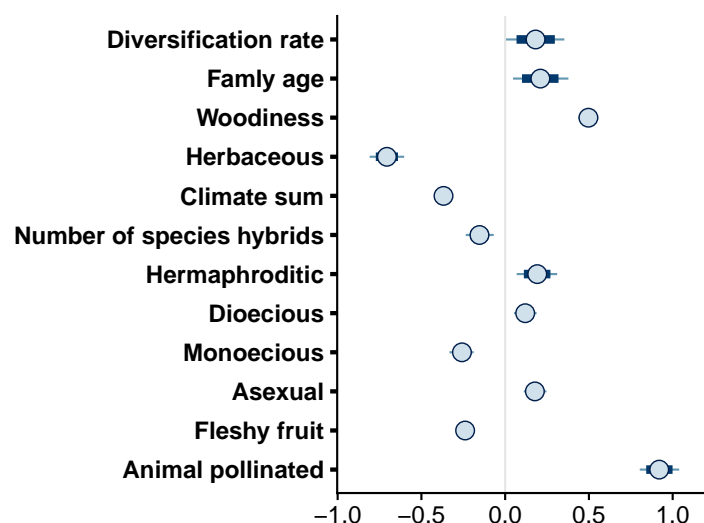
Figure SM 16: Forest plot for estimated β and σ parameters for the threatened model with no family effects. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.

16

3.2 Full threatened model: only non-Brownian family effects

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Diversification rate	0.47	0.32	-0.17	0.47	1.11	3299.25	1.00
Family age	0.49	0.29	-0.06	0.49	1.05	4013.44	1.00
Woodiness	0.65	0.12	0.41	0.65	0.88	3314.83	1.00
Herbaceous	-1.17	0.19	-1.55	-1.17	-0.80	3679.59	1.00
Climate sum	-0.27	0.09	-0.45	-0.27	-0.08	3554.38	1.00
Number of species hybrids	-0.17	0.13	-0.43	-0.17	0.09	4490.73	1.00
Hermaphroditic	0.13	0.23	-0.33	0.13	0.56	3460.29	1.00
Dioecious	-0.18	0.16	-0.47	-0.18	0.13	3374.37	1.00
Monoecious	-0.15	0.16	-0.46	-0.15	0.16	3365.90	1.00
Asexual	0.06	0.16	-0.26	0.06	0.37	3226.05	1.00
Fleshy fruit	0.04	0.15	-0.25	0.05	0.34	4319.92	1.00
Animal pollinated	0.63	0.24	0.16	0.63	1.10	4088.85	1.00
SD family specific effects	0.84	0.06	0.73	0.84	0.97	3310.40	1.00

Table SM 5: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=236; RMSE=5.94 (+/- 0.71 SD); NRMSE = 0.20 (+/- 0.02).
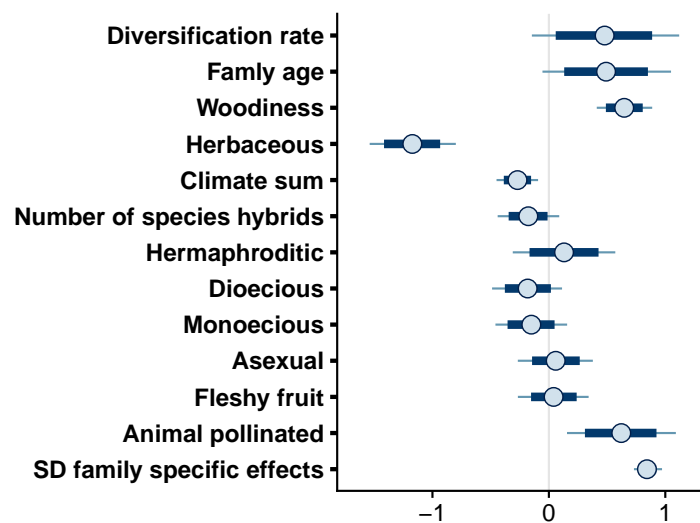

Figure SM 17: Forest plot for estimated β and σ parameters for the threatened model with non-Brownian family effects. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.

3.3   Full threatened model: only Brownian family effects

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Diversification rate	0.55	0.33	-0.12	0.55	1.18	3231.76	1.00
Famly age	0.35	0.33	-0.31	0.35	0.99	2949.88	1.00
Woodiness	0.57	0.14	0.30	0.57	0.84	2759.86	1.00
Herbaceous	-0.89	0.23	-1.35	-0.89	-0.44	2846.98	1.00
Climate sum	-0.24	0.09	-0.42	-0.24	-0.05	3148.80	1.00
Number of species hybrids	-0.19	0.13	-0.44	-0.18	0.06	4371.60	1.00
Hermaphroditic	0.10	0.22	-0.34	0.10	0.54	3358.10	1.00
Dioecious	-0.19	0.15	-0.49	-0.19	0.09	3263.29	1.00
Monoecious	-0.09	0.15	-0.38	-0.10	0.20	3056.87	1.00
Asexual	0.01	0.16	-0.30	0.01	0.32	2792.53	1.00
Fleshy fruit	-0.08	0.12	-0.32	-0.08	0.17	4134.48	1.00
Animal pollinated	0.44	0.25	-0.04	0.44	0.94	3792.81	1.00
SD brownian effects	1.31	0.10	1.13	1.31	1.51	3628.27	1.00

Table SM 6: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=236; RMSE=5.94 (+/- 0.71 SD); NRMSE = 0.20 (+/- 0.02).
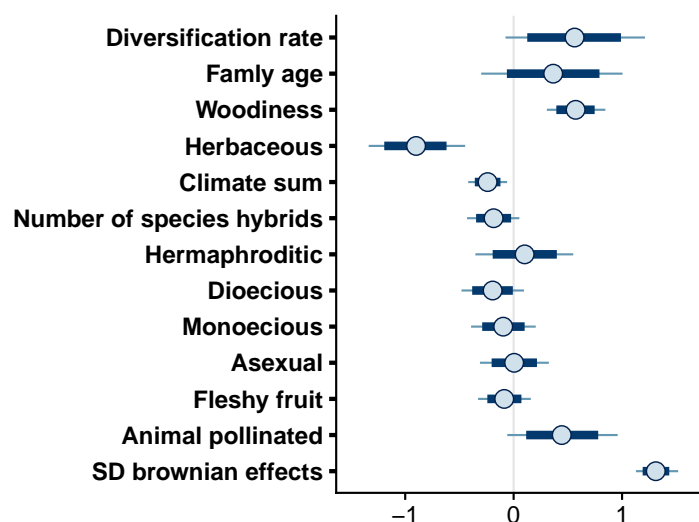

Figure SM 18: Forest plot for estimated β and σ parameters for the threatened model with Brownian family effects. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.

18

3.4 Full threatened model: including non-IUCN vetted species

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Diversification rate	0.69	0.36	0.00	0.69	1.40	4087.61	1.00
Famly age	0.51	0.32	-0.10	0.51	1.15	4584.84	1.00
Woodiness	-0.41	0.16	-0.73	-0.41	-0.10	3899.62	1.00
Herbaceous	-0.62	0.26	-1.12	-0.62	-0.13	2609.26	1.00
Climate sum	-0.18	0.11	-0.40	-0.18	0.03	4592.67	1.00
Number of species hybrids	0.03	0.16	-0.28	0.02	0.34	3696.19	1.00
Hermaphroditic	-0.11	0.27	-0.64	-0.12	0.41	4531.19	1.00
Dioecious	-0.19	0.19	-0.55	-0.19	0.16	4460.67	1.00
Monoecious	0.07	0.18	-0.29	0.07	0.44	4861.42	1.00
Asexual	0.15	0.21	-0.25	0.16	0.56	4629.34	1.00
Fleshy fruit	0.24	0.18	-0.12	0.24	0.60	5103.72	1.00
Animal pollinated	0.17	0.31	-0.41	0.17	0.78	4352.88	1.00
SD brownian effects	0.68	0.32	0.07	0.68	1.33	532.46	1.01
SD family specific effects	1.07	0.13	0.78	1.08	1.29	752.64	1.00

Table SM 7: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=395; RMSE=6.82 (+/- 0.76 SD); NRMSE = 0.49 (+/- 0.05).
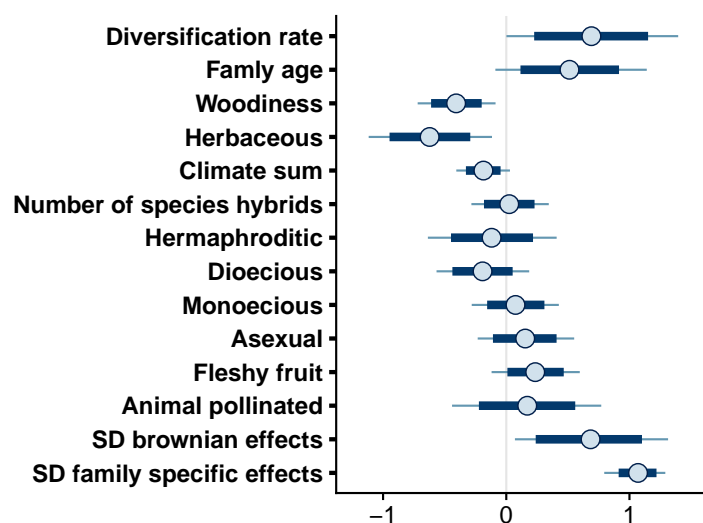
Figure SM 19: Forest plot for estimated β and σ parameters for the threatened model with non-IUCN vetted species. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.

3.5   Invasive model

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
Diversification rate	-0.97	0.31	-1.56	-0.98	-0.33	4530.29	1.00
Family age	-0.76	0.35	-1.46	-0.76	-0.09	4119.70	1.00
Hybrids	0.35	0.15	0.05	0.35	0.63	4144.23	1.00
Animal pollinated	-0.42	0.32	-1.04	-0.42	0.21	3702.58	1.00
Fleshy fruit	0.07	0.16	-0.25	0.07	0.39	4626.87	1.00
Climate sum	0.18	0.11	-0.04	0.18	0.41	4319.14	1.00
Herbaceous	0.66	0.25	0.18	0.66	1.15	4057.81	1.00
Annual	0.22	0.21	-0.19	0.21	0.63	4291.63	1.00
Axsexual	0.42	0.21	0.01	0.43	0.82	3729.91	1.00
SD brownian effects	1.69	0.22	1.20	1.71	2.07	677.12	1.01
SD family specific effects	0.40	0.25	0.02	0.40	0.89	400.47	1.01

Table SM 8: Summary of model output for β and σ parameters, including posterior means, posterior standard deviations, 2.5%, 50%, and 97.5% quantiles, the effective sample size (n_eff), and the potential scale reduction statistic (Rhat). n=395; RMSE=4.68 (+/- 0.66 SD); NRMSE = 0.78 (+/- 0.11).
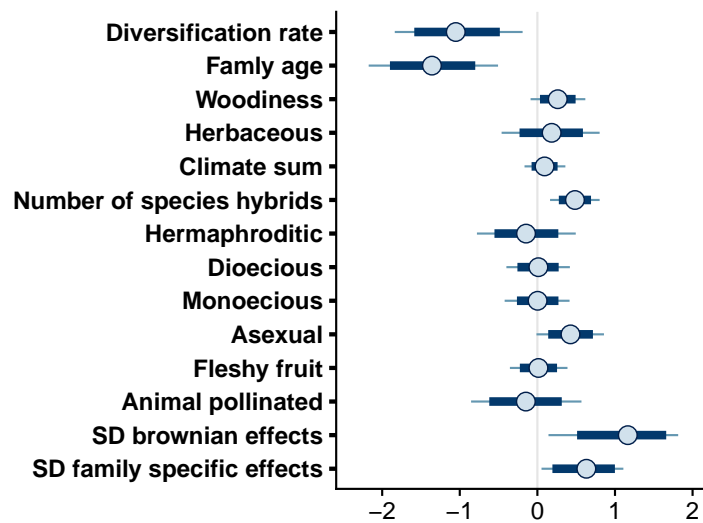

Figure SM 20: Forest plot for estimated β and σ parameters for the invasive model. Points represent posterior means, with thick lines represeting 80% credible intervals, and thin lines representing 95% credible intervals.