Big angiosperm genera are priorities for conservation

Matilda J.M. Brown[1]*, Tarciso C.C. Leão[1]*, Félix Forest[1], Eve Lucas[1], Barnaby E. Walker[1], Eimear Nic Lughadha[1]

[1]Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK

*These two authors contributed equally to this work

Authors for correspondence:
Matilda Brown m.brown2@kew.org
Eimear Nic Lughadha e.niclughadha@kew.org

ORCID:

Matilda J.M. Brown	0000-0003-2536-8365
Tarciso C.C. Leão	0000-0002-7029-1582
Félix Forest	0000-0002-2004-433X
Eve Lucas	0000-0002-7603-435X
Barnaby E. Walker	0000-0002-3884-671X
Eimear Nic Lughadha	0000-0002-8806-4345

Abstract

Species in big angiosperm genera (≥500 species) tend to be less well-known than those in smaller genera. Consequently, species in big genera are underrepresented in conservation at all levels, from analyses to actions. However, emerging evidence indicates that species in larger genera tend to have smaller geographic ranges (a strong correlate of increased extinction risk). Here, we test the global generality of this pattern, and the relationships between plant taxon size, extinction risk and climatic zone on a global scale. Species in larger genera are more likely to have small ranges, less likely to have a global Red List assessment, but are more likely to be threatened. Persistent obstacles to improving conservation documentation of big genera include taxonomic uncertainty, data shortfalls and number of species involved. We highlight the need for closer collaboration between taxonomy and conservation to enable effective conservation of the quarter of plant species that big genera encompass.

Key words: conservation, big genera, genus size, geographic range size, extinction risk, IUCN Red List, angiosperms

Introduction

'Big' plant genera (≥500 species *per* Frodin, 2004) contain over a quarter of angiosperm diversity (Moonlight *et al.*, 2024) and include several ecologically, culturally and economically important taxa (e.g. *Euphorbia*, >2,000 spp.; *Solanum*, >1,200 spp., *Eucalyptus*, >700 spp.). Despite the familiarity of many big plant genera, there remains a

pervasive knowledge shortfall for species in big genera (Moonlight *et al.*, 2024), largely due to the practical challenges associated with working with groups of ≥500 species. For example, big genera are less likely to have a global taxonomic monograph and have lower proportions of species with available DNA sequence data (Grace *et al.*, 2021). Longstanding latitudinal biases in taxonomic effort (the 'Latitudinal Taxonomy Gradient'; (Freeman & Pennell, 2021; Guedes *et al.*, 2025) mean that this shortfall is particularly severe in the tropics, where the majority of big plant genera occur (Moonlight *et al.*, 2024).

Both the practical and taxonomic challenges associated with big plant genera limit our knowledge of extinction risk in these taxa; for example, <5% of the >3,000 species of *Astragalus* have global Red List assessments (far fewer than the c. 20% coverage of angiosperms at large; IUCN, 2024). Undertaking a fully-documented species conservation assessment (e.g. a Red List assessment; IUCN, 2012; IUCN, 2022a) is expensive in terms of time and expertise. Thus, data-poor and taxonomically uncertain species, which are likely to be assessed as Data Deficient (Brummitt *et al.*, 2024) or re-circumscribed (requiring re-assessment, as assessments are specific to a particular species concept) tend to be low priorities for assessment. Assessments are also typically undertaken as part of projects of clearly defined geographic and/or taxonomic scope, often aiming to assess all species within that scope (e.g. the recent, comprehensive assessment of *Dalbergia*; Cowell, 2022). Comprehensive assessment of species in a big genus requires substantial resources, and usually many collaborators. The successful Global Tree Assessment (>45,000 species; Rivers, 2017) demonstrates the feasibility of large assessment projects, but also the enormity of the challenges involved (Newton *et al.*, 2015). We expect that the sheer numbers of species and the many uncertain species circumscriptions in big genera mean that they are systematically underrepresented on the global Red List, limiting our knowledge of extinction risk in a substantial, non-random portion of angiosperm diversity.

There is also emerging evidence that we may be underestimating extinction risk in big plant genera. Geographic range size is the strongest correlate of plant extinction risk (e.g. Nic Lughadha *et al.*, 2020; Walker *et al.*, 2022), and both region-specific and flora-based studies have found that taxa with larger numbers of species have, on average, species with smaller geographic ranges (Schwartz & Simberloff, 2001; Lozano & Schwartz, 2005; Leão *et al.*, 2020) and higher proportions of threatened species (Fu *et al.*, 2022). Furthermore, a

disproportionate number of species described in the last two decades – the majority of which are likely to be threatened (Brown *et al.*, 2023a) – are in the very largest genera (>1,000 species; Moonlight *et al.*, 2024). The global generality of the relationships between genus size, range size and extinction risk in flowering plants remain untested, particularly across climatic zones, lifeforms and higher taxonomic ranks, but we expect that an overall genus size - range size relationship, if supported, will also result in higher proportions of threatened species in larger genera.

Here, we test the generality of the negative relationship between genus size and species geographic range size (as number of botanical countries; Brummitt *et al.*, 2001) across climate, life form and family, using data from the World Checklist of Vascular Plants (Govaerts *et al.*, 2021). We then quantify the extent to which species in larger genera are more or less likely to have a Red List assessment, and their probability of, when assessed, being categorised as either Data Deficient or threatened with extinction, using data from the IUCN Red List (version 2024-1; IUCN, 2024) and comprehensive species-level extinction risk predictions from Bachman *et al.* (2024). Finally, we identify specific priority big genera for conservation assessment as those for which overall extinction risk on the Red List is most underestimated when compared to predictions.

Results

Nearly half of all big genera (and species therein) are tropical (Fig 1, Table S1). Of the 82 currently accepted big genera (from version 13 of the WCVP), 40 (49%) are entirely or mostly tropical, containing 40,631 species (48% of species in big genera; 13% of total angiosperm diversity). All big epiphytic genera (10 genera; 11,370 species) are predominantly tropical; similarly, all big tree genera are either tropical (4 genera; 3,211 species) or subtropical (*Eucalyptus*, 713 species). By contrast, most of the 15 temperate big genera are dominated by herbaceous perennials (12 genera; 16,239 species). There are no big genera dominated by annual species.

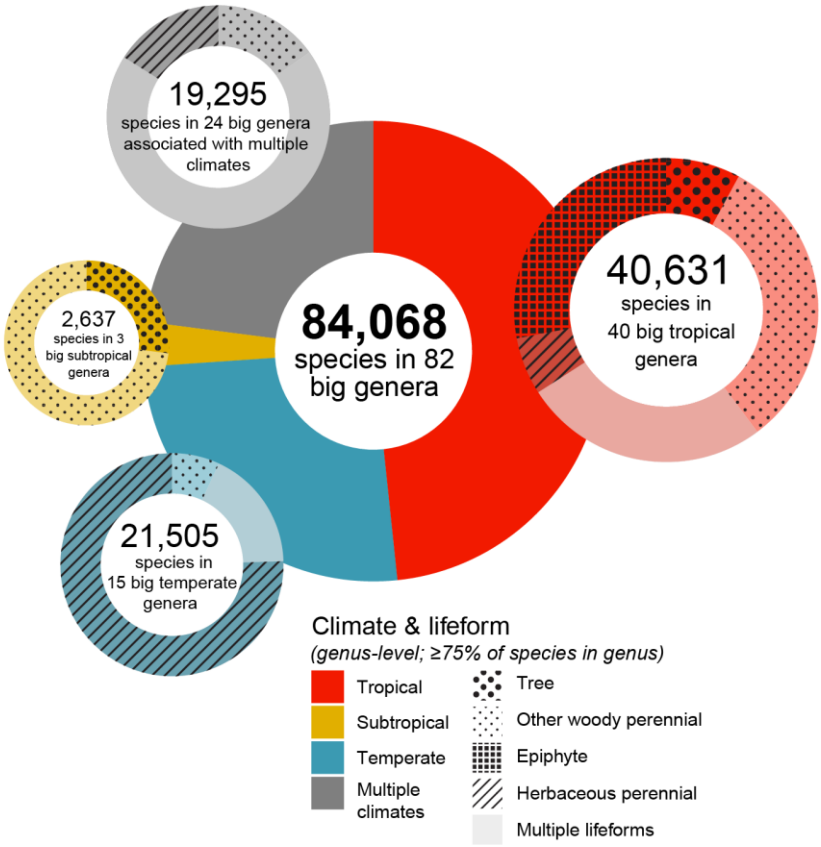


Figure 1. Climate and lifeform distribution in big angiosperm genera. Genera coded by their dominant (≥75% of species) climate/lifeform or as 'multiple climates/lifeforms' (see Methods). Colour and pattern show proportion of species across covariate combinations; note that overall circle sizes are not to scale. For full details on numbers of species and genera across covariates, see Table S1.

Covariates of genus size trends

The best-supported model formula (ranked using AIC; Table S2) across all response variables includes both additive and interaction terms for all covariates (climate, lifeform, family). Family-level interaction terms could not be estimated for families with very few species or genera; the results presented below are thus based on a reduced dataset (91.8% of angiosperm species; see Methods). Models fitted using the full dataset (excluding family effects) are consistent with the findings reported below (Tables S3-4).

Genus size and species range size

Overall, species in big genera tend to have smaller ranges (Fig 2). The probability of being restricted to a single botanical country is 49.7% (95% CI: 49.2-50.2%) for a species in a

monotypic genus, compared to 60.6% (CI: 60.3-60.9%) for species in the ‘smallest’ big genera (i.e. genera of exactly 500 species; from marginal mean estimate; Fig 2a).

The relationship between genus size and range size is consistent across scales (i.e. when modelling the proportion of species restricted to one, two, three, four, five and ten botanical countries, Fig 2a), climatic zones and lifeform (Fig 2b). It is strongest for annual, epiphytic, temperate, and subtropical genera, and comparatively weak (though still significantly positive) for tropical and woody (tree, other woody perennial) genera (Fig 2b).

Across families included in this analysis (100 families; 91.8% of species), the genus size - range size relationship was mixed (Fig 2c); 54 families have significant ($p<0.05$), positive slope coefficients (i.e. species in larger genera are more likely to be restricted to a single botanical country), five families have significant negative coefficients and 41 families had slope coefficients not significantly different from zero.

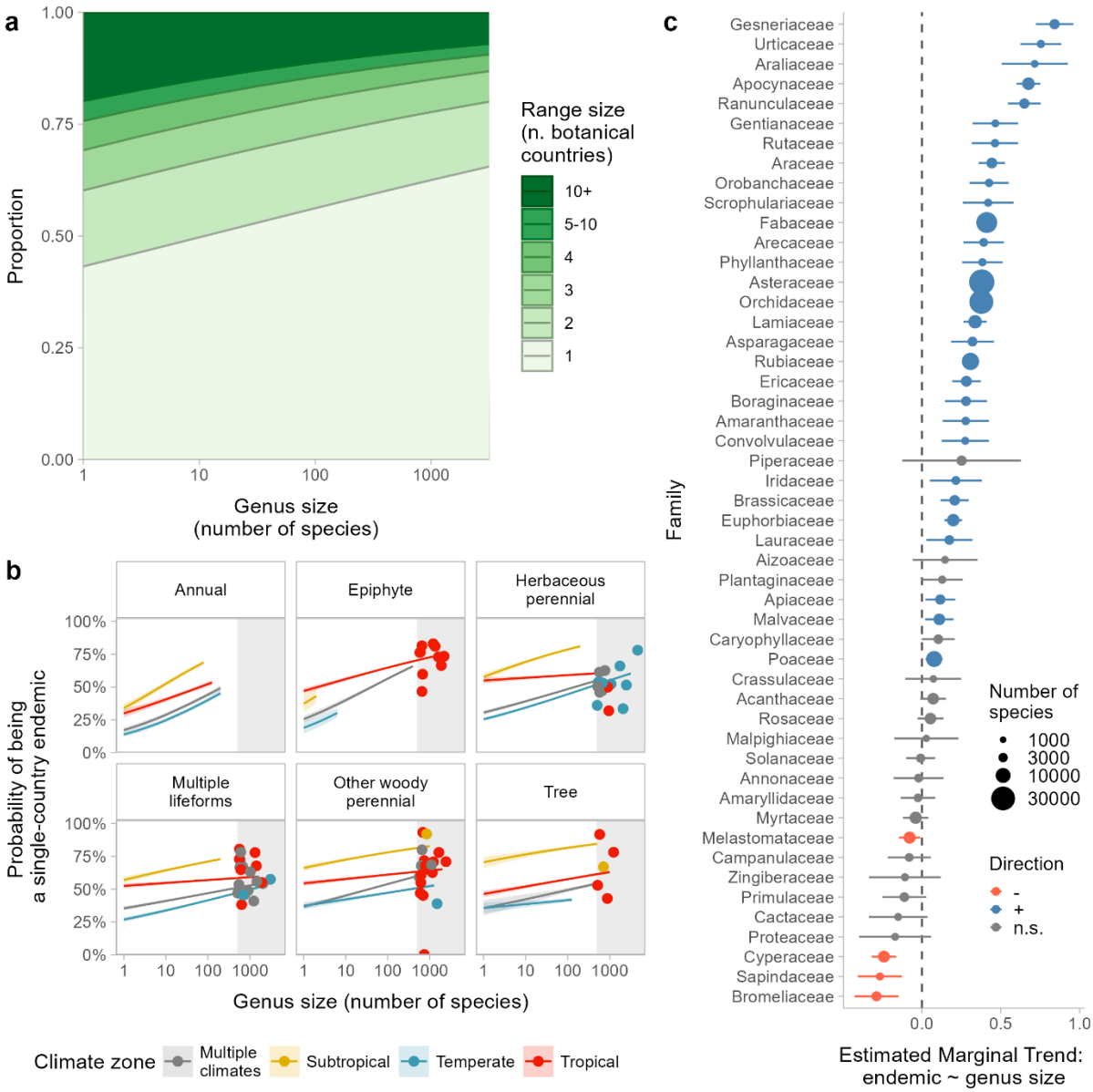


Figure 2. Genus size-range size relationship. Panel a) shows estimated marginal mean proportions of species under each range size threshold (as number of botanical countries) across genus size. The median range size is one botanical country for all but the very smallest genera. Panel b) shows the estimated marginal mean probability that species are restricted to a single botanical country across lifeform and climate. Shaded grey area shows the 'zone of big genera'; points show the observed proportions of single-country endemics in big genera. Panel c) shows the estimated marginal coefficients for family (the single-country endemic threshold; only largest 50 families shown); positive coefficients indicate that species in larger genera are more likely to be restricted to a single botanical country. Units of coefficients on the logit scale. Error bars show 95% CIs of coefficient estimates; colour of points denotes significance and direction; size of points denotes number of species per family. Marginal means and trends estimated across cell weights of the non-focal terms in all panels. For tabulated results with 95% CI values for climate, lifeform and overall marginal trends, see Table S3.

Genus size and extinction risk

Species in big genera are less likely to have a global Red List assessment (Fig 3); for a monotypic genus, the probability of being assessed is 15.3% (95% CI: 14.4-16.2%), compared to 11.5% (CI: 11.2-11.7%) for a species in a 500-species genus (estimated marginal means).

Marginal trends across lifeform and climate are generally consistent with the overall effect, except for subtropical genera – species in these genera are more likely to be assessed if they are in a larger genus (Fig 3a). This result is likely driven by the near-comprehensive (94%) assessment of *Eucalyptus* (Myrtaceae), the only large subtropical tree genus, comprising 27.0% of species (across all lifeforms) in big subtropical genera. The impact of the Global Tree Assessment is clearly visible in our data, with trees having the highest assessment coverage of any covariate (Fig 3a).

Marginal trends across families are less consistent: 49 families have significant negative trends as expected, 19 have positive trends and 32 are not significantly different from zero (50 largest families shown in Fig 3b). Regardless of lifeform, species in tropical genera are always more likely to have been assessed than for other climates (Fig. 3a).

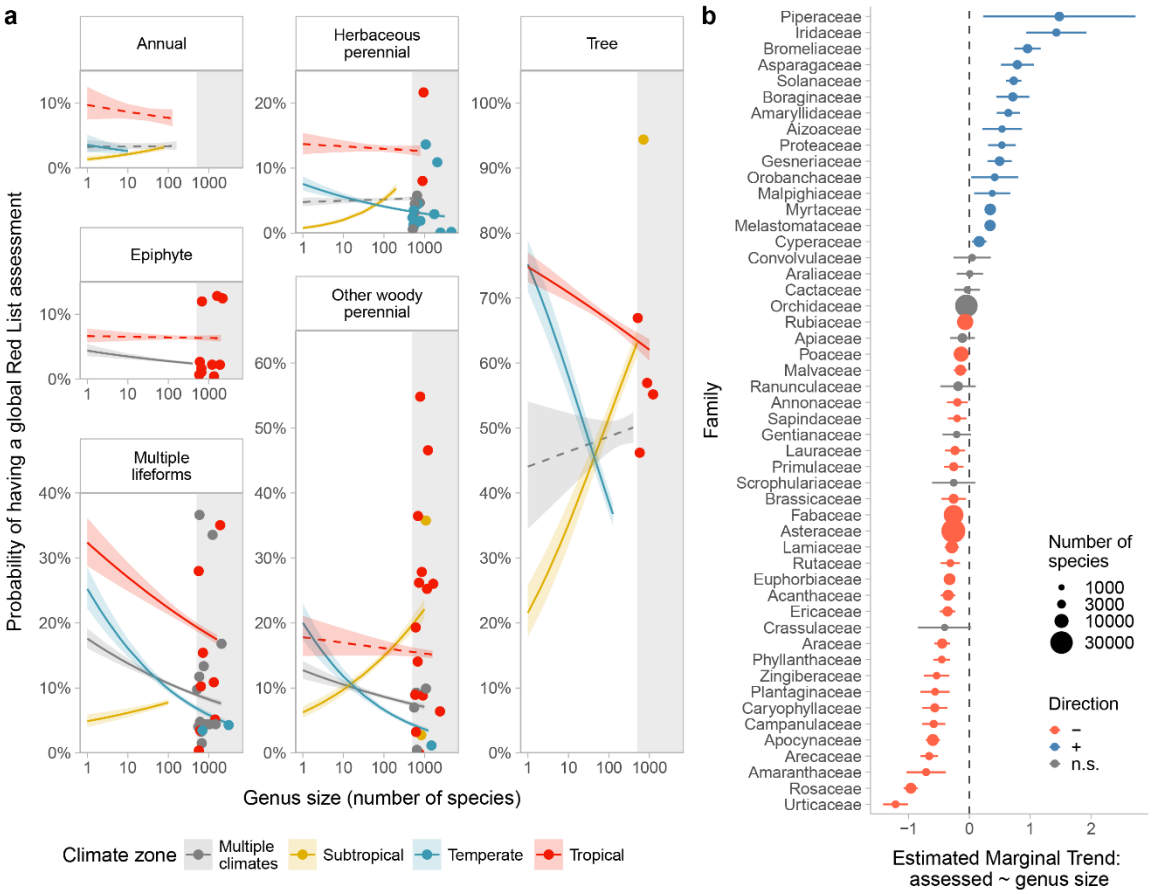


Figure 3. Genus size-assessment probability relationship. Panel a) shows estimated marginal means (i.e. predictions) of assessment probability across genus size, climate and lifeform. Upper limit of genus size values restricted to range of data for each covariate combination. Grey rectangle highlights the 'zone of big genera'; points show the observed proportions of assessed species in big genera. Dashed lines show n.s. trends; solid lines show trends statistically different from zero. Panel b) shows the estimated marginal coefficients for family (the single-country endemic threshold; only largest 50 families shown); negative coefficients indicate that species in larger genera are less likely to have a Red List assessment. Units of coefficients on the logit scale. Error bars show 95% CIs of coefficient estimates; colour of points denotes significance and direction; size of points denotes number of species per family. Marginal means and trends estimated across cell weights of the non-focal terms in all panels. For tabulated results with 95% CI values for climate, lifeform and overall marginal trends, see Table S3.

Using Red List data, we found that species in big genera are slightly more likely to be assessed as threatened (either Vulnerable, Endangered or Critically Endangered; Fig 4a). On average, an assessed species in a monotypic genus has a 32.6% (CI: 31.0-34.1%) chance of being assessed as threatened, compared to 37.5% (CI: 36.8-38.2%) if it is in a genus of 500 species (estimated marginal means). The effect is generally reversed in epiphytic and subtropical genera (Fig 4b), though this is likely driven by a small number of assessments, as

species from epiphytic or subtropical genera tend to be less likely to have assessments (Fig 3). Across the 51 families with sufficient assessment coverage for analysis (see Methods; Fig 4c), 13 have a significant positive relationship as expected, 9 have a significant negative relationship and 29 have relationships between genus size and assessment probability that are not significantly different from zero.

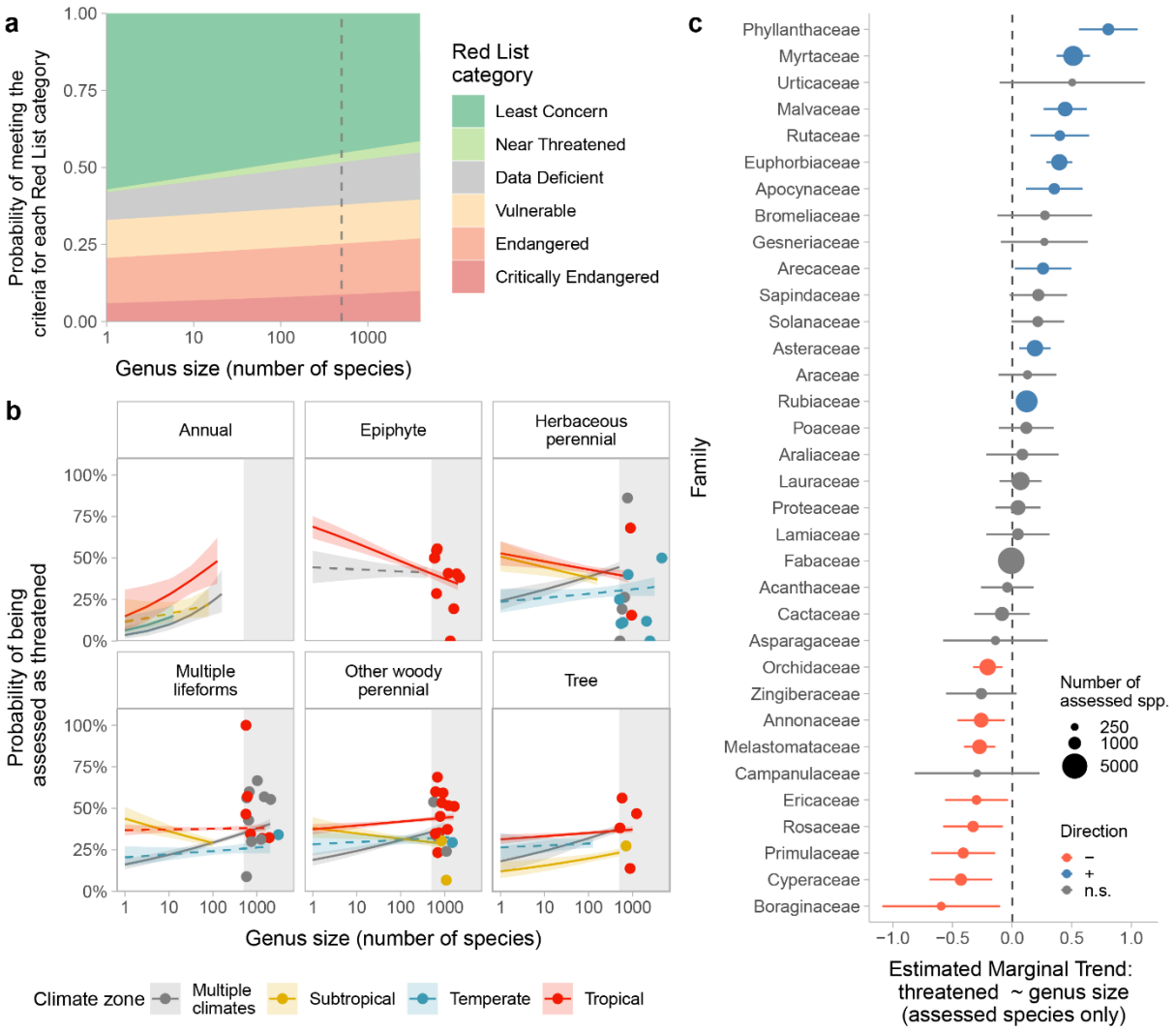


Figure 4. Genus size - documented extinction risk relationship (Red List assessments only). Panel a) shows estimated marginal means for the probability of meeting the criteria for each Red List category (note that categories are cumulative, e.g. all species assessed as Critically Endangered also meet the criteria for Endangered and Vulnerable. Panel b) shows estimated marginal means of being threatened (Vulnerable, Endangered or Critically Endangered) across genus size, climate and lifeform. Upper limit of genus size values restricted to range of data for each covariate combination. Grey rectangle highlights the 'zone of big genera'; points show the observed proportions of species assessed as threatened in big genera. Dashed lines show n.s. trends; solid lines show trends statistically different from zero. Panel c) shows the estimated marginal coefficients for family (only

34 of the largest 50 families shown; 16 of those families have <250 assessed species so were excluded from this analysis); positive coefficients indicate that species in larger genera are more likely to be assessed as threatened. Units of coefficients on the logit scale. Error bars show 95% CIs of coefficient estimates; colour of points denotes significance and direction; size of points denotes number of species per family. Marginal means and trends estimated across cell weights of the non-focal terms in all panels. For tabulated results with 95% CI values for climate, lifeform and overall marginal trends, see Table S3.

Species in larger genera are also more likely to be categorised as Data Deficient (Fig. 5); the probability that a species in a monotypic genus is categorised as Data Deficient is 4.5% (95% CI:4.1-4.9%), compared to 9.1% (CI: 8.7-9.4%) in a genus with 500 species. The effect on Data Deficiency of being in a large genus is strongest for tropical and epiphytic species, but reversed for annual and subtropical species (i.e. species in larger genera are less likely to be assessed as Data Deficient), though these groups have, respectively, only 18 and 94 species (zero and seven in big genera) that have been assessed as Data Deficient. Data were insufficient to fit the full model including family; full model for Data Deficiency fit with climate and lifeform covariates only.

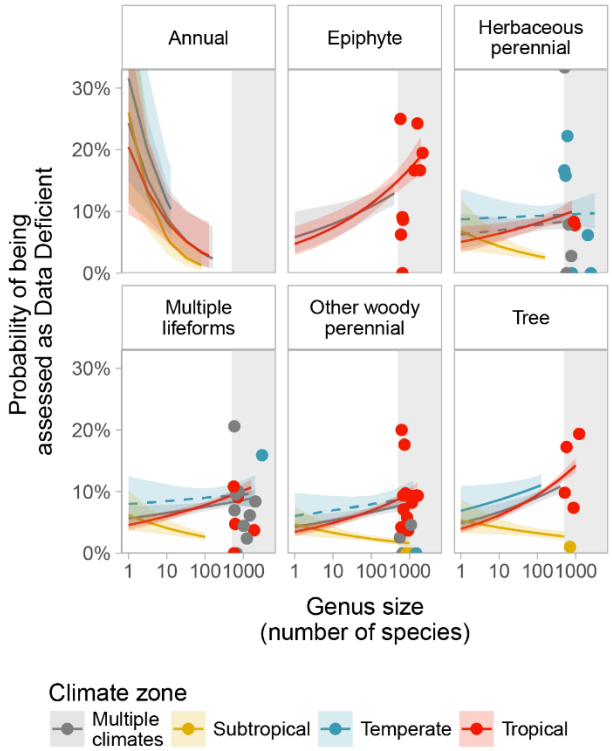


Figure 5. Genus size – Data Deficiency relationship (Red List assessments only). Estimated marginal means of being assessed as Data Deficient across genus size, climate and lifeform. Upper limit of genus size values restricted to range of data for each covariate combination. Note that y-axes are

specific to each panel. Shaded ribbons show 95% CIs. Dashed lines show n.s. trends; solid lines show trends that are statistically different from zero. For tabulated results with 95% CI values for climate, lifeform and overall marginal trends, see Table S3.

When we extended our analysis of extinction risk to include all angiosperms, by including species with extinction risk predictions (from Bachman *et al.* 2024), we observed a stronger overall relationship between genus size and extinction risk (Fig 6): a species in a monotypic genus has a 30.4% (CI: 29.8-31.1%) chance of being threatened (assessed or predicted with high confidence), compared to 45.6% (CI: 45.3-45.9%) for a genus of 500 species (estimated marginal means).

The relationship between genus size and probability of being threatened is consistent across climatic zones and lifeforms (Fig 6b). It is strongest for annual, epiphytic and genera found in multiple climates, and comparatively weak (though still significantly positive) for tropical, subtropical and woody genera (Fig 6b). Across families (Fig 6c), the genus size – extinction risk relationship was broadly consistent, though statistical power was limited; of the 100 families included in this analysis, 56 have significant ($p<0.05$), positive slope coefficients (i.e. species in larger genera are more likely to be either assessed as or predicted to be threatened), five have significant negative coefficients and 39 have slope coefficients that are not significantly different from zero. When we compared estimates of the probability that a species is threatened based on assessments versus assessments plus predictions (Extended Data Figure 1), we found strong evidence to suggest that current Red List data underestimate the proportion of species that are threatened with extinction in genera of eleven or more species, and that this underestimation is most severe for big genera.

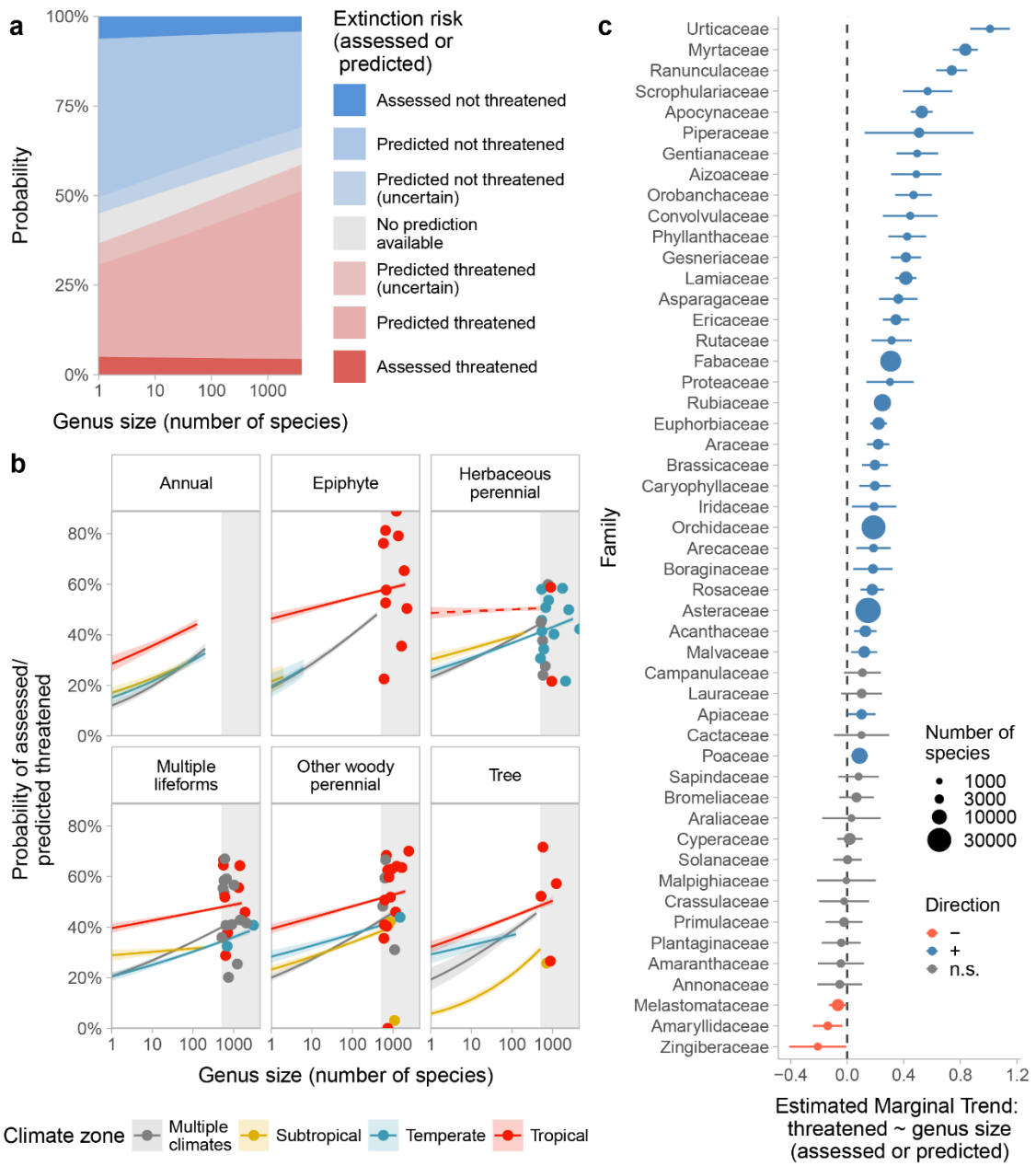


Figure 6. Genus size - extinction risk relationship (Red List assessments and extinction risk predictions from Bachman et al. 2024). Panel a) shows estimated marginal means across all thresholds of estimated extinction risk. Panel b) shows estimated marginal means of being threatened (assessed or predicted with high confidence) across genus size, climate and lifeform. Upper limit of genus size values restricted to range of data for each covariate combination. Grey rectangle highlights the 'zone of big genera'; points show the proportions of species in big genera that are assessed or predicted to be threatened. Dashed lines show n.s. trends; solid lines show trends that are statistically different from zero. Panel c) shows the estimated marginal coefficients for family (only largest 50 families shown); positive coefficients indicate that species in larger genera are more likely threatened (either assessed or predicted). Units of coefficients on the logit scale. Error bars show 95% CIs of coefficient estimates; colour of points denotes significance and direction; size of points denotes number of species per family. Marginal means and trends estimated across cell weights of the non-focal terms in all panels. For tabulated results with 95% CI values for climate, lifeform and overall marginal trends, see Table S3.

248 From the extinction risk predictions, 37,072 species in big genera are estimated (with high
249 confidence) to be threatened with extinction, but either lack a global Red List assessment
250 (36,411) or have been assessed as Data Deficient (661) – together representing 10.9% of all
251 angiosperm species. This number is conservative, as >13,000 species have been described or
252 otherwise added to the World Checklist of Vascular Plants since the release used as the
253 backbone for predictions. Of the species in big genera predicted to be threatened, 19,787
254 are in tropical genera (spanning 40 genera). We quantified the degree to which extinction
255 risk is likely to be underestimated in each large genus by subtracting the Red-List-derived
256 proportion of threatened species from that calculated using predictions (Extended Data
257 Table 1). Of the twenty big genera in which extinction risk is likely to be most
258 underestimated on the Red List, six are tropical genera in the family Orchidaceae. The most
259 extreme example is *Stelis* with <1% of species having a Red List assessment and none
260 threatened, while 79% of species are predicted to be threatened.

261 **Discussion**

262 Consistent with our expectations, species in big genera are more likely to have small ranges,
263 less likely to have Red List assessments, and if assessed, they are more likely to be assigned
264 to one of the threatened categories or categorised as Data Deficient. The relationships
265 between genus size, range size and extinction risk are broadly consistent across climatic
266 zones, lifeforms, and families; as over 25% of plant species are in 'big' genera (≥500 species)
267 and a further 35% are in moderately large genera (≥100 species), our findings are relevant
268 to the majority of known plant species and have significant implications for conservation
269 science, action and policy.

270 Our results support previous studies on the relationship between genus size and range size
271 in plants (Schwartz & Simberloff, 2001; Lozano & Schwartz, 2005; Leão *et al.*, 2020; family-
272 level Fu *et al.*, 2022), contrasting with earlier reports for both plants (Vamosi & Wilson,
273 2008) and animals (Purvis *et al.*, 2000) that smaller – especially monotypic – taxa tend to
274 have higher levels of extinction risk. However, the relationship between genus and range
275 size is not universal – in five families, species in larger genera are less likely to be restricted
276 to a single botanical country, contrary to expectations (Table S5). These families span seven
277 orders (both eudicots and monocots), vary in climate, lifeform and range sizes, and three

contain at least one big genus, so there is no clear pattern to these outliers. Similarly, we found groups for which the general pattern of higher extinction risk in larger genera is reversed, where species in larger genera are *less* likely to be either assessed as or predicted to be threatened (Figs 3-4; Tables S3, S5).

Results based on published assessments were less consistent than those based on predictions, likely due to the very low assessment coverage of many big genera. For example, in tropical epiphytes, where seven of the ten big genera have fewer than 50 published assessments (representing <3% of species in these seven genera), assessments show an apparent negative relationship between genus size and extinction risk (Fig 4), but the model including predictions showed the expected positive trend (Fig 5). There are five families with a negative relationship between genus size and extinction risk in prediction-based models. For one of these, Melastomataceae, this may be linked to range, as it was one of the five families where species in larger genera tended to have larger ranges (Table S5), but for the other four the reasons for this result are unclear; differences in both range size and extinction risk may be linked to the 'different ways to be a big genus' (an outstanding knowledge gap noted by Moonlight *et al.*, 2024). It is also possible that there are biases in the predictions that affect our results – this does not in any way diminish the need for species in big genera to be prioritised for conservation assessment, whether to formally document their predicted extinction risk, address the gaps in both knowledge and data for big genera or to better understand the drivers of extinction risk across taxon size.

The mechanisms underpinning the link between genus and range sizes may be related to evolutionary diversification patterns. More rapidly diversifying plant genera tend to have smaller ranges and higher extinction risk (Davies *et al.*, 2011; Leão et al., 2020; Tanentzap et al,. 2020), though conifers showed the opposite trend (Tanentzap et al. 2020). In animals, genus-level diversification rate appears to be a strong correlate of both range size and extinction risk (though there is significant variation across clades; Greenberg *et al.*, 2021), further supporting the hypothesis that rapid speciation tends to produce range-restricted species which are inherently more vulnerable to extinction (Davies *et al.*, 2011; Leão *et al.*, 2020).

Our results complement recent studies characterising biases in the representation of plants on the Red List and addressing how these can be quantified and corrected (Nic Lughadha *et*

309 *al.*, 2020) or circumvented (e.g. Bachman *et al.*, 2024) in future analyses to refine
310 understanding of global and regional patterns of plant extinction risk. For example, eight of
311 the 20 most under-estimated genera (when overall extinction risk is estimated using the Red
312 List; Table 1) are Orchidaceae (6 genera) or Asteraceae (2 genera). These two largest plant
313 families encompass c. 65,000 species and are recognised as being among the most under-
314 represented on the Red List (Nic Lughadha *et al.*, 2020).

315 Strikingly, the six orchid genera highlighted here for their underestimated threat levels are
316 all tropical, consistent with our original expectation that under-representation of big genera
317 in assessment processes might be greatest in the tropics. However, contrary to our
318 expectations, we found that species in tropical genera, across all lifeforms, were more likely
319 to have been assessed than species of other climates (Fig 3a). This encouraging result
320 suggests significant progress towards addressing the plant Red Listing deficit which was so
321 evident across many areas of the tropics as recently as a decade ago (Bachman *et al.*, 2019).
322 Much of this progress is attributable to the Global Tree Assessment, and the high
323 assessment coverage achieved for trees explains the relatively weak effect of genus size on
324 assessment in tree-dominated genera. Similarly concerted and large-scale endeavours are
325 urgently needed for other plants, and our findings suggest that an explicit focus on
326 comprehensive assessments of selected big genera could be an important and worthwhile
327 approach to start to redress the effects of neglect or avoidance of these groups over
328 decades.

329 Importantly, we are not suggesting that large genera be split, or that there should be an
330 upper limit on the size of genera. The size of genera is related to both evolutionary and
331 human processes (e.g. Bentham's preference for fewer, larger genera; Humphreys & Linder,
332 2009), and the highly skewed pattern of genus size frequency, with many small and
333 relatively few very large genera, is common across the tree of life (e.g. Strand & Panova,
334 2015; Sigwart *et al.*, 2018). Although ultimately arbitrary, there is a broad consensus that
335 'good' genera (monophyletic, stable and diagnosable) are useful divisions (Humphreys &
336 Linder, 2009; Vences *et al.*, 2013; Strand & Panova, 2015; Muñoz‐Rodríguez *et al.*, 2023),
337 so big plant genera (and their problems) are unlikely to be split out of existence (Moonlight
338 *et al.*, 2024). However, regardless of the basis of genus size, our results show that there are

339 tangible implications of being a member of a big genus that can influence species'
340 persistence.

341 We strongly encourage greater focus and resourcing of the further study of big plant genera
342 in taxonomic, ecological and conservation contexts. Although big genera continue to
343 represent an immense amount of work to monograph *in toto*, molecular data can be used to
344 divide large taxa into more manageable subclades; as in *Eugenia* or *Myrcia* (e.g., Mazine *et*
345 *al.*, 2014; Lucas *et al.*, 2018). Another approach may be the 'foundation monograph'
346 concept (Scotland & Wood, 2012) – a practical guide to this approach (based on the recent
347 monograph of *Ipomoea*; Muñoz-Rodríguez *et al.*, 2019) is detailed in Muñoz-Rodríguez *et al.*
348 (2024). We echo earlier recommendations for increased collaboration across conservation
349 and taxonomy (e.g. Nic Lughadha *et al.*, 2019; Brown *et al.*, 2023a). The success of any new
350 initiative focused on assessing big genera will depend on a clear-eyed, evidence-based
351 understanding of the challenges likely to be faced; among these is the higher proportion of
352 Data Deficient species in genera which lack recent monographs (Grace *et al.*, 2021) and are
353 likely to include unusually large proportions of species not yet described (Moonlight *et al.*,
354 2024). We found that species from big genera were at least twice as likely to be categorised
355 as Data Deficient as those in monotypic genera, a result consistent with our expectations.
356 However, this proportion remained relatively low overall (<10%, Fig 5) and could be further
357 minimised by effective collaboration between taxonomists, assessors and other local
358 experts to assemble the most complete information possible for species' assessments.
359 Where assessors identify species likely to be Data Deficient, this can feed back into
360 taxonomic prioritisation (e.g. targeted fieldwork or other data collection). Similarly,
361 taxonomists can provide input on the likelihood of species re-circumscriptions, facilitating
362 effective prioritisation of assessments in big genera – thus, a shared pipeline of activity
363 between taxonomists and assessors can allow extinction risk assessments of species in
364 speciose genera to be undertaken before a genus is fully monographed (Nic Lughadha *et al.*,
365 2019).

366 Species in big genera are undoubtedly conservation priorities; they are under-studied,
367 under-assessed, and their extinction risk has been underestimated. Over a quarter of
368 angiosperm diversity is in big genera, but overcoming the challenges of working with

speciose taxa will only be possible with support and further strengthening of joint efforts across taxonomy and conservation.

Methods

Data

We calculated genus size as the number of accepted species in each genus in version 13 of the World Checklist of Vascular Plants (WCVP; Govaerts *et al.*, 2021), accessed via the 'rWCVP' and 'rWCVPdata' R packages (Brown *et al.*, 2023b). Most species lack comprehensive and/or unbiased point data from which to measure range size; instead, as a proxy for range size, we used the number of Level 3 Areas (hereafter 'botanical countries', following e.g. Bachman *et al.*, 2024) of the World Geographic Scheme for Recording Plant Distributions (Brummitt *et al.*, 2001), accessed via WCVP. Although number of botanical countries is an imperfect proxy for range size (skewed towards single-country endemics, low spatial resolution and unequally-sized botanical countries), it remains the most comprehensive and least biased available data for angiosperms and shows a robust correlation with extent of occurrence, a common measure of range size (Bureš *et al.*, 2024).

We extracted lifeform information from WCVP, then simplified the Raunkiaer lifeforms into annual, herbaceous perennial, tree, other woody perennial, and epiphyte following Humphreys et al. 2019 with modifications (2019; see Table S1). For climatic zones, we derived the broad climate types (tropical, subtropical, temperate) from the WCVP 'climate description' field or species' distributions; species spanning multiple climate zones were combined into a 'multiple climates' category. We focused on the genus-level effects of climate and lifeform, rather than species-level (e.g. species in broadly tropical genera). To assign genus-level climate and lifeform categories, we aggregated species-level climate and lifeform data from WCVP to genus-level using a 75% dominance threshold – e.g. 89.3% of species in *Tillandsia* are epiphytic and 81.4% are tropical, so we treated *Tillandsia* as a tropical, epiphytic genus. Where no single lifeform or climate dominated, genera were assigned to a single 'multiple climates/lifeforms' category. This aggregated approach had the additional benefit of allowing us to include in our analyses the c. 78,480 species lacking lifeform information in WCVP version 13. We obtained extinction risk assessments from the IUCN Red List of Threatened Species version 2024-1 (hereafter Red List; IUCN, 2022b), and extinction risk predictions from Bachman *et al.* (2024) using 'rWCVP' (Brown *et al.*, 2023b) for taxonomic reconciliation.

402 **Analyses**

403 We used logistic regression to model five response variables as a function of genus size: (1)
404 range size, (2) probability of having a global Red List assessment, (3) probability of being
405 categorised as Data Deficient on the Red List, (4) probability of being assessed as threatened
406 on the Red List (i.e. categorised as Vulnerable, Endangered or Critically Endangered), and (5)
407 probability of being either assessed as or predicted to be threatened. Genus size was $\log_{10}$-
408 transformed prior to analysis; all effect sizes represent changes on this scale. To investigate
409 the consistency of the genus size-range size relationship across angiosperm diversity, we
410 included three covariates in our analyses: climatic zone, lifeform and family. For each of the
411 five response variables fitted, our overall approach was the same. We first fitted a 'full
412 model', which included all covariates and their interaction with genus size. We then
413 performed model selection using AIC (via the 'dredge' function implemented in 'MuMIn' ;
414 Barton, 2024) to identify the best model. To interpret the effect sizes and their associated
415 uncertainty for models with multiple categorical interaction terms, we used estimated
416 marginal means and estimated marginal trends for each value of covariates using the
417 'emmeans' and 'emtrends' functions (from the 'emmeans' R package; Lenth, 2024). To
418 account for imbalances in these variables, we weighted the non-focal covariates by their
419 proportional frequency in the dataset (weights = "cells"). For models that included a family-
420 level interaction term (including those used for model selection), we removed families with
421 fewer than five genera or fewer than 500 species. Although this retained only 25% of
422 angiosperm families, the effect on total species and generic diversity was small; the reduced
423 dataset of 100 families included 89.3% of angiosperm genera and 91.8% of angiosperm
424 species. Removing Aquifoliaceae (which comprises one genus), Balsaminaceae (two genera),
425 Begoniaceae (two genera), Dioscoreaceae (four genera) and Ebenaceae (three genera)
426 resulted in the removal of five big genera from our initial (full) models: *Begonia* (2,092
427 species), *Dioscorea* (635 species), *Diospyros* (778 species), *Ilex* (564 species) and *Impatiens*
428 (1,120 species). For models fitted to assessed species only, we removed families with fewer
429 than 250 assessments on the Red List (49 families). To check for bias resulting from these
430 reductions, we also fitted models without family effects to the full species dataset.

431 To account for the skew in range size distribution and lack of fine-scale resolution when
432 using botanical countries, we modelled range size as a function of genus size using a series

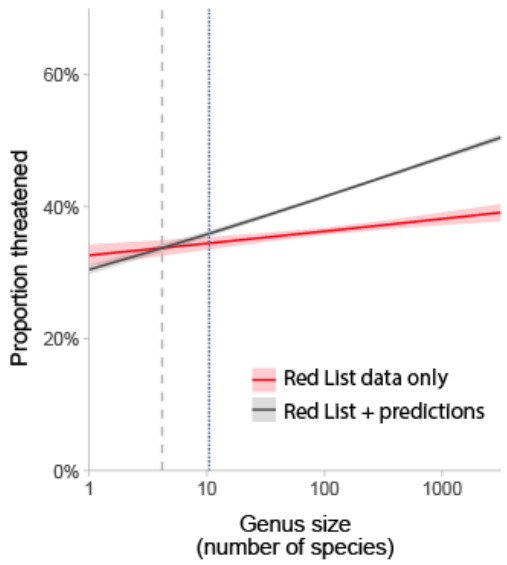
of logistic regressions. We repeated model-fitting at multiple range-size thresholds using a 'multiple threshold' approach to test the robustness of this relationship at different scales. We modelled the probability that species are endemic to a single botanical country, restricted to two or fewer botanical countries, then to three, four and ten or fewer botanical countries.

We used the same 'multiple threshold ' approach to model extinction risk as for previous studies (e.g. Brown *et al.*, 2023a; Soto Gomez *et al.*, 2024); for the full models including climate, lifeform and family we focused on a single threshold: the probability of being 'threatened' per the IUCN (Critically Endangered, Endangered, Vulnerable; IUCN, 2022a), or predicted to be threatened with high confidence (Bachman *et al.*, 2024). Species assessed as Extinct (84 species in reduced dataset) or Extinct in the Wild (31 species) could not be modelled separately, so were included under Critically Endangered for all analyses.

To highlight the implications of under-assessment of species in big genera, we directly compared the proportion of threatened species (using the threshold above, noting that this differs from the IUCN calculation in the handling of DD species) across genus sizes based on assessments only, and assessments plus predictions. From these data, we extracted the 20 big genera with the highest mismatch between their proportion threatened per the Red List, and their estimated proportion threatened from the Bachman et al. (2024) predictions. We present these genera as potential priorities for conservation assessment.

All models were fitted with species as the observation (i.e. data were not aggregated to genus prior to analyses), in R version 4.4.1 (R Core Team, 2024). All data, scripts and outputs required to reproduce our results are available at https://github.com/matildabrown/BigAngiospermGenera [NB: currently private, materials provided separately for review].

Extended Data



Extended Data Figure 1. Comparison of the estimated marginal means for Red List data (red) and combined (Red List + predictions) data (grey) for the relationship between genus size and probability of extinction risk. Dotted lines show where the lines intercept (grey; left) and where the 95% CIs diverge (blue; right), indicating the change from over- to under-estimation of threat.

466 *Extended Data Table 1. Numbers of species assessed as versus predicted to be threatened in big*
467 *genera, arranged in descending order of underprediction (Δthreat prop.). Top 20 genera shown here,*
468 *for full list of big genera see Table S4.*

Genus	Climate	# spp.	prop. spp. assessed	# spp. threatened (RL)	prop. threatened (RL)	# spp. predicted threatened	prop. predicted threatened	**Δthreat prop.**
Stelis	tropical	1340	0.4%	0	0.0%	1060	79.1%	79.1%
Taraxacum	temperate	2490	0.1%	0	0.0%	1243	49.9%	49.9%
Lupinus	multiple climates	580	11.7%	6	8.8%	333	57.4%	48.6%
Lepanthes	tropical	1203	2.2%	11	40.7%	1058	87.9%	47.2%
Saussurea	multiple climates	516	0.6%	0	0.0%	232	45.0%	45.0%
Pedicularis	temperate	679	4.4%	3	10.0%	342	50.4%	40.4%
Ranunculus	temperate	1750	2.9%	11	21.6%	1011	57.8%	36.2%
Elatostema	multiple climates	654	0.5%	1	33.3%	435	66.5%	33.2%
Primula	temperate	545	3.5%	2	10.5%	224	41.1%	30.6%
Corydalis	temperate	538	1.3%	2	28.6%	310	57.6%	29.0%
Pleurothallis	tropical	579	0.7%	2	50.0%	439	75.8%	25.8%
Masdevallia	tropical	651	1.7%	6	54.5%	523	80.3%	25.8%
Piper	tropical	2425	6.4%	65	41.9%	1634	67.4%	25.4%
Epidendrum	tropical	1878	2.2%	17	40.5%	1210	64.4%	24.0%
Maxillaria	tropical	652	1.1%	2	28.6%	341	52.3%	23.7%
Oxytropis	temperate	608	1.5%	1	11.1%	208	34.2%	23.1%
Anthurium	tropical	1325	10.9%	46	31.9%	691	52.2%	20.2%
Ardisia	tropical	737	26.2%	68	35.2%	393	53.3%	18.1%
Berberis	multiple climates	626	9.3%	22	37.9%	350	55.9%	18.0%
Allium	temperate	1078	13.6%	31	21.1%	403	37.4%	16.3%

469

470 **Acknowledgements**

471 TCCL was funded by a Future Leader Fellowship from the Royal Botanic Gardens, Kew.

472 **Author contributions**

473 MJMB: Conceptualisation, Formal analysis, Visualisation, Investigation, Writing - Original
474 Draft
475 TCCL: Conceptualisation, Investigation, Writing - Original Draft, Review & Editing
476 FF: Writing - Review & Editing
477 EL: Writing - Review & Editing
478 BEW: Writing - Review & Editing
479 ENL: Conceptualisation, Investigation, Writing - Review & Editing

Data availability

All source data are publicly available. Scripts and other materials required to reproduce our outputs are available at https://github.com/matildabrown/BigAngiospermGenera

Competing interests

The authors declare no competing interests.

References

Bachman SP, Brown MJM, Leao TCC, Nic Lughadha E, Walker BE. 2024. Extinction risk predictions for the world's flowering plants to support their conservation. *New Phytol* **242**(2): 797-808.

Bachman SP, Field R, Reader T, Raimondo D, Donaldson J, Schatz GE, Lughadha EN. 2019. Progress, challenges and opportunities for Red Listing. *Biological Conservation* **234**: 45-55.

Barton K 2024. MuMIn: Multi-Model Inference. R package.

Brown MJM, Bachman SP, Nic Lughadha E. 2023a. Three in four undescribed plant species are threatened with extinction. *New Phytol* **240**(4): 1340-1344.

Brown MJM, Walker BE, Black N, Govaerts RHA, Ondo I, Turner R, Nic Lughadha E. 2023b. rWCVP: a companion R package for the World Checklist of Vascular Plants. *New Phytol* **240**(4): 1355-1365.

Brummitt N, Cui Q, Bai P, Liu X, Scott B, De Palma A, Araujo AC. 2024. Reducing the number of Data Deficient plant species. *Systematics and Biodiversity* **22**(1): 2302180.

Brummitt RK, Pando F, Hollis S, Brummitt N. 2001. *World geographical scheme for recording plant distributions*. Carnegie Mellon University (Pittsburgh): Hunt Institute for Botanical Documentation.

Bureš P, Elliott TL, Veselý P, Šmarda P, Forest F, Leitch IJ, Nic Lughadha E, Soto Gomez M, Pironon S, Brown MJ. 2024. The global distribution of angiosperm genome size is shaped by climate. *New Phytologist* **242**(2): 744-759.

Cowell C. WE, Bullough L.-A., Grey J., Klitgaard B., Govaerts R., Andriambololonera S., Cervantes A., Crameri S., Lima, H.C., Lachenaud O., Li S.-J., Linares J.L., Phillipson P ., Rakotonirina N., Wilding N., van der Burgt X., Vatanparast M., Barker A., Barstow M., Beentje H., Plummer J. 2022. CITES Dalbergia Checklist. Commissioned by the CITES Secretariat. . Surrey, UK.

Davies TJ, Smith GF, Bellstedt DU, Boatwright JS, Bytebier B, Cowling RM, Forest F, Harmon LJ, Muasya AM, Schrire BD. 2011. Extinction risk and diversification are linked in a plant biodiversity hotspot. *PLoS biology* **9**(5): e1000620.

Freeman BG, Pennell MW. 2021. The latitudinal taxonomy gradient. *Trends in Ecology & Evolution* **36**(9): 778-786.

Fu Q, Huang X, Li L, Jin Y, Qian H, Kuai X, Ye Y, Wang H, Deng T, Sun H. 2022. Linking evolutionary dynamics to species extinction for flowering plants in global biodiversity hotspots. *Diversity and Distributions* **28**(12): 2871-2885.

Govaerts R, Nic Lughadha E, Black N, Turner R, Paton A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci Data* **8**(1): 215.

Grace OM, Pérez-Escobar OA, Lucas EJ, Vorontsova MS, Lewis GP, Walker BE, Lohmann LG, Knapp S, Wilkie P, Sarkinen T. 2021. Botanical monography in the Anthropocene. *Trends in Plant Science* **26**(5): 433-441.

Greenberg DA, Pyron RA, Johnson LG, Upham NS, Jetz W, Mooers AØ. 2021. Evolutionary legacies in contemporary tetrapod imperilment. *Ecology Letters* **24**(11): 2464-2476.

Guedes JJ, Moura MR, Jardim L, Diniz-Filho JAF. 2025. Global Patterns of Taxonomic Uncertainty and its Impacts on Biodiversity Research. *Systematic biology*: syaf010.

Humphreys AM, Linder HP. 2009. Concept versus data in delimitation of plant genera. *Taxon* **58**(4): 1054-1074.

IUCN. 2012. *IUCN Red List categories and criteria: version 3.1*. Gland, Switzerland and Cambridge, UK: Prepared by the IUCN Species Survival Commission.

IUCN 2022a. Guidelines for Using the IUCN Red List Categories and Criteria.

IUCN 2022b. The IUCN Red List of Threatened Species.

IUCN 2024. The IUCN Red List of Threatened Species.

Leão TC, Lughadha EN, Reich PB. 2020. Evolutionary patterns in the geographic range size of Atlantic Forest plants. *Ecography* **43**(10): 1510-1520.

Lenth R 2024. emmeans: Estimated marginal means, aka least-squares means. R package version 1.10.3.

Lozano FD, Schwartz MW. 2005. Patterns of rarity and taxonomic group size in plants. *Biological Conservation* **126**(2): 146-154.

Lucas E, Amorim B, Lima D, Lima-Lourenço A, Nic Lughadha E, Proença C, Rosa P, Rosário A, Santos L, Santos M. 2018. A new infra-generic classification of the species-rich Neotropical genus Myrcia sl. *Kew Bulletin* **73**: 1-12.

Mazine FF, Souza VC, Sobral M, Forest F, Lucas E. 2014. A preliminary phylogenetic analysis of Eugenia (Myrtaceae: Myrteae), with a focus on Neotropical species. *Kew Bulletin* **69**: 1-14.

Moonlight PW, Baldaszti L, Cardoso D, Elliott A, Särkinen T, Knapp S. 2024. Twenty years of big plant genera. *Proceedings of the Royal Society B* **291**(2023): 20240702.

Muñoz-Rodríguez P, Carruthers T, Wells T, Sumadijaya A, Wood JR, Scotland RW. 2024. The research behind a taxonomic monograph: a case study from Ipomoea (Convolvulaceae). *Kew Bulletin* **79**(4): 897-914.

Muñoz-Rodríguez P, Carruthers T, Wood JR, Williams BR, Weitemier K, Kronmiller B, Goodwin Z, Sumadijaya A, Anglin NL, Filer D. 2019. A taxonomic monograph of Ipomoea integrated across phylogenetic scales. *Nature plants* **5**(11): 1136-1144.

Muñoz‐Rodríguez P, Wood JR, Wells T, Carruthers T, Sumadijaya A, Scotland RW. 2023. The challenges of classifying big genera such as Ipomoea. *Taxon* **72**(6): 1201-1215.

Newton A, Oldfield S, Rivers M, Mark J, Schatz G, Garavito NT, Cantarello E, Golicher D, Cayuela L, Miles L. 2015. Towards a global tree assessment. *Oryx* **49**(3): 410-415.

Nic Lughadha E, Bachman SP, Leão TCC, Forest F, Halley JM, Moat J, Acedo C, Bacon KL, Brewer RFA, Gâteblé G, et al. 2020. Extinction risk and threats to plants and fungi. *PLANTS, PEOPLE, PLANET* **2**(5): 389-408.

Nic Lughadha EM, Graziele Staggemeier V, Vasconcelos TNC, Walker BE, Canteiro C, Lucas EJ. 2019. Harnessing the potential of integrated systematics for conservation of taxonomically complex, megadiverse plant groups. *Conserv Biol* **33**(3): 511-522.

Purvis A, Agapow P-M, Gittleman JL, Mace GM. 2000. Nonrandom extinction and the loss of evolutionary history. *science* **288**(5464): 328-330.

R Core Team 2024. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rivers M. 2017. The Global Tree Assessment–red listing the world's trees. *BGjournal* **14**(1): 16-19.

Schwartz MW, Simberloff. 2001. Taxon size predicts rates of rarity in vascular plants. *Ecology Letters* **4**(5): 464-469.

Scotland RW, Wood JR. 2012. Accelerating the pace of taxonomy. *Trends in Ecology & Evolution* **27**(8): 415-416.

Sigwart JD, Sutton MD, Bennett KD. 2018. How big is a genus? Towards a nomothetic systematics. *Zoological Journal of the Linnean Society* **183**(2): 237-252.

Soto Gomez M, Brown MJ, Pironon S, Bureš P, Verde Arregoitia LD, Veselý P, Elliott TL, Zedek F, Pellicer J, Forest F. 2024. Genome size is positively correlated with extinction risk in herbaceous angiosperms. *New Phytologist* **243**(6): 2470-2485.

Strand M, Panova M. 2015. Size of genera–biology or taxonomy. *Zoologica Scripta* **44**(1): 106-116.

Tanentzap AJ, Igea J, Johnston MG, Larcombe MJ. 2020. Does evolutionary history correlate with contemporary extinction risk by influencing range size dynamics? *The American Naturalist* **195**(3): 569-576.

Vamosi JC, Wilson JRU. 2008. Nonrandom extinction leads to elevated loss of angiosperm evolutionary history. *Ecology Letters* **11**(10): 1047-1053.

Vences M, Guayasamin JM, Miralles A, De La Riva I. 2013. To name or not to name: criteria to promote economy of change in supraspecific Linnean classification schemes.

Walker BE, Tucker A, Nicolson N. 2022. Harnessing large-scale herbarium image datasets through representation learning. *Frontiers in plant science* **12**: 806407.