ECOGRAPHY

Research

The Darwinian shortfall in plants: phylogenetic knowledge is driven by range size

Alexander V. Rudbeck, Miao Sun, Melanie Tietje, Rachael V. Gallagher, Rafaël Govaerts, Stephen A. Smith, Jens-Christian Svenning and Wolf L. Eiserhardt

A. V. Rudbeck (https://orcid.org/0000-0002-8028-1441), M. Sun (https://orcid.org/0000-0001-5701-0478), M. Tietje, J.-C. Svenning and W. L. Eiserhardt (https://orcid.org/0000-0002-8136-5233) ✉ (wolf.eiserhardt@bio.au.dk), Dept of Biology, Aarhus Univ., Aarhus C, Denmark. – R. V. Gallagher (https://orcid.org/0000-0002-4680-8115), Hawkesbury Inst. for the Environment, Western Sydney Univ., Penrith, NSW, Australia. – R. Govaerts and WLE, Royal Botanic Gardens, Kew, Richmond, Surrey, UK. – S. A. Smith, Dept of Ecology & Evolutionary Biology, Univ. of Michigan, Ann Arbor, MI, USA.

Ecography
2022: e06142
doi: 10.1111/ecog.06142
Subject Editor: Holger Kreft
Editor-in-Chief: Miguel Araújo
Accepted 30 March 2022

The Darwinian shortfall, i.e. the lack of knowledge of phylogenetic relationships, significantly impedes our understanding of evolutionary drivers of global patterns of biodiversity. Spatial bias in the Darwinian shortfall, where phylogenetic knowledge in some regions is more complete than others, could undermine eco- and biogeographic inferences. Yet, spatial biases in phylogenetic knowledge for major groups – such as plants – remain poorly understood. Using data for 337 023 species (99.7%) of seed plants (Spermatophyta), we produced a global map of phylogenetic knowledge based on regional data and tested several potential drivers of the observed spatial variation. Regional phylogenetic knowledge was defined as the proportion of the regional seed plant flora represented in GenBank's nucleotide database with phylogenetically relevant data. We used simultaneous autoregressive models to explain variation in phylogenetic knowledge based on three biodiversity variables (species richness, range size and endemism) and six socioeconomic variables representing funding and accessibility. We compared observed patterns and relationships to established patterns of the Wallacean shortfall (the lack of knowledge of species distributions). We found that the Darwinian shortfall is strongly and significantly related to the macroecological distribution of species' range sizes. Small-ranged species were significantly less likely to have phylogenetic data, leading to a concentration of the Darwinian shortfall in species-rich, tropical countries where range sizes are small on average. Socioeconomic factors were less important, with significant but quantitatively small effects of accessibility and funding. In conclusion, reducing the Darwinian shortfall and smoothen its spatial bias will require increased efforts to sequence the world's small-ranged (endemic) species.

Keywords: biodiversity, Darwinian shortfall, data bias, knowledge gaps, phylogenetic knowledge, plant phylogeny, socioeconomics, spatial bias, Spermatophyta

NORDIC SOCIETY OIKOS

www.ecography.org

Introduction

Evolutionary history is a major driver of spatial patterns in biodiversity, but the lack of phylogenetic data in many clades across the tree of life remains a major obstacle to large-scale biodiversity science and conservation. This is a fundamental aspect of the so-called Darwinian shortfall, the lack of knowledge about the evolution of species and their traits (Diniz-Filho et al. 2013, Hortal et al. 2015). Molecular phylogenetics as a field continues to rapidly grow, but dependable phylogenetic data is still only available for the minority of known species (Hinchliff et al. 2015). In plants, an ecologically important group that has been subject to extensive phylogenetic research, approximately two-thirds of all described species have no DNA sequence data in public repositories (RBG Kew 2016, Cornwell et al. 2019, Smith and Brown 2018). Some groups of plants are better sampled phylogenetically, and some spatial variation in phylogenetic knowledge has also been shown (Cornwell et al. 2019). However, the spatial variation in the Darwinian shortfall is still poorly explored and empirical explanations for this variation have not yet been provided.

Understanding spatial variation in the Darwinian shortfall is essential, as it may bias the results of ecogeographic studies that incorporate phylogenies. A map of phylogenetic knowledge could also guide future research towards closing significant data gaps by focusing on the regions of the greatest paucity in molecular data. As phylogenetic diversity tends to be geographically structured (Holt et al. 2013, Slik et al. 2018), spatial bias in the Darwinian shortfall may also translate into phylogenetic bias, i.e. certain lineages being under-represented in the tree of life, thus influencing research beyond biogeography. Yet, our understanding of the spatial variation in phylogenetic knowledge remains coarse. Cornwell et al. (2019) laid an important foundation by showing that the most disproportionately undersampled plant families differ among continents. They also revealed complex latitudinal patterns in phylogenetic data availability, highlighting a particular scarcity at low latitudes, albeit without testing the potential drivers of this pattern (Cornwell et al. 2019). Finer-scale patterns in phylogenetic knowledge and their causes remain unknown, making it difficult to account for the bias they may cause in downstream analyses.

The Darwinian shortfall has several components, the most fundamental of which is that most species have never been included in a phylogenetic analysis (Diniz-Filho et al. 2013, Hortal et al. 2015). Other components, such as lacking knowledge of internal branches of the tree of life, genetic variation within species, evolutionary rates and trait evolution are also important, but the proportion of a biota that is accessible to phylogenetic research is a useful first approximation of phylogenetic knowledge, the inverse of the Darwinian shortfall. Thus, the spatial distribution of the Darwinian shortfall depends on the spatial distribution of species sampling effort for DNA sequencing and phylogenetic analysis. This process is likely geographically non-random and may be governed by similar factors as the sampling of species for other types of biodiversity data. Studies on the barriers and biases of other information shortfalls, such as the Wallacean shortfall (the lack of distribution data, Hortal et al. 2015) have found a variety of socioeconomic and biological factors to influence the sampling of species (Amano and Sutherland 2013, Meyer et al. 2016a). Similar factors likely influence sampling for DNA sequencing. For instance, an economically wealthy region might provide better funding for the sequencing of the local biota, while many rare species may impede the probability for a broadly sequenced community. However, phylogenetic data differ in important aspects from distribution data, such as point occurrences. Obtaining the sequence of a single specimen is often much more time consuming and expensive than georeferencing an observed species occurrence. At the same time, a single relevant sequence can represent an entire species, while a single point occurrence tells little about the range of that species. We may therefore expect a different set of barriers to limit our phylogenetic knowledge relative to occurrence data.

Here, we provide a global assessment of the distribution of phylogenetic knowledge in seed plants (Spermatophyta, ~337 000 spp., Govaerts et al. 2021), by combining distribution data for all accepted plant species with all openly accessible phylogenetically relevant available molecular data. We define regional phylogenetic knowledge as the proportion of species in a region that have been sequenced for at least one of 128 widely used phylogenetic markers (Hinchliff and Smith 2014). This quantity is expected to vary substantially among regional floras, not least because species richness varies among regions by orders of magnitude (Kier et al. 2009). We assess two different null hypotheses (Table 1). The first (H0.a) assumes that species are chosen for phylogenetic research at random, i.e. without consideration of their geographic distribution. The second null hypothesis (H0.b) assumes that phylogenetic research effort is equal in all regions. We further test three alternative hypotheses (Table 1) related to species' range size, investment in research and accessibility of floras for DNA sampling. For comparison with the Wallacean shortfall, the geographic biases of which are relatively well documented (Meyer 2016), we also analyze variation in distribution knowledge, the extent to which species distributions are known.

Material and methods

Taxonomy

All analyses were conducted at the species level. We standardized species names to the World Checklist of Vascular Plants (WCVP hereafter; Govaerts et al. 2021), using the extensive synonymy included in the checklist. The standardization was done by converting all GenBank and BIEN entries that were listed as synonyms by the WCVP into their corresponding accepted names. We discarded any data that could not be assigned to an accepted species name.

Table 1. Potential mechanisms determining phylogenetic knowledge, the proportion of a regional flora that has been sequenced for at least one widely used phylogenetic marker, including two different null models of phylogenetically and geographically random sampling, respectively.

Hypothesis	Prediction
H0.a: Random species sampling. Species are chosen for sequencing without any consideration of their geographical distribution.	The number of sequenced species increases linearly with species richness. Spatial variation in phylogenetic knowledge is purely stochastic.
H0.b: Geographically uniform effort. Phylogenetic research effort is equal in all botanical countries and independent of species richness.	Spatial variation in the number of sequenced species is purely stochastic. Species-poor floras are phylogenetically more completely known than species-rich floras, and phylogenetic knowledge thus decreases linearly with species richness.
H1: Range size. Widely distributed species are usually easier to obtain for sampling than narrow endemics, and thus have a higher likelihood of being included in regional sequencing efforts.	Phylogenetic knowledge increases with the average range size, and decreases with the average level of endemism, of the species occurring in a regional flora.
H2: Funding. Species occurring in wealthier areas that invest heavily in education and research are more likely to be sequenced.	Phylogenetic knowledge increases with regional gross domestic product (GDP) as well as research and education expenditure.
H3: Accessibility. Species occurring in accessible, densely populated and safe areas are more likely to be sampled for sequencing.	Phylogenetic knowledge increases with population density, road density and security as measured by the Global Peace Index (GPI).

Geographic data

We used level 3 of the World Geographical Scheme for Recording Plant Distribution (formerly known as Taxonomic Databases Working Group, TDWG) as the main unit of analysis (Brummit 2001). These spatial units, in the following referred to as 'botanical countries', mostly correspond to political countries, but larger countries are split into lower-level administrative units (e.g. states of the USA) and some island units (e.g. Borneo) consist of parts of multiple countries. Data for presence or absence of plant species in the 369 botanical countries were obtained from the WCVP. These data have been recorded from published sources (primarily floras and regional checklists) following a workflow described by Govaerts et al. (2021). This workflow, which was in progress at the time of publication of Govaerts et al. (2021), is now completed, but the dataset is being continuously updated as new data sources are published. Updated versions of the database are currently available via Plants of the World Online (<www.plantsoftheworldonline.org>) and will soon also be available as part of the WCVP (<https://wcvp.science.kew.org/>). Our analyses use a download of the database from July 2021. Only presences of accepted, extant and native species were included, thus excluding infraspecific taxa, extinct or introduced species. This returned a total of 337 023 species of spermatophytes with geographic data (99.7% of the total number of accepted species). One botanical country (Bouvet Island) contained no recorded plant species and was thus excluded.

Response variables

We used phlawd_db_maker (<https://github.com/blackrim/phlawd_db_maker>) to create a SQLite database from the entire plant division of NCBI Genbank (<www.ncbi.nlm.nih.gov/genbank/>, GB Release Number 242, accessed in March 2021). To avoid confounding effects of sequences that are not actually phylogenetically useful or informative, we only considered data for 128 plastid, mitochondrial and nuclear markers identified as widely used in plant phylogenetics by Hinchliff and Smith (2014). From this database we extracted a list of species that have data for at least one marker. We considered these species to be 'phylogenetically known'. Genbank entries for subspecies were used at the species level. From this list, we calculated regional phylogenetic knowledge for each botanical country as the number of phylogenetically known species divided by the total number of species. We also used the raw number of phylogenetically known species per botanical country as an alternative response variable.

How well a biota is known phylogenetically depends on not only how many of its species have been sampled, but also how much sequence data has been produced for each species. Thus, we also calculated the number of phylogenetically relevant markers (sensu Hinchliff and Smith 2014) that had been sequenced for each species and built it into an alternative measure of phylogenetic knowledge. This alternative response variable was calculated as the number of unique species–marker combinations available for a botanical country divided by the total number of possible species–marker combinations for that country, thus representing both sampling and sequencing effort.

For comparison to the Wallacean shortfall, we also recorded 'distribution knowledge' for each botanical country based on point occurrence data from the BIEN database ver. 4.1.1 (Enquist et al. 2016), accessed in April 2020. In contrast to phylogenetic knowledge, where we disregarded the geographic provenance of the sequence data, we only counted species as 'geographically known' in a given botanical country if they had coordinate data in that country. Regional distribution knowledge was calculated as the percentage of species in each botanical country that were 'geographically known' in that country.

Explanatory variables

We included three biological explanatory variables to address the potential effect of variation in biodiversity on

phylogenetic knowledge. We also analyzed six socioeconomic variables to examine the role of funding and accessibility (Fig.1).

Biological variables included species richness, mean species range size and endemism. Plant species richness was determined as the total number of species for each botanical country. Because species richness was used as an explanatory variable, we did not area-standardize it despite the fact that species richness is clearly affected by the size of the botanical country. Mean species range was calculated as the average of the range size in km^2 of the species occurring in each botanical country, where range size was defined as the sum of the area of the countries in which each species occurs. Endemism was calculated as the proportion of species in a botanical country that were not found in any other botanical country.

Socioeconomic variables included population density, per capita gross domestic product (GDP), road density, security, research and development expenditure (research expenditure hereafter) and education expenditure.

We used gridded data for population density in 2010 (15 arcmin. resolution; Center for International Earth Science Information Network (CIESIN) 2018), subnational GDP in 2010 (20 arcsec. resolution; DECRG 2019) and road density (5 arcmin.; Meijer et al. 2018). We aggregated the data at the botanical country level using the tool 'Zonal Statistics' in

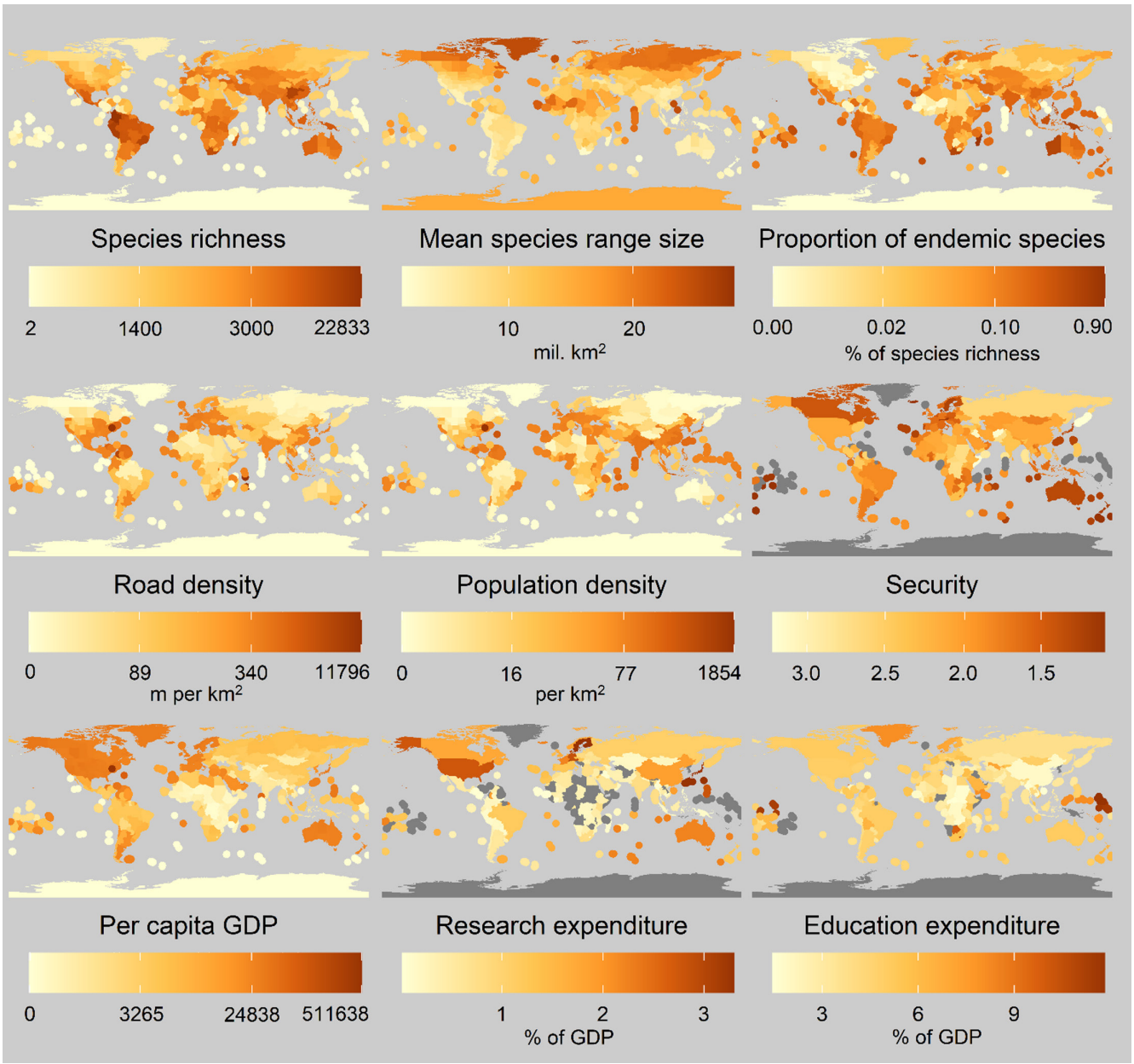

Figure 1. Maps of predictor variables.

ArcMap ver. 10.5.1 (ESRI 2011). We used mean values for population and road density. Due to the coarse resolution of this data, several very small botanical countries did not overlap with the centroids of any grid cells, precluding the calculation of mean values. Thus, we resampled these grids with a 100× higher resolution, allowing partly overlapping grid cells to enter the calculation of means. For per capita GDP, we first calculated the sum of the gridded subnational GDP values in each botanical country, and then divided this value by the botanical country's population, which was calculated as population density (above) times area.

Data regarding security, research expenditure and education expenditure, exist only at the national level. Security was measured using the Global Peace Index (GPI) (Inst. for Economics and Peace 2008–2019). Low GPI values indicate a high state of peace in a nation. Research expenditure and education expenditure data were acquired from The World Bank's open data catalog (World Bank World Development Indicators 2016, 2017). The values indicate the expenditure as a percentage of GDP for each nation. We used mean values for the years 2008–2019 for all three variables. National-level data were applied to botanical countries as follows. Where nations and botanical countries matched (n = 98), data were used directly. Where several botanical countries were nested in one nation (n = 239 botanical countries, e.g. states in the USA), the national value was applied to all nested botanical countries. The remaining cases (n = 31) where one botanical country contained several nations or parts of nations (e.g. the island of Borneo) were resolved by calculating a weighted mean value based on the proportion of land area belonging to each nation, following Gallagher et al. (2020).

Analysis

All analyses were performed in R ver. 3.5.0 (<www.r-project.org>). All variables were standardized to allow comparison between effect sizes. We used Kendall rank correlations to assess potential multicollinearity between explanatory variables. As all correlations were moderate (≤ 0.53), all variables were retained and used as predictors. We used multiple regression to separately model the effect of the predictors on each response variable. After detecting spatial autocorrelation in the residuals of preliminary ordinary least squares (OLS) regression models with a Moran's I test, we switched to simultaneous autoregressive (SAR) models (Cressie 1993). We tested several possible neighborhood structures, including queen's case, rooks case, K-nearest neighbor for 1, 2 and 3 neighbors and distance-based neighbors for great circle distances 1000, 1500, 2000 and 5000 km. Spatial correlograms using Moran's I to quantify autocorrelation showed that a distance-based neighborhood structure using 2000 km as distance removed spatial autocorrelation most effectively. A Lagrange multiplier comparing different types of SAR models indicated that a spatial error model fitted the data best. We performed model selection by building spatial error SAR models for all possible combinations of predictor variables, and selecting the best model based on the Akaike information criterion (AIC). We report Nagelkerke's pseudo R^2 (in the following referred to as R^2 for simplicity) as a measure of explained variation (Kissling and Carl 2008). We used model averaging based on AIC weights to infer model coefficients (slopes) across all candidate models (Diniz-Filho et al. 2008). To further investigate if species richness and mean species range represent independent drivers of phylogenetic knowledge, we fitted additional SAR models that included only one of those two predictors and observed the drop in R^2 compared to the full model. Moran's I tests and SAR models were performed using the R package 'spdep' ver. 0.8-1 (Bivand et al. 2013). We also tested the effect of range size on the probability of species to be 'phylogenetically known' at the species level using logistic regression. All data and code used for the analyses are available at <https://doi.org/10.5061/dryad.2547d7wrz> and <https://doi.org/10.5281/zenodo.6381989>, respectively.

Results

In March 2021, 119 405 seed plant species (35.7% of accepted species with distribution data sensu WCVP) had molecular data available in GenBank for at least one of the phylogenetically informative markers identified by Hinchliff and Smith (2014). Phylogenetic knowledge varied widely across botanical countries, ranging from 22.8% in New Guinea to 94.8% for New Brunswick in Canada and even 100% on six small sub-Antarctic islands or island groups (Table 2, Fig. 2). Botanical countries in the tropics generally possess the least complete inventories with regards to phylogenetic knowledge, while the most complete inventories are concentrated in North America (Fig. 2). Of the 368 botanical countries, 52 had > 90% inventory completeness, and these include exclusively areas with either low species richness, such as small islands or polar regions, or botanical countries of north-western North America. The six sub-Antarctic botanical countries with complete inventories all have a maximum of 24 species (Table 2).

The best fitting model to explain phylogenetic knowledge included all biological predictors (species richness, mean species range and endemism) as well as the socioeconomic predictors population density and research expenditure (Table 3). The combination of the five variables explained ~83% of the variance in the response variable ($R^2 = 0.831$). There were five alternative models with $\Delta AIC \leq 2$, which mainly differed from the best model in the inclusion of additional non-significant predictors (Supporting information). However, two of the five alternative models lacked population density as a significant predictor. Across models, mean species range had a strong positive effect on phylogenetic knowledge, while species richness and endemism had a negative effect (Table 3). Population density and research expenditure affected inventory completeness positively, albeit with a relatively shallow standardized slope compared to the biological variables (Table 3). Our alternative measure of phylogenetic knowledge (considering both sampling

Table 2. The 10 botanical countries with the highest/lowest phylogenetic knowledge and their significant predictors, defined as the proportion of a regional flora that has been sequenced for at least one widely used phylogenetic marker.

Botanical country	Phylogenetic knowledge	Species richness	Endemic proportion	Mean species range size (mil. km^2)	Population density (p km^{-2})	Resarch expenditure (% of GDP)
Most phylogenetically complete botanical countries						
Crozet Islands	100%	15	0%	5.8	0	2.2
Heard-McDonald Is.	100%	11	0%	5.9	0	2.2
Kerguelen	100%	24	4.2%	3.8	0.0	2.2
Marion-Prince Edward Is.	100%	16	6.3%	9.0	0	0.8
South Georgia	100%	15	0.0%	12.5	0	0.6
South Sandwich Islands	100%	2	0.0%	14.5	0	0.6
Prince Edward I.	96.5%	625	0.0%	19.9	22.7	1.8
Azores	95.9%	466	11.8%	18.1	104.0	1.4
New Brunswick	94.8%	996	0.0%	16.9	9.9	1.8
Manitoba	94.7%	1233	0.2%	18.4	1.8	1.8
Least phylogenetically complete botanical countries						
New Guinea	22.8%	12023	70.6%	2.7	9.2	0.1
Madagascar	34.4%	10720	85.3%	2.7	29.7	0.1
Borneo	34.8%	10738	55.7%	2.6	23.9	0.4
Philippines	35.2%	8124	55.9%	3.8	302.8	0.1
Colombia	37.1%	22833	31.4%	4.4	37.9	0.2
Fiji	37.5%	1383	55.2%	3.2	41.9	0
Peru	38.4%	19196	36.7%	4.7	20.2	0.1
New Caledonia	39.1%	16583	32.1%	4.0	54.6	0.4
Ecuador	39.5%	3008	81.3%	1.7	11.1	0
Sweden	40.5%	3580	34.7%	9.9	18.3	3.3

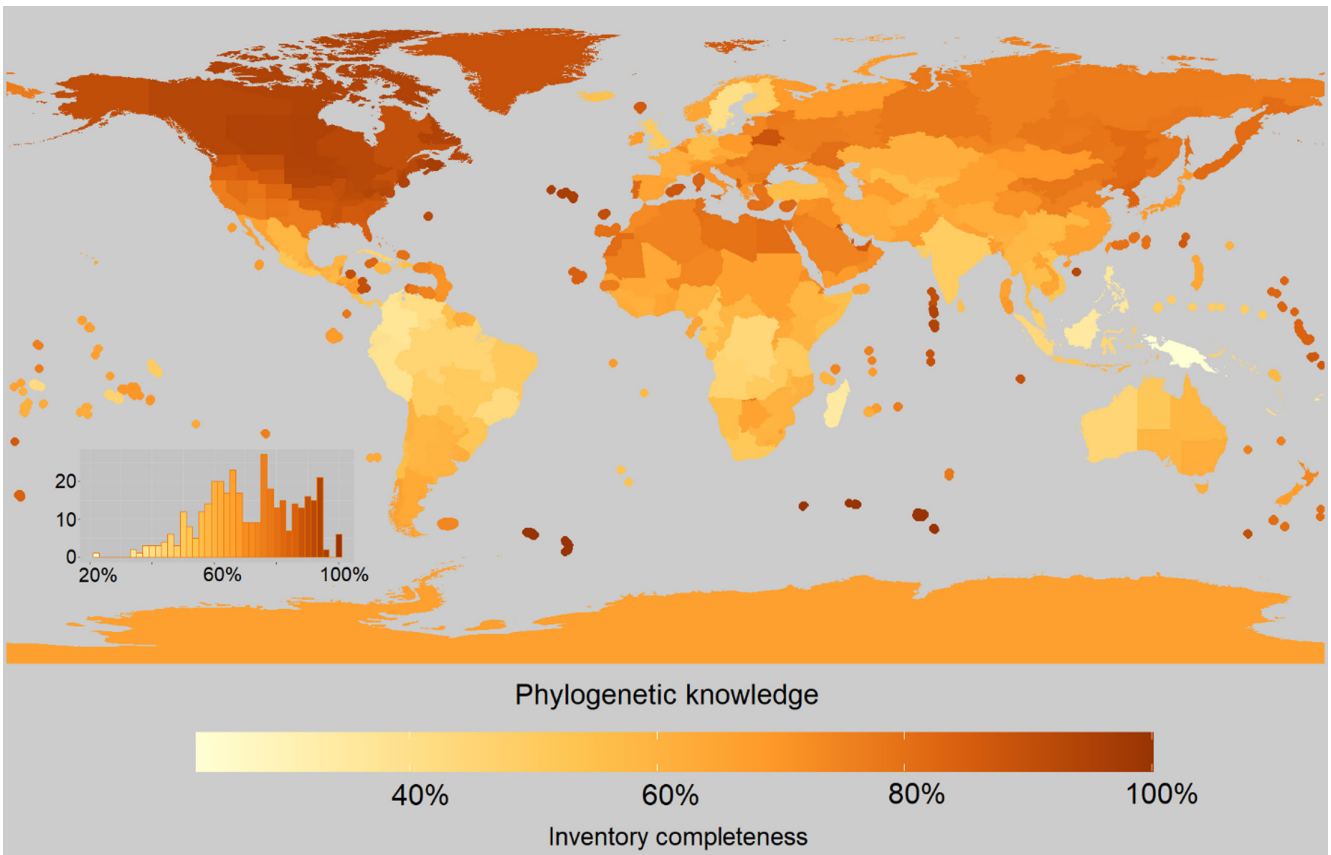

Figure 2. Map of phylogenetic knowledge, the proportion of a regional flora that has been sequenced for at least one widely used phylogenetic marker. The histogram shows the frequency distribution of phylogenetic knowledge across all spatial units.

Table 3. Standardized slopes (z) of predictors of phylogenetic knowledge as inferred from Akaike information criterion (AIC)-weighted model averaging across models with all possible combinations of predictor variables, and as inferred from the best model (as selected by AIC). NA = variable not included in model. Superscripts indicate significance of the predictor in the best model (***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$).

Predictor	Model averaged z	Best model
Species richness	−0.215	−0.217***
Mean species range size	0.385	0.384***
Proportion of endemic species	−0.211	−0.210***
Road density	0.014	NA
Population density	0.032	0.049*
Security	−0.008	NA
Per capita GDP	0.002	NA
Research expenditure	0.119	0.117***
Education expenditure	0.010	NA
Pseudo-R^2	–	0.831
AIC weight	–	0.101

and sequencing effort) was highly correlated with our main measure (Spearman's $\rho = 0.89$). Accordingly, the best model for the alternative measure included the same significant predictor variables with qualitatively similar effects (Supporting information). Of note, the effect of mean species range was quantitatively even stronger than for our main measure of phylogenetic knowledge.

The significant decline of phylogenetic knowledge with species richness results from a non-linear increase of the number of species sequenced with species richness (Fig. 3; Supporting information). For relatively species-poor assemblages, the number of species sequenced increases rapidly with species-richness, resulting in high values of phylogenetic knowledge. As species richness increases, the curve levels off, leading to much lower values of phylogenetic knowledge for more species-rich floras. This pattern is well-described by a quadratic effect of species richness on the number of phylogenetically known species, which corresponds to a linear effect of species richness on the proportion of phylogenetically known species (Fig. 3; Supporting information).

Alternative models that included only species richness or mean species range, but not both predictors, only had a marginally smaller R^2 than the full model ($R^2 = 0.77$ and 0.81, respectively, compared to $R^2 = 0.83$ for the full model). Logistic regression of species' sequencing status (1 = phylogenetically known, 0 = phylogenetically unknown) on their range size was highly significant ($p < 0.001$; Supporting information).

For distribution knowledge, the model with the best fit included two of three biological factors, species richness and mean species range, notably excluding endemism (Supporting information). It also included education expenditure and research expenditure. This model explained less variance ($R^2 = 0.52$) compared to the equivalent model for phylogenetic knowledge. The strongest predictors of distribution knowledge were species richness (positive), research expenditure (positive) and mean species range (negative), while education expenditure had a weaker positive effect.

Discussion

Our results show that the significant Darwinian shortfall in seed plants is by no means randomly distributed across the globe. In agreement with previous studies, we found that, globally, roughly one third of known seed plant species have phylogenetically informative sequence data in public repositories (Folk et al. 2018, Smith and Brown 2018, Cornwell et al. 2019). This knowledge is geographically biased. For example, many regions, especially at high latitudes, are approaching inventory completeness, while a considerable portion of the earth harbors floras that are still poorly known phylogenetically. Relevant sequences only exist for one in five seed plant species of New Guinea (Table 1), and most biodiversity hotspots do not exceed 50% completeness (Myers et al. 2000; Fig. 2). Meanwhile, the least known botanical country of the United States and Canada, California, has more than 76% completeness, though most botanical countries of this region

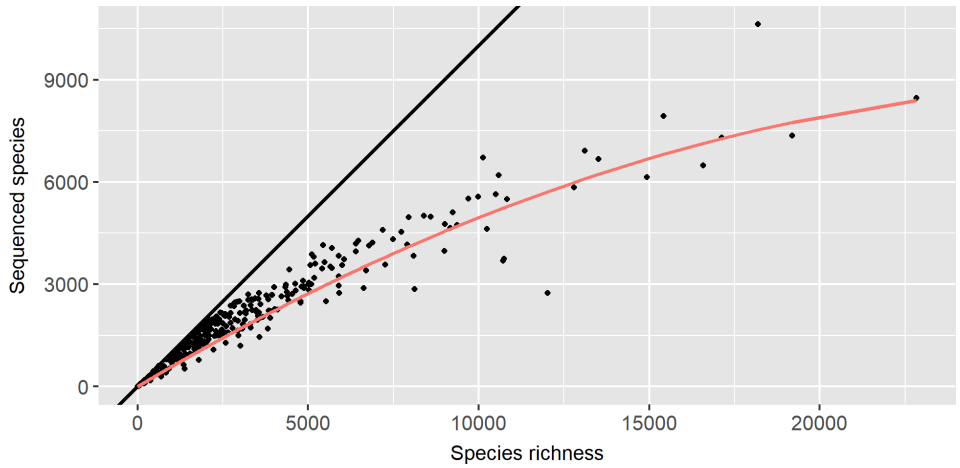

Figure 3. The number of species sequenced for at least one widely used phylogenetic marker as a function of species richness across regional floras. The red line is derived from the linear fit of phylogenetic knowledge (the proportion of a regional flora that has been sequenced for at least one widely used phylogenetic marker) as a function of species richness (Table 2).

have > 85% completeness. This geographic bias, which poses a major challenge for biodiversity research, seems to be chiefly driven by variation in species range size and its effect on the likelihood of species to be included in phylogenetic studies.

We rejected both our null hypotheses for the geographic distribution of the Darwinian shortfall. Random species sampling (H0.a) assumes that researchers target species without consideration of their geographic distribution. This scenario would be more plausible if access to DNA material was unrestricted, and institutions conducting phylogenetic research were distributed evenly over the globe and/or researchers had no preference for studying species closer to their home institutions. The consequence of such a scenario would be a linear dependence of the number of species sequenced on the species richness of any given flora, which is not what we found (Fig. 3). Instead, the gap between the number of species sequenced, and the total number of species in any given flora widens as species richness increases. In other words, the world's most species-rich floras are also the least phylogenetically known. This finding is in line with the latitudinal patterns of genetic sampling completeness shown by Cornwell et al. (2019), who also found the biodiverse tropics to be least well represented in Genbank, based on different distributional data. Our second null hypothesis, geographically uniform effort (H0.b) assumed that comparable numbers of species were sequenced in all botanical countries, with any variation among botanical countries being purely stochastic. This scenario would be plausible if sequencing efforts were mostly localised, and all botanical countries had roughly the same means for DNA sampling and molecular sequencing. This is also not supported by our analysis as there is a significant increase of sequenced species with species richness (Fig. 3, Table 2; Supporting information). What, then, drives the large observed differences in phylogenetic knowledge across the globe?

We found mean species range to be the quantitatively strongest driver of phylogenetic knowledge (Table 3), supporting our hypothesis that the probability of a species to be sampled for phylogenetic research should increase with its geographic range (H1, Table 1). Species that occur in more countries are more likely to be sampled and sequenced. Thus, floras containing many widespread species tend to be phylogenetically better known. We argue that this effect also explains the correlation between species richness and phylogenetic knowledge. In our dataset, average range size was only weakly correlated with species richness (Kendall's $\tau = -0.36$). This is consistent with previous findings showing that plant species' ranges are not always small at low latitudes (Morueta-Holme et al. 2013, Sheth et al. 2020). Thus, both variables were included in the final model. However, reduced models that included either species richness or mean species range, but not both, explained almost the same amount of variance as the full model, suggesting that the effects of those two variables on phylogenetic knowledge were largely shared. As there is no good reason to expect that researchers would shy away from species rich floras, we argue that range size is the more likely driver of this shared effect. Thus, we conclude that the low phylogenetic inventory completeness of the world's most species-rich floras is due to their high proportion of small-ranged, endemic species (Stevens 1989). This conclusion is supported by the effect of endemism, an alternative measure of the proportion of small-ranged species in each flora: the more endemic species a botanical country contains, the less well-known is its flora phylogenetically (Table 3).

Studies on limitations and barriers driving biases in data availability in the fields of biogeography and macroevolutionary biology often find a variety of socioeconomic factors important (Soberón and Peterson 2004, Yesson et al. 2007, Riddle et al. 2011, Meyer 2016). For phylogenetic data availability, however, it seems that socioeconomic factors only play a minor role. Both population density (a proxy for accessibility) and research expenditure (a proxy for investment in phylogenetic research) have a significant positive effect on phylogenetic knowledge (Table 3), supporting our hypotheses H2 and H3 (Table 1). These effects, however, are quantitatively minor compared to the effects of species range and endemism. This is unlikely to be due to the quality of our predictor variables, as we found a strong effect of research expenditure on distribution knowledge (Supporting information), as expected based on previous studies (Ahrends et al. 2011, Meyer et al. 2015, 2016a). It is also unlikely to be due to the scale of analysis, as most socioeconomic variables (GDP, security, education and research expenditure) actually are most meaningfully defined at a national level, which closely corresponds to our spatial resolution in most parts of the world. The relative weakness of the funding effect (H3) is more likely related to the spatial dynamics of phylogenetic research. As with taxonomy (Gaston and May 1992, Rodrigues et al. 2010), phylogenetic research efforts often are exported, and not necessarily focused on the home range of the researcher or institution performing the sequencing. Hence, high per capita GDP or research expenditure in a country may not translate to funding towards sequencing the flora of that same country. Instead, sequencing would be more affected by the socioeconomic circumstances of the home of the researcher. This, however, does not apply to measures of accessibility (H2) including road density, population density and security. We find a weak effect of population density, indicating that the floras of more densely populated botanical countries are phylogenetically better known. This is likely due to specimens being more easily available in those countries, either because fieldwork is easier, or because specimens have already been collected and are available from herbaria. Similar effects have been documented for distributional data in the past (Parnell et al. 2003, Ficetola et al. 2014).

Sampling species for phylogenetic research is only the first step on the route to increasing phylogenetic knowledge, which also depends on the amount of data generated for each species (Rokas and Carroll 2005). Thus, we explored if and how our results would change if the number of markers sequenced for each species is also included. We found that the effect of sequencing effort on the geographic patterns is small, suggesting that geographic biases operate mainly at the level of species sampling, not sequencing effort. However,

including sequencing effort amplified the effect of range size, suggesting that widespread species also tend to be sequenced for more different phylogenetic markers. It appears plausible that widely distributed species are more likely to be sampled by many studies employing different phylogenetic markers, emphasizing the importance of the sampling process. We also note that the effect of research expenditure was slightly stronger when taking sequencing effort into account, which is also plausible as researchers in more affluent countries may be able to include more phylogenetic markers in any given study. It is important to note that in practice, constructing usable phylogenies often rely on some commonality of phylogenetic markers in the flora. Although having multiple markers for each species increases the probability of having shared markers, a challenge still exists. The phylogenetic distribution of species with available markers also has the potential to disturb phylogenetic inference. If missing species are non-randomly dispersed in the phylogenetic tree, low phylogenetic knowledge is an even more problematic, as entire branches may be missing. These aspects would also have to be explored to produce a more accurate measure of the actual proportion of regional phylogenetic knowledge.

When exploring the map of phylogenetic knowledge (Fig. 2), a few botanical countries or regions show aberrant values compared to their surrounding countries. While these outliers are unlikely to influence our overall conclusion, they illustrate the sensitivity of this kind of analysis to taxonomy. The most prominent example of this is Sweden. Phylogenetically relevant data is only available for 40.5% of the Swedish flora, placing it nestled amid tropical regions as opposed to regional counterparts like Norway (65%) and Denmark (66%) and the Baltics (74%). Species composition in Sweden, shows a considerably higher number of endemic species than in neighboring countries. However, among the 1244 endemic species 68% belong to a single genus, *Hieracium* L. This is likely due to the intense effort to revise the taxonomy of this genus with apomictic tendencies in Sweden (Tyler 2007, 2017). The result of this local effort is a massive spike in species richness in a single genus for which not much genetic data is available (only 88 of 3026 species of *Hieracium* have phylogenetically relevant data available in GenBank). This case highlights the consequences of the Linnean shortfall (Lomolino et al. 2017) and variation in taxonomic opinion (Faurby et al. 2016) which must be taken into account for any comprehensive study on biodiversity. Like the Darwinian shortfall, the Linnean shortfall is thought to be highest in the tropics (Freeman and Pennell 2021). Thus, we are probably overestimating the proportion of sequenced species, and thus underestimating the Darwinian shortfall, in tropical areas. If so, our results would likely be even more pronounced if the Linnean shortfall was properly accounted for.

Although caveats must be considered when using the geographic and socioeconomic data used here to analyze the Darwinian shortfall, we argue that our results are robust to these limitations. For instance, the coarse resolution of our geographic units (botanical countries) inevitably leads to a systematic overestimation of species' range size, and the variable and biologically arbitrary size of the botanical countries introduces noise. However, the benefits of consistently recorded presences and absences outweigh these disadvantages. The alternative, point occurrence records, is heavily affected by the Wallacean shortfall which may vary with range size, thus potentially introducing bias (Meyer et al. 2016b). Improved estimates of species' range sizes would likely strengthen the relationship with phylogenetic knowledge documented here. As they mostly correspond to political nations, botanical countries are ideal for assessing the effect of socioeconomic factors such as GDP and political stability, which act and are recorded at the national level. Other socioeconomic factors, such as accessibility (road and population density), likely act at smaller scales, and we can thus not fully rule out that we have missed an existing effect of these factors on phylogenetic knowledge. Finally, phylogenetic progress may be influenced by socioeconomic properties that we were unable to record, such as conditions for obtaining permits for collection and export, or the policies of regional herbaria for allowing destructive sampling of their collections. Such factors may well explain additional variation in phylogenetic knowledge, although we note that the residual variation of our model is moderate (20%).

Identifying gaps in biodiversity data is an essential first step towards mending them. In this study, we showed that the Darwinian shortfall in plants varies substantially and systematically over the surface of the earth, with most species in temperate floras being phylogenetically known, whereas most species in tropical floras are not. Our results suggest that this pattern is chiefly driven by species' range size. It appears that wide-ranged plant species are 'low-hanging fruit' for phylogenetic research that have largely been sequenced already. Thus, a deliberate push to sequence the world's endemic plant species may be needed to remove the observed strong spatial bias in phylogenetic knowledge. Importantly, the least phylogenetically well-known regions are tropical and subtropical biodiversity hotspots with high concentration of threatened species, particularly due to habitat destruction (Myers et al. 2000, Baillie et al. 2004, Vamosi and Vamosi 2008). The paucity of phylogenetic data in these areas is a serious concern, as conservation efforts may benefit significantly from phylogenetic data (Lu et al. 2018, Velazco et al. 2020). Conversely, local and global extinction of species in those areas will make completing the plant Tree of Life increasingly difficult. Generating phylogenetic data for small-ranged, tropical plant species is thus not just important, but also urgent.

In the meantime, researchers incorporating phylogenetic data in their analyses must be aware of the existing biases that may significantly distort analyses comparing areas with high and low phylogenetic knowledge. For instance, a comparison of diversification rate or dispersal events between North- and South America using a phylogeny with all our currently sequenced species, may yield considerably misleading results. The phylogeny would contain ~80% of all North American plants, while only ~50% of all South American plants. This bias could result in a severe underestimation of diversification or dispersal in South America.

The paucity of phylogenetic knowledge in our most biodiverse regions seems to be a testament to the challenge that high biodiversity presents in many fields of life sciences. Both the Wallacean and Linnean shortfalls are most acute in biodiversity hotspots (Chapman 2006, Bush and Lovejoy 2007, Boakes et al. 2010, May 2010, Hortal et al. 2015). This study demonstrates that the Darwinian shortfall not only follows this pattern, but very strongly augments the role of biodiversity in the understanding of our ignorance.

Acknowledgements – We thank Bill Baker and Brian Maitner for valuable discussions. Brian Maitner also provided access to the BIEN database. We are grateful to Joaquín Hortal and Jose Alexandre Diniz-Filho for helpful comments on earlier versions of our manuscript. We acknowledge the herbaria that contributed data to the BIEN database: HA, CBS, FCO, MFU, UNEX, VDB, ASDM, AMD, BPI, BRI, DLF, CNPO, BRM, CLF, L, LPB, AD, TAES, AMO, CVRD, FURB, HPL, IAC, IB, WOLL, INPA, IPA, MW, MBML, KUN, MNHN, UESC, UFMA, UFRJ, UFRN, UFS, ULS, US, RAS, USP, RB, TRH, CGMS, ZMT, FEN, BRIT, MO, NCU, NY, TEX, FHO, U, UNCC, A, AAH, ACOR, ADW, AJOU, UI, AK, ALCB, AKPM, EA, AAU, ALTA, ALU, AMES, AMNH, ANA, GH, ANGU, ANSP, ARAN, ARM, AS, ASU, BAI, AUT, B, BA, BAA, BAB, BACP, BAF, BAJ, BAL, BARC, BAS, BBS, BC, BCF, BCN, BCRU, BEREA, BG, BH, BIO, BISH, SEV, BLA, DBN, BM, BOCH, BOG, MJG, BOL, BOLV, BONN, BOUM, BR, BREM, BRLU, BSB, BSIP, BTN, BUL, BUT, C, CALI, CAMU, CAN, CANB, CAS, CAY, CBG, CBM, CEN, CEPEC, CESJ, CHAP, CHI, CHL, CHR, CHRB, CICY, CIIDIR, CIMI, CINC, CLEMS, COA, COAH, COFC, CP, COL, COLO, CONC, CORD, CPAP, CPUN, CR, CRAI, CU, CRP, CS, CSU, CTES, CTESN, CUZ, DAO, HB, DAV, DNA, DR, DS, DUKE, DUSS, E, HUA, EAC, ECU, EIF, EIU, EKY, EMMA, ENCB, ERA, ESA, F, FAA, FAU, FB, UVIC, FI, FLAS, FLOR, FR, FTG, CMM, FUEL, G, GB, GDA, GENT, GEO, HUJ, CGE, GES, GI, GLM, GMNHJ, K, GOET, GUA, HAW, GZU, H, HAL, HAM, HAMAB, HAO, HAS, HAST, HASU, HBG, HBR, HEID, HGI, HIP, HNHM, HNT, HO, HRCB, HRP, HSS, HU, HUAL, HUEFS, HUEM, HUFU, HUSA, HUT, IAA, HXBH, HYO, IAN, IBGE, IBUG, ICEL, ICN, IEB, ILL, SF, ILLS, HAC, IFO, IPRN, FCQ, ABH, BAFC, BBB, BO, NAS, INB, INEGI, INM, EAN, ISKW, ISC, ISL, GAT, IBSC, UCSB, ISTC, ISU, IZAC, JACA, JBAG, JE, SD, JUA, JYV, KIEL, ECON, TOYA, MPN, USF, TALL, RELC, CATA, AQP, KMN, KMNH, KOELN, KOR, FRU, KPM, KSTC, LAGU, TRTE, KSU, GRA, IBK, KTU, ACAD, MISSA, KU, PSU, KYO, LA, LW, SUU, UNITEC, TASH, NAC, UBC, IEA, LAE, LAF, GMDRC, LAM, LCR, LD, LE, LEB, LI, LIL, LINN, AV, HUCP, QFA, LISE, NM, MT, FAUC, CNH, MACF, CATIE, LTB, LISI, LISU, LL, LOJA, LP, LPAG, MGC, LPD, LPS, IRVC, JOTR, LSU, LBG, BRY, LTR, CDBI, LYJB, LISC, DBG, AWH, NH, HSC, LMS, MELU, NZFRI, M, MA, UU, UBT, CSUSB, MAF, MAK, MAN, MB, MARY, MASS, MBK, MBM, UCSC, UCS, JBGP, DSM, OBI, BESA, LSUM, FULD, MCNS, ICESI, MEL, MEN, TUB, MERL, MEXU, MFA, FSU, MG, MICH, HIB, MIL, DPU, TRT, BABY, HITBC, ETH, YAMA, SCFS, SACT, ER, JCT, JROH, SBBG, SAV, PDD, MIN, SJSU, MISS, MMMN, PAMP, MNHM, SDSU, BOTU, MSE, MOR, USCH, MPU, MPUC, MSB, MSC, CANU, SFV, CNS, JEPS, BORH, WIN, BKF, MSUN, CIB, MUR, MTMG, VIT, UPNG, MU, MUB, MUCV, MVFA, MVFQ, PGM, MVJB, MVM, MY, PASA, N, UCMM, HGM, TAM, BOON, MHA, MARS, COI, NA, NCSC, ND, NU, NE, NHM, NHMC, NHT, NLH, NMB, NMC, NMNL, NMR, NMSU, NOU, NSPM, NSW, NT, NUM, NWOSU, O, OCLA, CHSC, LINC, CHAS, GA, ODU, OKL, OKLA, CDA, OS, OSA, OSC, OSH, OULU, OWU, OXF, P, PACA, PAR, UPS, PE, PEL, SGO, PEUFR, PFC, PH, PKDC, SI, PLAT, PMA, PNH, POM, PORT, PR, QM, PRC, TRA, PRE, PVNH, PY, QMEX, QCA, TROM, QCNE, QRS, UH, QUE, R, SAM, RBR, REG, RFA, RIOC, RM, RNG, RSA, RYU, S, SALA, SANT, SAPS, SASK, SBT, SEL, SIM, SING, SIU, SJRP, SMDB, SMF, SNM, SOM, SP, SRFA, SPF, SPSF, SQF, STL, STU, SUVA, SVG, SZU, TAI, TAIF, TAMU, TAN, TEF, TENN, TEPB, TFC, TI, TKPM, TNS, TO, TU, TULS, UADY, UAM, UAS, UB, UBA, UC, UCR, UEC, UFG, UFMT, UFP, UGDA, UJAT, ULM, UME, UMO, UNA, UNB, UNM, UNR, BRIU, UNSL, UPCB, UPEI, UPNA, USAS, USJ, USM, USNC, USON, USZ, UT, UTC, UTEP, UV, UWO, V, VAL, VALD, VEN, VMSL, VT, W, WAG, WAT, WII, WELT, WFU, WIS, WMNH, WOH, WS, WTU, WU, XAL, Z, ZSS, ZT, CUVC, LZ, AAS, AFS, BHCB, CHAM, FM, PERTH, SAN.
Funding – W.L.E. considers this work a contribution to his VILLUM Young Investigator project 'Explaining the hyperdiversity of Tropical rainforests using the Tree of Life (TropiToL)' funded by VILLUM FONDEN (grant no. 00025354). J.-C.S. considers this work a contribution to his VILLUM Investigator project 'Biodiversity dynamics in a changing world' funded by VILLUM FONDEN (grant no. 16549).

Author contributions

Alexander V. Rudbeck: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (equal). **Miao Sun**: Data curation (supporting); Investigation (supporting); Supervision (supporting); Writing – review and editing (equal). **Melanie Tietje**: Data curation (supporting); Investigation (supporting); Supervision (supporting); Writing – review and editing (equal). **Rachael V. Gallagher**: Methodology (supporting); Writing – review and editing (equal). **Rafaël Govaerts**: Resources (equal); Writing – review and editing (equal). **Stephen A. Smith**: Data curation (supporting); Software (equal); Writing – review and editing (equal). **Jens-Christian Svenning**: Conceptualization (equal); Supervision (equal); Writing – review and editing (equal). **Wolf L. Eiserhardt**: Conceptualization (lead); Investigation (supporting); Supervision (lead); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06142>.

Data availability statement

Data is available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.2547d7wrz> (Rudbeck et al. 2022).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Ahrends, A. et al. 2011. Funding begets biodiversity. – Divers. Distrib. 17: 191–200.

Amano, T. and Sutherland, W. J. 2013. Four barriers to the global understanding of biodiversity conservation: wealth, language, geographical location and security. – Proc. R. Soc. B 280: 20122649.

Baillie, J. E. M. et al. 2004. 2004 IUCN Red List of threatened species: a global species assessment. – IUCN.

Bivand, R. S. et al. 2013. Applied spatial data analysis with R, 2nd edn. – Springer.

Boakes, E. H. et al. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. – PLoS Biol. 8: e1000385.

Brummit, R. 2001. World geographical scheme for recording plant distributions, 2nd edn. – Biodiversity Information Standards (TDWG).

Bush, M. B. and Lovejoy, T. E. 2007. Amazonian conservation: pushing the limits of biogeographical knowledge. – J. Biogeogr. 34: 1291–1293.

Center for International Earth Science Information Network (CIESIN) 2018. Documentation for the gridded population of the world, ver. 4 (GPWv4), revision 11 data sets. – NASA Socioeconomic Data and Applications Center (SEDAC) 4, pp. 1–53.

Chapman, A. D. 2006. Numbers of living species in Australia and the world, 1st edn. – Australian Biological Resources Study.

Cornwell, W. K. et al. 2019. What we (don't) know about global plant diversity. – Ecography 42: 1819–1831.

Cressie, N. A. C. 1993. Statistics for spatial data, revised edn. – Wiley.

DECRG 2019. GIS processing World Bank DECRG. – Extrapolation UNEP/GRID-Geneva.

Diniz-Filho, A. F. et al. 2008. Model selection and information theory in geographical ecology. – Global Ecol. Biogeogr. 17: 479–488.

Diniz-Filho, A. F. et al. 2013. Darwinian shortfalls in biodiversity conservation. – Trends Ecol. Evol. 28: 689–695.

Enquist, B. J. et al. 2016. Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. – PeerJ 4: e2615v2.

ESRI 2011. ArcGIS Desktop: release 10. – Environmental Systems Research Inst.

Faurby, S. et al. 2016. Strong effects of variation in taxonomic opinion on diversification analyses. – Methods Ecol. Evol. 7: 4–13.

Ficetola, G. F. et al. 2014. Sampling bias inverts ecogeographical relationships in island reptiles. – Global Ecol. Biogeogr. 23: 13031313.

Folk, R. A. et al. 2018. Challenges of comprehensive taxon sampling in comparative biology: wrestling with rosids. – Am. J. Bot. 105: 433–445.

Freeman, B. G. and Pennell, M. W. 2021. The latitudinal taxonomy gradient. – Trends Ecol. Evol. 36: 778–786.

Gallagher, R. V. et al. 2020. Global shortfalls in extinction risk assessments for endemic flora. – Preprint: <www.biorxiv.org/content/10.1101/2020.03.12.984559v1>.

Gaston, K. J. and May, R. M. 1992. Taxonomy of taxonomists. – Nature 356: 281–282.

Govaerts, R. et al. 2021. The world checklist of vascular plants, a continuously updated resource for exploring global plant diversity. – Sci. Data 8: 1–10.

Hinchliff, C. E. and Smith, S. A. 2014. Some limitations of public sequence data for phylogenetic inference (in plants). – PLoS One 9: e98986.

Hinchliff, C. E. et al. 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. – Proc. Natl Acad. Sci. USA 112: 12764–12769.

Holt, B. G. et al. 2013. An update of Wallace's zoogeographic regions of the world. – Science 339: 74–78.

Hortal, J. et al. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. – Annu. Rev. Ecol. Syst. 46: 523–549.

Institute for Economics and Peace. 2019. Global Peace Index 2019: Measuring Peace in a Complex World, Sydney.

Kier, G. et al. 2009. A global assessment of endemism and species richness across island and mainland regions. – Proc. Natl Acad. Sci. USA 106: 9322–9327.

Kissling, W. D. and Carl, G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. – Global Ecol. Biogeogr. 17: 59–71.

Lomolino, M. V. et al. 2017. Biogeography, 5th edn. – Oxford Univ. Press, pp. 730.

Lu, L. et al. 2018. Evolutionary history of the angiosperm flora of China. – Nature 554: 234–238.

May, R. M. 2010. Tropical arthropod species, more or less? – Science 329: 41–42.

Meijer, J. R. et al. 2018. Global patterns of current and future road infrastructure. – Environ. Res. Lett. 13: 064006.

Meyer, C. 2016. Limitations in global information on species occurences. – Front. Biogeogr. 8: e28195.

Meyer, C. et al. 2015. Global priorities for an effective information basis of biodiversity distributions. – Nat. Commun. 6: 8221.

Meyer, C. et al. 2016a. Range geometry and socio-economics dominate species-level biases in occurrence information. – Global Ecol. Biogeogr. 25: 1181–1193.

Meyer, C. et al. 2016b. Multidimensional biases, gaps and uncertainties in global plant occurrence information. – Ecol. Lett. 19: 992–1006.

Morueta-Holme, N. et al. 2013. Habitat area and climate stability determine geographical variation in plant species range sizes. – Ecol. Lett. 16: 1446–1454.

Myers, N. et al. 2000. Biodiversity hotspots for conservation priorities. – Nature 403: 853–858.

Parnell, J. A. N. et al. 2003. Plant collecting spread and densities: their potential impact on biogeographical studies in Thailand. – J. Biogeogr. 30: 193–209.

RBG Kew 2016. The State of the World's Plants Report - 2016 – Royal Botanic Gardens, Kew.

Riddle, B. R. et al. 2011. Basic biogeography: estimating biodiversity and mapping nature. – In: Ladle, R. J. and Whittaker, R. J. (eds), Conservation biogeography. Wiley & Sons, pp. 47–92.

Rodrigues, A. S. L. et al. 2010. A global assessment of amphibian taxonomic effort and expertise. – BioScience 60: 798–806.

Rokas, A. and Carroll, S. B. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. – Mol. Biol. Evol. 22: 1337–1344.

Rudbeck, A. V. et al. 2022. Data from: The Darwinian shortfall in plants: phylogenetic knowledge is driven by range size. – Dryad Digital Repository, <https://doi.org/10.5061/dryad.2547d7wrz>.
Sheth, S. N. et al. 2020. Determinants of geographic range size in plants. – New Phytol. 226: 650–665.
Slik, J. W. F. et al. 2018. Phylogenetic classification of the world's tropical forests. – Proc. Natl Acad. Sci. USA 115: 1837–1842.
Smith, S. A. and Brown, J. W. 2018. Constructing a broadly inclusive seed plant phylogeny. – Am. J. Bot. 105: 1–13.
Soberón, J. and Peterson, T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. – Phil. Trans. R. Soc. B 359: 689–698.
Stevens, G. C. 1989. The latitudinal gradient in geographical range: how so many species coexist in the tropics. – Am. Nat. 133: 240–256.
Tyler, T. 2007. The Hawkweeds (*Hieracium* L. s. str., Asteraceae, Tracheophyta) of Sweden. – Svenska Artprojektet.
Tyler, T. 2017. The last step towards a full revision of *Hieracium* sect. *Vulgata* in Sweden. – Nord. J. Bot. 35: 305–321.
Vamosi, J. C. and Vamosi, S. M. 2008. Extinction Risk escalates in the tropics. – PLoS One 3: 8–13.
Velazco, S. J. E. et al. 2020. On opportunities and threats to conserve the phylogenetic diversity of Neotropical palms. – Divers. Distrib. 27: 512–523.
World Bank World Development Indicators 2016. Government expenditure on education, total (% of government expenditure) [data file]. – <https://data.worldbank.org/indicator/SE.XPD.TOTL.GB.ZS>.
World Bank World Development Indicators 2017. Research and development expenditure (% of GDP) [data file]. – <https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS>.
Yesson, C. et al. 2007. How global is the global biodiversity information facility? – PLoS One 2: e1124.