

ARTICLE



<https://doi.org/10.1038/s41467-021-27186-8>

OPEN

Widespread homogenization of plant communities in the Anthropocene

Barnabas H. Daru^{1,2}✉, T. Jonathan Davies³✉, Charles G. Willis⁴, Emily K. Meineke⁵, Argo Ronk⁶, Martin Zobel⁷, Meelis Pärtel⁷, Alexandre Antonelli^{7,8,9,10} & Charles C. Davis^{1,2}✉

Native biodiversity decline and non-native species spread are major features of the Anthropocene. Both processes can drive biotic homogenization by reducing trait and phylogenetic differences in species assemblages between regions, thus diminishing the regional distinctiveness of biotas and likely have negative impacts on key ecosystem functions. However, a global assessment of this phenomenon is lacking. Here, using a dataset of >200,000 plant species, we demonstrate widespread and temporal decreases in species and phylogenetic turnover across grain sizes and spatial extents. The extent of homogenization within major biomes is pronounced and is overwhelmingly explained by non-native species naturalizations. Asia and North America are major sources of non-native species; however, the species they export tend to be phylogenetically close to recipient floras. Australia, the Pacific and Europe, in contrast, contribute fewer species to the global pool of non-natives, but represent a disproportionate amount of phylogenetic diversity. The timeline of most naturalisations coincides with widespread human migration within the last ~500 years, and demonstrates the profound influence humans exert on regional biotas beyond changes in species richness.

¹ Department of Life Sciences, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412, USA. ² Department of Organismic and Evolutionary Biology, Harvard University Herbaria, Cambridge, MA 02138, USA. ³ Departments of Botany, and Forest & Conservation Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. ⁴ Department of Biology Teaching and Learning, University of Minnesota, Minneapolis, MN 55455, USA. ⁵ Department of Entomology and Nematology, University of California, Davis, CA 95616, USA. ⁶ Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA. ⁷ Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, EE-51005 Tartu, Estonia. ⁸ University of Gothenburg and Gothenburg Global Biodiversity Centre, Department of Biological and Environmental Sciences, Carl Skottsbergs gata 22B, SE 405 30 Gothenburg, Sweden. ⁹ Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK. ¹⁰ Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, UK.
✉email: barnabas.daru@tamucc.edu; j.davies@ubc.ca; cdavis@oeb.harvard.edu

Habitat conversion, biotic invasions, anthropogenic climate change, and pollution have contributed initially to dark diversity¹ and eventually to global species losses^{2–5}. There has been much focus on species extinctions^{6,7}; however, how these biodiversity changes manifest at local to regional scales is still unclear: some studies show declining local diversity⁸, while others suggest stable or even increasing species diversity through time⁹, and changes in the turnover of species diversity (β -diversity) have been less well studied. Biotic homogenization—declining β -diversity—reduces trait and phylogenetic differences between regions, and is perhaps a more characteristic feature of the Anthropocene than species loss¹⁰. The Anthropocene epoch encapsulates the profound effects of human activities on the land surface, oceans, atmosphere, and the evolution of life on Earth, with the implications that these changes have no geological analog¹⁰. Biotic homogenization is primarily driven by native species' extirpation and the introduction and spread of non-native species, commonly due to human activity^{11–13}. The balance and influence of these contributing factors remains largely untested across different scales.

For thousands of years, plants have been moved unintentionally or, more commonly, intentionally by humans from their native ranges as sources of food, ornament, medicine, fuel, and shelter^{14,15}. Plant invasions greatly accelerated ~500 years ago when the Eastern and Western Hemispheres were united by the Columbian Exchange¹⁶. One consequence of this widespread movement of species has been the increasing homogenization of plant communities across biomes (e.g.,^{17–20}). The magnitude and impact of these compositional changes on the evolutionary structure of native floras across the globe has received surprisingly little attention (see^{21,22} for regional assessments). Nonetheless, the importance of evolutionary history in determining the establishment and spread of non-native species has long been recognized. Charles Darwin proposed that introduced species were less likely to establish in communities if they were closely related to the native species—Darwin's naturalization hypothesis²³. If true, then non-native species would tend to add significantly to the phylogenetic diversity of a region, and reduce phylogenetic turnover between regions. However, if establishment success of non-natives reflects phylogenetically conserved environmental niche preferences (c.f. Darwin's naturalization conundrum)²⁴, then their addition to the native pool would likely add little to regional phylogenetic diversity, and changes in phylogenetic turnover between regions would be slight²⁵.

Species extirpations, especially those facilitated by human activities, may also contribute to increasing biotic homogenization of plant communities. While data on past plant extinctions remain sparse (but see²⁶), we can extrapolate future extinctions using current Red List assessments²⁷. In comparative cross-species analyses, the best predictor of species' extinction risk is geographic range: narrow-ranged endemics, in particular, have the highest risk of extinction²⁸. Thus, concomitant with the increasing spread of non-native species, there has been a decline in range-restricted species, which might also contribute to lower rates of species turnover across landscapes. Species extinctions will always result in a loss of evolutionary history, but if extinctions are random across the tree-of-life, then the loss of phylogenetic diversity may be small²⁹. There is growing evidence, however, that extinctions tend to be phylogenetically non-random³⁰, and that species in some clades are at higher risk of extinction³¹. There is as of yet no consensus on expected losses of phylogenetic diversity^{31,32}, nonetheless, there is some evidence that evolutionarily distinct plant species might be disproportionately at risk of extinction³³, which could elevate losses³⁴. A European study suggested that extinctions increased differentiation of regional floras, but was based on the loss of just

69 species²¹. The aggregate effect of more widespread losses on phylogenetic turnover between regions across the globe have yet to be quantified.

Using a global dataset on ~200,000 vascular plant species, we quantify how non-native naturalizations and recent native extinctions have impacted local (α) and between community (β) plant diversity across spatial scales. We then explore differences in biotic homogenization under varying future scenarios of increasing extinction intensity. By characterizing common routes of human assisted migration we also identify those biomes that are most susceptible to changes in community composition and ecological rearrangement in the Anthropocene^{10,35}. We map the distribution of each species using distribution models fitted to carefully curated species occurrence records, and contrast 'Holocene' and 'Anthropocene' species diversities around the globe. We define species composition in the Holocene as the native species' assemblages in each region before widespread migration by humans as initiated by the Columbian Exchange circa 1492¹⁶. Species composition in the Anthropocene post-date this seminal event, and includes non-native naturalizations, and recent past and projected plant extinctions²⁶. However, there is some evidence of non-native plant naturalizations by humans across regions in pre-Columbian times^{36,37}. We quantify changes in plant community diversity (α -diversity) between the Holocene and Anthropocene epochs, and examine the signature of increasing homogenization (lower β -diversity) at regional and global scales. We then evaluate the relative contribution of naturalizations vs extirpations in restructuring global plant diversity, and the macroecological correlates of changes in floristic composition across varying extinction scenarios.

Taken together, we reveal that regardless of extinction scenario, the strongest contributor to biotic homogenization results from non-native naturalizations. We show that the biogeographic histories of these recent, human-mediated plant movements between regions is imbalanced. Asia and North America are major sources of non-natives but the species they export are phylogenetically close to recipient floras. These results highlight yet another imprint of the Anthropocene and demonstrate the profound influence humans exert on regional biotas beyond changes in species richness.

Results and discussion

Temporal changes in α -diversity across plant communities. Under a 'best case' scenario defined as recent plant extinctions and naturalizations, but discounting possible future extinctions, we show that the magnitude of naturalizations is far greater than the magnitude of plant extinctions. Approximately 4.9% (10,138) of plant species have been naturalized to a region outside their native ranges (Fig. 1a), while an estimated 0.5% (1065) of species have gone extinct to date (Fig. 1a), leading to an estimated loss of >14,000 million years of evolutionary history (Fig. 1). The trend of declining species and phylogenetic diversity is not an artefact of the spatial resolution (Supplementary Fig. 1), with most losses in North America (particularly California and Florida), Mesoamerica, the Amazon, the Himalaya-Hengduan, Southeast Asia and southwest Australia (Fig. 1d, g, j). These regions are characterized by a number of spectacular clade radiations^{38–41}, but have also experienced high levels of threat and species invasion⁴².

Temporal changes in compositional turnover across floras. We found global decreases in β -diversity (the turnover of species and standardized phylogenetic diversity) across most regions (Fig. 2). Shifts towards increasing homogeneity and increasing α -diversity are most pronounced in regions with high elevations and greater rainfall under most scenarios (Supplementary Fig. 2). In

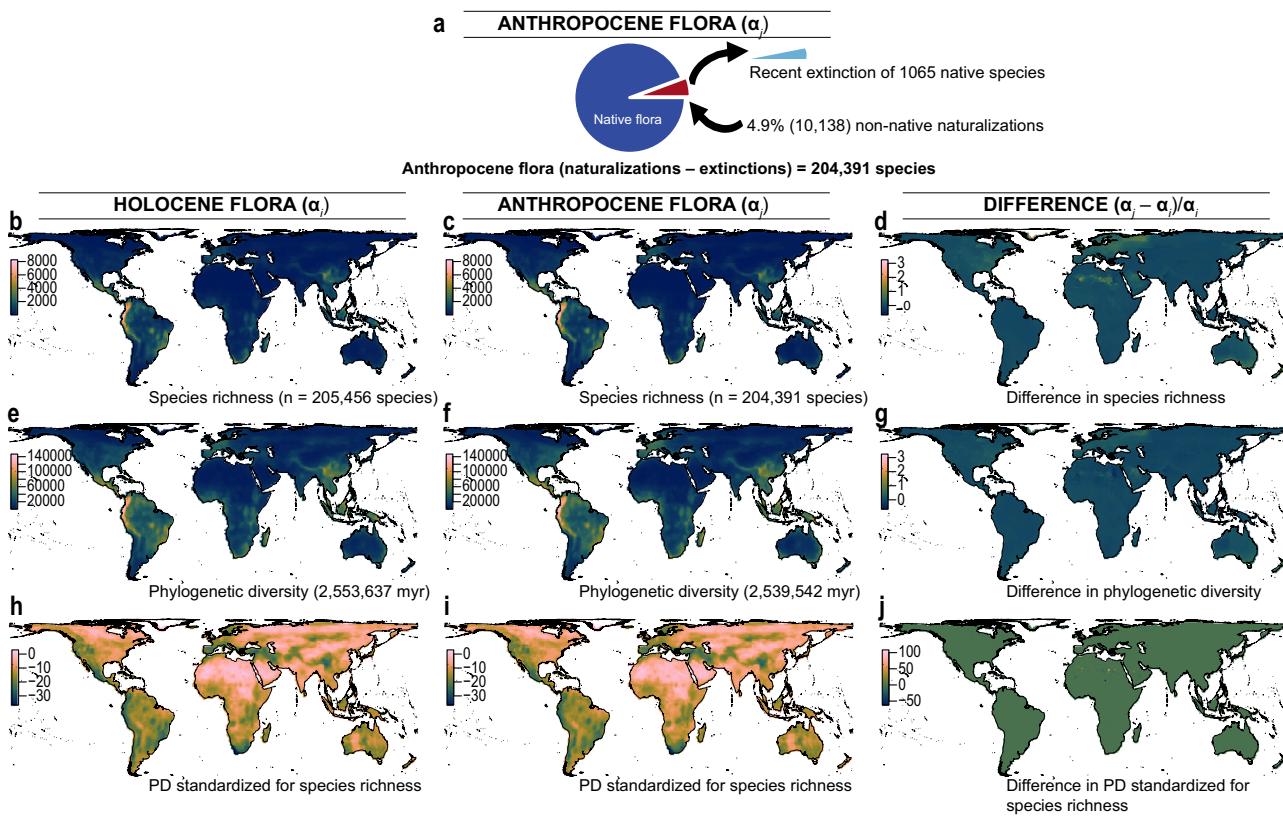


Fig. 1 Temporal and spatial changes in α -diversity across plant communities in the Anthropocene based on recent plant extinctions and naturalizations (best case scenario). Left panel shows the Holocene flora, middle the Anthropocene flora (based on recent extinctions and naturalizations) and right panel differences between Holocene and Anthropocene floras. **a** Schematic of the Anthropocene flora showing recent extinctions replaced by non-native naturalizations. **b-d** Spatial and temporal changes in species (α) diversity. **e-g** Spatial and temporal changes in observed phylogenetic (α) diversity. **h-j** Spatial and temporal changes in phylogenetic (α) diversity standardized for species richness (phylogenetic tip shuffling 1000 times). Species diversity was calculated as the numbers of species within 100 km \times 100 km grid cells (see Supplementary Fig. 1 for a different spatial scale). Phylogenetic diversity (PD) was calculated in million years (myr) as the sum of all phylogenetic branch lengths for the set of species within each grid cell. Species richness was corrected for by calculating the standardized effective size of phylogenetic (α) diversity based on 1000 randomizations (see Methods). Maps are in Behrmann equal-area projection.

particular, Northern Canada, the Saharan Desert (overlapping Chad and Libya), Saudi Arabia, Northern Russia, and Victoria (Australia) are the main epicenters of species and phylogenetic homogenization (Fig. 2c, f).

We demonstrate that changes in α - and β -diversity are driven predominantly by the naturalization of non-native species, rather than recent native species extinctions (Fig. 3). Even under future scenarios of increasing extinction intensity – assuming a ‘worst case’ when all currently threatened species become extinct – non-native and invasive species naturalizations are by far the strongest contributor to biotic reorganization (Fig. 3). Although our models did not account for non-native colonizations into the future, we suggest that our exploration of alternative extinction scenarios has strong parallels with the widespread use of climate projections to model future ecological scenarios^{43–45}. Previous work has indicated that the spread of non-natives might also compensate for biodiversity losses due to species extirpations (e.g.,⁹); however, we find that this is true only to a point, with most regions showing increases in alpha diversity and declines in beta diversity (Fig. 3).

Our results illustrating the disproportionate impact of non-native species are robust to choice of dissimilarity metric (Supplementary Fig. 3), and varying assumptions of extinction pressures (Fig. 3). We also explored whether our results were driven by a few species of large effect—superinvaders—i.e., non-native species with unusually large invaded ranges. While

superinvaders may have a substantial influence on changes in species diversity through time, they account for only 0–14% of the non-native species across all regions, and excluding them does not change our key findings that widespread reorganization of plant communities is primarily due to species naturalizations. The contribution of non-native species to biotic homogenization has been previously documented for birds⁴⁶, fish^{22,47}, insects⁴⁸, and plants at regional scales^{19,21,22,49,50}. To our knowledge, our study is the first global assessment of plants.

Exchange of non-native plant species and phylogenetic diversity across continents. We additionally illustrate how the exchange of species and phylogenetic diversity between regions is strongly asymmetrical (Fig. 4). The exchange and accumulation of non-native plants has been documented previously¹⁵, but their phylogenetic signature has been less well characterized. Here we show that while the principal donors of non-native species are Temperate Asia and North America, the species they export tend to be phylogenetically close to recipient floras (Fig. 4a; Supplementary Tables 2 and 3). Australasia, Pacific and Europe, in contrast, contribute fewer species to the global pool of non-native species, but a disproportionate amount of phylogenetic diversity (Fig. 4b, c). The biased pathways of naturalization we uncover here likely reflect major routes of human-mediated dispersal among regions (perhaps reflecting global trade), the climatic (dis)

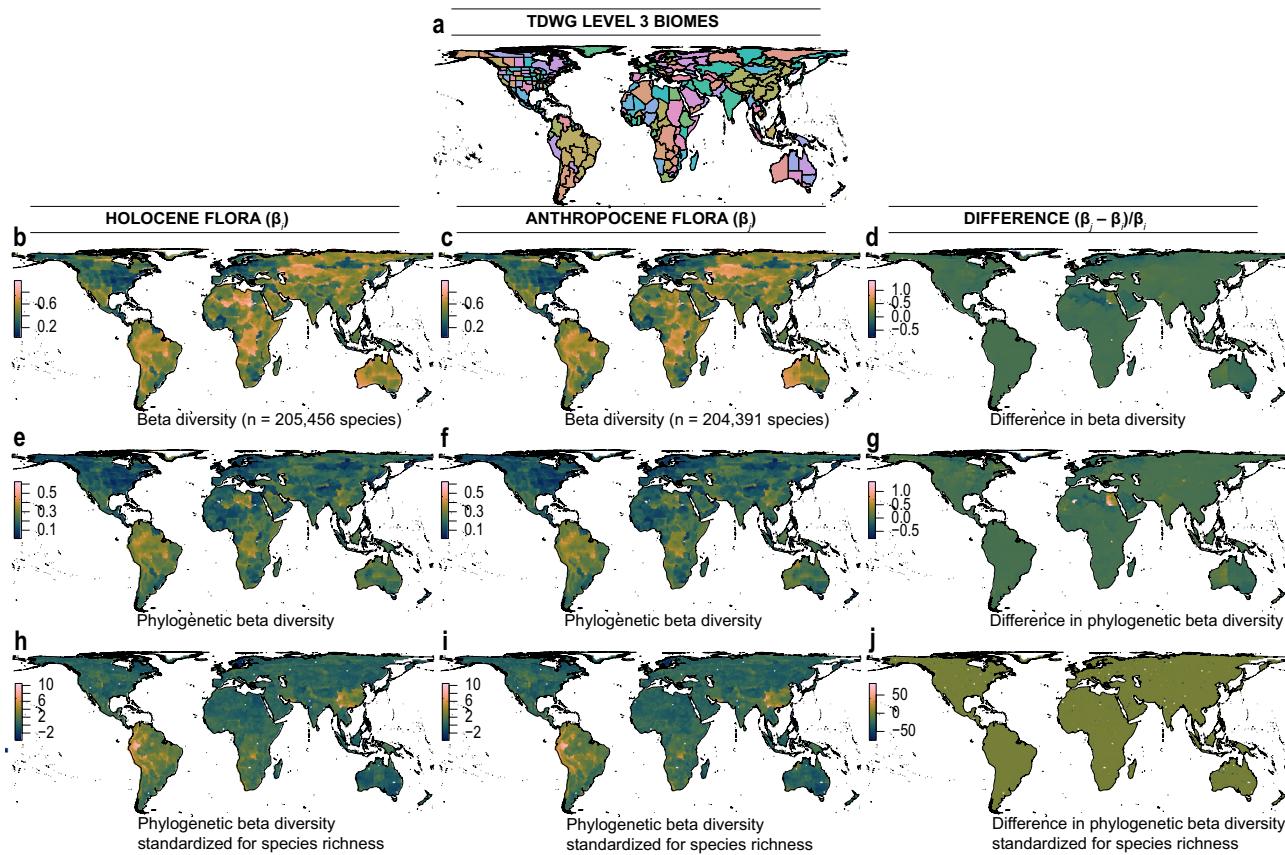


Fig. 2 Spatial and temporal changes in β -diversity between Holocene (pre-Columbian) and Anthropocene floras based on recent plant extinctions and naturalizations (best case scenario). Left panel shows the Holocene flora, middle the Anthropocene flora (based on recent extinctions and naturalizations) and right panel differences in turnover (homogenization) between Holocene and Anthropocene epochs. **a** The geographic sampling unit within level 3 regional classification as defined by the Biodiversity Information Standards Taxonomic Databases Working Group (TDWG). **b-d** Spatial and temporal changes in turnover (β -diversity) in species diversity. **e-g** Spatial and temporal changes in turnover (β -diversity) in phylogenetic diversity. **h-j** Spatial and temporal changes in phylogenetic β -diversity standardized for species richness (phylogenetic tip shuffling 1000 times). Both species and phylogenetic turnover were calculated using Simpson's metric of beta and phylogenetic beta diversity respectively, between 100 km \times 100 km grid cells aggregated across level 3 TDWG biomes. **a**, Maps are in Behrmann equal-area projection.

similarity between donor and recipient regions, and the vulnerability of regional floras to invasion.

Relationships of non-natives to native flora across spatial scales. Contrary to Darwin's naturalization hypothesis²³, we find that, on average, non-native species are not phylogenetically distinct from native plant communities (Supplementary Fig. 4). This is not true in all regions (exception Africa, Australasia and Pacific), and superinvaders tend to be more closely related to other non-natives than expected by chance (Supplementary Fig. 4). We also detect strong taxonomic structure in the familial membership of naturalized species. In particular, naturalized species in temperate Asia and North America cluster within similar families ($r = 0.830$; Spearman rank correlation). This is not true in Europe and South America, however, where naturalized species are represented among diverse families ($r = 0.20$; Spearman rank correlation; Supplementary Fig. 5). Such phylogenetic and taxonomic structuring emphasizes the importance of evolutionary history in species naturalization and establishment success, reflecting phylogenetic niche conservatism in environmental preferences and invasive potential⁵¹. Our analyses at the regional scale thus lend support to the pre-adaptation hypothesis of species invasion, also posited by Darwin⁵².

Species are not static in their geographic distributions; some may have been moved by people historically, and today many

species are tracking shifting climates. We recognize that generating a reliable estimation of the distribution ranges for extinct species would be challenging plus the historical naturalization of species beyond their native range may have already contributed to the homogenization of local floras. Likewise, data on species naturalizations might be more available than data on species extirpations, potentially biasing us to detect a stronger effect of naturalizations in our analyses. Our analyses thus capture the additional impact on biotic homogenization of more recent anthropogenic activities, and thus likely underestimates the true impact people have on native biodiversity. Further, we believe that as human populations have expanded only relatively recently, historical plant extinctions may have been less likely than historical translocations, and thus our findings that homogenization has been driven largely by naturalizations, rather than extinctions, is likely conservative.

We have demonstrated how recent native species extinctions and, more notably, non-native species naturalizations have reshaped native plant communities across tens of thousands of square kilometers, resulting in profound homogenization of global biodiversity. The floristic shifts we document largely result from human facilitated migrations during the past 500 years (and likely mostly within the past 200 years) and represent yet another imprint of the Anthropocene. Biodiversity change in the Anthropocene often manifests as habitat conversion for human

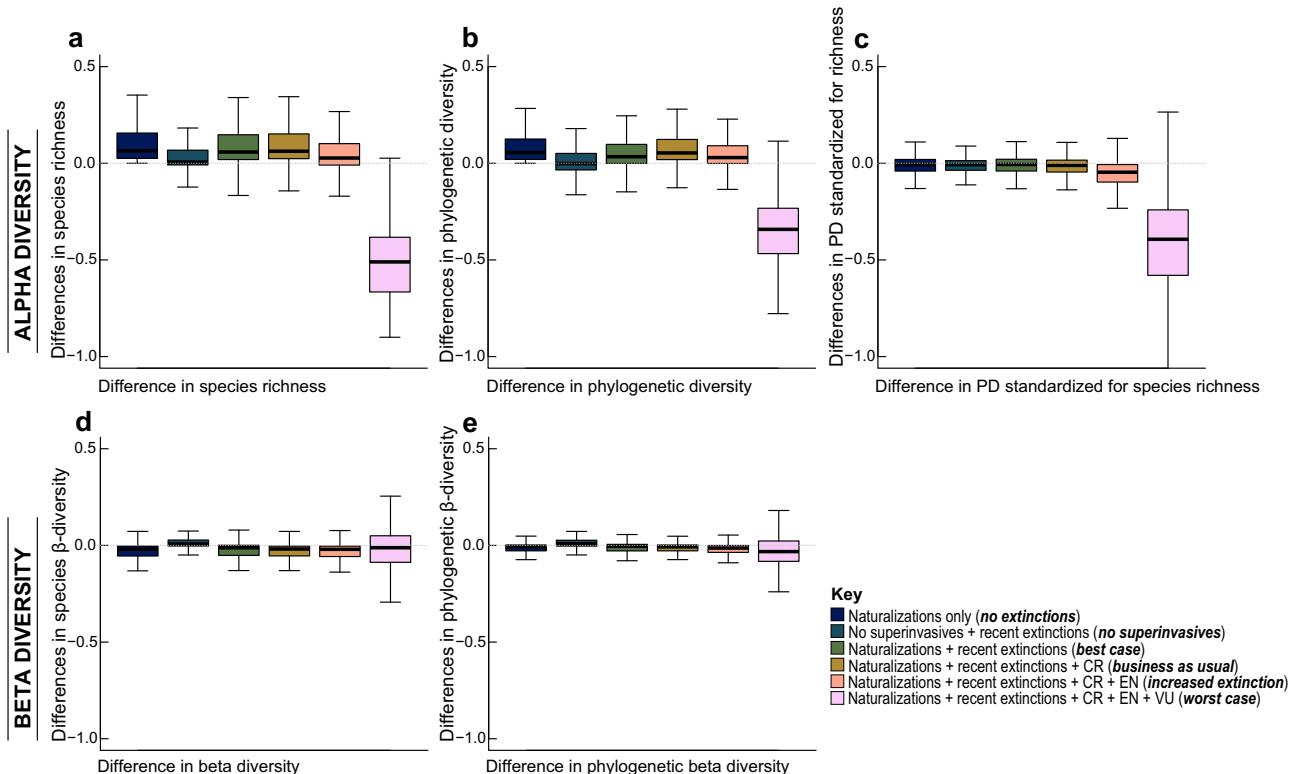


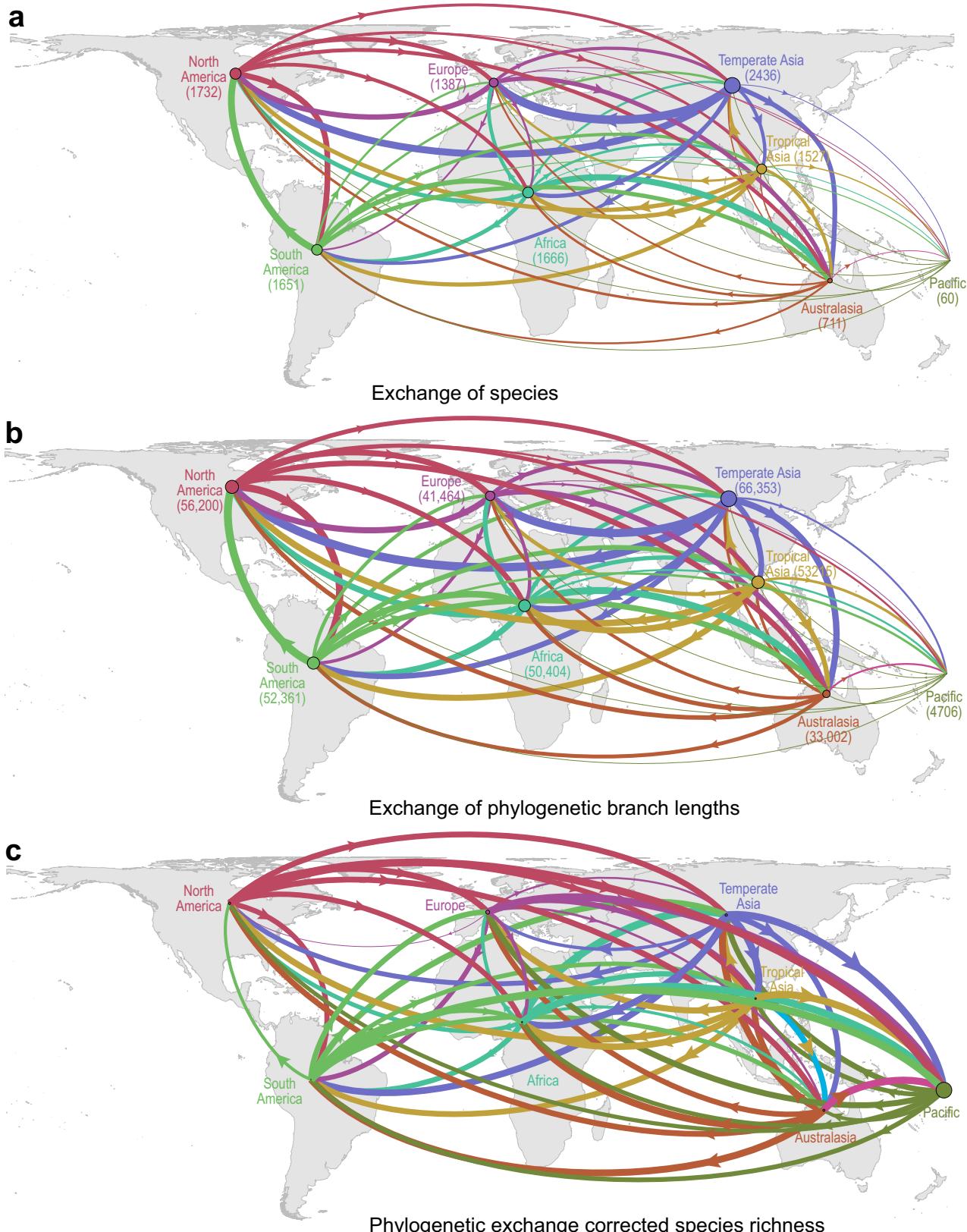
Fig. 3 Changes in plant communities under various scenarios of extinctions and naturalizations in the Anthropocene. Top row **a** ($n=13,218$ grid cells), **b** ($n=13,218$ grid cells), and **c** ($n=13,218$ grid cells) shows the differences in α -diversity and bottom row **d** ($n=13,218$ grid cells), and **e** ($n=13,218$ grid cells), shows differences in β -diversity. Comparisons are made across six scenarios: i) 'no extinctions' recent naturalizations only, ii) 'no superinvasives' based on the removal of non-native species with unusually large invaded ranges, iii) 'Best case' (based on recent extinctions and naturalizations that have occurred to date), iv) 'business as usual' projected extinction of critically endangered species (CR), v) 'increased extinction' based on projected extinction of endangered (EN) and CR species, and vi) 'worst case' based on projected extinction of all threatened species including vulnerable (VU), EN and CR species. Dashed line at zero corresponds to no change. Species richness was calculated as the numbers of species within $100\text{ km} \times 100\text{ km}$ grid cells. Phylogenetic diversity was calculated as the sum of all phylogenetic branch lengths for the set of species within each grid cell. The bottom and top of boxes show the first and third quartiles respectively, the median is indicated by the horizontal line, the range of the data by the whiskers. The dataset used for the analysis included 205,456 native species, 1065 recently extinct species, extinction projections for 150,000 species, and 10,138 naturalized species. Source data are provided as a Source Data file.

use, and it is driving the loss of wilderness areas, elevating species extinctions and promoting non-native naturalizations, at scales comparable to a global biodiversity crisis^{53,54}. The large-scale transport of species across the globe was likely facilitated initially by long-distance trade and travel via sea, beginning especially with the Columbian Exchange, when food crops, diseases, and populations started to be exchanged between hemispheres by humans¹⁶. While thousands of species have been spread to new areas unintentionally since this time—for example, as hitchhikers in ship ballast water⁵⁵ or dispersed to new climates by migratory animals⁵⁶—others were deliberately introduced to new areas for agriculture and horticulture. The increasing industrialization of agriculture and other drivers of biodiversity change have undoubtedly further accelerated the pace of floristic homogenization within recent decades⁵⁷. The consequences of this global biotic reorganization on ecosystems remain poorly understood, but there is increasing evidence that biotic heterogeneity provides insurance for the maintenance of ecosystem functioning⁵⁸ in the face of ongoing global change.

Methods

Estimating native plant species' distributions. We used the newly developed species database, GreenMaps, to estimate native plant species' distributions⁵⁹. GreenMaps includes global distribution maps for ~230,000 vascular plant species. Maps were generated using species distribution models—the statistical estimation of species geographic distributions based on only some known occurrences and

environmental conditions—derived from carefully curated species occurrence records. Occurrence records were obtained from a variety of sources, including herbarium specimens, primary literature, personal observation, and online data repositories including the Global Biodiversity Information Facility^{60–62}, and Integrated Digitized Biocollections (<https://www.idigbio.org/>). These records were thoroughly cleaned to reconcile names to follow currently accepted taxonomies [e.g., World Flora Online (www.worldfloraonline.org)], and to remove duplicates and records with doubtful or imprecise localities. Two stringent spatial filters were employed to restrict species' distributions to their known native ranges (i.e., realized niches) and to prevent erroneous records and predictions in areas that contain suitable habitat but are unoccupied by the species (i.e., fundamental niche). First, we applied the spatial constraint, APGfamilyGeo, which are expert drawn occurrence polygons ("expert maps") of plant family distributions^{63,64} (see Data availability) to restrict species to within these distributions. Second, we applied GeoEigenvectors, which are orthogonal variables representing spatial relationships among cells in a grid, encompassing the geometry of the study region at various scales⁶⁵. For the latter, we generated a pairwise geographical connectivity matrix among grid cells to establish a truncation distance for the eigenvector-based spatial filtering, returning a total of 150 spatial filters. These filters were then resampled to the same resolution as the input environmental variables, and were included with the bioclimatic variables in the species distribution modeling. Bioclimatic variables were derived from WorldClim⁶⁶ for a total of 19 variables (Supplementary Table 1). Species distribution models (SDMs) were fitted using four different algorithms: generalized linear models (GLM), generalized boosted models (GBM), maximum entropy (MaxEnt), and random forests (RF) with a binomial error distribution (with logit link). Model settings were chosen to yield intermediately complex response surfaces. Model performance was evaluated using area under the receiver operating curve (AUC) and true skill statistic (TSS) scores. AUC scores range from 0 to 1 and should be maximized whereas TSS scores range from -1 to 1. Prior to model building, all predictor variables were standardized. Univariate variable importance for each predictor was assessed in a 5-fold spatial block



cross-validation design. The ensemble predictions from species distribution models were derived using un-weighted ensemble means. Predictive model performance was assessed using a 5-fold spatial block cross-validation. We generated a total of 230,000 range maps, representing species within 382 families at a resolution of 50 × 50 km which was also resampled to 100 × 100 km. To our knowledge, this makes it the largest and only global assessment of geographic distributions for plants at the species-level. Our approach of modeling species distributions follows

the guidelines of ODMAP (Overview, Data, Model, Assessment, Prediction), a comprehensive framework of best practices for reporting species distribution models⁶⁷ (see Supplementary Material 1). These maps were stacked and converted to a community matrix for downstream analyses. We also provide a new R function, *sdm*, for performing the SDMs across four algorithms (random forest, generalized linear models, gradient boosted machines, and MaxEnt) tailored for SDMs of large datasets. The *sdm* function is included in our R package

Fig. 4 Asymmetrical exchange of phylogenetic diversity and non-native plant species across the world. **a** Non-native species originating (outbound arrow) or received (inbound arrows) between each continent. Line thickness is proportional to the number of species exchanged. **b** Phylogenetic diversity of non-native species originating (outbound arrow) or received (inbound arrows) between each continent. Line thickness is proportional to the sum of branch lengths exchanged. **c** Net donors and recipients of phylogenetic diversity after correcting for species richness, calculated as the difference in total phylogenetic diversity between the Holocene flora and the Anthropocene flora across continents divided by the number of species exchanged. Arrows indicate the direction of flows from donor to recipient continent, with line thickness proportional to the sum of shared branch lengths weighted by the inverse of species richness. The numbers within parenthesis and circle size represents the number of non-native species or phylogenetic branch lengths in each region. All phylogenetic analyses were run across 100 trees and the median reported. The maps are in Behrmann equal-area projection. A breakdown of nodes and edges exchanged is presented in Supplementary Tables 2 and 3. Source data are provided as a Source Data file.

*phyloregion*⁶⁸ along with improved documentation and vignettes to show practical application of this functionality under various modeling scenarios. The *sdm* function was designed with multiple checks such that any species that did not meet one or more checks were filtered out. A feature of novelty of the *sdm* function is the addition of an algorithm that allows a user to exclude records that occur within a certain distance to herbaria, museums or other infrastructure. By default, we used the most updated version of *Index Herbariorum*, a global directory of herbaria⁶⁹, but a user has the option to specify their own infrastructure to exclude.

We validated the output distribution maps against the Kew Plants of the World Online database (POWO; <http://www.plantsoftheworldonline.org/>), which includes native distribution maps for all plants of the world within major biogeographically defined areas recognized by the Biodiversity Information Standards (also known as the Taxonomic Databases Working Group (TDWG))⁷⁰. Although the Kew's distributions of native species are largely based on state/province level such that if a species was observed in any location within a state the whole state is marked as its distribution range, our GreenMaps approach only used the Kew distributions to restrict modeled species distributions within such biogeographic areas. See ref. ⁵⁹ for full description of the workflow. The range map rasters were converted to a community matrix using the function *raster2comm* in our new R package *phyloregion*⁶⁸ for downstream analysis.

Estimating non-native plant species' distributions. We used the Global Naturalized Alien Flora (GloNAF) database version 1.2^{71,72} to compile a checklist of non-native species, including documented records of alien plants that have dispersed into new regions largely by humans, and which have become successfully naturalized^{73,74}. The dataset includes non-native species distributions within TDWG regions. We generated species' distributions for these species using the GreenMaps approach⁵⁹ described above, but removing the spatial filters *APGfamilyGeo* and *GeoEigenvectors*. The non-native species ranges were modeled using occurrences that fell outside the boundaries of the native range of each species as determined by Plants of the World Online (POWO). Specifically, we used the following R code to subset occurrences falling outside of PWO as follows:

$$y <- x[!complete.cases(sp :: over(x, pwo)),] \quad (1)$$

where *x* is a data frame of occurrence of a species, and *pwo* a shapefile of the native range of the species. We then used the output *y* to model the distribution of non-native species using the *sdm* function in the R package *phyloregion*⁶⁸. We validated our non-native species distribution models against the GloNAF dataset by overlaying grid cells of non-native species predictions within GloNAF's TDWG levels, and selecting only those projected occurrences that fell within the naturalized range indicated by GloNAF. Such approach allowed us to capture the precise distribution of the non-native species within a state/province as opposed to broadly scoring them present or absent in a state/province as did GloNAF. From our dataset of non-native species, we also identified 'superinvasives', here defined as non-native species with $1.5 \times$ the interquartile range above the third quartile of their invaded range size within a TDWG region.

Recently extinct and threatened plant species. We compiled information on recent plant extinctions and conservation status of each mapped species. Our dataset of recent extinctions comes from a dataset that includes 1065 plant species that have become extinct since Linnaeus' Species Plantarum⁷⁵, derived from a comprehensive literature review and assessments of the International Union for Conservation of Nature (IUCN) Red List of Threatened Species^{26,76}. We also explored alternative scenarios of increasing future extinction intensity, considering future losses of currently extant native species, some of which are not currently recognized as of global concern (data from ref. ²⁷). For the latter analysis, we compiled information on the conservation status of each species and apply the term 'extinction' loosely, which included both native species lost from a region as well as native species that may still be present in some part or all of their native ranges, but they are unlikely to remain so in the near future if current trends continue (see ref. ²⁷). This dataset comes from machine-learning predictions of conservation status for over 150,000 land plant species²⁷ defined as the probability of each species as belonging to a Red List non-Least Concern category (i.e., likely being at risk on some level) based on geographic, environmental, and morphological trait data, variables that are key in predicting conservation risk²⁷. For our

purposes here, we assumed that Least Concern species were not at risk of extinction; although we recognize that a substantial proportion of these species may in fact be endangered^{27,77}. Within this framework, extinction risk is defined using the expected probability of extinction over 100 years of each taxon⁷⁸, scaled as follows: Least Concern = 0.001, Near Threatened and Conservation Dependent = 0.01, Vulnerable = 0.1, Endangered = 0.67, and Critically Endangered = 0.999. We used these statistical projections to estimate future extinction scenarios because they can be fit to over 150,000 land plant species, whereas formal IUCN Red List assessments are currently available for only 33,573 plant species (March 15, 2020).

The final dataset used for our analysis included 205,456 native species, 1065 recently extinct species, extinction projections for 150,000 species, and 10,138 naturalized species.

Phylogenetic data. We applied the dated phylogeny for seed plants of the world from ref. ⁷⁹, which includes 353,185 terminal taxa. The ref. ⁷⁹ phylogeny was assembled using a hierarchical clustering analysis of DNA sequence data of major seed plant clades and was resolved using data from the Open Tree of Life project. This represents one of the most comprehensive phylogenies of vascular plants at a global scale and includes all species in our analysis. It also provides divergence time estimates to facilitate downstream analytics.

Data analysis. We quantified changes in alpha and beta diversity between the Holocene (native species' assemblages in each region before widespread migration by humans as initiated by the Columbian Exchange circa 1492¹⁶) and Anthropocene (non-native naturalizations, and recent past and projected plant extinctions)²⁶ epochs across 100×100 km grid cells within major biogeographically defined areas recognized by the Biodiversity Information Standards (also known as the Taxonomic Databases Working Group (TDWG))⁷⁰. These TDWG geographic regions correspond to continents, countries, states and provinces. We then explored differences in biotic homogenization under varying future scenarios of extinction including naturalizations only, 'no superinvasives', 'best case' 'business as usual', 'increased extinction' and 'worst case'. Our definition of *best case* refers to recent plant extinctions and naturalizations, and assumes no future extinctions, *business as usual* assumes loss of Critically Endangered (CR) species, *increased extinction* assumes loss of Critically Endangered (CR) and Endangered (EN) species, and the *worst case* scenario assumes loss of all threatened species. Because biodiversity patterns are scale dependent, varying along spatial grains and geographic extents^{80,81}, we repeated all analyses at spatial grid resolution of 50×50 km.

Temporal changes in α -diversity across plant communities. For each grid cell, temporal and spatial change in α -diversity was quantified as the difference in species (or phylogenetic) diversity between the Anthropocene (*j*) and Holocene (*i*) periods (see above) expressed as:

$$\Delta\alpha = (\alpha_j - \alpha_i)/\alpha \quad (2)$$

Negative $\Delta\alpha$ values imply that alpha diversity has decreased and positive values indicate increased alpha diversity. Species α -diversity was calculated as the total count of species in each grid cell. Phylogenetic α -diversity was computed as the sum of the phylogenetic branch lengths connecting species from the tip to the root of a dated phylogenetic tree in each grid cell⁸². We also assessed changes in phylogenetic (α) diversity standardized for species richness by calculating standard effects sizes of phylogenetic diversity in communities by shuffling the tips in the phylogeny based on 1000 randomizations. For each iteration of the randomization, the analysis was regenerated using the same set of spatial conditions, but using the randomized version of the tree after which the z-score for each index value was calculated (observed - expected)/sqrt (variance). Temporal changes in α -diversity was assessed at the spatial grain resolution of 50 and 100 km to account for the effects of scale.

Temporal changes in compositional turnover across floras. Within TDWG geographic regions, we generated pairwise distance matrices of phylogenetic β -diversity (β_{phylo})⁸³ and species β -diversity (β_{tax}) between all pairs of grid cells, and

compared Holocene and Anthropocene epochs. We used Simpson's index for quantifying compositional turnover because it is insensitive to differences in total diversity among sites^{84,85}. The phylogenetic equivalent, β_{phylo} , represents the proportion of shared phylogenetic branch lengths between cells, and ranges from 0 (species sets are identical and all branch lengths are shared) to 1 (species sets share no phylogenetic branches). We calculated change in compositional turnover ($\Delta\beta$) as:

$$\Delta\beta = (\beta_j - \beta_i)/\beta_i \quad (3)$$

where j is the Anthropocene species pool and i refers to the Holocene species composition. Negative $\Delta\beta$ values imply that taxonomic/phylogenetic similarity has increased (i.e., biotic homogenization) and positive values indicate biotic differentiation. To assess sensitivity to our choice of diversity index, we re-ran all analyses using Sørensen and Jaccard dissimilarity indices. All (phylogenetic) β -diversity metrics were calculated using our new R package *phyloregion*⁶⁸.

Effect of superinvasive species. To determine the extent to which a small number of superinvasive non-native species may be driving patterns of homogenization, we re-ran the main analyses described above, but excluded non-native species with the widest ranges within biomes, i.e., species that are more than 1.5× the interquartile range above the third quartile of (invaded) range sizes (i.e., statistical outliers) within TDWG regions. Our definition of range size corresponds to the number of grid cells occupied by a species.

Phylogenetic structure of naturalizations. We evaluated whether naturalized species were more likely to have become naturalized in recipient communities in the absence of close relatives—Darwin's naturalization hypothesis—by comparing the mean phylogenetic distance between each non-native species and its nearest phylogenetic neighbor in the recipient flora. Larger mean phylogenetic distances indicate that non-native species tend to be less closely related to the native flora. We first ran each analysis on a set of 100 trees. Significance was assessed by comparing the distribution of observed phylogenetic distances to a null model shuffling non-native status randomly on the tips of the phylogeny (1000 replicates) as implemented in the R package *phyloregion*⁶⁸.

Drivers of change in composition across floras. To relate change in alpha and beta diversity to possible external drivers, we obtained three sets of variables for each site: (i) ecological: mean annual precipitation (MAP), mean annual temperature (MAT), and elevation; (ii) evolutionary: range size (as proxy for dispersal potential, defined as the average range size across species within a grid cell); and (iii) anthropogenic: wilderness index (inverse of human footprint index). MAP, MAT, and elevation were obtained from the WorldClim database⁶⁶; the geographic range of each species was calculated as the number of cells a species occupied. The Wilderness Index was obtained from ref. ⁸⁶, and describes the degree to which a place is remote from and undisturbed by the influences of modern society⁸⁶. These variables were converted to Behrmann equal-area projection using the function *projectRaster* in the R package *raster*⁸⁷.

We used a linear mixed effects (LME) model of temporal change in, separately, species (α) richness, phylogenetic (α) diversity, phylogenetic (α) diversity standardized for richness, β-diversity, and phylogenetic β-diversity between the Anthropocene and Holocene, against ecological, evolutionary and anthropogenic variables as predictors. We used level 3 regions as recognized by the Biodiversity Information Standards as a random effect, allowing us to account for idiosyncratic differences between regions. Changes in metrics of β-diversity were applied to grid cells by taking the average dissimilarity to other cells within a region as defined by the TDWG level 3 biomes, whereas changes in metrics of α diversity were applied directly to grid cells. We also included a spatial covariate of geographical coordinates as an additional predictor variable to account for spatial autocorrelation. Our model can be formulated as follows:

$$\Delta_i = \beta_0 + \beta_1 \text{MAT}_i + \beta_2 \text{MAP}_i + \beta_3 \text{elevation}_i + \beta_4 \text{range_size}_i + \beta_5 \text{wilderness}_i + e_i \quad (4)$$

where Δ_i is the temporal diversity change (temporal changes in metrics of α or β diversity) between the Anthropocene and Holocene in grid cell i , β_0 to β_5 are fixed effect parameters, and e_i is residual error. The LME model was fitted using the *lme* function in the *nlme* R package⁸⁸.

A vignette, with a worked example, data and R codes describing all the steps for the analyses, is also provided on Dryad (<https://doi.org/10.5061/dryad.f4qrjf6st>).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The plant species range maps included in this study come from a newly developed species database called GreenMaps (<https://doi.org/10.1101/2020.02.21.960161>). GreenMaps includes global distribution maps for ~230,000 vascular plant species. The maps were generated using species distribution models derived from carefully curated species occurrence records, and the dataset is archived on Dryad (<https://doi.org/10.5061/dryad.f4qrjf6st>). Occurrence records were obtained from a variety of sources,

including herbarium specimens, primary literature, personal observation, and online data repositories including the Global Biodiversity Information Facility (Accession: <https://doi.org/10.15468/dl.7ujp48>; <https://doi.org/10.15468/dl.jw4u5a>, and <https://doi.org/10.15468/dl.m8dzn5>), and Integrated Digitized Biocollections (<https://www.idigbio.org/>). The phylogeny used for the analyses is a published phylogeny that is already available in public repositories. Specifically, the plant phylogeny was downloaded from Smith and Brown (<https://doi.org/10.1002/ajb2.1019>). Source data are provided with this paper.

Code availability

All scripts and codes necessary to repeat our analyses have been made available in the Dryad database [<https://doi.org/10.5061/dryad.f4qrjf6st>] under the folder "R-CODES", and also in the R package *phyloregion*⁶⁸.

Received: 4 February 2021; Accepted: 8 November 2021;

Published online: 06 December 2021

References

- Trindade, D. P. F., Carmona, C. P. & Pärtel, M. Temporal lags in observed and dark diversity in the Anthropocene. *Glob. Change Biol.* **26**, 3193–3201 (2020).
- Barnosky, A. D. et al. Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).
- Ceballos, G. et al. Accelerated modern human-induced species losses: entering the sixth mass extinction. *Sci. Adv.* **1**, e1400253 (2015).
- Joppa, L. N. et al. Filling in biodiversity threat gaps. *Science* **352**, 416–418 (2016).
- Díaz, S. et al. Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* **366**, eaax3100 (2019).
- Wearn, O. R. et al. Extinction debt and windows of conservation opportunity in the Brazilian Amazon. *Science* **337**, 228–232 (2012).
- Veron, S. et al. Predicting loss of evolutionary history: where are we? *Biol. Rev.* **92**, 271–291 (2017).
- Cardinale, B. J. et al. Is local biodiversity declining or not? A summary of the debate over analysis of species richness time trends. *Biol. Conserv.* **219**, 175–183 (2018).
- Vellend, M. et al. Plant biodiversity change across scales during the Anthropocene. *Annu. Rev. Plant Biol.* **68**, 563–586 (2017).
- Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **519**, 171–180 (2015).
- McKinney, L. M. & Lockwood, J. L. Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends Ecol. Evol.* **14**, 450–453 (1999).
- Dornelas, M. et al. Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296–299 (2014).
- Nowakowski, A. J. et al. Phylogenetic homogenization of amphibian assemblages in human-altered habitats across the globe. *Proc. Natl Acad. Sci. USA* **115**, E3454–E3462 (2018).
- Levis, C. et al. Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science* **355**, 925–931 (2017).
- van Kleunen, M. et al. Global exchange and accumulation of non-native plants. *Nature* **525**, 100–103 (2015).
- Nunn, N. & Qian, N. The Columbian exchange: a history of disease, food and ideas. *J. Econ. Perspect.* **24**, 163–188 (2010).
- McKinney, M. L. Measuring floristic homogenization by non-native plants in North America. *Glob. Ecol. Biogeogr.* **13**, 47–53 (2004).
- Keith, S. A. et al. Taxonomic homogenization of woodland plant communities over 70 years. *Proc. Biol. Sci.* **276**, 3539–3544 (2009).
- Carvalho, G. O. & Castro, S. A. Invasions but not extinctions change phylogenetic diversity of angiosperm assemblage on southeastern Pacific Oceanic islands. *PLoS ONE* **12**, e0182105 (2017).
- Gossner, M. M. et al. Land-use intensification causes multitrophic homogenization of grassland communities. *Nature* **540**, 266–269 (2016).
- Winter, M. et al. Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. *Proc. Natl Acad. Sci. USA* **106**, 21721–21725 (2009).
- Li, D. et al. Changes in taxonomic and phylogenetic diversity in the Anthropocene. *Proc. R. Soc. B* **287**, 20200777 (2020).
- Daehler, C. C. Darwin's naturalization hypothesis revisited. *Am. Nat.* **158**, 324–330 (2001).
- Diez, J. M. et al. Duncan, Darwin's naturalization conundrum: dissecting taxonomic patterns of species invasions. *Ecol. Lett.* **11**, 674–681 (2008).
- Purvis, A. et al. Extinction. *Bioessays* **22**, 1123–1133 (2000).
- Humphreys, A. M. et al. Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nat. Ecol. Evol.* **3**, 1043–1047 (2019).

27. Pelletier, T. A. et al. Predicting plant conservation priorities on a global scale. *Proc. Natl Acad. Sci. USA* **115**, 13027–13032 (2018).
28. Enquist, B. J. et al. The commonness of rarity: Global and future distribution of rarity across land plants. *Sci. Adv.* **5**, eaaz0414 (2019).
29. Nee, S. & May, R. M. Extinction and the loss of evolutionary history. *Science* **278**, 692–694 (1997).
30. Davies, T. J. The macroecology and macroevolution of plant species at risk. *N. Phytol.* **222**, 708–713 (2019).
31. Purvis, A. et al. Nonrandom extinction and the loss of evolutionary history. *Science* **288**, 328–330 (2000).
32. Parhar, R. K. & Mooers, A. Ø. Phylogenetically clustered extinction risks do not substantially prune the tree of life. *PLoS ONE* **6**, e23528 (2011).
33. Daru, B. H. et al. A global trend towards the loss of evolutionarily unique species in mangrove ecosystems. *PLoS ONE* **8**, e66686 (2013).
34. Redding, D. W. et al. Evolutionarily distinctive species often capture more phylogenetic diversity than expected. *J. Theor. Biol.* **251**, 606–615 (2008).
35. Barnosky, A. D. et al. Approaching a state shift in Earth's biosphere. *Nature* **486**, 52–58 (2012).
36. Byrne, R. & McAndrews, J. Pre-Columbian purslane (*Portulaca oleracea* L.) in the New World. *Nature* **253**, 726–727 (1975).
37. McRostie, V. B. et al. The pre-Columbian introduction and dispersal of Algarrobo (*Prosopis*, Section Algarobia) in the Atacama Desert of northern Chile. *PLoS ONE* **12**, e0181759 (2017).
38. Sauquet, H. et al. Contrasted patterns of hyperdiversification in Mediterranean hotspots. *Proc. Natl Acad. Sci. USA* **106**, 221–225 (2009).
39. Pennington, R. T. & Dick, C. W. *Diversification of the Amazonian flora and its relation to key geological and environmental events: a molecular perspective in Amazonia, landscape and species evolution: A look into the past* 1st ed. pp. 373–385 (Blackwell Publishing, 2010).
40. Lu, L. M. et al. Evolutionary history of the angiosperm flora of China. *Nature* **554**, 234–238 (2018).
41. Yu, H. et al. Testing multiple hypotheses for the high endemic plant diversity of the Tibetan Plateau. *Glob. Ecol. Biogeogr.* **28**, 131–144 (2019).
42. Silvério, D. V. et al. Testing the Amazon savannization hypothesis: fire effects on invasion of a neotropical forest by native cerrado and exotic pasture grasses. *Philos. Trans. R. Soc. B* **368**, 20120427 (2013).
43. Sala, O. E. et al. Global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774 (2000).
44. Pecl, G. T. et al. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**, eaai9214 (2017).
45. Powers, R. P. & Jetz, W. Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. *Nat. Clim. Change* **9**, 323–329 (2019).
46. La Sorte, F. A. & McKinney, M. L. Compositional changes over space and time along an occurrence–abundance continuum: anthropogenic homogenization of the North American avifauna. *J. Biogeogr.* **34**, 2159–2167 (2007).
47. Rahel, F. J. Homogenization of fish faunas across the United States. *Science* **288**, 854–856 (2000).
48. White, P. J. T. & Kerr, J. T. Human impacts on environment–diversity relationships: evidence for biotic homogenization from butterfly species richness patterns. *Glob. Ecol. Biogeogr.* **16**, 290–299 (2007).
49. Rosenblad, K. & Sax, D. F. A new framework for investigating biotic homogenization and exploring future trajectories: Oceanic island plant and bird assemblages as a case study. *Ecography* **40**, 1040–1049 (2017).
50. Longman, E. K. et al. Extreme homogenization: The past, present and future of mammal assemblages on islands. *Glob. Ecol. Biogeogr.* **27**, 77–95 (2018).
51. Wiens, J. J. et al. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.* **13**, 1310–1324 (2010).
52. Duncan, R. P. & Williams, P. A. Darwin's naturalization hypothesis challenged. *Nature* **417**, 608–609 (2002).
53. Ellis, E. C. et al. All is not lost: plant biodiversity in the Anthropocene. *PLoS ONE* **7**, e30535 (2012).
54. Seddon, N. et al. Biodiversity in the Anthropocene: prospects and policy. *Proc. R. Soc. B* **283**, 20162094 (2016).
55. Hayes, K. R. Ecological risk assessment for ballast water introductions: a suggested approach. *ICES J. Mar. Sci.* **52**, 201–212 (1998).
56. Daru, B. H. Migratory birds aid the redistribution of plants to new climates. *Nature* **595**, 34–36 <https://doi.org/10.1038/d41586-021-01547-1> (2021).
57. Taylor, B. W. & Irwin, R. E. Linking economic activities to the distribution of exotic plants. *Proc. Natl Acad. Sci. USA* **101**, 17725–17730 (2004).
58. Loreau, M. et al. Biodiversity as spatial insurance in heterogeneous landscapes. *Proc. Natl Acad. Sci. USA* **100**, 12765–12770 (2003).
59. Daru, B. H. GreenMaps: a tool for addressing the Wallacean shortfall in the global distribution of plants. Preprint at <https://www.biorxiv.org/content/10.1101/2020.02.21.960161v1.abstract> (2020).
60. GBIF.org. GBIF Occurrence Download <https://doi.org/10.15468/dl.7ujp48> (2021).
61. GBIF.org. GBIF Occurrence Download <https://doi.org/10.15468/dlw4u5a> (2021).
62. GBIF.org. GBIF Occurrence Download <https://doi.org/10.15468/dl.m8dzn5> (2021).
63. Heywood, V. H. *Flowering plants of the world* (Batsford, 1993).
64. APG IV, An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).
65. Diniz-Filho, J. A. F. & Bini, L. M. Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Glob. Ecol. Biogeogr.* **14**, 177–185 (2005).
66. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
67. Zurell, D. et al. A standard protocol for reporting species distribution models. *Ecography* **43**, 1261–1277 (2020).
68. Daru, B. H. et al. phyloregion: R package for biogeographic regionalization and macroecology. *Methods Ecol. Evol.* **11**, 1483–1491 (2020).
69. Thiers, B. *Index herbariorum: a global directory of public herbaria and associated staff* (New York Botanical Garden's Virtual Herbarium, 2016). <http://sweetgum.nybg.org/science/ih/>.
70. Brummitt, R. K. *World geographical scheme for recording plant distributions*, 2nd ed. Biodiversity Information Standards (TDWG) <http://www.tdwg.org/standards/109> (2001).
71. van Kleunen, M. et al. The Global Naturalized Alien Flora (GloNAF) database. *Ecology* **100**, e02542 (2019).
72. van Kleunen, M. et al. Economic use of plants is key to their naturalization success. *Nat. Commun.* **11**, 3201 (2020).
73. Richardson, D. M. et al. Naturalization and invasion of alien plants: concepts and definitions. *Divers. Distrib.* **6**, 93–107 (2000).
74. Blackburn, T. M. et al. A proposed unified framework for biological invasions. *Trends Ecol. Evol.* **26**, 333–339 (2011).
75. Linnaeus, C. *Species plantarum* (Salvius, 1753).
76. Holmes, B. The recently extinct plants and animals database, <https://recentlyextinctspecies.com/> (2020).
77. Stévert, T. et al. A third of the tropical African flora is potentially threatened with extinction. *Sci. Adv.* **5**, eaax9444 (2019).
78. Redding, D. W. & Mooers, A. Ø. Incorporating evolutionary measures into conservation prioritization. *Conserv. Biol.* **20**, 1670–1678 (2006).
79. Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).
80. Jarzyna, M. A. & Jetz, W. Taxonomic and functional diversity change is scale dependent. *Nat. Commun.* **9**, 2565 (2018).
81. Daru, B. H. et al. Endemism patterns are scale dependent. *Nat. Commun.* **11**, 2115 (2020).
82. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
83. Holt, B. G. et al. An update of Wallace's zoogeographic regions of the world. *Science* **339**, 74–78 (2013).
84. Koleff, P. et al. Measuring beta diversity for presence–absence data. *J. Anim. Ecol.* **72**, 367–382 (2003).
85. Kreft, H. & Jetz, W. A framework for delineating biogeographical regions based on species distributions. *J. Biogeogr.* **37**, 2029–2053 (2010).
86. Carver, S. J. & Fritz, S. *Mapping wilderness: concepts, techniques and applications* (Springer, Dordrecht, 2016).
87. Hijmans, R. J. raster: geographic data analysis and modeling. R package version 3.0-12. <https://CRAN.R-project.org/package=raster> (2020).
88. Pinheiro, J. et al. nlme: linear and nonlinear mixed effects models. R package version 3.1-145 <https://CRAN.R-project.org/package=nlme> (2020).

Acknowledgements

We thank Texas A&M University–Corpus Christi and the Harvard University Herbaria for logistic and financial support. Funding: B.H.D. was supported by the U.S. National Science Foundation (awards 2031928, and 2113424) and Texas Parks and Wildlife (award F21AF03049-00), C.C.D. was supported by the U.S. National Science Foundation (awards 1208835, 1802209, and 1754584), M.Z. and M.P. were supported by the University of Tartu (PLTOM20903), Estonian Research Council (PRG609, PRG1065), and by the European Union through the European Regional Development Fund (Centre of Excellence EcolChange).

Author contributions

The study was conceived and designed by B.H.D. with early development and refinement by C.C.D. All analyses were carried out by B.H.D. The manuscript was written by B.H.D. with significant contributions from T.J.D. and C.C.D. The manuscript was revised by B.H.D., T.J.D., C.G.W., E.K.M., A.R., M.Z., M.P., A.A. and C.C.D. Final approval of the submitted version: B.H.D., T.J.D., C.G.W., E.K.M., A.R., M.Z., M.P., A.A. and C.C.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27186-8>.

Correspondence and requests for materials should be addressed to Barnabas H. Daru, T. Jonathan Davies or Charles C. Davis.

Peer review information *Nature Communications* thanks Duccio Rocchini and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021