

Regional uniqueness of tree species composition and response to forest loss and climate change

Received: 16 June 2023

Accepted: 26 April 2024

Published online: 31 May 2024

 Check for updates

Nina van Tiel ^{1,2} ✉, Fabian Fopp ^{3,4}, Philipp Brun ⁴, Johan van den Hoogen¹, Dirk Nikolaus Karger ⁵, Cecilia M. Casadei ^{6,7}, Lisha Lyu^{3,4}, Devis Tuia ², Niklaus E. Zimmermann ⁴, Thomas W. Crowther ^{1,8} & Loïc Pellissier ^{3,4,8}

The conservation and restoration of forest ecosystems require detailed knowledge of the native plant compositions. Here, we map global forest tree composition and assess the impacts of historical forest cover loss and climate change on trees. The global occupancy of 10,590 tree species reveals complex taxonomic and phylogenetic gradients determining a local signature of tree lineage assembly. Species occupancy analyses indicate that historical forest loss has significantly restricted the potential suitable range of tree species in all forest biomes. Nevertheless, tropical moist and boreal forest biomes display the lowest level of range restriction and harbor extremely large ranged tree species, albeit with a stark contrast in richness and composition. Climate change simulations indicate that forest biomes are projected to differ in their response to climate change, with the highest predicted species loss in tropical dry and Mediterranean ecoregions. Our findings highlight the need for preserving the remaining large forest biomes while regenerating degraded forests in a way that provides resilience against climate change.

The UN Decade on Ecosystem restoration has begun to catalyze interest in nature restoration, from the conservation and management of existing forests to the rejuvenation of ecosystems on degraded lands^{1,2}. Sustainable management offers an opportunity to address biodiversity loss and climate change, and at the same time enhance human well-being across the globe³. A key component in this global effort is the conservation and restoration of forests, which represent the largest repositories of biodiversity and carbon on land^{4,5} and are associated with essential ecosystem services⁶. Earth is home to ~73,000 tree species⁷ and understanding the spatial distribution of these species is a challenge to be tackled to guide regional to global

management policies. Conservation and restoration efforts should be guided by solid information on species' ecological preferences, taxonomic and phylogenetic diversity, and ecosystem resistance to climate change⁸.

The latest assessment indicates that at least 31% of all tree species are threatened with extinction globally⁹. This loss of biodiversity has far-reaching consequences¹⁰, and halting and reverting it requires the identification of priority conservation areas to target regions of greatest threat. The selection of such regions may consider different facets of biodiversity, such as taxonomic, functional, and phylogenetic diversity, in an effort to prevent the extinction of species, as well as

¹Global Ecosystem Ecology, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland. ²Environmental Computational Science and Earth Observation Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. ³Ecosystems and Landscape Evolution, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland. ⁴Land Change Science Research Unit, Swiss Federal Institute for Forest, Snow and Landscape Research, WSL Birmensdorf, Switzerland. ⁵Biodiversity and Conservation Biology, Swiss Federal Institute for Forest, Snow and Landscape Research, WSL Birmensdorf, Switzerland. ⁶Laboratory of Biomolecular Research, Biology and Chemistry Division, Paul Scherrer Institute, PSI Villigen, Switzerland. ⁷Institute of Molecular Biology and Biophysics, Department of Biology, ETH Zürich, Zürich, Switzerland. ⁸These authors jointly supervised this work: Thomas W. Crowther, Loïc Pellissier. ✉e-mail: nina.vantiel@epfl.ch

unique branches of evolutionary history^{11,12}. The strongest gradients in tree species composition are often observed at the boundary between major plant kingdoms following deep historical splits¹³, or between juxtaposed biomes, such as between tropical and temperate forests¹⁴, associated with rapid changes in climate, topography, and other environmental conditions¹⁵. These major shifts in tree lineage composition are complemented by gradients of community composition within biomes, influenced by a complex interplay of abiotic factors at regional scales, such as variations in freezing frequencies or moisture¹⁶. A better understanding of these gradients is not only important to comprehend the conditions sustaining biodiversity hotspots but also for conserving and restoring forest ecosystems and ensuring their continued contribution to biodiversity and ecosystem services. A few studies have investigated tree species composition at the global scale, using only the location of observations to construct species' ranges¹⁷ or to compute species richness¹⁸. Basing such analyses on habitat suitability maps may allow a better estimate of diversity in regions with limited observational data. However, suitability-based investigations of tree species composition have so far been limited to a relatively small number of tree species or local to regional scales^{19,20}. A global high-resolution mapping effort would offer a picture of the main gradients of taxonomic and phylogenetic composition in the forests from local to global scales.

Establishing effective conservation targets requires a sound understanding of the historical decline of species ranges associated with habitat loss²¹. For instance, Hubbell et al. used neutral theory to estimate how the decline of forest cover could have impacted tree species diversity²². However, the distribution of tree species is regulated by non-neutral ecological determinants²³, and the consequence of habitat loss could be informed by estimated species occupancy²⁴. In particular, species potential ranges can be matched with terrestrial land cover to map the species' area of habitat (AOH, also known as the extent of suitable habitat), which is considered a criterion for establishing species ranking on the IUCN red list²⁵. However, in contrast to vertebrates²⁶, the estimated species ranges for trees are limited to selected families²⁷ or available at coarse resolution only¹⁷, which limits assessments of the effect of habitat loss on tree diversity at the regional scale. To overcome this challenge, machine learning can support the mapping of species' habitat suitability with high predictive power²⁸. Then, combining these maps with data on forest cover estimated from remote sensing^{29,30} can provide insights into historical range loss and, therefore, support management decisions for conservation³¹.

The challenges of forest management are significantly compounded by the rapid changes in climate³². As climatic conditions change, the potential distribution of plant species will also shift^{33,34}, leading to changes in species composition¹⁴ and a potential decline in provided ecosystem services³⁵. Moreover, the response of species to those changes can largely vary across regions, where some might suffer more extinctions, while others might be able to track climate change along latitude or elevation³⁶. Thus, climate change-resilient forest and ecosystem management require an understanding of which species have the potential to thrive under future climatic conditions. However, predicting the potential distribution of tree species under changing climate conditions is a major challenge as model predictions incorporate various limitations and assumptions^{37,38}. Nevertheless, estimating future suitable ranges with statistical models remains a scalable approach to gain insights into the effects of climate change and inform conservation and restoration efforts.

The objective of this study is twofold: first, we estimated the spatial distribution of a large number of tree species at the global level at a 30-arc second resolution. We developed a cloud-based implementation of an environmental niche modeling pipeline with geographic dispersal constraints and over 26 million recorded observations of trees from a broad compilation of databases (Table 1)

to generate range maps for 10,590 tree species for which we had at least 90 occurrence records, which was found to be sufficient to train models with good predictive performance. Subsequently, and beyond the pure large-scale mapping effort, we used the predicted distributions to perform three analyses investigating the global distribution of tree species. First, we inferred global gradients of species composition and investigated the global variation of taxonomic and phylogenetic forest composition. Second, by combining potential species occupancy with tree cover data, we computed the reduction in species ranges when constrained to current forest cover, therefore assessing the impact of historical forest loss on species range reduction. Third, using climate change simulations, we investigated the extent of species turnover and shifts in latitude and elevation under projected climate change. This study provides a comprehensive overview of the global forest composition, the characterization of species' range sizes, and the resilience of tree biodiversity to climate change. Thanks to large publicly available occurrence databases and our cloud-based implementation, we are able to reach results of unprecedented scale for habitat suitability mapping, both in terms of spatial coverage and resolution and in terms of the number of species, allowing us to investigate the global composition of tree species at very high granularity in our subsequent analyses.

Results

Model performance

We applied a model combining geographic range polygons and species distribution modeling to map the ecological suitability of tree species, based on climatic and pedologic variables, at a 30-arc second resolution. Observational data was collated from 13 databases (Table 1) and aggregated to the same resolution as the model covariates. Geographic range polygons were constructed, using reported native countries and the location of occurrences, as coarse estimates of species' native ranges. They were used to filter observation records and to constrain the extents of the projected ranges. We modeled the distributions of 24,140 tree species with at least 20 distinct observations after aggregation, and with reported native countries available from GlobalTreeSearch³⁹. We assessed the performance of the models and found our modeling approach to obtain sufficient predictive performance for species with at least 90 spatially explicit observations (Supplementary Fig. 1). In total, 10,590 tree species met this criterion. When evaluated by 3-fold cross-validation, the models achieved an average true skill statistic (TSS) of 0.77 ± 0.11 and an average area under the ROC curve (AUC) of 0.93 ± 0.04 (Supplementary Fig. 2). The evaluation of the models on an independent presence-absence dataset from sPlot⁴⁰ yielded lower TSS (0.53 ± 0.30) with a high recall and low precision (Supplementary Fig. 3). This points towards a low binarization threshold, which is expected for low-prevalence species when optimizing the threshold on maximum-TSS. With an average false positive rate of 29.5%, false positive errors were the most common than false negative errors which had an average rate of 17.3%. False positive errors, where the model classified a location to be suitable for a species although no occurrence was recorded there, are not necessarily problematic as the species may be absent due to factors that were not considered by the model, such as biotic interactions, human influence, or stochasticity. Moreover, the comparison of our modeled distributions to maximum habitat suitability maps for 23 European species for which confident estimations were available in a published report⁴¹ showed some considerable differences with an average intersection over union (IoU) of 0.42 (Supplementary Fig. 4). However, our models were generally more conservative with ranges on average 36% smaller than those to which they were compared. Unlike previous large-scale modeling efforts⁴², our approach combined both species distribution modeling and geographic range limits, which can, for instance, represent dispersal limitations⁴³ and consider reported native ranges.

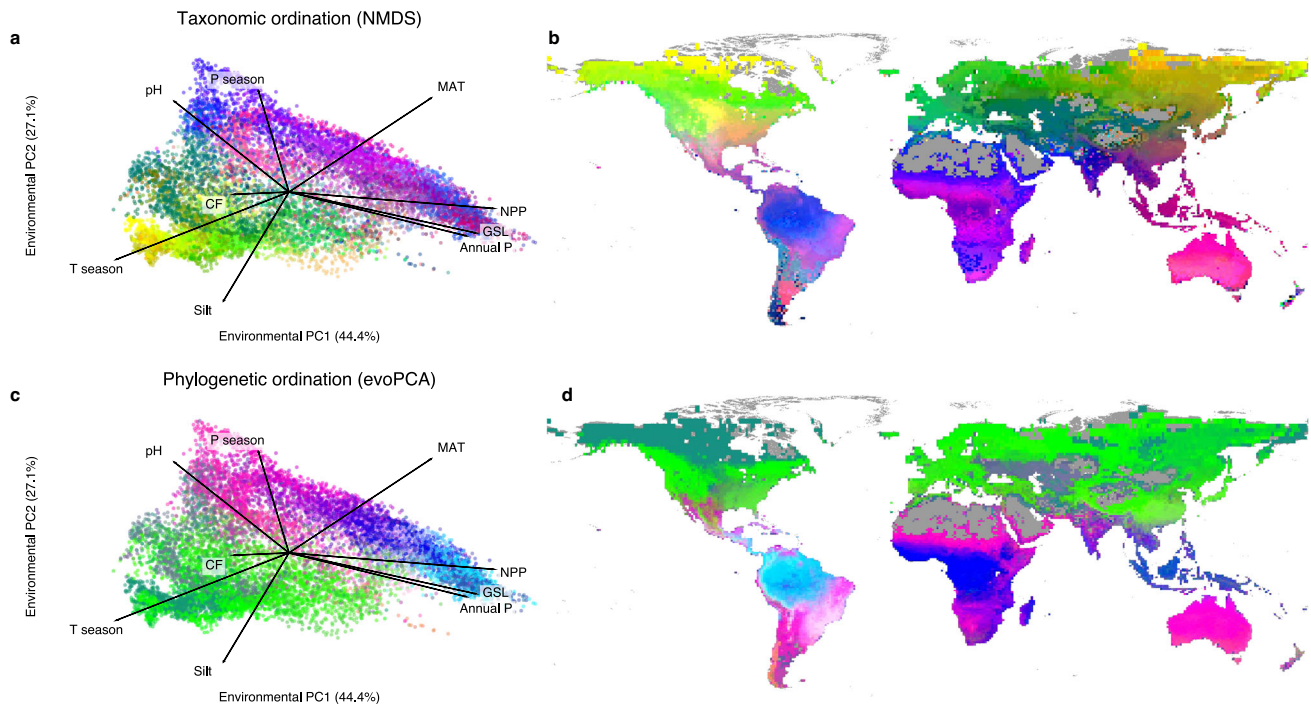


Fig. 1 | Gradients in taxonomic and phylogenetic composition show a near-unique biodiversity signature of every single location on the planet. Taxonomic composition is represented by a 3-axis non-metric dimensional scaling (NMDS) and phylogenetic beta-diversity is represented by the 3 first axes of a phylogenetic ordination (evoPCA). Both the taxonomic and phylogenetic ordinations are computed on the global community matrix derived from the modeled distributions of $n = 10,590$ tree species sampled at a resolution of 100 km, resulting in $n = 12,548$ sites. The 3 axes of each ordination are mapped to red, green, and blue with minimum and maximum values corresponding to the 10th and 90th percentiles.

a, c. Scatter plot of taxonomic and phylogenetic ordinations in environmental space, a 2-dimensional space made up of the 2 first axes of a PCA of the environmental variables used for species distribution modeling: mean annual temperature (MAT), temperature seasonality (T season), annual precipitation (Annual P), precipitation seasonality (P season), growing season length (GSL), net primary productivity (NPP), silt content (Silt), coarse fragments (CF), and soil pH (pH). **b, d.** Map of taxonomic and phylogenetic ordinations in geographical space. Source data are provided as a Source Data file. The maps were created with QGIS¹⁰ and the gray base map corresponds to all areas for which model predictors were available.

Unique species composition across the globe

Our models show that global tree taxonomic and phylogenetic composition is organized in complex biogeographic and environmental gradients worldwide. By computing non-metric multidimensional scaling (NMDS) on the community matrix based on our predicted distributions, sampled at using an equal area projection at a resolution of 100 km, we can visualize an estimate of tree species composition globally (Fig. 1a, b and Supplementary Fig. 5a–c). Combining the community matrix with a phylogeny of the 10,590 species considered⁴⁴, we computed a phylogenetic ordination^{45,46} thus obtaining an estimate of phylogenetic beta-diversity globally (Fig. 1c, d and Supplementary Fig. 5d–f). Taxonomic and phylogenetic composition show marked spatial gradients, where any given locations displays a unique signature in taxonomic and phylogenetic composition. Species composition also varies in environmental space, with similar taxonomic or phylogenetic compositions often found under similar environmental conditions, although we also observe dissimilar compositions under similar environmental conditions, such as across tropical moist forests.

Taxonomic turnover is described by smooth distributions both in geographical and ordination space (Supplementary Fig. 5c). The lack of structure in the underlying distributions can be characterized by the difficulty of finding an adequate clustering: we found that two clusters obtained the best score, albeit with high intra-cluster variance and massive cluster size (Supplementary Fig. 6a, b). Phylogenetic turnover showed clearer spatial boundaries and more peaky distributions (Supplementary Fig. 5f). One of the most striking peaks of these distributions corresponds to a very homogeneous phylogenetic composition across Northern Canada and Alaska which also corresponds to

one of the six clusters formed by the phylogenetic ordination (Supplementary Fig. 6c, d). Redundancy analysis indicates that gradients in phylogenetic beta-diversity are associated with shifts in climatic and edaphic factors, with 57.6% variance explained by the nine variables used for niche modeling. In contrast, the variation in taxonomic diversity was not well captured by the same abiotic variables, with only 11.8% variance explained. For both ordinations, variation partitioning analysis indicated that climatic variables had stronger effect than edaphic variables (Supplementary Fig. 7). This may be associated to scale-dependency or to the soil data being generated by models reliant on climate-related variables⁴⁷. Together, the lack of structured distributions and of fine-grained clusters in these ordinations point towards a near-unique biodiversity signature, as any given site is associated with a unique set of tree species and lineages resulting from a combination of historical and ecological factors.

Potential species occupancy restricted by tree cover

Combining remotely sensed tree cover and the potential habitat suitability of species described above, we documented that the suitable occupation area, considering ecological preference together with biogeographic and dispersal limitations, is significantly restricted by forest cover loss. We computed species' range sizes based on their modeled distributions, and range sizes of the distributions constrained to forested areas, defined as areas with at least 10% tree cover as estimated using remote sensing data from the year 2000²⁹. The relative decrease in range size reflects how species ranges are restricted by historical forest cover loss and we find a median range reduction of 22.0% across all considered species. All biomes showed significantly lower realized ranges compared to potential ranges of species

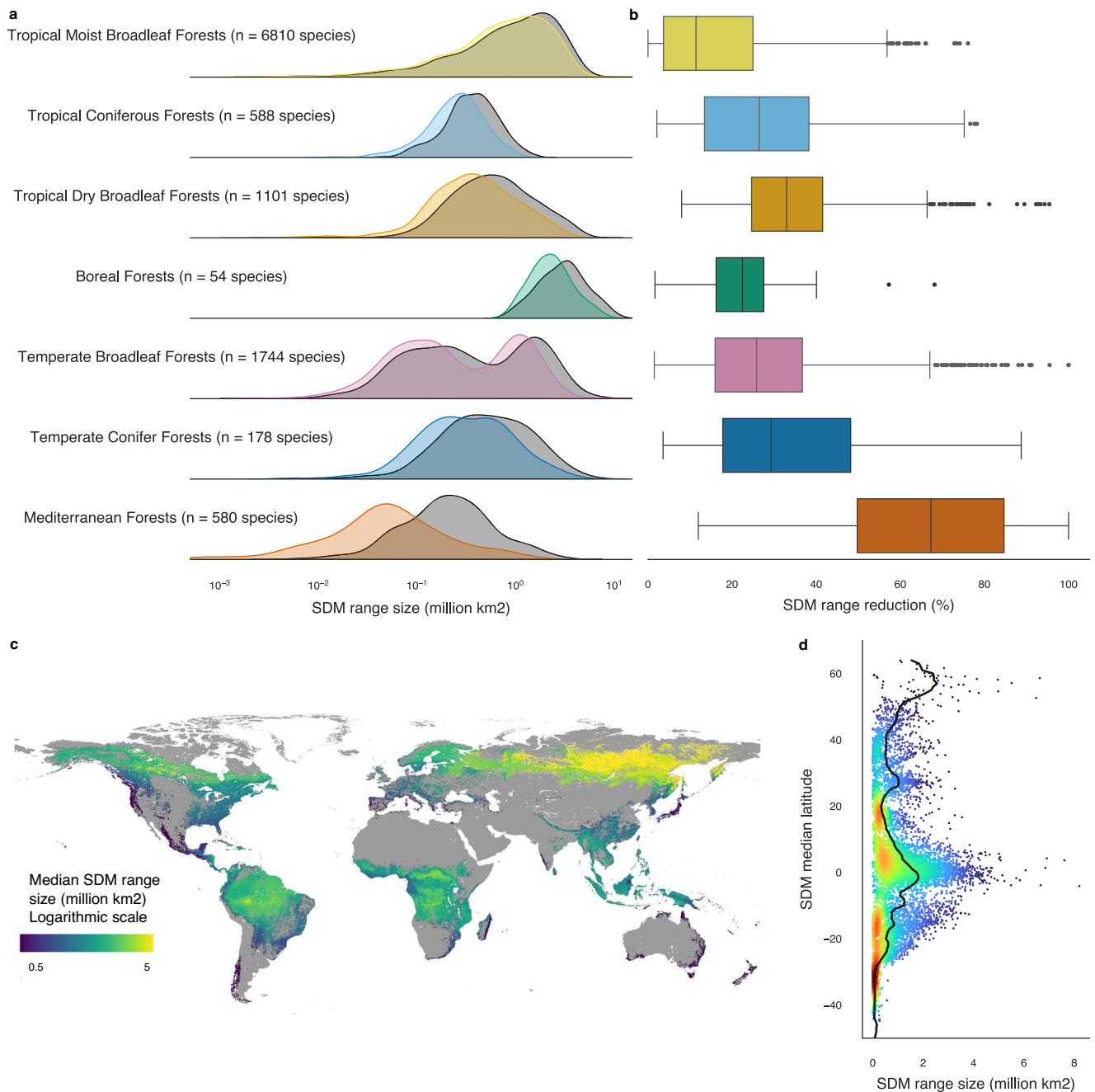


Fig. 2 | Species occupancy range distribution and loss. **a** Distributions of species occupancy range sizes globally (gray) and constrained to forests (at least 10% tree cover, color) for species in each forest biome. **b** Boxplot of relative range reduction across species in each forest biome with the center line showing the median, the box limits the quartiles, the whiskers 1.5 times the interquartile range, and the points the outliers. The distributions and boxplots are computed for $n = 6810$ species for Tropical Moist Broadleaf Forests, $n = 588$ species for Tropical Coniferous Forests, $n = 1101$ species for Tropical Dry Broadleaf Forests, $n = 54$ species for Boreal Forests, $n = 1744$ species for Temperate Broadleaf Forests, $n = 178$ species for Temperate Conifer Forests and $n = 580$ species for Mediterranean Forests. **c** Global map of median species range size constrained to forests, created with QGIS¹⁰. The gray base map corresponds to all areas for which model predictors were available. **d** Plot of species' median latitude against range size constrained to forests, colored by point density, where red indicates the highest density. Source data are provided as a Source Data file.

($p < 0.01$ for t-test between unrestricted and restricted range sizes over all species and per biome) with median range reduction per biome varying between 11.4% and 67.2%.

This decline was especially pronounced in the tropical dry broadleaf forests and temperate conifer forest biomes (33.0% and 29.3% median range reduction, respectively), and even more in the Mediterranean biomes with a 67.2% median range reduction (Fig. 2b). Despite increasing forest loss in the last two decades²⁹, some biomes had so far limited decline and those biomes still host species with

uniquely broad ranges, typically at high latitudes in the boreal forests, but also around the equator in the Amazon, Congo basin and South East Asian forests (Fig. 2c and Supplementary Fig. 8). While all species found in boreal forests were predicted to have large ranges, the ranges of species in tropical moist forests have a large spread in range size, spanning three orders of magnitude. For instance, the species with the smallest (*Palaquium sericeum*) and the largest (*Terminalia grandis*) range sizes with no tree cover restriction are found in the tropical moist broadleaf biome.

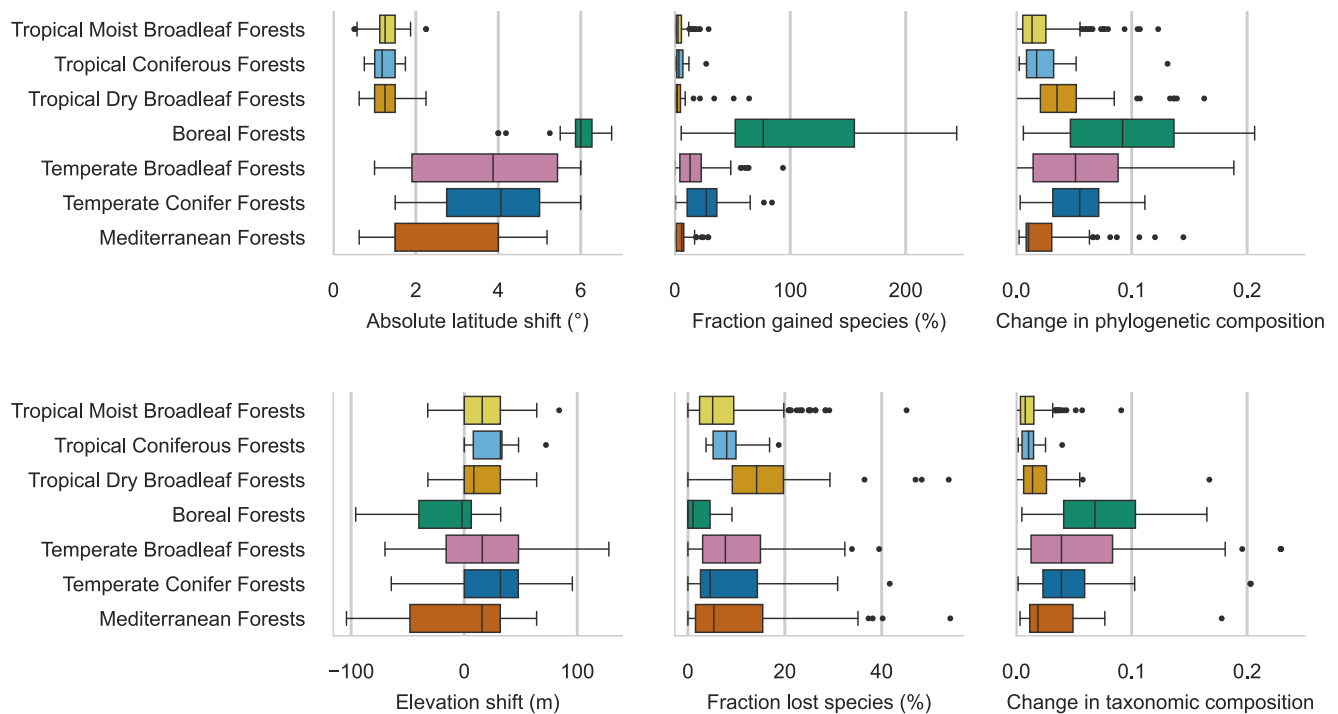


Fig. 3 | Response of tree species to climate change across biomes. The median absolute latitude and median elevation shift among species, fraction of lost and gained species, and change in taxonomic and phylogenetic composition under climate change were computed for each forest ecoregion. The boxplots show statistics for $n = 239$ ecoregions for Tropical Moist Broadleaf Forests, $n = 14$ ecoregions for Tropical Coniferous Forests, $n = 55$ ecoregions for Tropical Dry Broadleaf Forests, $n = 26$ ecoregions for Boreal Forests, $n = 91$ ecoregions for Temperate Broadleaf Forests, $n = 49$ ecoregions for Temperate Conifer Forests and $n = 61$

ecoregions for Mediterranean Forests. The center line of the boxplots shows the median, the box limits the quartiles, the whiskers 1.5 times the interquartile range, and the points the outliers. Changes are computed between predicted distributions with climate variables for 1981–2010 and climate projections for 2071–2100 under climate change scenario SSP 5.85. Changes in composition are computed as the Euclidean distance between scaled NMDS and evoPCA values computed at the ecoregion level. Source data are provided as a Source Data file.

Response of tree species to climate change across biomes

We compared species distributions mapped with climatic variables averaged over 1981–2010 with distributions with projected future climate for 2071–2100 with emissions corresponding to the shared socioeconomic pathway (SSP) 5.85. Predicting future species distributions has many limitations³⁷, yet it presents a scalable approach to grasp how climate change may affect species distributions in the future. Acknowledging that there are many uncertainties at the species level, we investigated trends at the scale of regional ecosystems, using the RESOLVE ecoregions⁴⁸. The results for each of the 497 forest ecoregions were then aggregated to their corresponding biome (Fig. 3). Our results predict that ecoregions in different biomes will show significant diverging behavior in their response to climate change (Pillai's Trace test, $p < 0.001$).

Species distributions in all forest ecoregions were predicted to undergo latitudinal shifts, with the smallest shifts found in tropical forest ecoregions, with a median of 1.25°, and particularly pronounced shifts in boreal forests with a median of 6.00°. Elevation shift presented a larger variation across ecoregions, with a coefficient of variation of 3.07, compared to 0.73 for shifts in absolute latitude, pointing towards a higher variability of available habitat in elevation than in latitude. We find relatively small median elevation shifts in most forest ecoregions, with an absolute median shift of less than five meters in about 24% of the considered ecoregions. Temperate and tropical coniferous forest biomes presented the highest positive median elevation shifts across ecoregions, with medians of 32 meters across ecoregions in both biomes. However, some ecoregions were estimated to have large downward median elevation shifts, in particular in the boreal forest biome with a median elevation shift of –1.95 meters. While most species are expected to shift their ranges upwards,

downward elevation shifts have been reported in multiple studies^{49,50} and may reflect both newly available suitable habitats and changes in precipitation regimes⁵¹.

Our results indicate that, in boreal forest ecoregions, many new species may find suitable habitats and few species may lose suitable habitats under future climate change. This is coherent with our predictions of larger shifts in latitude or elevation allowing species to track climate change thanks to continuously available habitat and is estimated to have a strong effect on the phylogenetic composition of these ecoregions. On the other hand, tropical dry and Mediterranean forests, with generally much more geographically restricted species' ranges, were found to suffer the highest proportion of extirpated species, with some ecoregions predicted to lose suitable habitat for over 50% of currently suitable species. In Mediterranean ecoregions, the turnover of species is estimated to affect taxonomic composition more strongly than the phylogenetic composition. Overall, more modest changes in composition were found in tropical forest ecoregions than in temperate and boreal forest ecoregions.

Discussion

Understanding tree species distribution is central to the conservation, restoration, and management of global forest biodiversity. Here, we provide species occupancy estimations for over 10,000 tree species, beyond regions where species range maps are already available and broadly used by foresters to manage forest resources⁵². Our analysis reveals the high uniqueness of local lineage composition, where any region across the globe is characterized by a unique tree composition signature associated partially with ecological factors. Residual variance may be attributed to regional and historical factors⁵³, such as plate tectonics and long-distance dispersal⁵⁴. By restricting species' suitable

habitat to forests with tree cover estimated from remote sensing²⁹, realized tree species distribution is significantly narrower than its potential, even when considering geographic range limits. Considering predicted species distributions under future climate projections shows that different areas will be affected differently by climate change and stresses the importance of climate-smart conservation efforts that are adapted to the specific region.

Modeling species distributions for such a large number of species was facilitated by the use of Google Earth Engine, a cloud-based platform for geospatial analysis, and its available implementation of machine learning algorithms⁵⁵. Additionally, models with adequate predictive performance can only be constructed if enough occurrences are available for a given species. This is made increasingly possible by the availability of large databases and citizen science efforts⁵⁶. Such global databases typically display important spatial bias^{57,58}, although efforts to concatenate different databases, as done here, may mitigate the acquisition bias from a single dataset. Nevertheless, to ensure satisfactory predictive performance, we excluded many rare and under-sampled species and our results are likely to present biases towards common and large-ranged species. Future work may consider the recent application of deep learning to species distributions modeling, which allows the joint estimation of the distributions of thousands of species with a single model and has been shown to obtain more reliable results for species with very few occurrence records^{59–61}. Moreover, the availability of global maps of environmental variables allowed us to make continent-wide predictions of species distributions. While we selected variables that are of general ecological relevance across all biomes, future work may consider more specific factors, such as the fire frequency and severity. The cloud-based data integration and modeling approach demonstrate the potential of high-performance computing infrastructure to support the documentation of global biodiversity patterns at scales and resolutions unattained before, therefore allowing fine-grained, global analyses such as those presented here.

With the distributions of thousands of tree species, we showed a striking uniqueness of local tree species and lineage composition with both large- and small-scale turnover. The phylogenetic beta-diversity showed a spatial organization matching known phylogeographic boundaries, such as the split between Gondwanan and Laurasian biotas¹³, or the clustering of tropical forests^{54,62,63}. In contrast, some regions displayed phylogenetic similarities despite the large geographic distance between them, which reflects historical connectivity. For instance, the phylogenetic similarity of the Indo-Pacific region may highlight the signal of its past connection before their segregation caused by plate tectonics⁶², or long-distance dispersal⁵⁴. The similarity across Palearctic and Nearctic temperate and boreal forests is likely to indicate the long history of connectivity between these regions⁶⁴. Within these larger biogeographic boundaries, our results further highlighted the role of environmental gradients in shaping tree lineage turnover, such as the compositional divide in tree assemblages in the Americas between tropical and extratropical environments, which is related to temperature¹⁶ and moisture gradients⁶⁵. Taxonomic composition showed more smooth gradients and was only weakly associated with environmental variables, suggesting a strong effect of historical factors, such as geographic range limits. Focusing on the most common tree species, our analyses uncovered marked local gradients in species and lineage composition. However, the global composition of tree species and the corresponding phylogenetic diversity may be even more strongly geographically structured when including rare species.

Beyond the current habitat suitability of local species, it is important for conservation and restoration efforts to take into account the past and the future with estimates about both the historical decline of species' ranges and to understand the spatial variation in climatic risks to forests to guide climate-smart ecosystem management^{66,67}. We

derived estimates of these effects from our modeled species distributions. However, it is important to acknowledge the limitations associated with our approach. To compute historical species' range reduction, we compare species' potential range size with their ranges constrained to current forests. Yet, computing potential ranges based only on ecological suitability and biogeographic limitations may lead to an overestimation of range size, as competition with other vegetation types is likely to restrict realized species distributions. Moreover, since trees may also be present outside of forests^{68,69}, constraining ranges to areas with at least 10% tree cover is likely to underestimate the realized range of some species. Nevertheless, the extent of the discrepancy between potential and actual distributions remains indicative of the extent of historical forest loss due to land use change⁷⁰. To predict future species distributions under climate change, we use projections of several climate models used by the IPCC (CMIP6) which are associated with substantial spatial and temporal uncertainty⁷¹ and do not directly account for extreme events such as floods, fires, droughts or floods which can cause large-scale forest die-backs⁷². Furthermore, our models present added uncertainties associated with limitations due to their correlative nature and the assumption that the relationships of species with environmental covariates remain constant through time. Although mechanistic models may present a better approach to predict range shifts under climate change^{73,74}, they are challenging to implement and require data that is not as widely available, as well as the identification of key limiting processes³⁸. Therefore, correlative models remain the most feasible option to predict the effects of climate change for a large number of species and at a global scale. Finally, the bias of our results towards well-sampled species should be carefully considered. The exclusion of species for which not enough occurrence records were available is likely to have led to an underestimation of the number of species with small ranges, in particular in tropical biomes due to sampling biases. Rare and small-ranged species may also be particularly vulnerable to change and our results regarding the response of ecosystems to climate change may be affected by their omission.

Despite these limitations, our results may serve as a guide for forest stakeholders, offering a global perspective that can be combined with more local observations and expert knowledge to conserve the preserved area and to decide which tree species to consider for restoration⁷⁵. For instance, in Mediterranean ecoregions, known for a long history of human land use and pastoralism⁷⁶, we find by far the strongest range reduction, leading to small realized species ranges and high predicted extirpation rates under climate change, highlighting the need for climate-smart restoration in these forests. In contrast, in accordance with a remote sensing-based estimation⁷⁷, we found that the tropical moist broadleaf forests had the greatest intactness. Specifically, Amazonia, the Congo Basin, and South East Asia are home to not only the highest species richness (Supplementary Fig. 9) but also some of the species with the largest ranges on Earth, alongside many small-ranged species. Our results relate to the finding of the hyperdominance of some species in the Amazon basin⁷⁸, suggesting that both extreme ecological generalism and specialism are viable strategies in tropical environments. However, small ranges may also reflect non-equilibrium dynamics rather than narrow ecological niches⁷⁹. Outside the tropics, the boreal forest shows the broadest ranges of species with the lowest decline associated with habitat loss. Moreover, the climate of boreal forests is predicted to become suitable for many species that are currently not found there.

Our results underline the uniqueness of species composition across the globe, and also the diversity of species' historical range reduction and response to climate change, reflecting that climate change displays distinct intensities and effects globally⁸⁰. While local forest protection efforts should be made everywhere, international efforts for forest conservation strategies should primarily focus on

Table 1 | Online databases used as data sources for occurrence data

Database name	URL
Botanical Information and Ecology Network (BIEN) ¹⁰⁴	https://bien.nceas.ucsb.edu/
BIOMASS ¹⁰⁵	https://www.nature.com/articles/sdata201770#data-records
Caudullo et al. 2017 ⁵²	https://doi.org/10.1016/j.dib.2017.05.007
CONIFER	https://herbaria.plants.ox.ac.uk/bol/conifers
DRYFLOR ¹⁰⁶	http://www.dryflor.info/data/datadownload108
GBIF ¹⁰⁷	https://www.gbif.org/occurrence/download/0032444-200221144449610
GFBI ¹⁸	https://www.gfbinitiative.org/
IDIGBIO	https://www.idigbio.org/portal/search
INDIABIODIVERSITY	https://indiabiodiversity.org/observation/list
MUSEUM	Field collection records and manually georeferenced herbarium data from Naturalis (L) and Paris (P) natural history museums. Unpublished data collected by J.S. Strijk.
PNG	http://www.pngplants.org/search.htm
PREDICTS ¹⁰⁸	https://data.nhm.ac.uk/dataset/the-2016-release-of-the-predicts-database
RAINBIO ¹⁰⁹	https://gdauby.github.io/rainbio/download_page.html

conserving the remaining most pristine extensive forest ecoregions. The strongest effort of restoration may be performed in temperate forests where historical losses were highest and recovery might bring the most direct benefits. Furthermore, tropical dry and Mediterranean regions may require the most acute conservation attention to avoid the extinction of a large fraction of species. Together, our findings emphasize the need for local restoration and conservation decisions that are tailored to the unique characteristics of each specific forest.

Methods

We used the python API for Google Earth Engine⁵⁵ to create a pipeline for high-resolution species distribution modeling. The distribution of each species is modeled separately. After data preparation, the pipeline consisted of two parts: (1) constructing the geographic range in which the species is considered, and (2) modeling ecological suitability with an ensemble of tree-based machine learning models, allowing us to map the species distribution using current climate covariates, as well as future climate projections. Further analyses were applied to the results of our models using the python API for Google Earth Engine v0.1.329⁵⁵ with python v3.8.13⁸¹ with helper packages: pandas v1.4.4⁸², numpy v1.23.4⁸³, matplotlib v3.5.3⁸⁴ and seaborn v0.12.1⁸⁵. Moreover, some analyses were performed using R v4.2.2⁸⁶ with helper packages: tidy v1.3.0⁸⁷, tidyverse v1.3.2⁸⁸, tibble v3.1.8⁸⁹, data.table v1.14.6⁹⁰, dplyr v1.1.0⁹¹, gridExtra v2.3⁹² and ggplot2 v3.4.1⁹³.

Data sources and cleaning

Tree species occurrence data were downloaded from 13 different online databases, published and unpublished datasets (Table 1). We kept only records with standardized tree species names, according to the global tree checklist GlobalTreeSearch³⁹ (v1.3, downloaded in January 2020). Then, records from all databases were merged and duplicates with identical coordinates for the same species were removed. We obtained almost 30 million (29,715,021) observations for 52,725 species. There were between 1 and 895,762 observations per species with a median of 196. The observations were then aggregated to the 30-arc second-pixel level to match the resolution of the predictor layers. Duplicate occurrences that fell within the same grid cell were removed and observations that fell off the pixel grid used for the predictors were removed (e.g. observations close to a coast that are aggregated to a pixel in the water). In subsequent steps, we consider the 24,140 species for which there were at least 20 spatially explicit observations after aggregation, which may be considered a reasonable minimum number of observations for modeling⁵⁸.

Geographic range polygon

For each species, we constructed a geographic range that was used to select observations to be included in the model training, thus excluding observations of non-native or invasive species, and for model projection, to prevent predictions outside the native range. The same range was used for model projection for all considered climate scenarios. The range for each species was constructed based on the reported native countries from GlobalTreeSearch³⁹ and the location of the observations.

The reported native range consisted of the geometries of the countries in which the species is listed as native with a 1000 km buffer to compensate for potential gaps in the database. The particularly large buffer size was also chosen to allow species to spread under projected future climates; however, this will consequently expand the range beyond the native countries. The geometries for country boundaries were obtained from the Global Administrative Unit Layers (GAUL) dataset, implemented by FAO. Particularly when the reported native range contained large countries, it was important to exclude the areas in which no observations were found. For this purpose, we intersected the reported native range with a polygon constructed around the ecoregions⁴⁸ in which there were observations that had at least 3 other observations within 1000 km. For small ecoregions (bounding box less than 1000 km in width or length), the ecoregion with a 1000 km buffer was included in the polygon. For larger ecoregions, a 1000 km buffer around the part of the ecoregion that is within 200 km of an observation was included in the polygon. For computational efficiency, if there were more than 10,000 observations, these were spatially aggregated for the construction of the range.

Training data

The training data consisted of the observations that fell within the considered range and pseudoabsences sampled uniformly at random within the considered range. For most species, the number of sampled pseudoabsences ($n_{PA,training}$) was 10 times the number of observations ($n_{obs,training}$). If there were less than 500 observations, we sampled 5000 pseudoabsences such that the environmental conditions of the region were well represented in the training data. The number of points in the training data was capped at 20,000 to ensure the model could run in Google Earth Engine without running out of memory. Therefore, if there were more than 1818 observations, we sampled fewer pseudoabsences, and if there were more than 10,000 observations we randomly sampled 10,000 observations and 10,000 pseudoabsences (Table 2).

Table 2 | The number of observations and pseudoabsences included in the training data ($n_{\text{obs,training}}$ and $n_{\text{PA,training}}$) dependent on the number of observations after aggregation (n_{obs})

n_{obs}	$n_{\text{obs,training}}$	$n_{\text{PA,training}}$	n_{total}
$n_{\text{obs}} \geq 10,000$	10,000	10,000	20,000
$1818 \leq n_{\text{obs}} < 10,000$	n_{obs}	$20,000 - n_{\text{obs}}$	20,000
$500 \leq n_{\text{obs}} < 1818$	n_{obs}	$n_{\text{obs}} * 10$	$5500 - 20,000$
$n_{\text{obs}} < 500$	n_{obs}	5000	$5000 - 5500$

Environmental variables and climate change scenarios

For the environmental niche model modeling, we selected nine environmental variables related to climate and soil conditions as predictive variables, based on preliminary variable importance analysis on a random subset of species using a random forests model. The climate variables consisted of average annual temperature, temperature seasonality, annual precipitation, precipitation seasonality, growing season length, and potential net primary production, and were obtained from CHELSA V2.1^{94,95}. These factors represent basic resource requirements, metabolic modifiers or disturbance constraints to plant growth and survival. The soil variables consisted of soil pH, coarse fragment content, and silt content, and were obtained from SoilGrids⁴⁷. We extracted all variables at a 30 arc-second resolution and projected them to the World Geodetic System 1984 (EPSG:4326) projection.

Species distribution modeling

For each species, we fitted an ensemble model consisting of two random forests and two gradient tree boost classifiers, with different complexity levels with regard to model formulation⁹⁶. The final distribution map was then achieved by taking the average over the predictions of the four individual models. We ran the model for 24,140 species for which at least 20 spatially explicit observations were available after aggregation.

To avoid overfitting when modeling rare species, we limited the number of predictor variables used in each model based on the number of observations for the corresponding species. We ensured that the number of observations available was at least 10 times the number of predictors used. For species that had less than 90 observations, the most important predictors for each species were selected based on the variable importance obtained when training a random forest model with all predictors. The number of selected predictors was determined by the number of occurrences in the training data (e.g., two predictors were selected for species with 20–29 occurrences). All nine predictors were used for species that had at least 90 observations.

The model output was probabilistic, and we used a 3-fold cross-validation with random fold assignment to determine the optimal threshold to transform the probabilistic output to binary, by assessing the threshold maximizing the true skill statistic (TSS). We evaluated our models by computing TSS, as well as the area under the ROC curve (AUC) during cross-validation.

Finally, the ensemble model was trained on the full training set and predictions were made on global maps of the covariates clipped to the considered range for each species. While soil variables were kept constant, we performed the predictions on historical averages of data from 1981–2010, referred to as current climate, and nine sets of variables for future climate scenarios: three time periods (2011–2040, 2041–2070, 2071–2100) and three different socio-economic pathways (SSP's), representing a sustainability scenario (SSP 1.26), a regional rivalry scenario (SSP 3.70), and a fossil fuel development scenario (SSP 5.85). We used five different global circulation models (GCMs) from the sixth coupled model intercomparison project (CMIP6): GFDL-ESM4,

UKESMI-0-LL, MPI-ESM1.2-HR, IPSL-CM6A-LR, and MRI-ESM2.0. They were bias-corrected⁹⁷ for the intersectoral-impact model inter-comparison project (ISIMIP)⁹⁸. For computational efficiency, the distributions were computed with a single model considering the average values over the five models, rather than taking the average of the predictions of five models each considering one GCM.

For further analyses, based on model evaluation through cross-validation (Supplementary Fig. 1), we considered only the 10,590 species for which at least 90 spatially explicit observations were included in the training data. The distributions were further evaluated on independent presence-absence data from sPlot⁴⁰. Data was available for 5939 of the 10,590 considered species and covered 47,479 plots across the globe. However, the comparison was computed only for the 3594 species that had at least 5 presences recorded in plots within the species' geographic range. For every species, a confusion matrix was constructed comparing the species' distribution model output to the survey data from plots within the species' geographic range, and the TSS, precision, and recall were computed. Finally, we compared our modeled distributions to maximum habitat suitability maps for 23 European species from the "Tree species distribution data and maps for Europe" report from the European Commission⁴¹ by computing their overlap with intersection over union (IoU).

Taxonomic and phylogenetic ordination

We extracted a global community matrix for the 10,590 considered tree species by extracting the predicted species distribution for the current climate at 100 km resolution using the equal area projection EPSG:6933⁹⁹, obtaining 12,548 sites. The distributions had to be sampled at a relatively low resolution due to computational feasibility of the ordinations. Distances between sites were computed with the Sorenson distance using the `vegdist()` function from the `vegan` R package¹⁰⁰ (2.6.4). The taxonomic ordination was computed as a 3-axis non-metric dimensionality scaling (NMDS) with the `metaMDS()` function with default parameters from the `vegan` R package¹⁰⁰ (2.6.4). We obtained a satisfactory stress value of 0.04. Each axis was mapped to the red, green, and blue values for visualizations. The phylogenetic ordination was computed with the `evopcahellinger()` function in the `adiv` R package⁴⁶ (2.2) using the same community matrix and a phylogenetic tree constructed with the `phylo.maker()` function in the `V.PhyloMaker` R package⁴⁴ (0.1.0). The first three axes of the ordination were retained. They explain 26.4%, 10.6% and 6.5% of the variation, respectively, and were also mapped to the red, green, and blue values.

The clustering analysis was performed using the `sklearn.cluster.Kmeans()` function and evaluated with the `sklearn.metrics.silhouette_score()` function from the `scikit-learn` python package¹⁰¹ (v1.1.3). The 3-axes of each ordination were clustered together after the removal of outliers, defined as points that fall outside of the range $[Q1-IQR, Q3+IQR]$ where $Q1$ and $Q3$ are the first and third quartiles and IQR is the interquartile range (ie. $Q3-Q1$). Maps of the clustered ordinations were created with the number of clusters that generated the highest silhouette score. Redundancy analyses were performed on the outputs from both ordinations with the `rda()` function and variation partitioning analyses with the `varpart()` function, both from the `vegan` R package¹⁰⁰ (2.6.4).

A principal component analysis was applied to the environmental variables extracted at a 100 km resolution, to match the sites of the ordinations. We used the `sklearn.decomposition.PCA()` function from the `scikit-learn` python package¹⁰¹ (v1.1.3). The two first principal components were used to represent the taxonomic and phylogenetic ordination results in a 2-dimensional representation of the environmental space.

Occupancy range size

We obtained the species range sizes by computing the surface area covered by the estimated species distribution for the current climate

variables (1981–2010). Range sizes were computed for the full distributions and for the distributions constrained to forested areas, defined as areas with at least 10% tree cover as estimated using remote sensing data from the year 2000²⁹. Although maps of tree cover corresponding to more recent years are available, we used the one from the year 2000 because it corresponds roughly to the period represented by the climatic data. Biome-level statistics were computed using the forest biomes from the RESOLVE classification⁴⁸ with species assigned to a biome if at least 20% of their estimated distribution fell within that biome. The global map of median range size was obtained by computing the median of the range sizes for all species that were predicted to be suitable in each 30-arc-second pixel.

Species response to climate change

To assess the response of tree species distributions to climate change, we computed differences between the predicted distribution maps with climatic variables for 1981–2010 and for 2041–2070 with SSP 5.85. The most distant and extreme climate change scenario was selected such that differences would be maximized. However, as species-level predictions in the future have many uncertainties, trends were investigated instead on the ecoregion-level, using the ecoregions from forest biomes using the RESOLVE classification⁴⁸. The fraction of species predicted to be gained and lost under climate change was computed for each ecoregion using the species predicted to be found within each ecoregion's boundaries for both sets of climatic variables. The median elevation shift and absolute latitude shift were computed for each ecoregion over the species predicted to be present in an ecoregion with both current and future climatic variables, by taking the shift in median latitude and elevation between current and future distributions for each species. Taxonomic and phylogenetic ordinations were computed on the ecoregion-level community matrix, considering both current and future compositions simultaneously. The same functions as for the global taxonomic and phylogenetic composition analysis were used. We obtained a stress value of 0.08 for the 3-dimensional NMDS, and the three first axes of the phylogenetic PCA explained 40%, 15%, and 7% of the variation. The change composition was computed as the Euclidean distance between scaled ordination values. Biome-level means and 95% confidence intervals were computed over the ecoregions assigned to each biome according to the RESOLVE ecoregion and biome classification⁴⁸.

Reporting summary

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

Data availability

The occurrence data used in this study are available from the online databases which are listed in Table 1. One unpublished dataset was used; it is available from the corresponding author upon request. The bioclimatic raster data used as used model covariates in this study are available from CHELSA 2.1^{94,95}. The edaphic raster data used as used model covariates in this study are available from Soilgrids⁴⁷. The reported native ranges data used in this study are available from GlobalTreeSearch³⁹. The country boundary data from the FAO used in this study are available through the Google Earthengine data catalog (https://developers.google.com/earth-engine/datasets/catalog/FAO_GAUL_2015_Level0). Plot data used as an independent validation data in this study area available from sPlot⁴⁰. The biome and ecoregion database used in this study are available through the Google Earthengine data catalog (https://developers.google.com/earth-engine/datasets/catalog/RESOLVE_ECOREGIONS_2017). The tree cover map used in this study are available through the Google Earthengine data catalog (https://developers.google.com/earth-engine/datasets/catalog/UMD_hansen_global_forest_change_2023_v1_11). Source Data for Figs. 1, 2 and 3 are provided with this paper as Source Data files. The raster

data of the modeled tree species' distributions generated in this study have been deposited in Zenodo: <https://doi.org/10.5281/zenodo.10911892>¹⁰². Source data are provided with this paper.

Code availability

The code used for both the modeling pipeline and the downstream analyses can be found in the following GitHub repository: https://github.com/ninavantiel/tree_sdms¹⁰³.

References

- Mrema, E. M. et al. Ten years to restore a planet. *One Earth* **3**, 647–652 (2020).
- Aronson, J., Goodwin, N., Orlando, L., Eisenberg, C. & Cross, A. T. A world of possibilities: six restoration strategies to support the United Nation's Decade on Ecosystem Restoration. *Restor. Ecol.* **28**, 730–736 (2020).
- Fischer, J., Riechers, M., Loos, J., Martin-Lopez, B. & Temperton, V. M. Making the UN decade on ecosystem restoration a social-ecological endeavour. *Trends Ecol. Evol.* **36**, 20–28 (2021).
- Lewis, S. L., Wheeler, C. E., Mitchard, E. T. A. & Koch, A. Restoring natural forests is the best way to remove atmospheric carbon. *Nature* **568**, 25–28 (2019).
- Pan, Y. et al. A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
- Jenkins, M. & Schaap, B. Forest ecosystem services. *Background analytical study, 1*. (2018).
- Cazzolla Gatti, R. et al. The number of tree species on Earth. *Proc. Natl Acad. Sci. USA* **119**, e2115329119 (2022).
- MacKenzie, W. H. & Mahony, C. R. An ecological approach to climate change-informed tree species selection for reforestation. *Ecol. Manag.* **481**, 118705 (2021).
- Rivers, M. The global tree assessment – red listing the world's trees. *BGjournal* **14**, 16–19 (2017).
- Cardinale, B. J. et al. Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
- Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C. & Baillie, J. E. M. Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE* **2**, e296 (2007).
- Owen, N. R., Gumbs, R., Gray, C. L. & Faith, D. P. Global conservation of phylogenetic diversity captures more than just functional diversity. *Nat. Commun.* **10**, 859 (2019).
- Carta, A., Peruzzi, L. & Ramirez-Barahona, S. A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *New Phytol.* **233**, 1494–1504 (2022).
- Morin, X. et al. Long-term response of forest productivity to climate change is mostly driven by change in tree species composition. *Sci. Rep.* **8**, 5627 (2018).
- Lavorel, S. & Garnier, E. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Funct. Ecol.* **16**, 545–556 (2002).
- Segovia, R. A. et al. Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Sci. Adv.* **6**, eaaz5373 (2020).
- Guo, W.-Y. et al. Paleoclimate and current climate collectively shape the phylogenetic and functional diversity of trees worldwide. Preprint at <https://doi.org/10.1101/2020.06.02.128975> (2020).
- Liang, J. et al. Co-limitation towards lower latitudes shapes global forest diversity gradients. *Nat. Ecol. Evol.* **6**, 1423–1437 (2022).
- Hidalgo, P. J., Marín, J. M., Quijada, J. & Moreira, J. M. A spatial distribution model of cork oak (*Quercus suber*) in southwestern Spain: a suitable tool for reforestation. *Ecol. Manag.* **255**, 25–34 (2008).
- Vessella, F. & Schirone, B. Predicting potential distribution of *Quercus suber* in Italy based on ecological niche models:

- conservation insights and reforestation involvements. *Ecol. Manag.* **304**, 150–161 (2013).
21. Guo, W. et al. High exposure of global tree diversity to human pressure. *Proc. Natl Acad. Sci. USA* **119**, e2026733119 (2022).
 22. Hubbell, S. et al. How many tree species are there in the Amazon and how many of them will go extinct? *Proc. Natl Acad. Sci. USA* **105**, 11498–11504 (2008).
 23. Jabot, F., Chave, J., Neubert, A. E. M. G. & McPeck, E. M. A. Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate bayesian inference. *Am. Nat.* **178**, E37–E47 (2011).
 24. Brooks, T. M. et al. Measuring terrestrial area of habitat (AOH) and its utility for the IUCN red list. *Trends Ecol. Evol.* **34**, 977–986 (2019).
 25. Syfert, M. M. et al. Using species distribution models to inform IUCN Red List assessments. *Biol. Conserv.* **177**, 174–184 (2014).
 26. 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).
 27. Lyu, L. et al. An integrated high-resolution mapping shows congruent biodiversity patterns of Fagales and Pinales. *New Phytol.* **235**, 759–772 (2022).
 28. Pecchi, M. et al. Species distribution modelling to support forest management. A literature review. *Ecol. Model.* **411**, 108817 (2019).
 29. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
 30. Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A. & Hansen, M. C. Classifying drivers of global forest loss. *Science* **361**, 1108–1111 (2018).
 31. Guisan, A. et al. Predicting species distributions for conservation decisions. *Ecol. Lett.* **16**, 1424–1435 (2013).
 32. Gray, L. K. & Hamann, A. Strategies for reforestation under uncertain future climates: guidelines for Alberta, Canada. *PLOS ONE* **6**, e22977 (2011).
 33. McKenney, D. W., Pedlar, J. H., Lawrence, K., Campbell, K. & Hutchinson, M. F. Potential impacts of climate change on the distribution of North American trees. *BioScience* **57**, 939–948 (2007).
 34. Dyderski, M. K., Paż, S., Frelich, L. E. & Jagodziński, A. M. How much does climate change threaten European forest tree species distributions? *Glob. Change Biol.* **24**, 1150–1163 (2018).
 35. Mina, M. et al. Future ecosystem services from European mountain forests under climate change. *J. Appl. Ecol.* **54**, 389–401 (2017).
 36. Elsen, P. R., Monahan, W. B. & Merenlender, A. M. Topography and human pressure in mountain ranges alter expected species responses to climate change. *Nat. Commun.* **11**, 1974 (2020).
 37. Thuiller, W. et al. Predicting global change impacts on plant species' distributions: Future challenges. *Perspect. Plant Ecol. Evol. Syst.* **9**, 137–152 (2008).
 38. Elith, J., Kearney, M. & Phillips, S. The art of modelling range-shifting species. *Methods Ecol. Evol.* **1**, 330–342 (2010).
 39. Beech, E., Rivers, M., Oldfield, S. & Smith, P. P. GlobalTreeSearch: the first complete global database of tree species and country distributions. *J. Sustain.* **36**, 454–489 (2017).
 40. Bruelheide, H. et al. sPlot – a new tool for global vegetation analyses. *J. Veg. Sci.* **30**, 161–186 (2019).
 41. Joint Research Centre (European Commission) et al. Tree Species Distribution Data and Maps for Europe (Publications Office of the European Union, LU, 2020).
 42. Thuiller, W. et al. Consequences of climate change on the tree of life in Europe. *Nature* **470**, 531–534 (2011).
 43. Merow, C., Wilson, A. M. & Jetz, W. Integrating occurrence data and expert maps for improved species range predictions. *Glob. Ecol. Biogeogr.* **26**, 243–258 (2017).
 44. Jin, Y. & Qian, H. V. PhylMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography* **42**, 1353–1359 (2019).
 45. Pavoine, S. A guide through a family of phylogenetic dissimilarity measures among sites. *Oikos* **125**, 1719–1732 (2016).
 46. Pavoine, S. adiv: an r package to analyse biodiversity in ecology. *Methods Ecol. Evol.* **11**, 1106–1112 (2020).
 47. Hengl, T. et al. SoilGrids250m: global gridded soil information based on machine learning. *PLoS ONE* **12**, e0169748 (2017).
 48. Dinerstein, E. et al. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience* **67**, 534–545 (2017).
 49. Freeman, B. G., Lee-Yaw, J. A., Sunday, J. M. & Hargreaves, A. L. Expanding, shifting and shrinking: the impact of global warming on species' elevational distributions. *Glob. Ecol. Biogeogr.* **27**, 1268–1276 (2018).
 50. Rumpf, S. B. et al. Range dynamics of mountain plants decrease with elevation. *Proc. Natl Acad. Sci. USA* **115**, 1848–1853 (2018).
 51. Lenoir, J. et al. Going against the flow: potential mechanisms for unexpected downslope range shifts in a warming climate. *Ecography* **33**, 295–303 (2010).
 52. Caudullo, G., Welk, E. & San-Miguel-Ayanz, J. Chorological maps for the main European woody species. *Data Brief.* **12**, 662–666 (2017).
 53. Ricklefs, R. E. A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.* **7**, 1–15 (2004).
 54. Slik, J. W. F. et al. Phylogenetic classification of the world's tropical forests. *Proc. Natl Acad. Sci. USA* **115**, 1837–1842 (2018).
 55. Gorelick, N. et al. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
 56. Telenius, A. Biodiversity information goes public: GBIF at your service. *Nord. J. Bot.* **29**, 378–381 (2011).
 57. Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inform.* **19**, 10–15 (2014).
 58. Serra-Diaz, J. M., Enquist, B. J., Maitner, B., Merow, C. & Svenning, J.-C. Big data of tree species distributions: how big and how good? *For. Ecosyst.* **4**, 30 (2018).
 59. Botella, C., Joly, A., Bonnet, P., Monestiez, P. & Munoz, F. A. *Multimedia Tools and Applications for Environmental & Biodiversity Informatics* (eds. Joly, A., Vrochidis, S., Karatzas, K., Karppinen, A. & Bonnet, P.) 169–199 (Springer International Publishing, Cham, 2018).
 60. Cole, E. et al. Spatial Implicit Neural Representations for Global-Scale Species Mapping. in *Proceedings of the 40th International Conference on Machine Learning* 6320–6342 (PMLR, 2023).
 61. Teng, M. et al. SatBird: a Dataset for Bird Species Distribution Modeling using Remote Sensing and Citizen Science Data. *Adv. Neural Inf. Process. Syst.* **36**, 75925–75950 (2023).
 62. Hagen, O., Skeels, A., Onstein, R. E., Jetz, W. & Pellissier, L. Earth history events shaped the evolution of uneven biodiversity across tropical moist forests. *Proc. Natl Acad. Sci. USA* **118**, e2026347118 (2021).
 63. Silva de Miranda, P. L. et al. Dissecting the difference in tree species richness between Africa and South America. *Proc. Natl Acad. Sci. USA* **119**, e2112336119 (2022).
 64. Scotese, C. R. An atlas of Phanerozoic Paleogeographic Maps: the seas come in and the seas go out. *Annu. Rev. Earth Planet. Sci.* **49**, 679–728 (2021).
 65. Ringelberg, J. J. et al. Precipitation is the main axis of tropical plant phylogenetic turnover across space and time. *Sci. Adv.* **9**, eade4954 (2023).
 66. Anderegg, W. R. L. et al. A climate risk analysis of Earth's forests in the 21st century. *Science* **377**, 1099–1103 (2022).
 67. Prober, S. et al. Climate-adjusted provenancing: a strategy for climate-resilient ecological restoration. *Front. Ecol. Evol.* **3**, 65 (2015).

68. Reiner, F. et al. More than one quarter of Africa's tree cover is found outside areas previously classified as forest. *Nat. Commun.* **14**, 2258 (2023).
69. Brandt, M. et al. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* **587**, 78–82 (2020).
70. Betts, M. G. et al. Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* **547**, 441–444 (2017).
71. Yazdandoost, F., Moradian, S., Izadi, A. & Aghakouchak, A. Evaluation of CMIP6 precipitation simulations across different climatic zones: Uncertainty and model intercomparison. *Atmos. Res.* **250**, 105369 (2021).
72. Brun, P. et al. Large-scale early-wilting response of Central European forests to the 2018 extreme drought. *Glob. Change Biol.* **26**, 7021–7035 (2020).
73. Case, M. J. & Lawler, J. J. Integrating mechanistic and empirical model projections to assess climate impacts on tree species distributions in northwestern North America. *Glob. Change Biol.* **23**, 2005–2015 (2017).
74. Kearney, M. & Porter, W. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol. Lett.* **12**, 334–350 (2009).
75. Gastón, A. et al. Species distribution models applied to plant species selection in forest restoration: are model predictions comparable to expert opinion? *New For.* **45**, 641–653 (2014).
76. Klein Goldewijk, K., Beusen, A., van Drecht, G. & de Vos, M. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Glob. Ecol. Biogeogr.* **20**, 73–86 (2011).
77. Li, W. et al. Human fingerprint on structural density of forests globally. *Nat. Sustain.* **6**, 368–379 (2023).
78. Fauset, S. et al. Hyperdominance in Amazonian forest carbon cycling. *Nat. Commun.* **6**, 6857 (2015).
79. Rosindell, J., Hubbell, S. P. & Etienne, R. S. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol. Evol.* **26**, 340–348 (2011).
80. Newbold, T., Oppenheimer, P., Etard, A. & Williams, J. J. Tropical and Mediterranean biodiversity is disproportionately sensitive to land-use and climate change. *Nat. Ecol. Evol.* **4**, 1630–1638 (2020).
81. Van Rossum, Guido & Drake, Fred L. *Python 3 Reference Manual* (CreateSpace, 2009).
82. McKinney, W. Data structures for statistical computing in python. in *Proceedings of the 9th python in science conference* vol. 445. p. 51–56 (2010).
83. Harris et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
84. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
85. Waskom, M. et al. seaborn: v0.8.1 (September 2017). Zenodo <https://doi.org/10.5281/zenodo.883859> (2017).
86. R. Core Team. *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/> (2022).
87. Wickham, H., Vaughan, D. & Girklich, M. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr> (2023).
88. Wickham, H. et al. Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
89. Müller, K. & Wickham, H. *Tibble: Simple Data Frames*. <https://CRAN.R-project.org/package=tibble> (2022).
90. Dowle, M. & Srinivasan, A. *Data.Table: Extension of 'data.Frame'*. <https://CRAN.R-project.org/package=data.table> (2022).
91. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr> (2023).
92. Auguie, B. *gridExtra: Miscellaneous Functions for 'Grid' Graphics*. <https://CRAN.R-project.org/package=gridExtra> (2017).
93. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
94. Karger, D. N. et al. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122 (2017).
95. Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L. & Karger, D. N. Global climate-related predictors at kilometer resolution for the past and future. *Earth Syst. Sci. Data* **14**, 5573–5603 (2022).
96. Brun, P. et al. Model complexity affects species distribution projections under climate change. *J. Biogeogr.* **47**, 130–142 (2020).
97. Lange, S. Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geosci. Model Dev.* **12**, 3055–3070 (2019).
98. Frieler, K. et al. A framework for the cross-sectoral integration of multi-model impact projections: land use decisions under climate impacts uncertainties. *Earth Syst. Dyn.* **6**, 447–460 (2015).
99. Brodzik, M. J., Billingsley, B., Haran, T., Raup, B. & Savoie, M. H. EASE-Grid 2.0: incremental but significant improvements for earth-gridded data sets. *ISPRS Int. J. Geo-Inf.* **1**, 32–45 (2012).
100. Oksanen, J. et al. Package 'vegan'. *Community Ecology Package Version 2*, p. 1–295 (2013).
101. Pedregosa, F. et al. Scikit-learn: Machine Learning. *Python. J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
102. van Tiel, N. et al. SDM results for 10,590 tree species from 'Regional uniqueness of tree species composition and response to forest loss and climate change'. Zenodo <https://doi.org/10.5281/zenodo.10911892> (2024).
103. van Tiel, N. Regional uniqueness of tree species composition and response to forest loss and climate change. Zenodo <https://doi.org/10.5281/zenodo.10908654> (2024).
104. Maitner, B. S. et al. The bien r package: a tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol. Evol.* **9**, 373–379 (2018).
105. Schepaschenko, D. et al. A dataset of forest biomass structure for Eurasia. *Sci. Data* **4**, 170070 (2017).
106. DRYFLOR et al. Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* **353**, 1383–1387 (2016).
107. GBIF. *Occurrence Download* <https://doi.org/10.15468/dl.ruu1ym> (2020).
108. Hudson, L. et al. *The 2016 release of the PREDICTS database [SUPERSEDED]* <https://doi.org/10.5519/0066354> (2016).
109. Dauby, G. et al. RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**, 1–18 (2016).
110. QGIS Development Team. *QGIS Geographic Information System* <http://qgis.osgeo.org> (2009).

Acknowledgements

This project was partly financed by the Internal WSL grant Treemap (LP), the Internal WSL grant ClimEx (D.N.K. and N.E.Z.), the BiodivERsA project FeedBaCks with the national funder Swiss National Science Foundation (20BD21_193907) (D.N.K. and N.E.Z.), BiodivERsA project Futureweb with the national funder Swiss National Science Foundation (20BD21_184131) (DNK), the deepHSM project with the national funder Swiss National Science Foundation (204057) (N.v.T. and D.T.). N.v.T., J.v.d.H. and T.W.C. acknowledge funding from DOB Ecology. N.v.T. acknowledges funding from the Restor Foundation. We thank to David Boehm for data cleaning and merging as part of his Swiss civil service. We thank the following organizations for making their data freely available: BIEN, BIOMASS, Chelsa, CONIFER, DRYFLOR, GBIF, GFBI, GlobalTreeSearch, IDIGBIO, INDIABIODIVERSITY, PNG, PREDICTS, RAINBIO, sPlot, SoilGrids.

Author contributions

N.v.T., T.W.C. and L.P. conceived and designed the analysis. N.v.T., F.F., P.B., J.v.d.H., D.N.K., C.M.C., L.L. and N.E.Z. contributed analysis tools. N.v.T. performed the analysis. N.v.T. and L.P. wrote the original draft of

the paper with assistance from T.W.C., D.T. and N.E.Z. All authors contributed to revisions and editing of the final version.

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/s41467-024-48276-3>.

Correspondence and requests for materials should be addressed to Nina van Tiel.

Peer review information *Nature Communications* thanks Kyle Dexter and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024