Check for updates

Received: 13 September 2023 | Revised: 2 January 2024 | Accepted: 25 March 2024

DOI: 10.1111/jbi.14840

Journal of Biogeography

WILEY

RESEARCH ARTICLE

Spatial phylogenetics of Fagales: Investigating drivers of temperate forest distributions

R. A. Folk[1] | C. M. Siniscalchi[1,2] | J. Doby[3] | H. R. Kates[3] | S. R. Manchester[3] | P. S. Soltis[3,4,5] | D. E. Soltis[3,4,5,6] | R. P. Guralnick[3,5] | M. Belitz[3,6]

[1]Department of Biological Sciences, Mississippi State University, Mississippi State, Mississippi, USA

[2]Mississippi State University Libraries, Mississippi State, Mississippi, USA

[3]Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA

[4]Genetics Institute, University of Florida, Gainesville, Florida, USA

[5]Biodiversity Institute, University of Florida, Gainesville, Florida, USA

[6]Department of Biology, University of Florida, Gainesville, Florida, USA

Correspondence
R. A. Folk, Department of Biological Sciences, Mississippi State University, Mississippi State, MS, USA.
Email: rfolk@biology.msstate.edu

Funding information
National Science Foundation, Grant/Award Number: DEB-1916632

Abstract

Aim: Quantifying the phylogenetic diversity of temperate trees is essential for understanding the processes that have shaped the modern distribution of temperate broadleaf forest and other major forest biomes. Here, we focus on Fagales, an iconic member of forests worldwide, to uncover global diversity and endemism patterns and investigate the distribution of root nodule symbiosis (RNS), an important morphological specialisation in this clade, as a key factor behind these patterns.

Location: Global.

Taxon: Fagales.

Methods: We combined phylogenetic data covering 60.2% of living species, fine-scale distribution models covering 90% of species, and nodulation data covering all species to investigate the distribution of species richness and phylogenetic diversity at fine spatial scales compared to the distribution of RNS. We identify abiotic environmental factors associated with RNS and with Fagales diversity in general.

Results: We find the highest species richness in temperate east Asia, eastern North America, and equatorial montane regions of Asia and Central America. By contrast, relative phylogenetic diversity (RPD) is highest at higher latitudes, where RNS also predominates. We found a strong spatial structuring of regionalisations of Fagales floras, reflecting distinct Northern and Southern Hemisphere floras (except a unique Afro-Boreal region), each with distinct RNS-environment relationships.

Main Conclusions: Although species richness and phylogenetic regionalisation for Fagales accord well with traditional biogeographic concepts for temperate forests, this is not the case for RPD. RNS is almost universal in the highest RPD regions, which may reflect ecological filtering promoting RNS in these regions. Our results highlight the utility of global-scale, clade-specific spatial phylogenetics and its utility for understanding drivers of diversity in species-rich clades.

KEYWORDS
Fagales, phylogenetic diversity, root nodule symbiosis, spatial phylogenetics, temperate broadleaf forest

R. P. Guralnick and M. Belitz contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2024 The Authors. *Journal of Biogeography* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

The distribution of today's forest biomes is profoundly shaped by the division of tropical and extratropical floras into distinct phytogeographical domains, with limited mutual migration imposed by niche conservatism (Donoghue, 2008; Folk et al., 2020; Takhtajan, 1986; Wiens & Donoghue, 2004) leading to distinct macroevolutionary histories in these regions (Axelrod, 1966; Economo et al., 2018; Edwards et al., 2017; Schubert et al., 2019; Sun et al., 2020; Wolfe, 1987). Although physiological challenges such as frost and elevated seasonality are important for governing the distribution of temperate floras, the diversity of tree species observed in today's temperate forests is more unevenly distributed than would be suggested by these factors alone (see Figure 2 in Folk et al., 2020). The discontinuous and imbalanced distribution of temperate forests has instead been primarily hypothesised to reflect the history of temperate forest biomes, with refugial areas remaining after the disappearance of ancient polar forests leading to modern-day endemism and diversity centres (Axelrod, 1983; Engler, 1905; Manchester, 1999; Wen, 1999; Wolfe, 1975). Persistence of the original lineage composition of an ancient northern temperate tree flora is considered greatest in eastern Asia, while aridification and glaciation resulted in the loss of many lineages elsewhere and a near-total loss in Western Europe and Western North America (Manchester, 1999; Wen, 1999; Wolfe, 1975). While refugial areas of ancient temperate forests have long been hypothesised to explain centres of angiosperm tree diversity, as informed by changes in distribution patterns documented in the fossil record, alternative hypotheses, such as ancestral origins in current diversity hotspots or ecological filtering (Cavender-Bares et al., 2004), are also plausible and largely untested.

As an iconic member of temperate forests worldwide, Fagales (recognised as an order by APG IV [2016]) may be the clade of choice for understanding the origin of forests in extratropical areas. As components of the major centres of diversity for temperate forests, Fagales include some of the most familiar Northern Hemisphere trees such as alder (*Alnus*), beech (*Fagus*), birch (*Betula*), hickory (*Carya*), oak (*Quercus*), and walnut (*Juglans*), and important plants of the Southern Hemisphere such as southern beech (*Nothofagus*) and she-oak (Casuarinaceae). Standing out among major woody clades for its ecological diversity in the Northern Hemisphere, Fagales are broadly present across major temperate to boreal forest biomes with a significant additional presence in tropical areas, both hot lowland areas (Casuarinaceae) and cooler upland regions (especially Juglandaceae and Ticodendraceae) (Wheeler et al., 2022). Fagales are similarly important within the limited extent of temperate and subtropical forests in the Southern Hemisphere, with Nothofagaceae occurring across southern South America to New Zealand and Papua New Guinea, forming an iconic Antarctic disjunction (Cook & Crisp, 2005; Hinojosa et al., 2016), and Myricaceae and Casuarinaceae covering the remaining southerly latitudes in Africa and Australasia, respectively.

A noteworthy morphological specialisation in Fagales that is unevenly distributed geographically across Earth is the presence in some members of root nodules, i.e., specialised root structures that house symbiotic diazotrophic bacteria. Fagales contain three of the 10 families (viz., Betulaceae, Casuarinaceae, and Myricaceae; Pawlowski & Bisseling, 1996; Pawlowski & Sprent, 2008) that nodulate, representing three of nine independent origins (Kates et al., 2024) of "actinorhizal" plants, those whose root nodule symbiosis (RNS) involves Actinomycetota (Actinobacteria) rather than the Alpha- or Betaproteobacteria ("rhizobia") found predominantly in the legumes (Fabaceae). Recently, differences in latitudinal diversity have been identified between plants with differing bacterial partners; compared to other forms of diazotrophic symbiosis, actinorhizal RNS is most prevalent in temperate to boreal environments (Tamme et al., 2021), while other forms of diazotrophic symbiosis achieve their greatest prevalence in the tropics and subtropics. This patterning suggests that symbiosis may shape continent-scale distributions and be an important factor in constraining plant distributions. Despite this importance, and while the distinct habitats of actinorhizal plants have been noted (Folk et al., 2020; Menge et al., 2019; Tamme et al., 2021), we do not yet understand the specific environmental factors (e.g., temperature, precipitation, soil; Doby et al., 2022; Tamme et al., 2021) that may be causative of differing latitudinal patterns among symbiosis types. Clarifying these factors within actinorhizal plants would shed light on the interaction between RNS and the abiotic environment and further identify which environments promote plant investment in symbiosis.

Mapping areas of diversity is critical to connecting geographic information to phylogenetic hypotheses and ultimately testing historical, trait-based, or other hypotheses of drivers of tree distributions (Allen et al., 2019; Mishler et al., 2020; Scherson et al., 2017; Thornhill et al., 2017). Phylogenetic measures of biodiversity help reveal additional aspects of lineage diversity relevant to distinguishing among such hypotheses (Doby et al., 2022; Li et al., 2020). Using spatial phylogenetic tools (Mishler et al., 2014) to distinguish areas of endemism promises to provide additional insight into areas that harbour ancient diversity (i.e., paleoendemism, which could indicate areas of reduced extinction) or that disproportionately contain recently evolved endemics (i.e., neoendemism, which could be the result of ecological filtering or recent in situ diversification). Integrating diversity estimates with contemporary environmental data and ecologically significant traits such as RNS yields additional power to distinguish ecological filtering from alternative explanations (Suissa et al., 2021; Thornhill et al., 2017).

While diversity mapping has seen intense interest in trees generally (e.g., Lyu et al., 2022; Segovia et al., 2020) and in focused groups of Fagales, such as oaks (Cavender-Bares et al., 2004), Fagales themselves have never been the subject of a global spatial phylogenetic analysis. Here, we assemble a view of fagalean phylodiversity that is fine-grained yet global in extent, via development of a new species distribution modelling pipeline that is semi-automated and uses best practices for modelling to produce a robust map of diversity. Using these models and multiple measures of phylogenetic and species diversity, we investigated and mapped the proportion of lineages

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

within Fagales that engage in RNS and investigated how this trait aligns with centres of diversity and endemism, asking whether this symbiotic strategy may be associated with ecological filtering of phylogenetic diversity. We ask the following major questions: (1) where are centres of Fagales diversity and endemism located globally, and (2) which environmental conditions best predict the distribution of these centres? Then, we ask (3) whether centres of RNS distribution coincide with overall centres of diversity or whether they are spatially concentrated in certain phylogenetic assemblages, and (4) which precipitation, temperature, or soil factors best predict centres of RNS.

2 | MATERIALS AND METHODS

2.1 | Phylogenetic framework

Fagales are well studied phylogenetically with numerous approaches to date, from ITS and small numbers of plastid loci (Li et al., 2004; Manos et al., 2007) to genomic data (Yang et al., 2021) and fossils (Larson-Johnson, 2016; Siniscalchi et al., 2023). Many of these studies focus primarily on higher-level relationships and are not densely sampled; for the purpose of spatial phylogenetics, maximising species presence is crucial (Li et al., 2019). We therefore elected to use a recent phylogenetic tree constructed across the nitrogen-fixing clade (Siniscalchi et al., 2022), which comprises Fagales as well as the orders Cucurbitales, Fabales, and Rosales. Siniscalchi et al. (2022) used 20 DNA regions from the nuclear, plastid, and mitochondrial genomes, recovered by aggregating all GenBank data available for these 20 markers, as well as extracting these same markers from a recent large-scale phylogenomic sequencing analysis (Kates et al., 2024). Tree inference, performed in RAxML-NG (Kozlov et al., 2019), was constrained by a high-quality backbone tree based on nuclear phylogenomic data derived from Kates et al. (2024) and calibrated with 11 secondary constraints placed across the nitrogen-fixing clade following Magallón et al. (2015); analytical details are available in Siniscalchi et al. (2022). Phylogenetic dating has proven controversial in Fagales (reviewed in Sauquet et al., 2012; Siniscalchi et al., 2022), but beyond our methodological decision to focus on an ultrametric branch length scheme (see Allen et al., 2019), the absolute time scale is irrelevant to hypothesis testing as conducted here (although the relative scaling of different clade dates is important and could have limited impacts; see also Li et al., 2019). The Fagales clade in this tree of the nitrogen-fixing clade, pruned for downstream analysis, comprises 707 species (60.2% of species-level diversity; Stevens, 2001 onwards) with complete genus-level representation.

2.2 | Occurrence records

Occurrence records of all Fagales found on the biodiversity discovery platforms GBIF (2020) and iDigBio were aggregated on May 21, 2020, compiling a dataset of 7,623,848 records. This search was restricted to georeferenced occurrence records, included both specimens and human observation-based records, and filtered GBIF-flagged geospatial issues. Next, we harmonised synonyms to a standardised species list (the NitFix names database, as reported in Folk et al., 2021) to deal with taxonomic changes that may not be up to date in GBIF or iDigBio repositories. Duplicate records and specimens, here defined as records or specimens collected at the same location, by the same collector, on the same day, were filtered to only retain single specimen records. Additionally, records were removed if they did not have coordinate values. Each species that had at least five occurrence records underwent a coordinate cleaning procedure using the *CoordinateCleaner* package v. 2.0–20 (Zizka et al., 2019). During this procedure, records were removed if they: 1) had identical latitude and longitude coordinates, 2) were within 500 m of the centroids of political countries or provinces, 3) were within 0.5 degree radius of the GBIF headquarters, 4) were within 100 m of zoos, botanical gardens, herbaria, and museums based on a global database of ~10,000 such biodiversity institutions (see Zizka et al., 2019), 5) had values of precisely zero for latitude or longitude (and therefore would be erroneous), or 6) were greater than 1000 km away from all other records of a species. Dot maps of occurrence records were then plotted, and based on manual inspection of species occurrence maps, species that retained obvious errors underwent manual occurrence record filtering. Ultimately, 455,704 records (6%) passed these strict filtering criteria; this likely reflects, in part, two dataset properties specific to Fagales: (1) the very large number of non-native and cultivated records in online repositories for this horticulturally important group; and (2) the existence of high spatial redundancy in the raw dataset, which was filtered by the removal of pixel-wise duplicates.

2.3 | Niche modelling approach

We built an ecological niche modelling pipeline in R to predict the ecological niche of 1045 species that had at least five occurrence records after undergoing the coordinate cleaning described above. This pipeline adapts the workflow outlined in Abbott et al. (2022) and was designed to automate the building of ecological niche models, while including steps that customize models for each species. First, the accessible area, which was the area where the distribution model was fit and projected, was determined by calculating a buffered alpha hull around occurrence records that passed all automated and manual filtering steps. The alpha hull was calculated using the getDynamicAlphaHull function from the R package *rangeBuilder* v. 1.5 (Davis Rabosky et al., 2016), and the alpha hull was then buffered by the larger value of either 75 km or the 80th percentile distance of each occurrence record to its nearest occurrence record. Next, we fit a Maxent model (Phillips et al., 2006, 2017) with default settings using the *dismo* package v. 1.3–3 in R (Hijmans et al., 2017).

Our initial model included 13 bioclimatic variables from 19 available in WorldClim (Fick & Hijmans, 2017), three soil variables

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

provided by the International Soil Reference and Information Center (Batjes et al., 2017), and two topography layers provided by EarthEnv (Amatulli et al., 2018). Variables were selected from these sources on the basis of biological relevance to plant distributions; collinearity was dealt with at the modelling stage as described below. The soil layers were downloaded at depths of 0–5 cm, 5–15 cm, and 15–30 cm, and the average value of each cell for these given depths was calculated for use in our models. In total, these predictor variables were bio1, bio2, bio4, bio5, bio6, bio8, bio9, bio12, bio13, bio14, bio15, bio16, bio17, elevation, ruggedness, soil nitrogen, sand, and soil organic carbon. Initial variables had a spatial resolution of approximately 1 km at the equator and were aggregated five-fold to the coarser resolution for model building. We calculated the variance inflation factors (VIF) of our initial global models with all 18 variables; if any predictor variable had a VIF greater than 5, we removed the variable with a VIF greater than 5 that contributed the least to the model given its permutation contribution value. This step was repeated until no variables were retained in the model with a VIF greater than five. These species-specific predictor variables were used in the following step below.

We used the R package *ENMeval* v. 2.0.1 (Muscarella et al., 2014) to evaluate many combinations of Maxent models with different tuning parameters to optimize model complexity while maintaining predictability. We fit models for each combination of tuning parameters within range multipliers of 0.5, 1, 2, 3, and 4 and feature classes of "linear", "linear + quadratic", "linear + quadratic + hinge + product", and "linear + quadratic + hinge + product + threshold". Occurrence and background localities were partitioned into training and testing bins using block partitioning. The model with the lowest AICc value was selected as the top model if it had training and validation area under the curve (AUC) greater than 0.7, while in the few cases where those values were less than 0.7, we selected the model with the highest validation AUC as the top model. Top models were converted to predicted presence/absence maps using the tenth percentile rule, where a model threshold was selected that would classify 90% of locations used for training as presences and the lowest 10% of values as absences. This threshold was chosen because our underlying species occurrence data used in model fitting may still have a small proportion of uncertain or poor-quality records despite several strict filtering criteria and manual examination, and thus allowing 10% of presumed presences to be outside predicted suitable habitats helps reduce commission error.

2.4 | Species richness, relative phylogenetic diversity, and endemism

The thresholded ecological niche models for all species and the phylogeny described above were imported into Biodiverse v. 3.1 (Laffan et al., 2010). These datasets were used to calculate species richness (SR) and relative phylogenetic diversity (RPD). RPD is the ratio of phylogenetic diversity (PD, measured as the sum of branch lengths connecting the terminal taxa present in each location) on the original phylogenetic tree compared to a phylogeny with the same topology but with a transformation imposed to equalize branch lengths (Mishler et al., 2014). Thus, low RPD represents more shallow branches compared to a tree with equal branch lengths, whereas high RPD represents more long branches. We opted to focus on RPD given that raw PD displays strong correlation with SR. Mapped raw PD is available in Figure S1.

We also calculated the proportion of species in each grid cell engaging in nodulation. This was performed by matching species to a recent comprehensive genus-level database of nodulating species (Kates et al., 2024). Kates et al. (2024) reported an ancestral reconstruction of RNS in a phylogenetic context; the determination of spatial distributions as reported in the present study was done by merging nodulation states recognised in Kates et al. (2024) with thresholded distributional models. A similar approach was used with occurrence records in Tamme et al. (2021), but the use of distribution models here has the benefit of reducing range omission error in poorly sampled regions. While data were only scored for genera and hence 100% species-level trait coverage represents an assumption of invariance within genera, the distribution of nodulation is thought to be fairly well understood in Fagales and reflect this assumption (Ardley & Sprent, 2021; Pawlowski & Sprent, 2008).

2.5 | Randomisations and endemism categorisation

We used spatially structured randomisations to determine geographic locations where RPD was significantly higher or lower than expected given a randomised distribution of species having equal branch lengths. Randomisations were calculated holding richness and range size of each species within a grid cell constant. Values of RPD were then calculated for each of 100 iterations, which creates a null distribution for each grid cell. A two-tailed test based on percentiles calculated from the null distribution is used to determine whether observed values are significantly higher or lower (alpha level 0.025 in each direction) compared to null distributions.

We also calculated relative phylogenetic endemism (RPE; Mishler et al., 2014), the ratio between measured phylogenetic endemism (PE) and the PE estimated from a phylogeny with equal branch lengths. The logic of including measures of endemism significance is that RPD does not include range size information, and therefore endemism analyses are required to identify hotspots of diversity made up of unique species rather than confluences of wide-ranging taxa (which would produce mismatches between RPD and RPE significance). We ran a randomisation on RPE as a means to categorize different types of phylogenetic endemism using the Categorical Analysis of Neo- And Paleo-Endemism (CANAPE) approach (Mishler et al., 2014). CANAPE is a hypothesis-based randomisation procedure that identifies areas with higher phylogenetic endemism than expected under a null model of RPE, based on a comparison tree with equal branch lengths, and classifies these significant areas in terms of neo- or paleoendemics. It first selects grid cells that are significantly high (one-tailed test,

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

alpha level 0.05) in either the numerator or denominator of RPE and then uses a two-tailed test of the RPE ratio (alpha level 0.025 in each direction) to categorize cells as having a high proportion of neoendemics, a high proportion of paleoendemics, a mixture of both types, or no significant endemism. Neoendemics have higher than expected concentrations of range-restricted short branches, while paleoendemics have higher than expected concentrations of range-restricted long branches. Endemism measures and randomisations were calculated in Biodiverse v. 3.1 (Laffan et al., 2010). A map of raw RPE is available in Figure S2.

2.6 | Phylogenetic regionalisation

Examining turnover of lineages across space offers the opportunity to test traditional hypotheses of biogeographic regions using quantitative methods (Daru et al., 2017). "Phyloregions" are defined here as clusters of areas of Earth possessing a similar phylogenetic composition of Fagales species, based on community distance metrics. We examined range-weighted phylogenetic turnover by calculating a pairwise distance matrix as a basis for clustering, with the purpose of identifying regions containing similar phylogenetic composition (Laffan et al., 2010). In a range-weighted phylogenetic turnover analysis, values of phylogenetic turnover (e.g., phylobeta) are first generated by comparing the lengths of branches of the overarching tree shared and unshared among pairs of cells. Then the phylobeta values are weighted by the fraction of their geographic range found in that location. We manually selected breaks in the dendrogram that determined well-defined groupings of contiguous sets of coloured grid cells. These analyses were also performed in Biodiverse; we aggregated the cells to a coarser resolution due to computational limitations in this software. The cells were coarsened to one tenth the resolution of the other analyses; this resulted in dropping 32 species (hence; this analysis reflects 1013 mapped species). In addition to comparing these recovered regions to the most authoritative traditional treatment of woody plant biogeographic regionalisation (the 35 floristic regions of Takhtajan 1986), we also used regionalisation hypotheses to identify shared lineage diversity among recognised species diversity and assess whether RNS centres share RNS lineage diversity or are independent phylogenetic assemblages.

2.7 | Environmental associates of grid cell metrics

We fit a series of models under a model choice paradigm to assess the best environmental predictors of RPD, CANAPE endemism categories, and proportion of grid cell species engaging in nodulation. We used these models to ask which environmental factors best explain these three responses, and particularly whether environmental factors shaping diversity are shared among the metrics. Environmental data were summarised by eight selected predictors chosen based on the biology of the plants following previous work in the clade (Doby et al., 2022; Siniscalchi et al., 2022): aridity (calculated as the UNEP aridity index following Folk et al., 2020), mean annual temperature (bio1), temperature annual range (bio7; included as a measure of temperature seasonality), annual precipitation (bio12), temperature of the driest quarter (bio17; included as a measure of precipitation seasonality), elevation, and three soil predictors: nitrogen content (chosen to study RNS distribution), soil pH, and soil carbon content (the latter two as additional factors representing edaphic ecology, well-known as constraining plant distributions).

Two classes of models were fit: standard generalised linear models (GLMs) and linear mixed models (LMMs). The logic behind including LMMs was as a natural framework for handling spatial autocorrelation as a random effect and partitioning variance separately from the main predictors, treated as fixed effects. Within each model class, we tested either the full eight-predictor set of environmental variables or a reduced set of five predictors, individually chosen within each model class and response, based on the highest magnitude of normalised coefficients in the full GLM. LMM models additionally fit cell centroid latitude and longitude as random effects. Finally, a LMM model was fit without environment, using only latitude and longitude as random effects, in order to verify the predictive power of environmental data for the response. To summarize, the model set therefore included five distinct models, GLM-full, GLM-reduced, LMM-full, LMM-reduced, and LMM-no environment, for each of three responses: RPD, CANAPE, and proportion of RNS species. For the continuous RPD and proportion of RNS responses, standard GLM and LMM were used. For the categorical CANAPE response, endemism significance categories were lumped to yield a binary comparison with non-significant cells, and the model family was specified as binary with a logit link function. Model choice used standard AIC (Akaike, 1974) in consideration of the large sample sizes.

3 | RESULTS

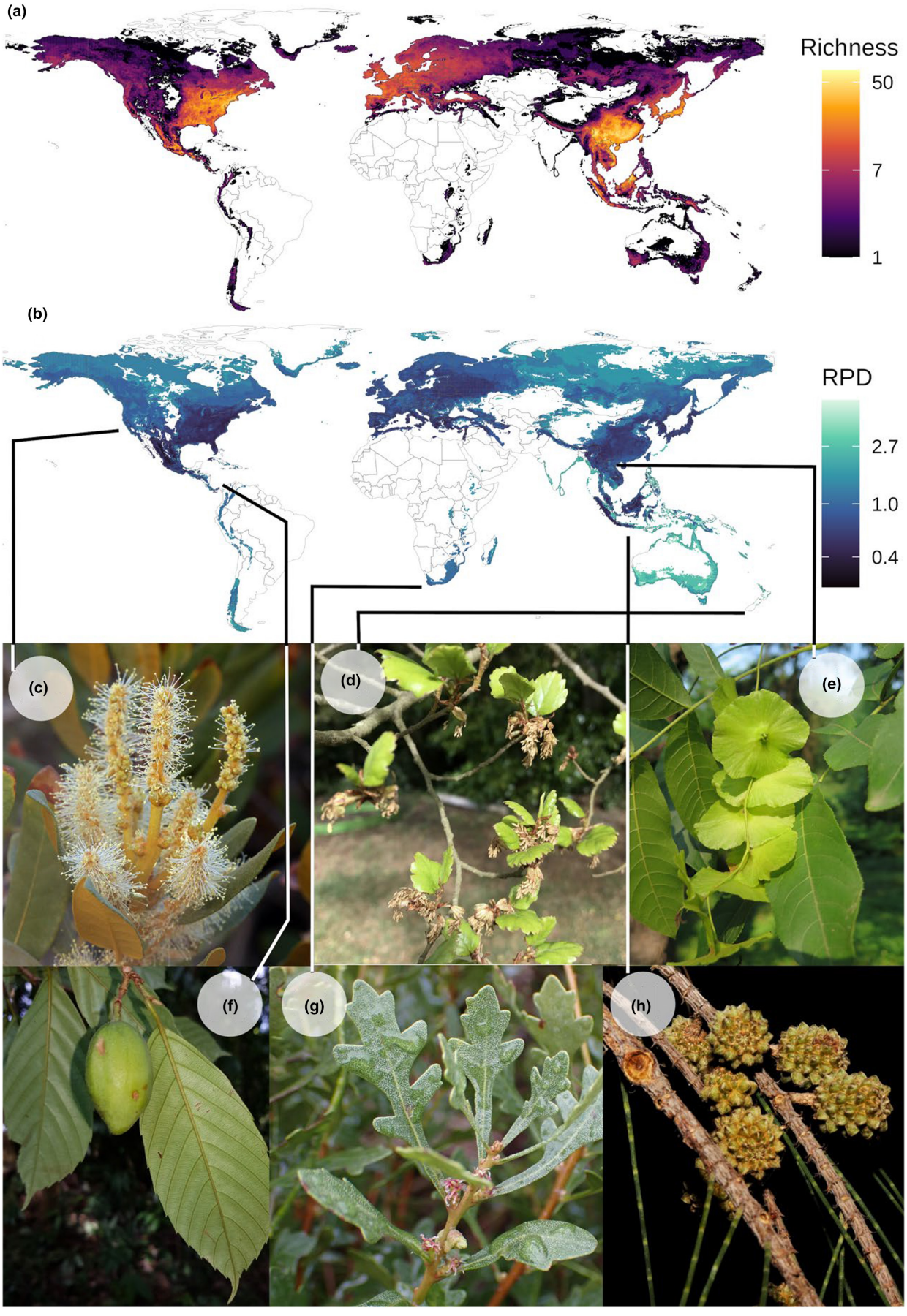
3.1 | Basic diversity metrics

Species distribution models covered 1045 species or 90% of all recognised Fagales species, suggesting robust estimates of species richness patterns. As expected, we recovered the highest species richness of extant Fagales in temperate eastern Asia, with secondary hotspots

FIGURE 1 Summary of diversity in Fagales. (a) Global species richness of Fagales, with warm colours representing more diverse areas. White areas of land indicate no mapped species. (b) Global distribution of relative phylogenetic diversity (RPD) in Fagales, with greener colours indicating more diverse areas. (c–g) Photos of representative plants: (c) *Chrysolepis sempervirens* (photo credit: Steve Matson). (d) *Nothofagus fusca* (photo credit: Nicola Baines). (e) *Cyclocarya paliurus* (photo credit: Yao Li). (f) *Ticodendron incognitum* (photo credit: Leonardo Álvarez-Alcázar). (g) *Myrica quercifolia* (photo credit: David Hoare). (h) *Casuarina equisetifolia* (photo credit: Savvas Zafeiriou).

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Journal of Biogeography



13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of species richness in eastern North America, montane Mexico, and Malesia (Figure 1a). In equatorial regions, species richness is high only in the Malesian biogeographic province (sensu Takhtajan 1986) from Indonesia to Papua New Guinea. Fagales communities in the Southern Hemisphere are relatively species-poor by contrast, with the highest diversity in southwestern Australia. Finally, although not as high in species richness as the foregoing regions, a further hotspot comprises much of Europe.

The distribution of RPD was markedly different from that for species richness, showing centres of RPD primarily in the Southern Hemisphere (Figure 1b). The highest RPD was seen in southern Australia, and relatively high RPD was also associated with eastern and southern Africa, the Andes, and to a lesser extent the boreal Northern Hemisphere. Further minor areas of high RPD, although associated with species-poor communities (Figure 1a), extend in a narrow area comprising coastal regions in South and Southeast Asia.

3.2 | Randomisation tests of diversity metrics

We implemented randomisation tests to assess whether phylogenetic diversity (specifically, tests of RPD and RPE) were outside of null expectations (that is, whether grid cells comprised taxa that were unstructured with regard to phylogenetic branch length and range restriction). RPD randomisations clearly demonstrated significant spatial structuring of lineage diversity. Areas of significantly high RPD (blue cells, Figure 2a), corresponding to communities with especially long branch lengths, occur in six biogeographic areas, listed in descending spatial extent: southern Australia, southeast Asia, western Europe, Malesia, the Sierra Madre Oriental south to Central America, and Tierra del Fuego. In the Northern Hemisphere, areas of significantly low RPD (red cells, Figure 2a) were most prevalent in the Americas, extending from the coastal plain of the southeastern United States to most of montane Mexico. In Eurasia, low RPD was associated with the Mediterranean basin, an additional northern region from Poland to the Volga, and further isolated Asian occurrences mostly in the Himalayan region to montane southeast Asia. In equatorial regions, significantly low RPD was prevalent in the western Malesia biogeographic province (i.e., Sundaland), but significantly low RPD does not occur in the Southern Hemisphere outside of equatorial Malesia.

CANAPE analyses (Figure 2b) were included to further contextualize high RPD regions by identifying areas that also have high concentrations of range-restricted taxa. Identified significant endemism regions were mostly similar in distribution to RPD randomisations (which, although similar in implementation, do not incorporate range size and therefore are not measures of endemism), with CANAPE showing that most significant regions of endemism are characterised by mixed endemism patterns. This alignment between the two analyses indicates that significance was largely driven by range-restricted taxa. Southern Mexico, southeast Asia, southwestern Australia, and the southern Andes were characterised by significant paleoendemism. RPD and CANAPE randomisations were mostly aligned, but the most important difference between the CANAPE and RPD results was that portions of southern Africa and Madagascar have strong Fagales neoendemism but no RPD significance (these neoendemic hotspots are also key areas of RNS distribution; see below). This significance misalignment indicates that southern Africa and Madagascar are distinguished by short phylogenetic branches in range-restricted taxa but not necessarily overall.

3.3 | Phylogenetic regionalisation

Fagales phylogenetic regionalisation accorded well with traditionally recognised biogeographic regions as summarised by Takhtajan (1986), which were rooted primarily in woody plant distributions. The most widely distributed region was a boreal region across the Northern Hemisphere (dark green, Figure 3), and three more major regions were at mid-higher latitudes in the south boreal region (light green, yellow, dark blue, Figure 3). The last of these (dark blue), termed here an "Afro-boreal region," was the only phyloregion present in large areas of both hemispheres, being present in both lower boreal latitudes and in southern Africa, and therefore closely matching areas of Fagales distribution dominated by *Myrica* of Myricaceae.

Eastern North America and temperate eastern Asia (respectively dark purple and light orange in Figure 3) were recognised as distinct phyloregions for Fagales and were not closely related to each other (see dendrogram, Figure 3b). North America comprises mostly the boreal and eastern North America regions discussed above, but the California Floristic Province was retrieved as part of a distinct "Mediterranean" phyloregion also occurring in the Mediterranean Basin (Figure 3, brown), and much of interior western North America was part of a semi-arid mid-latitude region distributed across the Northern Hemisphere (Figure 3, light blue). Montane Central America was primarily delimited as a phyloregion shared with portions of Malesia (Figure 3a, red, the second-most divergent phyloregion; Figure 3b), reflecting close phylogenetic relationships among small endemic genera of Fagaceae and Juglandaceae in these regions (e.g., *Formanodendron* and *Colombobalanus*; *Engelhardia* and *Alfaroa*). Finally, a small amount of the eastern North America region was present in the Sierra Madre Oriental of Mexico, reflecting a disjunction known in *Fagus grandifolia* (Fagaceae; Williams-Linera et al., 2003) and other taxa. The semi-arid phyloregion, along with a broadly distributed Eurasian phyloregion (Figure 3, turquoise), comprised continental Europe and the Mediterranean Basin.

Two phyloregions were unique to the Southern Hemisphere. One is a widespread region of Malesia and coastal Australia east to northern New Zealand (Figure 3, pink), representing Fagales communities dominated by Casuarinaceae. The second is a phyloregion unique to southern South America (Figure 3, black) that reflects the distribution of Nothofagaceae in Valdivian forests; this was the most distinct phyloregion (Figure 3b).

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

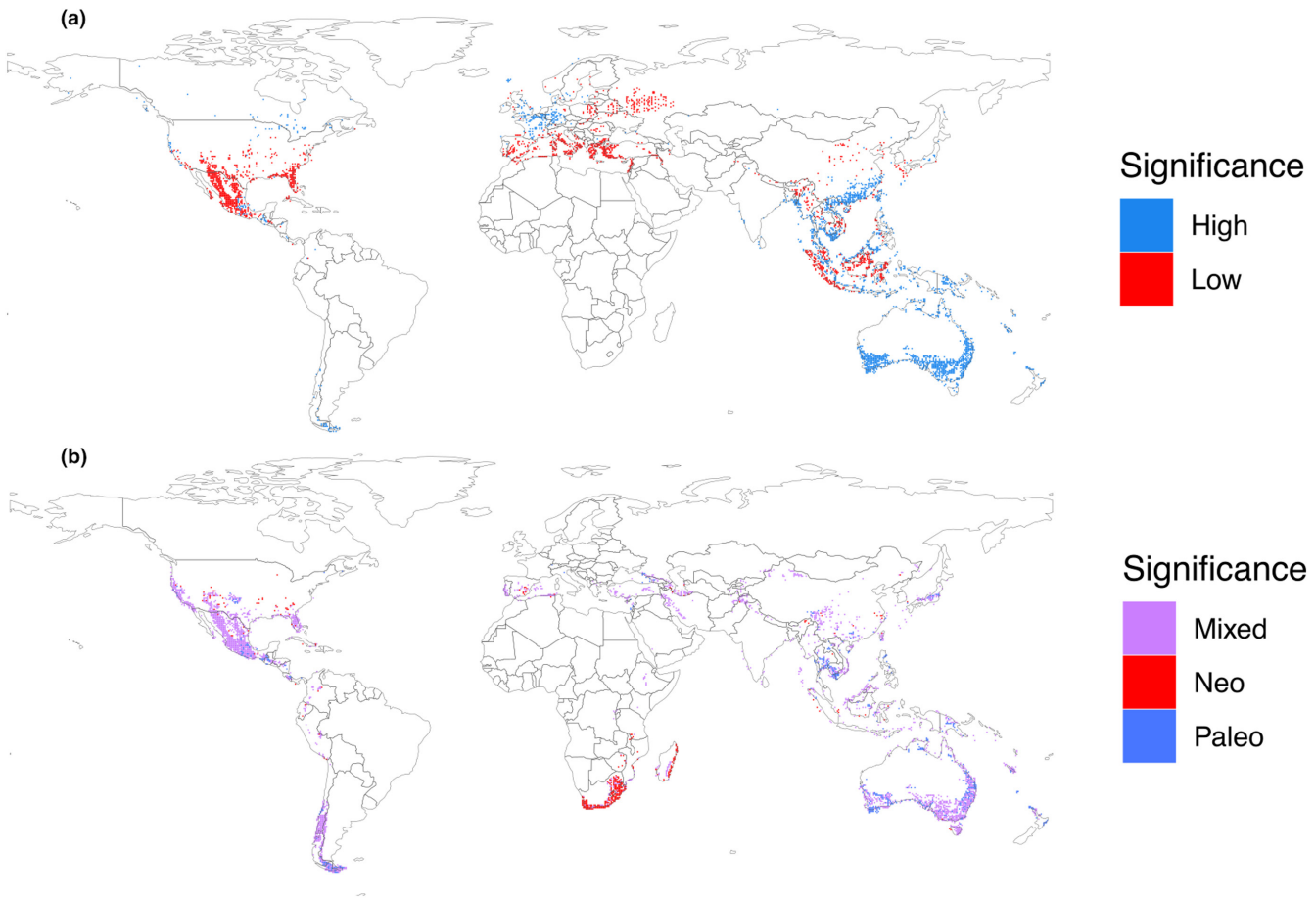


FIGURE 2 (a) Randomisation tests for RPD. "High" significance refers to cells above null expectations; "low" significance refers to cells below null expectations. (b) CANAPE analysis. Interpretation is similar to (a) except that CANAPE also distinguishes centres of mixed endemism, which contain species both above and below null expectations.

3.4 | Distribution of RNS

The distribution of species of Fagales engaged in RNS (Figure 4) is highly heterogeneous, and nodulating species (those displaying the characteristic root structures of RNS) are distributed in areas of low Fagales species richness but associated with areas of high RPD (Figures 1 and 2). A particularly high richness of Fagales nodulators (but not overall species richness) exists in southern Africa (reflecting a local radiation of Myricaceae) and the southern Malesian region and Australia (reflecting the primary distribution of Casuarinaceae); in these primarily semi-arid habitats, no non-nodulating Fagales are known. A secondary area of high RNS proportions occurs across boreal latitudes in the Northern Hemisphere, with lower RNS species diversity than the Southern Hemisphere but more site-level co-occurrence of distinctive lineages from both Betulaceae and Myricaceae. These two RNS areas comprise eight of the 12 recognised phyloregions.

3.5 | Environmental associates

The favoured full mixed model for RPD (ΔAIC = 14.7669; Table S1) had strong explanatory power (conditional $R^2 = 0.8656$), but the fixed environmental effects only had weak explanatory power for RPD (marginal $R^2 = 0.0211$). Nevertheless, models with environmental predictors are strongly favoured (best model vs. no-environment model, ΔAIC = 619.859). Comparing among the predictors with normalised coefficients, the most important predictor was aridity (positive relationship, meaning higher RPD in more mesic sites), followed by soil carbon and pH (also both positive relationships; Table S2).

Models of environmental factors vs. CANAPE were also implemented to understand whether phylogenetic endemism displayed similar responses. The favoured model had a moderate amount of explanatory power (marginal $R^2 = 0.3007$; conditional $R^2 = 0.6388$), where the complex mixed model was favoured (ΔAIC = 31.43719). Among these, arid environments were consistently associated with all forms of endemism significance (Figure 6a; significant endemism associated with more arid sites). Similar to RPD, models with environmental predictors are strongly favoured (best model vs. no-environment model, ΔAIC = 276.6233; Table S1).

The proportion of the community engaging in nodulation had similar explanatory power to RPD (marginal $R^2 = 0.0467$; conditional $R^2 = 0.8063$) with the full mixed model favoured (ΔAIC = 1602.14). Aridity was the most important predictor (Table S2), with more mesic environments associated with more nodulators (Figure 5d).

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

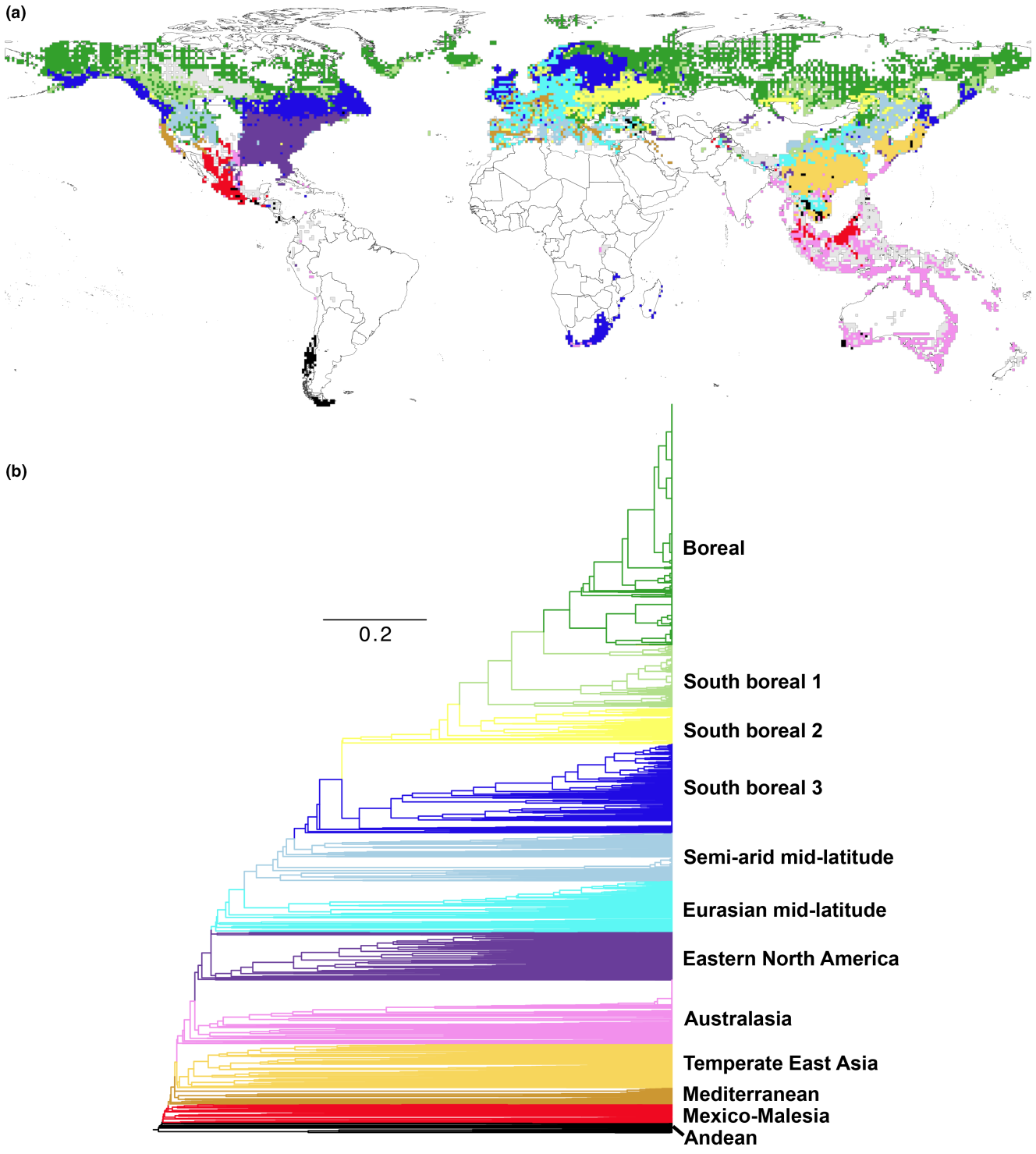


FIGURE 3 (a) Phylogenetic regionalisation. (b) Corresponding dendrogram showing group colorisation and distances. Each terminal represents a grid cell in the map.

Thus, similar aridity responses shape the distribution of nodulation and relative phylogenetic diversity and endemism. The result favouring nodulators in wetter conditions seemed to conflict with high-nodulator sites in arid areas of the Southern Hemisphere (Figure 4), so the analysis was rerun as two separate models partitioning sites above and below the equator. This confirmed that the aridity relationship differed by hemisphere, with the Northern Hemisphere showing a positive relationship (univariate R^2 0.03804, $p < 2.2e\text{-}16$; i.e., more nodulators in more mesic environments) and the Southern Hemisphere a negative relationship (univariate R^2 0.3885, $p < 2.2e\text{-}16$; i.e., more nodulators in more arid environments; see Figure 5d).

In summary, model selection via AIC favoured the full mixed model with latitude and longitude for all responses. Aridity was the most important predictor for all responses, indicating that lineage

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

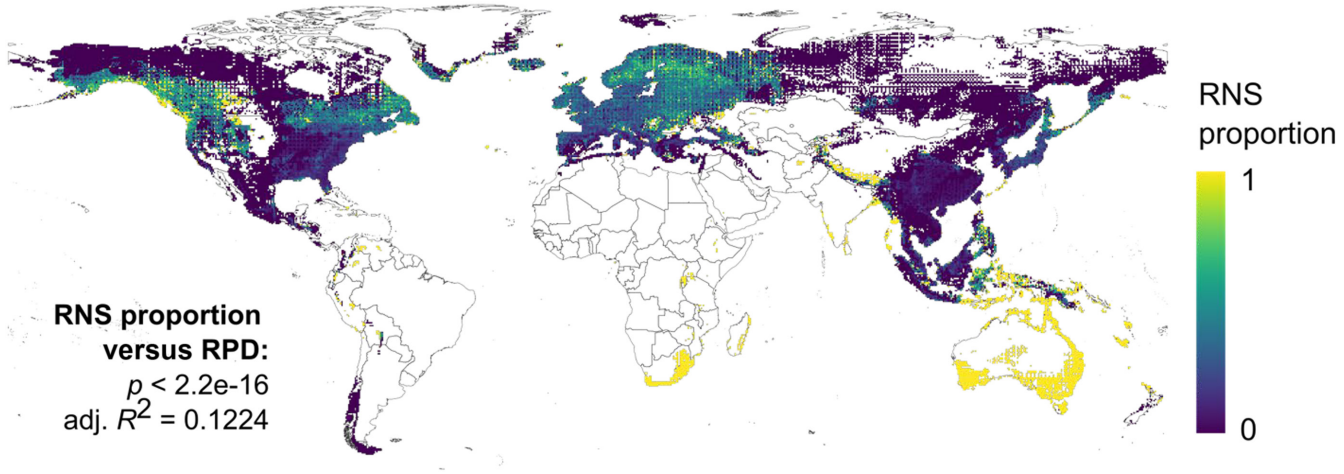


FIGURE 4 Proportion of species within a grid cell engaging in RNS (root nodule symbiosis). Note that areas with high RNS proportions correspond with areas of high RPD (Figure 1b). Significance and correlation (RPD~RNS proportion) is shown in the lower left inset, from a univariate linear model.

diversity, endemism, and RNS distributions are shaped by similar factors. However, the direction of the RNS relationship with aridity differed among hemispheres.

4 | DISCUSSION

4.1 | Distribution of Fagales diversity

A major finding of this study was that while centres of SR for Fagales reflect the distribution and spatial extent of temperate forest, RPD peaks in low SR regions of Fagales distributions in the Southern Hemisphere. These RPD results, with some of the most species-poor areas being highest in RPD, are unanticipated. High RPD was most spatially extensive at relatively high latitudes (Figure 1b). However, Australia and adjacent areas stood out as high-RPD outliers while boreal regions were within null expectations (Figure 2b). Significantly high Southern Hemisphere RPD therefore drives a surprising relationship with latitude, increasing monotonically southwards (Figure 5a), an unusual result given that, in other large global clades, the Southern Hemisphere is often lower in diversity (Economo et al., 2018). Accordingly, most Southern Hemisphere cells with significantly high RPD represent distinct lineages of only one family, Casuarinaceae. This result could be interpreted either as a result of a shared biogeographic history among taxa distributed across the Southern Hemisphere or as evidence for ecological filtering, as southern Africa and Australia (but not South America) primarily consist of nodulating species. Moreover, sites with significantly high RPD show a close correspondence with the distribution of nodulation (see below).

4.2 | Distribution of actinorhizal nodulators

Our results indicate an idiosyncratic distribution of RNS in Fagales, with two primary centres in the Northern and Southern Hemispheres. While occurring in similar latitudes in these separate areas of distribution, the relationship of RNS with environment in Fagales (Figure 1c) was unexpected. Previous studies have supported aridity as best explaining the distribution and diversity of RNS (Pellegrini et al., 2016; Siniscalchi et al., 2022) and especially the phylogenetic diversity of RNS lineages (Doby et al., 2022). A significant shortcoming of many previous studies has been the practice of treating all nodulators equally, which therefore primarily reflects the distribution of the more diverse and prevalent legumes (which are particularly successful in semi-arid habitats; Schrire et al., 2005). This study represents the first in-depth look at the spatial distribution of phylogenetic diversity specifically for actinorhizal RNS species. In Fagales, the global relationship between aridity and RNS distributions was the opposite of previous results (Doby et al., 2022; Pellegrini et al., 2016; Siniscalchi et al., 2022), with wetter environments here favouring a greater prevalence of RNS species. This result makes sense in light of habitat differences between nodulators in each hemisphere. The Southern Hemisphere RNS species of Fagales (Casuarinaceae and Myricaceae) specialize in semi-arid habitats and, although not found in the most arid areas, otherwise fit the spatial pattern seen in legumes. However, lineage diversity and species ranges are larger for Fagales nodulators in the Northern than Southern Hemisphere (Figure 4), and the species in the region (Myricaceae and Betulaceae) are often semi-aquatic.

Segregating the results by Northern vs. Southern Hemisphere confirms that RNS species have differing ecological strategies within Fagales, with more nodulators in more mesic environments in the Northern Hemisphere and more nodulators in more arid environments in the Southern Hemisphere. These modelling results focused on raw RPD but are robust to randomisations: significantly high RPD is strongly associated with more arid environments in the Southern Hemisphere, but with more mesic environments in the Northern Hemisphere (Figure 6b). Similarly, when contrasting sites that either possess or lack RNS plants, arid sites are overrepresented in the Southern Hemisphere but underrepresented in the Northern

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

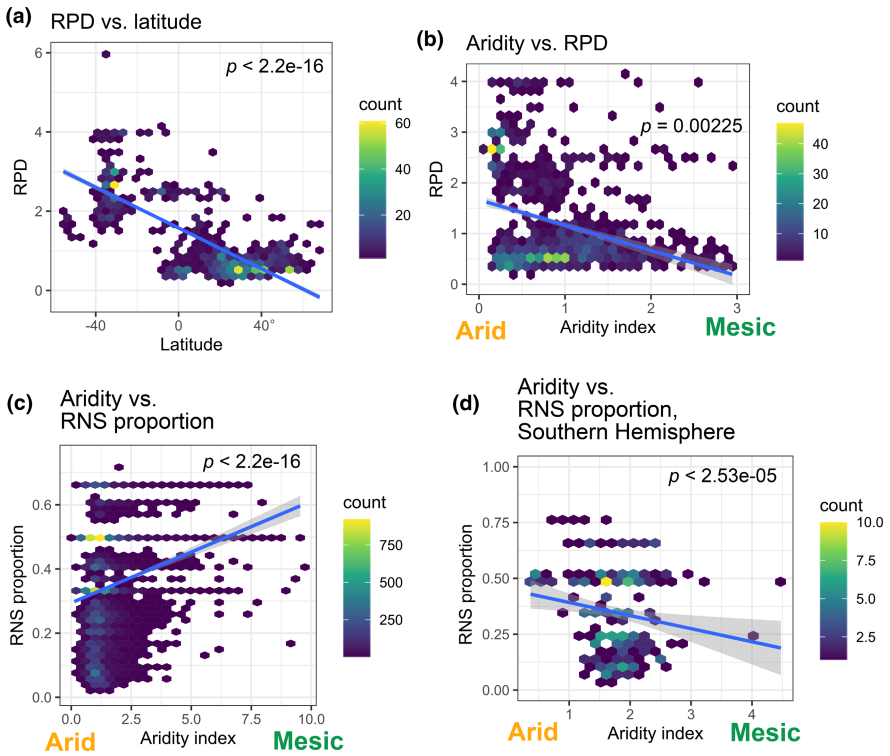


FIGURE 5 Environmental associates of richness and RPD. Hex grids in scatter plots represent aggregations of individual grid cells, with cell density in each grid indicated by the colour scale and regression lines indicated in blue. *p*-values represent predictor significance, in a univariate model (a) or one of the linear mixed models presented in the main text (b–d). (a) RPD vs. latitude; notice the monotonic behaviour of the response and the existence of two clear RPD clusters by latitude. (b) Aridity vs. RPD. (c) Aridity vs. RNS proportion; inflated 0 and 1 values omitted per Methods. (d) Aridity vs. RNS proportion, as in (c) but showing only the Southern Hemisphere.

Hemisphere (Figure 6c). These results suggest that arid specialisation is not universal in RNS species and is conditional on local radiations in different biogeographic provinces with differing ecological contexts and bacterial partnering. Significantly, Africa, an area of the Southern Hemisphere rich in RNS species, was also identified as the only continent with neoendemism, possibly reflecting just such a local radiation. These results accord with already recognised ecological differences between rhizobial and actinorhizal RNS species (Ardley & Sprent, 2021; Folk et al., 2020; Siniscalchi et al., 2022; Tamme et al., 2021) and highlight poorly understood facets of the ecological diversity of RNS species.

4.3 | Phylogenetic regionalisation

Our results are partly consistent with regionalisations previously identified in other clades, identifying Northern-Southern Hemisphere distinctions as a fundamental divide (Carta et al., 2022). The main finding of Carta et al. (2022), an analysis across vascular plants, was that phylogenetic clustering partly supports traditional biogeographic provinces but solely recognises a Northern-Southern Hemisphere divide. We found similar distinct communities in each hemisphere for the Americas and Australasia-Malesia, but we find that the Fagales flora of southern Africa is most similar to floras in boreal regions, together forming an Afro-Boreal flora. Likewise, the Asia/Pacific divide between hemispheres is different in this study as Malesia and Australasia are not recovered as distinct provinces. Malesia had two primary regionalisations, a small phyloregion associated with central America that reflects several endemic genera, and a much larger phyloregion recovered as part of Australasia. The southern rather than equatorial affinity of most of Malesia may be a special feature of Fagales biogeography as it is not generally recovered in vascular-plant-wide studies (Carta et al., 2022), although a partly similar regionalisation was recovered in mosses (Sanbonmatsu & Spalink, 2022). Carta et al. (2022) emphasised historical factors, interpreting distinct Northern vs. Southern Hemisphere floras as reflective of Laurasian vs. Gondwanan radiations, but several families in Fagales likely represent a mix of both Cretaceous and post-Cretaceous radiations and are therefore not completely consistent with this explanation (Larson-Johnson, 2016; Siniscalchi et al., 2023; but see Sauquet et al., 2012; Wheeler et al., 2022). For instance, *Gymnostoma*, one of the taxa spanning Australia and Southeast Asia, likely represents a Miocene or later dispersal with the Sahul-Sunda collision (Crayn et al., 2015). While RNS species showed high-latitude centres of diversity in the Northern and Southern Hemispheres, thus reflecting the separate regionalisation of these areas, phylogenetic regionalisation did not align with RNS distributions. This misalignment indicates that the full distribution of Fagales RNS species comprises independent assemblages of lineages, which likely reflect independent origins of RNS in Fagales (Kates et al., 2024).

In contrast to traditional concepts of temperate diversity, the temperate East Asia and Eastern North America phyloregions are not closely related and form separate, strongly distinct Fagales floras (see Figure 3b dendrogram). The boundaries of these two phyloregions accord well with their traditional definitions; eastern North America is almost identical to Takhtajan's (1986) North American Atlantic region. Our East Asia region is similar to but narrower than Takhtajan's Eastern Asiatic region, including Japan and the

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

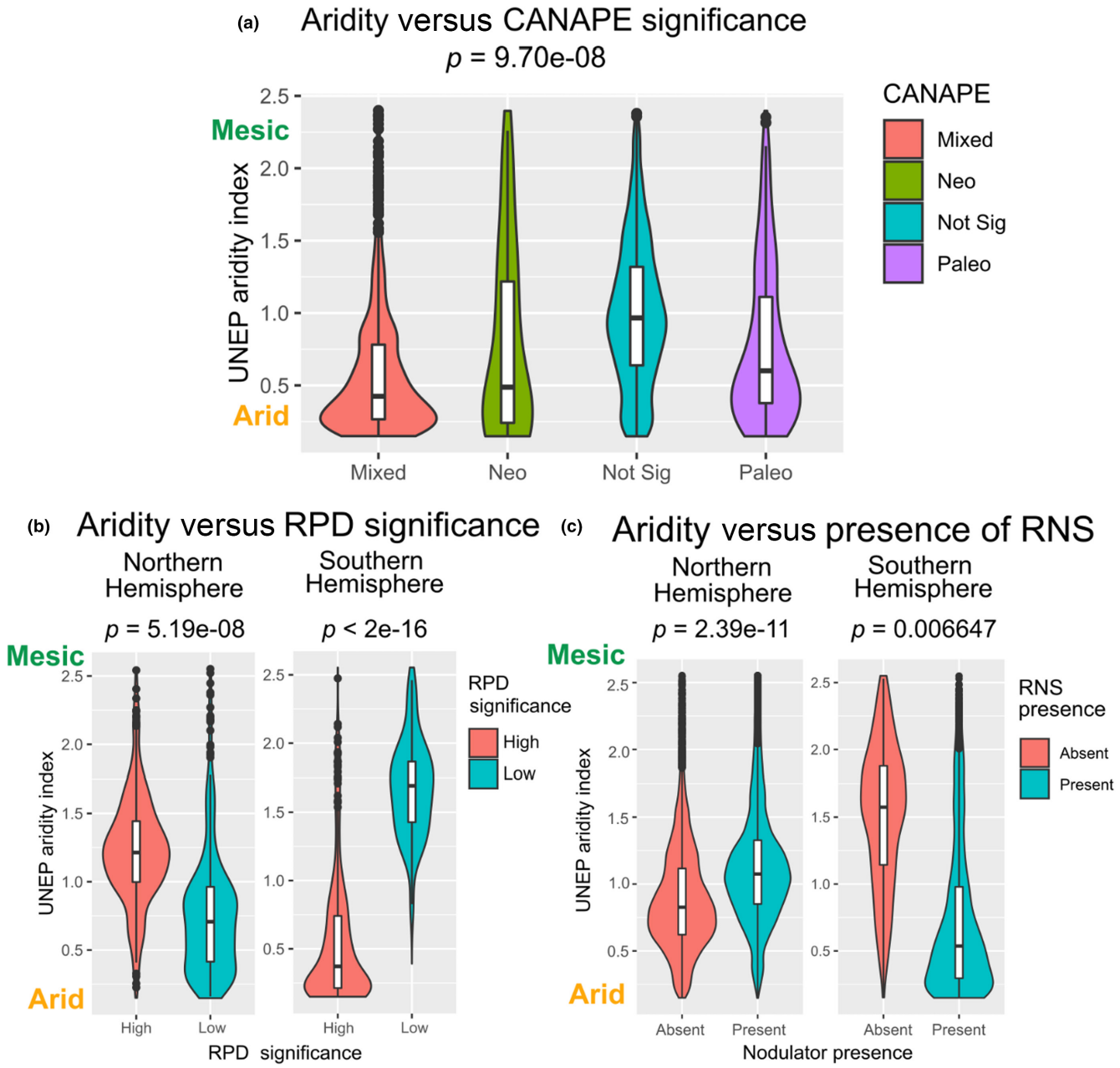


FIGURE 6 (a) Aridity vs. CANAPE significance categories. (b and c) breakdown of aridity relationships between the Northern (left) and Southern (right) Hemispheres. (b) Low RPD significance (branch lengths significantly less than random expectation) vs. aridity. (c) Site-based presence/absence of nodulators vs. aridity. *p*-values, noted in graph titles, represent predictor significance from linear mixed models.

Korean peninsula but excluding northeastern China, Sakhalin, and the Ryukyus, and including only portions of central Hokkaido. This north–south delimitation is very similar to a recent phyloregionalisation in China (Ye et al., 2019), but there was no sign of a major divide between eastern and western China (cf. Lu et al., 2018; Ye et al., 2019).

The main distinction of SR in Fagales compared to traditionally recognised centres of temperate forest diversity is the high diversity of montane equatorial areas in Central America and Malesia, more usually considered relictual and marginal areas of (climatically) temperate diversity (Miranda & Sharp, 1950). Our finding agrees, however, with the distribution of many isolated lineages in Fagales treated as species-poor or monotypic taxa. The montane Central American flora and much of the Malesia region were recovered as a single phyloregion, reflecting shared Fagales diversity between these distant regions. The remainder of Malesia is primarily part of an Australasian region, reflecting Casuarinaceae and other taxa primarily distributed in the Southern Hemisphere and reaching their northern limit in this area; this regionalisation solely reflects RNS species. Mainland Southeast Asia is similar to Malesia, with a strong north-Asian component and some influence from temperate East Asia, but also is a mosaic of other phyloregions including (most prominently) Australasian influence.

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Western and central Europe formed an additional hotspot of SR that could represent sampling intensity bias. However, high SR in this area has been identified before in Fagales (Lyu et al., 2022) and is also reflected in a recent investigation of actinorhizal RNS species including Fagales (Tamme et al., 2021: Figure 2c). Species richness patterns recovered here overall compare closely to those reported in Fagales in a recent methodological contribution (Lyu et al., 2022). Although the framework we present is not taxonomically complete, with 60% to 90% coverage of species depending on the analysis, we consider these results likely to be robust in terms of sampling, strict occurrence filtering criteria and presence/absence thresholding, and extensive model curation.

5 | CONCLUSIONS

Investigations of Fagales species richness and phylogenetic diversity support traditionally recognised hotspots for north temperate forests (East Asia, eastern North America), but also novel hotspots including Malesia and central and southern Mexico, not traditionally considered centres of ancient diversity for temperate trees. RNS in Fagales is particularly rich in mid- to high latitudes of the Northern and Southern Hemispheres and aligns with RPD rather than overall SR, pointing to a role for ecological filtering. Underlining the role of filtering, the geographic distribution of RNS in Fagales reflects three independent evolutionary origins (Kates et al., 2024), which is precisely the prediction under phylogenetic overdispersion (Cavender-Bares et al., 2004). Potential drivers of the distributions of nodulating plants disagree among these regions and with previous work, suggesting disparate ecological strategies in different biogeographic areas. Overall, our results point to the importance of clade-level investigations of phylogenetic diversity for investigating evolutionary processes (Cavender-Bares, 2019).

ACKNOWLEDGEMENTS

This work was funded by the National Science Foundation (DEB-1916632). No permits were required for this work.

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

Scripts to perform the analyses, as well as raw grid cell products, are available on GitHub (https://github.com/ryanafolk/fagales_phylodiversity). A static permanent version is deposited on Zenodo (https://doi.org/10.5281/zenodo.10428046).

ORCID

R. A. Folk https://orcid.org/0000-0002-5333-9273

REFERENCES

Abbott, J. C., Bota-Sierra, C. A., Guralnick, R., Kalkman, V., González-Soriano, E., Novelo-Gutiérrez, R., Bybee, S., Ware, J., & Belitz, M. W. (2022). Diversity of nearctic dragonflies and damselflies (Odonata). *Diversity*, *14*(7), 575.

Akaike, H. (1974). A new look at the statistical model identification. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Springer Series in Statistics* (pp. 215–222). Springer. https://doi.org/10.1007/978-1-4612-1694-0_16

Allen, J. M., Germain-Aubrey, C. C., Barve, N., Neubig, K. M., Majure, L. C., Laffan, S. W., Mishler, B. D., Owens, H. L., Smith, S. A., Whitten, W. M., Abbott, J. R., Soltis, D. E., Guralnick, R., & Soltis, P. S. (2019). Spatial phylogenetics of Florida vascular plants: The effects of calibration and uncertainty on diversity estimates. *iScience*, *11*, 57–70.

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., & Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, *5*, 180040.

APG IV (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, *181*(1), 1–20. Portico. https://doi.org/10.1111/boj.12385

Ardley, J., & Sprent, J. (2021). Evolution and biogeography of actinorhizal plants and legumes: A comparison. *The Journal of Ecology*, *109*(3), 1098–1121.

Axelrod, D. I. (1966). Origin of deciduous and evergreen habits in temperate forests. *Evolution: International Journal of Organic Evolution*, *20*(1), 1–15.

Axelrod, D. I. (1983). Biogeography of oaks in the Arcto-Tertiary Province. *Annals of the Missouri Botanical Garden*, *70*(4), 629–657.

Batjes, N. H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., & Mendes de Jesus, J. (2017). WoSIS: Providing standardised soil profile data for the world. *Earth System Science Data*, *9*(1), 1–14.

Carta, A., Peruzzi, L., & Ramírez-Barahona, S. (2022). A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *The New Phytologist*, *233*(3), 1494–1504.

Cavender-Bares, J. (2019). Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. *The New Phytologist*, *221*(2), 669–692.

Cavender-Bares, J., Ackerly, D. D., Baum, D. A., & Bazzaz, F. A. (2004). Phylogenetic overdispersion in Floridian oak communities. *The American Naturalist*, *163*(6), 823–843.

Cook, L. G., & Crisp, M. D. (2005). Not so ancient: The extant crown group of *Nothofagus* represents a post-Gondwanan radiation. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1580), 2535–2544.

Crayn, D. M., Costion, C., & Harrington, M. G. (2015). The Sahul-Sunda floristic exchange: Dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. *Journal of Biogeography*, *42*(1), 11–24.

Daru, B. H., Elliott, T. L., Park, D. S., & Davies, T. J. (2017). Understanding the processes underpinning patterns of phylogenetic regionalization. *Trends in Ecology & Evolution*, *32*(11), 845–860.

Davis Rabosky, A. R., Cox, C. L., Rabosky, D. L., Title, P. O., Holmes, I. A., Feldman, A., & McGuire, J. A. (2016). Coral snakes predict the evolution of mimicry across New World snakes. *Nature Communications*, *7*, 11484.

Doby, J. R., Li, D., Folk, R. A., Siniscalchi, C. M., & Guralnick, R. P. (2022). Aridity drives phylogenetic diversity and species richness patterns of nitrogen-fixing plants in North America. *Global Ecology and Biogeography: A Journal of Macroecology*, *31*, 1630–1642. https://doi.org/10.1111/geb.13535

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, *105*(Suppl 1), 11549–11555. https://doi.org/10.1073/pnas.0801962105

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Economo, E. P., Narula, N., Friedman, N. R., Weiser, M. D., & Guénard, B. (2018). Macroecology and macroevolution of the latitudinal diversity gradient in ants. *Nature Communications*, *9*(1), 1778.

Edwards, E. J., Chatelet, D. S., Chen, B.-C., Ong, J. Y., Tagane, S., Kanemitsu, H., Tagawa, K., Teramoto, K., Park, B., Chung, K.-F., Hu, J.-M., Yahara, T., & Donoghue, M. J. (2017). Convergence, consilience, and the evolution of temperate deciduous forests. *The American Naturalist*, *190*(S1), S87–S104.

Engler, A. (1905). Die pflanzengeographische Gliederung Nordamerikas erläutert an der nordamerikanischen Anlage des neuen Königlichen botanischen Gartens zu Dahlem-Steglitz bei Berlin, mit einer Verbreitungskarte und einem Orientierungsplan. *Notizblatt Des Königl. Botanischen Gartens Und Museums Zu Berlin*, *3*, 1–94.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315.

Folk, R. A., Kates, H. R., LaFrance, R., Soltis, D. E., Soltis, P. S., & Guralnick, R. P. (2021). High-throughput methods for efficiently building massive phylogenies from natural history collections. *Applications in Plant Sciences*, *9*(2), e11410.

Folk, R. A., Siniscalchi, C. M., & Soltis, D. E. (2020). Angiosperms at the edge: Extremity, diversity, and phylogeny. *Plant, Cell & Environment*, *43*(12), 2871–2893.

GBIF.org. (2020). GBIF occurrence download. https://doi.org/10.15468/dl.am52xh

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). Dismo: Species distribution modeling. R package version 1.1-4.

Hinojosa, L. F., Gaxiola, A., Pérez, M. F., Carvajal, F., Campano, M. F., Quattrocchio, M., Nishida, H., Uemura, K., Yabe, A., Bustamante, R., & Arroyo, M. T. K. (2016). Non-congruent fossil and phylogenetic evidence on the evolution of climatic niche in the Gondwana genus *Nothofagus*. *Journal of Biogeography*, *43*(3), 555–567.

Kates, H. R., O'Meara, B. C., LaFrance, R., Stull, G. W., James, E. K., Conde, D., Liu, S., Tian, Q., Yi, T., Kirst, M., Ané, J.-M., Soltis, D. E., Guralnick, R. P., Soltis, P. S., & Folk, R. A. (2024 [accepted]). Two shifts in evolutionary lability underlie independent gains and losses of root-nodule symbiosis in a single clade of plants. *Nature Communications*.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455.

Laffan, S. W., Lubarsky, E., & Rosauer, D. F. (2010). Biodiverse, a tool for the spatial analysis of biological and related diversity. *Ecography*, *33*(4), 643–647.

Larson-Johnson, K. (2016). Phylogenetic investigation of the complex evolutionary history of dispersal mode and diversification rates across living and fossil Fagales. *The New Phytologist*, *209*(1), 418–435.

Li, D., Olden, J. D., Lockwood, J. L., Record, S., McKinney, M. L., & Baiser, B. (2020). Changes in taxonomic and phylogenetic diversity in the Anthropocene. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1929), 20200777.

Li, D., Trotta, L., Marx, H. E., Allen, J. M., Sun, M., Soltis, D. E., Soltis, P. S., Guralnick, R. P., & Baiser, B. (2019). For common community phylogenetic analyses, go ahead and use synthesis phylogenies. *Ecology*, *100*(9), e02788.

Li, R., Chen, Z., Lu, A., Soltis, D. E., Soltis, P. S., & Manos, P. S. (2004). Phylogenetic relationships in Fagales based on DNA sequences from three genomes. *International Journal of Plant Sciences*, *165*(2), 311–324.

Lu, L.-M., Mao, L.-F., Yang, T., Ye, J.-F., Liu, B., Li, H.-L., Sun, M., Miller, J. T., Mathews, S., Hu, H.-H., Niu, Y.-T., Peng, D.-X., Chen, Y.-H., Smith, S. A., Chen, M., Xiang, K.-L., Le, C.-T., Dang, V.-C., Lu, A.-M., … Chen, Z.-D. (2018). Evolutionary history of the angiosperm flora of China. *Nature*, *554*(7691), 234–238.

Lyu, L., Leugger, F., Hagen, O., Fopp, F., Boschman, L. M., Strijk, J. S., Albouy, C., Karger, D. N., Brun, P., Wang, Z., Zimmermann, N. E., & Pellissier, L. (2022). An integrated high-resolution mapping shows congruent biodiversity patterns of Fagales and Pinales. *The New Phytologist*, *235*(2), 759–772.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *The New Phytologist*, *207*(2), 437–453.

Manchester, S. R. (1999). Biogeographical relationships of north American tertiary floras. *Annals of the Missouri Botanical Garden*, *86*(2), 472. https://doi.org/10.2307/2666183

Manos, P. S., Soltis, P. S., Soltis, D. E., Manchester, S. R., Oh, S.-H., Bell, C. D., Dilcher, D. L., & Stone, D. E. (2007). Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. *Systematic Biology*, *56*(3), 412–430.

Menge, D. N. L., Chisholm, R. A., Davies, S. J., Abu Salim, K., Allen, D., Alvarez, M., Bourg, N., Brockelman, W. Y., Bunyavejchewin, S., Butt, N., Cao, M., Chanthorn, W., Chao, W.-C., Clay, K., Condit, R., Cordell, S., Silva, J. B., Dattaraja, H. S., Andrade, A. C. S., … Fung, T. (2019). Patterns of nitrogen-fixing tree abundance in forests across Asia and America. *The Journal of Ecology*, *107*(6), 2598–2610.

Miranda, F., & Sharp, A. J. (1950). Characteristics of the vegetation in certain temperate regions of eastern Mexico. *Ecology*, *31*(3), 313–333.

Mishler, B. D., Guralnick, R., Soltis, P. S., Smith, S. A., Soltis, D. E., Barve, N., Allen, J. M., & Laffan, S. W. (2020). Spatial phylogenetics of the north American flora. *Journal of Systematics and Evolution*, *58*(4), 393–405.

Mishler, B. D., Knerr, N., González-Orozco, C. E., Thornhill, A. H., Laffan, S. W., & Miller, J. T. (2014). Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian acacia. *Nature Communications*, *5*, 4473.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, *5*, 1198–1205.

Pawlowski, K., & Bisseling, T. (1996). Rhizobial and actinorhizal symbioses: What are the shared features? *The Plant Cell*, *8*(10), 1899–1913.

Pawlowski, K., & Sprent, J. I. (2008). Comparison between actinorhizal and legume symbiosis. In K. Pawlowski & W. E. Newton (Eds.), *Nitrogen-fixing actinorhizal symbioses* (pp. 261–288). Springer.

Pellegrini, A. F. A., Staver, A. C., Hedin, L. O., Charles-Dominique, T., & Tourgee, A. (2016). Aridity, not fire, favors nitrogen-fixing plants across tropical savanna and forest biomes. *Ecology*, *97*(9), 2177–2183.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7), 887–893.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3), 231–259.

Sanbonmatsu, K. K., & Spalink, D. (2022). A global analysis of mosses reveals low phylogenetic endemism and highlights the importance of long-distance dispersal. *Journal of Biogeography*, *49*(4), 654–667.

Sauquet, H., Ho, S. Y. W., Gandolfo, M. A., Jordan, G. J., Wilf, P., Cantrill, D. J., Bayly, M. J., Bromham, L., Brown, G. K., Carpenter, R. J., Lee, D. M., Murphy, D. J., Sniderman, J. M. K., & Udovicic, F. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of *Nothofagus* (Fagales). *Systematic Biology*, *61*(2), 289–313.

Scherson, R. A., Thornhill, A. H., Urbina-Casanova, R., Freyman, W. A., Pliscoff, P. A., & Mishler, B. D. (2017). Spatial phylogenetics of the vascular flora of Chile. *Molecular Phylogenetics and Evolution*, *112*, 88–95.

Schrire, B. D., Lewis, G. P., & Lavin, M. (2005). Biogeography of the Leguminosae. In G. Lewis, B. Schrire, B. Mackinder, & M. Lock

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Eds.), *Legumes of the World* (pp. 21–54). Royal Botanic Gardens Kew.

Schubert, M., Grønvold, L., Sandve, S. R., Hvidsten, T. R., & Fjellheim, S. (2019). Evolution of cold acclimation and its role in niche transition in the temperate grass subfamily Pooideae. *Plant Physiology*, *180*(1), 404–419.

Segovia, R. A., Pennington, R. T., Baker, T. R., Coelho de Souza, F., Neves, D. M., Davis, C. C., Armesto, J. J., Olivera-Filho, A. T., & Dexter, K. G. (2020). Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Science Advances*, *6*(19), eaaz5373.

Siniscalchi, C. M., Correa-Narvaez, J., Kates, H. R., Soltis, D. E., Soltis, P. S., Guralnick, R. P., Manchester, S. R., & Folk, R. A. (2023). Fagalean phylogeny in a nutshell: Chronicling the diversification history of Fagales. *bioRxiv*, 2023.03.06.531381 https://doi.org/10.1101/2023.03.06.531381

Siniscalchi, C. M., Kates, H. R., Soltis, P. R., Soltis, D. E., Guralnick, R. P., & Folk, R. A. (2022). Testing the evolutionary drivers of nitrogen-fixing symbioses in challenging soil environments. *bioRxiv*, 2022.09.27.509719 https://doi.org/10.1101/2022.09.27.509719

Stevens, P. (2001). Angiosperm Phylogeny Website. http://www.mobot.org/MOBOT/research/APweb/

Suissa, J. S., Sundue, M. A., & Testo, W. L. (2021). Mountains, climate and niche heterogeneity explain global patterns of fern diversity. *Journal of Biogeography*, *48*(6), 1296–1308.

Sun, M., Folk, R. A., Gitzendanner, M. A., Soltis, P. S., Chen, Z., Soltis, D. E., & Guralnick, R. P. (2020). Recent accelerated diversification in rosids occurred outside the tropics. *Nature Communications*, *11*(1), 3333.

Takhtajan, AL. (1986). *Floristic regions of the world*. (Crovello, T.J., trans.; Cronquist, A., Ed.) University of California Press.

Tamme, R., Pärtel, M., Kõljalg, U., Laanisto, L., Liira, J., Mander, Ü., Moora, M., Niinemets, Ü., Öpik, M., Ostonen, I., Tedersoo, L., & Zobel, M. (2021). Global macroecology of nitrogen-fixing plants. *Global Ecology and Biogeography: A Journal of Macroecology*, *30*(2), 514–526.

Thornhill, A. H., Baldwin, B. G., Freyman, W. A., Nosratinia, S., Kling, M. M., Morueta-Holme, N., Madsen, T. P., Ackerly, D. D., & Mishler, B. D. (2017). Spatial phylogenetics of the native California flora. *BMC Biology*, *15*(1), 96.

Wen, J. (1999). Evolution of eastern Asian and eastern north American disjunct distributions in flowering plants. *Annual Review of Ecology and Systematics*, *30*(1), 421–455.

Wheeler, E. A., Baas, P., & Manchester, S. R. (2022). Wood anatomy of modern and fossil Fagales in relation to phylogenetic hypotheses, familial classification, and patterns of character evolution. *International Journal of Plant Sciences*, *183*(1), 61–86.

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, *19*(12), 639–644.

Williams-Linera, G., Rowden, A., & Newton, A. C. (2003). Distribution and stand characteristics of relict populations of Mexican beech (*Fagus grandifolia* var. *mexicana*). *Biological Conservation*, *109*(1), 27–36.

Wolfe, J. A. (1975). Some aspects of plant geography of the northern hemisphere during the late cretaceous and tertiary. *Annals of the Missouri Botanical Garden*, *62*(2), 264–279.

Wolfe, J. A. (1987). Late cretaceous-Cenozoic history of deciduousness and the terminal cretaceous event. *Paleobiology*, *13*(2), 215–226.

Yang, Y.-Y., Qu, X.-J., Zhang, R., Stull, G. W., & Yi, T.-S. (2021). Plastid phylogenomic analyses of Fagales reveal signatures of conflict and ancient chloroplast capture. *Molecular Phylogenetics and Evolution*, *107232*, 107232.

Ye, J., Lu, L., Liu, B., Yang, T., Zhang, J., Hu, H., Li, R., Lu, A., Liu, H., Mao, L., & Chen, Z. (2019). Phylogenetic delineation of regional biota: A case study of the Chinese flora. *Molecular Phylogenetics and Evolution*, *135*, 222–229.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Ritter, C. D., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*, 744–751.

BIOSKETCH

Ryan Folk is interested in the relationships and ecology of plants engaging in root nodule symbiosis in the nitrogen-fixing clade (NFC) of flowering plants, using niche modelling, phylogenomics, and other techniques. More information may be sought at https://www.ryanafolk.com/.

Author contributions: RAF, RPG, and MB conceived the study in conversation with all authors. MB performed occurrence record gathering and cleaning, species distribution modelling, randomisation, and clustering analyses. CMS performed phylogenetic analysis. RAF performed site-level statistical modelling and wrote the first draft with MB. All authors contributed to the final draft.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Folk, R. A., Siniscalchi, C. M., Doby, J., Kates, H. R., Manchester, S. R., Soltis, P. S., Soltis, D. E., Guralnick, R. P., & Belitz, M. (2024). Spatial phylogenetics of Fagales: Investigating drivers of temperate forest distributions. *Journal of Biogeography*, *51*, 1518–1532. https://doi.org/10.1111/jbi.14840

13652699, 2024, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14840 by Mississippi State University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License