Global patterns of Rhododendron diversity: The role of evolutionary time and diversification rates

Nawal Shrestha1 | Zhiheng Wang1,2 | Xiangyan Su1 | Xiaoting Xu1,3 | Lisha Lyu1,4 | Yunpeng Liu1 | Dimitar Dimitrov2,5 | Jonathan D. Kennedy2 | Qinggang Wang1 | Zhiyao Tang1 | Xiaojuan Feng6

1Institute of Ecology and Key Laboratory for Earth Surface Processes of the Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing, China
2Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark
3College of Life Science, Sichuan University, Chengdu, China
4School of Urban Planning and Design, Shenzhen Graduate School, Peking University, Shenzhen, China
5Natural History Museum, University of Oslo, Oslo, Norway
6State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China

Correspondence
Zhiheng Wang, Institute of Ecology and Key Laboratory for Earth Surface Processes of the Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. Email: zhiheng.wang@pku.edu.cn

Abstract
Aim: Understanding the evolution of the latitudinal diversity gradient (i.e. increase in species diversity towards the tropics) is a prominent issue in ecology and biogeography. Disentangling the relative contributions of environment and evolutionary history in shaping this gradient remains a major challenge because their relative importance has been found to vary across regions and taxa. Here, using the global distributions and a molecular phylogeny of Rhododendron, one of the largest genera of flowering plants, we aim to compare the relative contributions of contemporary environment, evolutionary time and diversification rates in generating extant species diversity patterns.

Time period: Undefined.

Major taxa studied: Rhododendron.

Methods: We compiled the global distributions of all Rhododendron species, and constructed a dated molecular phylogeny using nine chloroplast genes and seven nuclear regions. By integrating these two datasets, we estimated the temporal trends of Rhododendron diversification, and explored the global patterns of its species diversity, net diversification rates, and species ages. Next, we reconstructed the geographical ancestral area of the clade. Finally, we compared the relative contribution of contemporary environment, time-for-speciation, and diversification rates on the species diversity pattern of Rhododendron.

Results: In contrast to the predictions of the time-for-speciation hypothesis, we found that although Rhododendron originated at a temperate latitude, its contemporary species diversity is highest in the tropics/subtropics, suggesting an into-the-tropics colonization for this genus. We found that the elevated diversification induced by heterogeneous environmental conditions in the tropics/subtropics shapes the global pattern of Rhododendron diversity.

Main conclusions: Our findings support tropical and subtropical mountains as not only biodiversity and endemism hotspots, but also as cradles for the diversification of Rhododendron. Our study emphasizes the need of unifying ecological and evolutionary approaches in order to gain comprehensive understanding of the mechanisms underlying the global patterns of plant diversity.

Keywords
ancestral area reconstruction, evolutionary history, habitat heterogeneity, into-the-tropics, species richness, time-for-speciation, tropical mountains
The latitudinal gradient of species diversity (i.e. the increase in species diversity from the poles to the equator) has intrigued biologists continuously since the early 19th century. The mechanisms underlying this pattern remain one of the most debated issues in ecology and biogeography (Rosenzweig, 1995). One class of hypotheses explains this ubiquitous diversity gradient as a consequence of a variety of factors (e.g. water, energy, habitat heterogeneity, etc.; August, 1983; Currie & Paquin, 1987; Stein, Gerstner, & Kreft, 2014; Stein et al., 2015). However, there is also mounting evidence for the effects of evolutionary history in determining species diversity patterns (Kazak & Wiens, 2012; Latham & Ricklefs, 1993; Scholl & Wiens, 2016; Stephens & Wiens, 2003; Wiens & Donoghue, 2004). In order to fully understand the mechanisms that generate species diversity patterns, we must evaluate how environmental factors interact with evolutionary processes to produce geographical gradients in species diversity (Wiens, Graham, Moen, Smith, & Reeder, 2006).

One notable environmental hypothesis explains species diversity patterns as being predicted by habitat heterogeneity (Kerr & Packer, 1997; Pianka, 1967; Stein et al., 2014, 2015), in that species’ coexistence largely depends upon the availability of different habitats. According to this hypothesis, areas with a greater number of habitats provide more niche space and thus more opportunity for species diversification (Stein et al., 2014, 2015). There is ample empirical evidence from different taxonomic groups supporting positive diversity–habitat heterogeneity relationships (Hortal, Triantis, Meiri, Thébault, & Sfenthourakis, 2009; Tews et al., 2004), although the mechanisms underlying these relationships remain controversial (Stein et al., 2014).

Species diversity patterns are ultimately generated by the processes of speciation, extinction and dispersal that occur over evolutionary time-scales (Wiens & Donoghue, 2004) and species diversity–environment relationships may have originated as a consequence of the evolutionary histories of different clades (Ricklefs, 2006a). With the increasing use of phylogenetic data, two classic hypotheses explaining species diversity patterns in terms of evolutionary processes have regained wide attention in recent literature: the time-for-speciation hypothesis (TSH; Stephens & Wiens, 2003) and the diversification rate hypothesis (Condamine, Sperling, Wahlberg, Rasplus, & Kergoat, 2012). According to the TSH, the regions occupied by a clade for a longer time tend to accumulate more species. The TSH has often been used to explain the high tropical diversity of many groups (see Hawkins & DeVries, 2009; Hawkins, Diniz-Filho, Jaramillo, & Soeller, 2007; Latham & Ricklefs, 1993; Wiens & Donoghue, 2004; Wiens et al., 2006). However, the inverse latitudinal gradient of species diversity (diversity being highest in temperate regions) exhibited by some taxa originating in temperate areas (e.g. temperate taxa of frogs (Smith, Stephens, & Wiens, 2005) and snakes (Pyron & Burbrink, 2009)) suggests that the TSH is not restricted solely to explaining high tropical diversity (Supporting Information Figure S1). In contrast to the TSH, other studies based on palaeontological and phylogenetic data suggest that elevated diversification rates from the poles towards the tropics result in high tropical species diversity, that is, the diversification rate hypothesis (Condamine et al., 2012). Although recent studies suggest that high tropical diversification might be due to increased opportunities for reproductive isolation (Mittelbach et al., 2007) or faster molecular evolution (Allen, Gillooly, Savage, & Brown, 2006), the influences of environmental factors on the latitudinal gradient of diversification rate remain contentious. Moreover, previous studies have tested the effect of net diversification and time-for-speciation in isolation, and their relative contribution in comparison with the effects of environment on species diversity remain to be established.

Here, we evaluate the relative effects of contemporary environment, time-for-speciation, and diversification rates on large-scale patterns of plant diversity using one of the largest flowering plant genera *Rhododendron*. *Rhododendron* comprises > 1,000 species globally, among which c. 70% are classified to be vulnerable, threatened, endangered or critically endangered (Gibbs, Chamberlain, & Argent, 2011). We compiled the global distributions of extant *Rhododendron* species (see Supporting Information Appendix S3), and constructed a dated molecular phylogeny using nine chloroplast genes and seven nuclear regions (see Supporting Information Appendix S5). By integrating these two datasets, we estimated the temporal trends of *Rhododendron* diversification and the global patterns of its species diversity, net diversification rates and species ages. In addition to these analyses, we also reconstructed the geographical ancestral area of the clade. We then compared the relative predictive power of contemporary environment, time-for-speciation, and net diversification rates upon the species diversity patterns of *Rhododendron*, and evaluated the association between net diversification rates and environmental variables. We observed an ‘into-the-tropics’ pattern of colonization for *Rhododendron* and high diversification rates in the tropics/subtropics, which resulted in high tropical/subtropical diversity with relatively young species. Our findings suggest that the geographical species diversity gradient of the genus is likely determined by elevated diversification in the tropics/subtropics. Our study challenges the TSH and demonstrates an integrative framework incorporating ecological and evolutionary mechanisms to explain the global patterns of *Rhododendron* diversity.

## 2 | METHODS

### 2.1 | *Rhododendron* distributions

The global distributions of *Rhododendron* species were compiled from published continental and regional floras, online databases and available peer-reviewed papers (see Supporting Information Appendix S1 for the list of data sources). Some data sources provide coordinates of species occurrences, while most of them recorded species distributions at the level of administrative units, which differ in scale (e.g. villages, towns, cities, provinces/states and countries) or in physical geographical units (e.g. mountains and nature reserves). These records were georeferenced with the Global Administrative Areas Database (http://www.gadm.org/), which provides locations and boundaries of geopolitical units. To eliminate the effect of area on diversity estimation, smaller geopolitical units were merged together to create units of similar size (c. 150,000 km² ~ 4° longitude x 4° latitude). The final map divided
the global land area into 480 geographical units (see Supporting Information Figure S2 for the map). The final distributional database contained records of 1,010 Rhododendron species (see Supporting Information Appendix S3 for the list of species), which also included species of the formerly recognized genera Menziesia and Therorhodion, which are now recognized as members of Rhododendron (Craven, 2011; Goetsch, Eckert, & Hall, 2005).

2.2 | Environmental variables

Climatic data included mean annual temperature (MAT) and precipitation (MAP), mean temperature of the coldest quarter (MTCQ) and the warmest quarter (MTWQ), precipitation of the driest quarter (PDQ), annual range of temperature (ART), temperature seasonality (TSN), precipitation seasonality (PSN), potential evapotranspiration (PET), annual actual evapotranspiration (AET), moisture index (MI), warmth index (WI) and water deficit (WD). The data for MAT, MAP, MTCQ, MTWQ, ART, TSN and PSN were downloaded at a spatial resolution of 30 arc seconds (~1 km at the equator) from WorldClim (http://www.worldclim.org/). MI, PET and AET were calculated following Thornthwaite and Hare (1955). WD was calculated as the difference between PET and AET (Francis & Currie, 2003). WI represents the growing season warmth, and was calculated as Σ(T − 5), where T is the monthly mean temperature for months with T > 5 °C (Fang & Lechowicz, 2006). The climatic variable for each geographical unit was estimated by averaging all 1 km × 1 km cells within it.

Habitat heterogeneity variables included standard deviation of elevation (ELE_STD), and the range of MAT (MATR) and MAP (MAPR). ELE_STD was calculated as the standard deviation of elevation within each geographical unit using the GTOPO30 digital elevation model (https://lta.cr.usgs.gov/GTOPO30). MATR and MAPR were calculated as the differences between the maximum and minimum MAT and MAP within each geographical unit, respectively, and were used to represent climatic heterogeneity (Wang, Fang, Tang, & Lin, 2011).

All environmental variables were grouped into four categories, which represented (a) environmental energy (MAT, MTWQ, MTCQ, WI, PET); (b) water availability (MAP, PDQ, AET, MI, WD); (c) climate seasonality (ART, TSN, PSN); and (d) habitat heterogeneity (ELE_STD, MATR, MAPR).

Because soil conditions affect the growth of plants, we also used two soil variables in our analyses, namely soil organic carbon content (g/kg) and soil pH at 30 cm depth. The data of the soil variables were obtained from the SoilGrid database (https://www.soilgrids.org/) at a spatial resolution of 250 m. The soil parameters for each geographical unit were then estimated by averaging all 250 m × 250 m cells within it using the zonal statistics tool in ArcGIS 10.0 (ESRI, Redlands, CA).

2.3 | Phylogeny reconstruction

We selected nine chloroplast genes (atpB-rbcL, rbcL, matK, ndhF, psbA-trnH, trnL-trnF, trnL, trnT-trnL, trnS-trnG) and seven nuclear regions [internal transcribed spacer (ITS), retinol-binding proteins (RPB2I-1, RPB2I-2, RPB2I-3, RPB2I-4, RPB2I-5, RPB2I-6)] to reconstruct the phylogeny of Rhododendron. Forty-three species from 29 genera closely related to Rhododendron were used as outgroups. The sequence data for the 16 gene regions were exhaustively searched and downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) using NCBI-miner (Xu, Dimitrov, Rahbek, & Wang, 2015). In total, we found sequences for 467 Rhododendron species from all continents (c. 46.2% of the global Rhododendron species diversity). The sequences of each gene were aligned separately with MAFFT v. 7 (Katoh & Standley, 2013) using the L-INS-i algorithm, and were subsequently concatenated into a super matrix. After removing species with fewer than two genes or 1500 bp to reduce the amount of missing data, the super matrix contained 423 species × 16 gene regions (see Supporting Information Appendix S4 for GenBank accession numbers). The phylogeny was then reconstructed in BEAST v. 1.8 (Drummond, Suchard, Xie, & Rambaut, 2012) using a relaxed molecular clock, uncorrelated lognormal substitution rates and a discrete gamma model with four rate categories. The analysis was partitioned by genes/gene regions and the substitution model of molecular evolution for each partition was set to Generalised Time Reversible with Gamma distribution (GTR + I). We used both yule and birth/date speciation as priors in the phylogeny reconstruction (see Supporting Information Appendix S1 for details).

Three calibration points were used to date the phylogeny. (a) Following Popp, Mirró, and Brochmann (2011); Liu, Jolles, Zhou, Peng, and Milne (2014); Schwery et al. (2015), the fossil seeds of Rhododendron newburyanum (Collinson & Crane, 1978) dated to the Palaeocene (c. 54.5 Ma) were used to set the minimum age of the Rhododendron crown group [priors for time to the most recent common ancestor (tMRCA): lognormal distribution with logmean = 0, lognormal SD = 1 and offset = 54.5]. (b) Following Liu et al. (2014) and Schwery et al. (2015), leaves of Vaccinium creedens dated to 26.5 Ma (Axelrod, 1987) were used as a prior for the minimum age of the Vaccinium crown group (priors for tMRCA: lognormal distribution with logmean = 0, lognormal SD = 1 and offset = 26.5). (c) The divergence of Rhododendron from its closest relatives was set to have a normal distribution with a mean = 63.5 Ma and SD = 10.0 following Popp et al. (2011) (see Supporting Information Appendix S1 for details). The use of the Rhododendron seed fossil as a crown calibration is based on the consensus of available evidence (Liu et al., 2014; Popp et al., 2011; Schwery et al., 2015). Nonetheless, we also performed a preliminary dating analysis, using this fossil as a stem calibration of Rhododendron, which resulted in a crown age of Rhododendron < 45 Ma (see Supporting Information Figure S3). This age conflicts with the fossil evidence (Collinson, Hooker, & Grocke, 2003; Liu et al., 2014; Popp et al., 2011; Schwery et al., 2015) on the origin of Rhododendron. Hence, we do not present the results from the phylogeny where the Rhododendron seed fossil was placed on the stem of the genus Rhododendron. However, we found that the two phylogenies with different placements of the Rhododendron seed fossil generated geographically consistent spatial patterns of mean species ages and recovered the same biogeographical region as the possible ancestral area (see Supporting Information Figure S3), suggesting that our main findings are not influenced by the placement of this fossil. The global phylogeny of Rhododendron (Supporting
We divided the overall distributional range of *Rhododendron* into six biogeographical regions: A (North America), B (Europe and Caucasus), C (north-east Asia), D (south Asia), E (Malay Archipelago) and F (Australia) (see Figure 2b). This division was mainly based on areas of endemism for *Rhododendron* species and the existence of water barriers between different continents. Next, we used two methods to reconstruct the possible ancestral ranges and dispersal of *Rhododendron*: (a) statistical dispersal–vicariance analysis (S-DIVA) (Nylander, Olsson, Alström, & Sanmartín, 2008) and (b) likelihood analysis of geographical range evolution (Lagrange) implementing the dispersal–extinction–cladogenesis (DEC) model (Ree & Smith, 2008). As an expansion of Bayes-DIVA (Nylander et al., 2008), S-DIVA is a commonly used event-based method of biogeographical inference, and estimates the optimized number of dispersal and extinction events to explain contemporary distributional patterns. In contrast with many other biogeographical methods, S-DIVA explicitly uses a posterior distribution of trees to account for both phylogenetic uncertainty and uncertainty in ancestral states (Yu, Harris, Blair, & He, 2015). Unlike S-DIVA, Lagrange calculates the relative likelihood of possible ancestral ranges at each node, given a particular probability of dispersal and extinction. Both methods were implemented in RASP (reconstruct ancestral states at each node, given a particular probability of dispersal and extinction) v. 3.2 (Yu et al., 2015) using the dated phylogeny referred to above using two different methods: residual maximum likelihood (REML) and Felsenstein’s phylogenetic independent contrasts (PIC) (Felsenstein, 1985). REML first estimates the ancestral value at the root, and then subsequently the variance of the Brownian motion process is estimated by optimizing the residual log-likelihood. The ancestral values are inferred from the likelihood function given these two parameters. PIC is also a Brownian-motion based estimator, but takes only descendants of each node into account when reconstructing the ancestral state. The MAT and MTCQ of each species were calculated as the average of the values of all geographical units where the species is distributed. Both analyses were implemented in the R package ape (Paradis, Claude, & Strimmer, 2004), and the two methods generated consistent results.

### 2.5 Diversification rate and mean species age

We used the Bayesian analysis of macro-evolutionary mixtures (BAMM) to assess the historical diversification dynamics of *Rhododendron*. BAMM can infer diversification rate variations through time using reversible-jump Markov chain Monte Carlo (MCMC) to explore the best diversification models under a Bayesian framework. In addition, BAMM can account for extinction and incorporate incomplete taxon sampling directly into the likelihood calculations. We used BAMLMMODELS (Rabosky et al., 2014) to plot diversification rates through time for the six biogeographical regions. Next, we extracted the diversification rate and age of each species and calculated the average diversification rate and age of all species within each geographical unit. Subsequently, we arranged species in ascending order of their ages and divided them into four equal-sized groups (i.e. four age quartiles). The first quartile included species younger than 1.28 Myr, while the fourth quartile included species older than 6.65 Myr. We then calculated the proportions of species from each quartile in terms of the total species diversity within each geographical unit to show geographical variation in the proportions of old and young species (Supporting Information Figure S4). In order to validate the results from BAMM, we directly calculated the diversification rates following Magallón and Sanderson (2001) for two subclades corresponding to the Himalayan and Malayan species, respectively. We found the same spatial pattern of diversification rates as reported by BAMM. The Himalayan and the Malayan clades have higher diversification rates than the average diversification rates of all species under low and high extinction scenarios (see Supporting Information Table S2).

### 2.6 Statistical analyses

First, we used simple generalized linear models (GLMs) and simultaneous autoregressive (SAR) models to evaluate the effects of the 18 individual environmental variables on the geographical patterns in species diversity and net diversification rates. To eliminate the effects of spatial autocorrelation on the significance tests of GLM regressions, we used modified t tests (Dutilleul, Clifford, Richardson, & Hemon, 1993).

Second, we built two multiple SAR models of species diversity and mean net diversification rates against the combination of four environmental categories: energy, water availability, habitat heterogeneity and climate seasonality. We excluded soil variables in the multiple models because they had no significant effects on the geographical patterns of either species diversity or net diversification rates (see Supporting Information Table S1). Pearson correlations among all environmental variables were calculated to check for multicollinearity. Because each environmental category contained highly correlated variables (see Supporting Information Table S3), we conducted a principal component analysis (PCA) for each environmental category separately and extracted the first principal component, which was then used as the predictor in the two SAR models. The regression coefficients of the...
four predictors in these multiple SAR models were extracted and used to compare the relative effects of the four environmental categories on species diversity and net diversification rate.

Third, to further compare the effects of net diversification rate, mean species age and habitat heterogeneity on the geographical pattern of species diversity, we generated a path model by assuming that net diversification rate and mean species age influence species diversity directly, while habitat heterogeneity could influence species diversity directly and also indirectly via its effects on net diversification rate. To avoid the collinearity between the three habitat heterogeneity variables, we used their first principal component to represent habitat heterogeneity.

The correlations and GLM regressions were performed in R 3.4.1 (R Core Team, 2017), while the SAR models were built in SPATIAL ANALYSIS IN MACROECOLOGY (SAM) v. 4 (Rangel, Diniz-Filho, & Bini, 2010). The path models were built in IBM SPSS AMOS 22 (Amos Development Corporation, Pennsylvania, USA).

3 | RESULTS

3.1 | Geographical patterns of Rhododendron diversity

The pattern of Rhododendron species diversity is, in general, consistent with the distribution of mountains. The greatest diversity is found in the mountainous region extending from the southern Himalayas up to south-western China (Figure 1a), including Nepal, Bhutan, north-eastern India, Myanmar, south-eastern Tibet, Sichuan, and Yunnan. Diversity is slightly lower in central and eastern China. A large number of species occur in the islands between Asia and Australia, particularly in New Guinea. In contrast, southern India, north-east Asia, Europe, Caucasus, and North America have much lower species diversity.

3.2 | Historical ancestral areas and dispersal

The ancestral area reconstruction based on both S-DIVA and DEC methods supports a north-east Asian origin of Rhododendron (Figure 2 and Supporting Information Figure S5), likely occurring at c. 65 Ma. Our analyses indicate multiple dispersal events to other regions (Figure 2). The genus appears to have first dispersed out of north-east Asia into North America in the mid-Eocene, followed by dispersal to South Asia and the Malay Archipelago in the late Eocene. Multiple dispersal events were inferred throughout the late Eocene and Oligocene between north-east Asia and the other regions, leading to paraphyletic assemblages in these regions (Figure 2 and Supporting Information Figure S5). The ancestor of the Australian species likely occurred in the Malay Archipelago first, and dispersed to Australia c. 10 Ma (Figure 2 and Supporting Information Figure S5).

3.3 | Ancestral reconstruction of the thermal niche

The reconstructed ancestral thermal niche based on both REML and PIC methods suggests relatively cool temperatures at most early nodes (Supporting Information Figures S6 and S7). MAT at the root node was reconstructed to be ~7 °C (REML) and ~8 °C (PIC), while the reconstructed MTCQ was ~ ~5 °C (REML) and ~ ~3 °C (PIC) (Supporting Information Figures S6 and S7). In north-east Asia, these temperatures...
now occur in temperate deciduous broad-leaved forests between 40–43° N latitudes, but at much higher northern latitudes c. 65 Ma (Janis, 1993).

3.4 | Geographical patterns of mean species age

Species occurring in tropical and subtropical latitudes, particularly Southeast Asia and the Malay Archipelago (where the overall species diversity is high), appear to be much younger than those occurring at higher latitudes (Figure 1b). The average age of species in the tropics/subtropics appear to be less than 7 Myr. The proportions of the oldest 25% of species are found to be the highest in the northern temperate latitudes, and much lower in the tropics and subtropics (Figure 1c and Supporting Information Figure S4).

3.5 | Spatial and temporal patterns of diversification

The net diversification rate of all Rhododendron species estimated by BAMM increased slowly until c. 30–25 Ma, when an apparent acceleration in diversification occurred (Figure 3b). However, the temporal trends in Rhododendron diversification differ among the biogeographical regions. During the last 30 Myr, the Malay Archipelago and southern Asia exhibited much more rapid diversification than the other biogeographical regions. In the Malay Archipelago, the diversification rate abruptly increased from c. 30 Ma until c. 10 Ma. During this period, the Malay Archipelago had the highest diversification rate among all regions. However, the increase in diversification rate leveled off at c. 8 Ma, leading to a very high but constant diversification rate. This might reflect the effect of stabilized tectonic and orogeny events in Southeast Asia in the past few million years (Hall, 2009). An increase in diversification rate was also observed in southern Asia. However, this increase is estimated to have occurred at c. 18 Ma, much later than in the Malay Archipelago. Diversification in north-east Asia steadily increased throughout the evolutionary history of Rhododendron, and the increase accelerated at c. 10 Ma. In contrast, the diversification rates in North America and Europe/Caucasus show very slight increments throughout the evolutionary history of the clade, and are lowest.
during the last 10 Myr among all regions (Figure 3b). Geographically, the mean net diversification rate of extant species is higher in the tropics and subtropics than in temperate latitudes (Figure 3a).

### 3.6 Effects of contemporary environment, age and net diversification

Of the 18 environmental variables representing climate, habitat heterogeneity and soil conditions, the variables of habitat heterogeneity, particularly MATR and ELE_STD, were consistently the strongest predictors of species diversity in both the GLM and SAR models (Supporting Information Table S1). Variables representing climate seasonality including TSN and ART were the second-best predictors. The explanatory power of energy, water and soil variables on species diversity was low. The multiple SAR models created using the first principal components of the four environmental categories also showed that habitat heterogeneity had the highest effects on species diversity and net diversification rate (Table 1).

The path model involving net diversification rate, habitat heterogeneity and mean age showed that the direct effect of habitat heterogeneity on species diversity was lower ($0.16 \pm 0.07, p < 0.05$) than the direct effect of net diversification rate ($0.52 \pm 0.07, p < 0.001$), and was also lower than the indirect effect of habitat heterogeneity mediated through net diversification rate ($0.29 \pm 0.05, p < 0.001$) (Figure 4). The effect of habitat heterogeneity on diversification rate was stronger ($0.55 \pm 0.06, p < 0.001$) than its direct effect on species diversity. The effect of mean species age on species diversity was insignificant. Therefore, compared with mean species age, net diversification rate had a much stronger effect on species diversity (Figure 4 and Supporting Information Figure S8).

### 4 DISCUSSION

Our results show north-east Asia as the area of origin for *Rhododendron*. The reconstructed ancestral mean annual and winter temperatures at the root of the genus (see Supporting Information Figures S6 and S7) suggest its presence in a cool climate, which corresponds to the present day temperate broad-leaved forests in north-east Asia. Based on palaeontological evidence, temperate broad-leaved forests occurred at much higher latitudes, extending throughout the northern polar regions at around 65 Ma (Janis, 1993) when *Rhododendron* first originated. Therefore, these findings suggest an extra-tropical origin of *Rhododendron*, which is consistent with the age pattern of *Rhododendron* species (Figure 1b,c) and the ability of older species (e.g. members of subsection Pontica) to tolerate extreme coldness (Sakai & Malla, 1981). The fossil record shows that during its early history, *Rhododendron* was widely and more or less continuously distributed across Eurasia and North America (Collinson & Crane, 1978), which further supports the idea that *Rhododendron* might have originated in high northern latitudes. However, climatic cooling in the late Tertiary may...
have considerably reduced vast tracts of *Rhododendron* habitat, causing extinction of old lineages from Europe/Caucasus and North America (Qian & Ricklefs, 2000).

After its origination, lineages of *Rhododendron* likely dispersed on several occasions from north-east Asia to other regions. Dispersal occurred from north-east Asia to North America throughout the late Eocene and Oligocene (Figure 2). This is plausible because Siberia (north-east Asia) and Alaska (North America) were connected by the Bering land bridge throughout this period (Milne & Abbott, 2002). Our results further suggest dispersal from the late Oligocene to early Miocene into Europe/Caucasus. However, because this date is more recent than the age of the fossil species *R. newburyanum* from England (Collinson & Crane, 1978), some palaeo-lineages of *Rhododendron* may have reached Europe/Caucasus soon after the genus originated but subsequently become extinct, and thus recent species may be the result of an independent dispersal at a later time.

*Rhododendron* species colonized the subtropical and tropical mountains and diversified there relatively late in their history. This is consistent with previous findings for subclades of *Rhododendron*. For example, recent studies (Milne, 2004; Milne, Davies, Prickett, Inns, & Chamberlain, 2010) found that the Southeast Asian clades, especially the subgenus *Hymenanthes* (concentrated in southern China and the eastern Himalayas) have diversified more recently and are much younger than those occurring in the temperate regions of North America, south-west Eurasia, and north-east Asia. *Rhododendron*, therefore, exhibits an ‘into-the-tropics’ pattern of dispersal, which although rare, has been reported by studies on New World vertebrates (Kennedy et al., 2014; Smith, Bryson, Houston, & Klicka, 2012) and a few plant families (Chanderbali, van der Werff, & Renner, 2001; Lu-Irving & Olmstead, 2013). It is noteworthy that most *Rhododendron* in tropical and subtropical latitudes occur on mountain slopes, and their affinities for cooler climates might reflect their tendency to retain their ancestral climatic niche.

In contrast to the predictions of the TSH, our results indicate that although *Rhododendron* originated in temperate north-east Asia, its species diversity is relatively low in this region. The south-eastern Himalayas and the Malay Archipelago, in contrast, have the highest *Rhododendron* diversity, harbouring more than 80% of the known species (Ming & Fang, 1990). However, our results suggest that *Rhododendron* colonized these two regions c. 30–40 Ma, well after its origination and also after the time when North America was colonized. This pattern of *Rhododendron* diversity and the inferred time frame for its colonization is in contrast to the predictions of the TSH (Stephens & Wiens, 2003; Stevens, 2006). Along the continuous latitudinal gradient spanning from the Malay Archipelago through south Asia to north-east Asia, older lineages are primarily found at high latitudes, while younger ones are in the tropics and subtropics. Furthermore, mean species age did not predict the species diversity patterns (Figure 4 and Supporting Information Figure S8). These findings suggest that the TSH may not be a strong determinant of *Rhododendron* diversity. Recent studies exploring the relationships between age and species diversity within clades (Ricklefs, 2007; Wiens, 2011) also suggest that the TSH may not regulate species diversity. Rabosky, Slater, and Alfaro (2012) explored the relationship between clade age and species diversity across 1,397 clades of multicellular eukaryotes and found that clade age and species diversity were not significantly correlated. Our results, together with previous findings, suggest that processes other than time-for-speciation should be evaluated to understand the mechanisms underlying large-scale species diversity patterns.

In contrast to our findings, previous studies have shown that time-for-speciation could be a major determinant of the high species diversity of different animal groups, especially reptiles and amphibians with temperate origins (McPeek & Brown, 2007; Pyron & Burbrink, 2009; Smith et al., 2005; Stephens & Wiens, 2003; Wiens et al., 2006). For example, Smith et al. (2005) and Pyron and Burbrink (2009) found that clades of frogs and snakes that originated in temperate latitudes exhibit inverse latitudinal gradients, thus providing support for the TSH. Kozak and Wiens (2012) studied the species diversity pattern of plethodontid salamanders originating in temperate North America and found time-for-speciation to be the sole determinant of salamander diversity in the New World. The discrepancy between our study and previous findings might be due to two reasons. First, *Rhododendron* may have relatively higher net diversification rates and dispersal abilities than reptiles and amphibians, which could have led to a decoupling in the relationship between time-for-speciation and species diversity (Scholl & Wiens, 2016). Second, there is increasing evidence suggesting that the TSH may explain diversity patterns at smaller temporal and spatial scales, while other factors (e.g. diversification rates) determine species diversity at larger scales (Pontarp & Wiens, 2017; Wiens, 2015).

Indeed, the stronger correlation between *Rhododendron* diversity and net diversification indicates that the higher tropical diversity of *Rhododendron* is likely the consequence of higher rates of net diversification in these areas. This is further justified by the spatio-temporal trend of diversification rates across different regions (Figure 3a,b). These findings are consistent with several previous studies that have regarded elevated diversification rates in the tropics as the potential

**FIGURE 4** Path diagram showing interaction among the geographical patterns in habitat heterogeneity, net diversification rates, mean species age and species diversity. The numbers are path coefficients. *ns* = not significant; **p < .05; ***p < .001
driver of higher tropical diversity (Ricklefs, 2006b; Ricklefs, Schwarzbach, & Renner, 2006; Rolland, Condamine, Jiguet, & Morlon, 2014; Svenning, Borchsenius, Bjørnholm, & Balslev, 2008). A recent study based on phylogenetic simulations has shown that net diversification rate is important for explaining species diversity across the Tree of Life, and across different habitats and regions (Kozak & Wiens, 2016). It is noteworthy, however, that previous studies mainly explored the relationships between species diversity and diversification across different clades (e.g. Kozak & Wiens, 2016; Rabosky et al., 2012), and further studies exploring these relationships across habitats would enable us to better understand why species diversify faster in some areas but not elsewhere.

Why are diversification rates higher in the tropics? Previous studies suggest that clades in tropical regions diversify faster due to higher rates of speciation, which are a consequence of faster rates of reproductive isolation (Janzen, 1967), faster molecular evolution (Allen et al., 2006), increased biotic interactions (Fischer, 1960) or greater climatic variation (Dynesius & Jansson, 2000). Alternatively, these same trends may result from lower extinction rates (Fischer, 1960). In the case of *Rhododendron*, the formation of heterogeneous topography might have accentuated this process by creating an array of climatic niches, and providing increased opportunities for geographical isolation (Janzen, 1967; Shrestha, Su, Xu, & Wang, 2018). Of all the environmental variables, habitat heterogeneity variables were the strongest predictors of *Rhododendron* diversity (Hortal et al., 2009; Stein et al., 2014; Tews et al., 2004). We found that, in addition to direct effects, habitat heterogeneity indirectly affects species diversity by influencing net diversification rates. Specifically, the estimated diversification rates (Figure 3b) show significant increases in the two centres of species diversity [i.e. the south-eastern Qinghai–Tibetan Plateau (QTP) and the Malay Archipelago] at c. 30 Ma, which roughly coincides with the period of mountain formation in these regions (Hall, 1996; Harrison, Copeland, Kidd, & Yin, 1992). The development of complex topography following the orogeny events, especially the formation of high mountains and deep canyons in the south-eastern QTP and the Malay Archipelago might have generated heterogeneous habitats, which in turn, might have promoted the diversification of *Rhododendron* through allopatric speciation and adaptation to diverse environmental conditions (Antonelli & Sanmartín, 2011; Hughes & Eastwood, 2006; Xing & Ree, 2017). Similarly, the tectonic uplift and collision beginning at the end of the Miocene that shaped the Japanese Archipelago and its mountains (Barnes, 2003) might be responsible for an increase in diversification rate in north-east Asia (Figure 3b).

Our results differ from previous findings that have shown the limiting effect of water or energy availability on regional species diversity (Currie & Paquin, 1987; O'Brien, 1993). This is evident from the limited ability of energy and water variables to predict *Rhododendron* species diversity. Although previous studies on woody plants (e.g. Wang et al., 2011) have shown stronger effects of winter coldness on their species diversity, we did not find such a tendency here, suggesting that the distribution of *Rhododendron* may not be limited by cold winter temperatures. Indeed, studies on the cold tolerance of *Rhododendron* suggest that most extant *Rhododendron* species grow in cool climates and exhibit winter hardiness of −20 to −30 °C (Sakai & Malla, 1981). *Rhododendron* species might have acquired this cold adaptation from their ancestors that originated in cool climates.

Our phylogeny adequately represents all subgenera and sections and almost all subsections (~95%) of *Rhododendron* (Supporting Information Figure S9). At the species level, however, the sampling covers less than 50% of the extant diversity. Such a high level of incompleteness may raise concerns about the confidence of the inferred diversification rates. In the present analyses the very good representation of all major *Rhododendron* lineages from all biogeographical regions allows us to reconstruct the backbone topology and to use a strategy to deal with missing taxa. The missing species are associated with nodes near the tips of the phylogeny and constitute the majority of Malayan species that are more derived (Goetsch et al., 2011). Therefore, the potential negative effects of incomplete sampling at the species level are unlikely to have a significant effect on our conclusions. The proportion of species included in the phylogeny in each geographical unit is not significantly correlated with longitude (r² = .0007, p > .10) or latitude (r² = .22, p > .10), which suggests that sampled species tend to be randomly distributed across space (Supporting Information Figure S10). Moreover, the mean species age and diversification rate are not correlated with the proportion of sampled species (Supporting Information Figure S11). These results suggest that incomplete sampling in the phylogeny is unlikely to be a major influence on the inferred patterns of mean species age and diversification rates.

5 CONCLUSIONS

Our findings show that *Rhododendron* likely originated in temperate deciduous broad-leaved forests in high northern latitudes in the early Palaeocene. However, despite *Rhododendron*’s origin in temperate latitude, its contemporary species diversity is highest in the tropical and subtropical mountains of southern China and Southeast Asia. Therefore, these results strongly support an ‘into-the-tropics’ pattern of colonization for *Rhododendron*. The accumulation of high species diversity in areas that have been more recently colonized suggests that the time-for-speciation effect may not explain the latitudinal gradient of *Rhododendron* diversity. Instead, increasing diversification rates from temperate latitudes towards the tropics induced by heterogeneous environments provides a more likely explanation. This study underlines that tropical and subtropical mountains are not only biodiversity hotspots harbouring endemic and extremely rare plant species, but can also be cradles for the diversification of plants, such as *Rhododendron*. The integrative framework incorporating both ecological and evolutionary approaches demonstrated herein can allow better understanding of the factors that shape the global patterns of plant diversity.

ACKNOWLEDGMENTS

We thank Carsten Rabbek and Robert E. Ricklefs for their constructive comments. This work was supported by the National Key Research Development Program of China (no. 2017YFA0605101), the National Natural Science Foundation of China (nos 31650110471,
31522012, 31470564, 31621091), 111 Project (no. B14001) and Chinese Academy of Sciences-Peking University Pioneer Collaboration Team. XX was also supported by the Fundamental Research Funds for Central Universities (no. YJ201721). ZW, DD and JDK thank the support from the Danish National Research Foundation (DNRF96) to the Center for Macroecology, Evolution and Climate.

DATA ACCESSIBILITY

The global distribution of Rhododendron species, the GenBank accession numbers for each species and the phylogeny are available in Supporting Information Appendices S3, S4 and S5, respectively.

AUTHOR CONTRIBUTIONS

ZW and NS conceived the idea and designed the study; ZW, NS, XS, XX, YL and ZT collected the data; NS, LL, DD, QW, JDK and ZW analysed the data; NS and ZW led the writing. All authors approved the submission.

ORCID

Naval Shrestha http://orcid.org/0000-0002-6866-5100
Xiaoting Xu http://orcid.org/0000-0001-8126-614X

REFERENCES


BIOSKETCHES

NAWAL SHRESTHA is a post-doctoral researcher at Peking University and is broadly interested in historical biogeography and macroecology. His work focuses on investigating large-scale patterns of species distribution and evaluating how evolutionary processes and/or contemporary environment influence them.

ZHENG WANG has interests in macroecology and biogeography. His research is focused on ecological and evolutionary mechanisms underlying large-scale species diversity patterns and the impacts of climate change on species distributions.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.