Spatiotemporal evolution of the global species diversity of *Rhododendron*

Xiao-Mei Xia
Institute of Botany, Chinese Academy of Sciences
Miao-Qin Yang
State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences
Cong-Li Li
State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences
Si-Xin Huang
State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences
Fei Wang
West China Subalpine Botanical Garden, Institute of Botany, Chinese Academy of Sciences
Xiao-Hua Li
Lushan Botanical Garden, Institute of Botany, Chinese Academy of Sciences
Watanabe Yoichi
Graduate School of Horticulture, Chiba University
Le-Hua Zhang
Lushan Botanical Garden, Institute of Botany, Chinese Academy of Sciences
Yuanrun Zheng
Key Laboratory of Resource Plants, West China Subalpine Botanical Garden, Institute of Botany, Chinese Academy of Sciences
Xiao-Quan Wang (xiaoq_wang@ibcas.ac.cn)
State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences

---

Article

Keywords: Rhododendron, rediversification, spatiotemporal evolution

Posted Date: December 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-125967/v1

Version of Record: A version of this preprint was published at Molecular Biology and Evolution on October 28th, 2021. See the published version at https://doi.org/10.1093/molbev/msab314.

Abstract

How large cosmopolitan plant genera survived great environmental changes and rediversified remains largely unknown. Here we investigated mechanisms underlying the rediversification of *Rhododendron*, the largest genus of woody plants in the Northern Hemisphere. Using 3437 orthologous nuclear genes, we reconstructed the first completely resolved and dated phylogeny of *Rhododendron*. We found that most extant species of *Rhododendron* originated by Neogene rediversification from Paleogene relicts during southern migration. The geographically uneven rediversification of *Rhododendron* led to a much higher diversity in Asia than in other continents, which was driven by two main environmental variables, i.e., habitat heterogeneity represented by elevation range and annual precipitation related to the Asian monsoons, and can be explained by leaf functional traits that show strong phylogenetic signals and correspond well with leaf-forms and geographical regions. Our study highlights the importance of integrating phylogenomic and ecological analyses in revealing the spatiotemporal evolution of species-rich cosmopolitan plant genera.

Introduction

After each mass extinction event [1], the recovery of species diversity was achieved by rediversification of relict lineages and evolution of new groups [2]. The global patterns of species diversity were further shaped in the Cenozoic by mountain building events and great climatic oscillations [3, 4]. For example, the Arcto-Tertiary flora occupied wide areas of northern high latitudes in the Paleocene to Early Eocene, but subsequently migrated southward to the middle latitudes due to climatic cooling, giving rise to Tertiary relict floras with a disjunct distribution in refugia in the Northern Hemisphere [5]. Have some Tertiary relict floras greatly rediversified and expanded their distributions? It is of great interest to know what kinds of biological groups can survive great environmental changes and how they adapted to the new environments and rediversified.

Plants groups with wide geographic distributions, particularly in mountainous regions, generally require more adaptive strategies for high environmental heterogeneity and diverse selective pressures. Species may adapt to different environments through the evolution of suitable functional traits[6-9], for instance, some traits of leaf economics spectrum are good indicators for plant adaptive strategies[10, 11]. It remains unknown how the global patterns of species richness of large cosmopolitan plant genera have been jointly shaped by evolutionary processes, environmental variables, and functional traits.

Rhododendron, comprising 8 subgenera, 15 sections, 71 subsections, and more than 1000 species, is the largest genus of woody plants in the Northern Hemisphere and an important component of montane ecosystems [12], with a high species diversity in eastern Asia, the Himalaya-Hengduan Mountains, and Southeast Asia. Most extant species of *Rhododendron* very likely originated from rediversification given the fossil record of this genus from the Eocene to Miocene sediments in circumboreal areas[13-15]. This genus includes both deciduous and evergreen species, which show diverse life-forms, including small

creepers of several centimeters high, shrubs, trees up to 30 meters high, and some epiphytes. Ecologically, *Rhododendron* can adapt to low-nutrient habitats because of the unique ericoid mycorrhizal fungi and (or) tough evergreen leaves. In particular, leaves of *Rhododendron* are highly diverse in size and texture, showing that some leaf functional traits could have adapted to different habitats and altitudes. Therefore, *Rhododendron* provides an ideal system to investigate the mechanisms underlying rediversification of relict plant groups and the interactions of evolutionary and ecological forces in driving the global patterns of species richness.

A robust species phylogeny is essential to understand the spatiotemporal evolution of a plant group, but phylogenetic relationships of most subgenera, sections, subsections and species of *Rhododendron* remain unresolved due to the use of a few DNA markers[16-20]. In this study, we conducted a phylotranscriptomic analysis to reconstruct a highly resolved and dated species tree for *Rhododendron*, and then employed model-based biogeographical and diversification rate analyses to interpret its spatiotemporal diversification. We further explored which environmental variables and leaf functional traits had determined the global patterns of species diversity of this genus.

Results

Transcriptome/Chloroplast Genome Data Processing and Statistics

We obtained 248 transcriptomes from *Rhododendron* and two outgroups. Each transcriptome included 7,500,000–48,000,000 reads, which were de novo assembled into 19,900 – 467,000 transcripts, with the N50 scaffold length ranging from 400 to 1 600 bp (Supplementary Table 1). Our data include 201 species of *Rhododendron* that represent all eight subgenera (*Azaleastrum*, *Candidastrum*, *Hymenanthes*, *Mumeazalea*, *Pentanthera*, *Rhododendron*, *Therorhodion*, and *Tsutsusi*) and 15 sections, as well as nearly all multi-species subsections which comprise approximately 95% of the species richness of this genus [12]. In addition, 38 chloroplast genes with high-quality sequences were obtained from the transcriptome data for 175 species, using the six newly sequenced chloroplast genomes as reference sequences (Supplementary Table 1).

Well-resolved Phylogeny of *Rhododendron*

The concatenated alignments of the 3437 OGs (nuclear genes) and 38 chloroplast genes used in phylogenetic analyses include 3,903,298 bp for 202 species (200 *Rhododendron* species and two outgroups. *Rhododendron nipponicum* was excluded due to the low quality of its transcriptome sequences) and 31572 bp for 175 species, respectively. Both concatenated and coalescent trees generated from the nuclear genes are highly resolved and strongly supported (UFBoot = 100% and LPP = 100% for most nodes) with almost consistent topologies, but a few nodes of the coalescent tree have relatively lower support (Fig. 1 and Supplementary Fig. 2). The tree generated from the concatenated chloroplast genes shows consistent topologies as the nuclear gene tree at the section level, but has much lower resolution for interspecific relationships (Supplementary Fig. 3). Our results support the division of

Rhododendron into five subgenera (*Therorhodion*, *Tsutsusi*, *Rhododendron*, *Pentanthera*, and *Hymenanthes*) and eleven sections (*Therorhodion*, *Sciadorhodion*, *Choniastrum*, *Azaleastrum*, *Brachycalyx*, *Tsutsusi*, *Ledum*, *Vireya*, *Rhododendron*, *Pentanthera*, and *Ponticum*) (Fig. 1), and suggest a need to redefine some subgenera and sections. For example, the previously recognized subgenus *Choniastrum* was inferred as a sister to the large subgenus *Rhododendron* by some previous studies[16, 20], but was nested within subg. *Tsutsusi* and placed as sister to a clade comprising three sections (*Azaleastrum*, *Brachycalyx*, and *Tsutsusi*) in the present study. In addition, the previously recognized genus *Menziesia* was also nested within subg. *Tsutsusi* (Fig. 1). Therefore, subg. *Tsutsusi* defined by Chamberlain *et al*. [12] was not monophyletic, as found in previous molecular phylogenetic studies [16–20].

The transcriptome-based phylogeny also showed a very high resolution at the species level, even for the taxonomically difficult subgenus *Hymenanthes* which includes more than 300 species (Fig. 1). In subg. *Hymenanthes*, species at the basal positions were from the higher latitudes of Eurasia and North America, including *R. brachycarpum*, *R. aureum*, *R. maximum*, *R. smirnowii*, *R. ponticum*, *R. yakushimanum*, *R. japonoheptamerum*, *R. makinoi* and *R. degronianum*. Within subg. *Pentanthera* mainly distributed in Northern America, *R. molle* from North-East Asia was sister to *R. austrinum* from southeastern North America. In subg. *Rhododendron*, two species previously assigned to the genus *Ledum*, i.e., *R. hypoleucum* (*L. hypoleucum*) and *R. tomentosum* (*L. palustre*) from circumboreal regions, diverged first, and formed a clade sister to the other clade composed of two lineages, one comprising species of section *Vireya* mainly from the Malaysian Archipelago, and the other belonging to section *Rhododendron* that comprises the Asian species as well as *R. ferrugineum* and *R. hirsutum* from Europe. In section *Sciadorhodion*, *R. albiflorum* from Western North America was sister to species from North-East Asia. In particular, in section *Vireya* mainly distributed in the Malaysian Archipelago, the eastern and southern Asian species (*R.micranthum*, *R. charitopes*, *R. longistylum*, and *R. emarginatum*) representing the northernmost distribution of the section, branched off first, followed by *R. ericoides*, a species endemic to Malaysia, characterized by linear leaves. Moreover, the robust phylogeny of *Rhododendron* anchored the phylogenetic positions of some species with unique morphological characters, such as *R. pentandrum* (*Menziesia pentandra*), *R. goyozanense* (*Menziesia goyozanensis*), *R. benhalii* (*Menziesia ciliicalyx*), and *R. semibarbatum*. This phylogeny is important for investigating the evolutionary and biogeographic history, and provides a good framework for the taxonomic revision of *Rhododendron.*

Divergence Time Estimates and Biogeographical Analyses

The genus *Rhododendron* likely originated in the Paleocene at ~ 64 million years ago (MYA), with subg. *Therorhodion* as the earliest diverging lineage (~ 56 MYA), followed by subg. *Tsutsusi*, which diverged from the remaining subgenera at roughly 35.9 MYA during the late Eocene (Fig. 2 and Supplementary Fig. 4). The two subgenera *Hymenanthes* and *Pentanthera* diverged from each other in the early Oligocene (~ 33.5 MYA). Explosive diversifications occurred in four sections (*Tsutsusi*, *Vireya*, *Rhododendron*, and *Ponticum*) during the Miocene, particularly in the middle to late Miocene, leading to the occurrence of 94% of *Rhododendron*'s species diversity in these sections.

The Akaike information criterion (AIC) model selection supported the dispersal-extinction-cladogenesis (DEC) model (LnL = -192.25, AIC = 388.50), and the DEC + J model showed higher log-likelihood (LnL) values and lower AIC values (LnL = -186.09, AIC = 378.18) (Supplementary Table 2). The ancestral range estimates indicated that *Rhododendron* very likely originated in North-East Asia, subsequently expanded into Europe and North America, and finally into South Asia and the Malaysian Archipelago (Fig. 2 and Supplementary Fig. 5). The biogeographic stochastic mapping (BSM) using the DEC + J model suggested that most biogeographical events can be explained by within-area speciation (82.64%) and dispersals (15.05%), and only a few events involved vicariance (2.31%) (Supplementary Table 3). The high number of within-area speciation events can be expected given the large areas of the biogeographic regions (i.e., South Asia, North-East Asia, and the Malaysian Archipelago) we defined. Among the dispersal events, anagenetic events (i.e., range expansions) were more common (9.73%) than jump dispersal to a new area (i.e., founder events) (5.32%), and many dispersals occurred between North-East Asia and South Asia (Supplementary Table 3). Most of the inferred vicariance events also occurred between South Asia and North-East Asia. For example, the ancestor of subg. *Rhododendron* was inferred to have expanded its range from North-East Asia to South Asia and the Malaysian Archipelago, followed by vicariant evolution of the sections that are most species-rich in South Asia and the Malaysian Archipelago.

Diversification Analyses

We analyzed the global rates of speciation and extinction for *Rhododendron* using Bayesian analysis of macroevolutionary mixtures [21]. To alleviate the effect of incomplete and unequal sampling, we used clade-specific sampling fractions to ensure that each subgenus was sampled in proportion to the number of its described species. The net diversification rate and speciation rate increased rapidly during the Eocene, rose gradually during the early Oligocene to the middle Miocene, and increased sharply during the late Miocene (Fig. 3). Based on the BAMM analysis, the diversification rate was much higher in six nodes (1, the crown group of *Rhododendron*; 2–4, the three subgenera *Tsutsusi*, *Rhododendron* and *Hymenanthes*; 7, section *Vireya* in Southeast Asia; 8, sect. *Ponticum* in South Asia) than in other two nodes corresponding to the two species-poor subgenera *Pentanthera* (5) and *Therorhodion* (6) (Fig. 2).

Phylogenetic Signal and Divergence of Leaf Functional Traits

By analyzing the variation of seven traits critical to the growth of the studied *Rhododendron* species, we found that three traits, leaf N, leaf P and SLA (specific leaf area), were strongly related (Fig. 4a), and had strong phylogenetic signals based on Pagel's λ ($\lambda > 0.50$ and $P < 0.05$ for $\lambda = 0$) (Supplementary Table 4). Blomberg's *K* showed a similar variation pattern as Pagel's λ, but had a lower absolute magnitude (Supplementary Table 4). In addition, the three traits showed clear divergences among the two life-forms, the five subgenera and the five distribution ranges (Fig. 4b and Supplementary Fig. 8). For example, the deciduous species at lower altitudes from subg. *Pentanthera* and three sections (*Sciadorhodion*, *Brachycalyx* and *Tsutsusi*), the two species of subg. *Therorhodion* in the boreal-tundra, as well as a few deciduous species of section *Rhododendron* at extremely high altitudes or cold habitats, exhibited higher leaf N, leaf P, and SLA. In contrast, the evergreen species from subg. *Hymenanthes* and four sections

(*Choniastrum*, *Azaleastrum*, *Ledum*, and *Vireya*) as well as most species of section *Rhododendron* had lower leaf N, leaf P, and SLA.

Effects of Environmental Variables and Leaf Functional Traits on the Global Pattern of *Rhododendron* Species Diversity

According to the multiple regression analysis, all tested models had a lower R^2 at the global scale than the regional scale (Supplementary Table 7). The habitat heterogeneity represented by elevation range and the water availability represented by annual precipitation were positively correlated with global *Rhododendron* species richness (Fig. 5b; Supplementary Table 7). The driving factors of species richness were different among the three biogeographic regions (South Asia, North-East Asia, and the Malaysian Archipelago) with high species diversity (Fig. 5; Supplementary Table 7). In South Asia, the three variables about the energy availability including temperature annual range, isothermality and mean diurnal range, and the soil organic carbon content in the fine earth fraction were stronger predictors; In North-East Asia, the mean diurnal range, isothermality and elevation range were significantly correlated with species richness; In the Malaysian Archipelago, the annual mean temperature and the max temperature of warmest month showed negative correlation, whereas the mean temperature of wettest quarter, annual precipitation and elevation range showed positive correlation with species richness.

Although habitat shifts are numerous in *Rhododendron*, we only found phylogenetic constraints on elevations of species distribution and leaf functional traits (Supplementary Fig. 8). Moreover, we investigated whether leaf functional traits were niche-delimiting, and found a positive correlation of the pairwise disparity of geographical distance and elevation with leaf functional traits (Supplementary Figs. 9–10). For example, low trait values were mostly observed in evergreen trees at higher altitudes in South Asia and the Malaysian Archipelago, and high trait values were found in deciduous shrubs at lower altitudes in North-East Asia and North America, and deciduous creeping shrubs at extremely high altitudes in South Asia and North-East Asia.

Discussion

The present joint phylogenomic and ecological study has successfully reconstructed the evolutionary and biogeographical history of the large cosmopolitan genus *Rhododendron* and revealed its Neogene re-diversification from Paleogene relicts during its southern migration. Using 3437 orthologous nuclear genes (OGs) obtained from transcriptome sequencing, our phylogenomic analysis generated the first completely resolved and dated phylogeny of *Rhododendron* comprising 200 species that represent all subgenera, sections and nearly all subsections (Fig. 1), which is also supported by the analysis of 38 chloroplast genes (Supplementary Fig. 3). Notably, we found that the subgenus *Choniastrum* is sister to a clade comprising three sections (*Azaleastrum*, *Brachycalyx*, and *Tsutsusi*) of subg. *Tsutsusi* rather than sister to subg. *Rhododendron* as suggested by some previous studies [16, 20], and thus should be merged into subg. *Tsutsusi*. The transcriptome-based phylogeny also anchors the systematic positions of all subgenera and sections that remained controversial. This robust species phylogeny, together with

divergence time estimation, ancestral area reconstruction, diversification rate calculation, and the analysis of environmental variables, provides an excellent opportunity for understanding the spatiotemporal evolution of *Rhododendron*. The crown age estimate (~ 64 MYA) and reconstructed ancestral areas for different nodes (Fig. 2 and Supplementary Figs. 4–5), together with the fossil record [15], strongly suggest that *Rhododendron* originated in boreal regions in the early Paleocene and subsequently underwent several vicariance and dispersal events, as well as anagenetic events (Supplementary Table 3).This is corroborated by the most basal position of subg. *Therorhodion* (Fig. 1), which has peduncles bearing leaf-like bracts and only includes 2 species distributed in both sides of Beringia. However, molecular dating indicates that all extant lineages of *Rhododendron*, with the exception of subg. *Therorhodion*, did not diverge until the late Eocene (Fig. 2 and Supplementary Fig. 4). In particular, currently, the greatest diversity of *Rhododendron* occurs in the Himalaya-Hengduan Mountains and the Malaysian Archipelago, where at least 860 species have been recognized. Therefore, most extant species of *Rhododendron* very likely originated from rediversification of its Paleogene relicts that migrated southwards. A north-to-south migration history was also revealed in the two large subgenera *Hymenanthes* and *Rhododendron* as well as in the large section *Vireya* [22]. Considering that approximately five-sixth species became extinct in the Cretaceous-Paleogene mass extinction event [23], the earlier lineages of *Rhododendron* in Europe likely went extinct during the Paleocene based on the earliest fossil record [15]. It is particularly interesting that *Rhododendron* colonized the Himalaya-Hengduan Mountains and the Malaysian Archipelago in the Middle Miocene (Fig. 2), when several uplift episodes of the Qinghai-Tibetan Plateau (QTP) and the Himalayas occurred sequentially and the global climate cooled [24]. Previous studies suggested that the greatest diversity of *Rhododendron* in these regions may be due to the diverse and heterogeneous topography caused by the uplift of the QTP [24, 25], which created high mountains and deep valleys and led to the changes in drainage systems [26, 27]. We detected numerous within-area speciation events in the Himalaya-Hengduan Mountains, most of which occurred in sections *Ponticum* and *Rhododendron*, and the newly evolved species are much younger than those occurring in the high latitudes of Eurasia and North America. The other within-area speciation events mainly occurred in the uplands of Southeast Asia in section *Vireya* comprising approximately 312 species, of which 150 species from New Guinea, Australia and the Solomon Islands constitute a monophyletic group [28] that evolved with the Plio-Pleistocene orogeny in New Guinea [26, 27]. All of the above evidence suggests that *Rhododendron* survived the Tertiary and Quaternary climatic oscillations by migrating southwards from circumboreal areas to tropical/subtropical mountains and across the equator to Southeast Asia, accompanying by rapid radiations.

The uneven rediversification of *Rhododendron* in different continents led to a unique global pattern of its species diversity, i.e., a great diversity in Asia (including Southeast Asia) vs. a low species richness in Europe and North America, which can be explained by two main environmental variables and leaf functional traits. The species richness of *Rhododendron* is significantly positively correlated with elevation range and annual precipitation (Fig. 5). In the middle to late Miocene, active mountain building in East Asia, the Himalayas, and Southeast Asia to Oceania produced highly heterogeneous habitats

represented by a wide altitudinal range [26, 27, 29], which, together with high precipitation caused mainly by the intensification of Asian monsoons, greatly promoted the rapid rediversification of subgenera *Tsutsusi* and *Hymenanthes* and section *Vireya* of the subgenus *Rhododendron* and the evolution of new lineages, leading to the highest diversification rate of this genus. Although mountain building also occurred in southern Europe and western North America in the Neogene [29, 30], it was not active as in Asia, and the relatively dry habitats limited the rediversification of *Rhododendron*. This also explains why only a few intercontinental dispersals occurred within subgenera of this genus, except the deciduous subgenus *Pentanthera*, which grows in very low altitudes. Moreover, the energy availability, represented by mean diurnal range, isothermality, temperature annual range and annual mean temperature, also played an important role in shaping regional species richness of *Rhododendron* (Fig. 5). In contrast to the finding of previous studies that species richness of woody plants was often limited by cold winter temperatures [31], the species richness of *Rhododendron* in the Malaysian Archipelago is significantly negatively correlated with annual mean temperature (Fig. 5c), indicating a prefer of cold climate. Indeed, most extant *Rhododendron* species are clod-resistant and grow in cold and humid habitats, even at an altitude up to 5800 m for *R. nivale* [22]. The leaf functional traits, including SLA, leaf N, leaf P and leaf C, show different strategies for resource capture and phylogenetic conservatism, and are fundamental to survival and growth[10, 11, 32–35]. Previous studies have investigated the distributions, relationships and adaptation of leaf functional traits at both global and regional scales [10, 11, 32, 33], and reported phylogenetic conservatism of some traits [36–39], but rarely focused on the global patterns of these traits for a specific plant group. Based on the robust species phylogeny of *Rhododendron* (Fig. 1), our present study further analyzed seven leaf functional traits in these species, which have a common genetic background and similar divergence times, and thus can better understand the evolution and ecological adaptation of these traits. We found that three traits (SLA, leaf N, and leaf P) show strong phylogenetic signals at the subgenus level (Fig. 4), and these traits with phylogenetic conservatism are closely related to plant nutrient use, which might be caused by stabilizing selection of nutrient availability in different habitats and correlated adaptations between structural and functional traits. Moreover, the variation of SLA, leaf N, and leaf P corresponds well with leaf-forms and geographical distributions of species in *Rhododendron*. Leaves of deciduous species show higher nutrient concentration and greater SLA than evergreen species (Fig. 4), and particularly the North American species mainly belonging to the deciduous *Pentanthera* exhibit much higher SLA, leaf N and leaf P than the species of South Asia and Malaysian Archipelago mainly belonging to the evergreen sections *Ponticum* and *Vireya* (Fig. 4), which also display the highest diversification rates (Fig. 2). This suggests that plants adjust their ecophysiological traits to compete for resources in various environments, with trade-offs between co-adapted traits [40]. Compared to species with deciduous leaves, i.e., the “acquisitive” leaves that tend to exhibit photosynthetic capacity and growth rate[41–43], species with evergreen leaves (low-SLA and nitrogen-poor “conservative” leaves) are expected to have higher survival in the face of abiotic and biotic hazards [10, 44]. This might also explain the great species diversity of *Rhododendron* in the Himalaya-Hengduan Mountains and the Malaysian Archipelago.

Methods

Plant Sampling, Sequencing, and Data Processing

We sampled a total of 250 individuals for transcriptome sequencing, including 201 species of *Rhododendron* that represent all eight subgenera, all 15 sections and nearly all multi-species subsections of this genus [12], and two species (*Diplarche multiflora* and *Empetrum nigrum*) as outgroups (Supplementary Table 1), based on the most close relationships between *Rhododendron* and the two genera *Diplarche* and *Empetrum* [45]. To test the monophyly of some species of *Rhododendron*, each of them were represented by two individuals. In addition, six of the 250 individuals, representing six *Rhododendron* species (*R. championiae*, *R. tomentosum*, *R. mucronulatum*, *R. austrinum*, *R. jinggangshanicum*, and *R. hunnewellianum*) were also used for chloroplast genome sequencing (Supplementary Table 1). For transcriptome sequencing, total RNA was extracted from young leaves and leaf buds using the RNAprep Pure Plant Kit (Tiangen, Beijing, China). The cDNA libraries were constructed with NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England BioLabs Inc., Ipswich, MA, USA) and were sequenced on the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA), which generated paired-end reads of 100 bp and (or) 150 bp used in *de novo* assembly with Trinity v. 2.0.6 [46]. After discarding redundant transcripts using CD-HIT v. 4.6.5[47, 48] with a similarity threshold of 1, the longest transcript of each gene was retained. The coding sequence (CDS) identification was performed using TransDecoder v. 0.36 (https://github.com/TransDecoder/TransDecoder/releases)[49], and the CDS encoding the longest peptide was retained for the ortholog search. For chloroplast genome sequencing, total DNA was extracted from fresh young leaves using the cetyltrimethylammonium bromide (CTAB) method [50]. The DNA libraries with an average length of 350 bp were constructed using the NexteraXT DNA library preparation Kit (Illumina, San Diego, CA) and were sequenced on the Illumina Novaseq 6000 platform (Total Genomics Solution Limited,SZHT), generating reads with an average length of 150 bp. These raw reads were edited using NGS QC Tool Kit v2.3.3 [51], and the obtained high-quality reads were assembled *de novo* into contigs using SPAdes v.3.11.0 [52]. The PGA software [53] was used to annotate the chloroplast genomes and identify the protein-encoding genes.

Identification of Orthologous Genes for Phylogenomic Analysis

Orthologous genes (OGs) were searched from the transcriptome sequences by HaMStR v.13.2.6 [54]. First, 4740 core OGs were identified with OrthoFinder v.2.2.7 [55], based on 10 primer taxa, including *R. grande*, *R. glischrum*, *R. lanatum*, *R. fulvum*, *R. sinogrande*, *R. pingianum*, *R. hemsleyanum*, *R. sikangense*, *R. simiarum*, and *R. araiophyllum*. These OGs were then used as queries to search homologous sequences from the 250 samples by HaMStR v.13.2.6 [54] with strict parameters (-est -central -intron = remove -representative -strict -eval_blast = 1e-20 -eval_hmmer = 1e-20). The CDS and AA sequences of each OG were obtained using Perl scripts and aligned using MAFFT v7.407[56]. To obtain chloroplast genes for phylogenetic analyses, 68 protein-coding genes obtained from the complete chloroplast genome sequences of the six *Rhododendron* species were used as queries to search the transcriptome sequences

of the 200 *Rhododendron* species and two outgroups using HaMStR v.13.2.6 [54] based on strict parameters. The CDS and AA sequences of the chloroplast genes were also obtained using Perl scripts and aligned using MAFFT v7.407[56]. After discarding the low quality sequences and the species with over 50% missing data, 38 chloroplast genes of 175 species were used in the final analyses.

We obtained the optimal OG dataset by testing whether conspecific individuals formed highly supported monophyletic groups. A total of 3637 OGs, each of which covered at least 225 individuals (90% of the 250 samples), were concatenated by FASconCAT-G_v.1.04 [57] for the test. We used IQ-TREE v.2.0-rc1 [58] to perform maximum-likelihood (ML) analysis for the concatenated dataset using the GTR + F + I + G4 model, with 1000 ultrafast bootstrap replicates (UFBoot) [59]. The generated tree showed that all conspecific individuals formed monophyletic groups with high branch support (Supplementary Fig. 1). Due to computational limitations, to save calculation time, we selected one individual with most OGs in each monophyletic group to represent each species. After discarding the low-quality OGs, 3568 candidate OGs were retained, each of which covered 90% of all sampled species. The AA sequences and the corresponding CDS sequences of these candidate OGs were aligned with MAFFT v7.407[56]. To increase the robustness of the phylogenetic reconstruction, we identified the ambiguously aligned positions in the AA sequences using ZORRO [60] and removed the positions with a confidence score of less than 5. After manually inspecting and discarding low-quality sequences, a total of 3437 OGs, each having at least 90% species coverage and a length of at least 300 bp, were retained in the final analyses.

Phylogenetic Analyses and Divergence Time Estimation

We used 3437 OGs and 38 chloroplast genes in phylogenetic reconstruction with the maximum-likelihood (ML) method in IQ-TREE v.2.0-rc1[58]. For the 3437 OGs, both concatenation or coalescent approaches were used. In the concatenation analysis, all OGs were concatenated into a supermatrix, and the ML tree was generated under the GTR + F + I + G4 model. In the coalescent analysis, single-gene trees were estimated for all OGs, and the coalescent tree was reconstructed by ASTRAL v.5.7.3[61]. For the 38 chloroplast genes, only the concatenation analysis was conducted, and the ML tree was generated using the same method as mentioned above. For all phylogenetic analyses, all codon positions of the OGs/genes were included because no substitutional saturation was detected by the DAMBE [62], and 1000 ultrafast bootstrap replicates were used [59].

The divergence time estimation was performed based on the ML tree of the 3437 concatenated OGs, using the RelTime method [63] with the general time reversible (GTR) nucleotide substitution model in MEGA X [64]. We selected two fossils for time calibration. The seed fossil of *Rhododendron newburyanum* [15] dated to the Paleocene (c. 54.5 MYA) was used to set the minimum age of the *Rhododendron* crown group, based on the consensus of the available evidence [20, 45, 65, 66], and the leaf fossil of *Rhododendron protodilatatum* [67], dated to the Pliocene (c. 5.3 MYA), was used to set the minimum age of the section *Brachycalyx* crown group as in the study of [68].

Ancestral Area Reconstruction and Diversification Rate Estimation

To trace the biogeographic history of *Rhododendron*, the ancestral area reconstruction was conducted by BioGeoBEARS v.1.1.1 [69, 70] implemented in R v.3.6.2 [71]. The time-calibrated phylogenetic tree was generated by MEGA X [64], and outgroups were pruned. Six geographic regions were defined based on the current distribution of *Rhododendron*, including (A) Eastern North America, (B) Western North America, (C) South Asia, (D) Europe, (E) the Malaysian Archipelago, and (F) North-East Asia (Fig. 2). We tested six models, including the likelihood version of DIVA (DIVALIKE), the dispersal-extinction-cladogenesis model (LAGRANGE's DEC), and the BayArea likelihood version of the range evolution model (BAYAREA), as well as the "+J" versions of the three models, which consider founder events [69]. Based on a comparison of the AICc and LnL values, the DEC + j model was selected as the best model (Supplementary Table 2). We used biogeographical stochastic mapping (BSM) to calculate the number of dispersal events between the six geographic regions. This calculation estimated the mean and standard deviations for anagenetic dispersals, extinctions, cladogenetic range expansions (i.e., sympatry), and founder events from 50 stochastic maps as described by Matzke[70].

The diversification rates and their shifts were calculated by BAMM[21]. To avoid nonrandom unequal sampling, we specified the sampling fraction for each subgenus. We performed two runs for 10,000,000 and 1,000 generations, respectively. The convergence was assessed by the effective sample size of likelihood and the number of shifts (ESS > 200). After discarding 25% of samples as burn-in, the BAMM results were used as the input file for BAMMtools[72].

Environmental Data Assembly and Niche Analysis.

The global occurrence records of *Rhododendron* were collected from online databases (the global biodiversity information facility, GBIF, www.gbif.org), herbaria (PE, HAST, K, and TAI), and fieldwork. The entire dataset was closely scrutinized. All exact duplicates and all low-resolution coordinates without exact latitude and longitude information were removed. The coordinates that matched the occurrences of cultivated plants (urban and botanical gardens) were also removed. The occurrences of known invasive species were trimmed to only include its native range. Finally, we used a total of 13450 unique coordinates from 765 species, including the 200 *Rhododendron* species used in the phylogenetic analysis.

We compiled 35 environmental factors that represent the features of climate, soil, and land cover to study the relative contributions of these factors to the diversity pattern of *Rhododendron*. Nineteen bioclimatic variables (temperature and precipitation), eight soil layers, and six land cover classes were downloaded from CHELSA (Climatologies at High Resolution for the Earth's Land Surface Areas, http://chelsa-climate.org/) [73], SoilGrids project (https://soilgrids.org/), and Global 1-km Consensus Land Cover (https://www.earthenv.org/landcover), respectively. The elevation, the index of aridity (AR), and potential evapotranspiration (PET) were extracted from GMTED2010 (http://grasswiki.osgeo.org/wiki/Global_datasets) and the CGIAR-CSI website (www.cgiar-csi.org),

respectively. For the soil datasets, we utilized high-efficiency Python methods based on GDAL (https://www.gdal.org/) to extract relative layers and subsequently averaged them across 5-, 15-, and 30-cm sampling depths. The median values of the factors were used for downstream analyses.

Among the 35 datasets of environmental factors, 28 variables were retained for subsequent analysis after removing the highly correlated ones ($r > 0.8$) factors, including bio4 (Temperature Seasonality), bio6 (Min Temperature of Coldest Month), bio10 (Mean Temperature of Warmest Quarter), bio11 (Mean Temperature of Coldest Quarter), bio13 (Precipitation of Wettest Month), bio16 (Precipitation of Wettest Quarter) and bio17 (Precipitation of Driest Quarter). Multiple regression models were constructed based on the global *Rhododendron* dataset and the datasets for South Asia, North-East Asia, and the Malaysian Archipelago.

Measurement of Leaf functional Traits

The specific leaf area (SLA) data were assembled for the 200 species used in the the phylogenetic analysis. Up to five fully expanded leaves per species, representing 2–4 individuals, were scanned with a leaf area meter and dried at 60 ˚C for 36 h and the dry mass was determined to 0.0001 g. The SLA (cm^2g^{-1}) was computed by dividing the leaf area by the leaf dry mass. The dry leaves were grounded and homogenized for leaf nutrient measurements. The total leaf nitrogen content (leaf N) and leaf carbon content (leaf C) were determined by the fully automatic element analyzer Vario EL cube (Elementar Analysensysteme GmbH, Hanau, Germany) using one portion of dry leaves. The other portion of dry leaves was digested in HNO_3 ($p \approx 1.42$ g.cm^{-3}), and was used for determination of the concentrations of P (leaf P), K (leaf K), and Mg (leaf Mg) using inductively coupled plasma-optical emission spectrometer (iCAP 6300 ICP-OES spectrometer, Thermo Fisher, USA).

Investigation of Phylogenetic Signal and Evolution of Leaf Functional Traits

To test whether leaf traits are phylogenetically constrained, we used the *Blomberg's K* [74] and Pagel's lambda [75] with 1000 simulation replicates of trait data from the R package caper (Orme, 2013) and phytools [76]. For the quantitative traits, Pagel's λ was used to show the extent to which traits depart from the expected curve under a Brownian motion model of trait evolution [75]. The λ values range between 0 and 1, where $\lambda = 1$ indicates that the distribution of trait values across the phylogeny is exactly as expected under Brownian motion, whereas $\lambda = 0$ implies no phylogenetic signal. We interpreted the phylogenetic niche conservatism here as being represented by λ values > 0.50, which indicates that traits of closely related species are more similar than they would be expected by chance [77]. We also calculated the Euclidean distances for geographic location, environmental and trait variables. The Mantel test was applied to evaluate the correlation between trait distance and geographic distance, and between trait distance and distance in environmental variables from the R phytools [76].

Declarations

Reporting summary

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

Data availability

All newly generated raw sequence reads will be posted in the NCBI SRA database. SRA entries for each assembly will be listed in Supplementary Table 1. Other data, including phylogenetic trees, will be available from Dryad. All the data will be available before publication. All code is available from the authors upon request.

Acknowledgements

We sincerely thank Jian-Sheng Peng (Shangri-La Alpine Botanical Garden), Zheng-Yu Liu and Zhong-Wei An (Chongqing Institute of Medicinal Plant Cultivation), Zhi-Duan Chen, Jian-Fei Ye and Bing Liu (Institute of Botany, Chinese Academy of Sciences), Harry Wiriadinata (Bogor Botanical Gardens, Indonesia), and Arief Hidayat (Research Center for Plant Conservation and Botanic Gardens, Indonesian Institute of Sciences) for their help in plant sampling. We thank Hong Du for her assistance in lab work; and Wei-Tao Jin, Zhe Cai and Ting-Ting Shen (Institute of Botany, Chinese Academy of Sciences), and Yi Wang (Southern University of Science and Technology) for assistance with data analyses. This study was supported by the National Key Research and Development Program of China (grant no. 2017YFA0605100), the Strategic Priority Research Program, CAS (XDA23080000), and the Key Research Program of Frontier Sciences, CAS (QYZDJ-SSW-SMC027).

Author contributions

X.Q.W. conceived and designed the study. X.M.X. and M.Q.Y. performed the experiments. X.M.X., X.Q.W., C.L.L and S.X.H analysed the data. X.Q.W. and X.M.X. wrote the manuscript. X.M.X., X.Q.W., W.Y., Y.R.Z., F.W., L.H.Z and X.H.L provided plant materials of *Rhododendron*. All authors contributed to the final manuscript.

Corresponding author

Correspondence to Xiao-Quan Wang.

Competing interests

The authors declare no competing interests.

References

1. Barnosky, A.D. *et al*. Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51-57 (2011).

2. Marshall, C.R. Five palaeobiological laws needed to understand the evolution of the living biota. *Nat. Ecol.Evol.* **1** (2017).
3. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913 (2000).
4. Hughes, C.E. & Atchison, G.W. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytol.* **207**, 275-282 (2015).
5. Mai, D.H. Palaeofloristic changes in Europe and the confirmation of the Arctotertiary Palaeotropical geofloral concept. *Rev. Palaeobot. Palyno.* **68**, 29-36 (1991).
6. Ackerly, D.D. Adaptation, niche conservatism, and convergence: Comparative studies of leaf evolution in the California chaparral. *Am. Nat.* **163**, 654-671 (2004).
7. Losos, J.B. Convergence, adaptation, and constraint. *Evolution* **65**, 1827-1840 (2011).
8. Kostikova, A. *et al*. Scale-dependent adaptive evolution and morphological convergence to climatic niche in Californian eriogonoids (Polygonaceae). *J. Biogeogr.* **41**, 1326-1337 (2014).
9. Zanne, A.E. *et al*. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89-92 (2014).
10. Wright, I.J. *et al*. The worldwide leaf economics spectrum. *Nature* **428**, 821-827 (2004).
11. Diaz, S. *et al*. The global spectrum of plant form and function. *Nature* **529**, 167-U173 (2016).
12. Chamberlain, D., Hyam, R., Argent, G., Fairweather, G. & Walter, K. S. The Genus Rhododendron: Its Classification and Synonymy. (Royal Botanic Garden Edinburgh, Edinburgh, 1996).
13. Denk, T., Grimsson, F. & Kvacek, Z. The Miocene floras of Iceland and their significance for late Cainozoic North Atlantic biogeography. *Bot. J. Linn. Soc.* **149**, 369-417 (2005).
14. Hofmann, C.C. Light and scanning electron microscopic investigations of pollen of Ericales (Ericaceae, Sapotaceae, Ebenaceae, Styracaceae and Theaceae) from five lower and mid-Eocene localities. *Bot. J. Linn. Soc.* **187**, 550-578 (2018).
15. Collinson, M.E. & Crane, P.R. *Rhododendron* seeds from the Palaeocene of southern England. *Bot. J. Linn. Soc.* **76**, 195–205 (1978).
16. Goetsch, L., Eckert, A.J. & Hall, B.D. The molecular systematics of *Rhododendron* (Ericaceae): A phylogeny based upon RPB2 gene sequences. *Syst. Bot.* **30**, 616-626 (2005).
17. Yan, L.J. *et al*. DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya-Hengduan Mountains. *Mol. Ecol. Resour.* **15**, 932-944 (2015).
18. Berry, E., Sharma, S.K., Pandit, M.K. & Geeta, R. Evolutionary correlation between floral monosymmetry and corolla pigmentation patterns in *Rhododendron*. *Plant Syst. Evol.* **304**, 219-230 (2017).
19. Grimbs, A. *et al*. Bioactivity in *Rhododendron*: A systemic analysis of antimicrobial and cytotoxic activities and their phylogenetic and phytochemical origins. *Front. Plant. Sci.* **8**, 551 (2017).
20. Shrestha, N. *et al*. Global patterns of *Rhododendron* diversity: The role of evolutionary time and diversification rates. *Global. Ecol. Biogeogr.* **27**, 913-924 (2018).

21. Rabosky, D.L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *Plos One* **9** (2014).
22. Wu, Z.-Y., Raven, P. H. & Hong, D.-Y. (eds) Flora of China, Vol. 14 (Science Press & Missouri Botanical Garden Press, 2005).
23. Zachos, J.C., Dickens, G.R. & Zeebe, R.E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279-283 (2008).
24. Wen, J., Zhang, J.Q., Nie, Z.L., Zhong, Y. & Sun, H. Evolutionary diversificatons of plants on the Qinghai-Tibetan Plateau. *Front. Genet.* **5** (2014).
25. Shrestha, N., Su, X., Xu, X. & Wang, Z. The drivers of high *Rhododendron* diversity in south-west China: Does seasonality matter? *J. Biogeogr.* **45**, 438-447 (2018).
26. Royden, L.H., Burchfiel, B.C. & van der Hilst, R.D. The geological evolution of the Tibetan plateau. *Science* **321**, 1054-1058 (2008).
27. Clark, M.K. *et al*. Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. *Tectonics* **23** (2004).
28. Goetsch, L.A., Craven, L.A. & Hall, B.D. Major speciation accompanied the dispersal of *Vireya* Rhododendrons (Ericaceae, *Rhododendron* sect. *Schistanthe*) through the Malayan archipelago: Evidence from nuclear gene sequences. *Taxon* **60**, 1015-1028 (2011).
29. Morley, R.J. Assembly and division of the South and South-East Asian flora in relation to tectonics and climate change. *J. Tropic. Ecol.* **34**, 209-234 (2018).
30. Fernandes, V.M., Roberts, G.G., White, N. & Whittaker, A.C. Continental-Scale landscape evolution: A history of North American topography. *J. Geophy Res. Earth Surf.* **124**, 2689-2722 (2019).
31. Wang, Z.H., Fang, J.Y., Tang, Z.Y. & Lin, X. Patterns, determinants and models of woody plant diversity in China. *P. Roy. Soc. B.-Biol. Sci.* **278**, 2122-2132 (2011).
32. Chapin, F.S. The mineral-nutrition of wild plants. *Annu. Rev. Ecol. Syst.* **11**, 233-260 (1980).
33. Reich, P.B., Walters, M.B. & Ellsworth, D.S. From tropics to tundra: Global convergence in plant functioning. *Proc. Natl. Acad. Sci. USA* **94**, 13730-13734 (1997).
34. Grime, J.P. *et al*. Integrated screening validates primary axes of specialisation in plants. *Oikos* **79**, 259-281 (1997).
35. Diaz, S. *et al*. The plant traits that drive ecosystems: Evidence from three continents. *J. Veg. Sci.* **15**, 295-304 (2004).
36. Hao, G.Y. *et al*. Ecology of hemiepiphytism in fig species is based on evolutionary correlation of hydraulics and carbon economy. *Ecology* **92**, 2117-2130 (2011).
37. Cavender-Bares, J., Ackerly, D.D., Baum, D.A. & Bazzaz, F.A. Phylogenetic overdispersion in Floridian oak communities. *Am. Nat.* **163**, 823-843 (2004).
38. Ackerly, D.D. & Donoghue, M.J. Leaf size, sapling allometry, and Corner's rules: Phylogeny and correlated evolution in maples (Acer). *Am. Nat.* **152**, 767-791 (1998).

39. Liu, H. *et al*. Strong phylogenetic signals and phylogenetic niche conservatism in ecophysiological traits across divergent lineages of Magnoliaceae. *Sci. Rep.* **5**, 12 (2015).
40. Santiago, L.S. & Kim, S.C. Correlated evolution of leaf shape and physiology in the woody *Sonchus* alliance (Asteraceae: Sonchinae) in Macaronesia. *Int. J. Plant Sci.* **170**, 83-92 (2009).
41. Kikuzawa, K., Onoda, Y., Wright, I.J. & Reich, P.B. Mechanisms underlying global temperature-related patterns in leaf longevity. *Glob. Ecol. Biogeogr.* **22**, 982-993 (2013).
42. Santiago, L.S., Kitajima, K., Wright, S.J. & Mulkey, S.S. Coordinated changes in photosynthesis, water relations and leaf nutritional traits of canopy trees along a precipitation gradient in lowland tropical forest. *Oecologia* **139**, 495-502 (2004).
43. Givnish, T.J. Adaptive significance of evergreen vs. deciduous leaves: Solving the triple paradox. *Silva Fenn.* **36**, 703-743 (2002).
44. Defossez, E., Pellissier, L., Rasmann, S. & Swenson, N. The unfolding of plant growth form-defence syndromes along elevation gradients. *Ecol. Lett.* **21**, 609-618 (2018).
45. Schwery, O. *et al*. As old as the mountains: the radiations of the Ericaceae. *New Phytol.* **207**, 355-367 (2015).
46. Grabherr, M.G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652 (2011).
47. Fu, L.M., Niu, B.F., Zhu, Z.W., Wu, S.T. & Li, W.Z. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).
48. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).
49. Haas, B.J. *et al*. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494-1512 (2013).
50. Murray, M.G. & Thompson, W.F. Rapid isolation of high molecular-weight plant DNA. *Nucl. Acid. Res.* **8**, 4321-4325 (1980).
51. Patel, R.K. & Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *Plos One* **7** (2012).
52. Bankevich, A. *et al*. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
53. Qu, X.J., Moore, M.J., Li, D.Z. & Yi, T.S. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **15** (2019).
54. Ebersberger, I., Strauss, S. & von Haeseler, A. HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC. Evol. Biol.* **9**, 157 (2009).
55. Emms, D.M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
56. Katoh, K. & Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).

57. Kuck, P. & Longo, G.C. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front. Zool.* **11**, 81 (2014).
58. Nguyen, L.T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2015).
59. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518-522 (2018).
60. Wu, M., Chatterji, S. & Eisen, J.A. Accounting for alignment uncertainty in phylogenomics. *Plos One* **7**, e30288 (2012).
61. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinfor.* **19**, 153 (2018).
62. Xia, X. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* **35**, 1550-1552 (2018).
63. Tamura, K. *et al*. Estimating divergence times in large molecular phylogenies. *Proc. Natl. Acad. Sci. USA* **109**, 19333-19338 (2012).
64. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547-1549 (2018).
65. Liu, Z.W., Jolles, D.D., Zhou, J., Peng, H. & Milne, R.I. Multiple origins of circumboreal taxa in *Pyrola* (Ericaceae), a group with a Tertiary relict distribution. *Ann. Bot.-London* **114**, 1701-1709 (2014).
66. Popp, M., Mirre, V. & Brochmann, C. A single Mid-Pleistocene long-distance dispersal by a bird can explain the extreme bipolar disjunction in crowberries (Empetrum). *Proc. Natl. Acad. Sci. USA* **108**, 6520-6525 (2011).
67. Tanai, T. & Onoe, T. A Mio-Pliocene flora from the Ningyo-toge area on the border between Tottori and Okayama prefectures, Japan. *Geol. Surv. Japan Rep.* **187** (1961).
68. Yoichi, W., Jin, X.F., Peng, C.I., Tamaki, I. & Tomaru, N. Contrasting diversification history between insular and continental species of three-leaved azaleas (*Rhododendron* sect. *Brachycalyx*) in East Asia. *J. Biogeogr.* **44**, 1065-1076 (2017).
69. Matzke, N.J. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* (2013).
70. Matzke, N.J. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island Clades. *Syst. Biol.* **63**, 951-970 (2014).
71. R Core Team R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria; 2019).
72. Rabosky, D.L. *et al*. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* **5**, 701-707 (2014).
73. Karger, D.N. *et al*. Climatologies at high resolution for the earth's land surface areas. *Sci Data* **4**, 170122 (2017).

74. Adams, D.C. A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Syst. Biol.* **63**, 685-697 (2014).
75. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877-884 (1999).
76. Revell, L.J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217-223 (2012).
77. Cooper, N., Jetz, W. & Freckleton, R.P. Phylogenetic comparative approaches for studying niche conservatism. *J. Evol. Biol.* **23**, 2529-2539 (2010).
78. Zachos., J., Pagani., M., Sloan., L., Thomas., E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* (2001).
79. Hansen, J., Sato, M., Russell, G. & Kharecha, P. Climate sensitivity, sea level and atmospheric carbon dioxide. *Philos. T. R. Soc. A* **371** (2013).

Figures

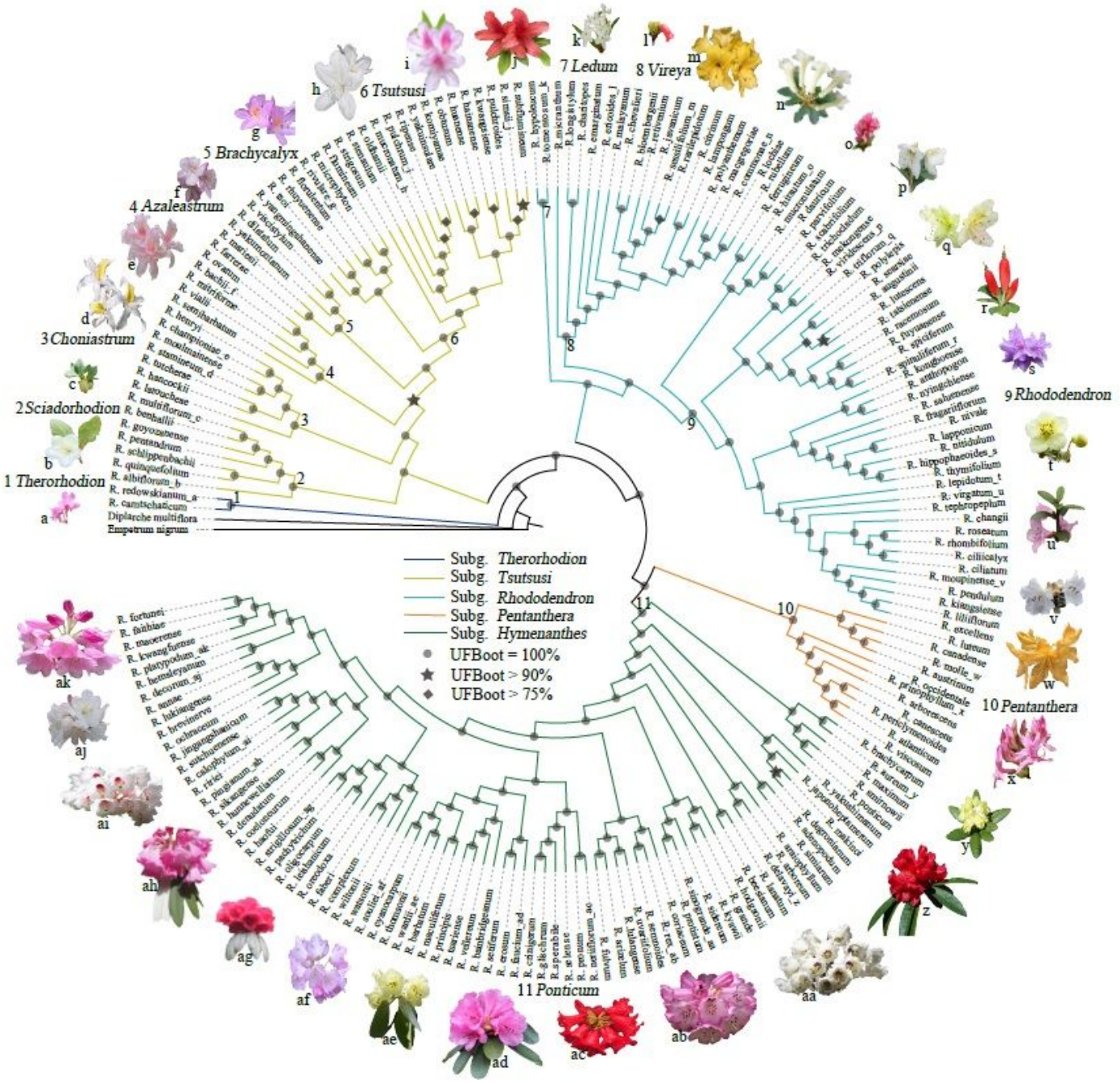

Figure 1

ML tree of Rhododendron reconstructed from the concatenated alignment of CDS sequences of 3437 OGs. The 11 sections are noted above the branches and around the tree by numbers 1-11. Flower photographs are not to scale. The lowercase letters (a-ak) following the species names correspond to the photographs of these species around the tree.

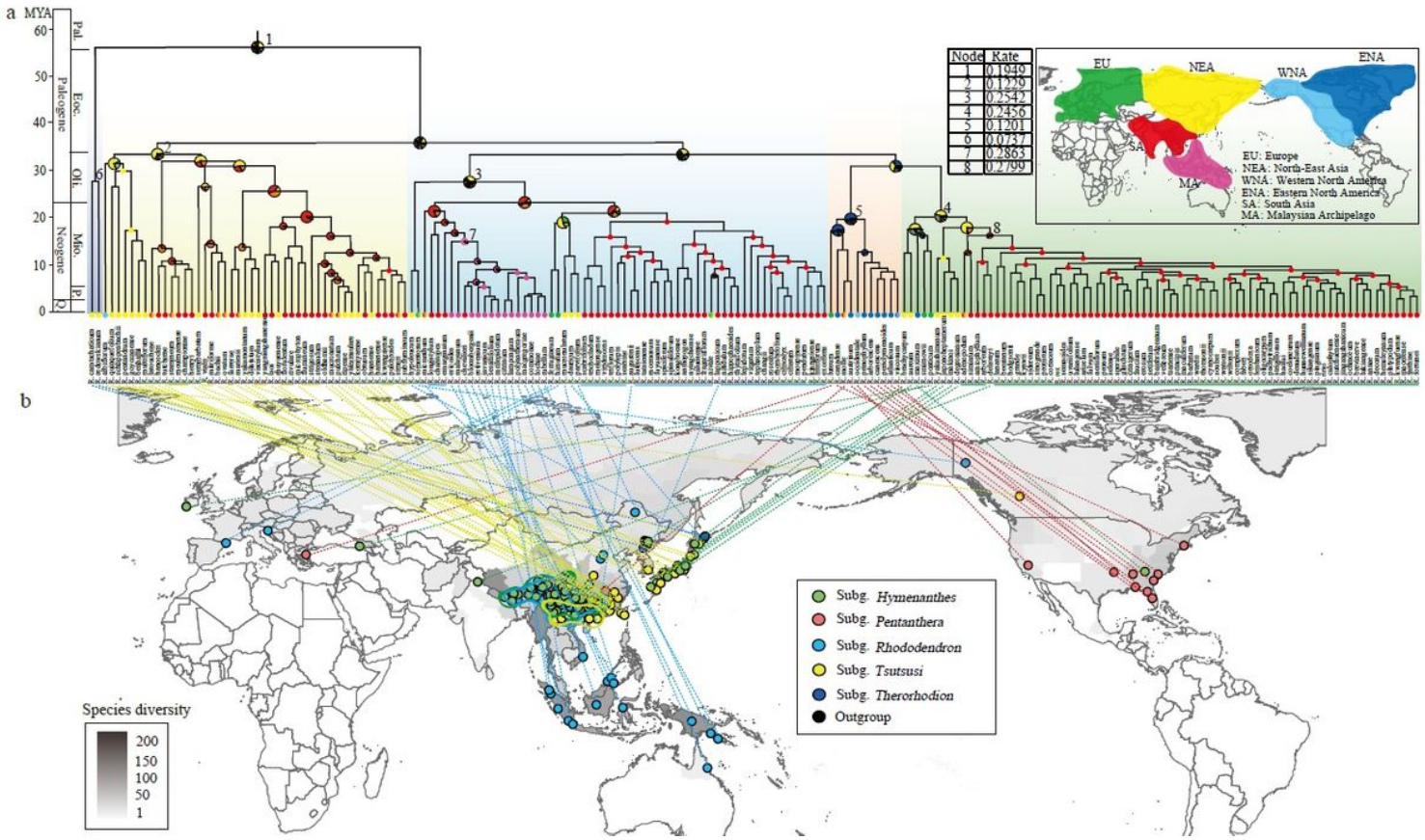

Figure 2

Chronogram and ancestral range estimates of Rhododendron. a, Divergence times estimated using RelTime based on the concatenated CDS sequences of 3437 OGs, and the ancestral areas reconstructed by BioGeoBEARS using the DEC + j model. The probabilities for the ancestral ranges are illustrated in pie charts color-coded by geographical regions on the world map. Diversification rate shifts are noted on the tree with numbers corresponding to those in the table above the phylogeny. Pal., Paleocene; Eoc., Eocene; Oli., Oligocene; Mio., Miocene; P., Pliocene; Q., Quaternary. b, Sample collection sites. The coordinates of the collection sites of 202 species are mapped to the world map of Rhododendron species diversity, which was constructed by the recorded species distributions at the administrative unit level. The colored solid circles in front of the names of subgenera correspond to the colored background on the phylogeny (a). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

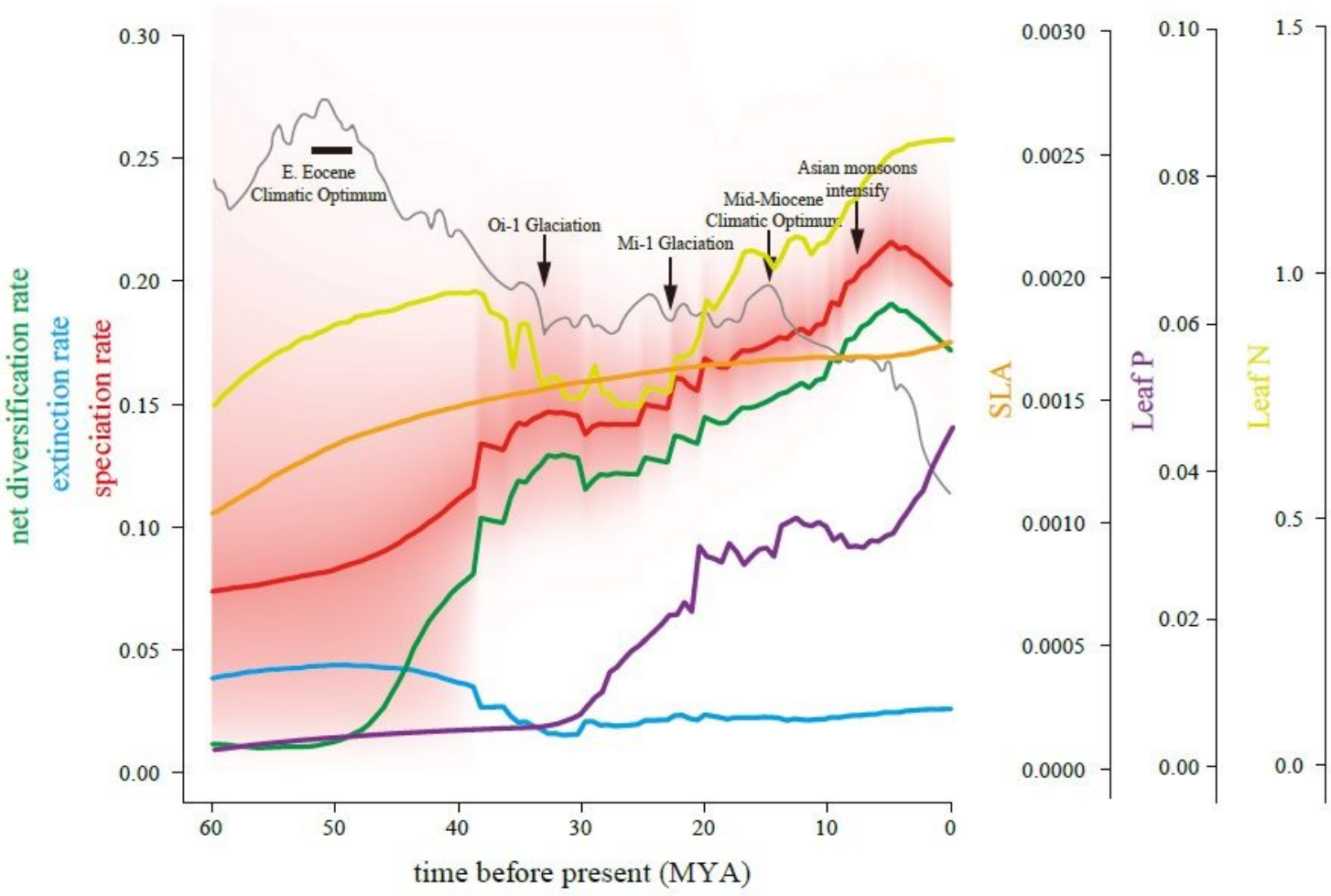

Figure 3

Median rates of net diversification, extinction and speciation, and macroevolutionary rates of leaf functional traits. The unit of diversification is speciation events per million years; the rates of leaf functional traits are unitless. Global surface temperature changes 78 are denoted by the dark gray curve, with several major geologic and climatic events 79 noted by black arrows or box.

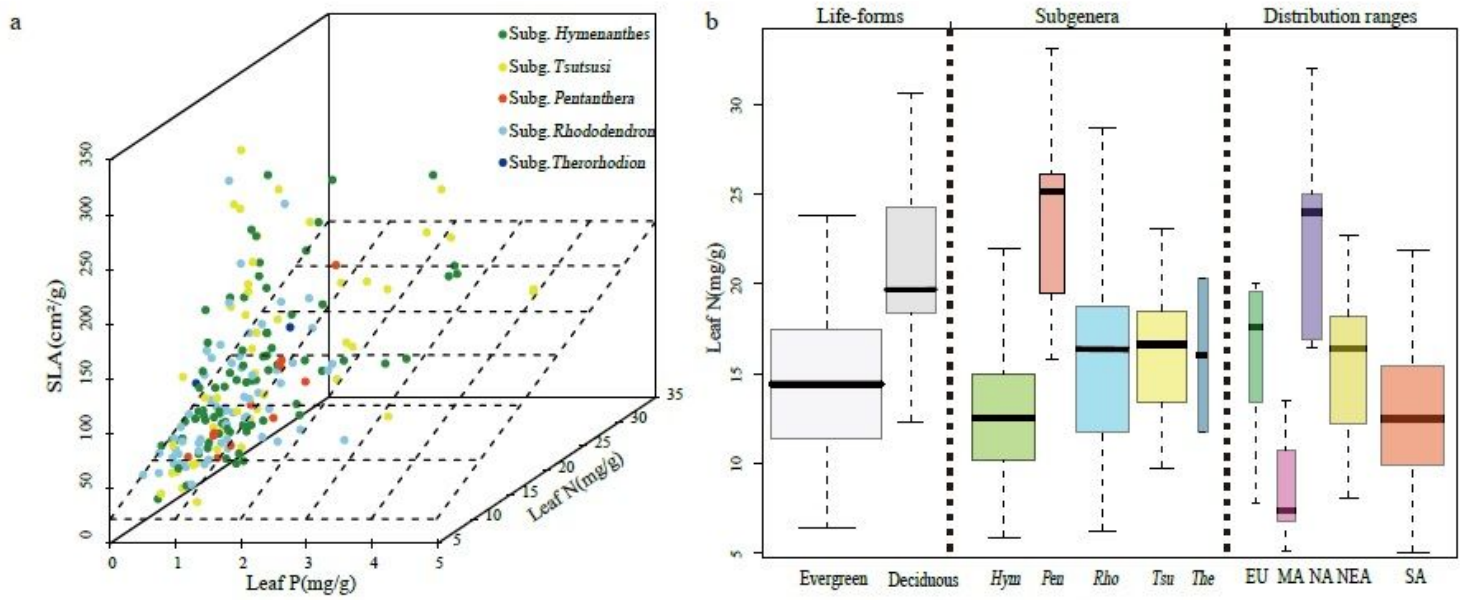

Figure 4

Scatterplot showing the relationships among three leaf traits in the leaf economics spectrum (a), and boxplot showing the variation of leaf N (mg/g) among different life-forms, subgenera, and distribution ranges (b). The width of the boxplot is proportional to the number of species. Hym, Sub. Hymennanthes; Pen, Sub. Pentanthera; Rho, Sub. Rhododendron; Tsu, Sub. Tsutsusi; The, Sub. Therorhodion; EU, Europe; MA, Malaysian Archipelago; NA, North America; NEA, North-East Asia; SA, South Asia.

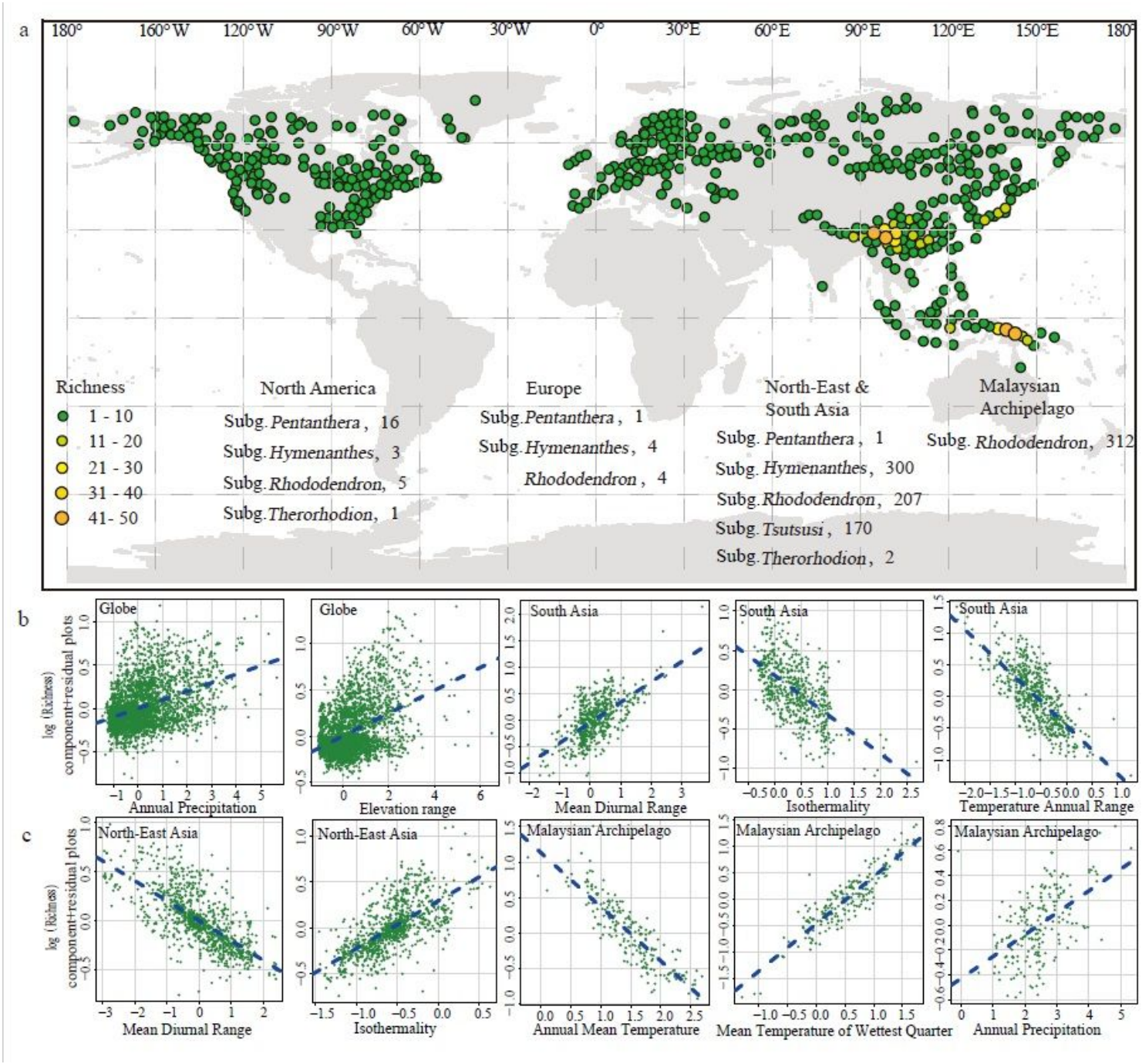

Figure 5

Global patterns of Rhododendron species diversity and the species composition in each geographical region (a), and the correlations between species richness and environmental variables at the global and regional (b and c) scales across grid cells of 50 km × 50 km. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf
- SupplementaryTable1Samplingandsequencedata.xlsx
- SupplementaryTable3Stochasticmappingstatistics.xlsx
- SupplementaryTable8Speciesrichnessanddataofenvironmentalvariables.xlsx
- SupplementaryTable9Statisticsofleaffunctionaltraits.xlsx
- SoftwareandCode.pdf
- nrreportingsummary.pdf