The origin and speciation of orchids

Oscar A. Pérez-Escobar[1]*[†], Diego Bogarín[2,3]*, Natalia A. S. Przelomska[1,4]*, James D. Ackerman[5], Juan A. Balbuena[6], Sidonie Bellot[1], Roland P. Bühlmann[7], Betsaida Cabrera[8], Jose Aguilar Cano[9], Martha Charitonidou[10], Guillaume Chomicki[11], Mark A. Clements[12], Phillip Cribb[1], Melania Fernández[2], Nicola S. Flanagan[13], Barbara Gravendeel[3], Eric Hágsater[14], John M. Halley[10], Ai-Qun Hu[15], Carlos Jaramillo[16], Anna Victoria Mauad[17], Olivier Maurin[1], Robert Müntz[18], Ilia J. Leitch[1], Lan Li[19], Raquel Negrão[1], Lizbeth Oses[2], Charlotte Phillips[1,4], Milton Rincon[9], Gerardo A. Salazar[20], Lalita Simpson[21], Eric Smidt[17], Rodolfo Solano-Gomez[22], Edicson Parra-Sánchez[23], Raymond L. Tremblay[5], Cassio van den Berg[24], Boris Stefan Villanueva Tamayo[9], Alejandro Zuluaga[25], Alexandre R. Zuntini[1], Mark W. Chase[1,26][†], Michael F. Fay[1][†], Fabien L. Condamine[27][†], Felix Forest[1][†], Katharina Nargar[19,21,28][†], Susanne S. Renner[29][†], William J. Baker[1][†] and Alexandre Antonelli[1,30,31,32,33][†]

[1]Royal Botanic Gardens, Kew, London, TW9 3AE, UK; [2]Lankester Botanical Garden, University of Costa Rica, P.O. Box 302-7050, Cartago, Costa Rica; [3]Naturalis Biodiversity Centre, Leiden, CR 2333, the Netherlands; [4]University of Portsmouth, Portsmouth, PO1 2DY, UK; [5]University of Puerto Rico – Rio Piedras, San Juan, PR, 00925-2537, USA; [6]ICBiBE, Universitat de València, Valencia, 13-46010, Spain; [7]Swiss Orchid Foundation, Schönenbuch, 4124, Switzerland; [8]Jardín Botánico Rafael Maria Moscoso, Santo Domingo, 21-9, Dominican Republic; [9]Jardín Botánico Jose Celestino Mutis, Bogota, 111071, Colombia; [10]University of Ioannina, Ioannina, GR 45110, Greece; [11]Durham University, Durham, DH13LE, UK; [12]Centre for Australian National Biodiversity Research (joint venture between Parks Australia and CSIRO), GPO Box 1700, Canberra, ACT, 2601, Australia; [13]Universidad Pontificia Javeriana, Seccional Cali, Cali, 760031, Colombia; [14]Herbarium AMO, Mexico City, 11000, Mexico; [15]Singapore Botanic Gardens, 1 Cluny Road, Singapore, 257494, Singapore; [16]Smithsonian Tropical Research Institute, Apartado, Panama City, 0843-03092, Panama; [17]Universidade Federal do Paraná, Curitiba, 19031, Brazil; [18]Reserva Biológica Guaitil, Eisenstadt, 7000, Austria; [19]National Research Collections Australia, Commonwealth Industrial and Scientific Research Organisation (CSIRO), GPO Box 1700, Canberra, ACT, 2601, Australia; [20]Universidad Nacional Autónoma de México, Mexico City, 04510, Mexico; [21]Australian Tropical Herbarium, James Cook University, GPO Box 6811, Cairns, Qld, 4878, Australia; [22]Instituto Politécnico Nacional, CIIDIR unidad Oaxaca, Oaxaca, 71230, Mexico; [23]University of Sheffield, Sheffield, S10 2AH, UK; [24]Universidade Estadual de Feira de Santana, Feira de Santana, 44036-900, Brazil; [25]Universidad del Valle, Cali, 760042, Colombia; [26]Department of Environment and Agriculture, Curtin University, Perth, WA, 6102, Australia; [27]Institut des Sciences de l'Evolution de Montpellier (Université de Montpellier|CNRS|IRD|EPHE), Place Eugène Bataillon, Montpellier, 34000, France; [28]Scientific Research Organisation (CSIRO), GPO Box 1700, Canberra, ACT, 2601, Australia; [29]Washington University in St. Louis, St Louis, MO 63130, USA; [30]Department of Biological and Environmental Sciences, Gothenburg Global Biodiversity Centre, Gothenburg, 417 56, Sweden; [31]University of Gothenburg, Gothenburg, 417 56, Sweden; [32]Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, 430074, China; [33]Department of Biology, University of Oxford, Oxford, OX1 3SZ, UK

Authors for correspondence:
Oscar A. Pérez-Escobar
Email: *o.perez-escobar@kew.org*

William J. Baker
Email: *w.baker@kew.org*

Alexandre Antonelli
Email: *a.antonelli@kew.org*

Summary

• Orchids constitute one of the most spectacular radiations of flowering plants. However, their origin, spread across the globe, and hotspots of speciation remain uncertain due to the lack of an up-to-date phylogeographic analysis.

• We present a new Orchidaceae phylogeny based on combined high-throughput and San-ger sequencing data, covering all five subfamilies, 17/22 tribes, 40/49 subtribes, 285/736 genera, and *c.* 7% (1921) of the 29 524 accepted species, and use it to infer geographic range evolution, diversity, and speciation patterns by adding curated geographical distributions from the World Checklist of Vascular Plants.

• The orchids' most recent common ancestor is inferred to have lived in Late Cretaceous Laurasia. The modern range of Apostasioideae, which comprises two genera with 16 species from India to northern Australia, is interpreted as relictual, similar to that of numerous other groups that went extinct at higher latitudes following the global climate cooling during the Oligocene. Despite their ancient origin, modern orchid species diversity mainly originated over the last 5 Ma, with the highest speciation rates in Panama and Costa Rica.

*Lead authors.
†Senior authors.

• These results alter our understanding of the geographic origin of orchids, previously proposed as Australian, and pinpoint Central America as a region of recent, explosive speciation.

Introduction

The angiosperm tree of life is characterised by the rise and demise of species, leading to species-rich and depauperate lineages coexisting in time and space (Magallón *et al.*, 2019; Tietje *et al.*, 2022). Investigating factors behind angiosperm diversification requires ancient, species-rich clades thriving across the globe and having a fossil record. One such clade is the Orchidaceae. With 29 524 species (Chase *et al.*, 2015; Christenhusz & Byng, 2016; Govaerts *et al.*, 2021), orchids are among the most species-rich groups of flowering plants, with molecular dating studies having estimated their initial diversification at 112–76 million years ago (Ma) (Ramírez *et al.*, 2007; Gustafsson *et al.*, 2010; Chomicki *et al.*, 2015; Givnish *et al.*, 2016; Serna-Sánchez *et al.*, 2021). This diversification time is partly supported by the orchid fossil record, which includes leaf compressions and pollinaria dated from the early Eocene through the mid-Miocene found in different deposits (Ramírez *et al.*, 2007; Conran *et al.*, 2009; Poinar Jr., 2016a,b; Poinar & Rasmussen, 2017). Most of these fossils have been credibly assigned to subfamilies and tribes.

An updated biogeographic study of this old, species-rich, and cosmopolitan family (Fig. 1a–d), which is uniquely diverse in animal and fungal interactions (Selosse *et al.*, 2022; Ackerman *et al.*, 2023; Karremans *et al.*, 2023), and adaptations to different habitats (Fig. 1e,f), will help our understanding of monocot clades with wide distribution ranges and diversity of adaptations (e.g. Arecaceae (Couvreur *et al.*, 2011), Bromeliaceae (Benzing, 2000)). Such a study can build on several in-depth analyses that have focused on subclades of orchids (Bouetard *et al.*, 2010; Guo *et al.*, 2012; Freudenstein & Chase, 2015; Pérez-Escobar *et al.*, 2017; Nauheimer *et al.*, 2018). In a benchmark study of the biogeographic history and diversification of the entire family, Givnish *et al.* (2016) used a fossil-calibrated plastid tree for 173 genera (out of 736), representing all five subfamilies, along with 10 outgroup species representing six families (Asteliaceae, Blandfordiaceae, Boryaceae, Hypoxidaceae, and Lanariaceae, with Iridaceae as a more distant outgroup). Using a likelihood approach (Matzke, 2013) and a single consensus ultrametric tree, their analyses relied on a matrix where terminals, usually representing genera, were coded for their entire distribution. The results pointed to an origin and initial diversification of orchids in Australia during the mid-Cretaceous, 120–90 Ma. Specifically, the range estimates for the Orchidaceae stem node indicated Australia as *c.* 40% likely, Neotropics plus Australia as *c.* 20%, and all other ranges together as *c.* 40%. At 90 Ma, Australia was at high latitudes connected to Antarctica and part of the southern supercontinent, Gondwana, with connections to South America, and the authors, therefore, stressed the likely importance of expansion across Antarctica. A Gondwanan origin for Orchidaceae was also suggested by Chase (2001).

Givnish *et al.* (2016: table 1) also addressed the correlates of diversification rates through time and inferred Southeast Asia as the region with the highest net diversification rates. Studying how diversification rates are linked to geographical variation helps us understand not only the pace at which extant species diversity has accumulated but also the correlated biotic and abiotic variables (Condamine *et al.*, 2013; Velasco & Pinto-Ledezma, 2022). Although orchid species diversity clearly is unevenly distributed across their distribution (Vitt *et al.*, 2023), no study has yet assessed the relationship between the distribution of orchid species richness and underlying speciation rates at a global scale.

Here, we revisit the biogeographic history of Orchidaceae and infer geographic patterns of speciation, using a greatly expanded taxon sampling (1921 species, 285 genera) as compared to previous studies, with a particular focus on early divergences. We generate a phylogenomic framework by combining high-throughput and Sanger sequencing data. Our spatial information derives from herbarium-vouchered, georeferenced occurrence data, sourced from the Global Biodiversity Information Facility (GBIF; https://www.gbif.org) and the RAINBIO mega-database (Dauby *et al.*, 2016). This dataset was further vetted using the spatial distribution data of the World Checklist of Vascular Plants (WCVP; Govaerts *et al.*, 2021). Based on these datasets, we: (1) infer the evolutionary and biogeographical history of the orchids; (2) test the hypothesis of an Australasian origin for their most recent common ancestor; and (3) revisit the hypothesis that Southeast Asia is the region with the highest current diversification rate of orchids.

Materials and Methods

Taxon sampling, DNA library preparation and sequencing

Our phylogenomic framework was assembled using a high-throughput-sequencing dataset, complemented with Sanger data for single genes. The high-throughput-sequencing dataset focused on the genes targeted by the Angiosperms353 probe set (Johnson *et al.*, 2019; Baker *et al.*, 2022) and includes 448 species, representing 285/736 genera, 40/49 subtribes, 17/22 tribes, and 5/5 subfamilies (in the classification of Chase *et al.*, 2015). These data were generated by the Plant and Fungal Trees of Life Project (Baker *et al.*, 2022) at the Royal Botanic Gardens, Kew, and the Genomics for Australian Plants Consortium (https://www.genomicsforaustralianplants.com/). Nineteen monocot species were included as outgroup taxa, with *Dioscorea caucasica* (Dioscoreales) used as a rooting terminal, for a total of 467 species in the high-throughput-sequencing dataset. Illumina sequencing reads were newly produced for 377 samples (vs Pérez-Escobar *et al.*, 2021b) from expertly curated specimens housed in the Kew (K) and Australian National Herbarium (CANB),
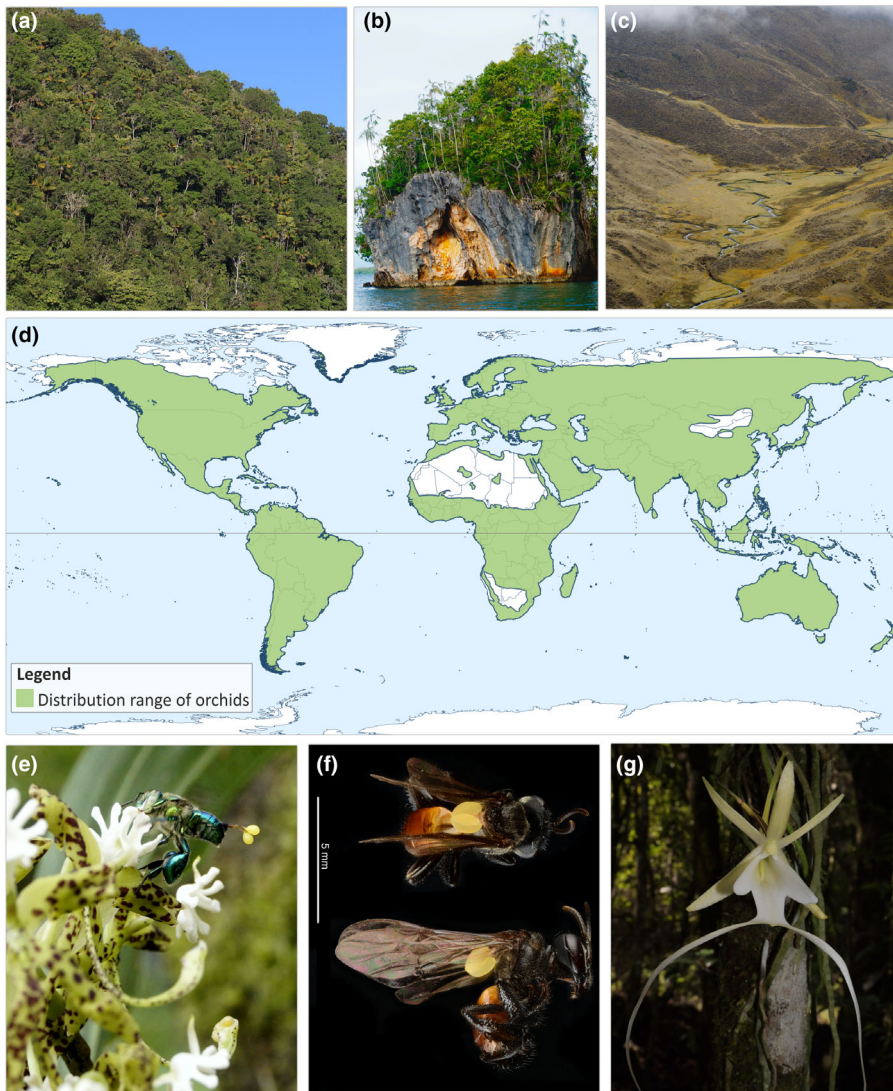
Fig. 1 Diversity of orchid habitats (a–c), modern distribution range (d) and pollination systems in Orchidaceae (e–g). (a) Tropical cloud forests; (b) Tropical lowland wet forests; (c) High-elevation grasslands; (e) the Neotropical epiphyte *Cycnoches guttulatum* Schltr., pollinated by male euglossine bees collecting aromatic compounds; (f) *Trigona* bee carrying a pollinarium of a Neotropical *Xylobium elongatum* (upper panel: dorsal view; lower panel: side view); (g) the epiphytic and leafless *Dendrophylax sallei* (Rchb.f.) Benth. ex-Rolfe., a Neotropical orchid pollinated by moths. Photos: Oscar A. Pérez-Escobar & Diego Bogarín.

including 23 species sampled from type material and silica-gel dried tissue (Supporting Information Table S1).

DNA extraction was conducted from 1.5 to 4 mg of silica-dried fresh tissue or herbarium specimen tissue, constituting leaf or occasionally flower. We used a sterile steel bead for mechanical disruption in a SPEX® sample prep tissue homogeniser (SPEX Inc., Metusen NJ, USA), applying two to four disruption cycles of 1 min at 1350 Hz. Next, we added 750 µl CTAB with 2% v/v 2-mercaptoethanol (Doyle & Doyle, 1990) for a 30-min incubation at 65°C, followed by either a 4 h incubation at 57°C (silica-dried samples) or an overnight incubation at 57°C (herbarium specimens). We then added 750 µl SEVAG (24 : 1 chloroform : isoamyl alcohol), placed the samples on a shaking incubator for 30 min, and centrifuged the samples to induce phase separation. Next, we separated off the upper phase containing the DNA fraction and added to this a 0.7× volume of isopropanol for a 4–5-d incubation at −20°C to encourage DNA precipitation. Finally, we subjected the samples to two rounds of centrifugation and

supernatant removal, first to remove the isopropanol and second to wash the DNA pellet with 75% ethanol. Subsequently, the DNA pellet was eluted into 70 µl 10 mM Tris-Cl and purified using a paramagnetic bead clean-up method with a 2 : 1 ratio of Ampure XP beads (Beckman Coulter, Brea, CA, USA) to DNA elution. To gauge the size distribution of the genomic DNA and, on this basis, determine which samples might need shearing before library preparation, we visualised the samples on a 1% agarose gel. Using 18–200 ng starting material, we prepared DNA libraries using NEB Next Ultra II Library Prep Kits (New England Biolabs, Ipswich, MA, USA), according to the manufacturer's protocol, but with half-volume reactions. Where possible (silica-dried material and herbarium specimens with DNA of higher integrity), we aimed for insert sizes of *c.* 350 bp, by shearing the DNA with a Covaris ME220 Focussed Ultrasonicator (Covaris LLC, Woburn, MA, USA). The libraries were indexed with NEBNext Multiplex Oligos for Illumina (New England Biolabs) and amplified using 6–14 PCR cycles. The yield was

estimated using a Quantus fluorometer (Promega), and fragment size distribution was estimated using either a 4200 TapeStation system or an Agilent 2100 BioAnalyser (Agilent Technologies, Santa Clara, CA, USA). Finally, we used the Angiosperms353 probe set to enrich these genomic libraries for 353 low-copy nuclear genes (Johnson *et al.*, 2019), modifying the equimolar pooling of libraries into groups of 10–20 libraries per capture reaction and total input of up to 1 μg. Furthermore, for the herbarium samples, we applied a lowered hybridisation temperature – of 62°C and a prolonged hybridisation time – of 40 h. The sequencing of paired-end genomic libraries (150 bp × 2) was conducted on an Illumina HiSeq by Macrogen (Geumcheon, South Korea).

The Sanger sequencing dataset was assembled through the SuperCRUNCH pipeline (Portik & Wiens, 2020). This pipeline was executed using an initial set of 24 172 sequences of the nuclear ribosomal ITS and the plastid *mat*K markers – two commonly sequenced, informative loci for Orchidaceae and other angiosperms (Grace *et al.*, 2021) – obtained from GenBank (https://www.ncbi.nlm.nih.gov/nuccore). The following parameters were used: (1) search terms of 'ITS; ITS1; ITS2; internal transcribed spacer 1; partial sequence; 5.8S ribosomal RNA gene; complete sequence; internal transcribed spacer 2' for ITS, and '*matK*; *trnK*; maturase K; *trnK* gene, intron; and maturase K *matK* gene' for *matK*; (2) a list of 285 generic (accepted) names that were sampled in the high-throughput-sequencing dataset; (3) similarity searches between sequences using MEGABLAST (Camacho *et al.*, 2009), retaining a maximum of 200 hits per search. After removing duplicated sequences and obvious contaminants (defined as sequences with misplaced positions given current taxonomic views), we retained sequences of nrITS and plastid *matK* for 2060 species, of which 64% had sequences produced from the same voucher specimen (Table S2).

High-throughput and Sanger sequencing data analyses

Illumina libraries were quality-assessed with FASTQC software v.0.11 (available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Paired-end reads were quality-filtered and adapter-trimmed using TRIMGALORE! v.0.6.4 available at (https://github.com/FelixKrueger/TrimGalore), using the following parameters: (1) *-q 30* (minimum Phred score), (2) *length* 20 (minimum read length), (3) retaining read pairs that passed the quality thresholds. *In silico* retrieval of the Angiosperms353 coding loci was conducted using the pipeline HYBPIPER v.2.1.6 (Johnson *et al.*, 2016) and the same parameters and software described in Pérez-Escobar *et al.* (2021b; also see Methods S1). Next, low-copy nuclear and Sanger markers were aligned using the software MAFFT v.7.4 (Katoh & Standley, 2013) in conjunction with the iterative refinement method FFT-NS-I, implementing a maximum of 1000 iterations. To ameliorate potential bias introduced by missing data in phylogenetic analyses, sequences shorter than 50% of the total alignment length were excluded. Subsequently, alignments were filtered for misaligned positions using TAPER v.1.0 (Zhang *et al.*, 2021), with a cut-off value of 1 (flag *−c*), and inspected by

eye with GENEIOUS v.8.0 (available at https://www.geneious.com/). The Sanger/Angiosperms353 nucleotide alignments are accessible at doi: 10.6084/m9.figshare.22245940.

Distance-based, maximum likelihood, Bayesian, and multispecies coalescence phylogenomic inferences

We computed maximum likelihood (ML) trees from the individual Sanger and Angiosperms353 nucleotide alignments using RAxML v.8.0 (Stamatakis, 2014) with 500 rapid bootstrap replicates (flags -# 500 and -x) and the GTR+ Γ nucleotide substitution model. Additionally, a multispecies coalescent (MSC) tree was inferred from the individual Angiosperms353 gene trees produced by RAxML using the software ASTRAL-III v.5.6 (Zhang *et al.*, 2018) and the flag *−t* 2 (full tree annotation), after collapsing bipartitions with a likelihood bootstrap percentage (LBP) < 20% (achieved with the function new_ed of the package *newick_utils* v.1.6.0, available at https://bioweb.pasteur.fr/packages/pack@newick-utils@1.6). To visualise the proportion of gene tree quartets that agreed with the species tree, we produced quartet support pie charts for every branch represented in the species tree, using the full annotation produced by ASTRAL-III. We also visualised the proportion of gene trees in agreement with each species' tree bipartition using PHYPARTS v.1.0, by labelling as informative any gene tree bipartition with a LBP > 20% (Smith *et al.*, 2015).

Because of the overall topological congruence between the ITS- and *matK*-derived trees and the Angiosperms353 low-copy nuclear MSC trees, as well as a goodness-of-fit test (Balbuena *et al.*, 2013; Pérez-Escobar *et al.*, 2016, see Methods S1; Notes S1), we proceeded to compute ML phylogenetic trees from: a supermatrix derived by concatenating the ITS and *matK* sequences; and a supermatrix derived from the concatenation of the Angiosperms353 loci, again using RAxML v.8.0 with the same settings previously specified, considering each of the two supermatrices as a single partition. Lastly, we inferred a consensus network using the 500 RAxML bootstrap replicates produced from the Angiosperms353 supermatrix and SPLITSTREE v.4.0 (Huson & Bryant, 2006), with a mean edge weight and a threshold value of 0.75 to filter out any splits not found in at least 75% of the bootstrap trees. The trees and consensus network are accessible at doi: 10.6084/m9.figshare.22245940.

Molecular-clock-dating analyses and species-level phylogeny assembly

Absolute age estimation analysis was conducted in two stages using the Bayesian framework implemented in BEAST v.2.6 (Bouckaert *et al.*, 2019), as follows:
(**1**) The backbone was inferred by subsampling the Angiosperms353 low-copy nuclear gene alignments using SORTADATE v.1.0 (Smith *et al.*, 2018). Here, we selected the top 25 low-copy nuclear gene alignments with the lowest root-to-tip variance coefficient (i.e. highest clock-likeness). This selection ensured the representation of the entire generic diversity as sampled by

the ML high-throughput dataset (i.e. 285 genera; see Methods S1; Tables S3, S4). This data subset was imported in BEAUTI v.2.6 (Bouckaert *et al.*, 2019) as unlinked partitions, with the same priors employed by Pérez-Escobar *et al.* (2021a), as follows: (1) the GTR nucleotide substitution model and a rate heterogeneity among sites modelled by a Γ distribution with four categories; (2) an uncorrelated log-normal relaxed molecular clock in combination with a prior clock rate interval of 0.0001–0.001 substitutions/site/Ma, modelled by a uniform distribution; (3) a birth–death tree process, modelled by a uniform distribution for the birth and relative death rates; (4) three secondary calibration points from Givnish *et al.* (2015), modelled by normal distributions ($\sigma = 1$) and located at the root node, that is, the most recent common ancestor (MRCA) of Dioscoreales, Asparagales, and Liliales (125 Ma), the MRCA of Orchidaceae (89 Ma), and the MRCA of subtribe Goodyerinae (32 Ma); (5) the ML consensus phylogram of generic lineages produced in RAxML v.8.0 as a starting tree; (6) 500 million generations, sampling every 100 000 generations and ensuring that all posterior values reached effective sample sizes (ESS) > 100. The proportion of informative and missing sequences per taxon sample is provided in Table S4. Finally, the support for the backbone ultrametric trees derived from BEAST was assessed by first computing a maximum credibility consensus tree from the Markov chain Monte Carlo (MCMC) posterior trees using TREEANNOTATOR v.2.6 (https://www.beast2.org/treeannotator/), using a burn-in threshold of 10%, and then counting the proportion of bipartitions falling on different support intervals.

(**2**) A species-level-ultrametric phylogeny of Orchidaceae was produced by inputting the ITS and *matK* alignments as unlinked partitions in BEAUTI v.2.6 (Bouckaert *et al.*, 2019), using the same priors as in (1). Furthermore, the ML consensus phylogram produced in RAxML v.8.0 from the ITS-*matK* supermatrix was used as a starting tree.

(**3**) To try to accommodate phylogenetic uncertainty, we constructed 10 ultrametric species trees through a novel pipeline (Fig. 2) instead of using a single consensus maximum clade credibility tree. We first randomly sampled 10 posterior trees derived from the BEAST analyses conducted on the genus-level Angiosperms353 and the ITS-*matK* Sanger species-level datasets. Then, for each genus represented in the Angiosperms353 chronograms, we pruned its counterpart clade sampled on the species-level Sanger chronogram and proceeded to graft it onto the corresponding stem of the Angiosperms353 chronogram. This operation was conducted on each pair of randomly sampled posterior trees (hence called posterior probability (PP) species trees). A detailed description of the pipeline is provided in the Methods S1.

Biogeographic analyses

Biogeographical state estimations were conducted on the 10 PP species trees (see previous section). We estimated areas of origin and geographic range using the ML approach of dispersal-extinction-cladogenesis (DEC, Ree & Smith, 2008) as implemented in the C++ version (Beeravolu & Condamine, 2016). We relied on DEC because in its C++ implementation, is a model that is scalable to the number of terminals and biogeographical areas that our study involves and because it models the main biogeographic phenomena that could have shaped the range evolution of the orchid family at higher taxonomic levels. The eight geographic areas considered were: (1) Palearctic, defined as Europe, Siberia, Central Asia, and Western Asia; (2) Nearctic, including all North America to the north of Tehuantepec in Mexico and excluding the southern tip of Florida; (3) Neotropics, including Central America, the Caribbean Islands, and South America; (4) Africa, defined as the whole African continent including Madagascar, and Arabian Peninsula; (5) Indomalaya, including India, southern China, and Wallacea; and (6) Australasia, defined as everything east of Heilprin-Lydekker's Line (Ali & Heaney, 2021). A time-stratified geographic model specified constraints on area connectivity by coding 0 if any two areas were not connected or 1 if they were connected during a given period based on paleogeographic reconstructions. Our first time slice covered the Late Cretaceous and early Palaeocene (100.5–60 Ma), corresponding to a time when Gondwana and Laurasia had already separated (de Lamotte *et al.*, 2015). The second time slice covered the late Paleocene to early Oligocene (60–30 Ma), and the third the early Oligocene to the present (30–0 Ma). To avoid biased results supporting ancestral areas in favour of hyperdiverse biomes, our sampling per biogeographic area used proportional sampling that reflected known diversity across biogeographic realms, with the Asian and American tropics hosting the highest levels of species richness (Vitt *et al.*, 2023). Our analyses included 905 species distributed in the Neotropics (6% of total known diversity in the area; Govaerts *et al.*, 2021), 623 from the Indomalaya region (6%), 347 from the Palearctic (5%), 221 from the Afrotropics (9%), 180 from Australasia (15%), and 44 from the Nearctic (10%).

Spatial analysis of species diversity and speciation rates

We queried 29 524 accepted orchid species names obtained from the WCVP through the GBIF (https://www.gbif.org/) and RAINBIO (https://gdauby.github.io/rainbio/index.html; Dauby *et al.*, 2016) databases. Our initial step involved accessing the GBIF database through the R package SPOCC (available at https://github.com/ropensci/spocc). Using the *occ* function, we downloaded up to 1000 records for each queried species name, ensuring that we only selected records linked to geographical coordinates and preserved specimens. The RAINBIO repository was manually accessed. Next, we conducted an automated filtering using the SPECIESGEOCODER package (Töpel *et al.*, 2017) to exclude duplicated records and those located within urban areas. Subsequently, we removed occurrences that fell outside global coastlines. Lastly, to mitigate the influence of misidentified records on downstream analyses (Maldonado *et al.*, 2015), we filtered out records for which the distribution did not match with the species distribution per botanical country (level 3 of the World Geographical Scheme for Recording Plant Distributions; Brummitt, 2001) provided by the WCVP. The original record database sourced from GBIF is available at doi: 10.15468/dl.
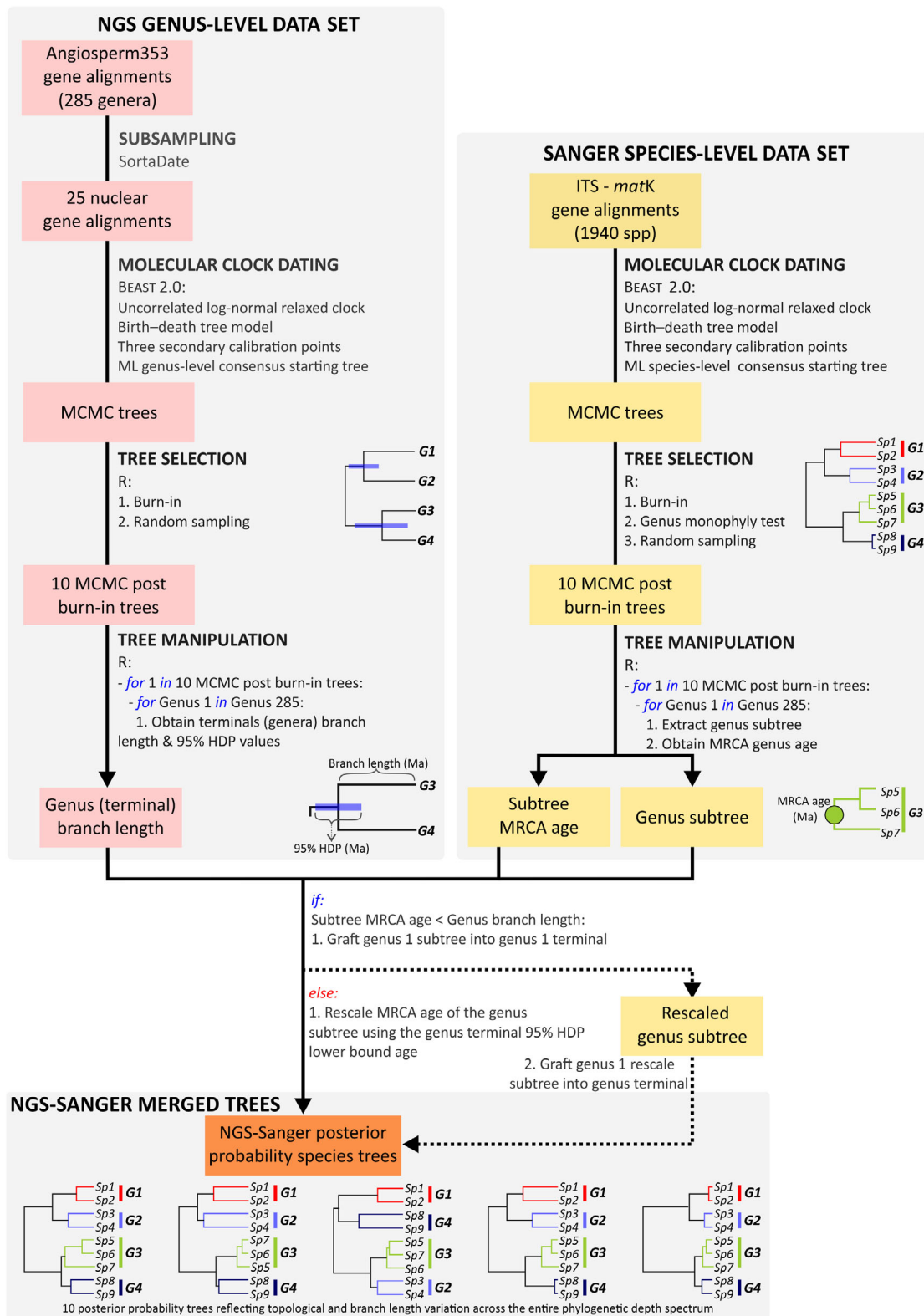
Fig. 2 An overview of the pipeline used to produce ultrametric NGS-Sanger merged posterior probability species trees (PP species trees). The pipeline uses as input a custom set of randomly sampled MCMC trees derived from independently conducted molecular-clock dating analyses on the genus-level Angiosperms353 and species-level Sanger gene alignments. It then produces a custom set of species-level trees that reflect branch length variation and support in deep (tribe, subtribe and genus) and recent (species) time as informed by the Angiosperms353 and Sanger datasets. A detailed description of the pipeline is provided in the Supporting Information Methods S1.

Fig. 3 Phylogenetic relationships of Orchidaceae. (a) Consensus tree network inferred from 200 bootstrap replicate maximum likelihood (ML) trees derived from the concatenated alignment of 339 low-copy nuclear genes. Circles at nodes represent bipartitions present in > 75% of the bootstrap ML trees (see Supporting Information Figs S1–S6 for a detailed on the consensus tree network). Non monophyletic groups are highlighted in bold and grey. Samples sequenced from typological material are highlighted in bold and pink. (b) Number of bipartitions (bp) with different likelihood bootstrap support (LBS) values derived from an ML tree inferred from the 339 low-copy nuclear gene alignment. (c) Number of bp with different LBS values derived from 339 nuclear ML gene trees. Grey bars represent poorly supported bipartitions, whereas yellow bars represent moderately to maximally supported bipartitions. (d) Normalised Q-score value derived from an ASTRAL-III analysis, computed from 339 nuclear ML gene trees. Photos: Oscar A. Pérez-Escobar, Diego Bogarín, Sebastian Vieria, Kerry Dressler, William J. Baker.

v2gwxv, and the filtered geographical records used in all downstream analyses will be available at doi: 10.6084/m9.figshare.22245940.

Estimating net diversification rates (speciation (λ) minus extinction (μ)) in plant lineages remains challenging (Louca & Pennell, 2021). Our approach focuses on speciation dynamics, as inferred from tip λ rates. This metric represents contemporary rates of speciation for a given lineage and is less prone to bias than net diversification rates (Title & Rabosky, 2019). To obtain tip λ rates, we fitted a time-dependent model using BAMM v.2.5.0 (Rabosky et al., 2013). This was informed by sampling fractions, contrasting accepted species number per genus from the WCVP database with the number of species sam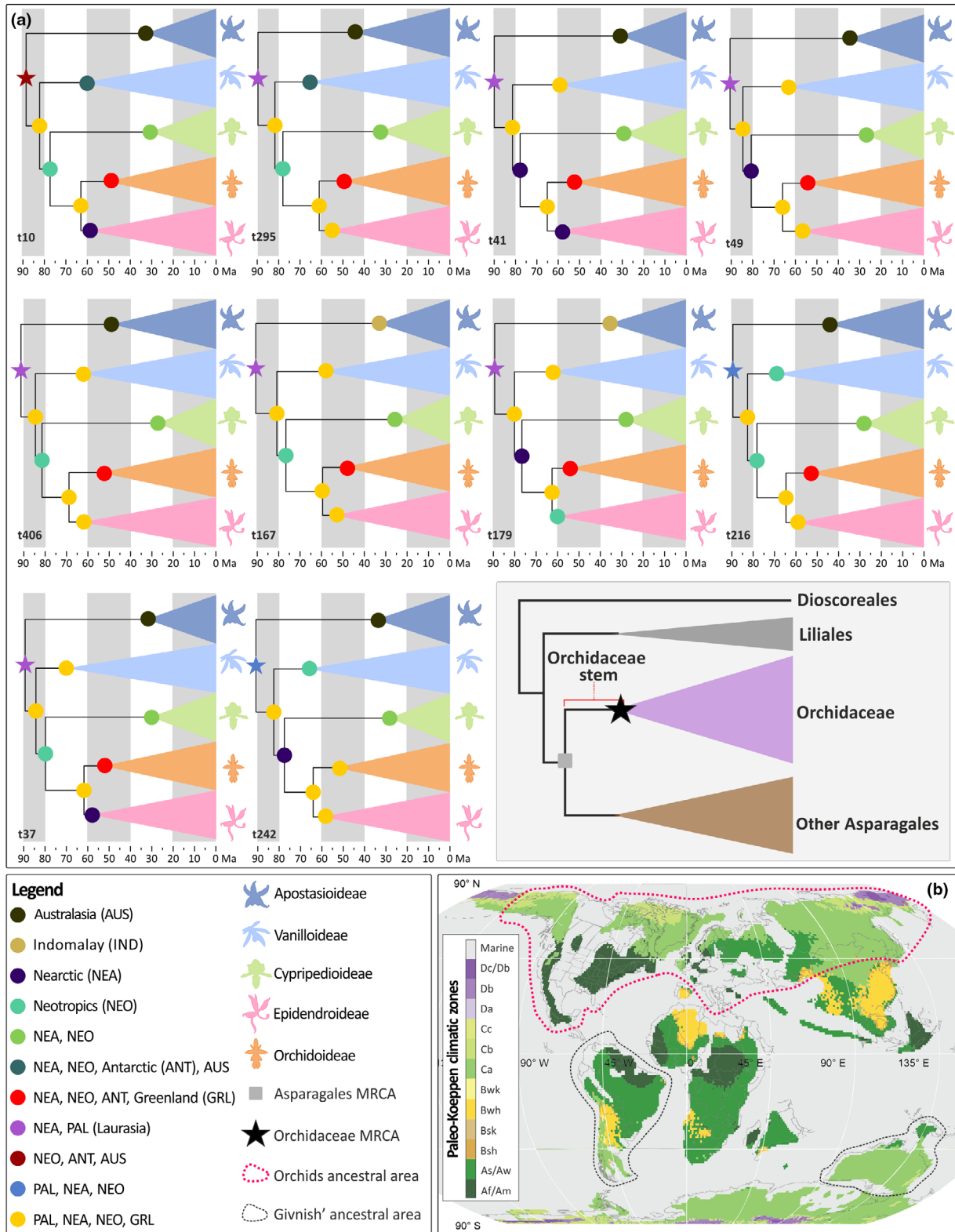pled in our phylogenetic analyses. Here, 54% of the genera sampled included sequences from 10 to 50% of their known species, 21% included > 50%,

and 25% included < 10% of their known species. This analysis was conducted on each of the 10 PP species trees generated by pruning and grafting the genus- and species-level chronograms produced by BEAST. We initially computed prior values for the initial λ and shift parameters, using the R package BAMMTOOLS (Rabosky et al., 2014), and performed 10 million generations, sampling the MCMC simulations every 10 000 generations. The BAMM output was analysed in BAMMtools to ensure that all analyses reached convergence (ESS values > 100). Mean speciation rates at present time (λ tip rates) were extracted for all terminals of the phylogenetic trees using the *getTipRates* function in BAMMtools, and then linked to the filtered geographical species occurrences.

Global patterns of species richness were inferred by calculating the average number of species per political and botanical country,

Fig. 4 Biogeographic history of the orchid family. (a) Ancestral areas at nodes inferred on the 10 posterior probability species trees as estimated by a DEC model (see the Materials and Methods section) and summarised to the five orchid subfamilies (Inset: A summary of the outgroup sampling considered in our study (the MRCA of orchids is indicated with a black star)). (b) A palaeoclimatic and tectonic plate reconstruction at 90–80 Ma from Burgener *et al.* (2023) showing the possible ancestral range of the orchid MRCA as estimated by Givnish *et al.* (2016) and this study. A detailed account of the first and second most likely ancestral areas estimated to the tribe level is provided on Supporting Information Table S5 (complete results including annotated trees with the most likely ancestral area at nodes and alternative ancestral areas, likelihoods and probabilities are available at 10.6084/m9.figshare.22245940).

and per grid cell (100 km × 100 km), using the filtered occurrence records and the average and maximum λ rate values per grid cell. All calculations were performed in the open-source software QGIS v.3.0 (available at https://www.qgis.org/en/site/forusers/download.html) and the R package KEWR (Walker, 2022).
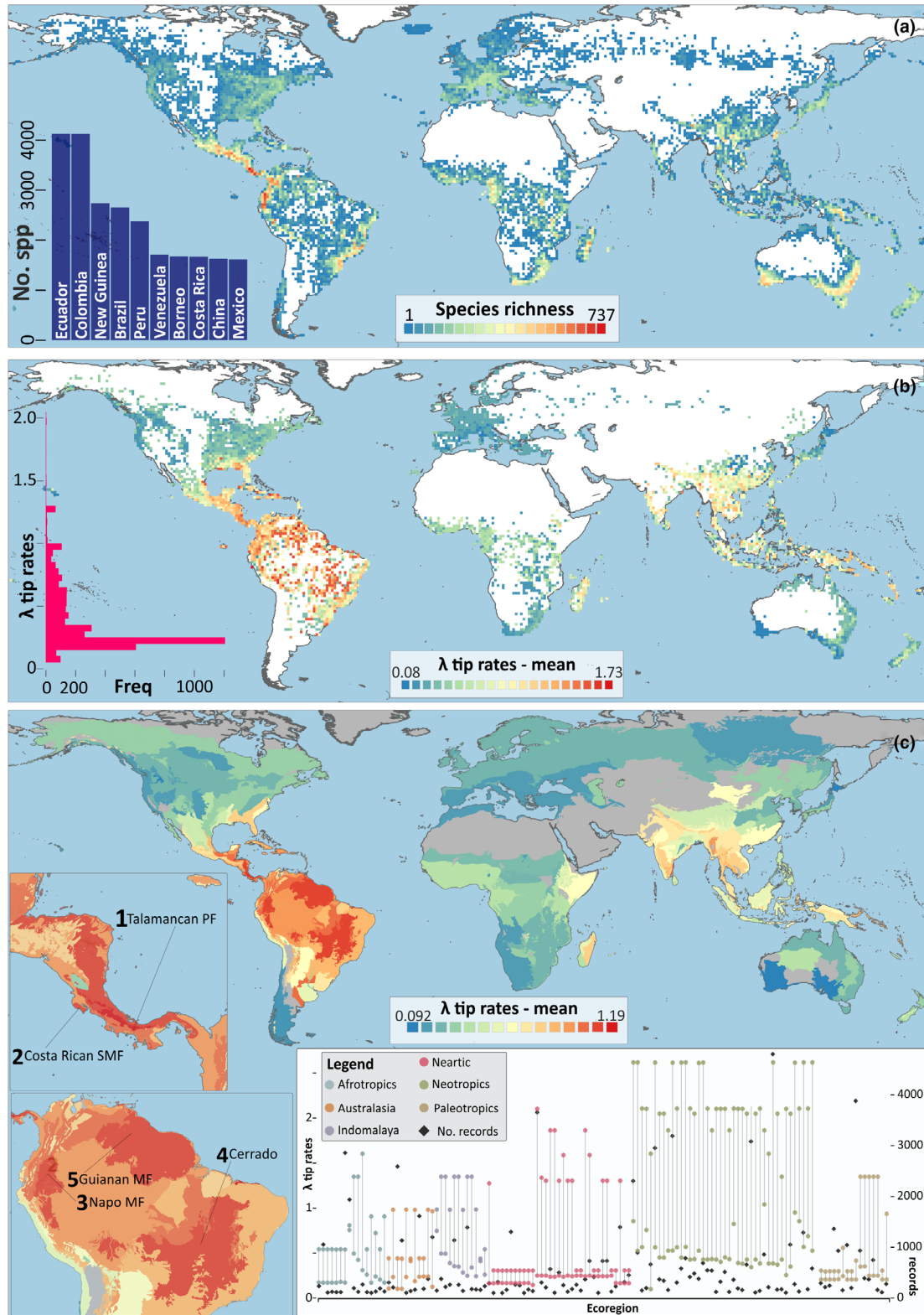
Fig. 5 Geography of speciation and species diversity of Orchidaceae. (a) Global patterns of species richness per grid cell (100 × 100 km), calculated from a curated database of geographical distribution records; reddish colours indicate higher numbers of species per grid cell whereas bluish colours indicate lower numbers of species per grid cell. A map of species richness per botanical country is provided in Supporting Information Fig. S10. (Inset: orchid species numbers reported for the 10 most orchid biodiverse countries). (b) Global patterns of mean λ tip rates (linear scale) per grid cell (100 × 100 km) as derived from the BAMM software; warm colours indicate higher λ tip rates whereas cold colours indicate lower λ tip rates (Inset: a histogram of the mean λ tip rates attained across all grid cells). (c) Global patterns of mean λ tip rates per ecological region, as defined by the WWF, derived from the BAMM software; warm colours indicate higher numbers of mean λ tip rates per ecological region whereas cold colours indicate lower mean λ tip rates. The highlighted geographical areas indicate the five ecoregions with the highest mean λ tip rates (Inset: Maximum and minimum λ tip rate values for ecoregions containing 100 or more geographical distribution records for which λ tip rates were linked).

Results

A new phylogenomic framework for the orchid family

This study combines high-throughput data with Sanger sequencing data to achieve denser taxon sampling at relatively low cost. Our MSC and ML analyses yielded a well-supported phylogenomic framework for the Orchidaceae, containing 38% of the currently accepted orchid genera (Figs 3, S1–S8; see Notes S1). The 10 posterior probability (PP) species trees showed no major conflicts with the MSC tree produced in ASTRAL, with the sole exception of the position of *Corymborkis*, which in some instances was recovered as nested within tribe Gastrodieae (Figs S7, S8). Please refer to the Discussion section below for a detailed comparison of our topology with earlier phylogenomic studies.

Our results reveal the placements of the monospecific genus *Cooktownia* D.L.Jones (Orchidoideae) and the small genus *Claderia* (two species; Epidendroideae), distributed, respectively, from the western Indian Ocean to the western Pacific and from the south-western Pacific to New Zealand. Previous classifications had placed *Cooktownia robertsii* in the tribe Orchideae, based on stigmatic chamber characters (Jones, 1997). Our findings now show that *Cooktownia* is most closely related to *Habenaria* Willd., an almost cosmopolitan genus from subtribe Orchidinae. *Claderia viridiflora* had been included in the predominantly African subtribe Eulophiinae based on morphological similarities in the gynostemium and unpublished nrITS sequences (Cribb & Pridgeon, 2009). Our data place *Claderia* as sister to *Agrostophyllum* Blume and *Earina* Lindl. (Agrostophyllinae, Epidendreae), based on sequences from the type specimen of *Claderia viridiflora* collected in 1867. Though the Angiosperms353 gene recovery was predictably low for this old specimen, we successfully identified 40 genes, with 23 being informative (Table S1). The placement as sister to *Agrostophyllum* was also found by Niissalo *et al.* (2023) based on plastid genomes and the nrITS loci.

Molecular clock dating and biogeographic reconstructions point to a Laurasian origin of Orchidaceae during the Late Cretaceous

Our molecular clock models (Materials and Methods section) estimated the stem age of Orchidaceae to 120 ± 6 Ma (Fig. S7) and the crown age as 83 ± 10 Ma, corresponding to the Early and Late Cretaceous, respectively. The results also indicate that most orchid species originated over the past 5 million years (Fig. S9). Comparison of stem ages and tip branch lengths derived from the Angiosperm353 chronograms and crown node ages of the Sanger chronograms indicated that age discordance between both datasets was minimal (Methods S1; Figs S10, S11), thus suggesting that absolute ages obtained from the 10 posterior probability trees were reliable. Seven of the 10 biogeographical reconstructions (conducted on the 10 PP species trees) supported Laurasia (Nearctic + Palearctic) as the place of origin for the most recent common ancestor of orchids (relative probability = 0.15–0.30; Table S5; Fig. 4a). The three others supported Laurasia + Neotropics or Gondwana (Neotropics + Australasia + Antarctica) as the most likely place of origin for the most recent common ancestor of orchids (relative probability = 0.14–0.19), followed by Laurasia (relative probability = 0.14–0.18; Table S5).

Subsequent colonisation of other biogeographical realms apparently occurred through long-distance dispersals, for example, from Indomalaya to the Afrotropics in tribe Vandeae (Epidendroideae) between 35 and 10 Ma, or through stepping-stone processes from the Palearctic and Nearctic to the Neotropics in tribe Cymbidieae (Epidendroideae, 30–20 Ma). The early diversification of Orchidoideae appears to have taken place 55–40 Ma, with dispersals to Australasia and the Palearctic and subsequent *in situ* Neotropical diversifications (e.g. Cranichidinae), Nearctic (Spiranthinae), Paleartic, Indomalaya (Goodyerinae) and Afrotropics (e.g. Disinae).

Orchid speciation in space and time

We downloaded 795 735 records from the GBIF and RAINBIO repositories of which 495 755 accessions were retained after filtering out duplicate records and inaccurate distributions *sensu* the WCVP database (see Materials and Methods section). Analyses of political country species richness indicated that Ecuador, Colombia, and Papua New Guinea are the top three countries in terms of species richness. Notably, seven out of 10 most orchid species-rich countries are located in the Neotropics. Analyses based on the botanical country species richness (as inferred from the WCVP) yielded similar results (Figs 5a, insert, S12; Table S6). An analysis of species richness per grid cell derived from the curated GBIF-RAINBIO dataset showed that Central America (especially Costa Rica) and the northern Andean region (particularly Ecuador and Colombia) have the highest levels of species richness. These geographical patterns of species richness are in agreement with the species richness distributions independently obtained through the WCVP database and support findings of studies conducted at the family level (Vitt *et al.*, 2023) and in Orchidoideae (Thompson *et al.*, 2023).

Geographical speciation (λ) patterns, as informed from tip λ rates, did not always coincide with the areas of highest current species richness. Notably, when considering minimum and maximum values of speciation rates, virtually all ecoregions assessed here had one- or even two-fold variation in their speciation rates (Fig. 5c, inset). Our speciation rate analyses unveiled multiple accelerations within each subfamily with the fastest tip rates in Orchidoideae and Epidendroideae, starting from the early
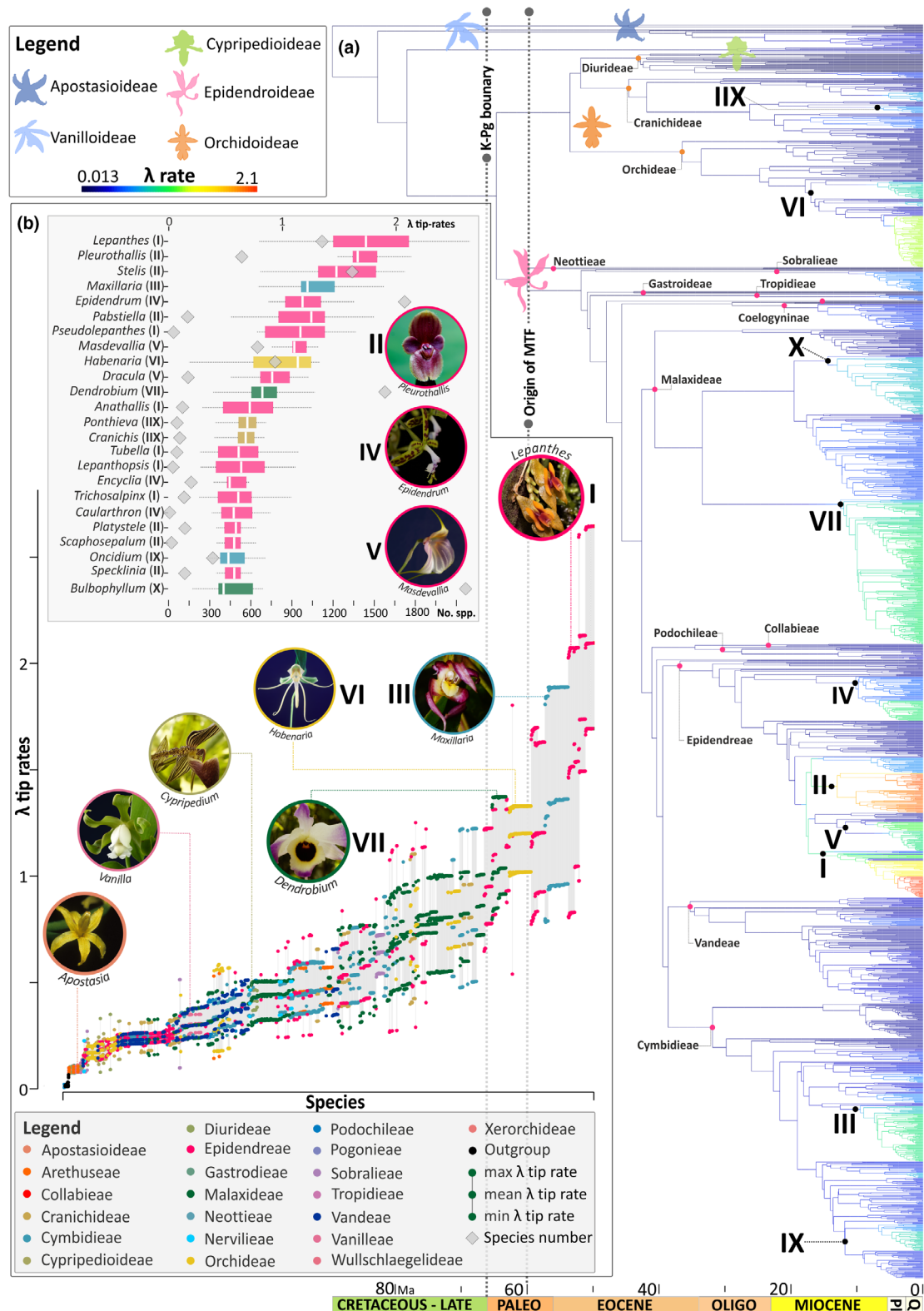
Fig. 6 Speciation dynamics of Orchidaceae. (a) One of 10 posterior probability species trees with modelled speciation (λ) rates along branches (blue colours denote low λ rates, reddish colours indicate high λ rates). Small circles at nodes denote the most recent common ancestors (MRCA) of orchid tribes sampled in this study. Numbers at nodes indicate the lineages with the highest λ tip rates across the entire family. (b) Tip λ rate values (maximum, minimum, mean) derived from the 10 posterior probability (PP) species trees for all terminals sampled. The dots are colour-coded by taxonomic groups (tribes and subfamilies). (Inset: boxplot of λ tip rates summarised from the 10 PP species trees for the genera associated with the highest λ tip rates (their respective totals of accepted species are provided)). Boxplots are colour-coded by tribes (the lower and upper whisker bounds were computed from the largest and smallest values that are within the 1.5 $\times$ interquartile range). Photos: Oscar A. Pérez-Escobar, Diego Bogarín, Kerry Dressler.

Miocene (Fig. 6a). These shifts coincided with the initial diversification of species-rich genera such as *Maxillaria* and *Dendrobium*. In other instances, however, rate increases preceded the divergence of species-rich and depauperate clades as seen with *Lepanthes* and *Lepanthopsis* or occurred within genera like *Bulbophyllum* and *Habenaria*. The lineages with the highest speciation rates are Neotropical epiphytes, including the tribe Epidendreae, and the subtribes Maxillariinae and Oncidiinae. The sole exceptions to this trend are the nearly cosmopolitan terrestrial genus *Habenaria* (Batista *et al.*, 2011) and the genera *Bulbophyllum* and *Dendrobium*, which are mostly epiphytic and distributed mainly in tropical Asia (Fig. 6b, Xiang *et al.*, 2016; Simpson *et al.*, 2022).

The most rapidly speciating lineages predominantly occur in south-eastern Central America where they span diverse habitats, from lowland dry and wet forests through cloud forests to high-elevation grasslands, as exemplified by genera like *Epidendrum*, *Lepanthes*, *Maxillaria*, *Pleurothallis*, and *Stelis* (Fig. 6a).

Discussion

Congruence of our global orchid phylogenomic framework with recent orchid phylogenomic studies

The relationships among subfamilies and tribes found here largely agree with the findings of Zhang *et al.* (2023), a study that used between 639 and 1195 nuclear genes from 610 species representing 297 genera. Notable exceptions include the monophyly of Bletiinae within Epidendreae, and the position of Eriopsidinae within Cymbidieae (fig. S2 in Zhang *et al.*, 2023). A detailed view of the summary trees produced by Zhang *et al.* (2023) reveals evidence of intragenomic conflicts (see fig. S3 in Zhang *et al.*, 2023): 46 bipartitions in the MSC trees attained low local posterior probabilities (< 0.5), and LBPs (< 70%), suggesting that dominant alternative relationships exist for the branches in question. Zhang *et al.* (2023) recovered the subtribe Bletiinae as paraphyletic (i.e. *Bletia* and *Chysis* as successive sisters to Ponerinae, Laeliinae and Pleurothallidinae), a relationship also found, albeit with low support, by van den Berg *et al.* (2005) and Górniak *et al.* (2010). Unfortunately, voucher information and corresponding sequencing data of Zhang *et al.* (2023, first available online on February 2023) have not been made publicly available (at the time of writing), which limits the extent to which these phylogenomic discrepancies can be evaluated (PRJNA923320 accession code, queried on 26 November 2023 in https://www.ncbi.nlm.nih.gov/, returned zero results).

A Laurasian, Cretaceous origin of Orchidaceae

Our dating and biogeographic analyses inferred an origin of orchids in the late Cretaceous, in agreement with previous estimates (Chase, 2001; Ramírez *et al.*, 2007; Chomicki *et al.*, 2015; Givnish *et al.*, 2016; Zhang *et al.*, 2023), and suggested Laurasia as the likely place of initial diversification of the family. This matches the important role that Laurasia played during the late Cretaceous, fostering the initial diversification of many tropical and warm-temperate flowering plant and animal lineages. Notable examples include the palm family Arecaceae (Baker & Couvreur, 2012) and the yams, *Dioscorea* (Dioscoreaceae: Viruel *et al.*, 2015). During the late Cretaceous, Laurasian angiosperms were mostly herbs to small trees with early successional strategies restricted to unstable habitats (Wing & Boucher, 1998; Wing *et al.*, 2012). Subsequently, throughout the Paleogene, much of Laurasia had a monsoon-influenced humid subtropical climate that supported closed canopy, broad-leaved deciduous angiosperm forests that extended into high paleolatitudes beyond the Arctic Circle (Eiserhardt *et al.*, 2017; Korasidis *et al.*, 2022).

Proposing an initial diversification of the orchids in Laurasia contrasts with prior inferences of an Australian origin of the family, followed by entry into the Neotropics via Antarctica (Fig. 4b; Givnish *et al.*, 2016). Those results were due mostly to the outgroup sampling, which used nine species as representatives of Asteliaceae, Blandfordiaceae, Boryaceae, Hypoxidaceae, and Lanariaceae (most of them from Australia), plus one species of Iridaceae. The orchid stem node was, therefore, placed in Australia with a likelihood of *c.* 40%, in Australia + the Neotropics (linked via Antarctica) with *c.* 20%, and in other ranges with together *c.* 40%. Their orchid crown node was placed in Australia with a likelihood of *c.* 20%, in Australia + the Neotropics with *c.* 30%, and in all other ranges with together *c.* 50%. Our broader sampling of orchid groups and biogeographic analyses on multiple trees may lead to less phylogenetic and geographic uncertainty (Rangel *et al.*, 2015).

The subfamily Apostasioideae, sister to all other Orchidaceae and comprising only 16 species in two genera distributed in Japan, India, Nepal, Bhutan, Southeast Asia, New Guinea and northern Australia (Govaerts *et al.*, 2021; Li *et al.*, 2023), is here interpreted as having a relictual range. This would mirror the relictual ranges of numerous other plant groups that were widespread in the Eocene in Eurasia, but which survive only in tropical Southeast Asia today (Manchester *et al.*, 2009; Meseguer & Condamine, 2020). The pattern is known from at least 50 groups of gymnosperms and angiosperms that formerly occurred in Europe and/or North America (Manchester

et al., 2009). Another example in the Orchidaceae of an ancient relictual range is the subfamily Cypripedioideae (slipper orchids), with five genera and 169 extant species (136 Old World and 33 New World). The ancestor of the slipper orchids likely had a continuous distribution in the boreotropics from where slipper orchids migrated southwards to both sides of the Pacific Ocean due to the climate cooling in the late Cenozoic (Guo *et al.*, 2012; Fig. 4).

The Indomalayan region harbours *c.* 30% of the extant orchid species, mostly within the Epidendroideae. The so-far oldest orchid fossil is a hard epidendroid pollinarium found attached to a fungus gnat preserved in Baltic amber, dated to 55–40 Ma (Poinar & Rasmussen, 2017), documenting the presence of Epidendroideae at high latitudes in the early Eocene. During this period, evergreen forests covered northern Europe (Collinson & Hooker, 2003), and epiphytism likely had already evolved in Epidendroideae (Chomicki *et al.*, 2015; Collobert *et al.*, 2022). Our biogeographic models, which did not use the Baltic amber pollinarium as either a fossil or geographic constraint, inferred that Epidendroideae were present in the Palearctic (55–50 Ma), consistent with the fossil evidence. Hard pollinia are prevalent in Epidendroideae and probably appeared early in the history of the subfamily (e.g. in the MRCA of Sobralieae + remainder of Epidendroideae; Dressler, 1990; Mosquera-Mosquera *et al.*, 2019).

A Central American hotspot of orchid speciation

The discrepancy between the orchid's extent species richness and speciation rates agrees with the findings of a broad-scale study on global plant diversification rates and species richness, which are often unlinked (Tietje *et al.*, 2022). The low λ tip rates in Australia can be attributed to the predominance of terrestrial orchids in southern Australia (Ackerman, 2019), which tend to have lower speciation rates than epiphytic orchids (Givnish *et al.*, 2015). Contrary to the findings of Pérez-Escobar *et al.* (2017), the northern Andes do not appear to host the fastest speciating orchid clades in the American tropics despite being one of the most species-rich areas world-wide (Pérez-Escobar *et al.*, 2022; Parra-Sánchez *et al.*, 2023). Instead, southern Mesoamerica, comprising the moist and seasonal forests of Costa Rica and Panama, has the highest orchid speciation rates per grid cell, matching its outstandingly high levels of species richness also of other groups (Myers *et al.*, 2000; Mittermeier *et al.*, 2011; Crain & Fernández, 2020; Fig. 5c; Table S7).

Central American cloud forests and high-elevation grasslands are of Pliocene origin, making them some of the youngest Neotropical biomes (Kirby, 2011). The Late Cenozoic shallow subduction of the young hotspot Cocos ridge under the Panama block (part of the Caribbean plate) is controlling the rapid uplift of the Cordillera de Talamanca in south-eastern Costa Rica and eastern Panama (Morell *et al.*, 2011). This ongoing process started between 5.5 and 3.5 Ma (Grafe *et al.*, 2002), with elevated mountain uplift rates of *c.* 1 km per 1 Ma (Driese *et al.*, 2007). The rapid creation of mountain ranges and valleys under a tropical climate could generate multiple microhabitats that might have enhanced speciation rates. Additionally, the geographic position

of Costa Rica and Panama acts as a biological crossroads between the two biodiversity hotspots of northern Mesoamerica and the northern Andes, resulting in the current coexistence of northern and southern groups (Burger, 1980; Kapelle, 2016). These findings support and expand the hypothesis of Central American tropical forests having evolved rapidly and recently (Cano *et al.*, 2022). Nevertheless, although our analyses of geographical speciation rates offer new perspectives on the evolution of orchid diversity, we urge caution in interpreting the relationship between λ and species richness (Tietje *et al.*, 2022). Additionally, our taxon sampling and species-level spatial distributions are likely biased due to uneven collection efforts in certain regions, such as the historically understudied New Guinea (Camara-Leret *et al.*, 2020).

Conclusions

This study places the initial diversification of orchids in Laurasia during the Late Cretaceous (83 ± 10 Ma) and interprets the modern geographic range of Apostasioideae, comprising 16 species in Southeast Asia and Northern Australasia, as relictual. The presence of Orchidaceae in Eurasia and North America during the Paleogene, followed by extinction and survival in more southern regions resembles the history of at least 50 genera of seed plants that once occurred in Eurasia, but survive with just a few species in tropical Southeast Asia. We reject the hypothesis that Southeast Asia is the region with highest orchid speciation rates (Givnish *et al.*, 2016). Instead, our results show that southern Central America, which contains 4.5% of the world's flora and fauna in just 0.5% of its land surface, has been a hotspot for orchid speciation since the Pliocene with the highest speciation rates per grid cell on a global scale.

Competing interests

None declared.

Author contributions

OAP-E, DB, SSR, AA, and WJB were involved in conceptualisation. OAP-E, FLC, DB, and SB were involved in methodology and formal analysis. OM, LS, and NASP were involved in data curation. OAP-E, NASP, AZ, DB A-QH, BC, KN, MAC, and LL were involved in resources. OAP-E, FLC, and SB were involved in software. NASP, A-QH, KN, LS, ARZ, and LL were involved in investigation. WJB, FF, IJL, AA, KN, MAC, OAP-E, RPB, and RM were involved in funding acquisition. OAP-E, DB, MC, LO, and SB were involved in visualisation. OAP-E and SSR were involved in writing – original draft. OAP-E, SSR, CJ, GC, AA, MC, JDA, RLT, FLC, DB, NASP, WJB, with further contributions from AVM, BG, BSVT, CP, CvdB, EH, EP-S, ES, GAS, JAB, JAC, JMH, MF, MFF, MR, MWC, NSF, PC, RN, and RS-G were involved in writing – review and editing. OAP-E, DB, and NASP are lead authors on this work. OAP-E, MWC, MFF, FLC, FF, KN, SSR, WJB, and AA are senior authors on this work.

ORCID

Alexandre Antonelli https://orcid.org/0000-0003-1842-9297
William J. Baker https://orcid.org/0000-0001-6727-1831
Sidonie Bellot https://orcid.org/0000-0001-6355-237X
Diego Bogarín https://orcid.org/0000-0002-8408-8841
Martha Charitonidou https://orcid.org/0000-0002-9657-8362
Guillaume Chomicki https://orcid.org/0000-0003-4547-6195
Fabien L. Condamine https://orcid.org/0000-0003-1673-9910
Nicola S. Flanagan https://orcid.org/0000-0002-4909-8710
Barbara Gravendeel https://orcid.org/0000-0002-6508-0895
Carlos Jaramillo https://orcid.org/0000-0002-2616-5079
Ilia J. Leitch https://orcid.org/0000-0002-3837-8186
Robert Müntz https://orcid.org/0000-0001-9635-2884
Olivier Maurin https://orcid.org/0000-0002-4151-6164
Katharina Nargar https://orcid.org/0000-0002-0459-5991
Oscar A. Pérez-Escobar https://orcid.org/0000-0001-9166-2410
Edicson Parra-Sánchez https://orcid.org/0000-0003-2670-3882
Charlotte Phillips https://orcid.org/0000-0002-2200-9677
Natalia A. S. Przelomska https://orcid.org/0000-0001-9207-4565
Susanne S. Renner https://orcid.org/0000-0003-3704-0703
Eric Smidt https://orcid.org/0000-0002-1177-1682
Alejandro Zuluaga https://orcid.org/0000-0002-5874-6353

Data availability

References

Ackerman J. 2019. Orchids and the persistent instability principle. In: Pridgeon AM, Arosemena AR, eds. *Proceedings of the 22th World Orchid Conference, vol. 1*. Guayaquil, Ecuador: Asociación Ecuatoriana de Orquideología.

Ackerman J, Phillips RD, Tremblay RL, Karremans A, Reiter N, Peter CI, Bogarín D, Pérez-Escobar OA, Liu H. 2023. Beyond the various contrivances by which orchids are pollinated: global patterns in orchid pollination biology. *Botanical Journal of the Linnean Society* 202: 295–324.

Ali JR, Heaney LR. 2021. Wallace's line, Wallacea, and associated divides and areas: history of a tortuous tangle of ideas and labels. *Biological Reviews* 96: 922–942.

Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, Botigué LR, Brewer G, Carruthers T, Clarkson JJ *et al.* 2022. A comprehensive phylogenetic platform for exploring the angiosperm tree of life. *Systematic Biology* 71: 301–319.

Baker WJ, Couvreur TLP. 2012. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages. *Journal of Biogeography* 40: 286–298.

Balbuena JA, Miguez-Lozano R, Blasco-Costa I. 2013. PACo: a novel Procustres application to cophylogenetic analysis. *PLoS ONE* 8: e61408.

Batista J, de Bem BL, Gonzalez-Tamayo R, Figueroa XM, Cribb P. 2011. A synopsis of the New World *Habenaria* (Orchidaceae) I. *Harvard Papers in Botany* 16: 1–47.

Beeravolu R, Condamine F. 2016. An extended maximum likelihood inference of geographic range evolution by dispersal, local extinction and cladogenesis. *BioRxiv.* doi: 10.1101/038695.

Benzing DH. 2000. *Bromeliaceae: profile of an adaptive radiation*. Cambridge, UK: Cambridge University Press.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushina A, Heled J, Jones G, Kuhnert D, De Maio N *et al.* 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15: e1006650.

Bouetard A, Lefeuvre P, Gigant R, Séverine Bory S, Pignal M, Besse P, Grisoni M. 2010. Evidence of transoceanic dispersal of the genus *Vanilla* based on plastid DNA phylogenetic analysis. *Molecular Phylogenetics and Evolution* 55: 621–630.

Brummitt K. 2001. *World geographical scheme for recording plant distributions, 2nd edn.* Pittsburgh, PA, USA: Hunt Institute for Botanical Documentation, Carnegie Mellon University.

Burgener L, Hyland E, Reich RJ, Scotese C. 2023. Cretaceous climates: mapping paleo-Köppen climatic zones using a Bayesian statistical analysis of lithologic,

paleontologic and geochemical proxies. *Palaeogeography, Palaeoclimatology, Palaeoecology* **613**: 111373.

Burger WC. 1980. Why are there so many kinds of flowering plants in Costa Rica? *Brenesia* **17**: 371–388.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Camara-Leret R, Frodin DG, Adema F, Anderson C, Appelhans M, George A, Guerrero SA, Ashton P, Baker WJ, Barfod AS *et al.* 2020. New Guinea has the world's richest Island flora. *Nature* **584**: 579–583.

Cano A, Stauffer FW, Andermann T, Liberal IM, Zizka A, Bacon CD, Lorenzi H, Christe C, Töpel M, Perret M *et al.* 2022. Recent and local diversification of Central American understorey palms. *Global Ecology and Biogeography* **31**: 1513–1525.

Chase M. 2001. The origin and biogeography of Orchidaceae. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN, eds. *Genera Orchidacearum: vol. 2. Orchidoideae (part one)*. Oxford, UK: Oxford University Press.

Chase MW, Cameron KM, Freudenstein JV, Pridgeon AM, Salazar G, van den Berg C, Schuiteman A. 2015. An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society* **177**: 151–174.

Chomicki G, Bidel LPR, Ming F, Coiro M, Zhang X, Wang Y, Jay-Allemand C, Renner SS. 2015. The velamen protects photosynthetic orchid roots against UV-B damage, and a large dated phylogeny implies multiple gains and losses of this function during the Cenozoic. *New Phytologist* **205**: 1330–1341.

Christenhusz MJ, Byng JW. 2016. The number of known plant species in the world and its annual increase. *Phytotaxa* **261**: 201–217.

Collinson ME, Hooker JJ. 2003. Paleogene vegetation of Eurasia: framework for mammalian faunas. *Deinsea* **10**: 41–83.

Collobert G, Perez-Lamarque B, Dubuisson J-Y, Martos F. 2022. Gains and losses of the epiphytic lifestyle in epidendroid orchids: review and new analyses with succulent traits. *BioRxiv.* doi: 10.1101/2022.09.30.510324.

Condamine F, Rolland J, Morlon H. 2013. Macroevolutionary perspectives to environmental change. *Ecology Letters* **16**: 72–85.

Conran JG, Bannister JM, Lee DE. 2009. Earliest orchid macrofossils: early Miocene *Dendrobium* and *Earina* (Orchidaceae: Epidendroideae) from New Zealand. *American Journal of Botany* **96**: 466–474.

Couvreur TLP, Forest F, Baker WJ. 2011. Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biology* **9**: 44.

Crain BJ, Fernández M. 2020. Biogeographical analyses to facilitate targeted conservation of orchid diversity in Costa Rica. *Diversity and Distributions* **26**: 853–866.

Cribb P, Pridgeon A. 2009. *Claderia*: phylogenetics. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN, eds. *Genera Orchidacearum: vol. 5. Epidendroideae (part two)*. Oxford, UK: Oxford University Press.

Dauby G, Zaiss R, Blach-Overgaard A, Catarino L, Damen T, Deblauwe V, Dessin S, Dransfield J, Droissart V, Duarte MC *et al.* 2016. RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**: 1–18.

De Lamotte DF, Fourdan B, Leleu S, Francois L, Clarens P. 2015. Style of rifting and the stages of Pangea break-up. *Tectonics* **34**: 1009–1029.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.

Dressler RL. 1990. *The orchids: natural history and classification*. Cambridge, UK: Harvard University Press.

Driese GS, Kenneth HO, Sally PH, Zheng-Hua L, Debra SJ. 2007. Paleosol evidence for Quaternary uplift and for climate and ecosystem changes in the Cordillera de Talamanca, Costa Rica. *Palaeogeography, Palaeoclimatology, Palaeoecology* **248**: 1–23.

Eiserhardt WL, Couvreur TLP, Baker WJ. 2017. Plant phylogeny as a window on the evolution of hyperdiversity in the tropical rainforest biome. *New Phytologist* **214**: 1408–1422.

Freudenstein JV, Chase MW. 2015. Phylogenetic relationships in Epidendroideae (Orchidaceae), one of the great flowering plant radiations: progressive specialization and diversification. *Annals of Botany* **115**: 665–681.

Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Doucette A, Giraldo G, McDaniel J, Clements MA *et al.* 2016. Orchid historical

biogeography, diversification, Antarctica and the paradox of orchid dispersal. *Journal of Biogeography* **43**: 1905–1916.

Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJD, Clements MA, Arroyo MTK, Leebens-Mack J *et al.* 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences* **282**: 20151553.

Górniak M, Paun O, Chase MW. 2010. Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, *Xdh*: congruence with organellar and nuclear ribosomal DNA results. *Molecular Phylogenetics and Evolution* **56**: 784–795.

Govaerts R, Lughadha EN, Black N, Turner R, Paton A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data* **8**: 215.

Grace OM, Pérez-Escobar OA, Lucas EJ, Vorontsova MS, Lewis GP, Walker BE, Lohmann LG, Knapp S, Wilkie P, Sarkinen T *et al.* 2021. Botanical monograph in the Anthropocene. *Trends in Plant Science* **26**: 433–441.

Grafe KW, Frisch IM, Villa MM. 2002. Geodynamic evolution of southern Costa Rica related to low-angle subduction of the Cocos Ridge: constraints from thermochronology. *Tectonophysics* **348**: 187–204.

Guo Y-Y, Luo Y-B, Liu Z-J, Wang X-Q. 2012. Evolution and biogeography of the slipper orchids: eocene vicariance of the conduplicate genera in the Old and New World Tropics. *PLoS ONE* **7**: e38788.

Gustafsson ALS, Verola CF, Antonelli A. 2010. Reassessing the temporal evolution of orchids with new fossils and a Bayesian relaxed clock, with implications for the diversification of the rare South American genus *Hoffmannseggella* (Orchidaceae: Epidendroideae). *BMC Evolutionary Biology* **10**: 1–13.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJ, Wicket NJ. 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4**: 1600016.

Johnson MG, Pokorny LP, Dodsworth SD, Botigué LR, Cowan RS, Devault A, Eiserhardt WL, Epitawalage N, Forest F, Kim JT *et al.* 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* **68**: 594–606.

Jones DL. 1997. *Cooktownia robertsii*, a remarkable new genus and species of Orchidaceae from Australia. *Austrobaileya* **5**: 71–78.

Kapelle M. 2016. The montane cloud forests of the Cordillera de Talamanca. In: Kapelle M, ed. *Costa Rican ecosystems*. Chicago, IL, USA: The University of Chicago Press.

Karremans A, Watteyn C, Scaccabarozzi D, Pérez-Escobar OA, Bogarín D. 2023. Evolution of seed dispersal modes in the Orchidaceae: has the *Vanilla* mystery been solved? *Horticulturae* **9**: 1270.

Katoh K, Standley DM. 2013. MAFFT: multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kirby SH. 2011. Active mountain building and the distribution of "core" Maxillariinae species in tropical Mexico and Central America. *Lankesteriana* **11**: 275–291.

Korasidis VA, Wing SL, Shields CA, Kiehl JT. 2022. Global changes in terrestrial vegetation and continental climate during the Paleocene-Eocene Thermal Maximum. *Paleoceanography and Paleoclimatology* **37**: e2021PA004325.

Li Y, Ma L, Liu D-K, Zhao X-WZD, Ke S, Chen G-Z, Zheng Q, Liu Z-J, Lan S. 2023. *Apostasia fujianica* (Apostasioideae, Orchidaceae), a new Chinese species: evidence from morphological, genome size and molecular analyses. *Phytotaxa* **583**: 277–284.

Louca S, Pennell MW. 2021. Why extinction estimates from extant phylogenies are so often zero. *Current Biology* **31**: 3168–3173.

Magallón S, Sánchez-Reyes LL, Gómez-Acevedo SL. 2019. Thirty clues to the exceptional diversification of flowering plants. *Annals of Botany* **123**: 491–503.

Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Alban J, Chilquillo E, Ronsted N, Antonelli A. 2015. Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography* **24**: 973–984.

Manchester SR, Chen Z-D, Lu A-M, Uemura K. 2009. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the northern hemisphere. *Journal of Systematics and Evolution* 47: 1–42.

Matzke NJ. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossil allow improved accuracy and model-testing. *Frontiers of Biogeography* 5: 243–248.

Meseguer AS, Condamine FL. 2020. Ancient tropical extinctions at high latitudes contributed to the latitudinal diversity gradient. *Evolution* 74: 1966–1987.

Mittermeier RA, Turner WR, Larsen FW, Boorks TM, Gascon C. 2011. Global biodiversity conservation: the critical role of hotspots. In: Zachos FE, Habel JC, eds. *Biodiversity hotspots: distribution and protection of conservation priority areas.* Heidelberg, Germany: Springer.

Morell KD, Fisher DM, Gardner TW, La Femina P, Davidson D, Teletzke A. 2011. Quaternary outer fore-arc deformation and uplift inboard of the Panama Triple Junction, Burica Peninsula. *Journal of Geophysical Research* 116: B05402.

Mosquera-Mosquera HR, Valencia-Barrera RM, Acedo C. 2019. Variation and evolutionary transformation of some characters of the pollinarium and pistil in Epidendroideae (Orchidaceae). *Plant Systematics and Evolution* 305: 353–374.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Nauheimer L, Schley RJ, Clements MA, Micheneau C, Nargar K. 2018. Australian orchid biogeography at continental scale: molecular phylogenetic insights from the sun orchids (*Thelymitra*, Orchidaceae). *Molecular Phylogenetics and Evolution* 127: 304–319.

Niissalo MA, Leong PKF, Tay FEL, Choo LM, Kurzweil H, Khew GS. 2023. A new species of *Claderia* (Orchidaceae). *Gardens' Bulletin Singapore* 75: 21–41.

Parra-Sánchez E, Pérez-Escobar OA, Edwards DP. 2023. Neutral-based processes overrule niche-based processes in shaping tropical montane orchid communities across spatial scales. *Journal of Ecology* 111: 1614–1628.

Pérez-Escobar OA, Balbuena JA, Gottschling M. 2016. Rumbling orchids: how to assess divergent evolution between chloroplast endosymbionts and the nuclear host. *Systematic Biology* 65: 51–65.

Pérez-Escobar OA, Bellot S, Przelomska NAS, Flowers JM, Nesbitt M, Ryan P, Gutaker RM, Gros-Balthazard M, Wells T, Kuhnhäuser BG et al. 2021a. Molecular clocks and archaeogenomics of a late period Egyptian date palm leaf reveal introgression from wild relatives and add timestamps on the domestication. *Molecular Biology and Evolution* 38: 4475–4492.

Pérez-Escobar OA, Chomicki G, Condamine FL, Karremans AP, Bogarín D, Matzke NJ, Silvestro D, Antonelli A. 2017. Recent origin and rapid speciation of Neotropical orchids in the world's richest plant biodiversity hotspot. *New Phytologist* 215: 891–905.

Pérez-Escobar OA, Dodsworth S, Bogarín D, Balbuena JA, Schley RJ, Kikuchi IZ, Morris SK, Epitawalage N, Cowan R, Maurin O et al. 2021b. Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *American Journal of Botany* 108: 1166–1180.

Pérez-Escobar OA, Zizka A, Bermúdez MA, Meseguer AS, Condamine FL, Hoorn C, Hooghiemstra H, Pu Y, Bogarín D, Boschman LM et al. 2022. The Andes through time: evolution and distribution of Andean floras. *Trends in Plant Science* 27: 1–12.

Poinar G, Rasmussen FN. 2017. Orchids from the past, with a new species in Baltic amber. *Botanical Journal of the Linnean Society* 183: 327–333.

Poinar G Jr. 2016a. Orchid pollinaria (Orchidaceae) attached to stingless bees (Hymenoptera: Apidae) in Dominican amber. *Neues Jahrbuch für Geologie Und Paläontologie – Abhandlungen* 279: 287–293.

Poinar G Jr. 2016b. Beetles with orchid pollinaria in Dominican and Mexican amber. *American Entomologist* 62: 172–177.

Portik DM, Wiens JJ. 2020. SuperCRUNCH: a bioinformatics toolkit for creating and manipulating supermatrices and other large phylogenetic datasets. *Methods in Ecology and Evolution* 11: 7763–7772.

Rabosky DL, Grundler M, Anderson C, Title P, Shi JF, Brown JW, Huang H, Larson JG. 2014. BAMMTools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5: 701–707.

Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, Alfaro ME. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* 4: 1958.

Ramírez SR, Gravendeel B, Singer RB, Marshall CR, Pierce NE. 2007. Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* 448: 1042–1045.

Rangel TF, Colwell RK, Graves GR, Fucikova K, Rahbek C, Diniz-Filho JF. 2015. Phylogenetic uncertainty revisited: implications for ecological analyses. *Evolution* 69: 1301–1312.

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.

Selosse M-A, Petrolli R, Mujica MI, Laurent L, Perez-Lamarque B, Figura T, Bourceret A, Jacquemyn H, Li T, Gao J et al. 2022. The waiting room hypothesis revisited by orchids: were orchid mycorrhizal fungi recruited among root endophytes? *Annals of Botany* 129: 259–270.

Serna-Sánchez M, Pérez-Escobar OA, Bogarín D, Torres-Jimenez MF, Alvarez-Yela AC, Arcila-Galvis JE, Hall C, de Barros D, Pinheiro F, Dodsworth S et al. 2021. Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Scientific Reports* 11: 6858.

Simpson L, Clements MA, Orel HK, Crayn DM, Nargar K. 2022. Plastid phylogenomics clarifies broad-level relationships in *Bulbophyllum* (Orchidaceae) and provides insights into range evolution of Australasian section *Adelopetalum. BioRxiv.* doi: 10.1101/2022.07.24.500920.

Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS ONE* 13: e0197433.

Smith SA, Moore MJ, Brown JW, Ya Y. 2015. Analysis of phylogenomic datasets reveal conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.

Stamatakis A. 2014. RAxML v.8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Thompson JB, Davis KE, Dodd HO, Priest NK. 2023. Speciation across the Earth driven by global cooling in terrestrial orchids. *Proceedings of the National Academy of Sciences, USA* 120: e2102408120.

Tietje M, Antonelli A, Baker WJ, Govaerts R, Smith SA, Eiserhardt WL. 2022. Global variation in diversification rate and species richness are unlinked in plants. *Proceedings of the National Academy of Sciences, USA* 119: e2120662119.

Title PO, Rabosky DL. 2019. Tip rates, phylogenies and diversification: what are we estimating, and how good are the estimates? *Methods in Ecology and Evolution* 10: 821–834.

Töpel M, Zizka A, Maria Fernanda Calió MF, Scharn R, Silvestro D, Antonelli A. 2017. SpeciesGeoCoder: fast categorization of species occurrences for analyses of biodiversity, biogeography, ecology, and evolution. *Systematic Biology* 66: 145–151.

Van den Berg C, Goldman DH, Freudenstein JV, Pridgeon AM, Cameron KM, Chase MW. 2005. AN overview of the phylogenetic relationships within Epidendroideae inferred from multiple DNA regions and re-circumscription of Epidendreae and Arethuseae (Orchidaceae). *American Journal of Botany* 92: 613–624.

Velasco JA, Pinto-Ledezma JN. 2022. Mapping species diversification metrics in macroecology: prospects and challenges. *Frontiers in Ecology and Evolution* 10: 1–18.

Viruel J, Segarra-Moragues JG, Raz L, Forest F, Wilkin P, Sanmartín I, Catalán P. 2015. Late Cretaceous – Early Eocene origin of yams (Dioscorea, Dioscoreaceae) in the Laurasian Palaeartic and their subsequent Oligocene-Miocene diversification. *Journal of Biogeography* 43: 672–750.

Vitt P, Taylor A, Rakosy D, Kreft H, Meyer A, Wigelt P, Knight TM. 2023. Global conservation prioritization of the Orchidaceae. *Scientific Reports* 13: 6718.

Walker B. 2022. *Kewr: R package to access kew data APIs.* [WWW document] URL https://barnabywalker.github.io/kewr/, https://github.com/barnabywalker/kewr/ [accessed 1 November 2023].

Wing SL, Boucher LD. 1998. Ecological aspects of the Cretaceous flowering plant radiation. *Annual Reviews of Earth and Planetary Science* 26: 379–421.

Wing SL, Strömberg C, Hickey LJ, Tiver F, Willis B, Burnham RJ, Behrensmeyer AK. 2012. Floral and environmental gradients on a Late Cretaceous landscape. *Ecological Monographs* 82: 23–457.

Xiang X-G, Mi X-C, Zhou H-L, Li J-W, Chung S-W, Li D-Z, Huang W-C, Jin W-T, Li Z-Y, Huang L-Q *et al.* **2016.** Biogeographical diversification of mainland Asian *Dendrobium* (Orchidaceae) and its implications for the historical dynamics of evergreen broad-leaved forest. *Journal of Biogeography* **43:** 1310–1323.

Zhang C, Rabiee M, Sayyari E, Mirarab S. **2018.** ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19:** 153.

Zhang C, Zhao Y, Braun EL, Mirarab S. **2021.** TAPER: pinpointing errors in multiple sequence alignments despite varying rates of evolution. *Methods in Ecology and Evolution* **20:** 1–14.

Zhang G, Hu Y, Huang M-Z, Huang W-C, Liu D-K, Zhang D, Hu H, Downing JL, Liu Z-J, Ma H. **2023.** Comprehensive phylogenetic analyses of Orchidaceae using nuclear genes and evolutionary insights into epiphytism. *Journal of Integrative Plant Biology* **65:** 1204–1225.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 A detailed view of phylogenetic relationships between the Apostasioideae, Vanilloideae, and Cypripedioideae subfamilies.

Fig. S2 A detailed view of phylogenetic relationships within the Orchidoideae subfamily.

Fig. S3 A detailed view of phylogenetic relationships between early diverging Epidendroideae lineages.

Fig. S4 A detailed view of phylogenetic relationships within the Cymbidieae.

Fig. S5 A detailed view of phylogenetic relationships within the Vandeae.

Fig. S6 A detailed view of phylogenetic relationships within the Epidendreae.

Fig. S7 A genus-level chronogram of the Orchidaceae produced from 25 low-copy nuclear, clock-like genes in 339 samples.

Fig. S8 Species-level Maximum Clade Credibility ultrametric tree of the orchid family, derived from the grafting of 500 MCMC trees produced from a *matK*-ITS supermatrix (1940 samples) onto 500 MCMC trees produced from 25 clock-like genes (339 samples).

Fig. S9 The age of modern orchid diversity (genera and species) as inferred from branch lengths obtained from the 10 PP species trees.

Fig. S10 Genera crown and stem node age congruence in the Sanger species-level and NGS genus-level consensus phylogenies (as shown in Figs S7, S8), and the frequency of stem and crown node ages obtained, respectively, from the 10 randomly sampled NGS genus-level and Sanger species-level posterior probability trees.

Fig. S11 Dot plot and linear regression Pearson tests of terminal lengths (Mega annum (Ma)) and MRCA ages (Ma) from Maximum Clade Credibility (MCC) consensus trees inferred from 10 PP species trees and 500 PP species trees (same tree presented on Fig. S8).

Fig. S12 Global patterns of species richness per botanical country calculated from the World Checklist of Vascular Plants database.

Methods S1 Extended methods on high-throughput sequencing data, gene tree incongruence, molecular clock dating and species-level phylogeny assembly analyses.

Notes S1 Extended results and discussion on phylogenetic relationships of the Orchidaceae and stem and crown node absolute time discordance.

Table S1 Voucher information, SRA accession nos., taxonomic rank, and proportion of informative and missing Angiosperms353 sequences of the plant material included in the construction of the NGS phylogenomic backbone.

Table S2 GenBank accession nos. of the samples mined from GenBank and retained for subsequent analysis.

Table S3 Root-to-tree variance, tree length, and bipartition support of the Angiosperm353 genes as inferred by SortaDate.

Table S4 The number of informative and missing sequences included in the inference of absolute age estimation analyses for the genus-level orchid backbone.

Table S5 Ancestral areas and their corresponding probabilities of key nodes of the orchid species PP trees, as inferred by dispersal-extinction-cladogenesis.

Table S6 Orchid species richness per botanical country as calculated from the WCVP.

Table S7 Maximum, minimum, and mean tip speciation rates derived from the 10 PP species trees using the software BAMM.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.