Supplemental Information

The Andes through time: Evolution and distribution of Andean floras

Oscar Alejandro Pérez-Escobar[1*¶], Alexander Zizka[2¶], Mauricio A Bermúdez[3], Andrea S. Meseguer[4], Fabien L. Condamine[5], Carina Hoorn[6], Henry Hooghiemstra[6], Yuanshu Pu[2], Diego Bogarín[7,8], Lydian M. Boschman[9], R. Toby Pennington[10,11], Alexandre Antonelli[1,12,13], Guillaume Chomicki[14¶*]

*Correspondence: O.PerezEscobar@kew.org (O.A.P.E.) or g.chomicki@sheffield.ac.uk (G.C.)

¶These authors contributed equally to the study.

[1]Royal Botanic Gardens, Kew, TW9 3AB, Surrey, UK

[2]German Center for Integrative Biodiversity Research (iDiv) Puschstraße 4 D 04317 Leipzig, Germany

[3]Escuela de Ingeniería Geológica, Universidad Pedagógica y Tecnológica de Colombia, Colombia

[4]Real Jardín Botánico de Madrid (RJB-CSIC), Madrid, Spain

[5]CNRS, Institut des Sciences de l'Evolution de Montpellier (Université de Montpellier), Place Eugène Bataillon, 34095 Montpellier, France

[6]Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Science Park 904, 1098XH Amsterdam, the Netherlands

[7]Jardín Botánico Lankester, Universidad de Costa Rica, P.O. Box 302-7050, Cartago, Costa Rica

[8]Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands

[9]Department of Environmental Systems Science, ETH Zurich, Universitätstrasse 16, 8092 Zurich, Switzerland

[10]Department of Geography, University of Exeter, Amory Building, Rennes Drive Exeter, EX4 4RJ, UK

[11]Royal Botanic Garden, Edinburgh, 20a Inverleith Row, Edinburgh EH3 5LR UK

[12]Gothenburg Global Biodiversity Centre, Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden

[13]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

[14]Ecology and Evolutionary Biology, Alfred Denny Building, University of Sheffield, Western Bank, Sheffield S10 2TN

Further supplemental information can be found at

Supplemental dataset S1: Working list of Andean plants. Available at: https://doi.org/10.6084/m9.figshare.16540173

Supplemental dataset S2: Trees used for biogeography analyses. Available at: https://10.5281/zenodo.5336848

Supplemental dataset S3 (online, affiliated with main article): List of Andean plant fossils.

Materials and Methods

Compilation of the working list of Andean plants

We relied on the taxonomic checklist of vascular plant species of the Americas by Ulloa et al. [S1] as a starting point for the compilation of geographical distribution records. The starting species list was taxonomically updated and refined to include species names reported for Andean countries (i.e. Venezuela, Colombia, Ecuador, Peru, Bolivia, Chile and Argentina) with the International Plant Names Index (IPNI) (https://www.ipni.org), TROPICOS (www.tropicos.org) and the World Checklist of Selected Plant Families (https://wcsp.science.kew.org) databases.

Geographical distribution records were downloaded from the GBIF repository. To define altitudinal boundaries of the Andes, we produced contours of elevational values (100 m, 500 m and 1000 m) along the Andes from raster files of relieve for all Andean-Americas countries, using the free software QGIS v. 3.4. Geographical records where subsequently filtered by retaining geographical occurrences falling inside the two elevational contours. The taxonomic validity and orthography of the names reported to grow on the Andes at elevations of 100 and 500 m was checked against the World Checklist of Vascular Plants (https://wcvp.science.kew.org/taxon/306049-1).

Biogeographic meta-analyses

To quantify the number of transitions in and out of the Andes, we first obtain biogeographic regions for the Neotropics from Antonelli et al. [S2] and modified the representation of the Andes following our definition, with the 100 m altitudinal threshold (see main text). We then obtained georeferenced point occurrences for Neotropical plant species [S1], relying on gbif (www.gbif.org) and the Botanical Information and Ecology Network (https://bien.nceas.ucsb.edu/bien/). We cleaned the occurrence records to exclude common geographic errors [S3] using the clean_coordinates_function of the CoordinateCleaner package [S4] and after cleaning retained 329536 records for 89736 species. We then obtained a large-scale plant phylogeny [S5] and matched 14501 species between our occurrence records and the phylogeny. We classified species as Neotropic, if they had at least one occurrence records within any of our Neotropic bioregions (73324 species) and then used the CladeByTrait function of the speciesgeocodeR v2.0-10 package [S6] to obtain 194 clades with between 9-100 tips where at least 85% of the species where Neotropical. We then classified each species as present in a bioregion if at least 5% of the species records were recorded from this region (see ref. 2 for a justification of this threshold) and used an unconstraint DEC model [S7] as implemented in BioGeoBEARS [S8] to estimate ancestral ranges. We then extracted all clades that showed at least one bioregions shift (171 clades) and used the methodology of Antonelli et al. [S2] to count the number of shifts among regions.

Our analysis follows the Antonelli et al. [S2] analysis closely but differs in four details, which cause the Andes to feature more prominent in our Fig. 4A compared to Figure 3 of Antonelli et al. [S2]: 1) The definition of the Andes. Following our general definition of the Andes throughout the manuscript and the 100m height cut-off we use a considerably broader definition of the Andes than Antonelli et al. [S2], as these authors follow the WWF definition relatively strictly. Particularly, our analysis includes parts of what Antonelli et al. [S2] consider "Dry Western South America" and "Amazonia" in the Central Andes, and part of what Antonelli et al. [S2] consider "Amazonia" in the Northern Andes. This change in definition means that more lineages are considered "Andean" in our analyses (compare Fig. 1 in Antonelli et al. [S2], and the inlet map in our Fig. 4). 2) We include areas not at all included by Antonelli et al. [S2] in southern South America. 3) Following this change in definition of the Andes and the overall extension of the study our sampling includes lineages not included in Antonelli et al. [S2] (171 vs 104 clades) and considers more lineages Andean. Hence our sampling is much greater. 4) We split the Andes into three separate regions (Northern, Central, Southern). This aspect is central to our analysis since we want to highlight the interchange among regions within the Andes, and the difference in interchanged of the different parts of the Andes with surrounding ecoregions. For instance, the high interchange of the northern and central Andes with Amazonia compared to the isolation of the Southern Andes from Amazonia. Since the network nodes in our Fig. 4 and Antonelli et al. [S2] Fig. 2 are positioned based on the number of total connections, the Northern and Central Andes feature more prominent in our Figure. However, this does not bias the analyses to give a more prominent role to the Andes, because our methodology is not enriched for Andean species since it includes all Neotropical species (based on the Ulloa Ulloa et al. [S1] which were sampled in the largest plant phylogeny available [S5].

Searching for DNA sequence gaps in Andean plants

One key factor precluding a wider understanding on the species richness and spatio-temporal evolution of Andean floras is the absence of densely sampled DNA sequence dataset. We used our list to identify where are the gaps in the DNA sequence sampling of Andean plants. We searched the NCBI GenBank repository (https://www.ncbi.nlm.nih.gov/genbank/) to determine the fraction of Andean plants with available DNA sequence data of 10 widely used DNA markers in phylogenetic studies such as (ITS, *Xdh*, *atp*B, *mat*K, *nad5*, *psb*A, *rbc*L, *trn*L-F, *rpl*32, NADH).

Building a list of Andean fossils

All Plantae fossil occurrences in Argentina, Chile, Peru, Bolivia, Venezuela, Colombia, and Ecuador were searched in the Paleobiology Database (https://paleobiodb.org/#/), using the web-based portal Fossilworks [S9]. Then all references of the searching results were scanned to confirm age, taxony, locality and other information to form our list of Andean fossils. The list was then manually curated by Dr. Mario Coiro (U. Fribourg). We excluded all records before the Cenozoic as there is substantially less curation on them (Mario Coiro, pers. Comm. To G.C. March 2021).

Supplemental References

S1. Ulloa Ulloa, Carmen, Pedro Acevedo-Rodríguez, Stephan Beck, Manuel J. Belgrano, Rodrigo Bernal, Paul E. Berry, Lois Brako, Marcela Celis, Gerrit Davidse, Rafaela C. Forzza, S. Robbert Gradstein, Omaira Hokche, Blanca León, Susana León-Yánez, Robert E. Magill, David A. Neill, Michael Nee, Peter H. Raven, Heather Stimmel, Mark T. Strong, José L. Villaseñor, James L. Zarucchi, Fernando O. Zuloaga, and Peter M. Jørgensen. 2017. 'An Integrated Assessment of the Vascular Plant Species of the Americas'. *Science (New York, N.Y.)* 358(6370):1614–17. doi: 10.1126/science.aao0398.

S2. Antonelli, A. *et al.* (2018). Amazonia is the primary source of Neotropical biodiversity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6034-6039.

S3. Zizka, Alexander, Fernanda Antunes Carvalho, Alice Calvente, Mabel Rocio Baez-Lizarazo, Andressa Cabral, Jéssica Fernanda Ramos Coelho, Matheus Colli-Silva, Mariana Ramos Fantinati, Moabe F. Fernandes, Thais Ferreira-Araújo, Fernanda Gondim Lambert Moreira, Nathália Michellyda Cunha Santos, Tiago Andrade Borges Santos, Renata Clicia dos Santos-Costa, Filipe C. Serrano, Ana Paula Alves da Silva, Arthur de Souza Soares, Paolla Gabryelle

Cavalcante de Souza, Eduardo Calisto Tomaz, Valéria Fonseca Vale, Tiago Luiz Vieira, and Alexandre Antonelli. 2020. 'No One-Size-Fits-All Solution to Clean GBIF'. *PeerJ* 8:e9916. doi: 10.7717/peerj.9916.

S4. Zizka, Alexander, Daniele Silvestro, Tobias Andermann, Josué Azevedo, Camila Duarte Ritter, Daniel Edler, Harith Farooq, Andrei Herdean, María Ariza, Ruud Scharn, Sten Svantesson, Niklas Wengström, Vera Zizka, and Alexandre Antonelli. 2019. 'CoordinateCleaner: Standardized Cleaning of Occurrence Records from Biological Collection Databases' edited by T. Quental. *Methods Ecol. Evol.*10, 744–51. doi: 10.1111/2041-210X.13152.

S5. Smith, S.A. and Brown, J.W. (2018). Constructing a broadly inclusive seed plant phylogeny. *Amer. J. Bot.* 105, 302-314.

S6. Töpel, Mats, Alexander Zizka, M. F. Maria Fernanda Calió, Ruud Scharn, Daniele Silvestro, and Alexandre Antonelli. 2017. 'SpeciesGeoCoder: Fast Categorization of Species Occurrences for Analyses of Biodiversity, Biogeography, Ecology, and Evolution'. *Systematic Biology* 66(2):145–51. doi: 10.1093/sysbio/syw064.

S7. Ree, Richard H., and Stephen A. Smith. 2008. 'Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis'. *Systematic Biology* 57(1):4–14. doi: 10.1080/10635150701883881.

S8. Matzke, Nicholas J. 2014. 'Model Selection in Historical Biogeography Reveals That Founder-Event Speciation Is a Crucial Process in Island Clades'. *Systematic Biology* 63(6):951–70. doi: 10.1093/sysbio/syu056.

S9. Occurrences of Plantae in Argentina, Chile, Peru, Bolivia, Venezuela, Colombia, and Ecuador recorded in the Paleobiology Database. Fossilworks. http://fossilworks.org. Accessed on January 2020.