The Geography of Speciation Across Andean-Centred Clades

Andrés Felipe Ramos-Torres | Y. Franchesco Molina-Henao

Research Group in Ecogenetic Studies and Molecular Biology. Department of Biology, Universidad del Valle, Cali, Colombia.

Correspondence: Andrés Felipe Ramos-Torres (thbioandres@gmail.com) |

Y. Franchesco Molina-Henao (yherson.molina@correounivalle.edu.co)

Keywords: Andes | mountain biodiversity | speciation geography | range overlap | species distribution models | comparative phylogenetics | approximate Bayesian computation

ABSTRACT

Aim: Mountains harbour a disproportionate share of global biodiversity, yet the geographic pathways through which this diversity arises remain debated. In the Andes, diversification has been linked both to geographic isolation generated by complex topography and climatic dynamics, and to ecological divergence across steep environmental gradients. Here, we assess the predominant geographic modes of speciation across Andean-centred clades using a large-scale comparative framework, while evaluating how alternative representations of species distributions influence macroevolutionary inference.

Location: Andes, South America.

Taxon: A total of 102 Andean-centred genera of animals, plants, and fungi.

Methods: We combined time-calibrated phylogenies with species distribution data derived from two alternative spatial representations: concave polygons and potential distribution models. For each genus, we estimated 13 summary metrics capturing range overlap, range asymmetry, and phylogenetic structure. We then used simulation-trained, likelihood-free classification approaches, including linear discriminant analysis and two approximate Bayesian computation methods, to assign each clade to its most probable geographic mode of speciation: vicariant, peripatric, parapatric, sympatric, or mixed geographic scenarios.

Results: Across clades, overlap-associated scenarios, corresponding to parapatric or sympatric classifications, were inferred as predominant in 57–67% of genera, whereas strictly allopatric scenarios, that is, vicariant or peripatric speciation, accounted for 23–30%, depending on the spatial representation and classification method used. Sympatric scenarios alone were inferred in 38–52% of genera, particularly when potential distribution models were used. At least one-third of clades showed congruent classifications across both analytical methods and spatial representations. However, several genera were sensitive to

how species distributions were represented, with some shifting between allopatric and sympatric classifications depending on the spatial input used.

Main Conclusions: Our results suggest that Andean diversification has not followed a single dominant geographic pathway, but instead reflects a mosaic of spatial contexts involving both isolation-driven and overlap-associated scenarios. However, classifications as parapatric or sympatric should be interpreted as hypotheses of geographic overlap during lineage divergence, rather than as direct evidence of speciation with gene flow. By shifting from single-clade case studies to a broad multi-taxon comparative perspective, this study suggests that the geography of Andean speciation reflects a heterogeneous process shaped by multiple spatial pathways to lineage divergence, while identifying candidate clades for future phylogeographic, ecological, cytogenetic, and population genomic studies of mountain biodiversity.

1 | Introduction

Mountains are among the most biodiverse regions on Earth. Although they occupy less than one-quarter of the terrestrial surface, they harbour nearly one-third of global species diversity and contain many of the world's most prominent centres of endemism (Price and Butt, 2000; Spehn, Rudmann-Maurer and Körner, 2012; Körner et al., 2017). This disproportionate concentration of biodiversity has long motivated a central question in biogeography and evolutionary biology: why do mountains generate and maintain so much species diversity? Mountain systems combine steep environmental gradients, complex topography, climatic heterogeneity, and dynamic geological histories. As a result, they may act simultaneously as biodiversity reservoirs, dispersal corridors, and geographic barriers, creating multiple opportunities for population isolation, ecological divergence, and lineage diversification (Perrigo, Hoorn and Antonelli, 2020).

The evolutionary role of mountains has often been linked to the interaction between landscape complexity and speciation (Rahbek et al., 2019). Geological uplift can generate new habitats, alter climatic regimes, and reshape connectivity among populations, while elevational gradients can promote ecological differentiation over short geographic distances (Muellner-Riehl, 2019). These processes have been associated with rapid diversification in several mountain systems, including the Andes, the Himalayas, and other topographically complex regions (Madriñán, Cortés and Richardson, 2013; Hughes and Atchison, 2015; Esquerré et al., 2019; Zhao et al., 2021). However, the relative contribution of different geographic pathways to mountain diversification remains unresolved. In particular, it is still unclear whether mountain biodiversity has arisen predominantly through spatial isolation among populations, divergence across environmental gradients in the presence of geographic proximity, or a combination of both.

The Andes provide an exceptional setting for addressing this question. Extending along western South America, the Andes constitute the longest continental mountain chain and one of the most species-rich regions on the planet (Graham, 2009). They are recognized as a global biodiversity hotspot, with exceptionally high levels of species richness and endemism across multiple taxonomic groups (Myers et al., 2000; Pérez-Escobar et al., 2022). Their complex geological history, including substantial uplift during the last 20–30

million years, has generated strong spatial heterogeneity in climate, topography, and habitat availability (Gregory-Wodzicki, 2000; Luebert and Muller, 2015). This environmental complexity is especially evident in high-elevation ecosystems such as the páramo, where rapid diversification has been documented in several plant lineages (Madriñán, Cortés and Richardson, 2013). More broadly, the Andes offer a natural laboratory for examining how geographic isolation, ecological gradients, and landscape dynamics interact to shape the origin of species.

Speciation is commonly classified according to the geographic context in which population divergence occurs and the degree to which gene flow is reduced during lineage splitting (Coyne and Orr, 2004; Butlin, Galindo and Grahame, 2008). Under allopatric speciation, divergence proceeds after geographic barriers strongly restrict or prevent gene flow among populations. Such barriers may arise through vicariance, when formerly connected populations become separated, or through peripatric isolation, when small peripheral populations become geographically isolated from the ancestral range. In contrast, parapatric and sympatric speciation involve divergence under geographic adjacency or overlap, where some degree of gene flow may persist. In these contexts, ecological differentiation, divergent selection, or other isolating mechanisms are expected to play a greater role in counteracting the homogenizing effects of gene flow (Turelli, Barton and Coyne, 2001; Bolnick and Fitzpatrick, 2007). Although these categories represent simplified endpoints along a continuum of spatial and demographic scenarios, they provide a useful comparative framework for evaluating how geography contributes to the buildup of biodiversity.

Two broad, non-mutually exclusive hypotheses have been proposed to explain Andean diversification. The first emphasizes geographic isolation, suggesting that asynchronous uplift, valley formation, riverine barriers, habitat fragmentation, and climatic oscillations have repeatedly subdivided populations, promoting allopatric divergence (Kattan et al., 2004; Ramírez-Barahona and Eguiarte, 2013). The second emphasizes ecological opportunity and environmental heterogeneity, proposing that the rise of the Andes created novel habitats and steep ecological gradients that facilitated adaptation and divergence, potentially under parapatric or sympatric geographic settings (Bates and Zink, 1994; Amico, Vidal-Russell and Nickrent, 2007). These hypotheses are often discussed as alternatives, but the spatial complexity of the Andes suggests that multiple geographic modes of speciation may have operated across clades, regions, and time periods.

Despite extensive research on Andean diversification, most studies have focused on individual clades or particular taxonomic groups. As a result, we still lack a broad comparative assessment of the geographic modes of speciation across the Andean biota. This gap limits our ability to evaluate whether Andean diversification is best explained by a dominant geographic mode or by a mosaic of spatial pathways to speciation. Here, we address this question using a large-scale comparative analysis of 102 Andean-centred genera of animals, plants, and fungi. By integrating time-calibrated phylogenies with species distribution data and simulation-based classification approaches, we infer the predominant geographic mode of speciation across clades. Given the topographic, climatic, and ecological heterogeneity of the Andes, we expect allopatric, parapatric, and sympatric scenarios to be represented, but with their relative prevalence varying among clades. More broadly, our study aims to clarify whether the geography of Andean speciation reflects a single dominant mechanism or a heterogeneous set of spatial pathways underlying mountain biodiversity.

2 | Methods

2.1 | Taxonomic sampling

We compiled a candidate list of Andean-centred genera of animals, plants, and fungi using occurrence records available through the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/). To identify genera represented in the Andes, we queried GBIF occurrence data using the following filters: occurrence status = "present", records with geographic coordinates, continent = "South America", and geographic location within the spatial extent of the Andean mountain system. The Andean spatial extent was defined using the mountain polygon available from the Global Mountain Biodiversity Assessment (GMBA; https://www.gmba.unibe.ch/). The resulting occurrence dataset was downloaded as a CSV file and processed in R. Records were retained only when they corresponded to accepted species-level taxa belonging to animals, plants, or fungi.

Candidate genera were then filtered according to three main criteria: (1) availability of a time-calibrated phylogeny for the genus or, when appropriate, for the Andean clade within the genus; (2) sufficient species richness to allow inference of geographic speciation mode; and (3) availability of georeferenced occurrence records for the species included in the clade. Phylogenetic availability was evaluated using the TimeTree database (http://www.timetree.org/), which synthesizes published molecular timetrees from more than 4,000 studies (Kumar et al., 2022). For each candidate genus, we searched for genus-level synthetic timetrees and downloaded available trees in Newick format. Genera lacking phylogenetic information in TimeTree were excluded from further analyses. Because the calculation of several summary metrics requires resolved sister-lineage relationships, species involved in unresolved polytomies were excluded from the corresponding downstream analyses. The final set of species retained for each genus therefore depended on the overlap between the resolved phylogeny and the cleaned occurrence dataset.

We retained genera with at least 12 species, following the minimum clade size used by Skeels and Cardillo (2019) for inferring geographic modes of speciation from summary statistics. Candidate genera represented biologically relevant Andean-centred clades, defined here as clades in which the majority of species in the available phylogeny occur within the Andes. Specifically, after cleaning occurrence data, we excluded genera containing fewer than 11 Andean species with at least five occurrence records each. We also excluded genera where Andean species represented less than 75% of species in the phylogeny, ensuring our analyses focused on strictly Andean-centred clades rather than those with marginal representation. Finally, we excluded genera with a sampling fraction below 25%; the total number of species per genus was calculated using accepted species with more than two georeferenced occurrences in GBIF.

2.2 | Occurrence data processing

Species occurrence records for each selected genus were downloaded from GBIF using the R package 'rgbif'. Occurrence data were cleaned before spatial analyses to reduce taxonomic, geographic, and sampling errors. We removed duplicate records, records with coordinates equal to zero, and records whose taxonomic status was not "accepted". Records

falling outside terrestrial areas, including marine or oceanic localities, were excluded through spatial inspection against land boundaries using QGIS software (https://qgis.org/). To reduce spatial clustering and minimize the probability of multiple records falling within the same environmental pixel, we applied a spatial thinning procedure that retained occurrence points separated by a minimum distance of 1 km using the *thin_data* function from the 'ellipsenm' package in R.

The cleaned occurrence records were used to estimate species ranges using two alternative spatial representations: concave polygons and potential distribution models. These two approaches were implemented to evaluate how different assumptions about species distributions influence the inference of geographic speciation mode. Concave polygons provide a conservative approximation of occupied geographic space based directly on known occurrences, whereas potential distribution models estimate areas of environmental suitability and may therefore represent broader potential ranges. Together, these approaches capture different balances between omission and commission error in species range estimation.

2.3 | Concave polygons

For each species, we generated concave polygons from cleaned occurrence records using the R package 'concaveman', with the concavity parameter set to 3. This approach provided a spatially constrained approximation of species distributions based on known occurrences, reducing overestimation relative to convex hulls. Each polygon was then rasterized at a spatial resolution of 5 km to match the spatial format required for downstream analyses of range overlap and phylogenetic structure.

2.4 | Environmental predictors

Potential distribution models were calibrated using bioclimatic variables from WorldClim version 2.1 (https://www.worldclim.org/data/worldclim21.html) at 30 arc-second resolution. We initially considered 15 of the 19 standard bioclimatic variables, excluding variables BIO 8, BIO 9, BIO 18 and BIO 19 previously identified as prone to spatial artefacts following Escobar et al. (2014). To reduce dimensionality and collinearity among environmental predictors, we performed a principal component analysis (PCA) using environmental data for the Americas. The PCA was implemented with the *prcomp* function in the R 'stats' package, adapting the workflow used in the Principal Component Analysis function of NicheA. The first three principal components, which together explained 91% of the total environmental variation, were retained as predictors for potential distribution modelling.

2.5 | Potential distribution models

Potential distribution models were calibrated using Maxent (Phillips et al., 2006) through the R package 'kuenm' (Cobos et al., 2019). For each species, occurrence records were randomly partitioned into two equal subsets: 50% of the records were used for model calibration and the remaining 50% were used for model evaluation. Candidate models were calibrated using five combinations of feature classes: linear (L), linear-quadratic (LQ), linear-quadratic-product (LQP), linear-quadratic-product-threshold (LQPT), and

linear-quadratic-product-threshold-hinge (LQPTH) and a range of regularization multipliers from 0.1 to 1 in increments of 0.1 and from 2 to 5 in increments of 1. Model selection was performed in 'kuenm' using statistical significance, omission rate, and Akaike's information criterion corrected for small sample size (AICc). Among statistically significant models with omission rates below the default threshold of 5%, the model with the lowest AICc was selected as the final model. When multiple models showed similar support ($\Delta AICc \leq 2$), the model with the lowest regularization multiplier was retained, and if several candidates shared that value, the one with the simplest feature class was selected. For each species, the accessible area (*M*) was defined as a 200-km buffer around its cleaned occurrence records, following the conceptual framework of Barve et al. (2011). This accessible area was used for both model calibration and spatial projection, thereby restricting predictions to regions considered biogeographically accessible to each species.

Models were calibrated using the first three bioclimatic principal components as environmental predictors. After calibration, continuous suitability outputs were converted into binary presence–absence maps using the minimum training presence logistic threshold provided by Maxent. The resulting binary maps were used as an alternative representation of species distributions for subsequent analyses of range overlap, range asymmetry, and geographic speciation mode. Given that several species were represented by few occurrence records, this modelling strategy was intended to limit overfitting by combining low-dimensional environmental predictors, spatially restricted accessible areas, and a conservative threshold that retained all training occurrences in the final binary predictions. Accordingly, these models were not intended to provide definitive estimates of realized distributions, but rather to generate a standardized alternative representation of potentially suitable areas for comparative analyses.

2.6 | Inference of geographic modes of speciation

We inferred the predominant geographic mode of speciation for each clade by combining species distribution data with time-calibrated phylogenies. For each genus, analyses were performed independently using the two alternative spatial representations: concave polygons and potential distribution models. From each spatial dataset and its corresponding phylogeny, we calculated 13 summary metrics to characterize geographic and phylogenetic patterns associated with alternative modes of speciation. These metrics summarize range overlap, range asymmetry, and the phylogenetic distribution of spatial relationships among species (Table 1).

To classify genera according to their most probable geographic mode of speciation, we used the simulation framework of Skeels and Cardillo (2019). Specifically, we downloaded the simulated summary-statistic dataset made available by those authors in the Dryad Digital Repository (https://dx.doi.org/10.5061/dryad.d9j09bm) and used it to train likelihood-free model classification approaches. Following Skeels and Cardillo (2019), we considered four geographic modes of speciation: vicariant, peripatric, parapatric and sympatric, in addition to a mixed model. In their simulation framework, the mixed model represents clades in which all four modes occur during diversification, implemented by assigning equal speciation-rate parameters (λ) to each mode. We therefore interpreted support for the mixed model as indicating geographically heterogeneous speciation histories within a clade, rather than a single predominant geographic mode.

We implemented three likelihood-free classification methods. First, we used linear discriminant analysis implemented in the R package ‘caret’ (Kuhn, 2008). Second, we used two approximate Bayesian computation approaches implemented in the R package ‘abc’ (Csilléry et al., 2012): multinomial logistic regression and neural networks. For each genus, these methods were applied separately to the two spatial representations to estimate the most probable geographic mode of speciation. Because these approaches differ in their statistical interpretation, we refer to their outputs collectively as classification support. We then compared classifications across methods and spatial representations to assess the robustness of inferred speciation modes to the analytical approach and species-range estimation. Congruence was defined as the recovery of the same geographic mode across all six classifications for a given genus, corresponding to the three classification methods applied to both spatial representations.

Table 1. Summary metrics used to infer geographic modes of speciation. The 13 metrics proposed by Skeels and Cardillo (2019) to summarize patterns of range overlap, range asymmetry, and phylogenetic structure and used to classify clades according to alternative geographic modes of speciation.

Metric	Abbreviation	Description
Age-range correlation	ARC_{slope} $ARC_{intercept}$	Slope and intercept of the regression between phylogenetic node age and geographic range overlap (RO) among descendant lineages (Fitzpatrick and Turelli 2006).
Sister-species RO x divergence time	$RO_{intercept}$	Intercept of the regression between RO among sister species and their divergence times. This metric is analogous to ARC but is calculated only for sister-species pairs.
Mean sister-species RO	RO_{mean}	Mean proportion of the range of the smaller-ranged sister species that overlaps with the range of the larger-ranged sister species.
Sister-species sympatry proportions	RO_{90} RO_{100}	Proportion of sister-species pairs with RO $\geq$ .9 or RO equal to 1.00, respectively.
Range asymmetry x divergence time	$Asym_{slope}$ $Asym_{intercept}$	Slope and intercept of the regression between range-size asymmetry among sister species and their divergence times.
Mean and standard deviation of the standardized range size	RS_{mean} RS_{SD}	Mean and standard deviation of standardized species range size across all terminal taxa in the phylogeny. Standardized range size is calculated as each species' range size divided by the largest species range size in the clade.
Standard deviation of the standardized distance between sister ranges	RD_{SD}	Standard deviation of standardized geographic distances between sister-species ranges. Standardized range distance is calculated as the minimum distance between sister-species ranges divided by the maximum distance between any two species ranges in the clade.
Range isolation x divergence time	$RD_{intercept}$	Intercept of the regression between standardized geographic distance among sister-species ranges and their divergence times.
Sackin's index	SI	Measure of phylogenetic tree imbalance (Blum and François, 2006)

3 | Results

3.1 | Predominant geographic mode of speciation inferred at the genus level

For each genus, we estimated the relative support of five alternative geographic contexts of speciation: vicariant, peripatric, mixed, parapatric, and sympatric scenarios. Full genus-level results are provided in Supporting Information. To visualize how geographic modes were inferred across clades, Figure 1 presents three illustrative genera, *Menonvillea*, *Diglossa* and *Nicotiana,* representing contrasting predominant geographic scenarios. For each genus, classification probabilities are shown under two spatial representations, concave polygons and potential distribution models (PDM), and three classification approaches, linear discriminant analysis (LDA), multinomial logistic regression (mnL), and neural networks (NN).

In *Menonvillea*, the vicariant scenario received the strongest relative support across all analytical combinations (Figure 1a). Under the polygon-based representation, its classification probability reached 0.97 (LDA), 0.81 (mnL), and 0.79 (NN), with the latter two methods assigning moderate support (0.15) to the mixed geographic scenario. The PDM-based representation yielded similar values of 0.99 (LDA), 0.88 (mnL), and 0.76 (NN), with mnL and NN also favouring the mixed scenario at 0.1 and 0.21, respectively. The consistency between concave polygon- and PDM-based representations indicates a robust inference of vicariant speciation for this genus.

In Diglossa, parapatric speciation received the highest classification support as the predominant geographic scenario (Figure 1b). For the polygon-based representation, the parapatric scenario received high support, with classification probabilities of 0.93 (mnL), 0.96 (LDA), and 0.81 (NN). The PDM-based representation likewise supported parapatric speciation, albeit at lower probabilities of 0.84 (mnL), 0.77 (LDA), and 0.64 (NN). Under the PDM representation, the sympatric scenario received secondary support, with classification probabilities of 0.15 (mnL), 0.22 (LDA), and 0.26 (NN), indicating some sensitivity to the choice of distributional representation.

In *Nicotiana*, the sympatric geographic scenario consistently received the highest support across spatial representations and classification methods (Figure 1c). Support was particularly strong under the PDM representation, with classification probabilities of 0.97 (mnL and NN) and 0.92 (LDA). The polygon-based representation likewise supported sympatric speciation, albeit at lower probabilities of 0.86 (mnL), 0.58 (LDA), and 0.83 (NN). Under this representation, the mixed, parapatric, and vicariant scenarios received minor secondary support. This pattern indicates that *Nicotiana* represents a case in which overlap-associated scenarios are consistently favoured, though with greater uncertainty when distributions are represented more conservatively.

Together, these genus-level examples illustrate the range of geographic scenarios recovered across Andean-centred clades and show that some inferences were robust across spatial representations and classification methods, whereas others were more sensitive to the input distribution data. The predominant geographic mode inferred for each of the 102 clades is summarized in the Supporting information (Table S1), with full classification probabilities provided in Table S2.

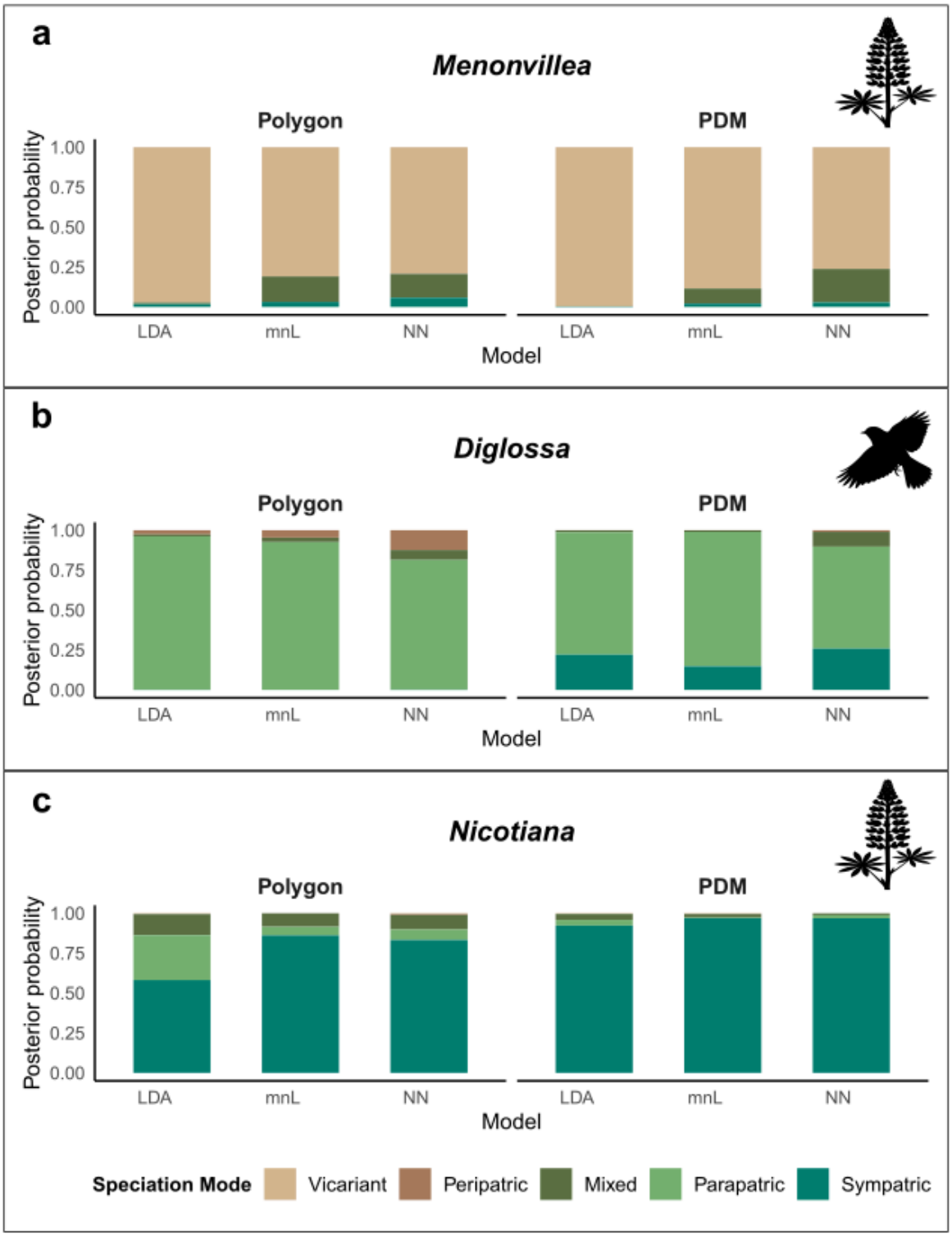


Figure 1. Classification support for alternative geographic modes of speciation inferred for three example clades: (a) *Menonvillea*, (b) *Diglossa*, and (c) *Nicotiana*. For each clade, results are shown under two spatial representations of species distributions, concave polygons and potential distribution models (PDMs), and three classification approaches: linear discriminant analysis (LDA), multinomial logistic regression (mnL), and neural

networks (NN). Bars represent relative classification support for each geographic scenario: vicariant, peripatric, mixed, parapatric and sympatric.

3.2 | Overall patterns of inferred geographic modes

We inferred the predominant geographic mode of speciation for 102 Andean genera using three likelihood-free classification methods and two alternative spatial representations of species distributions: concave polygons and PDMs (Figure 2). Overall, classifications consistent with sympatric geographic scenarios were the most frequent across clades, although their prevalence varied depending on the spatial input and classification method used. When concave polygons were used, sympatric geographic scenarios were inferred as predominant in 38–41% of all genera, depending on the method: 41% (NN), 39% (LDA), and 38% (mnL). When PDMs were used, the proportion of genera classified as predominantly sympatric increased to 50–52% across methods: 52% (NN) and 50% with both LDA and mnL.

Strictly allopatric scenarios, including both vicariant and peripatric speciation, were inferred as predominant in 23–28% of genera when concave polygons were used as spatial input: 23% (mnL), 26% (NN), and 28% (LDA). Under PDM-based representations, allopatric scenarios were inferred in 29%–30% of genera across the three classification methods: 29% with both LDA and mnL and 30% (NN). Parapatric speciation was less frequently inferred than sympatric or allopatric modes. Using concave polygons, parapatric speciation was inferred as predominant in 19–28% of genera, depending on the classification method. This proportion decreased under PDMs, where parapatric speciation obtained the highest classification support in 8–13% of genera. In addition, some genera showed mixed signals, with no single geographic mode clearly favoured by the classification methods. These mixed classifications occurred in 5–21% of genera using concave polygons and in 8–13% of genera using PDMs.

3.3 | Taxonomic variation in inferred speciation modes

Patterns varied among major taxonomic groups. In arthropods, allopatric speciation (vicariant + peripatric) was the most frequently inferred mode across methods and spatial representations, accounting for at least 44% of genera. Under the concave polygon representation, allopatric scenarios were inferred in 67% of genera (NN) and 55% with both LDA and mnL; under the PDM representation, this proportion was 44% (NN) and 55% with both LDA and mnL. In chordates, sympatric speciation was generally the most frequently inferred mode, though this pattern was not consistent across all combinations of method and spatial representation. Under the concave polygon representation, parapatric speciation was inferred as the most frequent mode under both LDA (41%) and NN (30%), whereas sympatric speciation predominated under mnL (33%) and all PDM-based representations: 33% (LDA) and 41% with both mnL and NN.

Among plants, results were summarized separately for angiosperms and non-flowering plants. In angiosperms, the sympatric scenario was consistently inferred as the predominant speciation mode across all methods and spatial representations. Under the concave polygon representation, sympatric speciation showed the highest classification support in 46% of genera (LDA), 44% (mnL), and 48% (NN), followed by parapatric speciation (31%, 17%, and 17%, respectively) and allopatric speciation (20%, 15%, and 19%, respectively). Under

the PDM representation, the proportion assigned to the sympatric scenario increased to 60% (LDA), 54% (mnL), and 58% (NN), with vicariant speciation emerging as the second most frequent mode (23%, 29%, and 27%, respectively).

A comparable pattern was recovered in non-flowering plants, in which the sympatric scenario again predominated but at generally higher proportions. Under the concave polygon representation, sympatric speciation received the highest classification support in 62% of genera (LDA), 38% (mnL), and 54% (NN), whereas under the PDM representation it reached 62% (LDA) and 69% with both mnL and NN. Allopatric speciation was consistently the second most frequent mode under the PDM representation, inferred in 31% of genera (LDA) and 23% under both mnL and NN (in the latter two, only vicariant speciation showed the highest classification support). Overall, both plant groups showed a shift toward higher sympatric support under the PDM representation relative to concave polygons, although the magnitude of this shift varied among methods.

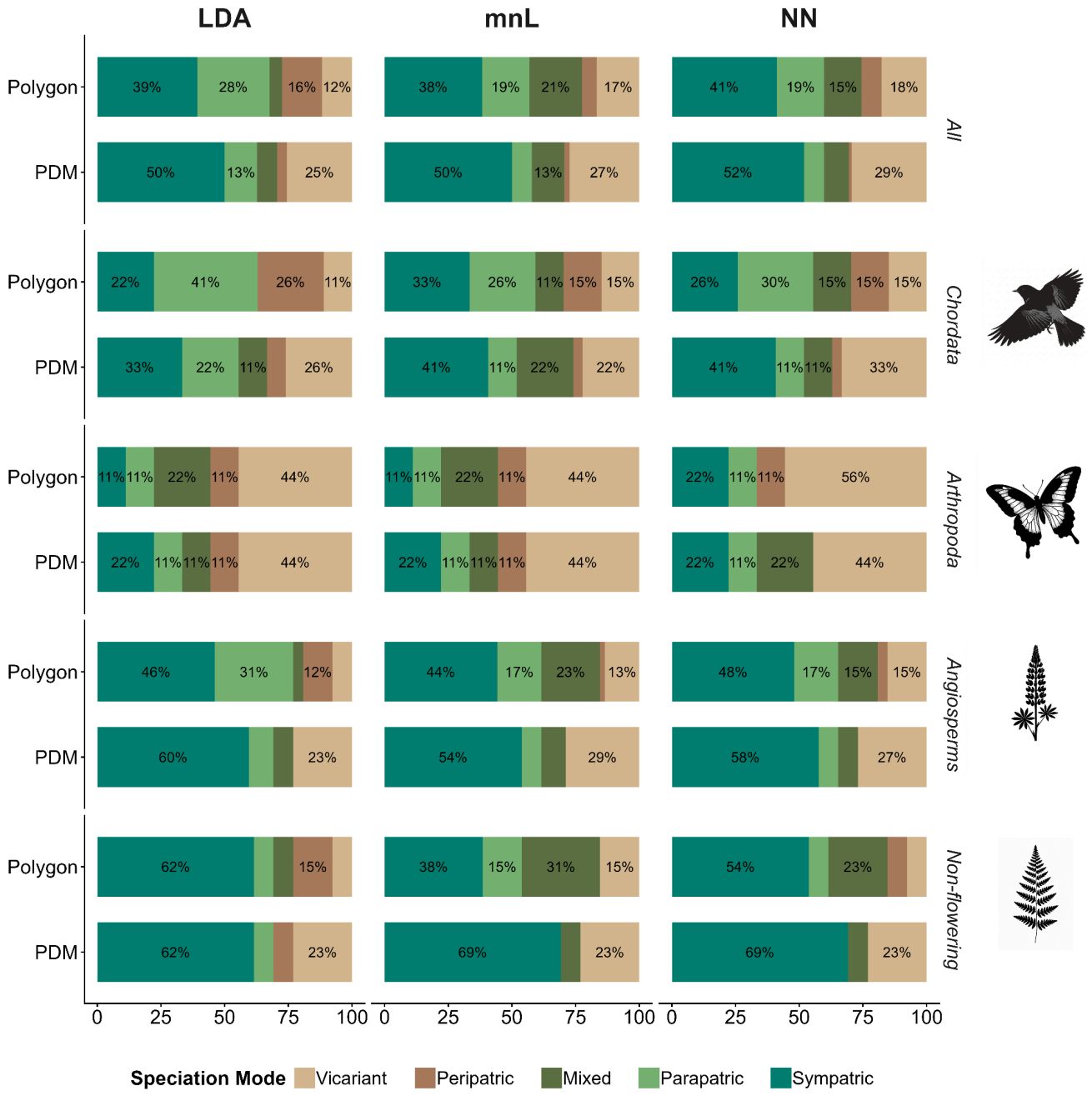


Figure 2. Relative frequency of inferred geographic modes of speciation across major taxonomic groups. Bar plots show the proportion of genera assigned to each predominant geographic mode of speciation for the full dataset and for major taxonomic subsets: vertebrates, insects, angiosperms, and non-flowering plants. Results are shown for the three model classification methods used to estimate classification support for each geographic mode: neural networks, linear discriminant analysis, and multinomial logistic regression. Within each method, bars are grouped by the two spatial representations used to calculate geographic summary metrics: concave polygons and potential distribution models (PDMs). Values below 10% are not displayed.

3.4 | Congruence among methods and spatial representations

Across the full dataset, 37 genera showed congruent classifications across all three model classification methods and both spatial representations, although this number decreases to 14 when considering only clades with sampling fraction >50% (Table 2). Sympatric geographic scenarios were consistently inferred for the amphibian genus *Craugastor*; the mammal genus *Akodon*; the insect genus *Pseudomyrmex*; the dicot genera *Bunchosia, Capsicum, Centropogon, Drymonia, Heliotropium, Inga, Lupinus, Nicotiana, Nolana, Psammisia, Senna, Stachys, Symplocos* and *Viburnum*; the fern genera *Adiantum, Asplenium, Lindsaea* and *Serpocaulon;* and a fungal genus *Sticta*. Parapatric speciation was consistently inferred as predominant in the bird genera *Diglossa*, *Pachyramphus,* and *Synallaxis*; the insect genus *Napeogenes*; and the monocot genera *Aristida* and *Bactris*. Vicariant speciation was consistently inferred as predominant in the amphibian genera *Gastrotheca* and *Rhinella*; the insect genera *Ithomia*, *Morpho,* and *Trigona*; the dicot genera *Fuchsia* and *Menonvillea;* and the conifer genus *Podocarpus*. Finally, peripatric speciation was consistently inferred as predominant in the bird genus *Polioptila* (Table S1). The remaining 65 genera showed discrepancies in the predominant inferred mode, either among classification methods, between spatial representations, or both.

3.5 | Sensitivity to species distribution representation

The comparison between concave polygons and PDMs indicates that the representation of species distributions can strongly influence inferred speciation modes in some clades. In several genera, the predominant inferred mode shifted between sympatric and allopatric speciation depending on the type of spatial input used. This pattern was observed in *Atlapetes*, *Eresia*, *Burmeistera*, *Canavalia*, *Chaetanthera*, *Cobaea,* and *Danaea*. Other genera also showed differences between spatial representations, although these involved less pronounced shifts, such as changes between sympatric and parapatric or between parapatric and allopatric classifications (Table 2; Table S2). Together, these results suggest that inferred geographic modes of speciation are partly contingent on how species ranges are estimated, highlighting the importance of evaluating the sensitivity of macroevolutionary inferences to spatial data representation.

Table 2. Distribution of genera across congruent geographic modes of speciation. For the two summary categories ("Congruent across all methods and spatial inputs" and "Non-congruent across methods and/or spatial representations"), values are given as absolute frequencies of genera, both for the full dataset and for the subset of genera with ≥ 50% sampling fraction. For the allopatric (vicariant, peripatric) and overlap-associated (parapatric, sympatric) sub-modes, values are given as percentages of the total number of genera in each subset

Category	Number of genera	Number of genera (≥ 50% sampling fraction)	Main interpretation
Congruent across all methods and spatial inputs	37	14	Robust candidates
Non-congruent across methods and/or spatial representations	65	43	Methodologically informative clades
Strictly allopatric scenarios			Isolation-associated divergence
Vicariant	23%	21%	
Peripatric	3%	7%	
Overlap-associated scenarios			Candidate non-allopatric / secondary-contact systems
Parapatric	15%	29%	
Sympatric	59%	43%	

4 | Discussion

4.1 | Multiple geographic pathways to Andean speciation

Our comparative analysis of 102 Andean-centred clades indicates that diversification in the Andes has not been dominated by a single geographic mode of speciation. Depending on the classification method and spatial representation used, strictly allopatric scenarios, including vicariance and peripatry, were inferred as predominant in 23–30% of clades, whereas overlap-associated scenarios, corresponding to parapatric or sympatric classifications, were inferred in 57–67% of clades. The remaining clades showed mixed signals, with no single geographic mode consistently favoured. Together, these results suggest that Andean diversification has proceeded through a mosaic of geographic contexts, involving both spatial isolation and geographic overlap among diverging lineages.

This finding is consistent with the exceptional spatial complexity of the Andes. Mountain systems can simultaneously promote isolation, connectivity and ecological divergence by combining steep environmental gradients, complex topography, climatic heterogeneity and dynamic geological histories. In the Andes, asynchronous uplift, valley formation, habitat fragmentation and Quaternary climatic oscillations may have repeatedly isolated populations, favouring allopatric divergence (Kattan et al., 2004; Ramírez-Barahona and

Eguiarte, 2013; Muñoz-Ortiz et al., 2015). At the same time, elevational gradients, ecological turnover and environmental heterogeneity have created opportunities for divergence among geographically adjacent or partially overlapping populations, particularly where divergent selection acts across steep environmental gradients (Turelli et al., 2001; Rundle and Nosil, 2005; Bolnick and Fitzpatrick, 2007; Nosil, 2012; Rodríguez-Cajarville et al., 2022). Our results therefore support a view of the Andes not as a system governed by a single speciation mechanism, but as a heterogeneous evolutionary landscape in which different clades may have diversified through different spatial pathways.

According to Hernández‐Hernández et al. (2021), allopatry may be the most frequent geographic mode of speciation across the Tree of Life, although the authors also noted that "there are few groups in which sympatric speciation might be as or more common." This variation in the predominant mode of speciation across taxa is consistent with our results, and both suggest that the prevailing geographic mode of speciation may be shaped by lineage-specific ecological and life-history traits rather than by a general process. Our analysis did not recover strictly allopatric scenarios as the most frequent genus-level predominant speciation mode. However, this result should not be interpreted as evidence that allopatric speciation has been less frequent across all Andean lineages or individual speciation events. Several considerations must be taken into account when assessing speciation in the presence of gene flow, which we discuss throughout the following sections.

The multi-pathway interpretation is particularly relevant because Andean diversification has often been framed around alternative explanations emphasizing either geographic isolation or ecological divergence. The geographic-isolation perspective highlights the role of uplift, climatic oscillations and landscape fragmentation in subdividing populations, whereas the ecological-divergence perspective emphasizes the creation of novel habitats and adaptive opportunities along elevational and climatic gradients (Bates and Zink, 1994; Kattan et al., 2004; Rundle and Nosil, 2005; Amico et al., 2007; Ramírez-Barahona and Eguiarte, 2013). Our results suggest that these explanations should not be treated as mutually exclusive. Instead, the same mountain system may generate repeated opportunities for both isolation-driven and overlap-associated divergence, depending on the spatial scale, taxonomic group, dispersal ability, ecological specialization and historical context of each clade. By showing that allopatric and overlap-associated scenarios coexist across a broad set of Andean-centred clades, our results shift the emphasis from identifying a single dominant geographic mode in the region to understanding how different spatial pathways are assembled across taxa, landscapes and evolutionary timescales.

4.2 | Interpreting overlap-associated classifications with caution

A central result of our study is the relatively high frequency of clades classified as predominantly sympatric or parapatric. However, these classifications require careful interpretation. The methods used here infer geographic modes of speciation from present-day species distributions and phylogenetic relationships; they do not directly estimate historical gene flow, reproductive isolation or the demographic conditions under which species originated. Consequently, classifications as “sympatric” or “parapatric” should be interpreted as indications that current spatial and phylogenetic patterns are

consistent with overlap-associated divergence, rather than as definitive proof of speciation with gene flow.

This distinction is important because sympatric speciation remains one of the most debated modes of species formation (Coyne and Orr, 2004; Bolnick and Fitzpatrick, 2007; Smith, 1966). Demonstrating sympatric speciation requires evidence that sister species are geographically overlapping, reproductively isolated, historically unlikely to have passed through an allopatric phase and differentiated through mechanisms capable of overcoming the homogenizing effects of gene flow (Coyne and Orr, 2004). Our study can address only part of this evidentiary framework, namely the association between phylogenetic relatedness and geographic overlap. It does not evaluate reproductive isolation, historical demography, genomic introgression or the ecological basis of divergence. Therefore, the high proportion of clades classified as sympatric should be regarded as a hypothesis-generating result rather than as a direct estimate of the frequency of true sympatric speciation in the Andes.

The same caution applies to parapatric classifications. Parapatric speciation implies divergence across adjacent ranges, potentially under reduced but non-zero gene flow (Coyne and Orr, 2004; Butlin et al., 2008). However, present-day adjacency may also result from secondary contact following an earlier allopatric phase, a well-recognized challenge in the study of speciation geography (Smadja and Butlin, 2011; Pettengill and Moeller, 2012; Dong et al., 2020). Although the simulation-based framework used here was designed to distinguish among geographic modes under scenarios that include post-speciation range change, no method based exclusively on current distributions can fully reconstruct the spatial history of divergence (Losos and Glor, 2003; Skeels and Cardillo, 2019). For this reason, we use the broader term “overlap-associated scenarios” to describe clades classified as sympatric or parapatric, while recognizing that these patterns may reflect either divergence with gene flow, secondary contact after allopatry or a combination of both across different speciation events within a clade. Accordingly, the value of these classifications lies not in resolving the demographic history of each clade, but in identifying systems in which overlap-associated divergence, secondary contact or recurrent spatial transitions can be explicitly tested.

4.3 | Spatial representation as a source of macroevolutionary uncertainty

Our results also show that inferred geographic modes of speciation can be sensitive to how species distributions are represented. In several genera, classifications differed between concave polygons and potential distribution models, and in some cases the inferred predominant mode shifted between allopatric and sympatric scenarios. This sensitivity highlights a key methodological issue for studies of speciation geography: estimates of range overlap, range asymmetry and range isolation are not intrinsic properties of clades, but depend on the spatial representation used to approximate species distributions.

Comparative approaches to speciation geography have long relied on present-day range configurations to infer historical processes of lineage divergence (Lynch, 1989; Chesser and Zink, 1994; Barraclough and Vogler, 2000; Losos and Glor, 2003). However, the reliability of such inferences depends strongly on how species distributions are represented. Range maps, point localities, spatial polygons and species distribution models capture different

dimensions of geographic distributions and differ in their susceptibility to omission and commission errors (Graham and Hijmans, 2006; Rondinini et al., 2006; Hurlbert and Jetz, 2007; La Sorte and Hawkins, 2007; O'Connor et al., 2019; Hughes et al., 2021). Expert-based range maps have been widely used, particularly in conservation biology, but concerns remain regarding their objectivity, reproducibility, accuracy assessment and potential inclusion of unsuitable areas (Mainali et al., 2020; Aronsson et al., 2024).

The two spatial representations used here therefore bracket different assumptions about species distributions. Concave polygons provide conservative approximations of occupied areas based directly on known occurrences, but they are sensitive to sampling gaps and may underestimate ranges for poorly collected species. Potential distribution models, by contrast, estimate areas of environmental suitability and may reduce omission error, but they can overestimate realized occupancy when dispersal limitation, biotic interactions, historical barriers, or non-climatic habitat requirements are not accounted for (Graham and Hijmans, 2006; Rondinini et al., 2006; Barve et al., 2011). These differences are especially relevant in the Andes, where sampling is spatially uneven and species ranges are shaped by fine-scale topography, elevational turnover and ecological specialization.

The contrast between spatial inputs therefore provides more than a methodological caveat; it is itself a biogeographical result. Genera whose inferred mode changes depending on spatial representation may represent cases in which current knowledge of distributions is insufficient, occurrence records are geographically biased or potential suitable areas are not equivalent to realized ranges. Conversely, genera with congruent classifications across both spatial inputs and all classification methods provide stronger candidates for future evolutionary investigation. We therefore suggest that sensitivity to spatial representation should be treated explicitly in comparative studies of speciation geography, rather than as a secondary technical issue.

4.4 | Congruent clades and testable hypotheses

Despite methodological sensitivity in some genera, 37 clades showed congruent classifications across all three classification methods and both spatial representations. These clades are particularly valuable because their inferred geographic mode was robust to both analytical approach and range-estimation method. Among them, several genera were consistently classified under overlap-associated scenarios, including clades of amphibians, mammals, birds, insects, angiosperms, ferns and fungi. These lineages provide a prioritized set of candidates for future studies aimed at testing whether geographic overlap reflects divergence with gene flow, secondary contact after allopatric divergence or recurrent combinations of both.

The three example clades highlighted in Figure 1 illustrate this range of outcomes. *Menonvillea* represents a robust vicariant signal, *Diglossa* a predominantly parapatric signal with secondary support for sympatry under PDMs, and *Nicotiana* a consistently overlap-associated signal that warrants further ecological, cytogenetic and genomic investigation. More broadly, several congruent clades already have independent evidence consistent with ecological differentiation, coexistence among close relatives or spatially complex divergence. For example, ecological divergence and coexistence among congeners have been discussed in genera such as *Centropogon* (Lagomarsino and Muchhala, 2019),

Senna (Robbiati et al., 2022) and *Diglossa* (Mauck and Burns, 2009). Similarly, non-allopatric or geographically complex divergence has been suggested for lineages such as *Heliotropium* (Luebert et al., 2014) and *Napeogenes* (Elias et al., 2009). However, for many genera included in this study, the geography of speciation has not been evaluated explicitly. Our results therefore provide a comparative framework for identifying clades in which fine-scale phylogeographic, ecological and population genomic analyses would be especially informative.

Such studies should focus on testing the mechanisms that cannot be resolved by the present comparative approach. These include the timing and direction of gene flow, the strength of reproductive isolation, the role of divergent ecological selection, the history of range shifts and the possibility of secondary contact. In particular, genomic analyses of sister species, combined with ecological niche comparisons and explicit demographic modelling, would allow researchers to distinguish among alternative histories that can produce similar present-day patterns of range overlap (Pinho and Hey, 2010; Smadja and Butlin, 2011; Seehausen et al., 2014). Thus, the most robust contribution of our comparative framework may be not the assignment of any individual genus to a definitive mode of speciation, but the prioritization of clades in which competing hypotheses about isolation, secondary contact and divergence with gene flow can now be tested explicitly.

The prevalence of congruent sympatric classifications among dicotyledonous and fern genera is particularly intriguing because plant lineages may diversify through mechanisms capable of producing reproductive isolation under geographic overlap. Polyploidy is one such mechanism: whole-genome duplication can promote reproductive isolation, generate phenotypic novelty and facilitate ecological differentiation among cytotypes or closely related lineages (Soltis et al., 2004; Soltis et al., 2015; Van de Peer et al., 2021). Because polyploidy is often associated with environmentally heterogeneous, stressful or mountain environments, it represents a plausible hypothesis for some Andean plant clades classified under overlap-associated scenarios (Brochmann et al., 2004; Theodoridis et al., 2013; Rice et al., 2019). However, our analyses do not test polyploidy directly. Future studies integrating chromosome counts, genome-size estimates, cytogenomic data, ecological niche comparisons and elevational range analyses could assess whether polyploidy has contributed to overlap-associated divergence in Andean plant lineages.

4.5 | Implications for the biogeography of mountain biodiversity

Our findings refine the common view of mountains as engines of diversification primarily through geographic isolation. Although barriers, fragmentation and climatic oscillations remain central to Andean diversification (Kattan et al., 2004; Ramírez-Barahona and Eguiarte, 2013; Muñoz-Ortiz et al., 2015), the frequent recovery of overlap-associated classifications suggests that mountain landscapes may also promote divergence under geographic proximity. In topographically complex systems, geographic proximity does not necessarily imply ecological similarity: nearby populations may experience sharply contrasting climatic, elevational or habitat conditions, allowing divergent selection to contribute to lineage divergence even in the absence of complete spatial isolation (Turelli et al., 2001; Rundle and Nosil, 2005; Bolnick and Fitzpatrick, 2007; Nosil, 2012; Rodríguez-Cajarville et al., 2022).

This perspective supports a more integrative view of mountain speciation, in which isolation, environmental heterogeneity and range dynamics interact across spatial and temporal scales (Rahbek et al., 2019; Muellner-Riehl, 2019; Perrigo et al., 2020). Under this view, allopatric, parapatric and sympatric classifications are best understood as simplified representations of a more complex geographic continuum. The relative importance of each geographic scenario is likely to vary across organisms with different dispersal capacities, habitat requirements, generation times and ecological tolerances. Future comparative studies could test whether overlap-associated classifications are more common in clades with low dispersal ability, narrow elevational ranges, specialized habitat requirements or strong ecological turnover across elevational gradients (Hughes and Atchison, 2015; Muellner-Riehl, 2019; Niu et al., 2019; Rodríguez-Cajarville et al., 2022).

4.6 | Limitations and future directions

Several limitations should be considered when interpreting our findings. First, present-day species distributions may not preserve a perfect signal of the geography of speciation. Range expansion, range contraction, extinction, dispersal and secondary contact can all alter spatial patterns after species formation (Losos and Glor, 2003). Although the simulation-based approach used here accounts for some degree of post-speciation range change (Skeels and Cardillo, 2019), historical range dynamics remain a major source of uncertainty.

Second, species distributions in the Andes are affected by uneven sampling. Many Andean regions remain under-collected, and sampling intensity varies strongly among countries, elevations, habitats and taxonomic groups. This may affect both concave polygons and potential distribution models, although in different ways. Third, our analyses were conducted at the genus level, and genera may contain multiple speciation events that occurred under different geographic contexts. Therefore, the inferred “predominant” mode should not be interpreted as applying uniformly to every speciation event within a clade. This issue is particularly relevant for species-rich genera, where the inferred predominant mode may summarize heterogeneous histories across many independent divergence events (Losos and Glor, 2003; Skeels and Cardillo, 2019).

Fourth, our analyses relied on available synthetic timetrees and did not propagate phylogenetic uncertainty across alternative tree topologies or divergence-time estimates. This may influence summary metrics based on sister-lineage relationships and node ages, particularly in poorly resolved clades (Kumar et al., 2022; Skeels and Cardillo, 2019). Fifth, our taxonomic sampling depends on the availability of phylogenies and occurrence data, which may bias the dataset toward better-studied groups. Because sampling fractions varied substantially among genera, inferences for poorly sampled clades should be treated with greater caution than those for clades with more complete phylogenetic and occurrence sampling. This limitation is common in large-scale comparative analyses, but it should be acknowledged because it may influence the apparent prevalence of particular speciation modes.

An additional caveat concerns the interpretation of the mixed model. In the simulation framework of Skeels and Cardillo (2019), the mixed model represents clades in which all four geographic modes contribute to diversification. However, the original cross-validation

analyses showed that this category had lower reclassification accuracy than single-mode models. Consequently, mixed histories may be difficult to detect and could be underrepresented in empirical classifications. For this reason, clades classified under a single predominant mode should not necessarily be interpreted as having diversified exclusively through that mode.

Finally, our classification framework does not incorporate ecological traits, dispersal traits, reproductive systems or genomic evidence. Including these additional data types would allow future studies to move from geographic inference toward mechanistic explanations of Andean diversification. In this sense, the classifications reported here should be interpreted as comparative, hypothesis-generating inferences about the geographic context of diversification, rather than as direct reconstructions of individual speciation events. Future studies integrating spatially explicit demographic models, ecological niche comparisons, genomic data, cytogenetic information and trait-based analyses will be essential to test the mechanisms underlying the geographic patterns identified here.

5 | Conclusions

Our study provides a broad comparative assessment of speciation geography across Andean-centred clades and shows that Andean diversification is best understood as a geographically heterogeneous process. Strictly allopatric scenarios explain a substantial fraction of clades, but overlap-associated scenarios were inferred more frequently across methods and spatial representations. These results do not demonstrate that speciation with gene flow has predominated in the Andes, but they do indicate that many clades exhibit spatial and phylogenetic patterns consistent with such scenarios.

By shifting the focus from single-clade case studies to a multi-taxon comparative perspective, this study highlights the Andes as an evolutionary landscape shaped by multiple spatial pathways to lineage divergence. The congruent clades identified here provide a set of candidate systems for future phylogeographic, ecological, cytogenetic, and population genomic research aimed at testing the mechanisms underlying mountain biodiversity.

Author Contributions: Y.F.M.H. and A.F.R.T. conceived and designed the study. A.F.R.T. collected the data, performed the analyses, conducted the literature review, and compiled the background information. A.F.R.T. wrote the first draft of the manuscript. Y.F.M.H. and A.F.R.T. contributed to manuscript revisions and approved the final version.

Acknowledgements: We are grateful to CIBioFi at Universidad del Valle for providing access to its computational infrastructure and facilities, where multiple analyses were conducted.

Funding: This research received no external funding and was conducted using institutional resources only.

Disclosure: Fieldwork permission statements: This study is based entirely on previously published data and did not involve new fieldwork. As such, no permits were required.

Conflicts of Interest: The authors declare no conflicts of interest.

Data Availability Statement: All data and code used in the analysis are available at GitHub (https://github.com/AndresTorres7/Bsc-thesis-Speciation-in-the-Andes) for purposes of reproducing or extending the analysis.

References

1. Amico, G.C., Vidal-Russell, R. and Nickrent, D.L. 2007 'Phylogenetic relationships and ecological speciation in the mistletoe Tristerix (Loranthaceae): the influence of pollinators, dispersers, and hosts', American journal of botany, 94(4), pp. 558–567. https://doi.org/10.3732/AJB.94.4.558.
2. Aronsson, H., Zizka, A., Antonelli, A., Faurby, S. 2024. Expert-based range maps cannot be replicated using data-driven methods but macroecological conclusions arising from them can. Journal of Biogeography 51, 1549–1559. https://doi.org/10.1111/jbi.14847
3. Barraclough, T.G., Vogler, A.P. 2000. Detecting the Geographical Pattern of Speciation from Species‐Level Phylogenies. Am. Nat. 155, 419–434. https://doi.org/10.1086/303332
4. Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J. and Villalobos, F. 2011 'The crucial role of the accessible area in ecological niche modeling and species distribution modeling', Ecological Modelling, 222(11), pp. 1810–1819. https://doi.org/10.1016/J.ECOLMODEL.2011.02.011.
5. Bates, J.G. and Zink Tm, R.G. 1994 'Evolution into the Andes: Molecular Evidence for Species Relationships in the Genus Leptopogon', The Auk, 111(3), pp. 507–515. https://doi.org/10.1093/auk/111.3.507.
6. Bolnick, D.I. and Fitzpatrick, B.M. 2007 'Sympatric Speciation: Models and Empirical Evidence', 38, pp. 459– 487. https://doi.org/10.1146/annurev.ecolsys.38.091206.095804.
7. Butlin, R.K., Galindo, J. and Grahame, J.W. 2008 'Sympatric, parapatric or allopatric: the most important way to classify speciation?', Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1506), pp. 2997–3007. https://doi.org/10.1098/RSTB.2008.0076.
8. Blum, M.G.B., François, O., 2006. Which Random Processes Describe the Tree of Life? A Large-Scale Study of Phylogenetic Tree Imbalance. Syst. Biol. 55, 685–691. https://doi.org/10.1080/10635150600889625
9. Brochmann, C., Brysting, A.K., Alsos, I.G., Borgen, L., Grundt, H.H., Scheen, A.-C. and Elven, R. 2004. Polyploidy in arctic plants. Biol J Linn Soc 82, 521–536. https://doi.org/10.1111/j.1095-8312.2004.00337.x
10. Chesser, R.T., Zink, R.M., 1994. Modes of Speciation in Birds: A Test of Lynch's Method. Evolution 48, 490–497. https://doi.org/10.2307/2410107
11. Cobos, M.E., Peterson, A.T., Barve, N. and Osorio-Olvera, L. 2019 'Kuenm: An R package for detailed development of ecological niche models using Maxent', PeerJ, 2019(2), p. e6281. https://doi.org/10.7717/PEERJ.6281/SUPP-2.
12. Coyne, J.A. and Orr, H.A. 2004 Speciation. Sinauer Associates, Sunderland, MA.
13. Csilléry, K., François, O. and Blum, M.G.B. 2012 'abc: an R package for approximate Bayesian computation (ABC)', Methods in Ecology and Evolution, 3(3), pp. 475– 479. https://doi.org/10.1111/J.2041-210X.2011.00179.X.

14. Dong, F., Hung, C.M. and Yang, X.J. 2020 ‘Secondary contact after allopatric divergence explains avian speciation and high species diversity in the Himalayan-Hengduan Mountains’, Molecular Phylogenetics and Evolution, 143, p. 106671. https://doi.org/10.1016/J.YMPEV.2019.106671.
15. Elias, M., Joron, M., Willmott, K., Silva-Brandão, K.L., Kaiser, V., Arias, C.F., Piñerez, L.M.G., Uribe, S., Brower, A.V.Z., Freitas, A.V.L., Jiggins, C.D., 2009. Out of the Andes: patterns of diversification in clearwing butterflies. Mol. Ecol. 18, 1716–1729. https://doi.org/10.1111/j.1365-294X.2009.04149.x
16. Escobar, L.E., Lira-Noriega, A., Medina-Vogel, G. and Peterson, A.T. 2014 ‘Potential for spread of the white-nose fungus (Pseudogymnoascus destructans) in the Americas: use of Maxent and NicheA to assure strict model transference’, Geospatial health, 9(1), pp. 221–229. https://doi.org/10.4081/GH.2014.19.
17. Esquerré, D., Brennan, I.G., Catullo, R.A., Torres-Pérez, F. and Keogh, J.S. 2019 ‘How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America’s most species rich lizard radiation (Squamata: Liolaemidae)’, Evolution, 73(2), pp. 214– 230. https://doi.org/10.1111/EVO.13657.
18. Fitzpatrick, B.M., Turelli, M., 2006. The Geography of Mammalian Speciation: Mixed Signals from Phylogenies and Range Maps. Evolution 60, 601–615. https://doi.org/10.1111/j.0014-3820.2006.tb01140.x
19. Hernández-Hernández, T., Miller, E.C., Román-Palacios, C. and Wiens, J.J. 2021 ‘Speciation across the Tree of Life’, Biological Reviews, 96(4), pp. 1205–1242. https://doi.org/10.1111/brv.12698.
20. Hurlbert, A.H., Jetz, W., 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proceedings of the National Academy of Sciences 104, 13384–13389. https://doi.org/10.1073/pnas.0704469104
21. Hughes, A.C., Orr, M.C., Yang, Q. and Qiao, H. 2021 ‘Effectively and accurately mapping global biodiversity patterns for different regions and taxa’, Global Ecology and Biogeography, 30(7), pp. 1375– 1388. https://doi.org/10.1111/GEB.13304.
22. Hughes, C.E. and Atchison, G.W. 2015 ‘The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains’, New Phytologist, 207(2), pp. 275–282. https://doi.org/10.1111/NPH.13230.
23. Graham, C.H., Hijmans, R.J., 2006. A comparison of methods for mapping species ranges and species richness. Glob. Ecol. Biogeogr. 15, 578–587. https://doi.org/10.1111/j.1466-8238.2006.00257.x
24. Graham, A. 2009 ‘The Andes: a geological overview from a biological perspective’, 96(3), pp. 371–385. https://doi.org/10.3417/2007146.
25. Gregory-Wodzicki, K. M. 2000 ‘Uplift history of the Central and Northern Andes: A review’, GSA Bulletin, 112(7), pp. 1091–1105. https://doi.org/10.1130/0016-7606(2000)112<1091:UHOTCA>2.0.CO;2
26. Kattan, G.H., Franco, P., Rojas, V. and Morales, G. 2004 Biological diversification in a complex region: a spatial analysis of faunistic diversity and biogeography of the Andes of Colombia. *Journal of Biogeography*, 31, 1829–1839. https://doi.org/10.1111/j.1365-2699.2004.01109.x
27. Körner, C., Jetz, W., Paulsen, J., Payne, D., Rudmann-Maurer, K. and Spehn, E.M. (2017) A global inventory of mountains for bio-geographical applications. Alpine Botany, 127, 1–15. https://doi.org/10.1007/s00035-016-0182-6

28. Kuhn, M. 2008. “Building Predictive Models in R Using the caret Package.” Journal of Statistical Software, 28(5), 1–26. doi:10.18637/jss.v028.i05. https://www.jstatsoft.org/index.php/jss/article/view/v028i05.
29. Kumar, S., Suleski, M., Craig, J.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G. and Hedges, S.B. 2022. TimeTree 5: An Expanded Resource for Species Divergence Times. Molecular Biology and Evolution, 39, msac174. DOI: 10.1093/molbev/msac174.
30. Lagomarsino, L.P. and Muchhala, N. 2019 ‘A gradient of pollination specialization in three species of Bolivian Centropogon’, American Journal of Botany, 106(5), pp. 633–642. https://doi.org/10.1002/AJB2.1276.
31. La Sorte, F., Hawkins, B., 2007. Range maps and species richness patterns: Errors of commission and estimates of uncertainty. Ecography 30, 649–662. https://doi.org/10.1111/j.2007.0906-7590.05084.x
32. Losos, J.B. and Glor, R.E. 2003 ‘Phylogenetic comparative methods and the geography of speciation’, Trends in Ecology and Evolution, 18(5), pp. 220–227. https://doi.org/10.1016/S0169-5347(03)00037-5.
33. Luebert, F., Jacobs, P., Hilger, H.H., Muller, L.A.H., 2014. Evidence for nonallopatric speciation among closely related sympatric Heliotropium species in the Atacama Desert. Ecol. Evol. 4, 266–275. https://doi.org/10.1002/ece3.929
34. Lynch 1989 ‘The gauge of speciation: on the frequencies of modes of speciation. ’, in Otte, D. and Endler, J.A. (eds) *Speciation and its consequences*. Sinauer Associates.
35. Luebert, F., Muller, L.A.H., 2015. Effects of mountain formation and uplift on biological diversity. Frontiers in Genetics. 6. https://doi.org/10.3389/fgene.2015.00054
36. Madriñán, S., Cortés, A.J. and Richardson, J.E. 2013 ‘Páramo is the world’s fastest evolving and coolest biodiversity hotspot’, Frontiers in Genetics, 4(OCT). https://doi.org/10.3389/fgene.2013.00192.
37. Mainali, K., Hefley, T., Ries, L., Fagan, W.F., 2020. Matching expert range maps with species distribution model predictions. Conservation Biology 34, 1292–1304. https://doi.org/10.1111/cobi.13492
38. Mauck, W.M. and Burns, K.J. 2009 ‘Phylogeny, biogeography, and recurrent evolution of divergent bill types in the nectar-stealing flowerpiercers (Thraupini: Diglossa and Diglossopis)’, Biological Journal of the Linnean Society, 98(1), pp. 14–28. https://doi.org/10.1111/j.1095-8312.2009.01278.x.
39. Muellner-Riehl, A.N., 2019. Mountains as Evolutionary Arenas: Patterns, Emerging Approaches, Paradigm Shifts, and Their Implications for Plant Phylogeographic Research in the Tibeto-Himalayan Region. Frontiers in Plant Science. https://doi.org/10.3389/fpls.2019.00195
40. Muñoz-Ortiz, A., Velásquez-Álvarez, Á.A., Guarnizo, C.E. and Crawford, A.J. 2015 ‘Of peaks and valleys: testing the roles of orogeny and habitat heterogeneity in driving allopatry in mid-elevation frogs (Aromobatidae: Rheobates) of the northern Andes’, Journal of Biogeography, 42(1), pp. 193–205. https://doi.org/10.1111/jbi.12409.
41. Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. and Kent, J. 2000 ‘Biodiversity hotspots for conservation priorities’, Nature 2000 403:6772, 403(6772), pp. 853–858. https://doi.org/10.1038/35002501.
42. Niu, Y., Yang, S., Zhou, J., Chu, B., Ma, S., Zhu, H. and Hua, L. 2019 ‘Vegetation distribution along mountain environmental gradient predicts shifts in plant community response to climate change in alpine meadow on the Tibetan Plateau’, Science of The Total Environment, 650, pp. 505–514. https://doi.org/10.1016/j.scitotenv.2018.08.390.

43. Nosil, P., 2012. Ecological Speciation, Oxford Series in Ecology and Evolution. Oxford University Press, Oxford, New York.
44. O'Connor, D., Stacy-Dawes, J., Muneza, A., Fennessy, J., Gobush, K., Chase, M.J., Brown, M.B., Bracis, C., Elkan, P., Zabeirou, A.R.M., Rabeil, T., Rubenstein, D., Becker, M.S., Phillips, S., Stabach, J.A., Leimgruber, P., Glikman, J.A., Ruppert, K., Masiaine, S. and Mueller, T. 2019 'Updated geographic range maps for giraffe, *Giraffa spp*., throughout sub-Saharan Africa, and implications of changing distributions for conservation', Mammal Review, 49(4), pp. 285–299. https://doi.org/10.1111/mam.12165.
45. Pérez-Escobar, O.A., Zizka, A., Bermúdez, M.A., Meseguer, A.S., Condamine, F.L., Hoorn, C., Hooghiemstra, H., Pu, Y., Bogarín, D., Boschman, L.M., Pennington, R.T., Antonelli, A., Chomicki, G., 2022. The Andes through time: evolution and distribution of Andean floras. Trends in Plant Science 27, 364–378. https://doi.org/10.1016/j.tplants.2021.09.010.
46. Perrigo, A., Hoorn, C. and Antonelli, A. 2020 'Why mountains matter for biodiversity', Journal of Biogeography, 47(2), pp. 315–325. https://doi.org/10.1111/JBI.13731.
47. Pettengill, J.B. and Moeller, D.A. 2012 'Phylogeography of speciation: allopatric divergence and secondary contact between outcrossing and selfing Clarkia', Molecular Ecology, 21(18), pp. 4578–4592. https://doi.org/10.1111/j.1365-294X.2012.05715.x.
48. Phillips, S.J., Anderson, R.P. and Schapire, R.E. 2006 'Maximum entropy modeling of species geographic distributions', Ecological Modelling, 190(3–4), pp. 231–259. https://doi.org/10.1016/J.ECOLMODEL.2005.03.026.
49. Pinho, C. and Hey, J. 2010 'Divergence with Gene Flow: Models and Data', 41, pp. 215–230. https://doi.org/10.1146/annurev-ecolsys-102209-144644.
50. Price, M.F. and Butt, N. 2000 Forests in sustainable mountain development: a state of knowledge report for 2000. Task Force on Forests in Sustainable Mountain Development., Forests in sustainable mountain development: a state of knowledge report for 2000. Task Force on Forests in Sustainable Mountain Development. CABI Publishing. https://doi.org/10.1079/9780851994468.0000.
51. Rahbek, C., Borregaard, M.K., Antonelli, A., Colwell, R.K., Holt, B.G., Nogues-Bravo, D., Rasmussen, C.M.Ø., Richardson, K., Rosing, M.T., Whittaker, R.J., Fjeldså, J., 2019. Building mountain biodiversity: Geological and evolutionary processes. Science 365, 1114–1119. https://doi.org/10.1126/science.aax0151
52. Ramírez-Barahona, S. and Eguiarte, L.E. 2013 'The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum', Ecology and Evolution, 3(3), pp. 725–738. https://doi.org/10.1002/ECE3.483.
53. Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., Meiri, S., Belmaker, J., Mayrose, I., 2019. The global biogeography of polyploid plants. Nat Ecol Evol 3, 265–273. https://doi.org/10.1038/s41559-018-0787-9
54. Robbiati, F.O., Moro Cordobés, F., Fortunato, R.H., Anton, A.M., Urdampilleta, J.D., 2022. Niche conservatism, divergence and polyploidy in Senna series Aphyllae (Fabaceae: Caesalpinioideae) from arid zones of South America. Bot. J. Linn. Soc. 199, 773–789. https://doi.org/10.1093/botlinnean/boab105
55. Rodríguez-Cajarville, M.J., Campagna, L., Lima-Rezende, C.A., Carboni, M., Tubaro, P.L. and Cabanne, G.S. 2022 'Genomic and phenotypic divergence-with-gene-flow across an ecological and elevational gradient in a neotropical bird', Journal of Biogeography, 49(8), pp. 1535–1548. https://doi.org/10.1111/jbi.14449.
56. Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H. and Possingham, H.P. 2006 'Tradeoffs of different types of species occurrence data for use in systematic conservation planning', Ecology Letters, 9(10), pp. 1136–1145. https://doi.org/10.1111/j.1461-0248.2006.00970.x.

57. Rundle, H.D., Nosil, P., 2005. Ecological speciation. Ecology Letters 8, 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x
58. Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C.S., Eroukhmanoff, F., Feder, J.L., Fischer, M.C., Foote, A.D., Franchini, P., Jiggins, C.D., Jones, F.C., Lindholm, A.K., Lucek, K., Maan, M.E., Marques, D.A., Martin, S.H., Matthews, B., Meier, J.I., Möst, M., Nachman, M.W., Nonaka, E., Rennison, D.J., Schwarzer, J., Watson, E.T., Westram, A.M., Widmer, A., 2014. Genomics and the origin of species. Nat Rev Genet 15, 176–192. https://doi.org/10.1038/nrg3644
59. Skeels, A., Cardillo, M., 2019. Reconstructing the Geography of Speciation from Contemporary Biodiversity Data. Am. Nat. 193, 240–255. https://doi.org/10.1086/701125
60. Smadja, C.M. and Butlin, R.K. 2011 'A framework for comparing processes of speciation in the presence of gene flow', Molecular Ecology, 20(24), pp. 5123–5140. https://doi.org/10.1111/j.1365-294X.2011.05350.x.
61. Smith, J.M. 1966 'Sympatric Speciation', https://doi.org/10.1086/282457, 100(916), pp. 637–650. https://doi.org/10.1086/282457.
62. Soltis, D.E., Soltis, P.S., Tate, J.A., 2004. Advances in the study of polyploidy since Plant speciation. New Phytologist 161, 173–191. https://doi.org/10.1046/j.1469-8137.2003.00948.x
63. Soltis, P.S., Marchant, D.B., Van de Peer, Y., Soltis, D.E., 2015. Polyploidy and genome evolution in plants. Current Opinion in Genetics and Development, Genomes and evolution 35, 119–125. https://doi.org/10.1016/j.gde.2015.11.003
64. Spehn, E.M., Rudmann-Maurer, K. and Körner, C. 2012 'Mountain biodiversity', 4(4), pp. 301–302. https://doi.org/10.1080/17550874.2012.698660.
65. Theodoridis, S., Randin, C., Broennimann, O., Patsiou, T., Conti, E., 2013. Divergent and narrower climatic niches characterize polyploid species of European primroses in Primula sect. Aleuritia. Journal of Biogeography 40, 1278–1289. https://doi.org/10.1111/jbi.12085
66. Turelli, M., Barton, N.H. and Coyne, J.A. 2001 'Theory and speciation', Trends in Ecology and Evolution, 16(7), pp. 330–343. https://doi.org/10.1016/S0169- 5347(01)02177-2.
67. Van de Peer, Y., Ashman, T.-L., Soltis, P.S., Soltis, D.E., 2021. Polyploidy: an evolutionary and ecological force in stressful times. Plant Cell 33, 11–26. https://doi.org/10.1093/plcell/koaa015
68. Zhao, Y.-J., Yin, G.-S., Pan, Y.-Z., Tian, B. and Gong, X. 2021 'Climatic Refugia and Geographical Isolation Contribute to the Speciation and Genetic Divergence in Himalayan-Hengduan Tree Peonies (Paeoniadelavayi and Paeonia ludlowii)', Frontiers in Genetics, 11, p. 1808.https://doi.org/10.3389/fgene.2020.595334.

Supplementary material for

The Geography of Speciation Across Andean-Centred Clades

Table S1. Predominant geographic mode of speciation inferred for each genus across classification methods and spatial inputs. Inferences are shown for three classification methods—linear discriminant analysis (LDA), neural networks (NN), and multinomial logistic regression (mnL)—and two spatial representations: concave polygons and potential distribution

models (PDMs). When no method is indicated, all three classification methods inferred the same geographic mode. n = number of species in the phylogeny. f = the sampling fraction, calculated as the ratio of species included in the phylogeny to the total number of accepted species with more than 2 georeferenced occurrences in GBIF. Genera with congruent geographic modes of speciation are shown in bold.

	Genus	Mode Polygon	Mode PDM	n	f
Amphibia	*Allobates*	parapatric	sympatric, parapatric (LDA)	14	0.33
	Alsodes	peripatric	vicariant	14	0.88
	Craugastor	**sympatric**	**sympatric**	32	0.33
	Dendropsophus	parapatric	vicariant	36	0.39
	Gastrotheca	**vicariant**	**vicariant**	21	0.38
	Hyloxalus	sympatric, parapatric (LDA)	sympatric	16	0.47
	Leptodactylus	sympatric	mixed, sympatric (LDA)	34	0.48
	Pristimantis	vicariant, sympatric (LDA)	sympatric	122	0.30
	Rhinella	**vicariant**	**vicariant**	22	0.31
Aves	*Atlapetes*	sympatric	vicariant	18	0.56
	Cinclodes	parapatric, sympatric (mnL)	sympatric, parapatric (LDA)	14	1
	Diglossa	**parapatric**	**parapatric**	18	1
	Elaenia	sympatric	mixed	19	0.95
	Grallaria	peripatric	mixed, sympatric (NN)	30	0.68
	Myrmotherula	peripatric, mixed (mnL)	peripatric (LDA), vicariant (NN), mixed (mnL)	23	0.96
	Pachyramphus	**parapatric**	**parapatric**	17	0.94
	Polioptila	**peripatric**	**peripatric**	12	0.80
	Scytalopus	mixed, peripatric (LDA)	vicariant	42	0.86
	Synallaxis	**parapatric**	**parapatric**	31	0.86
	Thamnophilus	parapatric	mixed, vicariant (NN)	30	1
Mammalia	***Akodon***	**sympatric**	**sympatric**	29	0.83
	Artibeus	parapatric (LDA), mixed (NN), sympatric (mnL)	mixed, parapatric (LDA)	16	0.73
	Sturnira	parapatric	sympatric	13	0.62
Reptilia	*Anolis*	sympatric, parapatric (LDA)	sympatric	154	0.62
	Kinosternon	vicariant	vicariant, sympatric (mnL)	18	0.82

	Liolaemus	mixed, peripatric (LDA)	sympatric	117	0.54
	Stenocercus	peripatric, mixed (NN)	sympatric	27	0.39
Insecta	*Centris*	mixed, sympatric (NN)	mixed	49	0.40
	Eresia	vicariant	sympatric	17	0.57
	Euglossa	peripatric	vicariant	62	0.64
	Ithomia	**vicariant**	**vicariant**	20	0.74
	Morpho	**vicariant**	**vicariant**	20	0.49
	Napeogenes	**parapatric**	**parapatric**	16	1
	Parides	mixed, vicariant (NN)	peripatric, mixed (NN)	20	0.63
	Pseudomyrmex	**sympatric**	**sympatric**	36	0.31
	Trigona	**vicariant**	**vicariant**	21	0.66
Monocots	***Aristida***	**parapatric**	**parapatric**	46	0.39
	Bactris	**parapatric**	**parapatric**	25	0.32
	Chamaedorea	parapatric	sympatric	47	0.47
	Fosterella	mixed, peripatric (LDA)	vicariant	18	0.58
	Geonoma	parapatric	sympatric	40	0.59
	Heliconia	mixed	sympatric	107	0.63
Dicots	***Bunchosia***	**sympatric**	**sympatric**	31	0.47
	Burmeistera	vicariant	sympatric	30	0.26
	Caiophora	peripatric	parapatric	17	0.61
	Canavalia	sympatric	vicariant	23	0.66
	Capsicum	**sympatric**	**sympatric**	19	0.46
	Cayaponia	mixed	mixed, vicariant (LDA)	16	0.25
	Centropogon	**sympatric**	**sympatric**	63	0.38
	Chaetanthera	sympatric	mixed, vicariant (LDA)	21	0.78
	Cobaea	sympatric	vicariant	16	0.94
	Coursetia	mixed, peripatric (LDA)	vicariant	23	0.66
	Cremastosperma	sympatric, parapatric (LDA)	sympatric	20	0.61
	Crotalaria	parapatric, mixed (NN)	sympatric, mixed (mnL)	45	0.63
	Disterigma	mixed, peripatric (LDA)	vicariant	31	0.86
	Drymonia	**sympatric**	**sympatric**	31	0.41
	Eschweilera	vicariant, sympatric (LDA)	sympatric	50	0.51
	Fuchsia	**vicariant**	**vicariant**	29	0.30
	Glossoloma	mixed, parapatric (LDA)	mixed, sympatric (LDA)	18	0.64
	Guarea	sympatric	vicariant, sympatric (LDA)	37	0.58
	Hedyosmum	parapatric	sympatric	25	0.64

	Heliotropium	**sympatric**	**sympatric**	30	0.43
	Hypericum	vicariant, sympatric (LDA)	vicariant, sympatric (LDA)	105	0.58
	Hypochaeris	mixed, sympatric (LDA)	vicariant, mixed (LDA)	16	0.36
	Inga	**sympatric**	**sympatric**	74	0.28
	Iochroma	sympatric, parapatric (LDA)	parapatric	12	0.57
	Jaltomata	parapatric	sympatric	20	0.50
	Lecythis	sympatric, parapatric (LDA)	sympatric, parapatric (LDA)	25	0.86
	Lupinus	**sympatric**	**sympatric**	95	0.26
	Lycianthes	parapatric (LDA), sympatric (NN), mixed (mnL)	sympatric	28	0.31
	Macrocarpaea	peripatric (LDA), parapatric (NN), mixed (mnL)	vicariant, mixed (LDA)	27	0.33
	Menonvillea	**vicariant**	**vicariant**	16	0.89
	Mimosa	sympatric, parapatric (LDA)	sympatric	219	0.43
	Nicotiana	**sympatric**	**sympatric**	36	0.71
	Nolana	**sympatric**	**sympatric**	34	0.64
	Portulaca	parapatric	sympatric	28	0.50
	Pouteria	sympatric	sympatric, mixed (LDA)	46	0.25
	Protium	vicariant, sympatric (LDA)	vicariant, sympatric (LDA)	72	0.63
	Psammisia	**sympatric**	**sympatric**	13	0.25
	Ranunculus	sympatric, mixed (mnL)	mixed	55	0.42
	Senna	**sympatric**	**sympatric**	62	0.30
	Siphocampylus	parapatric (LDA), vicariant (NN), mixed (mnL)	vicariant	48	0.32
	Stachys	**sympatric**	**sympatric**	31	0.31
	Symplocos	**sympatric**	**sympatric**	54	0.35
	Valeriana	vicariant	vicariant, sympatric (NN)	71	0.37
	Vasconcellea	parapatric	sympatric	13	0.65
	Viburnum	**sympatric**	**sympatric**	61	0.56
	Vicia	peripatric, mixed (mnL)	vicariant	30	0.56
Ferns	***Adiantum***	**sympatric**	**sympatric**	52	0.36
	Ascogrammitis	peripatric (LDA), mixed (NN), vicariant (mnL)	vicariant	12	0.8

	Asplenium	**sympatric**	**sympatric**	56	0.25
	Danaea	sympatric	vicariant	19	0.37
	Lellingeria	mixed, peripatric (LDA)	mixed, peripatric (LDA)	22	0.49
	Lindsaea	**sympatric**	**sympatric**	38	0.76
	Melpomene	parapatric, peripatric (NN)	sympatric, parapatric (LDA)	18	0.64
	Pleopeltis	sympatric, mixed (mnL)	sympatric	41	0.49
	Polystichum	sympatric, mixed (mnL)	sympatric	41	0.61
	Serpocaulon	**sympatric**	**sympatric**	30	0.67
Other Non-flowering plants	***Podocarpus***	**vicariant**	**vicariant**	21	0.42
	Selaginella	mixed, sympatric (LDA)	sympatric	81	0.30
	Zamia	parapatric, mixed (LDA)	sympatric	47	0.64
Fungi	***Sticta***	**sympatric**	**sympatric**	22	0.36