

Title

Pleistocene glacial cycles drive isolation, gene flow and speciation in the high elevation Andes

Authors

Bruno Nevado¹, Natalia Contreras-Ortiz^{2,3}, Colin Hughes^{4,*} and Dmitry A. Filatov^{1,*}

Author affiliations

1 Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

2 Laboratorio de Botánica y Sistemática, Departamento de Ciencias Biológicas, Universidad de los Andes, Apartado Aéreo 4976, Bogotá, Colombia

3 Jardín Botánico de Bogotá “José Celestino Mutis”, Avenida Calle 63 No. 68-95, Bogotá D.C, Colombia

4 Department of Systematic & Evolutionary Botany, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland

* Joint senior authors

Corresponding author

Bruno Nevado

Email: Bruno.nevado@plants.ox.ac.uk

Telephone: (044) 01865 275119

Word count (max 6,500)

Main body: 6,580

Introduction: 1,356

Materials and Methods: 1,317

Results: 1,938

Discussion: 1,825

Number of figures: 5 (3 in colour: Figures 1, 3 and 5)

Number of tables: 4

Supporting information: 2 (Supporting Fig. S1, Supporting Table S1)

Summary

- Mountain ranges are amongst the most species-rich habitats, with many large and rapid evolutionary radiations. The tempo and mode of diversification in these systems are key unanswered questions in evolutionary biology. We study the Andean *Lupinus* radiation to understand the processes driving very rapid diversification in montane systems.
- We use genomic and transcriptomic data of multiple species and populations, and apply phylogenomic and demographic analyses to test whether diversification proceeded without interspecific geneflow – as expected if Andean orogeny and geographic isolation were the main drivers of diversification; or if diversification was accompanied by gene flow – in which case other processes were likely involved.
- We uncover several episodes of gene flow between species, including very recent events likely prompted by changes in habitat connectivity during Pleistocene glacial cycles. Furthermore, we find that gene flow between species was heterogeneously distributed across the genome.
- We argue that exceptionally fast diversification of Andean *Lupinus* was partly due to Late Pleistocene glacial cycles, with associated cycles of expansion and contraction driving geographic isolation or secondary contact of species. Furthermore, heterogeneous gene flow across the genome suggests a role for selection and ecological speciation in rapid diversification in this system.

Keywords (5-8): Evolutionary radiation, Pleistocene glacial cycles, *Lupinus* spp., Andes Mountains, geneflow, speciation, phylogenomics, demographic modelling.

Introduction

Rapid evolutionary radiations are ubiquitous across the Tree of Life, but there is still little understanding about the tempo and mode of diversification or the evolutionary mechanisms responsible for rapid diversification. This is especially true for very recent evolutionary radiations confined to the Pleistocene, exemplified by the many radiations of high elevation plants in mountains across the globe (Linder, 2008; Hughes & Atchison, 2015). Rapid diversification in high elevation montane habitats has often been assumed to result from increased ecological opportunities due to the high physiographical heterogeneity in mountains prompting frequent episodes of geographic isolation and allopatric speciation (e.g. Simpson, 1964; Hughes & Eastwood, 2006). However, these highly heterogeneous montane habitats may also promote adaptive divergence via ecological speciation (e.g. Chapman *et al.*, 2013; Ruiz-Sanchez & Specht, 2014; Contreras-Ortiz *et al.*, 2018). Furthermore, patterns of geographic isolation in mountain ranges have changed dramatically over time – most notably during Pleistocene glacial cycles (e.g. Flantua & Hooghiemstra, 2018).

The Andes are one of the richest biodiversity hotspots in the world (Myers *et al.*, 2000; Olson & Dinerstein, 2002; Mittermeier *et al.*, 2005; Orme *et al.*, 2005), and the high elevation Andean grasslands of the Páramo and Puna in particular present an excellent system for investigating the processes driving rapid speciation and recent evolutionary radiations for several reasons. First, they harbour a remarkable number of large plant radiations exhibiting very high rates of species diversification and encompassing 50–100+ species (Madriñán *et al.*, 2013; Luebert & Weigend, 2014; Hughes & Atchison, 2015). Second, the fact that these large plant radiations are monophyletic and consist predominantly of endemic species suggests that they evolved via *in situ* diversification – contrary to other mountain systems where migration and colonization apparently played a more dominant role (Hughes & Atchison, 2015; Hughes, 2017; Kadereit, 2017; Xing & Ree, 2017). And third, high elevation Andean plant radiations are typically very recent, all of them within the last 5Myr, most of them < 2.5 Myr, and with the majority of species diversity arising in the Pleistocene (Madriñán *et al.*, 2013; Luebert & Weigend, 2014), in line with the recency of the final uplift of the Andes (Kroonenberg *et al.*, 1990; Ghosh *et al.*, 2006).

These high elevation Andean plant radiations have been attributed to increased opportunity for geographic isolation and allopatric speciation, driven by the emergence of new and unoccupied island-like habitats following the final uplift of the Andes 2–5 Myr ago (Kroonenberg *et al.*, 1990; Ghosh *et al.*, 2006), coupled with extremely high physiographic heterogeneity and extended multi-dimensional environmental gradients found in the Andes (Hughes & Eastwood, 2006; Drummond *et al.*, 2012). While these scenarios of allopatric and clinal speciation are clearly relevant for explaining the remarkable concentration of young and fast radiations in the high elevation Andes, they largely ignore the impacts of cyclical fragmentation and connectivity of high elevation grassland habitats through the Pleistocene (Hooghiemstra & Van der Hammen, 2004; Flantua *et al.*, 2014; Flantua & Hooghiemstra, 2018). Pleistocene glacial cycles have strongly affected habitat connectivity in the high elevation

Andes, although in different ways compared to north temperate regions: while glacial cycles in the latter regions resulted in long distance horizontal displacement of habitats (e.g. up to 2000 km northward shift in the treeline in Europe) (Hewitt, 1996; Hewitt, 2004; Gavin *et al.*, 2014), in the high elevation Andes glaciations were primarily associated with vertical and only minimal horizontal displacement of vegetation zones, with e.g. a downward shift of 1200 – 1400 m in the treeline at the Last Glacial Maximum (van der Hammen & Cleef, 1986; Hooghiemstra & Van der Hammen, 2004; Hooghiemstra *et al.*, 2006; Graham, 2009; Jomelli *et al.*, 2014). These vertical displacements resulted in expansion and contraction of available habitat areas and changes in the degree of connectivity (Simpson, 1974; Flantua *et al.*, 2014; Flantua & Hooghiemstra, 2018), presenting a highly dynamic and rapidly flickering connectivity landscape over the last 1Myr (Flantua & Hooghiemstra, 2018). In general, glacial periods correspond to lowering of vegetation belts throughout the Andes and expansion of distribution ranges of many high-elevation species, while warmer interglacial periods resulted in population fragmentation, isolation and reduced available area of high elevation grassland habitats (Simpson, 1975; van der Hammen & Cleef, 1986; Hooghiemstra *et al.*, 2006; Sklenář *et al.*, 2011). However, the inverse scenario likely occurred in the Altiplano, the high-elevation plateau covering part of the south-central Andes (Simpson, 1975), and timing of glacial advances and retreats during glacial cycles may have been somewhat different in different parts of the Andes (Zech *et al.*, 2008), reflecting the variable and complex mountain topography across the Andes.

The relative contributions of Andean orogeny and more recent Pleistocene glacial cycles in driving diversification of high elevation Andean plants remain unclear, partly due to the difficulties associated with reconstructing the evolutionary history of rapidly diversifying clades. However, “next generation” sequencing (NGS) data sets for high elevation Andean plants are providing phylogenetic resolution in these groups for the first time (Nevado *et al.*, 2016; Uribe-Convers *et al.*, 2016; Vargas *et al.*, 2017; Contreras-Ortiz *et al.*, 2018), and recent studies suggest that interspecific gene flow may have occurred in some Andean radiations (Kolar *et al.*, 2016; Duskova *et al.*, 2017; Vargas *et al.*, 2017; Pouchon *et al.* 2018). While these reports are at odds with a strict allopatric speciation scenario driven by Andean orogeny, thus far the timing of species divergence and hybridisation, and the effect of gene flow between species at a genome-wide scale, has not been studied. In this work, we address this gap by focusing on one of the most prominent high elevation Andean plant radiations, for which the largest genome scale datasets are available – the legume genus *Lupinus* (Hughes & Eastwood, 2006; Drummond *et al.*, 2012).

The Andean clade of *Lupinus* comprises c. 85 species with an estimated crown age of 1.19-3.5 Myr (Drummond, 2008; Drummond *et al.*, 2012) and presents one of the highest documented rates of net species diversification in plants (Hughes & Eastwood, 2006; Drummond *et al.*, 2012). The majority of Andean *Lupinus* species diversity occurs in the high elevation grasslands above 3200 m from Venezuela to northern Argentina, with only a few species found at lower elevations. The substantial variation in plant growth forms (Hughes & Eastwood, 2006; Hughes & Atchison, 2015; Contreras-Ortiz

et al., 2018) shows clear correlation with elevation, with annuals at lower elevations below 3000 m, and a wide range of perennials mainly above 3000 m – including woody shrubs of various dimensions, small treelets to 6m height, and dwarf prostrate mat-forming and acaulescent or stem rosette species. The Andean *Lupinus* species have a uniform $2n=48$ chromosome number (Conterato & Schifino-Wittmann, 2006), with no evidence that polyploidy played any role in the diversification of species in this clade. We have previously shown that adaptive evolution in this clade is widespread (Nevado *et al.*, 2016), and presented evidence to suggest both geographical and adaptive components of this radiation (Contreras-Ortiz *et al.*, 2018). However, the relative importance of purely geographical i.e. non-adaptive vs adaptive axes in the rapid diversification of Andean lupins has yet to be resolved (Givnish, 2015; Nevado *et al.*, 2016; Contreras-Ortiz *et al.*, 2018).

To investigate the evolutionary processes underpinning Andean *Lupinus* diversification, we analyse the patterns of genetic diversity and divergence of multiple species and populations using genomic and transcriptomic datasets. We test whether diversification in this system proceeded without gene flow between species – as expected if geographic isolation is the main driver of diversification – or if diversification was accompanied by gene flow – in which case processes other than geographic isolation and allopatric speciation would be inferred. Our results support several episodes of gene flow between different *Lupinus* species, including very recent events of gene flow likely prompted by changes to habitat connectivity during Late Pleistocene glacial cycles. Importantly, gene flow between species occurs heterogeneously across the genome – with gene flow restricted to some regions of the genome – which implies ecological divergence as an important process driving (or maintaining) species diversity. Taken together, these results suggest that rapid diversification in Andean *Lupinus* is at least partly due to ecological speciation, and that Pleistocene glacial cycles played an important role in diversification in this system.

Materials and Methods

Analysis of phylogenetic incongruence

To infer to what extent gene flow between species has occurred during the diversification of Andean *Lupinus* we analysed the 6,013 orthologous genes previously identified by Nevado *et al.* (2016). These genes were obtained by RNAseq and a clustering and phylogenetic-based orthology inference approach as detailed in Nevado *et al.* (2016). We focused on the wild Andean species – i.e., excluding the domesticated *L. mutabilis* – and sampled 25 species, and used the Mexican lupin *L. campestris* as an outgroup (Supporting Information Table S1). As positive selection was shown to be widespread in this clade (Nevado *et al.*, 2016), we considered only four-fold degenerate sites from each gene, and reduced the dataset by excluding genes with fewer than four phylogenetically-informative sites and fewer than 10 species sequenced.

This dataset was used to infer whether a phylogenetic tree (without hybridisation events) or a phylogenetic network (with one or several hybridisation events) better describe the evolutionary history

of Andean *Lupinus* using Species Networks applying Quartets (SNaQ) (Solís-Lemus & Ané, 2016) implemented in the software PHYLONETWORKS (Solís-Lemus *et al.*, 2017). In this approach, phylogenetic trees from multiple loci are used to estimate quartet concordance factors (CFs), i.e. the proportion of genes that support each possible relationship between each set of four species. These CFs are then used to reconstruct phylogenetic networks under incomplete lineage sorting (ILS) and a different number of hybridisation events and to calculate their respective pseudo-likelihoods. We used RAXML v. 8 (Stamatakis, 2014) to obtain gene trees with bootstrap support for each gene (with the GTR+GAMMA model and 100 bootstrap replicates), and used these trees to generate a species tree with ASTRAL (Mirarab & Warnow, 2015), and estimate CFs with PHYLONETWORKS (function readTrees2CF). Using the phylogeny obtained with ASTRAL as a starting tree, we then estimated the best phylogenetic network with varying number of hybridisation events allowed (h between 0 and 4) using the function snaq! in PHYLONETWORKS. To ensure convergence we performed 12 independent runs under each value of h . To infer the support for the best network identified, we analysed 100 bootstrap replicates, each performing 10 independent searches starting from the best network identified previously, and summarised the results with PHYLONETWORKS.

As the SNaQ approach estimates pseudo-likelihoods, we cannot readily compare the fit of models with different number of hybridisation events. Instead, we used the log pseudolikelihood profile with increasing number of hybridisation events: pseudolikelihood is expected to increase sharply with h until it reaches the optimal value, and then to increase more slowly with increasing h (Solís-Lemus & Ané, 2016). In addition, we compared the fit of the observed CFs to those expected under a phylogenetic tree without hybridisation (obtained with ASTRAL) and those expected under the best phylogenetic network with a single hybridisation event (obtained with SNaQ), using the TICC test (Stenz *et al.*, 2015) with the function “test.one.species.tree” from the R package PHYLLOLM (Ho & Ané, 2014). This test calculates, for each quartet of species, a p-value that expresses how unlikely are the CFs observed to have arisen under a model allowing for ILS alone. The p-values calculated for each quartet are then binned using four categories ($p < 0.01$, 0.01-0.05, 0.05-0.1, 0.1-1), and this binned distribution is compared to the expected values of 0.01, 0.04, 0.05 and 0.9 using a chi-square test with 3 df.

Demographic modelling of divergence events

To infer the demographic history of divergence between species, we analysed RAD data from two pairs of robustly supported sister lineages from the northern Andes, obtained with the nextRAD approach as detailed in Contreras-Ortiz *et al.* (2018). First, we analysed two high elevation acaulescent rosette forming sister species occupying disjunct distributions: *L. alopecuroides*, restricted to the central cordillera of Colombia and the contiguous ranges in north-central Ecuador, between 3600 m and 4700 m elevation; and *L. trianaanus*, which is endemic to the southern páramos of the eastern Cordillera of Colombia, between 3300 and 3700 m (Figure 1). These two species are geographically isolated by the deep Magdalena Valley separating these two Cordilleras. Second, we analysed two robustly supported

geographically disjunct sister lineages (here referred to as *Ocetá* and *Pisba/Cocuy*), within an as yet undescribed acaulescent rosette forming species of *Lupinus* (Contreras-Ortiz *et al.*, submitted) from the geographically adjacent páramos of Pisba/Cocuy and Ocetá in the northern part of the eastern Cordillera of Colombia (Contreras-Ortiz *et al.*, 2018). This dataset included between 14 and 17 individuals from each lineage, sequenced from between 2 and 4 locations in the Andes in Colombia and Northern Ecuador (Figure 1).

The raw Illumina sequences from (Contreras-Ortiz *et al.*, 2018) were trimmed with CUTADAPT v 1.8 (Martin, 2011) to remove adaptors and low quality ends (Phred score below 20), and the resulting trimmed reads were checked with FASTQC v 0.11 (Andrews, 2010). Trimmed reads were mapped to a draft genome assembly of the Andean crop lupin, *L. mutabilis* (Nevado *et al.*, *in prep*) using BWA v 0.7 (Li & Durbin, 2010) with the mem algorithm and default settings. Reads with mapping quality below 20 were excluded, and the package STACKS v 1.42 (Catchen *et al.*, 2013) was used to identify RAD loci along the genome. Individual genotypes with a minimum stack depth of 10 per individual (-m 10) and a minimum of two species represented at each locus (-p 2) were inferred in STACKS. For subsequent analyses, we kept a single representative locus for each overlapping site (--ordered_export) and a single random SNP from each locus (--write_random_snp). Population structure within each species was investigated with FASTSTRUCTURE v 1.0 (Raj *et al.*, 2014) assuming a varying number of clusters (1-5), while an analysis of molecular variance within each species was performed in ARLEQUIN v 3.5 (Excoffier & Lischer, 2010).

To infer the demographic history of divergence of each lineage pair we analysed their joint Site-Frequency-Spectrum (SFS) under several models of Isolation with Migration (IM), including previously published (Tine *et al.*, 2014; Filatov *et al.*, 2016) and newly implemented models (Table 3), in the program DADI v 1.7 (Gutenkunst *et al.*, 2009). Given the large geographic distance between sampling localities of *L. alopecuroides* (Figure 1) and that only 2 individuals of these species were sampled from the two most southern localities, we excluded these two individuals from demographic inference. In the analysis, we used only SNPs that showed no sign of being under selection. To identify RAD loci putatively under selection we used BAYESCAN v 2.1 (Foll & Gaggiotti, 2008; Fischer *et al.*, 2011) running 50,000 iterations as burnin and 50,000 iterations thereafter. The demographic history reconstruction was based on the so-called “unfolded” SFS – that is, only the SNPs with known ancestral states were used. The ancestral state of each SNP was reconstructed in a comparison with up to six outgroups, all of them New World lupin species (*L. albifrons*, *L. huaronensis*, *L. luteolus*, *L. mutabilis*, *L. nanus* and *L. nevadensis*), for which low coverage genomic data are available (Nevado *et al.* *in prep*). Only SNPs with alleles scored in at least two outgroup species, and where all outgroup species were homozygous for one of the two alleles seen in the RAD loci, were considered further. To deal with missing data we “projected” the SFS down to 16 alleles (for ‘*Ocetá*’ – ‘*Pisba/Cocuy*’ analysis) and to 14 alleles (*L. alopecuroides* – *L. triananus*). This means that all SNPs with less than 14 (or 16) alleles scored in each species were removed, and SNPs with more than 14 (or 16) alleles were subsampled to

this number of alleles. To confirm the convergence of analyses in DADI, for each species pair we performed 20 runs from random starts using large search ranges for all parameters. After confirming that most runs returned similar results for all parameters, we performed another set of 20 runs for each pairwise comparison with shorter search ranges around the optimal values identified before. Python code of the IM models used in this study is available from <https://github.com/brunonevado/IM-models>.

Results

To test whether rapid diversification of the Andean *Lupinus* has been primarily driven by geographic isolation, and to understand the possible impacts of Pleistocene climatic oscillations on diversification in this group, we used two complementary analyses: i) analysis of phylogenetic incongruence across the phylogeny of Andean lupins and ii) demographic modelling of divergence events.

Analysis of phylogenetic incongruence

Analysis of phylogenetic incongruence relies on comparison of phylogenetic trees from multiple loci across the genome. To obtain such trees we analysed the 6,013 orthologous genes previously identified (Nevado *et al.*, 2016). After extracting all four-fold degenerate sites from each orthologous gene, and excluding genes with fewer than 10 Andean species and fewer than 4 phylogenetically-informative sites, we retained 867 genes for analysis of phylogenetic incongruence. The concordance factors (CFs) estimated from phylogenetic trees obtained from these genes were used to estimate phylogenetic networks with different numbers of hybridisation events (h), and to calculate their respective pseudo-likelihood. The pseudo-likelihood increased sharply from $h=0$ to $h=2$, while increasing the number of hybridisation events past 2 resulted in smaller improvements in pseudo-likelihood values (Figure 2). This suggests that the best fitting phylogenetic model involved two hybridisation events.

Analysis of the resulting phylogenetic network with $h=2$ (Figure 3) showed that this network is similar to the coalescent-based phylogeny reported previously (Nevado *et al.*, 2016), with the following differences: the sister-group relationship between the three most derived clades changed (the sister clade to clade6 is now clade4, instead of clade5 as before), species relationships between members of clade 4 changed, and the second accession of *L. misticola*, previously resolved within clade 4, is now resolved as basal in clade 6. In addition, the phylogenetic network obtained herein received lower bootstrap support than our previous phylogenetic reconstruction (Nevado *et al.* 2016). This is likely due to inclusion in this study of only four-fold degenerate sites from 867 genes, while our previous work used the entire coding sequences of all 6,013 genes. Reconstructing the phylogenetic history of very young and rapidly diversifying species is notoriously difficult, yet the high degree of congruence between the phylogenies obtained here and in our previous work supports our phylogenetic reconstruction, and suggests strong phylogenetic signal is present in the dataset.

One of the hybridisation events detected occurred within clade 5, suggesting that about 9% of genetic material in the ancestor of *L. bangii* S and *L. subacaulis* originated from the lineage leading to

L. misticola (Figure 3). This event was well supported in the bootstrap analysis, with 85% of bootstrap replicates identifying the same donor and recipient clades, with a mean contribution of genetic material into the ancestor of *L. bangii* and *L. subacaulis* of 6.7% (max=18.7, min=3.8, sd= 2.1).

The second hybridisation event that was detected involved members of clades 3 and 4, with the ancestor of *L. semperflorens* and *L. ramosissimus* resolved as a hybrid clade with approximately equal contributions from both lineages (Figure 3). This hybridisation event was not well supported in the bootstrap analysis, with only 8% of bootstraps recovering the same two clades as being involved in the hybridisation event. Despite this uncertainty over the exact placement of the hybridization event of the tree, the clade consisting of *L. semperflorens* and *L. ramosissimus* was identified as of hybrid origin in 69% of the bootstrap replicates, although the contributing lineages differed between replicates: the most frequently found clades involved in this hybridisation were *L. ballianus* (44% of replicates), the ancestor of *L. ballianus* and *L. praestabilis* (20% of replicates), *Lupinus* sp CEH3079A1 (17%) and the ancestor of *Lupinus* sp CEH3079A1, *L. bangii* N and *L. mantaroensis* (16%).

To further test whether the evolutionary history of Andean lupins is better represented by a phylogenetic network or a phylogenetic tree, we used the Tree Incongruence Checking in R (TICR) method. This test uses the CFs observed across the entire Andean clade, and revealed a significant excess of outlier quartets with highly significant values ($Pval = 1.316508e-08$), showing that the observed CFs are very unlikely to have arisen under incomplete lineage sorting alone.

Demographic modelling of divergence events

To reconstruct demographic history of speciation in the Andean *Lupinus* we compiled population-level RAD-seq datasets for two pairs of sister lineages from the northern Andes (Contreras-Ortiz *et al.*, 2018). Pre-processing of raw illumina reads resulted in a very low percentage of reads lost due to low quality (average 0.6%, maximum 1.37%). Using the draft genome of *L. mutabilis* as the reference, the percentage of reads mapped averaged 88%, with only 2 samples having less than 80% of reads mapped. The STACKS pipeline identified 22,936 variable loci present in at least 2 species, and 5,435–11,689 polymorphic sites in each lineage analysed (Table 1). Nucleotide diversity per site within each lineage, estimated by STACKS across all loci sequenced (Hohenlohe *et al.*, 2010), was between 0.0004–0.0007 (Table 1).

The analysis of population structure with FASTSTRUCTURE revealed that the number of clusters that maximized the marginal likelihood was 1 for *Pisba/Cocuy* and *Ocetá*, 2 for *L. trianaus* and 3 for *L. alopecuroides* (Supporting Information Fig. S1). Analysis of Molecular Variance (AMOVA) accordingly revealed that the percentage of variation between localities was higher in *L. alopecuroides* and *L. trianaus* compared to *Pisba/Cocuy* and *Ocetá* (Table 2). Using low-coverage re-sequencing data from six other Andean lupins, we confirmed the ancestral allele in 15,869 of the 22,936 SNPs obtained with STACKS. BAYESCAN analysis of selection identified 60 loci with significant traces of

selection (FDR q-value < 0.05), with all loci exhibiting low F_{st} and negative alpha parameter, indicative of purifying or balancing selection. We removed these 60 loci before demographic modelling in DADI.

To infer the most likely demographic history of divergence for each lineage pair we investigated 12 models (Table 3, Figure 4) chosen to represent a range of plausible and relevant demographic scenarios underpinning divergence in Andean *Lupinus*, and particularly the extent and timing of gene flow between diverging lineages. All scenarios are based on the Isolation with Migration model, in which two populations split at some point in the past (T) after which gene flow between populations (M_{12} and M_{21}) is restricted. The key difference between the scenarios is how migration between diverging populations is modelled. For instance, in the simplest model (Isolation) gene flow between diverging species does not occur after splitting ($M_{12}=M_{21}=0$). More complex models allow for asymmetric migration between populations ($M_{12}\neq M_{21}$, e.g. IM or IMpre models), symmetric but time-dependent migration between populations (e.g. splitExpMig, eSplitExpMig), migration occurring at different rates in different loci (M_A and M_B , e.g. IM2M) or times (e.g. ancestral migration [AM], secondary contact [SC]), or a combination of these (Table 3, Figure 4). In addition, some models allow for population size changes during and after splitting (e.g. IM, IMpre), while others assume constant population size throughout (e.g. AM, SC).

Maximum-likelihood parameter estimates for each model, and their respective likelihoods, are given in Table 4. Because not all models are nested, we first used the Akaike's Information Criterion (AIC) to rank the different models. This approach assumes that the different SNPs are unlinked, which is a reasonable assumption given we used a single SNP per RAD locus. The best-fitting demographic model for the split between 'Ocetá' and 'Pisba/Cocuy' was the IM2MSC, and for *L. alopecuroides* and *L. triananus* it was the IM2M (Table 4). In both cases all alternative models received considerably less support (Delta-AIC > 3). The best fitting models resulted in a reasonable fit to the observed 2D-SFS in both lineage pairs (Figure 5), although there was a slight deficit of high frequency variants in *L. triananus* and medium to high frequency in *L. alopecuroides*. For both divergence events investigated, demographic modelling supports a scenario of post-divergence gene flow with two classes of sites with very different amount of gene flow (Table 4).

For the split between *Ocetá* and *Pisba/Cocuy* the best fitting model was the IM2MSC, in which gene flow occurs after an initial period of complete isolation. To formally test that a period of secondary contact results in a significantly better fit than a model with constant gene flow after initial divergence, we performed a Likelihood Ratio Test (LRT) between the full IM2MSC model and a simpler, nested model. In the simpler model, the time period following divergence, during which there is no gene flow (T_A in Figure 4), was fixed at 0, and LRT between models was performed assuming 0.5 degrees of freedom. Given that DADI uses composite likelihoods to approximate the full likelihood of each model, LRT p-values were adjusted using the Godambe function (Coffman *et al.*, 2016), with the more complex IM2MSC being significantly supported ($P=0.023$). Similarly, a simpler nested model in which

migration occurred homogeneously across the genome (i.e. with M_A fixed to 0, Figure 4) could also be rejected in favour of the full IM2MSC model ($P=0.022$).

For the split between *L. alopecuroides* and *L. triananus* the best fitting model was IM2M. This scenario implies gene flow between species after initial split, but without an initial period of complete isolation. Nevertheless, demographic models including periods of secondary contact usually provided a better fit than comparable models assuming older migration, in particular: SC vs AM models (logLik=-497.80 vs -505.78 respectively) and IM2MSC vs IM2MAM (logLik=-312.73 vs -327.52). A LRT between the full IM2M model and a simpler model assuming a single migration value for all sites (M_A fixed to 0, performed as described above) provided significant support for the more complex model ($P=0.004$), confirming that gene flow between species was heterogeneous across the genome.

Having found the best-fitting demographic model for each split, it is also possible to estimate the absolute times of divergence in each case. The split times estimated in DADI are in units of $2*N_A$ generations, with N_A representing the effective population size of the ancestral population before split. We also know, from the parameter estimates of the best fitting model in each species, the relative sizes of current populations compared to their ancestral population sizes (N_1 and N_2 in Table 4). Thus, to obtain absolute estimates for speciation times we need estimates of current population sizes and generation times of each species. If we use a mutation rate of 7.0×10^{-9} per generation per site, which is similar to the available estimates in plants (*Arabidopsis thaliana* (Ossowski *et al.*, 2010) and *Prunus* (Xie *et al.*, 2016)), to convert the genome-wide estimates of nucleotide diversity of each species (Table 1) into effective population sizes, we get estimates of current $N_e=14,285$ (*L. alopecuroides*), $N_e=21,428$ (*Pisba/Cocuy*) and $N_e = 25,000$ (*Ocetá* and *L. triananus*). From these current population sizes, we can estimate the ancestral size of each species pair using the relative population size parameters (N_1 and N_2 in Table 4), which gives an average $N_A=15,458$ (*L. alopecuroides* and *L. triananus*) and $N_A=7,673$ (*Ocetá* and *Pisba/Cocuy*). We can now convert the split times (Table 4) into number of generations using the estimated ancestral population sizes: for *L. alopecuroides* and *L. triananus* this would be $2 \times N_A \times T = 2 \times 15,458 \times 1.725 = 53,330$ generations ago; and for *Ocetá* and *Pisba/Cocuy* $2 \times N_A \times (T_A + T_S) = 2 \times 7,673 \times (0.72+0.95) = 25,628$ generations ago. Definitive information about generation times for these species is lacking, but these species are certainly not annuals and are likely to start flowering only in the second year (N. Contreras, pers. obs.). Using a two-year generation time we can thus estimate that *L. alopecuroides* and *L. triananus* started to diverge around 106 thousand years ago (kya), and *Ocetá* and *Pisba/Cocuy* around 51 kya. With the same assumptions, secondary contact between *Ocetá* and *Pisba/Cocuy* can be dated to around 29 kya.

Discussion

While it has been suggested that geographic isolation may be the predominant driver of rapid diversification in Andean *Lupinus* (Givnish, 2015), our results suggest that gene flow after speciation may be common in this system. We show that interspecific gene flow resulted in clades of hybrid origin,

that gene flow occurred in the recent evolutionary past (following recent divergence events c. 51-106 Kya), and that gene flow has occurred at disparate localities across Andes (Figure 1). Furthermore, our results are minimum estimates for the number of hybridisation events occurring during diversification of Andean *Lupinus*. In fact, our taxonomic sampling included only 25 species, i.e. <30% of the estimated 85 species of Andean lupins. Increasing taxonomic sampling may lead to detection of additional cases of hybridisation. Furthermore, the analysis of phylogenetic incongruence with SNaQ can only detect a restricted subset of hybridisation events (Solís-Lemus & Ané, 2016): hybridisation events between sister lineages can only be detected if both lineages have at least 2 descendant species; and only level-1 networks are considered, i.e. each hybrid node can only be part of a single hybridisation event. As such, the two hybridisation events identified by analysis of phylogenetic incongruence are most likely an underestimate of the true amount of gene flow among Andean lupins. In this regard, it is noteworthy that our demographic modelling detected recent episodes of post-split gene flow associated with both lineage pairs analysed. These results are in line with recent studies suggesting that gene flow may have occurred between currently geographically isolated Andean cordilleras (Kolar *et al.*, 2016) and that hybridization may have played an important role in other Andean plant radiations (Duskova *et al.*, 2017; Vargas *et al.*, 2017; Pouchon *et al.*, 2018).

Hybridisation in Andean Lupinus

The finding that hybridisation might be common in Andean *Lupinus* is surprising because it is likely that both pollen and seed long-distance dispersal are limited. Although there are no data on seed dispersal and few observations of pollinators for Andean *Lupinus*, the dispersal mechanism in almost all species in the genus is ballistic dispersal – the seed pods dry and explosively dehisce, ejecting the seeds at most a few meters away from the parent plants (e.g. 0-4 m for the North American species *L. texensis* (Turner *et al.*, 2017). Furthermore, the seeds of wild Andean lupins are relatively large and smooth and not easily transported by wind, while pollination is typically dependent on bees (Langridge & Goodman, 1977), which are also infrequent in the high-elevation Andes.

In addition to low dispersal ability, many Andean lupins are narrow endemics, or exhibit larger distribution ranges but consist of multiple small isolated populations. For instance, *L. alopecuroides* is known from only a handful of isolated mountain tops (mainly above 4,000m) in Colombia and Ecuador (Figure 1), with microsatellite data showing a high degree of differentiation between populations (Vásquez *et al.*, 2016). Lower altitude valleys separating these populations are likely effective barriers to gene flow. Yet, our demographic modelling shows that gene flow occurred between populations of *L. alopecuroides* and *L. trianaus*, which are currently isolated on the central and eastern Cordilleras separated by the deep Magdalena valley in Colombia (Figure 1).

Gene flow between species could result from speciation with ongoing gene flow, or secondary contact after allopatric speciation. Analysis of the fit of alternative demographic models supports the scenario of secondary contact for the divergence between *Oceta* and *Pisba/Cocuy*, suggesting that

divergence was originally driven by geographic isolation. For *L. alopecuroides* and *L. trianaus* the best fitting model (IM2M) includes a single period of gene flow post speciation, and although this suggests continuous gene flow since divergence it could also reflect the difficulties in discerning the timing of gene flow after speciation.

A role for natural selection in diversification in Andean Lupinus

It has been suggested that the exceptionally high speciation rates estimated for recent evolutionary radiations is evidence for an ephemeral speciation model whereby speciation is very frequent, but most species rapidly go extinct or are reabsorbed into parental forms (Rosenblum *et al.*, 2012). This is a possible scenario for the high elevation *Lupinus* species radiation, whereby many recognized extant species could be viewed as ephemeral. However, in the case of Andean *Lupinus*, our results suggest that even in the face of geneflow precipitated by secondary contact species are not immediately reabsorbed into parental forms. Our results show that, in both pairs investigated, gene flow occurs at two different rates, with some sites introgressing freely while others introgress at very low rates (Table 4). The proportion of sites that freely introgress between lineages is high in both cases: 43% of sites in *L. alopecuroides* and *L. trianaus*, and 14% of sites in *Ocetá* and *Pisba/Cocuy*. Yet, a large part of the genome appears to be protected from introgression between species.

Heterogeneous gene flow across the genome can be explained in three different ways. First, regions that do not introgress may harbour genetic incompatibilities, which can cause hybrid inviability or sterility (Coyne & Orr, 2004). However, hybrid sterility between plant lineages accrues over much longer time scales than considered here, with most plant species studied to date showing little to no hybrid sterility within 1 Myr of divergence (Levin, 2012). Second, regions that experience no migration between species may harbour loci responsible for species-specific adaptations, with selection preventing gene flow in linked regions of the genome and leading to “genomic islands of divergence” (Wu, 2001; Nosil *et al.*, 2009). This scenario has been extensively studied in recent years (e.g. Renaut *et al.*, 2013; Marques *et al.*, 2016; Han *et al.*, 2017; Larson *et al.*, 2017; Vijay *et al.*, 2017), however it remains unclear in which cases the observed heterogeneity is due to selection, drift, or variation of recombination rate across the genome (Noor & Bennett, 2009; Cruickshank & Hahn, 2014; Ravinet *et al.*, 2017). Third, heterogeneity in gene flow may result from selection acting on introgressed genes, possibly promoting introgression of advantageous genes (e.g. Staubach *et al.*, 2012; Dannemann *et al.*, 2016; Bechsgaard *et al.*, 2017).

Diversification of Andean *Lupinus* was previously found to be associated with increased rates of positive selection genome-wide (Nevado *et al.*, 2016). Our results suggest this widespread positive selection may play a role in the large species diversity of Andean *Lupinus*: without a sizeable number of genes evolving species-specific adaptations, gene flow between species would not be impeded, and the gene pools of both species could merge (Seehausen, 2006; Taylor *et al.*, 2006). Thus, while geographical isolation may initiate speciation, species differentiation in the face of changing

environmental conditions is only achieved once positive selection drives fixation of different adaptive alleles in each species. This form of ecological speciation in geographic isolation is likely driven by the extreme heterogeneity of the Andes: in addition to the typical changes in temperature and insolation in mountains, the Andes are notable for dramatic differences also at smaller scales, e.g. within the same mountain peak the slopes facing the inter-Andean valleys can be much drier than those facing the Amazonian forest or the Pacific Ocean (van der Hammen & Cleef, 1986). In the case of the species studied herein, *L. alopecuroides* and *L. trianaus* occupy slightly different elevation ranges: 3600–4700m 3300–3700m respectively – suggesting possible ecological divergence related to altitude. Morphologically, the floral bracts of *L. alopecuroides* are very long, strongly exerted and persistent such that they largely conceal the flowers even at full anthesis, while in *L. trianaus* the bracts are smaller and wither quickly, and do not conceal the flowers at full anthesis. It is likely that the long bracts of *L. alopecuroides* insulate the flowers from the daily freeze-thaw regime that prevails at very high elevations in the Andes. Whether these habitat and ecological differences are indeed responsible for divergence of these two species remains to be tested.

Isolation and gene flow driven by Pleistocene climatic oscillations

That *Lupinus* species were already present in the Andes by 2.25 Myr is evident from the pollen record ((Hooghiemstra, 1983), with the age model of (Torres *et al.*, 2013)), in line with molecular clock estimates of 1.19–3.5 Myr for the crown node of the Andean clade (Hughes & Eastwood, 2006; Drummond *et al.*, 2012). However, how many and which species were present at that time is unknown. It is clear that many present-day species must have diverged much more recently, and in agreement with this scenario our demographic modelling suggests very young divergence times in both cases: c. 106 Kya (*L. alopecuroides* and *L. trianaus*) and 51 Kyr (*Oceta* and *Pisba/Cocuy*). While these ages must be interpreted with caution given the large uncertainty involved, they suggest speciation occurred long after the final uplift of the Andes and emergence of the Páramo, and is more likely to have resulted from climatic oscillations in the Late Pleistocene. In this case, the exceptionally fast speciation rates in Andean *Lupinus* might in part result from cyclic climatic oscillations that periodically change species' suitable habitat, promoting continuous cycles of expansion and contraction coupled with geographic isolation and secondary contact of species – a mechanism sometimes referred to as the 'species pump'.

Given their low dispersal abilities, gene flow between *Lupinus* species currently isolated by large areas of unsuitable habitat is likely to have been brought about by environmental changes that affected connectivity of the high elevation Andean grassland habitats. In this regard, it is notable that the Last Glacial Maximum in the Andes occurred significantly earlier than in temperate zones, at c. 34 Kyr (Smith *et al.*, 2005). At this time, with the treeline around 2100 m, there would have been maximal connectivity across the north Andean high elevation grassland habitats, including numerous closely adjacent stepping stone areas of grasslands linking the central and eastern Cordilleras in Colombia (see

Fig. 12.3a in Flantua & Hooghiemstra, 2018), and potentially coinciding with the estimated period of secondary contact between *Oceta* and *Pisba/Cocuy* around 29 Kya.

While a role for Late Pleistocene glacial cycles in species diversification has long been suggested (Haffer, 1969; Simpson, 1974), the impacts of Pleistocene glacial cycles in different environments seems highly variable (Haffer, 1969; Hewitt, 1996; Klicka & Zink, 1997; Rull, 2011; Weir *et al.*, 2016). Diversification caused by Pleistocene sea level fluctuations on oceanic island archipelagos appear to be mainly associated with infraspecific divergence (Papadopoulou & Knowles, 2015a; Papadopoulou & Knowles, 2015b). However, in other settings there is growing evidence for ice age or post-glacial speciation across diverse organismal groups, including fish in post-glacial lakes (e.g. Hudson *et al.*, 2011), birds in boreal North America (e.g. Weir & Schluter, 2004) and temperate coastal regions of New Zealand (e.g. Weir *et al.*, 2016), grasshoppers in western North American montane sky islands in the Rockies (Knowles, 2000), and plants in the high Arctic (Brochmann *et al.*, 2004). Here we add to this growing tally of examples of recent Late Pleistocene speciation, and suggest an important role for Pleistocene glacial cycles in driving the very high rates of speciation associated with the Andean *Lupinus* radiation.

Acknowledgments

This work was supported by funds from the Natural Environment Research Council, UK (Grant NE/K004352/1 to DAF) and from the Swiss National Science Foundation (Grants 31003A_135522 and 31003A_156140 to CH), and the Claraz-Schenkung Foundation, Switzerland, Stanley Smith Horticultural Trust and the Royal Society, U.K. for fieldwork in the Andes. We acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work (<http://dx.doi.org/10.5281/zenodo.22558>). We thank the United States Department of Agriculture and the Desert Legume Program (University of Arizona) for seedlots, the national authorities in Colombia, Ecuador, Peru, Bolivia and Argentina for permission to collect plant material and Petr Sklenář for collecting and contributing plant material from Ecuador. We thank Santiago Madriñán and Guy Atchison for help with sampling, and Varvara Fazalova for comments on an earlier version of this manuscript.

Author contributions

BN and DAF designed the research. BN and NCO generated the data. BN performed the research. BN, CH and DAF wrote the manuscript. All authors have read and agree with the manuscript.

References

- Andrews S 2010. FastqQC: A quality control tool for high throughput sequence data.
- Bechsgaard J, Jorgensen TH, Schierup MH. 2017. Evidence for Adaptive Introgression of Disease Resistance Genes Among Closely Related Arabidopsis Species. *G3 (Bethesda)* 7(8): 2677-2683.

536 **Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen AC, Elven R. 2004.**
537 Polyploidy in arctic plants. *Biological Journal of the Linnean Society* **82**(4): 521-536.

538 **Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013.** Stacks: an analysis tool set
539 for population genomics. *Molecular ecology* **22**(11): 3124-3140.

540 **Chapman MA, Hiscock SJ, Filatov DA. 2013.** Genomic Divergence during Speciation Driven by
541 Adaptation to Altitude. *Molecular biology and evolution* **30**(12): 2553-2567.

542 **Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN. 2016.** Computationally Efficient Composite
543 Likelihood Statistics for Demographic Inference. *Molecular biology and evolution* **33**(2): 591-593.

544 **Conterato IF, Schifino-Wittmann MT. 2006.** New chromosome numbers, meiotic behaviour and
545 pollen fertility in American taxa of *Lupinus* (Leguminosae): contributions to taxonomic and
546 evolutionary studies. *Botanical Journal of the Linnean Society* **150**(2): 229-240.

547 **Contreras-Ortiz N, Atchison GW, Hughes CE, Madriñán S. 2018.** Convergent evolution of high
548 elevation plant growth forms and geographically structured variation in Andean *Lupinus*
549 (Leguminosae). *Biological Journal of the Linnean Society* **In press**.

550 **Coyne JA, Orr HA. 2004.** *Speciation*: Sinauer.

551 **Cruikshank TE, Hahn MW. 2014.** Reanalysis suggests that genomic islands of speciation are due
552 to reduced diversity, not reduced gene flow. *Molecular ecology* **23**(13): 3133-3157.

553 **Dannemann M, Andres AM, Kelso J. 2016.** Introgression of Neandertal- and Denisovan-like
554 Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *The American Journal*
555 *of Human Genetics* **98**(1): 22-33.

556 **Drummond CS. 2008.** Diversification of *Lupinus* (Leguminosae) in the western New World: derived
557 evolution of perennial life history and colonization of montane habitats. *Molecular phylogenetics and*
558 *evolution* **48**(2): 408-421.

559 **Drummond CS, Eastwood RJ, Miotto ST, Hughes CE. 2012.** Multiple continental radiations and
560 correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete
561 taxon sampling. *Systematic biology* **61**(3): 443-460.

562 **Duskova E, Sklenar P, Kolar F, Vasquez DLA, Romoleroux K, Fer T, Marhold K. 2017.** Growth
563 form evolution and hybridization in *Senecio* (Asteraceae) from the high equatorial Andes. *Ecology*
564 *and Evolution* **7**(16): 6455-6468.

565 **Excoffier L, Lischer HE. 2010.** Arlequin suite ver 3.5: a new series of programs to perform
566 population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**(3): 564-
567 567.

568 **Filatov DA, Osborne OG, Papadopoulos AS. 2016.** Demographic history of speciation in a *Senecio*
569 altitudinal hybrid zone on Mt. Etna. *Molecular ecology* **25**(11): 2467-2481.

570 **Fischer MC, Foll M, Excoffier L, Heckel G. 2011.** Enhanced AFLP genome scans detect local
571 adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular ecology* **20**(7):
572 1450-1462.

573 **Flantua S, Hooghiemstra H 2018.** Historical connectivity and mountain biodiversity. In: Hoorn C,
574 Perrigo A, Antonelli A eds. *Mountains, Climate and Biodiversity*: John Wiley.

575 **Flantua S, Hooghiemstra H, Van Boxel JH, Cabrera M, Gonzalez-Carranza Z, Gonzalez-**
576 **Arango C 2014.** Connectivity dynamics since the Last Glacial Maximum in the northern Andes: a
577 pollen-driven framework to assess potential migration. In: Stevens WD, Montiel OM, Raven P eds.
578 *Paleobotany and Biogeography: a Festschrift for Alan Graham in his 80th Year*. St. Louis: Missouri
579 Botanical Garden Press, 98-123.

580 **Foll M, Gaggiotti O. 2008.** A genome-scan method to identify selected loci appropriate for both
581 dominant and codominant markers: a Bayesian perspective. *Genetics* **180**(2): 977-993.

582 **Gavin DG, Fitzpatrick MC, Gugger PF, Heath KD, Rodriguez-Sanchez F, Dobrowski SZ,**
583 **Hampe A, Hu FS, Ashcroft MB, Bartlein PJ, et al. 2014.** Climate refugia: joint inference from
584 fossil records, species distribution models and phylogeography. *New Phytologist* **204**(1): 37-54.

585 **Ghosh P, Garzione CN, Eiler JM. 2006.** Rapid uplift of the Altiplano revealed through 13C-18O
586 bonds in paleosol carbonates. *Science* **311**(5760): 511-515.

587 **Givnish TJ. 2015.** Adaptive radiation versus 'radiation' and 'explosive diversification': why
588 conceptual distinctions are fundamental to understanding evolution. *New Phytologist* **207**(2): 297-303.

589 **Graham A. 2009.** The Andes: a geological overview from a biological perspective. *Annals of the*
590 *Missouri Botanical Garden* **96**(3): 371-385.

591 **Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009.** Inferring the joint
592 demographic history of multiple populations from multidimensional SNP frequency data. *PLoS*
593 *genetics* **5**(10).

594 **Haffer J. 1969.** Speciation in amazonian forest birds. *Science* **165**(3889): 131-137.

595 **Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. 2017.** Gene flow,
596 ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among
597 Darwin's finches. *Genome Research* **27**(6): 1004-1015.

598 **Hewitt GM. 1996.** Some genetic consequences of ice ages, and their role in divergence and
599 speciation. *Biological Journal of the Linnean Society* **58**: 247-276.

600 **Hewitt GM. 2004.** Genetic consequences of climatic oscillations in the Quaternary. *Philosophical*
601 *transactions of the Royal Society of London. Series B, Biological sciences* **359**(1442): 183.

602 **Ho Ls, Ané C. 2014.** A linear-time algorithm for Gaussian and non-Gaussian trait evolution models.
603 *Systematic biology* **63**(3): 397-408.

604 **Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010.** Population
605 genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*
606 **6**(2): e1000862.

607 **Hooghiemstra H. 1983.** *Vegetational and climatic history of the Bogotá high plain, Colombia*. PhD
608 Phd thesis, University of Amsterdam Amsterdam.

609 **Hooghiemstra H, Van der Hammen T. 2004.** Quaternary Ice-Age dynamics in the Colombian
610 Andes: developing an understanding of our legacy. *Philosophical transactions of the Royal Society of*
611 *London. Series B, Biological sciences* **359**(1442): 173-181.

612 **Hooghiemstra H, Wijninga VM, Cleef AM. 2006.** The paleobotanical record of Colombia:
613 implications for biogeography and biodiversity. *Annals of the Missouri Botanical Garden* **93**(2): 297-
614 325.

615 **Hudson AG, Vonlanthen P, Seehausen O. 2011.** Rapid parallel adaptive radiations from a single
616 hybridogenic ancestral population. *Proceedings of the Royal Society B* **278**(1702): 58-66.

617 **Hughes C, Eastwood R. 2006.** Island radiation on a continental scale: exceptional rates of plant
618 diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the*
619 *United States of America* **103**(27): 10334-10339.

620 **Hughes CE. 2017.** Are there many different routes to becoming a global biodiversity hotspot?
621 *Proceedings of the National Academy of Sciences of the United States of America* **114**: 4275-4277.

622 **Hughes CE, Atchison GW. 2015.** The ubiquity of alpine plant radiations: from the Andes to the
623 Hengduan Mountains. *New Phytologist*.

624 **Jomelli V, Favier V, Vuille M, Braucher R, Martin L, Blard PH, Colose C, Brunstein D, He F,**
625 **Khodri M, et al. 2014.** A major advance of tropical Andean glaciers during the Antarctic cold
626 reversal. *Nature* **513**: 224-228.

627 **Kadereit JW. 2017.** The role of in situ species diversification for the evolution of high vascular plant
628 species diversity in the European Alps—A review and interpretation of phylogenetic studies of the
629 endemic flora of the Alps. *Perspectives in Plant Ecology, Evolution and Systematics* **26**: 28-38.

630 **Klicka J, Zink RM. 1997.** The Importance of Recent Ice Ages in Speciation: A Failed Paradigm.
631 *Science* **277**(5332): 1666-1669.

632 **Knowles LL. 2000.** Tests of pleistocene speciation in montane grasshoppers (genus *Melanoplus*)
633 from the sky islands of western North America. *Evolution* **54**(4): 1337-1348.

634 **Kolar F, Duskova E, Sklenar P. 2016.** Niche shifts and range expansions along cordilleras drove
635 diversification in a high-elevation endemic plant genus in the tropical Andes. *Molecular ecology*
636 **25**(18): 4593-4610.

637 **Kroonenberg SB, Bakker JGM, van der Wiel AM. 1990.** Late Cenozoic uplift and paleogeography
638 of the Colombian Andes: constraints on the development of high-Andean biota. *Geologie*
639 *en Mijnbouw* **69**: 279-290.

640 **Langridge D, Goodman R. 1977.** A study on pollination of lupins (*Lupinus angustifolius*).
641 *Australian Journal of Experimental Agriculture* **17**(85): 319-322.

642 **Larson WA, Limborg MT, McKinney GJ, Schindler DE, Seeb JE, Seeb LW. 2017.** Genomic
643 islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular ecology* **26**(2): 554-
644 570.

645 **Levin DA. 2012.** The long wait for hybrid sterility in flowering plants. *New Phytologist* **196**(3): 666-
646 670.

647 **Li H, Durbin R. 2010.** Fast and accurate long-read alignment with Burrows-Wheeler transform.
648 *Bioinformatics* **26**(5): 589-595.

649 **Linder HP. 2008.** Plant species radiations: where, when, why? *Philosophical transactions of the*
650 *Royal Society of London. Series B, Biological sciences* **363**(1506): 3097-3105.

651 **Luebert F, Weigend M. 2014.** Phylogenetic insights into Andean plant diversification. *Frontiers in*
652 *Ecology and Evolution* **2**(27).

653 **Madriñán S, Cortés AJ, Richardson JE. 2013.** Páramo is the world's fastest evolving and coolest
654 biodiversity hotspot. *Frontiers in genetics* **4**: 192.

655 **Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2016.**
656 Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLoS genetics* **12**(2):
657 e1005887.

658 **Martin M. 2011.** Cutadapt removes adapter sequences from high-throughput sequencing reads.
659 *EMBnet.journal* **17**(1): 10-12.

660 **Mirarab S, Warnow T. 2015.** ASTRAL-II: coalescent-based species tree estimation with many
661 hundreds of taxa and thousands of genes. *Bioinformatics* **31**(12): 52.

662 **Mittermeier R, Gil PR, Hoffmann M, Pilgrim J, Brooks T, Mittermeier CG, Lamoreux J,**
663 **Fonseca G. 2005.** *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial*
664 *Ecoregions*. Chicago, IL: University of Chicago Press.

665 **Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. 2000.** Biodiversity hotspots
666 for conservation priorities. *Nature* **403**(6772): 853-858.

667 **Nevado B, Atchison GW, Hughes CE, Filatov DA. 2016.** Widespread adaptive evolution during
668 repeated evolutionary radiations in New World lupins. *Nature communications* **7**: 12384.

669 **Noor MA, Bennett SM. 2009.** Islands of speciation or mirages in the desert? Examining the role of
670 restricted recombination in maintaining species. *Heredity* **103**(6): 439-444.

671 **Nosil P, Funk DJ, Ortiz-Barrientos D. 2009.** Divergent selection and heterogeneous genomic
672 divergence. *Molecular ecology* **18**(3): 375-402.

673 **Olson DM, Dinerstein E. 2002.** The Global 200: Priority Ecoregions for Global Conservation.
674 *Annals of the Missouri Botanical Garden* **89**(2): 199.

675 **Orme CD, Davies RG, Burgess M, Eigenbrod F, Pickup N, Olson VA, Webster AJ, Ding T-SS,**
676 **Rasmussen PC, Ridgely RS, et al. 2005.** Global hotspots of species richness are not congruent with
677 endemism or threat. *Nature* **436**(7053): 1016-1019.

678 **Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D,**
679 **Lynch M. 2010.** The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*.
680 *Science* **327**(5961): 92-94.

681 **Papadopoulou A, Knowles LL. 2015a.** Genomic tests of the species-pump hypothesis: Recent island
682 connectivity cycles drive population divergence but not speciation in Caribbean crickets across the
683 Virgin Islands. *Evolution* **69**(6): 1501-1517.

684 **Papadopoulou A, Knowles LL. 2015b.** Species-specific responses to island connectivity cycles:
685 refined models for testing phylogeographic concordance across a Mediterranean Pleistocene
686 Aggregate Island Complex. *Molecular ecology* **24**(16): 4252-4268.

687 **Raj A, Stephens M, Pritchard JK. 2014.** fastSTRUCTURE: variational inference of population
688 structure in large SNP data sets. *Genetics* **197**(2): 573-589.

689 **Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlovic M, Noor MAF, Mehlig B,**
690 **Westram AM. 2017.** Interpreting the genomic landscape of speciation: a road map for finding
691 barriers to gene flow. *Journal of Evolutionary Biology* **30**(8): 1450-1477.

692 **Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM,**
693 **Rieseberg LH. 2013.** Genomic islands of divergence are not affected by geography of speciation in
694 sunflowers. *Nature communications* **4**: 1827.

695 **Rosenblum EB, Sarver BA, Brown JW, Des Roches S, Hardwick KM, Hether TD, Eastman JM,**
696 **Pennell MW, Harmon LJ. 2012.** Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model
697 Explains Patterns of Diversification Across Time Scales. *Evolutionary Biology* **39**(2): 255-261.

698 **Pouchon C, Fernández A, Nassar JM, Boyer F, Aubert S, Lavergne S, Mavárez J. 2018.**
699 Phylogenomic Analysis of the Explosive Adaptive Radiation of the Espeletia complex (Asteraceae) in
700 the Tropical Andes. *Systematic biology*: syy022-syy022.

701 **Ruiz-Sanchez E, Specht CD. 2014.** Ecological speciation in *Nolina parviflora* (Asparagaceae):
702 lacking spatial connectivity along of the Trans-Mexican Volcanic Belt. *PLoS One* **9**(6): e98754.

703 **Rull V. 2011.** Neotropical biodiversity: timing and potential drivers. *Trends in Ecology and Evolution*
704 **26**(10): 508-513.

705 **Seehausen O. 2006.** Conservation: losing biodiversity by reverse speciation. *Current Biology* **16**(9):
706 334-337.

707 **Simpson BB. 1974.** Glacial migrations of plants: island biogeographical evidence. *Science*
708 **185**(4152): 698-700.

709 **Simpson BB. 1975.** Pleistocene changes in the flora of the high tropical Andes. *Paleobiology* **1**(3):
710 273-294.

711 **Simpson GG. 1964.** Species Density of North American Recent Mammals. *Systematic Zoology*
712 **13**(2): 57-73.

713 **Sklenář P, Dušková E, Balslev H. 2011.** Tropical and temperate: evolutionary history of páramo
714 flora. *The Botanical Review* **77**: 71-108.

715 **Smith JA, Seltzer GO, Farber DL, Rodbell DT, Finkel RC. 2005.** Early local last glacial
716 maximum in the tropical Andes. *Science* **308**(5722): 678-681.

Solis-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS genetics* **12**.

Solis-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular biology and evolution* **34**: 3292-3298.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9): 1312-1313.

Staubach F, Lorenc A, Messer PW, Tang K, Petrov DA, Tautz D. 2012. Genome Patterns of Selection and Introgression of Haplotypes in Natural Populations of the House Mouse (*Mus musculus*). *PLoS genetics* **8**(8): e1002891.

Stenz NW, Larget B, Baum DA, Ané C. 2015. Exploring tree-like and non-tree-like patterns using genome sequences: an example using the inbreeding plant species *Arabidopsis thaliana* (l.) heynh. *Systematic biology* **64**(5): 809-823.

Taylor EB, Boughman JW, Groenenboom M, Sniatynski M, Schluter D, Gow JL. 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular ecology* **15**(2): 343-355.

Tine M, Kuhl H, Gagnaire P-AA, Louro B, Desmarais E, Martins RS, Hecht J, Knaust F, Belkhir K, Klages S, et al. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature communications* **5**: 5770.

Torres V, Hooghiemstra H, Lourens L, Tzedakis PC. 2013. Astronomical tuning of long pollen records reveals the dynamic history of montane biomes and lake levels in the tropical high Andes during the Quaternary. *Quaternary Science Reviews* **63**: 59-72.

Turner KG, Huang DI, Cronk QCB, Rieseberg LH. 2017. Homogenization of Populations in the Wildflower, Texas Bluebonnet (*Lupinus texensis*). *Journal of Heredity*: esx094-esx094.

Uribe-Convers S, Settles ML, Tank DC. 2016. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the south american species of *Bartsia* l. (Orobanchaceae). *PLoS One* **11**(2): e0148203.

van der Hammen T, Cleef AM 1986. Development of the high Andean Paramo flora and vegetation. In: Vuilleumier F, Monasterio M eds. *High Altitude Tropical Biogeography*. New York, USA: Oxford University Press.

Vargas OM, Ortiz EM, Simpson BB. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high Andean diversification (Asteraceae: Astereae: Diplostephium). *New Phytologist* **214**(4): 1736-1750.

Vásquez DLA, Balslev H, Hansen MM, Botany S-P. 2016. Low genetic variation and high differentiation across sky island populations of *Lupinus alopecuroides* (Fabaceae) in the northern Andes. *Alpine Botany* **126**: 135-142.

Vijay N, Weissensteiner M, Burri R, Kawakami T, Ellegren H, Wolf JBW. 2017. Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Molecular ecology* **26**(16): 4284-4295.

Weir JT, Haddrath O, Robertson HA, Colbourne RM, Baker AJ. 2016. Explosive ice age diversification of kiwi. *Proceedings of the National Academy of Sciences of the United States of America* **113**(38): E5580-5587.

Weir JT, Schluter D. 2004. Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society B* **271**(1551): 1881-1887.

Wu C-I. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**(6): 851-865.

Xie Z, Wang L, Wang L, Wang Z, Lu Z, Tian D, Yang S, Hurst LD. 2016. Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. A low rate in woody perennials and a higher mutagenicity in hybrids. *Proceedings of the Royal Society B* **283**(1841).

Xing Y, Ree RH. 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences of the United States of America* **114**(17): E3444-E3451.

Zech R, May J, Kull C, Ilgner J, Kubik PW, Veit H. 2008. Timing of the late Quaternary glaciation in the Andes from 15 to 40 S. *Journal of Quaternary Science* **23**(6 7): 635-647.

Figure legends

Figure 1: Map of South America showing the sampling locations of the *Lupinus* specimens used in this study. For the RNA dataset, species putatively involved in the two hybridisation events detected (see results) are denoted with yellow diamonds and squares, remaining species with blue triangles. Species putatively involved in hybridisation event 1: the ancestor of *L. sp. bangii* S and *L. subacaulis*; *L. misticola* and *L. sp. GA13*. Species putatively involved in hybridisation event 2: the ancestor of *L. semperflorens* and *L. ramosissimus*; *L. ballianus*; the ancestor of *L. mantaroensis*, *L. sp. bangii* N and *L. sp. CEH3079A1*. Map produced with R package MARMAP. Inset: Map of the northern Andes showing the geographically isolated distributions of *L. alopecuroides* and *L. trianaus* at high elevations on the Eastern and Central cordilleras, and the geographically more proximal Páramos of Cocuy-Pisba and Ocetá in Colombia for which detailed demographic analysis of gene flow were carried out. Data points are georeferenced herbarium specimen records. Map produced with QGIS.

Figure 2: Pseudo-likelihood profile with increasing number of hybridisation events allowed, obtained with the SNaQ pipeline.

Figure 3: Phylogenetic network estimated with the SNaQ pipeline with two hybridisation events. The blue edges denote the hybridisation events identified, with numbers next to edges denoting the proportion of genes that were transferred from each lineage. Bootstrap support for non-hybrid edges are plotted above corresponding edges (not shown when support is below 50). Clades are numbered as mentioned in the text.

Figure 4: Schematic representation of the demographic models investigated in this study. Models are named as in Table 3, and the free parameters in each model are depicted in *italics* in the respective figures. In each diagram, the top shows the two current population sizes (N_1 and N_2), and moving down (which is towards the past) the assumed demographic history since the split. Models that allow for exponential population size changes since the split have curved lines and include the parameter s , which is the relative size of population 1 just after split (relative size of population 2 just after split is $1-s$). Models that assume constant population size since split are represented with straight lines, and if the sum of the two current population sizes (N_1+N_2) is different from the ancestral size (N_A), then an instant population size change is assumed to occur during the split. For all models, N_A is the ancestral population size before split, and is not a free parameter. All population sizes (N_1 , N_2 and N_{PRE}) are expressed in units of N_A . All parameters related to time are given in units of $2*N_A$ generations. All migration parameters (M) are represented with horizontal arrows, and expressed in units of $2*N_A*m$, where m is the proportion of the receiving population consisting of immigrants in each generation. Models AM2M, SC2M, IM2M, IM2MAM and IM2MSC assume two classes of sites which experience different migration rates (M_A and M_B), and include an extra parameter (P) not shown in the figures, and denoting the proportion of sites that experience a migration rate of M_A (proportion of sites that experience M_B is $1-P$).

Figure 5: Fit of the best fitting demographic models for each species pair obtained with DADI: (A) the IM2MSC model for *Pisba-Cocuy* vs *Ocetá*, and (B) the IM2M model for *L. alopecuroides* vs *L. triananus*. Within each panel, top graphs show the two-dimensional Site-Frequency-Spectrum observed (left) and expected under the best-fitting model (right); bottom graphs show the residuals between the observed and the expected Site-Frequency-Spectra.

Figure S1: Result of clustering analysis of individuals of *L. triananus* and *L. alopecuroides*.

Table S1: Details of the *Lupinus* species used for phylogenetic analysis.

822 Table 1 – Number of polymorphic sites (S), and nucleotide diversity per site (Pi) estimated in STACKS
 823 using all RAD loci identified in each species.

# Pop ID	N	S	Pi
<i>L. alopecuroides</i>	14	5435	0.0004
<i>L. ocetá</i>	15	11689	0.0007
<i>L. pisba-cocuy</i>	15	11082	0.0006
<i>L. triananus</i>	17	10871	0.0007

824

825 Table 2 – Analysis of molecular variance showing the distribution of variance within each species.

# Pop ID	df	Sum of squares	Variance components	Percentage of variation
<i>L. alopecuroides</i> between localities	3	1133.075	52.15096	20.77
<i>L. alopecuroides</i> within localities	24	4773.318	198.88826	79.23
<i>L. ocetá</i> between localities	1	315.124	4.01617	1.32
<i>L. ocetá</i> within localities	28	8403.643	300.13010	98.68
<i>L. pisba-cocuy</i> between localities	1	364.933	15.20217	4.70
<i>L. pisba-cocuy</i> within localities	28	8629.000	308.17857	95.30
<i>L. triananus</i> between localities	3	1833.478	60.28913	17.00
<i>L. triananus</i> within localities	30	8830.375	294.34583	83.00

826

827 Table 3: The demographic models investigated in this study.

Model	Description	Source
<i>Isolation</i>	Model without migration after split.	Default model in dadi
<i>IM</i>	Isolation-with-migration model with asymmetric migrations and exponential population growth.	Default model in dadi
<i>IMpre</i>	Isolation-with-migration model with exponential population growth and a size change prior to split.	Default model in dadi
<i>splitExpMigr</i>	Constant population size and exponentially changing symmetric migration after split.	Filatov et al. 2016
<i>eSplitExpMigr</i>	Exponential population growth with exponentially changing symmetric migration after split.	This study, modified from Filatov et al. 2016
<i>AM</i>	Ancient migration model, asymmetric migration persists for some time after initial split, constant population sizes after split.	Tine et al. 2014
<i>SC</i>	Secondary contact model, no migration immediately after split but includes a recent period of asymmetric migration. Constant population sizes after split.	Tine et al. 2014
<i>AM2M</i>	Ancient migration model with two classes of loci experiencing different and asymmetric migration rates. Constant population sizes after split.	Tine et al. 2014
<i>SC2M</i>	Secondary contact model with two classes of loci experiencing different and asymmetric migration rates. Constant population size after split.	Tine et al. 2014
<i>IM2M</i>	The Isolation-with-migration model with two classes of sites experiencing different and symmetric migration rates. Exponential population growth after split.	This study, modified from Tine et al. 2014
<i>IM2MSC</i>	The isolation-with-migration model with two classes of sites experiencing different and symmetric migration rates and a period of no migration after split. Exponential population growth after split.	This study, modified from Tine et al. 2014
<i>IM2MAM</i>	The isolation-with-migration model with two classes of sites experiencing different and symmetric migration rates, migration persists for some time after split. Exponential population growth after split.	This study, modified from Tine et al. 2014

Table 4 – The likelihood of each demographic model investigated in DADI for each species pair, their respective parameters estimates, AIC scores, Δ -AIC scores (relative to best model), and the relative likelihood of each model compared to the best model. Parameter names as described in Fig. 4. Best-fit models (according to the AIC score) denoted in bold.

<i>L. osetá vs L. pisba-cocuy</i>						
Model	df	logLik	Parameters ML estimates	AIC	Δ -AIC	L(Mi y)
Isolation	4	-639.98	s=0.674, N1=0.899, N2=0.607, T=0.581	1287.97	530.62	0.00
IM	6	-477.00	s=0.911, N1=2.110, N2=1.417, T=3.074, M12=0.054, M21=0.096	965.99	208.65	0.00
splitExpMig	5	-500.22	N1=1.563, N2=0.923, T=2.845, MB=0.027, ME=0.085	1010.45	253.10	0.00
eSplitExpMig	6	-479.91	s=0.874, N1=2.828, N2=2.012, T=4.960, MB=0.073, ME=0.050	971.83	214.48	0.00
IMpre	8	-477.00	NPRe=0.025, TPRe=0.160, s=0.901, N1=0.054, N2=0.035, T=0.076, M12=2.14, M21=3.77	969.99	212.65	0.00
AM	6	-498.95	N1=2.344, N2=1.319, M12=0.047, M21=0.060, TA=4.744, TS=0.006	1009.90	252.55	0.00
SC	6	-497.87	N1=1.652, N2=0.922, M12=0.071, M21=0.092, TA=1.633, TS=1.161	1007.74	250.40	0.00
AM2M	9	-411.66	N1=1.191, N2=0.817, MA12=0.051, MA21=2.4e-10, MB12=5.822, MB21=3.5e-5, TA=2.104, TS=0.009, P=0.657	841.32	83.98	0.00
SC2M	9	-390.61	N1=1.994, N2=1.209, MA12=0.031, MA21=0.022, MB12=5.466, MB21=6.0e-7, TA=0.201, TS=4.010, P=0.680	799.22	41.88	0.00
IM2M	7	-374.26	s=0.911, N1=1.871, N2=1.362, T=2.841, MA=2.958, MB=0.042, P=0.196	762.52	5.17	0.08
IM2MSC	8	-370.67	s=0.66, N1=3.99, N2=2.36, TA=0.72, TS=0.95, MA=0.04, MB=8.25, P=0.86	757.34	0.00	1.00
IM2MAM	8	-414.522	s=0.661, N1=1.297, N2=0.775, TA=1.108, TB=1.04e-11, MA=8.89e-54, MB=3.104, P=0.807	845.04	87.70	0.00
<i>L. alopecuroides vs L. triananus</i>						
Model	df	logLik	Parameters ML estimates	AIC	Δ -AIC	L(Mi y)
Isolation	4	-610.32	s=0.053, N1=1.280, N2=1.294, T=0.271	1228.63	597.20	0.00
IM	6	-460.66	s=2.6e-4, N1=1.283, N2=2.912, T=2.106, M12=0.205, M21=0.089	933.32	301.89	0.00
splitExpMig	5	-500.64	N1=0.605, N2=2.298, T=2.215, MB=0.010, ME=0.191	1011.28	379.85	0.00
eSplitExpMig	6	-470.29	s=1.471, N1=3.631, N2=3.534, T=0.100, MB=0.111, ME=0.002	952.57	321.15	0.00
IMpre	8	-460.95	NPRe=1.945, TPRe=1.776, s=0.001, N1=1.769, N2=4.162, T=3.007, M12=0.126, M21=0.056	937.90	306.47	0.00
AM	6	-505.78	N1=0.743, N2=2.859, M12=0.116, M21=0.094, TA=3.341, TS=4.2e-4	1023.56	392.13	0.00
SC	6	-497.80	N1=0.579, N2=2.279, M12=0.179, M21=0.172, TA=1.542, TS=0.561	1007.59	376.16	0.00
AM2M	9	-350.19	N1=1.266, N2=4.845, MA12=0.007, MA21=0.035, MB12=3.6e-5, MB21=3.299, TA=8.610, TS=2.8e-4, P=0.408	718.38	86.95	0.00
SC2M	9	-350.46	N1=1.445, N2=5.486, MA12=0.007, MA21=0.033, MB12=1.0e-13, MB21=2.551, TA=4.670, TS=5.269, P=0.397	718.92	87.49	0.00
IM2M	7	-308.71	s=1.0e-4, N1=0.790, N2=1.948, T=1.725, MA=0.081, MB=5.978, P=0.573	631.43	0.00	1.00
IM2MSC	8	-312.73	s=0.005, N1=2.389, N2=2.535, TA=0.300, TS=1.036, MA=9.819, MB=0.077, P=0.475	641.46	10.03	0.01
IM2MAM	8	-327.52	s=0.054, N1=0.578, N2=2.593, TA=1.540, TS=2.7e-5, MA=0.061, MB=9.581, P=0.639	671.03	39.60	0.00

Figure 1

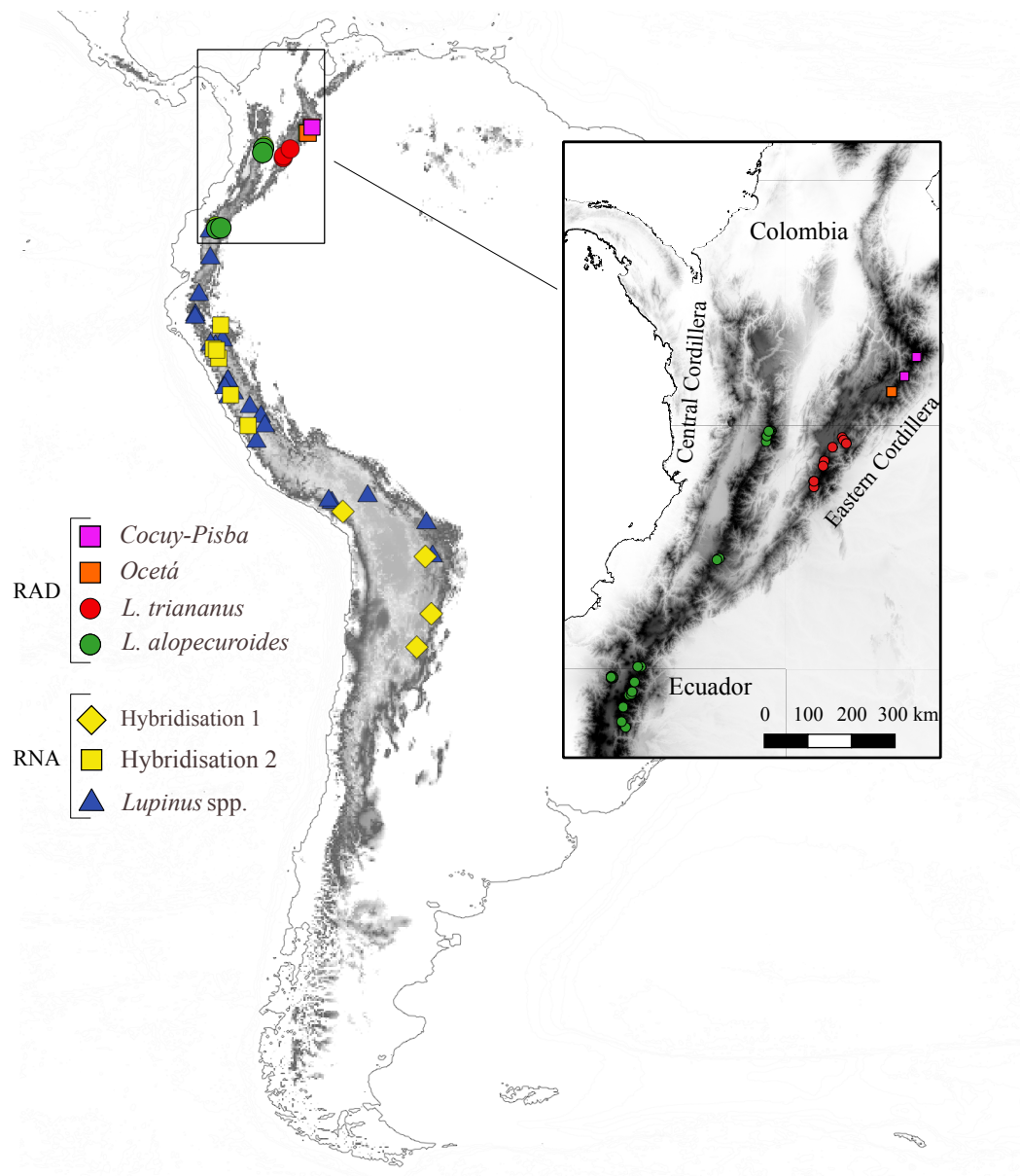


Figure 2

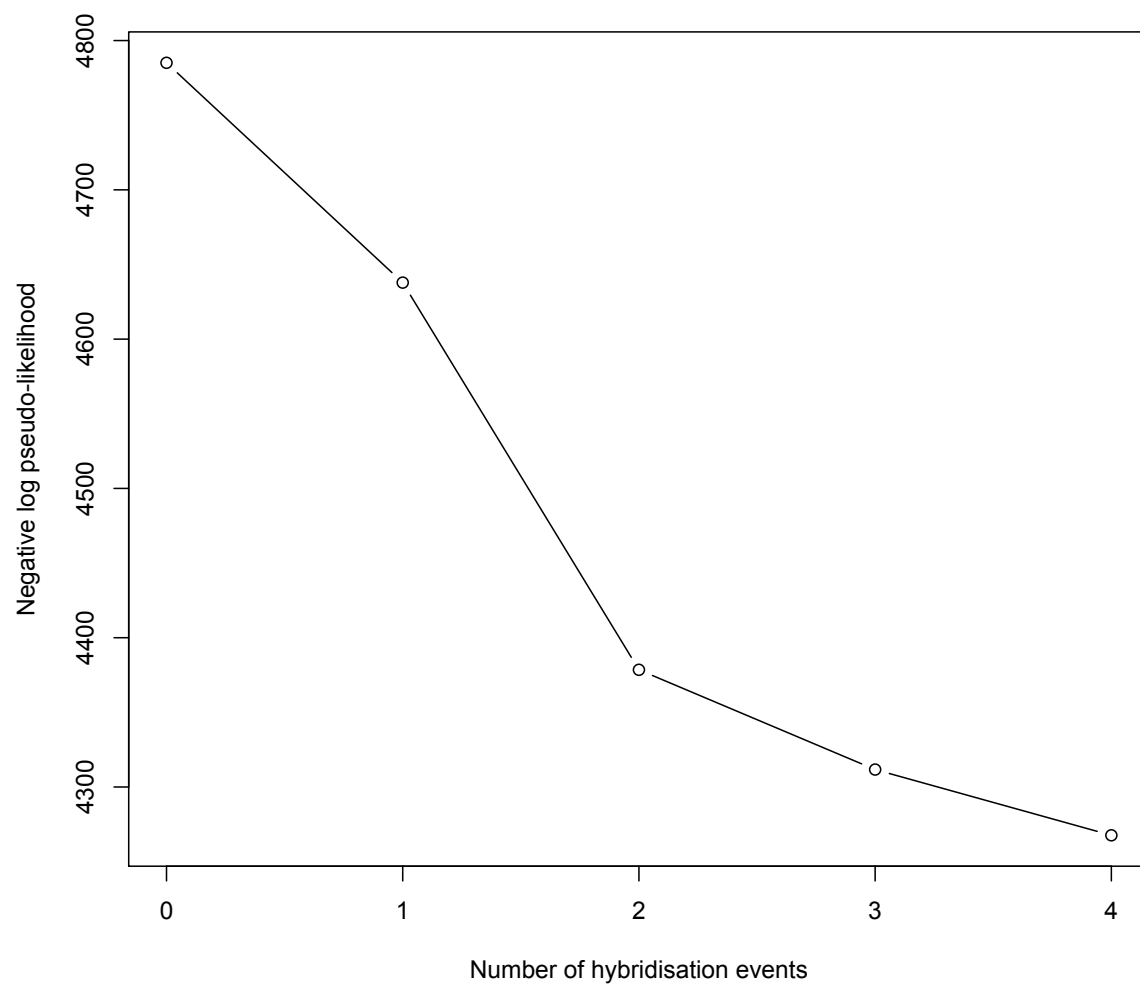


Figure 3

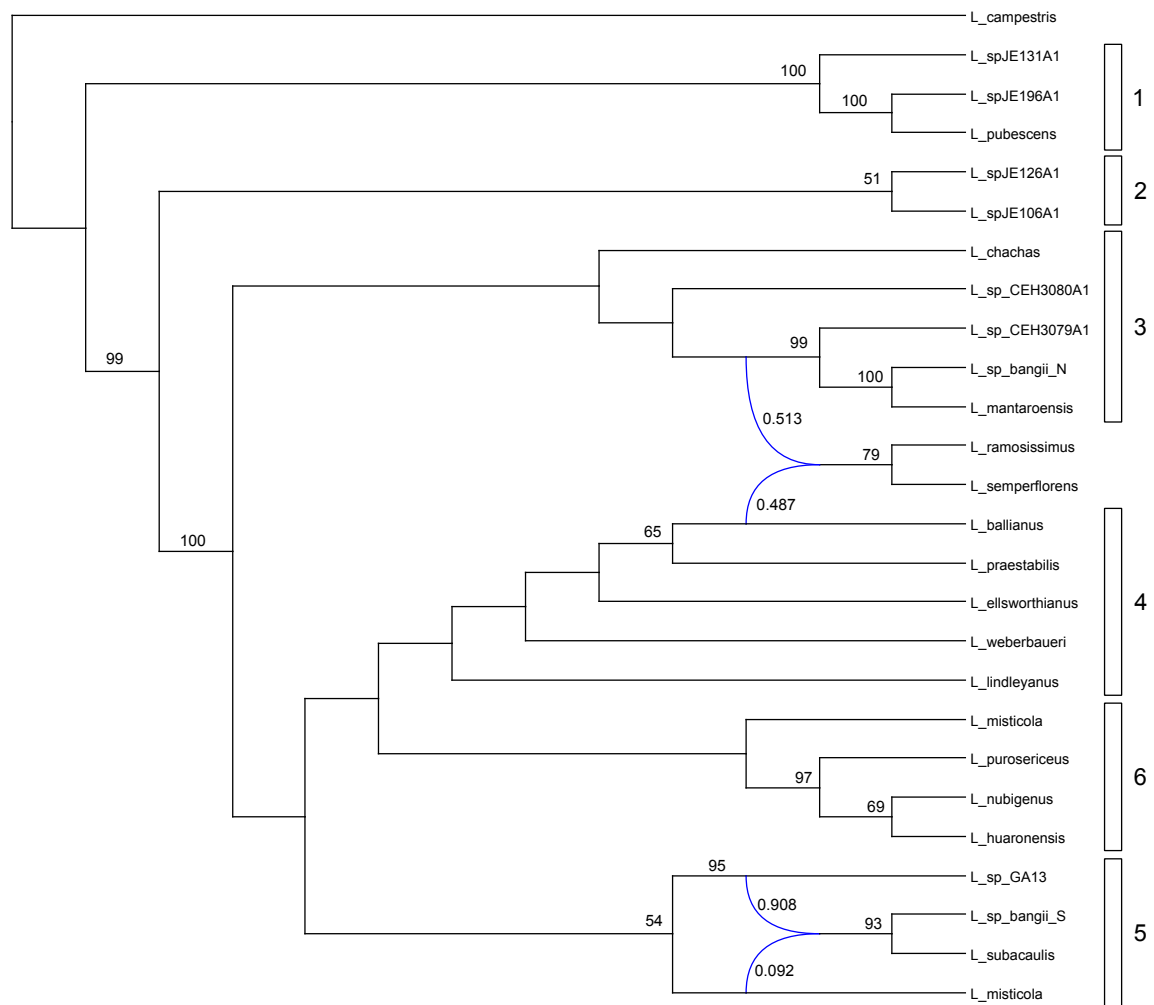


Figure 4

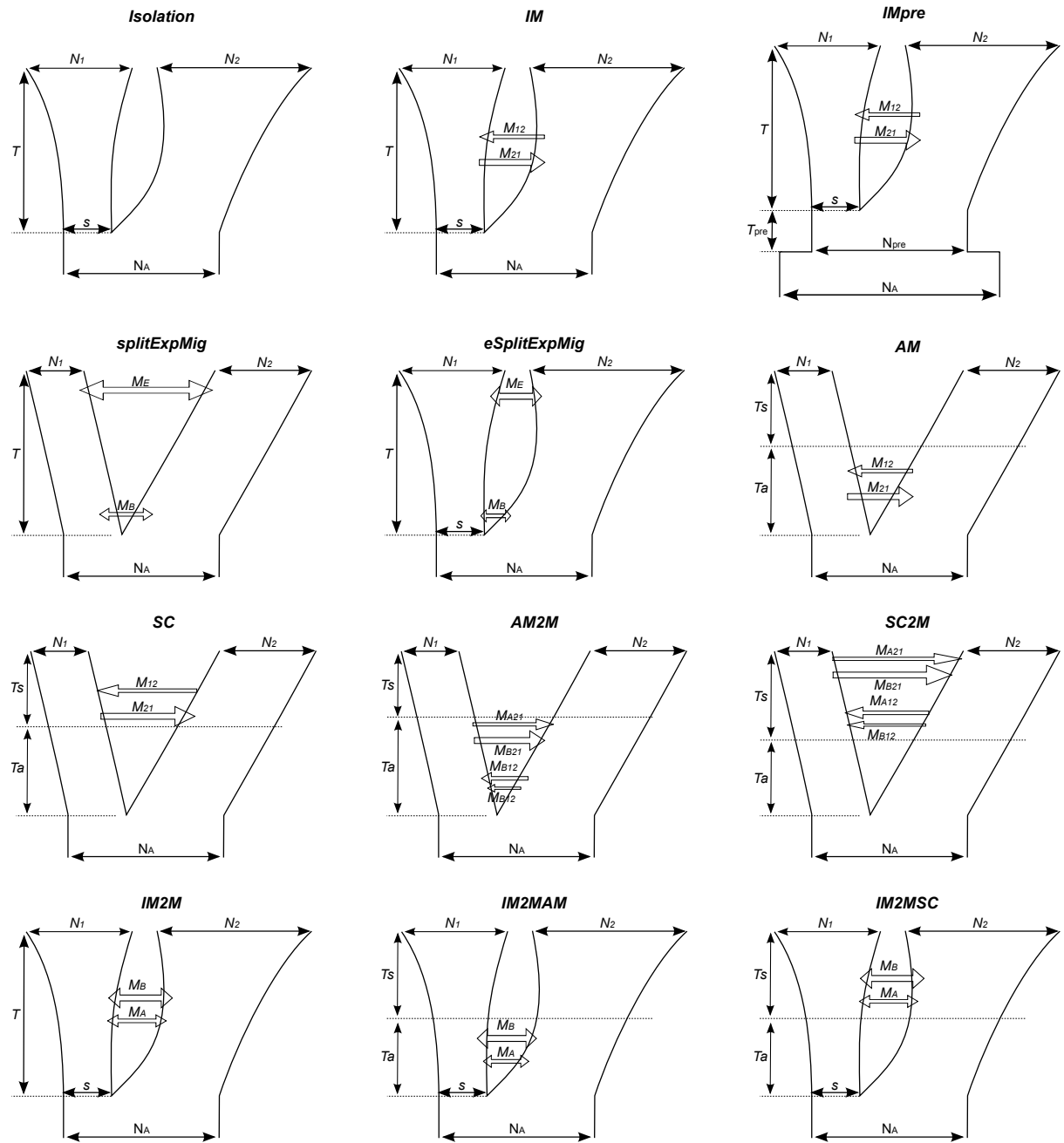


Figure 5

