Phylogenomic Analysis of the Explosive Adaptive Radiation of the *Espeletia* Complex (Asteraceae) in the Tropical Andes

Charles Pouchon[1], Angel Fernández[2], Jafet M. Nassar[3], Frédéric Boyer[1], Serge Aubert[1,4], Sébastien Lavergne[1], and Jesús Mavárez[1,*]

[1]*Laboratoire d'Ecologie Alpine, UMR 5553, Université Grenoble Alpes-CNRS, Grenoble, France;* [2]*Herbario IVIC, Centro de Biofísica y Bioquímica, Instituto Venezolano de Investigaciones Científicas, Apartado 20632, Caracas 1020-A, Venezuela;* [3]*Laboratorio de Biología de Organismos, Centro de Ecología, Instituto Venezolano de Investigaciones Científicas, Apartado 20632, Caracas 1020-A, Venezuela;* [4]*Station alpine Joseph-Fourier, UMS 3370, Université Grenoble Alpes-CNRS, Grenoble, France*
Correspondence to be sent to: Laboratoire d'Ecologie Alpine, UMR 5553, Université Grenoble Alpes-CNRS, BP 53, 2233 rue de la piscine, 38041 Grenoble Cedex 9, France;
E-mail: jesus.mavarez@univ-grenoble-alpes.fr.
Serge Aubert: 1966–2015.

Abstract.—The subtribe Espeletiinae (Asteraceae), endemic to the high-elevations in the Northern Andes, exhibits an exceptional diversity of species, growth-forms, and reproductive strategies. This complex of 140 species includes large trees, dichotomous trees, shrubs and the extraordinary giant caulescent rosettes, considered as a classic example of adaptation in tropical high-elevation ecosystems. The subtribe has also long been recognized as a prominent case of adaptive radiation, but the understanding of its evolution has been hampered by a lack of phylogenetic resolution. Herein, we produce the first fully resolved phylogeny of all morphological groups of Espeletiinae, using whole plastomes and about a million nuclear nucleotides obtained with an original *de novo* assembly procedure without reference genome, and analyzed with traditional and coalescent-based approaches that consider the possible impact of incomplete lineage sorting and hybridization on phylogenetic inference. We show that the diversification of Espeletiinae started from a rosette ancestor about 2.3 Ma, after the final uplift of the Northern Andes. This was followed by two independent radiations in the Colombian and Venezuelan Andes, with a few trans-cordilleran dispersal events among low-elevation tree lineages but none among high-elevation rosettes. We demonstrate complex scenarios of morphological change in Espeletiinae, usually implying the convergent evolution of growth-forms with frequent loss/gains of various traits. For instance, caulescent rosettes evolved independently in both countries, likely as convergent adaptations to life in tropical high-elevation habitats. Tree growth-forms evolved independently three times from the repeated colonization of lower elevations by high-elevation rosette ancestors. The rate of morphological diversification increased during the early phase of the radiation, after which it decreased steadily towards the present. On the other hand, the rate of species diversification in the best-sampled Venezuelan radiation was on average very high (3.1 spp/My), with significant rate variation among growth-forms (much higher in polycarpic caulescent rosettes). Our results point out a scenario where both adaptive morphological evolution and geographical isolation due to Pleistocene climatic oscillations triggered an exceptionally rapid radiation for a continental plant group. [Adaptive radiation; caulescent rosette; explosive diversification; Espeletiinae; hybridization; paramo; phylogenomics; tropical high-elevation.]

There has been a persistent debate about the causes of the outstanding species richness in northern South America, and particularly in the Andean-Amazonian region (reviewed in Hoorn et al. 2010; Antonelli and Sanmartín 2011a; Hoorn et al. 2011; Rull 2011a, 2011b; Madriñán et al. 2013; Luebert and Weigend 2014). Two broad types of explanations have been traditionally proposed. On one hand, the Miocene/Pliocene Andean orogeny is generally considered to be an important driver of species diversification in the Neotropics, as it provided new mountain habitats and deep valleys that increased habitat heterogeneity and acted as both intra and inter Andean vicariant barriers, e.g. *Hedyosmum* (Antonelli and Sanmartín 2011b), *Macrocarpaea* (Struwe et al. 2009) and Rubiaceae (Antonelli et al. 2009). On the other hand, the Pleistocene climatic (glacial) cycling undoubtedly impacted the adaptation, diversity and distribution of organisms, particularly at high elevation habitats in the Northern Andes (a.k.a. paramo) (Van der Hammen and Cleef 1986; Hooghiemstra et al. 2006; Flantua and Hooghiemstra 2018). Thus, during the last 2.4 Myr, glacial periods during which paramos were relatively larger and inter-connected alternated with interglacial periods during which paramos tended to be relatively smaller and isolated "sky islands" (Van der Hammen and Cleef 1986). These repeated cycles of ecological connectivity and spatial isolation could have acted as a "species pump," increasing significantly the speciation rates in high-elevation Andean taxa as consequence of the joint action of allopatry, diversifying selection and adaptation, e.g., *Halenia* (von Hagen and Kadereit 2003), *Hypericum* (Nürk et al. 2013) and *Lupinus* (Drummond et al. 2012b). More recently, biotic factors such as plant-pollinator interactions have also been considered as important drivers in the diversification of Andean taxa (Schmidt-Lebuhn et al. 2007; Abrahamczyk et al. 2014; Lagomarsino et al. 2016).

A proper assessment of the relative importance of these factors on the evolution and diversification of Andean taxa requires integrative analyses that combine morphologic, geographic, and ecological data into a time-calibrated species-level phylogenetic framework. Several examples of spectacular high-elevation Andean diversification have been documented

recently, e.g. *Bartsia* (Uribe-Convers and Tank 2015), Espeletiinae (Rauscher 2002; Diazgranados and Barber 2017), *Hypericum* (Nürk et al. 2013), *Jamesonia-Eriosorus* (Sánchez-Baracaldo and Thomas 2014), *Lupinus* (Drummond et al. 2012b) and *Valeriana* (Bell et al. 2015), but only a few have provided fully-resolved phylogenetic trees, e.g. *Lupinus* (Nevado et al. 2016), *Diplostephium* (Vargas et al. 2017). In fact, the lack of resolution has been so far a prominent feature of most phylogenetic studies of species-rich high-elevation Andean plant taxa, surely due to both the rapid diversification yielding low phylogenetic signal in molecular data and the low resolution of the molecular markers used (Maddison and Knowles 2006). Some of these problems can now be overcome thanks to the advent of next-generation sequencing (NGS), which offers unprecedented possibilities in phylogenomics by increasing greatly both taxonomic coverage and the amount of sequence data, while remaining at a relatively low cost.

The increase in sequence data has also reinforced the view that phylogenetic trees built from different genes sometimes show conflicting branching patterns due to incomplete lineage sorting (ILS), gene flow or gene duplication/loss (Maddison 1997; Maddison and Knowles 2006; Degnan and Rosenberg 2009). Some of these phenomena are presumably important among plant groups that have experienced rapid and recent diversification. For instance, ILS is expected to be common because the short phylogenetic branches exhibited by rapidly diverging taxa provide insufficient time for lineages to coalesce within species between speciation events. Recurrent gene flow is also expected to be important after the secondary contact of interfertile taxa that diverged through ecological or geographical isolation, and that frequently shift their distribution ranges due to climate oscillations. Fortunately, several methods based on the multispecies coalescent have been developed recently to reconstruct the evolutionary history of species in the presence of gene-tree discordance, e.g. ILS (*ASTRAL*, Mirarab et al. 2014) and hybridization (*SNaQ*, Solís-Lemus et al. 2016). Other sources of phylogenetic discordance, such as gene duplication/loss, are often treated upstream during the selection of studied loci, in order to maximize the proportion of orthologous markers in the sequence dataset (Dalquen et al. 2013; Gabaldón and Koonin 2013). Despite these methodological advances, there is still a persistent debate about the best strategy for species tree inference (Springer and Gatesy 2016; Edwards et al. 2016), with some studies showing that coalescent-based multispecies methods that use the ensemble of markers perform better than the concatenation of those markers into a single sequence (Kubatko and Degnan 2007; Roch and Steel 2015), whereas other studies have showed the opposite under some circumstances, e.g. low phylogenetic signal (Bayzid and Warnow 2013; Mirarab et al. 2014). In consequence, most studies use both approaches and perform statistical tests to infer the extent of conflict between them and among genes, e.g. Nevado et al. (2016).

The Exceptional Diversity of Espeletiinae (Asteraceae)

The subtribe Espeletiinae comprises about 144 species characterized by a unique combination of: diploid karyotype ($N = 19$), spiral phyllotaxis, xeromorphic structure and organization of leaves and inflorescences, fertile ray flowers, functionally male disc flowers and, with a single exception, lack of specialized seed dispersal mechanisms (Cuatrecasas 1986; 2013). These plants are distributed in the high Andes of Venezuela, Colombia and Northern Ecuador, with an important diversity in the Colombian Cordillera Oriental (~76 spp.) and the Venezuelan Cordillera de Mérida (~59 spp.) (Fig. 1a). Other areas within the distribution of Espeletiinae harbor considerably lower species diversity, a pattern that has been attributed to their isolation from the presumed center of origin in the Venezuelan Cordillera de Mérida and the lack of long-distance dispersal mechanisms in these plants (Cuatrecasas 2013). Espeletiinae originated from a common ancestor likely around 3–5 Ma (Rauscher 2002; Madriñán et al. 2013) and are thus considered as one of the fastest examples of continental plant diversification. In this relatively short time they have accumulated a remarkable ecological variation in regards with 1) elevation ranges: from upper montane cloud forests at about 2500 m to the edge of glaciers at 4800 m, 2) moisture requirements: from wet paramo bogs to xeric periglacial talus slopes and rocky outcrops, 3) light conditions: from montane forest understory to open grasslands of paramo, and 4) reproductive strategies: with polycarpic species that reproduce repeatedly across adult life and monocarpic species that reproduce only once before death. In addition, Espeletiinae also exhibit a large morphological disparity: caulescent (stemmed) rosettes with dense foliar pubescence and thick layers of dead leaves covering the stem, caulescent rosettes with leafless stems, stem-less and dwarf rosettes, large and profusely branched trees, dichotomous trees, and shrubs (Cuatrecasas 1986; 2013) (Fig. 1b).

Throughout this article, the rosette growth-form is defined as a compact crown of green leaves arranged spirally around the apical meristem, each leaf possessing a flat sheath tightly appressed under and over the sheaths of nearby leaves. The stem consists of a large central pith surrounded by secondary vascular tissues, entirely covered in its apical part by the sheaths of the green leaves. Most rosette species have a unique, unbranched stem, but a few can be slightly branched nearby ground level, each branch producing a terminal crown of rosette leaves. In contrast, tree/shrub growth-forms have a distinct woody trunk with a small or absent pith in comparison with the surrounding wood. They are profusely branched at a considerable distance from the ground and show a less compact leaf arrangement, although with some crowding at the end of the branches. The leaf sheaths of these forms are either curved with free margins and embracing part of the branch, or tubular and encircling the branch entirely. All Espeletiinae fit into these two broad types, with only
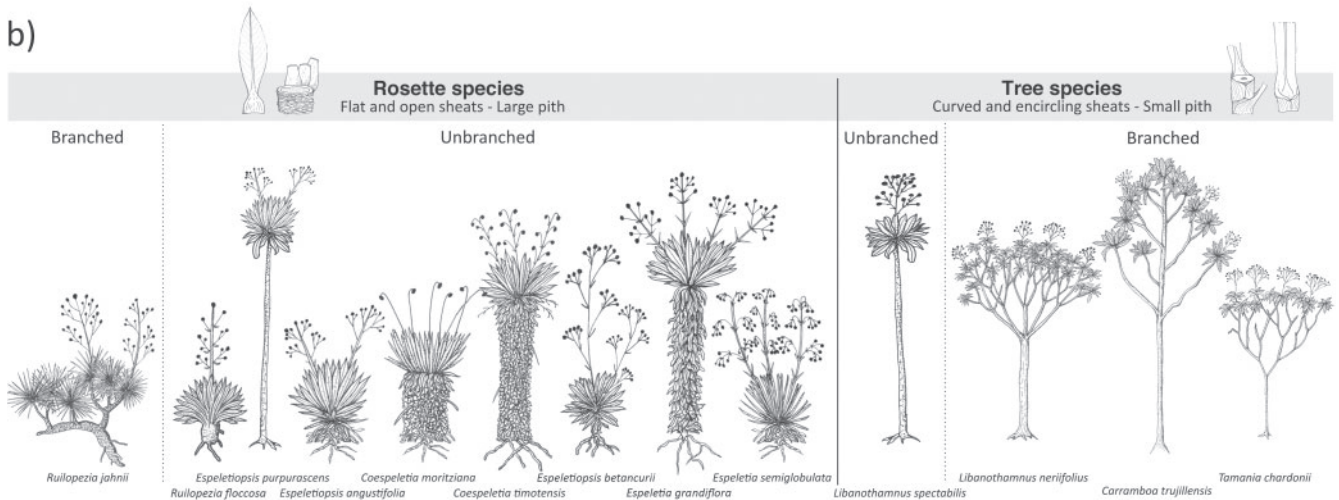
a)
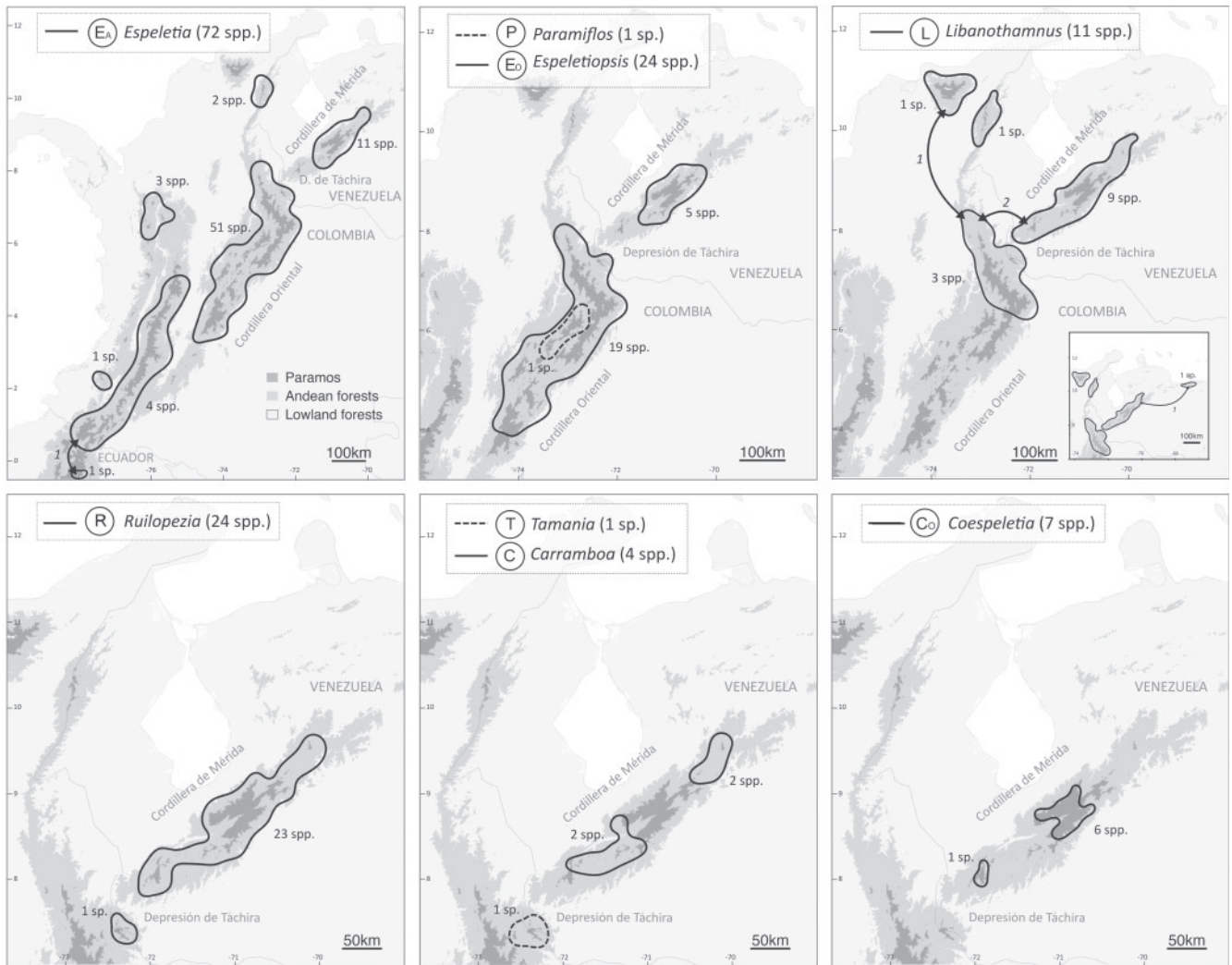
b)

FIGURE 1. a) Species diversity and approximate distributions of the eight genera of Espeletiinae. Arrows provide numbers of shared species between main geographic areas. b) Schematic representations of main growth-forms.

two known exceptions: *Libanothamnus spectabilis* is a tall caulescent rosette with curved and embracing sheaths, while *Tamania chardonii* is a treelet with relatively flat leaf sheaths (Fig. 1b).

The latest systematic treatment of Espeletiinae divides the subtribe in eight genera regrouped in two large clades split according to the position of the inflorescence (Diazgranados 2012; Cuatrecasas 2013): 1) with terminal inflorescences in the trees *Libanothamnus* Ernst (11 spp) and *Tamania* Cuatrec. (1 sp), and the rosettes *Ruilopezia* Cuatrec. (24 spp) and 2) with lateral inflorescences in the trees *Carramboa* Cuatrec. (4 spp) and the rosettes *Coespeletia* Cuatrec. (7 spp), *Espeletia* Mutis ex Bonpl. (72 spp), *Espeletiopsis* Cuatrec. (24 spp) and *Paramiflos* Cuatrec. (1 sp). This system contrasts with a previous one by the same author (Cuatrecasas 1986), in which the same genera were grouped according to the structure of the inflorescence: 1) with alternate (monochasial) inflorescences in *Libanothamnus, Tamania, Espeletiopsis, Paramiflos* and *Ruilopezia* and 2) with opposite (dichasial) inflorescences in *Carramboa, Espeletia* and *Coespeletia* (the alternate inflorescences in *Coespeletia* would be a secondary condition). In this context, molecular phylogenetic methods can be used to assess the relative importance of the position and the structure of the inflorescence on the evolution and the systematic treatments of the group.

Earlier Works of Phylogenetics and Evolution in Espeletiinae

Despite the long-standing interest generated by their exceptional morphological and ecological diversity, the diversification dynamics of Espeletiinae and its potential drivers have remained largely unknown. The earlier works of Panero et al. (1999) on cpDNA and Rauscher (2002) on ITS1-5.8s-ITS2 dealt more with the closest relatives and positioning of the subtribe Espeletiinae within the tribe Heliantheae than with its internal relationships and recent evolution. Indeed, the evolutionary relationships within Espeletiinae have been particularly difficult to disentangle due to lack of phylogenetic resolution (Rauscher 2002; Sánchez-Andrade 2005; Diazgranados and Barber 2017). Nonetheless, a certain number of hypotheses regarding patterns of molecular, morphologic and geographic evolution in Espeletiinae have been formulated. For instance, first Smith and Koch (1935), and later Cuatrecasas (1986) considered that the group evolved from a branched woody tree or shrub ancestor, similar to the tree genus *Libanothamnus* in the Cuatrecasas' classification, an idea that reflects the supposed ancestral habit of the tribe Heliantheae (Smith and Koch 1935; Cuatrecasas 2013). However, as noted by Rauscher (2002), most of the closest relatives of Espeletiinae are herbs, therefore, if the most recent common ancestor of the subtribe was a woody tree, it must have been the result of a recent evolution from herbaceous ancestors near the origin of the subtribe.

It has also been hypothesized that the radiation of Espeletiinae must have started in the Venezuelan Cordillera de Mérida given the greater morphological diversity occurring there (Smith and Koch 1935; Cuatrecasas 1986, 2013) or a supposed older age of species in this range (Diazgranados and Barber 2017). According to this view, Espeletiinae lineages that are non-endemic to the Cordillera de Mérida, mainly the trees *Libanothamnus* and the rosettes *Espeletia* and *Espeletiopsis* (Fig. 1a), would have later crossed the ~30 km wide Táchira depression and spread to the Colombian Cordillera Oriental, presumably during Pleistocene glacial periods that favored connectivity among paramos. However, such view has no support in previous phylogenetic (Rauscher 2000; Diazgranados and Barber 2017) or paleobotanical reconstructions of the distributions of paramos during the Pleistocene (Van der Hammen and Cleef 1986; Flantua and Hooghiemstra 2018). Finally, some conflicting views have as well emerged in regard to the evolution of the caulescent rosette growth-form within Espeletiinae. Smith and Koch (1935) suggested that it evolved from the tree ancestor only once, followed by repeated evolutions of monocarpic and polycarpic rosette growth-forms. On the other hand, Cuatrecasas (1986, 2013) suggested that the caulescent rosette evolved multiple times from the tree ancestor lineage and without repeated evolution of monocarpy/polycarpy (Fig. 2). Unfortunately, the lack of phylogenetic resolution in previous analyses has prevented a conclusive test of these ideas, e.g. Rauscher (2000); Diazgranados and Barber (2017).

In this work, we produce the first fully resolved phylogenetic tree of all the morphological groups of Espeletiinae, based on the analysis of genome shotgun sequence data. The study provides sequences for both the plastome and a large number of nuclear contigs obtained with an original *de novo* assembly procedure that does not require a reference genome. We use this data to run phylogenomic analyses to test previous ideas on the diversification of this emblematic group of Andean plants, particularly with regard to the following questions: 1) When did the radiation of Espeletiinae start? Was it associated with Andean Pliocene tectonics, Pleistocene climatic cycles or both? 2) What was the ancestral morphological growth-form of Espeletiinae? 3) Is there evidence for long-distance dispersal events between the Colombian Cordillera Oriental and the Venezuelan Cordillera de Mérida, particularly among rosette lineages? and 4) Do the different growth-forms in the subtribe represent unique evolutionary events or instead multiple/repeated evolutions of morphologically similar growth-forms? Did these changes occur in a predictable way, e.g. from trees to rosettes and from lower to higher elevations?

MATERIALS AND METHODS

Sampling, DNA Extraction, and Sequencing

A total of 41 samples were taken from an equal number of Espeletiinae taxa in the Andes of Colombia and Venezuela, representing the eight
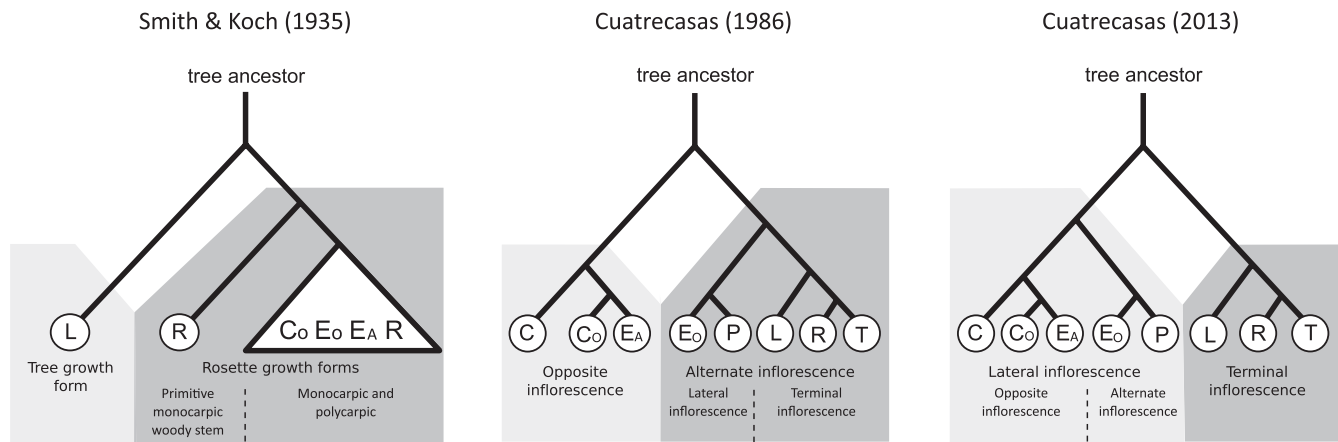
FIGURE 2. Hypotheses of morphological evolution in Espeletiinae (codes for genera as in Fig. 1).

previously recognized genera and all the known morphological diversity of the subtribe in terms of growth-form (trees and rosettes), inflorescence structure (opposite and alternate), inflorescence position (terminal and lateral) and reproductive strategy (monocarpic and polycarpic) (Supplementary Table S1 available on Dryad at http://dx.doi.org/10.5061/dryad.m1t32). Species were identified following the keys provided in Cuatrecasas (1996, 2013). Following Rauscher (2002), *Smallanthus* aff. *pyramidalis* and *Rumfordia penninervis* in the closely related subtribe Milleriinae were also sampled and considered as outgroups for phylogenetic reconstructions. To assess the possible impact of using a single sample per taxon on the phylogenetic reconstruction of such a recent evolutionary group, two samples were used in species exhibiting considerable intra-specific variation, such as *Carramboa badilloi* (*C. badilloi* var. *badilloi* and *C. badilloi* var. *pittieri*) and *Espeletia schultzii* (*E. schultzii* var. *schultzii* and *E. schultzii* var. *mucurubana*).

Leaf samples were collected, shaved with a razor blade to eliminate as much foliar pubescence as possible and placed in airtight plastic bags with silica-gel. To increase yield, DNA extractions were done with a modified protocol of the DNeasy Plant Mini Kit (QIAGEN), using 40% more than recommended grinded dried leaf tissue (and reagents) and overnight incubations at 65°C for the cell lysis step. Genomic shotgun libraries were prepared and sequenced according to the manufacturer's instructions in an Illumina HiSeq 2000 at Fasteris (Geneva, Switzerland).

Assembly of Genomic Regions

Paired-end 2×100 bp Illumina reads were filtered out using a minimum Phred score quality threshold value of 20 (i.e., 99% correct base calling) with *FASTX* (http://hannonlab.cshl.edu/fastx_toolkit/). The remaining reads were used for a *de novo* assembly of the plastid genomes and nuclear contigs (referred to as Total Number of Reads, TNR, in Supplementary Table S1 available on Dryad).

Plastomes.—Assembly of plastid genomes was performed using the Organelle Assembler described in http://metabarcoding.org/asm. The sequence seeds necessary for the *buildgraph* command consisted of chloroplast proteins from *Arabidopsis thaliana* provided with the assembler. In order to optimize the assembly procedure, the *cutlow* coverage value was readjusted for each species after inspection of its assembly graph. Plastomes that could not be assembled with this approach were obtained by aligning the genomic reads of concerned species onto a reference composed by the *de novo* assembled plastomes using the *BWA* v0.7.5 (Li and Durbin 2009). We considered only reads with maximal MAPQ scores of 60 for variant calling using *SAMtools* v0.1.19 and *BCFtools* v0.1.19 (Li et al. 2009). To ensure calling quality, SNPs were filtered using the *vcfutils* script (provided by *SAMtools*) with a root-mean-square of the SNP mapping quality ≥ 15 and with a minimum coverage of 10 per sample.

Nuclear contigs.—Given the unavailability of reference genomes in the Heliantheae at the time of this study (July 2016), we used a *de novo* assembly for nuclear contigs of Espeletiinae according to the pipeline described schematically in Supplementary Fig. S1 available on Dryad. In short, to isolate nuclear sequences, *BWA* was used to remove all the reads of each species that aligned to the mitochondrial genome of the sunflower (*Helianthus annuus*, GenBank accession number KF815390) and the assembled plastomes of Espeletiinae. Reads that aligned to an available database of the repeated elements of the sunflower genome (http://www.agr.unipi.it/ricerca/plant-genetics-and-genomics-lab/sequence-repository.html) were also removed, in order to avoid problems during SNP base calling steps (step 1, Supplementary Fig. S1 available on Dryad). All remaining unmapped reads were considered as potential nuclear reads (referred to as PNR in Supplementary Table S1 available on Dryad) and assembled into contigs using *Velvet* v.1.2.10 (Zerbino and Birney 2008) (step 2,

Supplementary Fig. S1 available on Dryad). To optimize the assembly, the hash length (or k-mer length) was automatically selected using *VelvetOptimiser* v.2.2.5 (https://github.com/tseemann/VelvetOptimiser), according to the N50 value. Coverage *cutoff* and expected coverage were automatically selected by *Velvet* but only contigs >200 bp were considered (Contigs, Supplementary Table S1 available on Dryad). Contigs that still matched mitochondrial (*Helianthus*) and plastid (Espeletiinae) sequences were removed using a Python script. Only the remaining sets of nuclear contigs were retained for further analyses (referred to as Final Number of Contigs, FNC, in Supplementary Table S1 available on Dryad).

In order to find shared nuclear regions, nuclear contigs from all species were pooled and assembled into composite multispecies contigs (hereafter metacontigs) using *VelvetOptimiser* (step 3, Supplementary Fig. S1 available on Dryad). Only metacontigs assembled from at least three different species were retained. Next, the nuclear reads of all species were pooled and aligned to this set of metacontigs, from which coverage and position were extracted using *BEDtools* v2.17.0 (Quinlan and Hall 2010). In order to identify and eliminate repeated elements that could still be present in the data, metacontigs with at least one region within the 10% most covered FNC were arbitrarily removed from the list using *SAMtools* and a custom Python script (step 4, Supplementary Fig. S1 available on Dryad). The nuclear reads of each species were aligned to the remaining metacontigs using *BWA*, retaining only those that mapped to a single metacontig with a maximal MAPQ score of 60 for the variant calling using *SAMtools* and *BCFtools* (step 5, Supplementary Fig. S1 available on Dryad). SNPs were further filtered with *vcfutils* using a root-mean-square ≥ 15 and a minimum SNP coverage of 10 per marker and per species. Finally, to guarantee the quality of alignments, only metacontigs with less than 5% of ambiguities were kept for further phylogenetic analyses (step 6, Supplementary Fig. S1 available on Dryad). The final number of composite nuclear metacontigs in the reference list was 9880.

Phylogenetic Analyses

Plastomes.—Plastid sequences were aligned using *MAFFT* v7.222 (Katoh 2002). Ambiguous sites from the alignment matrices were trimmed with *trimAl* using the heuristic algorithm *automated1* (Capella-Gutiérrez et al. 2009), which uses the distributions of gaps and similarities to determine the thresholds for removing the poorly aligned sites of an alignment. Phylogenies were next reconstructed using Bayesian Inference (BI) and Maximum Likelihood (ML) approaches. BI analyses were conducted using a standard non-clock approach with *ExaBayes* v.1.4.1 (Aberer et al. 2014). We used the default values for the temperatures of chain heating and the number of Metropolis-coupled Markov chain Monte Carlo chains for the replicate runs with 500,000 generations. Convergence and mixing

were assessed using summary statistics from the *postProcParam* command. We ensured a minimum value of 200 for the effective sample sizes of each parameter, and the similarity within and between chain-variance with a potential scale reduction factor close to 1 and below 1.1. 50% majority-rule consensus phylograms and posterior probabilities were obtained using the *consense* command with a burnin fraction of 25%. ML analyses were performed using *RAxML-HPC* (Stamatakis 2006). The best-fitting model of evolution, GTR+G+I, was determined with the Akaike Information Criterion (AIC) as implemented in the R package *Phangorn* (Schliep 2011) (Supplementary Table S2 available on Dryad). 1000 ML bootstrap replicates were performed.

Nuclear metacontigs.—Phylogenies based on the nuclear dataset were built with two approaches: using concatenated alignments of metacontigs and coalescent-based multispecies tree methods. In order to use the same data in the two phylogenetic approaches, only nuclear metacontigs shared by at least four taxa were used (i.e., quartets), which is a requisite of the coalescent-based approach. This dataset, consisting of all metacontigs shared by one outgroup and at least three other taxa, contained 1877 metacontigs and 933,394 nucleotide positions.

Concatenated alignment approach. The alignments of metacontigs were concatenated into a single sequence, trimmed with *trimAl* and used for BI and ML phylogenetic analyses performed with the same software described above. The ML analysis used a GTRCAT substitution model and 1000 bootstrap replicates.

Coalescent-based multispecies tree approach I: ASTRAL. In order to assess for the sensitivity of the concatenated approach to conflicts among gene trees due to ILS, phylogenetic reconstructions were also based on a coalescent-based multispecies tree approach as implemented in *ASTRAL II* (Mirarab and Warnow 2015). This software has been shown to enable highly accurate species tree estimation in the presence of ILS. 1877 gene trees were first estimated using *RAxML* under the GTRCAT model with 200 bootstrap replicates performed on the alignment of each metacontig. To assess for gene conflicts among metacontigs, we used the multilocus bootstrapping procedure from *ASTRAL II* with 1000 replicates combining each *RAxML* best tree and its bootstrap replicates. A majority-rule consensus tree was generated with the R package *ape* (Popescu et al. 2012).

Coalescent-based multispecies tree approach II: SNaQ. We used another approach to discriminate between ILS and gene flow as possible sources of phylogenetic discordance, implemented in the software *SNaQ* (Solís-Lemus et al. 2017). *SNaQ* performs maximum pseudolikelihood estimation of phylogenetic networks from multilocus data using the coalescent-based multispecies model and quartet-based concordance analyses. This software allows for the simultaneous estimation of both a major phylogenetic topology depicting the main evolutionary

history ("major edge") and past events of hybridization ("minor edges"), while integrating optimization of branch lengths, inheritance probabilities and a heuristic search in the space of phylogenetic networks based on a starting topology and on the observed quartet concordance factors (CF). A relatively fast procedure to run *SNaQ* consists in using the *ASTRAL* tree as the starting topology and a table of the CF values obtained from the best trees inferred with *RAxML* for each metacontig. Unfortunately, modelling of the multispecies coalescent with gene flow has a heavy computational cost so that the application of *SNaQ* (and other equivalent software) is severely hampered with more than ~25 taxa (Hejase and Liu 2016). To deal with this limitation, we used the fast procedure for *SNaQ* on 29 datasets with <25 taxa each (Supplementary Fig. S2 available on Dryad). The first dataset, consisting of at least two members of each major clade identified in *ASTRAL* plus an outgroup, was used to identify the main topology of the phylogenetic tree connecting all major clades. The remaining 28 datasets, each consisting of all members of two major clades identified in *ASTRAL* (e.g., I+II, I+III, … VII+VIII), were used to identify within clade topologies and to detect all possible hybridization events within and between major clades. A supertree was built with *Phangorn* combining all 29 *SNaQ* trees. *SNaQ* was run on all 1877 *RAxML* best trees, initiated in every case with a starting topology based on the *ASTRAL* species tree pruned to the taxa list of the specific dataset. Five to ten runs were used for each value of the maximum expected number of hybridization events (parameter *h*, increased sequentially from 0 to 5). The optimal *h* was chosen using a heuristic approach by plotting pseudo-likelihoods against *h* values, following *SNaQ*'s authors recommendation (Cécile Ané, comm. pers).

Phylogenetic tests.—To assess the significance of discordances between phylogenetic trees built with the concatenated and the coalescent-based multispecies approaches, we used *RAxML* to calculate site-by-site likelihoods for both a bootstrap set of 100 unconstrained phylogenetic trees and bootstrap sets of 100 trees generated for each alternative constrained topologies. This was used to perform Shimodaira and Hasegawa (1999) test as implemented in *CONSEL* (stat.sys.i.kyoto-u.ac.jp/prog/consel). These tests were also used to compare the phylogenetic hypotheses depicted in Fig. 2 and the monophyly of Cuatrecasas' genera of Espeletiinae (Cuatrecasas 1986, 2013). For the Smith and Koch (1935) hypothesis, a pruned tree was produced by retaining only the 14 species known at that time that matched with our species sampling. In addition, the correlation between genetic (plastid or nuclear) and geographic distances was assessed using the R package *Ecodist* (Goslee and Urban 2007).

Divergence Times and Diversification Rates of Espeletiinae

Divergence times of Espeletiinae were estimated following a two-step procedure involving: 1) estimation of age of the crown node using plastid genomes of additional Asteraceae with ages known from literature and 2) calibration of the divergence time analysis within Espeletiinae using the nuclear data and the crown node age estimated in (1).

(i) The age of the crown node of the subtribe was estimated with a phylogenetic analysis of plastid sequences of Espeletiinae and some members of the Asteraceae tribes Millerieae, Heliantheae, Anthemidae and Cichorieae (Supplementary Table S3 available on Dryad), using known ages extracted from fossil data (pollen) and/or previous molecular divergence analyses. Three calibration dates were employed (Tremetsberger et al. 2013): 22.0–28.4 Ma interval for the age of Cichorieae based on fossil pollen, 10.6–17.8 Ma interval for the divergence *Sonchus/Crepis* and 6.9–16.5 Ma interval for the divergence *Scorzonera/Tragopogon*. Plastomes were sequenced, assembled and analyzed as described above for the plastid BI phylogeny of Espeletiinae. In order to estimate a time-calibrated phylogeny, we used the relaxed-clock penalized likelihood (PL) method implemented in *r8s* v1.7.1 (Sanderson 2003). As implemented in Buerki et al. (2010), the uncertainty in the estimation of divergence times was assessed by generating 1000 trees during the Bayesian Markov chain Monte Carlo process, discarding the first 100 (10% burnin) and retaining the remaining 900 trees. The cross-validation method was used on each tree replicate to determine the optimal level of rate-smoothing for the PL analyses by fitting smoothing parameters with 100 estimates (starting with a log10 value of 1 and increments of 0.1). Dating analyses were then performed on each tree according to the optimal smoothing value. Mean values and 95% confidence intervals for each node age were estimated from the chronograms of the 900 time-calibrated trees using *TreeAnnotator* (Drummond et al. 2012a).

(ii) A time-calibrated analysis of divergences within the subtribe was produced in *r8s* using the concatenated nuclear BI phylogenetic tree, and minimum and maximum age estimations for the crown node of Espeletiinae obtained from the previous analysis. Uncertainty in the estimation of node ages was assessed as above but with 200 trees randomly selected from the Bayesian Markov chain Monte Carlo stationary distribution. Cross-validation was conducted for 100 estimates starting with a log10 value of −5 and increments of 0.1. Mean values and 95% confidence intervals of the divergence dates were estimated from the 200 time-calibrated trees in *TreeAnnotator*.

Diversification rates of Espeletiinae were analysed by testing the fit to models with density-dependence birth-death processes using the R package *DDD*

(Etienne et al. 2012), and calculating rates of speciation, extinction and carrying capacities. Lineages-through-time (LTT) plots were built with the R package *ape* using the 200 nuclear chronograms.

Reconstruction of Ancestral States

Ancestral state reconstruction was performed on the consensus of the nuclear chronograms for four morphological traits considered as binary characters: growth-form (GF, rosette/non-rosette), branching of the stem (STEM, branched/unbranched), inflorescence position on stems (IPOS, lateral/terminal) and inflorescence structure (ISTR, alternate/opposite). Information about these characters was collected from Cuatrecasas (2013). Inferences about trait evolution were done using Markov models (Mk) in the R package *Diversitree* (FitzJohn 2012). Two possible discrete Markov models were fitted and compared using AIC values: a full model with different transition rates (ARD) and a constrained model with equal transition rates (ER). For the growth-form trait, the tree root was constrained to have the non-rosette character state owing to the absence of rosettes among close relatives of Espeletiinae within Heliantheae. No other constraints were imposed in the analyses; however, due to limited sampling of outgroup taxa, only Espeletiinae were considered for inferences about the evolution of the two inflorescences traits. The best model was used to run simulations of stochastic character mapping in order to infer most likely scenarios of trait evolution (Supplementary Table S4 available on Dryad).

RESULTS

Illumina sequencing returned on average 12.8×10^6 (sd. $= 6.6 \times 10^6$) quality-filtered reads per species, of which 0.77% (91,055 reads, sd. = 71,839) served to assemble plastid genomes and 83.9% (10.7×10^6 reads, sd. $= 55.5 \times 10^6$) were used to build nuclear contigs (Supplementary Table S1 available on Dryad). The nuclear data allowed the assembly of on average 7883 (sd. = 3048) contigs per species, with an average N50 of 389 bp (sd. = 32).

Phylogenetic Analysis of Plastomes

We were able to assemble *de novo* 25 complete plastomes from 23 Espeletiinae species and the two outgroups. Plastomes of the remaining 18 species were obtained by mapping their reads on the consensus sequence of the *de novo* assemblies. The alignment contained 151,388 nucleotide sites (mean coverage of 60.1), of which 2.46% were variable and 0.60% were parsimony informative (Supplementary Table S5 available on Dryad). The BI phylogenetic tree showed well-supported clades (mean BPP = 0.98, sd. = 0.06, Fig. 3) and only one basal polytomy. The ML

phylogenetic tree was topologically similar (SH-test, $P = 0.477$), although with slightly lower bootstraps values (mean PP = 89.12, sd. = 19.9, Fig. 3). In spite of this high phylogenetic signal, plastid-based analyses do not support the monophyly of morphological growth-forms, since rosettes and branched species are present in all major clades. The same morpho-phylogenetic discordances were also observed for both the structure and the position of the inflorescences. Accordingly, analyses of plastomes do not support the monophyly of any of the genera proposed by Cuatrecasas (1986, 2013). Finally, there is a strong spatial influence in the clustering of plastid diversity in Espeletiinae, as shown by the highly significant correlation between geographic and genetic (plastid) distances (Mantel test, $P = 0.001$, Fig 3b)

Phylogenomic Analysis of Nuclear Contigs

Concatenated alignment approach.—The BI phylogenetic reconstruction appeared fully resolved with only one node with a BPP <1.0 (Fig. 4). The ML reconstruction provided an identical topology (SH-test, $P = 0.928$, Table 1) although with slightly lower values for node support (PP = 91.3, sd. = 18.7). These analyses indicated an early split of Espeletiinae between two clades composed of lineages from the Colombian Cordillera Oriental and the Venezuelan Cordillera de Mérida, respectively. The Colombian lineage (clade I) is composed of three groups of rosette growth-forms in the genera *Espeletia*, *Espeletiopsis* and the monotypic *Paramiflos glandulosus*. The remaining lineages (clades II to VIII) also formed a strongly supported clade with an early divergence of the lineage composed of tree growth-forms in the genus *Carramboa* (clade II). The tree growth-form species currently grouped in the genus *Libanothamnus* formed a monophyletic group (clade V), not related to the *Carramboa* lineages but rather nested within a larger clade composed exclusively by rosette growth-forms. The third tree growth-form lineage, the monotypic genus *Tamania (T. chardonii)*, was not closely related to any other tree lineage but rather nested within a clade of rosette growth-forms (clade III). These results suggest therefore that taxa exhibiting the tree growth-forms in Espeletiinae are polyphyletic. Phylogenetic analyses indicated that the monocarpic rosettes in the genus *Ruilopezia* are polyphyletic as well (clades III, IV and VI). Accordingly, the SH-test showed significant support for the unconstrained BI tree compared to the constrained topology in which *Ruilopezia* is forced to be monophyletic ($P < 0.001$, Table 1). Furthermore, with only one exception (*Espeletia semiglobulata*, see below) the rosette growth-forms in the Venezuelan members of the genera *Espeletia* (clade VII) and *Espeletiopsis* (within clade IV) represented two monophyletic groups respectively, but polyphyletic at the scale of the subtribe as they appeared unrelated to Colombian taxa grouped under the same generic names. The SH-tests showed significant support for
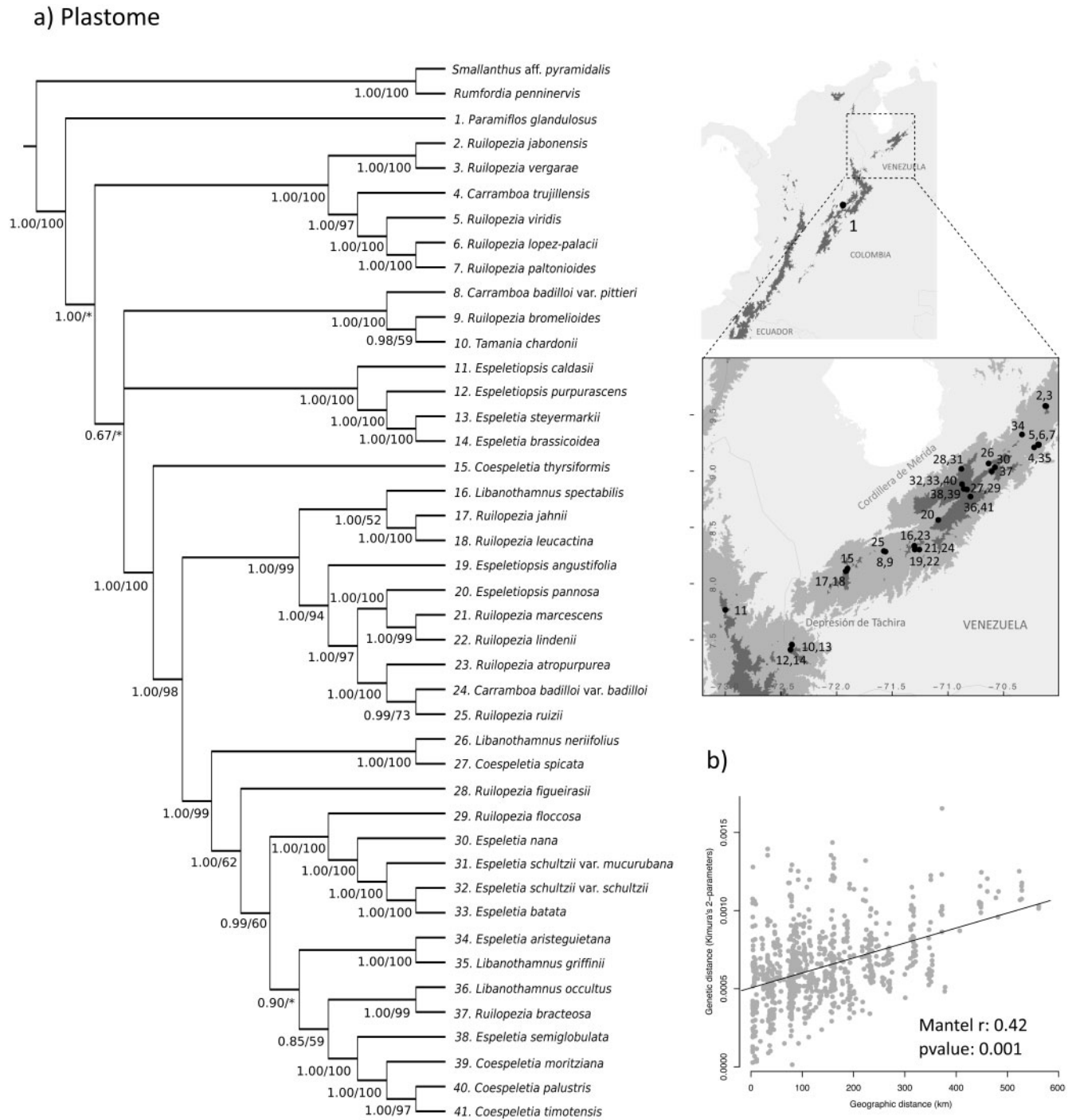
a) Plastome

FIGURE 3. a) Phylogeny of Espeletiinae inferred from analyses of whole plastomes. Phylogenetic supports are represented for each node as Bayesian inference (BI/-) and maximum likelihood (-/ML) values (*indicate support <50). Maps show sampling localities in the Colombian and Venezuelan Andes. b) Mantel tests of the correlation between geographic and genetic (plastid) distances in Espeletiinae.

the unconstrained BI trees compared to the constrained topologies in which *Espeletia* and *Espeletiopsis* are forced to be monophyletic ($P < 0.001$ in both cases, Table 1). Finally, all species in the high-elevation genus *Coespeletia* appeared grouped within a clade that also included *E. semiglobulata* (clade VIII). The SH-test showed significant support for the unconstrained tree compared to the

constrained topology in which *Coespeletia* is forced to be monophyletic by placing *E. semiglobulata* within the Venezuelan *Espeletia* (clade VII) (p = 0.037, Table 1). Finally, conspecific taxa appeared related to each other within species *C. badilloi* and *E. schultzii*, supporting the strength of our approach at recovering phylogenetic relationships even between closest relatives.
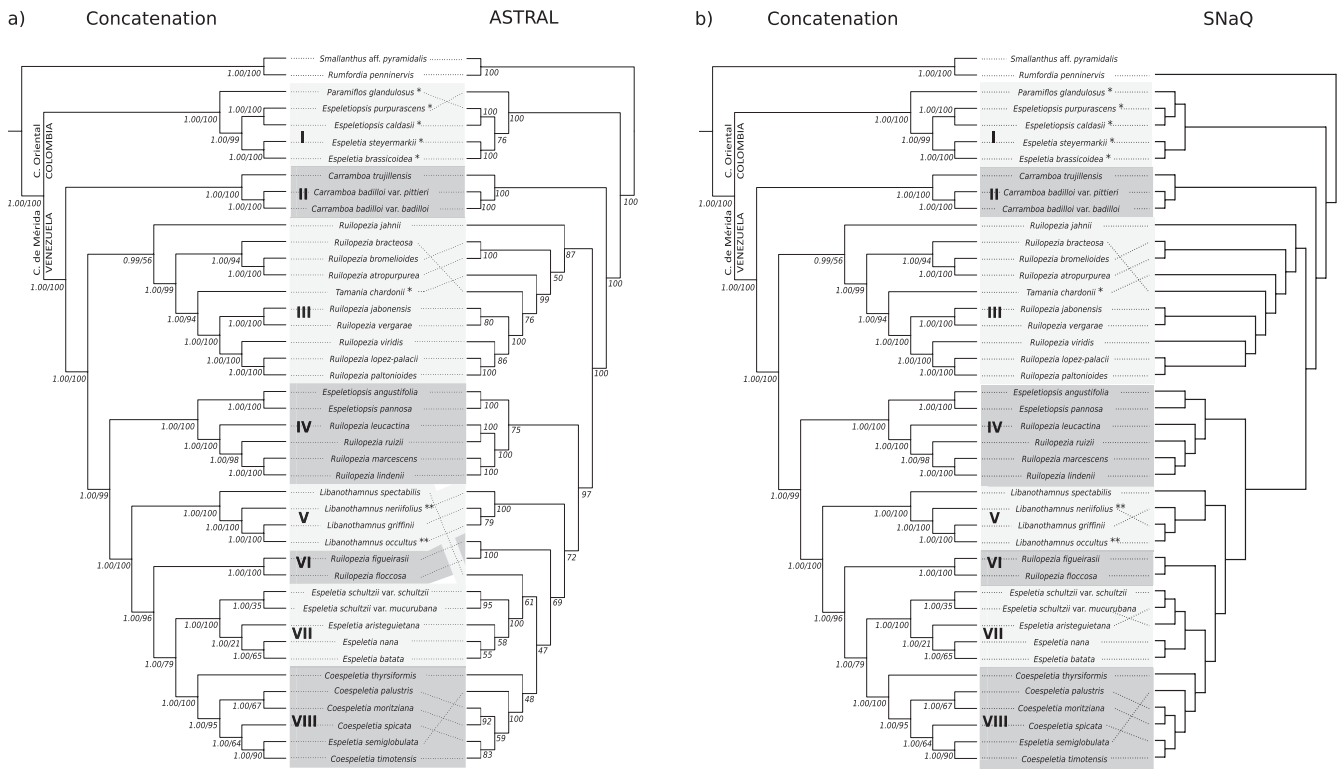
FIGURE 4.   Comparisons of nuclear phylogenies of Espeletiinae. a) Concatenation *vs. ASTRAL*. b) Concatenation *vs. SNaQ*. Node supports are given for concatenation as Bayesian inference (BI/-) and maximum likelihood (-/ML), and for *ASTRAL* as bootstrap probabilities. Species distributions: Venezuelan Andes (no marks), Colombian Andes (*), both countries (**).

TABLE 1.   Phylogenetic comparisons among alternative hypotheses of evolution in Espeletiinae

Phylogenetic topology	Rank	$\Delta_{\log}L$	P-value
BI concatenated nuclear	1	best	0.990
ML concatenated nuclear	2	21.6	0.928
SNaQ species tree	3	678.2	0.237
ASTRAL species tree	4	1145.7	0.065
Monophyly of *Coespeletia*	5	1266.9	0.037
Monophyly of *Ruilopezia*	6	3749.1	< 0.001
Monophyly of *Espeletiopsis*	7	23,125.4	< 0.001
Monophyly of *Espeletia*	8	24,102.6	< 0.001
Cuatrecasas (2013)	9	43,111.1	< 0.001
Cuatrecasas (1986)	10	43,655.2	< 0.001
ML plastid	11	57,295.8	< 0.001
BI pruned nuclear	1	Best	1.000
Smith and Koch (1935)	2	7335.4	< 0.001

$\Delta \log$ L = likelihood difference between the best and alternative topologies; *P*-value = *P*-value for the SH tests.

In agreement with the results above, phylogenetic reconstructions allowed the rejection of all the scenarios previously considered for the evolution of Espeletiinae (Fig. 2). Thus, SH-tests performed rejected Smith and Koch's (1935) hypothesis proposing the monophyly of the rosette growth-form ($P < 0.001$, Table 1). Cuatrecasas' hypotheses about the reciprocal monophyly of alternate/opposite inflorescence structures (Cuatrecasas 1986) and terminal/lateral inflorescence positions (Cuatrecasas 2013) were also

highly unlikely given our phylogenomic reconstructions ($P < 0.001$ in both cases, Table 1). In addition, unlike plastid data, Mantel tests on nuclear data show no significant correlation between genetic and geographic distances ($P = 0.393$).

Coalescent-based multispecies tree approaches.—The tree topology obtained with *ASTRAL* (Fig. 4a) was not significantly different to concatenation according to the SH-test ($P = 0.065$, Table 1). It had, however, lower node support values (mean PP = 86.1, sd. = 17.8), and it exhibited five topological differences. Four of these differences involved within-clade changes without effect on the composition of clades: the position of terminal lineages *P. glandulosus* and *E. purpurascens* (clade I), *R. bracteosa* (clade III) and *E. semiglobulata* (clade VIII). The fifth difference involved an inter-clade change in regard with the position of the unbranched tree *Libanothamnus spectabilis*, which moved from a tree growth-form clade (clade V) to the base of a rosette growth-form clade (clade VII).

The phylogenetic tree recovered with *SNaQ* (Fig. 4b) did not differ significantly from the concatenated tree according to the SH-test ($P = 0.237$, Table 1), although it exhibited five differences, all concerning within-clade changes: the position of terminal lineages *P. glandulosus* (clade I), *R. bracteosa* (clade III), *L. griffinii* (clade V), *E. aristeguietana* (clade VII) and *E. semiglobulata* (clade
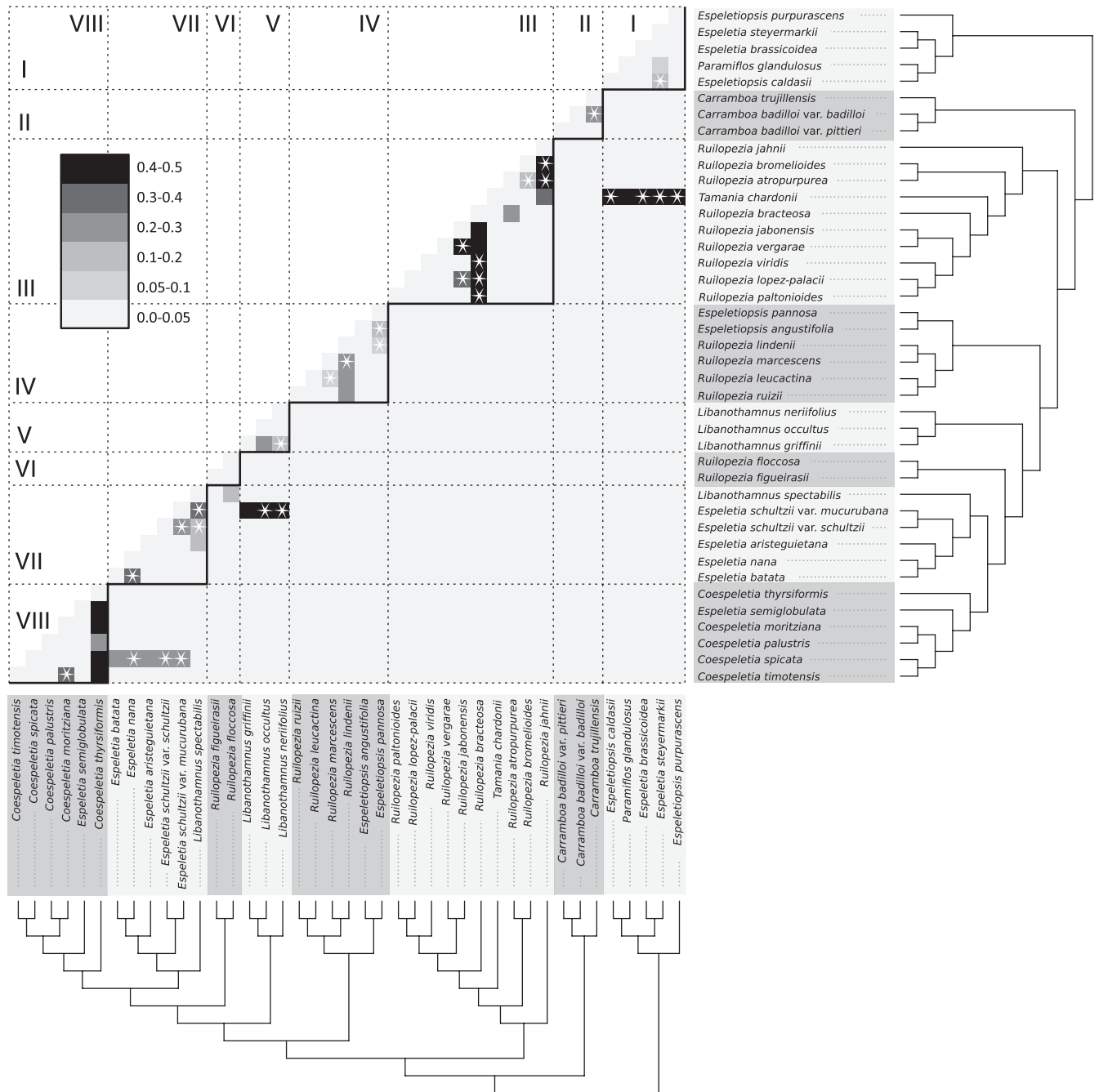
FIGURE 5. Gene flow events among Espeletiinae detected by *SNaQ*. Grey tones show the percentage of nuclear contigs with significant evidence of gene flow between the concerned taxa. Asterisks indicate sympatric/parapatric taxa. The black line separates intra and inter-clade gene flow events. Main clades defined according to the *ASTRAL* gene tree.

VIII). It is worth noting that *SNaQ* analyses did not place *L. spectabilis* within the Venezuelan *Espeletia* (clade VII), but at the base of the *Libanothamnus* clade (clade V). However, *SNaQ* detected significant evidence of gene flow between this species and some rosettes in clade VII, suggesting that hybridization may indeed have played a role during the speciation of *L. spectabilis*. In fact, *SNaQ* allowed detection of an important number of inter-specific gene flow events, but most represented hybridizations between very closely related species within major clades and living in sympatry (Fig. 5). *SNaQ* detected as well some inter-clade hybridization events: 1) between *T. chardonii* and Colombian rosettes in clade I which occur in sympatry in the eastern part of the Colombian Cordillera Oriental, 2) between *E. schultzii* var. *mucurubana* and sympatric *Libanothamnus spp* (clade V), and 3) between *C. spicata* and mostly sympatric *Espeletia spp* (clade VII).
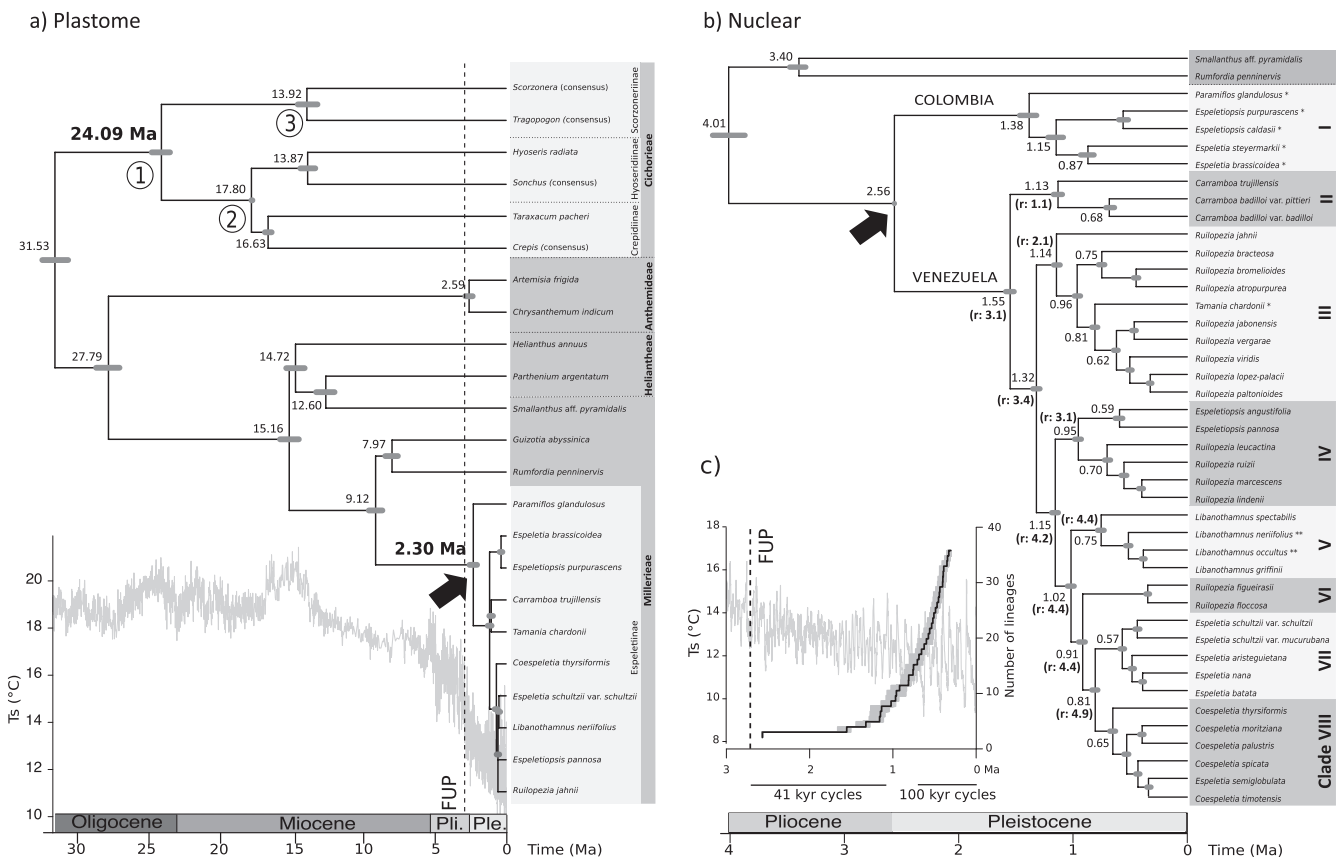
FIGURE 6.   Divergence times of Espeletiinae based on (a) whole plastomes and (b) BI analysis of concatenated nuclear contigs. Age estimations for the crown node of Espeletiinae obtained in (a) were used to calibrate the analysis in (b) (black arrows). Bars on nodes represent 95% confidence intervals. Numbers in parentheses are diversification rates estimated with DDD (see text). c) Lineage through time (LTT) plot of the diversification of Venezuelan Espeletiinae for 200 time-calibrated trees (grey) and the median time-calibrated tree (black line). In (a) and (c), the dashed vertical line represents the age for the final uplift of the paramo ecosystem (FUP, Gregory-Wodzicki 2000) and the light grey lines show sea surface temperatures (Zachos et al. 2001). Calibration nodes are highlighted with circles.

Divergence Times of Espeletiinae

The age recovered for the crown node of Espeletiinae in plastid dataset was 2.3 Ma (2.0–2.6), suggesting that the divergence of the group started quite quickly after paramos reached their modern elevation (~2.7 Ma, Gregory-Wodzicki 2000) (Fig. 6a). Divergence time estimations of Espeletiinae based on the nuclear dataset provided narrow age estimates for all nodes, a few tens of thousands of years on average (mean 63.0 kyr, sd. 46.3) (Fig. 6b). The age of the crown node of Espeletiinae was estimated with this analysis at 2.6 Ma, which is slightly older than the plastid-based age estimation and even closer to the proposed time for the final uplift of paramos. In the Venezuelan radiation, the most thoroughly sampled in this study, no further diversification occurred for about a million years following the initial split of the subtribe, after which the number of lineages accumulated very rapidly and concomitantly with the increase in the amplitude of climate cycles of the second-half of the Pleistocene (6c).

The analysis of diversification in the Venezuelan Espeletiinae (clades II–VIII) showed that the best model corresponds to a pure-birth with an average

speciation rate = 3.09 spp/Myr (Supplementary Table S6 available on Dryad). However, there is a high degree of rate variation among clades in this radiation, with a minimum rate in trees in clade II (1.12), intermediate rates in monocarpic caulescent rosettes in clades III (2.06) and IV (3.09), and trees in clade V (4.37), and a maximum rate observed in polycarpic caulescent rosettes in the highest elevations of the Cordillera de Mérida, clade VII–VIII (4.86) (Fig. 6b, Supplementary Table S6 available on Dryad).

Morphological Evolution of Espeletiinae

According to AIC values, the ER model was used to run evolutionary simulations for traits GF, STEM, and ISTR, whereas the ARD model was used for the IPOS trait (Supplementary Table S4 available on Dryad). The reconstructions of ancestral states, displayed in Fig. 7a and Supplementary Fig. S3 available on Dryad, indicated that the most recent common ancestor of Espeletiinae likely had a woody rosette-like growth-form with a branched stem and a terminal inflorescence with alternate structure. This analysis showed that the tree
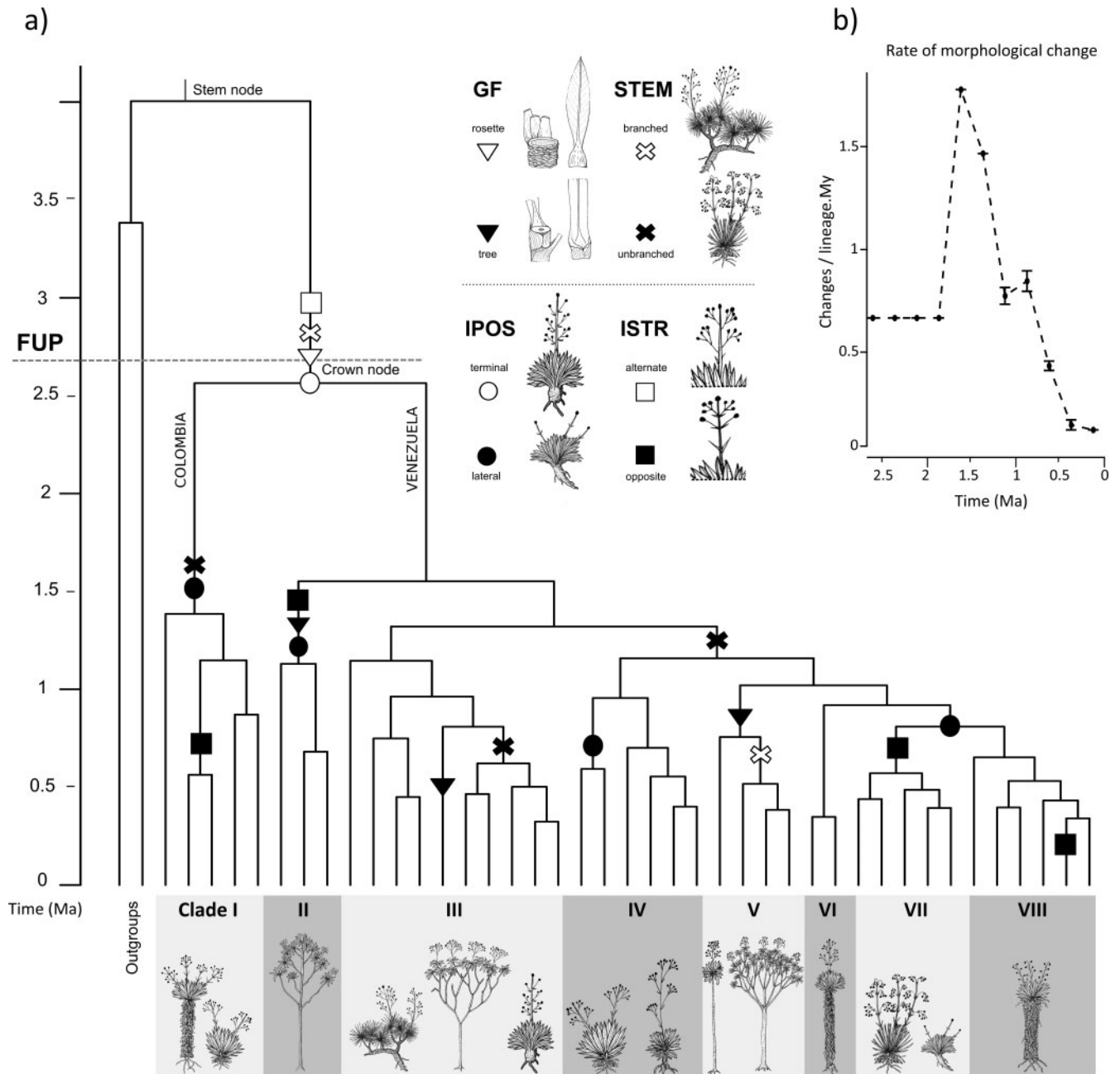
FIGURE 7.   a) Reconstruction of ancestral states of four morphological traits in Espeletiinae. The horizontal dashed line represents the age for the final uplift of the paramo ecosystem (FUP, Gregory-Wodzicki 2000. GF = growth-form; STEM = branching type; IPOS = inflorescence position; ISTR = inflorescence structure. b) Temporal changes in the rate of morphological evolution in Espeletiinae.

growth-form evolved from rosette ancestors three times: in the ancestors of clades II, V and within clade III (GF trait, triangles). The opposite trend, from trees to rosettes, does not seem to have occurred within Espeletiinae. The branching in the ancestral rosette-form seems to have been lost once in the Colombian Andes (clade I) and twice in the Venezuelan Andes; in the common ancestor of clades IV–VIII and within clade III (STEM trait, crosses). The reverse trend did occur once, i.e. the re-evolution of branched forms from unbranched ancestors in clade V. According to these results, with the exception of *L. spectabilis*, the members of clade V are the consequence of two major morphological changes: from unbranched rosettes to branched trees. The change from terminal to lateral inflorescences has occurred four times in the group: in Colombia, and in the Venezuelan clades II, VII + VIII, and within clade IV (IPOS trait, circles). The opposite trend, from lateral to terminal inflorescences, has not occurred. Altogether with the analysis of trait STEM, these results indicate that the monocarpic reproductive strategy of clades IV, VI, some members of clade III and *L. spectabilis* (clade V) is the

result of the loss of branching in lineages that already possessed a terminal inflorescence and not the other way around. Finally, the transition from alternate to opposite inflorescences occurred at least four times, once in Colombia (a number that might increase upon analysis of more Colombian taxa), and in Venezuelan clades II, VII and *Espeletia semiglobulata* (clade VIII) (ISTR trait, squares).

Taken together, these results indicate complex, repetitive, and rarely bi-directional trends in the morphological evolution of Espeletiinae. This analysis also showed that after a period of relative stasis following the start of the Espeletiinae radiation, the rate of morphological change per lineage and per million years increased considerably between 1.75 and 1.25 Ma, after which it decreased rather steadily towards the minimum rates observed in the last 0.5 Myr (Fig. 7b).

DISCUSSION

We developed here a new approach of phylogenomic reconstruction that uses informative data identified from alignments of shotgun sequences without the need for a reference genome in the group of interest and that requires neither annotation nor synteny for the identification of conserved regions across taxa. Although with some important differences, it is conceptually close to the whole-genome resequencing of multiple individuals from a population to a low depth (lcWGR) described in Fuentes-Pardo and Ruzzante (2017). It is also close to the one of Schwartz et al. (2015), who used similar methods and available NGS data to resolve the phylogenetic relationships within apes and mammals. In practical terms, our approach is analogous to Genotyping by Sequencing (GBS) and Restriction-site Associated DNA sequencing (RAD-seq) in that it provides a large number of short and anonymous nuclear markers (Cronn et al. 2012). Compared to RAD-seq or GBS approaches, ours has the advantages of requiring minimum sample preparation before sequencing, producing full organelle sequences, yielding data more evenly scattered across the nuclear genome and, possibly, generating more data. On the other hand, it requires more sequencing effort in order to keep reliable read coverage levels, particularly for nuclear markers, as well as more intensive bioinformatics methods during the post-sequencing locus selection steps. From a phylogenetic perspective, our approach also shares the rationale of studies based on GBS and RAD-seq, which is that even if the phylogenetic signal of individual markers may be small, the phylogenetic signal of the sum can indeed be very large, provided that markers are conserved across taxa, i.e. orthologous (Rubin et al. 2012; Eaton and Ree 2013).

The issues of orthology and paralogy are central to molecular phylogenetics and deserve further comment. Whenever possible, orthology and paralogy should be assessed using a reference genome (Koonin 2005), but

several theoretical and empirical studies have shown that, in the absence of this resource, the clustering by sequence similarity provides a simple and generally effective way to identify orthologous markers (Rubin et al. 2012; Dalquen et al. 2013; Gabaldón and Koonin 2013; but see Smith and Pease 2017). However, orthology assessment by similarity clustering is clearly not a perfect method given that some sequences may be similar but not orthologous (e.g., gene duplication), or orthologous but not similar (e.g., sequence divergence). Nevertheless, the importance of these processes is expected to increase with divergence time. Hence, as long as the study deals with groups of recently diverged taxa such as Espeletiinae, clustering by similarity should perform particularly well (Rubin et al. 2012). Yet, in order to overcome biases in similarity scores and to maximize the number of presumably orthologous markers in our dataset we excluded all species' sequences matching more than one contig, species' contigs matching more than one metacontig and metacontigs with more than 5% ambiguous sites. Further work will however be needed in order to assess thoroughly the issues of orthology and paralogy, as well as the overall benefits/shortfalls of this approach and its performances in regard with different phylogenetic scales, genome sizes and complexities.

Phylogeny and Gene Conflicts

Nuclear concordance.—In spite of the important conceptual differences underlying concatenation and coalescent-based approaches, with the only exception of the position of *L. spectabilis* in the *ASTRAL* tree, no evidence for important conflicts among nuclear genes due to ILS or inter-specific gene flow could be detected in relation with the phylogenetic relationships among major clades of Espeletiinae. In the case of ILS, it should be noted that the incidence of this phenomenon decreases with phylogenetic branch lengths measured in coalescent time units (i.e., T/Ne, where T is the number of generations and Ne is the effective population size). Thus, the small number of generations between speciation events expected in Espeletiinae due to their young age does not necessarily imply a short branch in coalescent time units provided that their effective population sizes are small (Degnan and Rosenberg 2009). There are no available estimations of effective population sizes in Espeletiinae, but given that an important number of species are only known from very small areas or even single paramo summits (Cuatrecasas 2013), it is possible that actual values of this parameter be indeed quite small in these plants. In regard with gene flow, our study did find significant evidence of this process among the taxa examined, but most cases represent recent events of nuclear introgression between closely related and sympatric species within terminal clades, which is not expected to generate much phylogenetic conflicts for the relationships between clades or among the older nodes of the Espeletiinae tree (Fig. 5). Similar results have been recently found

in phylogenomic analyses of diversification in the genera *Heuchera* (Folk et al. 2017) and *Diplostephium* (Vargas et al. 2017), in which low levels of phylogenetic signal conflict were observed among nuclear markers despite frequent inter-specific hybridization. More generally, the ensemble of these results agree with a growing body of empirical and theoretical studies which found that concatenation and coalescent-based methods can recover the true tree topology under low to moderate levels of ILS and/or gene flow (Bayzid and Warnow 2013; Mirarab et al. 2014; Stephens et al. 2015; Solís-Lemus et al. 2016).

Our finding of a low impact of gene flow on the phylogenetic inference of main clades of Espeletiinae should not be interpreted as implying a diminished importance of hybridization in the evolution of this group. Hybridization can clearly be a strong evolutionary force with profound impacts in the processes of speciation (Mallet 2007; Mavarez and Linares 2008) and adaptive radiation (Seehausen 2004; Schwarzer et al. 2012). In Espeletiinae, inter-specific hybridization may be in part responsible for the important morphological variation observed within species known to hybridize frequently, e.g. *E. schultzii*, *E. schultzii* var. *mucurubana*, *C. spicata* (Berry et al. 1988; Cuatrecasas 2013). Moreover, some species exhibiting mixtures of traits from different morphologic groups such as *L. spectabilis* (a rosette with closed sheaths), but also *T. chardonii* (a tree with open sheaths) and *C. thyrsiformis* (a *Coespeletia*-like rosette with *Espeletia*-like inflorescences), also appear to be genetically admixed in our analyses (Fig. 5). Further work will however be necessary to clarify whether this is the result of hybrid speciation or post-speciation events of introgressive hybridization.

Cyto-nuclear discordance.—In contrast with the phylogenetic signal homogeneity among nuclear markers discussed above, this study showed that plastid and nuclear phylogenies are highly discordant in Espeletiinae (SH-test, $P < 0.001$). In fact, as can be seen in the tanglegram in Supplementary Fig. S5 available on Dryad, the backbones of the two phylogenies and even a large fraction of their terminal nodes are very different. Yet, strong cyto-nuclear discordances as the observed here are relatively common among plants (Bruun-Lund et al. 2016; Schmickl et al. 2016; Liu et al. 2017), including some paramo taxa (e.g. *Diplostephium*, Vargas et al. 2017), and they are also usually attributed to ILS and reticulate evolution processes such as interspecific hybridization (Rieseberg and Soltis 1991). Under the light of the *ASTRAL/SNaQ* results discussed above, and given that the effective population size of plastids is ¼ of the nuclear genome due to its haploid nature and maternal inheritance, ILS does not seem to be a likely explanation for the cyto-nuclear discordance observed in Espeletiinae. Furthermore, ILS is a random process that should not necessarily lead to the geographic footprint in the phylogenetic clustering of plastid

markers that we demonstrated here. Indeed, the joint results of the phylogenetic analyses, *SNaQ* and isolation-by-distance indicate that the cyto-nuclear discordance in Espeletiinae is probably caused by differences between plastid and nuclear markers in regard with the strength of barriers to inter-specific gene flow. Thus, plastids seem to be exchanged between geographically close taxa regardless of their morphologic or nuclear phylogenetic distance, while nuclear gene flow appears mostly restricted to sympatric and closely related species within major phylogenetic groups (Figs. 3 and 5). Such heterogeneity in the permeability of barriers is not unexpected given the important differences in size, composition and inheritance of plastid and nuclear genomes (Rieseberg and Soltis 1991; Tsitrone et al. 2003). For instance, relatively common biological phenomena such as sex-biased hybridization, unidirectional backcrossing and cytoplasmic male sterility can allow the complete transfer of uniparentally-inherited markers across species with little or no traces of concomitant nuclear exchange (Tsitrone et al. 2003; Folk et al. 2017). Since all these phenomena depend on inter-specific hybridization, they generate patterns of cyto-nuclear discordance that are usually structured geographically, e.g. Acosta and Premoli (2010). Although no analysis has been specifically designed to study these processes in Espeletiinae, the data in Berry et al. (1988) indicates that some inter-specific crosses can be completely sterile in one direction and partially fertile in the other, and we have observed exclusively unidirectional backcrossing in a *E. schultzii/C. moritziana* hybrid zone (Mavárez et al. unpublished data).

A Phylogenomic Analysis of the Radiation of Espeletiinae

The dataset of nuclear contigs generated in this study, consisting of 1877 metacontigs with >31,000 informative sites, provided enough polymorphic genomic data to recover for the first time a fully-resolved phylogeny of major morphological groups of Espeletiinae (Fig. 4). Some of these relationships agree with previous views of the evolution of the group, while others appear to be rather in conflict. For example, in agreement with the recent work of Diazgranados and Barber (2017), Espeletiinae experienced two geographically delimited and mostly disconnected radiations in Venezuela and Colombia, respectively, instead of an initial radiation in Venezuela followed by multiple trans-cordilleran dispersal events towards Colombia, as proposed by Smith and Koch (1935) and Cuatrecasas (1986, 2013). More specifically, our study shows that dispersal across the Táchira depression has indeed occurred quite rarely in Espeletiinae, always from Venezuela to Colombia, each time involving lineages with the tree growth-forms and occurring at relatively lower elevations. These events may have concerned *L. neriifolius* (min. 1800 m), *L. occultus* (min. 2700 m) and *T. chardonii* (min. 2200 m), which can sometimes thrive in the warmer and wetter conditions of the upper Andean forest habitats (Supplementary Fig. S4 available on Dryad)

(Cuatrecasas 2013). However, the actual number of dispersal events could rise once all the phylogenetic relationships within Venezuelan clade V are fully resolved, because some of its members are distributed in the Colombian Cordillera Oriental (*L. tamanus*), in its northward extension the Sierra de Perija (*L. divisoriensis*) or in the Sierra Nevada de Santa Marta (*L. occultus glossophyllus*).

Our study also confirms a result implicit in Diazgranados and Barber (2017), which is that there have been no trans-cordilleran dispersal events in the Cuatrecasas' genera *Espeletia* or *Espeletiopsis.* These taxa represent independent evolutions in each country of the polycarpic caulescent rosette with opposite or alternate inflorescences, respectively. This result explains one of the most puzzling mysteries unsolved by previous scenarios of evolution in Espeletiinae (Cuatrecasas 1986, 2013): how did the dispersal-limited paramo-endemic *Espeletia* and *Espeletiopsis* manage to cross the >30 km of land below 2000 m in the Táchira depression between the southern border of the Venezuelan Cordillera de Mérida and the eastern border of the Colombian Cordillera Oriental which, even during glacial maxima, never sheltered a paramo habitat? The answer appears now simple: they never did.

The two radiations of Espeletiinae in Colombia and Venezuela differ notably in terms of morphological diversity—much larger in Venezuela, a pattern that Diazgranados and Barber (2017) suggested to result from a supposed older age of the Venezuelan Espeletiinae lineages. Although our species sampling remains partially incomplete, our results suggest otherwise, as the onsets of the radiation in each country appear quite simultaneous (Fig. 6b), with the possibility of the Colombian radiation becoming even older following the analysis of more Espeletiinae taxa from this country. This agrees with thermochronological studies of the exhumation of the Northern Andes that show similar ages and uplift dynamics for the Venezuelan Cordillera de Mérida and the Colombian Cordillera Oriental (Bermúdez et al. 2010; Hoorn et al. 2010).

Rapid Radiation During the Pleistocene   We estimated the age for the crown node of Espeletiinae at 2.3 (2.0–2.6) Ma (Fig. 6a), which is both much more recent and precise than the previous estimate of 4.0 (2.4–5.9) Ma provided by Madriñan et al. (2013). According to this result, the diversification dynamics of Espeletiinae appears to have been driven most likely by Pleistocene climatic oscillations than by Neogene tectonics. Further support for this idea comes from our finding that, in the better-sampled Venezuelan radiation, no diversification occurred for about the first million years of its evolution, during a period in which glacial dynamics were dominated by the low-amplitude cycles of the Early Pleistocene. Later, the rate of diversification increased dramatically in this radiation, when climate variability changed to the large cycles characteristic of the Late Pleistocene (Hönisch et al. 2009). On the

other hand, even though the Andes had reached current elevations by about 2.7 Ma, cold-adapted high-elevation paramo lineages (clades VII and VIII) appeared in the Venezuelan Cordillera de Mérida only by 0.65 Ma, when glacial cycles reached the lowest temperatures. Moreover, although still preliminary, the estimated age for the crown node of the Colombian Espeletiinae (1.38 Ma), composed entirely of caulescent rosettes, seems quite synchronous with that of the most recent common ancestor of all Venezuelan caulescent rosettes (1.32 Ma, clades III-VIII, Fig. 6b). This suggests that similar time-coordinated processes could have been at play at the onset of the radiations in the two cordilleras. Further studies specifically designed to understand the drivers of diversification in species-rich paramo taxa such as *Lupinus* (Nevado et al. 2016), *Hypericum* (Nürk et al. 2013), *Diplostephium* (Vargas et al. 2017) or the Colombian Espeletiinae will however be necessary before a link with Pleistocene climatic instability or Miocene/Pliocene tectonics can be firmly established.

Models of diversification rates on phylogenetic trees show that with an average rate $r = 3.09$ spp/Myr, the Venezuelan radiation of Espeletiinae ranks among the fastest evolving plant groups on Earth, just behind *Dianthus* ($r = 2.20$–7.60, Valente et al. 2010) and Andean *Lupinus* ($r = 1.56$–5.21, Hughes and Atchison 2015), and perhaps other Andean taxa that deserve further studies (Madriñán et al. 2013; Lagomarsino et al. 2016). However, there is ample rate heterogeneity among clades in the Venezuelan radiation, suggesting that an average rate does not necessarily reflects changes in the diversification dynamics neither through space nor across lineages. For example, the diversification rate in the unbranched polycarpic caulescent rosettes in the highest elevations of the Cordillera de Mérida (clades VII–VIII) is faster than of trees in clades II or V, and than of monocarpic caulescent rosettes in clades III and IV. This suggests an interesting link between growth-form, elevation and diversification rate in Espeletiinae, a pattern also observed in the Andean *Lupinus* and that has been attributed to a possible ecological release resulting from reduced inter-specific competition in recently formed habitats in the high Andes (Hughes and Atchison 2015; Nevado et al. 2016).

Morphological Evolution

The ancestor of Espeletiinae.—Both Smith and Koch (1935) and Cuatrecasas (1986) suggested that the ancestor of Espeletiinae was an evergreen perennial shrub/tree exhibiting the xeromorphic foliage with spiral phyllotaxy characteristic of the subtribe, likely similar to the trees *Libanothamnus*. Our study suggests, however, an alternative scenario, with a branched rosette with imbricated open/flat leaf sheaths as the crown ancestor of the subtribe (Fig. 7a). This result is coherent with our finding of a Pleistocene onset for the radiation of Espeletiinae, because an older age would imply that the evolution of this rosette ancestor pre-dated the paramo habitat, in contradiction with the

known distributions and ecological restrictions of this growth-form in the subtribe and other Northern Andes plants (Cuatrecasas 2013). The synchronicity between the evolution of Espeletiinae and the rising of paramo conditions, together with the fact that the diversification of the subtribe started from a rosette, provide additional support to the widely accepted notion that this growth-form, and particularly the caulescent type, represents a key morphological trait or an ensemble of traits that triggered the adaptive radiation of Espeletiinae (Monasterio and Sarmiento 1991). This is also likely the case of other caulescent rosette taxa across the World, e.g. *Dendrosenecio* and *Lobelia* in the East African highlands, *Argyroxiphium*, *Wilkesia* and the lobeliads in the Hawaiian mountains (Givnish et al. 2009; Givnish 2010).

Trees.—Our study revealed that there have been three independent evolutions of the tree growth-form in the subtribe; two from branched rosette ancestors in clade II (*Carramboa*) and within clade III (*Tamania*), and one from unbranched rosettes in clade V (*Libanothamnus*) (Fig. 7a). In opposition to the dominant ideas about the evolution of this group, the reverse trend from trees to rosettes does not seem to have occurred. Furthermore, given that rosettes are distributed at significantly higher elevations and with a marked preference for the colder/drier conditions of the paramo habitat, the evolution of the tree growth-form in the subtribe appears to be associated with a downward colonization of the warmer/wetter conditions in the forest-paramo ecotone and upper Andean forest, in which tree forms are dominant (Supplementary Fig. S4 available on Dryad). This trend contradicts another common sentiment regarding the evolution of Espeletiinae and other paramo plants, which claims that habitat colonizations have occurred from lowlands to highlands (Cuatrecasas 1986; Van der Hammen and Cleef 1986). A somewhat similar trend has been proposed recently for the evolution of the Neotropical genus *Diplostephium*, in which lower-elevation tree species appear to have evolved from higher elevation shrubby ancestors (Vargas and Madriñán 2012).

Rosettes.—This study also highlighted three independent evolutions of the unbranched caulescent rosettes in Espeletiinae; one in the Colombian radiation (clade I) and two in the Venezuelan; within clade III and in the ancestor of clades IV–VIII (Fig. 7a). This conclusion is unlikely to change upon analysis of more Colombian rosette taxa, because all species in this country belong to the unbranched caulescent type and appear to be monophyletic (Diazgranados and Barber 2017; this study). The numerous parallelisms between the giant caulescent rosettes in Espeletiinae of the high Andes of Colombia and Venezuela in regard to, for instance, plant size and form, foliar pubescence and insulation of the stem with dead leaves, provide strong support to the idea of a convergent evolution

of this growth-form in response to the cold, dry, and irradiated conditions that prevail at high elevations in both countries. These traits confer protection against freezing of the apical meristem, reduced water loss through transpiration, increased volume of water stored in the large pith of an insulated stem, etc. (Smith 1981; Monasterio and Sarmiento 1991). A similar idea has been recently put forward to explain correlations between geographic origin and metabolic fingerprints in Colombian and Venezuelan *Espeletia* caulescent rosettes by Padilla-González et al. (2017). Analogous arguments were considered by Loeuille et al. (2015) to explain the independent evolution of three lineages of caulescent rosettes in the subtribe Lychnophorinae (Asteraceae) endemic to the herbaceous/shrubby vegetation known as *campos rupestres* in the Brazilian Cerrado, which the authors describe as an stressful habitat with paramo-like conditions such as intense solar radiation, important fluctuations in daily temperature and seasonal water scarcity, in addition to frequent fires.

We note however that some of the tallest caulescent rosettes in Espeletiinae, such as *Ruilopezia marcescens* (clade IV), *R. paltonioides* (clade III) and *Espeletiopsis purpurascens* (clade I) (Fig. 1b), do not live in the proper paramo but at lower elevations in the forest-paramo ecotone. The gigantism in these species with naked stems, reduced foliar pubescence and comparatively smaller piths cannot be considered as a protection against freezing and/or droughts, but perhaps as the response of the shade-intolerant Espeletiinae (Baruch and Smith 1979) to selection favoring rapid vertical growth under the conditions of strong competition for light typical of forest habitats.

Monocarpy.—In theory, monocarpy in Espeletiinae could be achieved through two paths: by a change from lateral to terminal inflorescences in unbranched species or by the loss of branching in species with terminal inflorescences. Our study only found examples of the latter, which produced at least four independent evolutions of the monocarpy in clades III, IV, V (*L. spectabilis*) and VI (Fig. 7a). A detailed analysis of the causes of changes of reproductive strategy in Espeletiinae lies beyond the scope of this study, but the pattern of repeated evolution suggests that some form of selection could have been at play. Indeed, monocarpic species in the subtribe exhibit several traits expected to be favored by natural selection as part of a "monocarpic syndrome", such as large inflorescences (Cuatrecasas 2013), big seeds (Smith 1981) and synchronized reproduction (Mavarez pers. obs). On the other hand, monocarpy in Espeletiinae is common among unbranched rosettes but rare or inexistent among branched rosettes and trees. This association between branching pattern and reproductive strategy holds for all known long-lived monocarpic rosettes, which has been suggested to represent a predisposition for monocarpy in plants with this growth-form (Young and Augspurger 1991; Jabaily and Sytsma 2013). Further

studies will therefore be necessary in order to depict whether monocarpy in Espeletiinae evolves as a direct consequence of selection or as a by-product of other processes, e.g. loss of branching.

A New Systematic Proposal for Espeletiinae

Our study shows that, as currently circumscribed, the three largest genera in the Cuatrecasas' classification system of Espeletiinae, *Espeletia*, *Ruilopezia* and *Espeletiopsis*, representing about 84% of the subtribe, are clearly polyphyletic. Furthermore, the genus *Coespeletia* is paraphyletic in regard to *E. semiglobulata*, while the two monotypic genera *Tamania* and *Paramiflos* are nested within one group of *Ruilopezia* and the Colombian *Espeletiopsis*, respectively. Only the genera *Carramboa* and *Libanothamnus* have been resolved as monophyletic. As a consequence, the classification system with eight-genera of Espeletiinae (Cuatrecasas 1976, 1986, 2013) appears largely artificial and must be modified. One such modification, following a "splitter approach", could be: 1) preserve the monophyletic genera *Carramboa* and *Libanothamnus*; 2) modify the definition of *Coespeletia* so that it can accept *Espeletia semiglobulata*; 3) create a new genus for the Venezuelan *Espeletiopsis*; 4) create a new genus for the Venezuelan *Espeletia*; 5) decide whether *Tamania* is preserved or merged into *Ruilopezia*; and 6) split *Ruilopezia* in three or four genera depending on the decision taken in (5). Additional changes might be necessary for the Colombian radiation, particularly in regard to the status of *Paramiflos*. The problem with this modification is that it would require a considerable amount of time and effort in order to identify the synapomorphies that define the new taxa, as well as the creation of many new name combinations.

We propose a more workable and stable solution, adopting the view that only the genus *Espeletia* Mutis ex Bonpl. should be recognized in the subtribe, with the other seven genera considered as (heterotypic) synonyms. Given that the majority of the species were originally described under the genus *Espeletia* (~130 spp.), the change proposed here implies the resurrection of species names discarded by Cuatrecasas (1976, 2013) and about a dozen new combinations. Moreover, several influential workers on the ecology and physiology of the group did not accept Cuatrecasas' nomenclature and a large number of important studies have been published using only *Espeletia* (Smith 1981; Goldstein et al. 1984, 1985; Berry and Calvo 1989; Monasterio and Sarmiento 1991). Following Rauscher (2002), to acknowledge for the large specific and morphological diversity in the genus, the ensemble of the species can also be collectively referred to as the "*Espeletia* complex".

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.m1t32.

REFERENCES

Aberer A.J., Kobert K., Stamatakis A. 2014. ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31:2553–2556.

Abrahamczyk S., Souto-Vilarós D., Renner S.S. 2014. Escape from extreme specialization: passionflowers, bats and the sword-billed hummingbird. Proc. R. Soc. B Biol. Sci. 281:20140888.

Acosta M.C., Premoli A.C. 2010. Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). Mol. Biol. Evol. 54:235–242.

Antonelli A., Nylander J.A.A., Persson C., Sanmartín I. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. Proc. Natl. Acad. Sci. USA. 106:9749–9754.

Antonelli A., Sanmartín I. 2011a. Why are there so many plant species in the Neotropics? Taxon 60:403–414.

Antonelli A., Sanmartín I. 2011b. Mass extinction, gradual cooling, or rapid radiation? Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. Syst. Biol. 60:596–615.

Baruch Z., Smith A.P. 1979. Morphological and physiological correlates of niche breadth in two species of *Espeletia* (Compositae) in the Venezuelan Andes. Oecologia. 38:71–82.

Bayzid M.S., Warnow T. 2013. Naive binning improves phylogenomic analyses. Bioinformatics. 29:2277–2284.

Bell C.D., Calderon G., Gonzalez L., Scholz A. 2015. Resolving Relationships within Valerianaceae (Dipsacales): New Insights and Hypotheses from Low-Copy Nuclear Regions. Systematic Botany. 40:327–335.

Bermúdez M.A., Kohn B.P., van der Beek P.A., Bernet M., O'Sullivan P.B., Shagam R. 2010. Spatial and temporal patterns of exhumation across the Venezuelan Andes: Implications for Cenozoic Caribbean geodynamics. Tectonics. 29:TC5009.

Berry P., Beaujon S., Calvo R.N. 1988. La hibridización en la evolución de los frailejones (*Espeletia*, Asteraceae). Ecotropicos. 1:11–24.

Berry P.E., Calvo R.N. 1989. Wind Pollination, Self-Incompatibility, and Altitudinal Shifts in Pollination Systems in the High Andean genus *Espeletia* (Asteraceae). American Journal of Botany. 76: 1602–1614.

Bruun-Lund S., Clement W.L., Kjellberg F., Rønsted N. 2016. First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). Mol. Biol. Evol. 109: 93-104.

Buerki S., Forest F., Alvarez N., Nylander J.A.A., Arrigo N., Sanmartín I. 2010. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. Journal of Biogeography. 38:531–550.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.

Cronn R., Knaus B.J., Liston A., Maughan P.J., Parks M., Syring J.V., Udall J. 2012. Targeted enrichment strategies for next-generation plant biology. American Journal of Botany. 99:291–311.

Cuatrecasas J. 1976. A new subtribe in the Heliantheae (Compositae): Espeletiinae. Phytologia. 35:43–61.

Cuatrecasas J. 1986. Speciation and radiation of the Espeletiinae in the Andes. In: Vuilleumier F., Monasterio M., editors. High Altitude Tropical Biogeography. Oxford University Press. p. 267–303.

Cuatrecasas J. 1996. Clave provisional de las especies del género *Espeletiopsis* Cuatrec. (Espeletiinae, Compositae). Anales del Jardín Botánico de Madrid. 54:370–377.

Cuatrecasas J. 2013. A Systematic Study of the Subtribe Espeletiinae (Heliantheae, Asteraceae). The New York Botanical Garden Press.

Dalquen D.A., Altenhoff A.M., Gonnet G.H., Dessimoz C. 2013. The impact of gene duplication, insertion, deletion, lateral gene transfer and sequencing error on orthology inference: a simulation study. PLoS ONE. 8:e56925.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution. 24:332–340.

Diazgranados M. 2012. A nomenclator for the frailejones (Espeletiinae Cuatrec., Asteraceae). PhytoKeys. 16:1–52.

Diazgranados M., Barber J.C. 2017. Geography shapes the phylogeny of frailejones (Espeletiinae Cuatrec., Asteraceae): a remarkable example of recent rapid radiation in sky islands. PeerJ. 5:e2968.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012a. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution. 29:1969–1973.

Drummond C.S., Eastwood R.J., Miotto S.T.S., Hughes C.E. 2012b. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. Syst. Biol. 61:443–460.

Eaton D.A.R., Ree R.H. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 62:689–706.

Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Biol. Evol. 94:447–462.

Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. R. Soc. B Biol. Sci. 279:1300–1309.

FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.

Flantua S., Hooghiemstra H. 2018. Historical connectivity and mountain biodiversity. In: Hoorn C., Antonelli A., Perrigo A., editors. Mountains, climate, and biodiversity. p. 171–186.

Folk R.A., Mandel J.R., Freudenstein J.V. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. Syst. Biol. 66:320–337.

Fuentes-Pardo A.P., Ruzzante D.E. 2017. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. Mol. Ecol. 26: 5369-5406.

Gabaldón T., Koonin E.V. 2013. Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14:360–366.

Givnish T.J. 2010. Giant lobelias exemplify convergent evolution. BMC Biol. 8:3.

Givnish T.J., Millam K.C., Mast A.R., Paterson T.B., Theim T.J., Hipp A.L., Henss J.M., Smith J.F., Wood K.R., Sytsma K.J. 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). Proc. R. Soc. B Biol. Sci. 276:407–416.

Goldstein G., Meinzer F., Monasterio M. 1984. The role of capacitance in the water balance of Andean giant rosette species. Plant Cell Environ. 7:179–186.

Goldstein G., Rada F., Azocar A. 1985. Cold hardiness and supercooling along an altitudinal gradient in andean giant rosette species. Oecologia 68:147–152.

Goslee S.C., Urban D.L. 2007. The ecodist package for dissimilarity-based analysis of ecological data. J. Stat. Softw. 22:1–19.

Gregory-Wodzicki K.M. 2000. Uplift history of the Central and Northern Andes: a review. Geol. Soc. Am. Bull. 112:1091–1105.

Hejase H.A., Liu K.J. 2016. A scalability study of phylogenetic network inference methods using empirical datasets and simulations involving a single reticulation. BMC Bioinformatics 17:422.

Hooghiemstra H., Wijninga V.M., Cleef A.M. 2006. The Paleobotanical record of Colombia: Implications for Biogeography and Biodiversity. Ann. Mo. Bot. Gard. 93:297–324.

Hoorn C., Wesselingh F.P., Steege ter H., Bermudez M.A., Mora A., Sevink J., Sanmartín I., Sanchez-Meseguer A., Anderson C.L., Figueiredo J.P., Jaramillo C., Riff D., Negri F.R., Hooghiemstra H., Lundberg J., Stadler T., Särkinen T., Antonelli A. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330:927–931.

Hoorn C., Wesselingh F.P., Steege ter H., Bermudez M.A., Mora A., Sevink J., Sanmartín I., Sanchez-Meseguer A., Anderson C.L., Figueiredo J.P., Jaramillo C., Riff D., Negri F.R., Hooghiemstra H., Lundberg J., Stadler T., Särkinen T., Antonelli A. 2011. Origins of Biodiversity. Authors reply. Science 331:399–400.

Hönisch B., Hemming N.G., Archer D., Siddall M., McManus J.F. 2009. Atmospheric carbon dioxide concentration across the mid-Pleistocene transition. Science 324:1551–1554.

Hughes C.E., Atchison G.W. 2015. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. New Phytol. 207:275–282.

Jabaily R.S., Sytsma K.J. 2013. Historical biogeography and life-history evolution of Andean *Puya* (Bromeliaceae). Bot. Rev. 171:201–224.

Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Koonin E.V. 2005. Orthologs, paralogs, and evolutionary genomics. Ann. Rev. Genet. 39:309–338.

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Lagomarsino L.P., Condamine F.L., Antonelli A., Mulch A., Davis C.C. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). New Phytol. 210:1430–1442.

Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Liu Y., Li D., Zhang Q., Song C., Zhong C., Zhang X., Wang Y., Yao X., Wang Z., Zeng S., Wang Y., Guo Y., Wang S., Li X., Li L., Liu C., McCann H.C., He W., Niu Y., Chen M., Du L., Gong J., Datson P.M., Hilario E., Huang H. 2017. Rapid radiations of both kiwifruit hybrid lineages and their parents shed light on a two-layer mode of species diversification. New Phytol. 215: 877–890.

Loeuille B., Semir J., Lohmann L.G., Pirani J.R. 2015. A phylogenetic analysis of Lychnophorinae (Asteraceae: Vernonieae) based on molecular and morphological data. Syst. Bot. 40:299–315.

Luebert F., Weigend M. 2014. Phylogenetic insights into Andean plant diversification. Front. Ecol. Evol. 2:27.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.

Madriñán S., Cortés A.J., Richardson J.E. 2013. Páramo is the world's fastest evolving and coolest biodiversity hotspot. Front. Genet. 4:192.

Mallet J. 2007. Hybrid speciation. Nature 446:279–283.

Mavarez J., Linares M. 2008. Homoploid hybrid speciation in animals. Mol. Ecol. 17: 4181–4185.

Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548.

Mirarab S., Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31:i44-i52.

Monasterio M., Sarmiento L. 1991. Adaptive radiation of *Espeletia* in the cold andean tropics. Trends Ecol. Evol. 6:387–391.

Nevado B., Atchison G.W., Hughes C.E., Filatov D.A. 2016. Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. Nat. Commun. 7:12384.

Nürk N.M., Scheriau C., Madriñán S. 2013. Explosive radiation in high Andean *Hypericum*-rates of diversification among New World lineages. Front. Genet. 4:175.

Padilla-González G.F., Diazgranados M., Da Costa F.B. 2017. Biogeography shaped the metabolome of the genus *Espeletia*: a phytochemical perspective on an Andean adaptive radiation. Sci. Rep. 7:8835.

Panero J.L., Jansen R.K., Clevinger J.A. 1999. Phylogenetic relationships of subtribe Ecliptinae (Asteraceae: Heliantheae) based on chloroplast DNA restriction site data. Am. J. Bot. 86:413–427.

Popescu A.-A., Huber K.T., Paradis E. 2012. Ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. Bioinformatics 28:1536–1537.

Quinlan A.R., Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Rauscher J.T. 2000. Molecular systematics of the *Espeletia* complex: evidence from nrITS sequences on the evolution of an Andean adaptive radiation [Ph.D thesis]. Washington University, Saint-Louis, USA.

Rauscher J.T. 2002. Molecular phylogenetics of the *Espeletia* complex (Asteraceae): evidence from nrDNA ITS sequences on the closest relatives of an Andean adaptive radiation. Am. J. Bot. 89: 1074–1084.

Rieseberg L.H., Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants. 5:65–84.

Roch S., Steel M. 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100:56–62.

Rubin B.E.R., Ree R.H., Moreau C.S. 2012. Inferring phylogenies from RAD sequence data. PLoS One 7:e33394.

Rull V. 2011a. Neotropical biodiversity: timing and potential drivers. Trends Ecol. Evol. 26:508–513.

Rull V. 2011b. Origins of biodiversity. Science 331:398–399.

Sánchez-Andrade A. 2005. Filogenética molecular de los Espeletiinae, una radiación adaptativa andina [M.Sc. thesis]. Universidad de los Andes, Bogotá, Colombia.

Sánchez-Baracaldo P., Thomas G.H. 2014. Adaptation and convergent evolution within the *Jamesonia-Eriosorus* complex in high-elevation biodiverse andean hotspots. PLoS One 9: e110618.

Sanderson M.J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.

Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593.

Schmickl R., Liston A., Zeisek V., Oberlander K., Weitemier K., Straub S.C.K., Cronn R.C., Dreyer L.L., Suda J. 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). Mol. Ecol. Resour. 16(5): 1124–1135.

Schmidt-Lebuhn A.N., Kessler M., Hensen I. 2007. Hummingbirds as drivers of plant speciation? Trends Plant Sci. 12:329–331.

Schwartz R.S., Harkins K.M., Stone A.C., Cartwright R.A. 2015. A composite genome approach to identify phylogenetically informative data from next-generation sequencing. BMC Bioinformatics 16:193.

Schwarzer J., Swartz E.R., Vreven E., Snoeks J., Cotterill F.P., Misof B. & Schliewen U.K. 2012. Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. Proc. Biol. Sci. 279:4389–4398.

Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19:198–207.

Shimodaira H., Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution. 16:114–116.

Smith A.C., Koch M.F. 1935. The genus *Espeletia*: a study in phylogenetic taxonomy. Brittonia 1:479–530.

Smith A.P. 1981. Growth and population dynamics of *Espeletia* (Compositae) in the Venezuelan Andes. Smithsonian Contributions to Botany 48. Smithsonian Institution Press.

Smith S.A., Pease J.B. 2017. Heterogeneous molecular processes among the causes of how sequence similarity scores can fail to recapitulate phylogeny. *Brief. Bioinform.* 18:451–457.

Solís-Lemus C., Yang M., Ané C. 2016. Inconsistency of species tree methods under gene flow. Syst. Biol. 65:843–851.

Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34:3292–3298

Springer M.S., Gatesy J. 2016. The gene tree delusion. Mol. Biol. Evol. 94:1–33.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stephens J.D., Rogers W.L., Heyduk K., Cruse-Sanders J.M., Determann R.O., Glenn T.C., Malmberg R.L. 2015. Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. Mol. Biol. Evol. 85:76–87.

Struwe L., Haag S., Heiberg E., Grant J.R. 2009. Andean speciation and vicariance in Neotropical *Macrocarpaea* (gentianaceae-Helieae). Ann. Mo. Bot. Gard. 96:450–469.

Tremetsberger K., Gemeinholzer B., Zetzsche H., Blackmore S., Kilian N., Talavera S. 2013. Divergence time estimation in Cichorieae (Asteraceae) using a fossil-calibrated relaxed molecular clock. Org. Divers. Evol. 13:1–13.

Tsitrone A., Kirkpatrick M., Levin D.A. 2003. A model for chloroplast capture. Evolution 57:1776–1782.

Uribe-Convers S., Tank D.C. 2015. Shifts in diversification rates linked to biogeographic movement into new areas: An example of a recent radiation in the Andes. Am. J. Bot. 102:1854–1869.

Valente L.M., Savolainen V., Vargas P. 2010. Unparalleled rates of species diversification in Europe. Proc. R. Soc. B Biol. Sci. 277:1489–1496.

Van der Hammen T., Cleef A.M. 1986. Development of the High Andean Páramo flora and vegetation. In: Vuilleumier F., Monasterio M., editors. High altitude tropical biogeography. New York: Oxford University Press. p. 153–201.

Vargas O.M., Madriñán S. 2012. Preliminary phylogeny of *Diplostephium* (Asteraceae): speciation rate and character evolution. Lundellia 15:1–15.

Vargas O.M., Ortiz E.M., Simpson B.B. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). New Phytol. 214:1736–1750.

von Hagen K.B., Kadereit J.W. 2003. The diversification of *Halenia* (Gentianaceae): ecological opportunity versus key innovation. Evolution 57:2507–2518.

Young T.P., Augspurger C.K. 1991. Ecology and evolution of long-lived semelparous plants. Trends Ecol. Evol. 6:285–289.

Zachos J., Pagani M., Sloan L., Thomas E. and Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292:686–693.

Zerbino D.R., Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.