EUROPEAN JOURNAL OF ENTOMOLOGY
ISSN (online): 1802-8829
http://www.eje.cz
Eur. J. Entomol. **121**: 124–133, 2024
doi: 10.14411/eje.2024.015
ORIGINAL ARTICLE

Evidence of a South American origin for the *Drosophila repleta* group (Diptera: Drosophilidae)

Andrea E. Acurio[1, 2]

[1] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

[2] Laboratory of Evolutionary Genetics, Institute of Entomology, Biology Centre, Czech Academy of Sciences, Branišovská 31, 370 05 České Budějovice, Czech Republic; e-mail: andrea.acurio@entu.cas.cz

Key words. *repleta* lineage, evolutionary relationships, systematics, phylogeny, biogeography

Abstract. The *Drosophila repleta* group is one of the most widely used model systems for studying adaptation and speciation. Traditionally, five subgroups are recognized within the *repleta* group: *fasciola*, *hydei*, *mercatorum*, *repleta* and *mulleri*. A sixth subgroup, *inca*, was the last to be defined. The *inca* species subgroup includes three species: *Drosophila inca*, *Drosophila huancavilcae* and *Drosophila yangana*, all of which are endemic to Ecuador and Peru. Previous molecular phylogenetic studies have been inconclusive regarding the geographic location, time and mode of diversification of lineages within the *repleta* group. By applying a phylogenetic and biogeographical analysis of 54 taxa belonging to the *repleta*, *nannoptera*, *atalaia* and *virilis* groups, the aim of this study was to: (1) determine the relationships between the *inca* subgroup and the other five subgroups within the *repleta* group, (2) improve the unresolved branching and low supports of the early divergent lineages in the *repleta* group phylogeny and (3) estimate the geographic and temporal context of the early divergence within the *repleta* group. Based on these findings, it is proposed that the *Drosophila repleta* group first diversified during the mid-late Miocene, most likely following the uplift of the Northwestern Andes.

INTRODUCTION

For over a century, studies on ecological adaptation, evolution and speciation have used the *repleta* species group of the subgenus *Drosophila* Fallén as a model system (Sturtevant, 1915; Wharton, 1942; Vilela, 1983; Wasserman, 1992; Oliveira et al., 2012; Stefanini et al., 2021). As in Throckmorton's definition of 1975, this group belongs to the *virilis-repleta* radiation, one of the three main radiations in the *Drosophila* subgenus. With 106 species (O'Grady & DeSalle, 2018), the *repleta* group is one of the most speciose in the genus. Many of these species are cactophilic, so they inhabit dry zones and deserts throughout the Americas. Six subgroups are recognized within the *repleta* species group: *fasciola*, *hydei*, *mercatorum*, *repleta*, *mulleri* and *inca,* with the last one being the most recent to be defined (Rafael & Arcos, 1989).

The most recent revision of the phylogenetic relationships within the *repleta* group (Oliveira et al., 2012), estimated that the diversification of the crown group started around 16 million years ago (Mya) and concludes it is of a monophyletic origin. However, this study did not include representatives of the *inca* subgroup; therefore, the phylogenetic position of the *inca* subgroup within the *repleta* group remains unclear.

Historically, the Mexican Trans-Volcanic Region was considered the centre of diversification for the *repleta* group (Patterson & Stone, 1952; Throckmorton, 1975). Oliveira et al. (2012), suggest that the centre of diversification of the *repleta* group could have been South America and associated with the radiation of its cactus hosts, although the authors did not provide statistical support for this hypothesis. Another study by Morales-Hojas & Vieira (2012) analysed the patterns of diversification across the subgenus *Drosophila* and supports the monophyly of the *repleta* group. Despite the substantial contribution of these two studies, the geographical origin of the group, as well as the phylogenetic position of the *inca* subgroup remains unknown.

As in many cases, a major issue in understanding the diversification of the *repleta* group was the bias regarding the localities sampled for *Drosophila* collections. For many years, efforts to collect samples of *Drosophila* focused on regions in North and Central America with an emphasis on arid zones in Mexico (Sturtevant, 1921; Patterson & Mainland, 1945; Oliveira et al., 2005). In contrast, there are relatively few collections from Andean countries of South America (Oliveira et al., 2012). As a result, the *Drosophila* fauna is well known in North America, while new species and records are still being described in South America

(Acurio & Rafael, 2009b; Figuero et al., 2012; Acurio et al., 2013; Llangarí-Arizo & Rafael, 2020).

Preliminary morphological and cytological data support the idea that the *inca* subgroup diverged from the *repleta* group very early (Rafael & Arcos, 1989; Mafla & Romero, 2009). The *inca* subgroup comprises three species known to inhabit only Northwestern South America. *Drosophila huancavilcae* (Rafael & Arcos, 1989) and *Drosophila yangana* (Rafael & Vela, 2003) are endemic to valleys in the Ecuadorian Andes with very narrow distributions, whereas *Drosophila inca* (Dobzhansky & Pavan, 1943) has a less restricted distribution and is widespread in the inter-Andean desert valleys in Ecuador and Peru (Dobzhansky & Pavan, 1943; Acurio & Rafael, 2009a). Other members of the *repleta* group with restricted Andean distributions are *D. huaylasi* (Fontdevila et al., 1990) of the *mulleri* subgroup, endemic to Ecuador and Peru and *D. guayllabambae* (Rafael & Arcos, 1989) of the *hydei* subgroup endemic to Ecuador.

Understanding the earliest events in speciation remains a challenge for evolutionary biology; therefore, the *inca* subgroup is a unique opportunity to assess questions regarding the centre of origin of the *repleta* species group. Here, the phylogenetic relationships of the six subgroups inside the *repleta* group were analysed using a multilocus dataset of gene coding regions of two nuclear and three mitochondrial genes. Phylogenetic analysis, estimates of time of divergence and biogeographic analysis indicate that the earliest diversification events in the *Drosophila repleta* group occurred in the mid-late Miocene, most likely as a result of the uplift of the Northwestern Andes.

MATERIAL AND METHODS

Taxon sampling

To obtain a representative sample of the *repleta* species group, the material analysed included 54 taxa from the *fasciola*, *hydei*, *mercatorum*, *repleta* and *mulleri* subgroups, all species of the *inca* subgroup (*inca*, *huancavilcae* and *yangana*), along with five outgroup taxa thought to be closely related: *Drosophila machalilla* (*atalaia* group), *D. virilis* (virilis group), and *D. wassermani*, *D. nannoptera* and *D. acanthoptera* (*nannoptera* species group).

To obtain samples of the *inca* subgroup, fieldwork was carried out at localities in Northern, Central and Southern Ecuador. The sampled localities were established using published information on the distribution of these species (Rafael & Arcos, 1989; Rafael & Vela, 2003; Acurio & Rafael, 2009a). The method of collection is that previously described by Acurio et al. (2019), with the only change being the addition of rotting prickly pear (*Opuntia ficus-indica*) cladodes as bait. Once individuals were identified in the field using morphological characters, isofemale strains were established by adding a piece of fresh *Opuntia* cladode to the culture medium. Individuals from each isofemale strain were killed and stored in ethanol at –20°C for molecular analysis.

Molecular analyses

Genomic DNA was extracted from three flies per isofemale strain using a modified Cetyltrimethylammonium Bromide (CTAB) protocol (Calderón-Cortés et al., 2010). Partial regions of two nuclear genes (*marf* and *sinA*) and three mitochondrial genes (*ND2, COI, COII*) were amplified using PCR and gene-specific primers (Table 1**)**. These markers were selected because they provide a good phylogenetic signal at the deepest taxonomic level within *Drosophila* (Oliveira et al., 2008, 2012). Gene regions of interest were amplified using 0.4 μ_M oligonucleotides, 1 U DNA Taq DNA polymerase (Roche, Germany) per 35 µL reaction volume, 2 m_M $MgCl_2$ and 200 μ_M dNTP and ~ 25 ng of genomic DNA. The reactions were carried out under standard thermocycle conditions. The PCR products were purified using NucleoSpin Extract II kit (Clontech Laboratories Inc., USA). When necessary, the PCR product was cloned using the StrataClone PCR Cloning Kit (Stratagene, La Jolla, USA) according to the manufacturer's instructions. Sequencing was carried out at Macrogen Inc. (Seoul, South Korea) using Sanger sequencing with gene-specific primers or T7 and SP6 universal primers. Chromatograms were compiled using Geneious v 5.0.4 (Biomatters, Ltd., Auckland, New Zealand). The sequences generated in this study were deposited in GenBank under the accession numbers KC011819-KC011843. The identifiers for all the sequences used in this study are listed Table S1.

Phylogenetic reconstruction

DNA sequences were aligned using MAFFT (Katoh et al., 2009). The generated dataset, contained homologous genomic regions, comprising 2,462 aligned sites, including 147 constant characters, 468 parsimony-uninformative characters and 1847 parsimony informative (75%) sites.

Phylogenetic inference was determined using Maximum Parsimony (MP), Maximum Likelihood (ML) and Bayesian Inference (BI). MP analysis was done using MESQUITE 2.74 (Maddison & Maddison, 2010). A search for the most parsimonious tree was done based on the tree-length criterion using Subtree Pruning and Regrafting (SPR). A consensus tree was obtained from the trees using Majority Rule Consensus, considering tree weights with a frequency of clades of 0.5 in unrooted trees. ML analysis was done using SATé software (Liu et al., 2012), which is a multilocus analysis. Models of nucleotide substitution were selected using Akaike Information Criteria, calculated in jModelTest 2.1.3 (Posada, 2008; Darriba et al., 2012). The model of nucleotide evolution that was the best fit for the data set was General Time Reversible (GTR). Alignment and merger steps were performed separately for each locus and tree inferences made on a single tree for all loci. BI analysis was performed using BEAST v.1.7.5 (Drummond et al., 2012), locus specific substitution models and molecular clocks for nuclear and mitochondrial partitions. The nucleotide substitution model for the mitochondrial partition (including *ND2*, *COI*, *COII*) was (GTR), with empirical base frequencies plus the Gamma model of site heterogeneity (four cat-

Table 1. Summary of the partially coded regions used in this study and reference for each primer.

Genes	Abbreviation	Length	Primer design reference
Cytochrome C oxidase subunit I	COI	367	(Oliveira et al., 2005)
Cytochrome C oxdase subunit II	COII	706	(Liu & Beckenbach, 1992)
Mitochondrial-ubiquinone oxidoreductase chain	NADH	782	(Oliveira et al., 2005)
Mitochondrial assembly regulatory factor	Marf	552	(Bonacum et al., 2001)
Seven in Absentia	SinA	397	(Bonacum et al., 2001)

egories). The nuclear partition (*SinA*, *Marf*) had the same settings, but without a codon partition. The same concatenated dataset was used in all three (MP, ML, BI) analyses.

Estimates of time of divergence

To estimate the time of divergence of the *inca* clade, calibration points were chosen from Oliveira et al. (2012): the TMRCA of the split between *D. mojavensis* and *D. arizonae* (1.83 Mya) and the TMRCA of the *repleta* group (16.3 Mya). Priors were assumed to follow a normal distribution with mean and standard deviation, according to Oliveira et al. (2012). Due to differences in the rate variation of mitochondrial and nuclear genes (Moriyama & Powell, 1997), the analysis was run on a concatenated data set with two partitions, nuclear and mitochondrial. The clock models were linked and a common strict clock was assumed for all partitions. A starting tree was randomly generated during the Yule process. Four independent runs using Markov Chain Monte Carlo (MCMC) chains with 10 million generations were carried out and sampled every 1000 generations. The resulting output file was processed using TreeAnnotator v.1.5.3 (Drummond & Rambaut, 2007) with a burn-in parameter setting of 1000. Effective sample sizes were reviewed using Tracer v. 1.4 to ensure that they were greater than 500 for each parameter. Independent runs were compared to ensure that they converged on the same posterior distribution and were stationary.

Biogeographical analyses

The historical biogeographic ranges of the *Drosophila repleta* group were reconstructed using BioGeoBEARS (Matzke, 2013) in R (R Development Core Team, 2011). First, a three-state presence-absence matrix was constructed that represented the known distribution of each species in North, Central and South America. The historical ranges were then estimated under two different unconstrained models (1) Dispersal-Extinction-Cladogenesis (DEC) (Ree & Smith, 2008) and (2) Dispersal-Extinction-Cladogenesis-Jump (DEC+*j*) (Matzke, 2013), first implemented using maximum likelihood. A comparison of these two models allowed an assessment of the relative roles of range expansion, range extinction and founder events (defined in this model as the acquisition of a new range without the parent lineage having already spread into it) in the evolution of the ranges in this group. Model performance was assessed using the likelihood ratio test. The overall likelihood scores, *d*, *e* and *j* parameters for the two biogeographic models were as follows: (1) DEC = LnL = –87.8, *d* = 0.02, *e* = 0, *j* = 0 and (2) DEC+j = LnL = –81, *d* = 0.014, *e* = 0, *j* = 0.076. The DEC+*j* model performed significantly better than the DEC model (LRT *pval* = 0.0002). The difference between the two biogeographic models tested is that in addition to allowing range expansions and extinctions (*d* and *e*), the DEC+*j* model also allows for founding events (*j*). Both models support no role of range extinction, but the addition of *j* parameter in the DEC+*j* model appears to result in a better fit to the data. Reconstructions were conditioned in absolute time using a chronogram from BEAST.

RESULTS

The *inca* subgroup was the earliest clade to diverge

Phylogenetic analysis of five partial coding regions (*marf*, *sinA*, *ND2, COI, COII*) using three methods of phylogenetic reconstruction (MP, ML and BI) generated trees with highly congruent topologies (Fig. S1). The *inca* subgroup was retrieved as monophyletic and supported by different methods of phylogenetic reconstruction. In the *inca* clade (Fig. 1), *Drosophila huancavilcae* is a sister taxon of *D. inca* and both species are closely related to *D. yangana*. The *inca* subgroup was the earliest clade to diverge within the *repleta* group. In the phylogeny eight clades with good statistical support (ML Bootstrap > 0.75, ML, Shimodaira – Hasegawa values on MP and BI Posterior Probabilities = 1) were retrieved as follows: *nannoptera* group, *inca* subgroup, *eremophila* complex, *fasciola* subgroup, *hydei* subgroup, *anceps* complex, *repleta* subgroup and *mulleri* subgroup.

Table 2. Estimates (with 95% Credibility Interval, CI) of the times of divergence (Mya) for the main nodes recovered in the phylogenetic analysis of the *repleta* group using BEAST.

Dated nodes	Mean node age (Mya) + 95%CI
inca subgroup	13.11 (11.53–14.63)
fasciola subgroup	10.73 (9.28–12.29)
eremophila complex	7.8 (6.87–8.99)
hydei subgroup	10.61 (9.32–11.93)
anceps complex	12.82 (11.7–13.95)
repleta subgroup	12.12 (11.07–10.10)
mulleri subgroup	11.38 (12.32–14.08)
repleta group	17.00 (16.35–17.85)

Timing and place of first diversifications within the *repleta* group

The estimated times of divergence for the main nodes recovered from the phylogenetic analysis are listed in Table 2. The splitting of the *repleta-nannoptera* groups was dated as about 21.94 million years ago (Mya). The earliest diversification event within the *repleta* group was estimated to be 17 Mya. In the *inca* clade, *D. yangana* diverged from its sister species 13 Mya, whereas the split between *D. huancavilca* and *D. inca* occurred around 9.39 Mya.

The result of the biogeographical analysis is shown in Fig. 2. Biogeographical assessment of 54 taxa, including the *repleta*, *atalaia*, *virilis* and *nannoptera* groups indicate that the *repleta* group evolved in South America (prob = 0.66). There was a relatively small amount of support for the origin of the group in North America (prob = 0.17) or both North and South America (prob = 0.17; Fig. 2). Divergence of the *inca* species subgroup also occurred in South America (prob = 1.0).

DISCUSSION

The key role of the Northwestern Andes uplift

Based on the hypothesis proposed by Oliveira et al., (2012) the evolutionary radiation of the *repleta* species group was facilitated by transitions among the major cactus host lineages. However no evidence was found for a causal relationship between a stable or expanding population and host plant shifts from prickly -pear cactus to columnar cacti (Pfeiler, 2019). Host plant switches are likely to have had a role in the diversification of the *repleta* group; however, based on timing, biogeographic reconstructions and the distribution data used in this study, the uplift of the Northwestern Andes, a region that currently includes Colombia, Ecuador and Peru, is thought to be the context for the earliest diversification events in the *repleta* group.

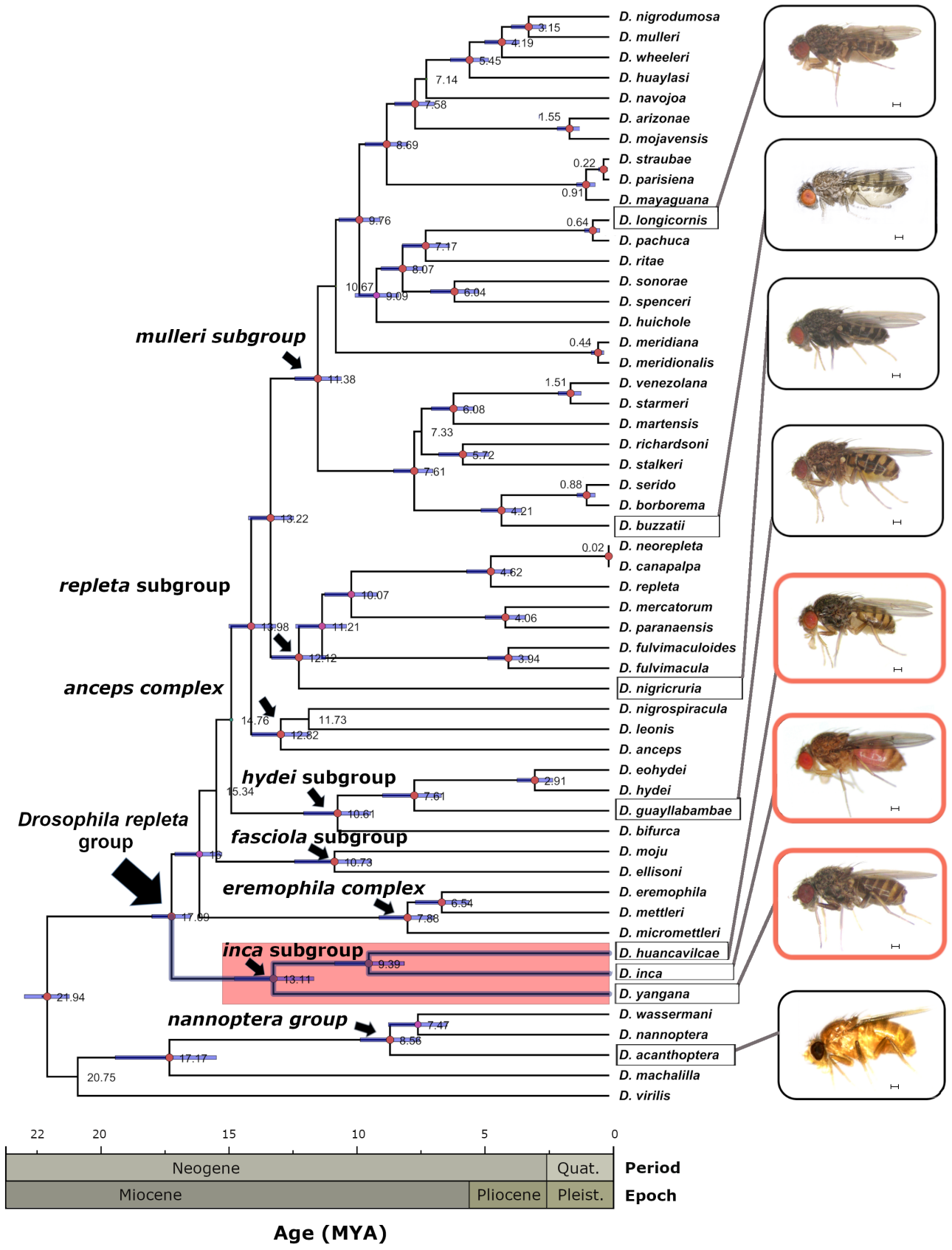

Fig. 1. Phylogenetic tree based on the analysis of five partially coded regions of 54 species of *Drosophila* from the *repleta*, *nannoptera*, *atalaia* and *virilis* groups. The *inca* clade is shown in red, estimated dates of divergence (Mya) are indicated on nodes, blue bars denote confidence intervals and geological scale is shown at the bottom. The pictures are of: *Drosophila longicornis* (female), *D. buzzatii* (female), *D. nigricruria* (male), *D. guayllabamabae* (male), *D. huancavilcae* (male), *D. yangana* (male) and *D. acanthoptera* (male).

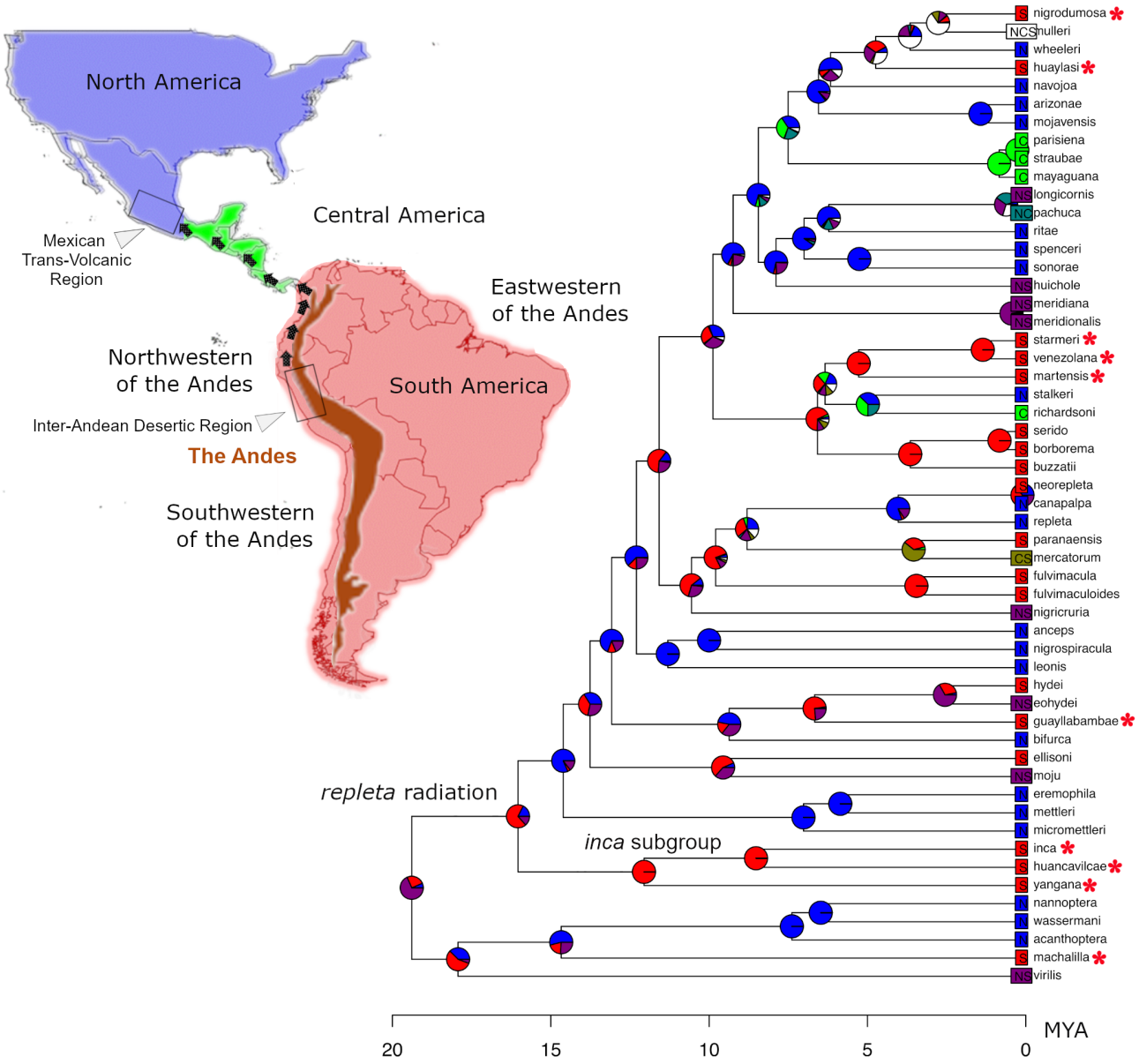

Fig. 2. Historical range reconstructions for the *Drosophila repleta* radiation. Ranges are in blue for North America (N), light-green for Central America (C), dark-green for Central and South America (CS), aquamarine for North and Central America (NC), red for South America (S), purple for North and South America (NS) and white for North, Central and South America (NCS). Red asterisk indicates species endemic to Northwestern South America. The arrows on the map shows the most probable direction of migration based on the biogeographical analysis.

Drosophila guayllabambae, *Drosophila martensis*, *Drosophila venezolana*, *Drosophila starmeri*, *Drosophila huaylasi* and *Drosophila nigrodumosa* are other species endemic to the Northwestern Andes, in addition to the species from the *inca* subgroup. The phylogenetic hypothesis presented here is based not only on preliminary morphological and cytological data for the *inca* subgroup (Rafael & Arcos, 1989; Rafael & Vela, 2003; Mafla & Romero, 2009), but also the evolutionary relationships with other early divergent clades such as the *eremophila* complex, *fasciola* and *hydei* subgroups that were included in earlier phylogenetic studies (Oliveira et al., 2005, 2012; O'Grady & DeSalle, 2018).

To determine whether the *eremophila* and *anceps* complexes are better classified as species subgroups is dependent on a taxonomic revision of the *mulleri* subgroup. A previous phylogenetic assessment of the *mulleri* subgroup (Durando et al., 2000) described it as polyphyletic, suggesting that the close relationships between the *mulleri*, *repleta* and *mercatorum* subgroups might account for the intermingling of species in the phylogeny. These authors point out that the *mulleri* subgroup includes several species that could not be classified into other *repleta* subgroups. The close evolutionary relationships previously reported for the *atalaia*, *nannoptera* and *repleta* groups (Acurio et al., 2013, 2019; Lang et al., 2014), are consistent with the phylogenetic hypothesis presented here.

The Andes uplift had a key role in shaping biodiversity patterns in animals and plants in South America, particularly in mutualistic species relationships (Hoorn et al.,

2022). The biogeography of the cacti also appears to have been influenced by the Andean uplift. According to Arakaki et al. (2011), most of the extant diversity in cacti was generated throughout the mid to late Miocene and during the Pliocene, resulting in three main centres of cactus diversity and endemism: Mexico, central Andes and Brazil. The temporal concordance between major diversification events within cacti and the crown diversification of the *repleta* group date from the same period as the Andean uplift in the middle Miocene (Oliveira et al., 2012; Morales-Hojas & Vieira, 2012, and results from this study).

The addition of new taxa improved phylogenetic accuracy

The results of this study indicate that the addition of taxa presumed to have diverged earlier, such as the *inca* subgroup, based on morphological synapomorphies and cytological data, may enhance the robustness of phylogenetic assessment and correct the problematic branching in the phylogeny. Oliveira et al. (2012) based on ten molecular markers infer a phylogenetic hypothesis with good statistical support (MP bootstrap > 98%, ML bootstrap = 100%, BI posterior probability = 1) for three of the five subgroups examined. In the current study, by using five molecular markers and a larger number of taxa, good statistical support (MP values > 0.75; ML values > 90; BI posterior probability = 1) was obtained for the six subgroups analysed. In large datasets, the phylogenetic signal is increased because of the large number of taxa (Hedtke et al., 2006), which results in better estimates of the evolution of characters (Townsend & Leuenberger, 2011; Nabhan & Sarkar, 2012) and improves the precision of estimates of the times of divergence using Bayesian molecular dating analysis (Soares & Schrago, 2015).

Timing of the initial divergence events in the *repleta* group

Various estimates of the time of divergence are proposed for the most recent common ancestor (TMRCA) of the *virilis-repleta* radiation, which vary depending on the model used and the number of points chosen to calibrate the molecular clock. Obbard et al. (2012), estimate that the ancestor of the *repleta* group and related species groups split approximately 12 ± 3 Mya. However, most estimates are significantly older and are in general agreement. Oliveira et al. (2012) propose a date of 26 ± 6 million years ago (Mya), Morales-Hojas & Vieira (2012) provide two estimates based on different calibration strategies of 23 ± 4 and 31 ± 4 Mya, and Russo et al. (2013) estimate the split to be 27 ± 5 Mya. The estimated time of divergence between the *nannoptera* group and *D. machalilla* is 7.4–16.9 Mya, according to Lang et al. (2014). The ancestor of the *nannoptera* species group may have migrated over the isthmus from South America, since this period coincides with the closing of the Panama isthmus (Montes et al., 2012). The results reported here are in line with the diversification pattern of drosophilids proposed by Russo et al. (2013), who hypothesize that the radiation of the family Drosophilidae began during the Palaeogene, peaked during the Miocene and was associated with the exploitation of the newly diversified fleshy fruits of angiosperms. Members of the *repleta* group occupy a great diversity of habitats, ranging from wet tropical forests to temperate environments (Vilela & Bächli, 1990; Acurio & Rafael, 2009b), but most *repleta* specialize on cacti (Oliveira et al., 2012; Guillén et al., 2014; O'Grady & DeSalle, 2018).

The Huancabamba region, a centre of endemism and species richness

The present Andes may be divided into several discrete areas that have undergone a unique evolution and have a complex geological history. Based on several studies (Daly, 1989; Ramos, 2009; Hoorn et al., 2022), the general tectonic evolution of the oceanic rocks in the western sector of the Northern Andes should have different ages, starting in the Jurassic-Early Cretaceous (185–130 Mya), continuing in the Late Cretaceous (90–80 Mya), the latest Cretaceous-Paleogene (45–35 Mya) and the Miocene to the Present (15–0 Mya). Geological changes can result in barriers and filters that affect the migration of biota. Andean uplift dated to the mid-Miocene (Gregory-Wodzicki, 2000; Capitanio et al., 2011) had a major role in the distributions of a variety of animals, such as rodents (Reig, 1986), butterflies (Descimon, 1986) and amphibians (Duellman & Wild, 1993).

The distribution of the *inca* clade is restricted to desert areas in inter-Andean valleys of Southern Ecuador and Northern Peru with xerophytic vegetation that is particularly rich in columnar cacti and agaves. In Ecuador, this region is reported to have the highest diversity and endemism of Cactaceae, including 13 genera (81.3%) and 28 species (58.3%) (Loaiza & Morrone, 2011). *Drosophila inca* and *D. yangana* occur in the Huancabamba region, which is an Andean centre of endemism and species richness (Young & Reynel, 1997). Studies on birds and amphibians by Vuilleumier (1969) and Duellman & Wild (1993) propose that the high level of endemism and species richness is associated with the dynamic and changing environment presented by the growing Andes as they rose to their current altitude. The orogenic sequence of the Andes proceeded in a south-to-north fashion (Gregory-Wodzicki, 2000) allowing the southern species to move northwards through Central America. Patterns in endemic bird species in the trans-Andean region (Weir & Price, 2011) indicate that the Andean uplift promoted the build-up of biodiversity in the lowland neotropical fauna through vicariance-based speciation during uplift and dispersal-based speciation following uplift. This pattern also holds for the *Drosophila repleta* group, as there are several species within this group endemic to the lowland tropics east of the Andes, *Drosophila vicentinae*, *Drosophila peninsularis* and members of the *fasciola* species subgroup (Vilela, 1983; Acurio & Rafael, 2009a).

CONCLUSIONS

For the first time, six subgroups within the *Drosophila repleta* group were assessed phylogenetically and biogeographically. The results reported here support the hypothesis that the *inca* subgroup was the earliest clade to diverge

in the *repleta* group. Based on the phylogenetic assessment, estimates of time of divergence and biogeographic analysis, the *Drosophila repleta* group experienced its first episode of diversification in the mid-late Miocene, most likely due to the uplift of the Northwestern Andes. With increasing sampling in South America and whole genome research, it may be possible to address questions regarding the mechanisms of speciation and adaptation in the *repleta* and other *Drosophila* lineages.

ACKNOWLEDGEMENTS. I owe a debt of gratitude to V. Rafael and A. Ruiz for their advice and support throughout this study. Many thanks go to D. CSG Oliveira, P. O'Grady and M. Lang for their feedback on the many rewrites of this manuscript. I am very grateful to K. Roesch-Goodman for helping with biogeographical analyses. The FI-DGR Doctoral Fellowship (FI-B100197) from L'Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya provided funding for this research. The fieldwork on *Drosophila* in Ecuador was only possible because of the assistance of A. Acurio, M. Armas and A. Barragan. Voucher material was deposited in the Museum of Zoology QCAZ and our collection of samples was authorised by research permit 0016-07IC-FAU-DNBAPVS/MA.

REFERENCES

Acurio A. & Rafael V. 2009a: Diversity and geographical distribution of *Drosophila* (Diptera, Drosophilidae) in Ecuador. — *Drosophila Inform. Serv.* **92**: 20–25.

Acurio A. & Rafael V. 2009b: Taxonomic survey of Drosophilidae (Diptera) in the Yasuni National Park, Ecuadorian Amazon. — *Acta Amazon.* **39**: 713–718.

Acurio A.E., Rafael V., Céspedes D. & Ruiz A. 2013: Description of a new spotted-thorax *Drosophila* (Diptera: Drosophilidae) species and its evolutionary relationships inferred by a cladistic analysis of morphological traits. — *Ann. Entomol. Soc. Am.* **106**: 695–705.

Acurio A.E., Rhebergen F.T., Paulus S., Courtier-Orgogozo V. & Lang M. 2019: Repeated evolution of asymmetric genitalia and right-sided mating behavior in the *Drosophila nannoptera* species group. — *BMC Evol. Biol.* **19**: 109, 14 pp.

Arakaki M., Christin P.A., Nyffeler R., Lendel A., Eggli U., Ogburn R.M., Spriggs E., Moore M.J. & Edwards E.J. 2011: Contemporaneous and recent radiations of the world's major succulent plant lineages. — *Proc. Natn. Acad. Sci. U. S. A.* **108**: 8379–8384.

Bonacum J., DeSalle R., O'Grady P., Oliveira D., Wintermute J. & Zilversmit M. 2001: New nuclear and mitochondrial primers for systematics and comparative genomics in Drosophilidae. — *Drosophila Inform. Serv.* **84**: 201–204.

Calderón-Cortés N., Quesada M., Cano-Camacho H. & Zavala-Páramo G. 2010: A simple and rapid method for DNA isolation from xylophagous insects. — *Int. J. Mol. Sci.* **11**: 5056–5064.

Capitanio F.A., Faccenna C., Zlotnik S. & Stegman D.R. 2011: Subduction dynamics and the origin of Andean orogeny and the Bolivian orocline. — *Nature* **480**: 83–86.

Daly M. 1989: Correlations between Nazca-Farallon Plate kinematics and forearc basin evolution in Ecuador. — *Tectonics* **8**: 769–790.

Darriba D., Taboada G.L., Doallo R. & Posada D. 2012: jModelTest 2: more models, new heuristics and parallel computing. — *Nat. Meth* **9**: 772.

Descimon H. 1986: Origins of lepidopteran faunas in the high tropical Andes. In Vuilleumier F. & Monasterio M. (eds): *High Altitude Tropical Biogeography.* Oxford University Press, Oxford, pp. 500–532.

Dobzhansky T. & Pavan C. 1943: Studies on Brazilian species of *Drosophila*. — *Bol. Fac. Filos. Cienc. Letr. Univ. S. Paulo (Biol. Ger.)* **36**: 7–72.

Drummond A.J. & Rambaut A. 2007: BEAST: Bayesian evolutionary analysis by sampling trees. — *BMC Evol Biol*, *7*.

Drummond A.J., Suchard M.A., Xie D. & Rambaut A. 2012: Bayesian phylogenetics with BEAUti and the BEAST 1.7. — *Mol. Biol. Evol.* **29**: 1969–1973.

Duellman W.E. & Wild E.R. 1993: Anuran amphibians from the Cordillera de Huancabamba, northern Peru: systematics, ecology, and biogeography. — *Occas. Pap. Mus. Nat. Hist. Univ. Kansas* **157**: 1–53.

Durando C.M., Baker R.H., Etges W.J., Heed W.B., Wasserman M. & DeSalle R. 2000: Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. — *Mol. Phylogen. Evol.* **16**: 296–307.

Figuero M.L., Rafael V. & Céspedes D. 2012: Grupo *Drosophila asiri* (Diptera, Drosophilidae), un nuevo grupo de especies andinas con la descripción de dos nuevas especies y la redescripción de *Drosophila asiri*. — *Iheringia (Zool.)* **102**: 33–42.

Fontdevila A., Wasserman M., Pla C., Pilares L., De Armengol R., Suyo M. del P., Sanchez A., Vasquez J., Ruiz A. & Garcia J.L. 1990: Description and evolutionary relationships of two species of the *Drosophila mulleri* cluster (Diptera: Drosophilidae). — *Ann. Entomol. Soc. Am.* **83**: 444–452.

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

Guillén Y., Rius N., Delprat A., Williford A., Muyas F., Puig M., Casillas S., Ràmia M., Egea R., Negre B. et al. 2014: Genomics of ecological adaptation in cactophilic *Drosophila*. — *Genome Biol. Evol.* **7**: 349–366.

Hedtke S.M., Townsend T.M. & Hillis D.M. 2006: Resolution of phyloegenetic conflict in large data sets by increasing taxon sampling. — *Syst. Biol.* **55**: 522–529.

Hoorn C., Palazzesi L. & Silvestro D. 2022: Editorial preface to special issue: Exploring the impact of Andean uplift and climate on life evolution and landscape modification: From Amazonia to Patagonia. — *Glob. Planet. Change* **211**: 103759, 6 pp.

Katoh K., Asimenos G. & Toh H. 2009: Multiple alignment of DNA sequences with MAFFT. In Posada D. (ed.): *Bioinformatics for DNA Sequence Analysis. Vol. 537.* Humana Press, Totowa, NJ, pp. 39–64.

Lang M., Polihronakis Richmond M., Acurio A.E., Markow T.A. & Orgogozo V. 2014: Radiation of the *Drosophila nannoptera* species group in Mexico. — *J. Evol. Biol.* **27**: 575–584.

Liu H. & Beckenbach A.T. 1992: Evolution of the mitochondrial cytochrome oxidase II gene among ten orders of insects. — *Mol. Phylog. Evol.* **41**: 31–52.

Liu K., Warnow T.J., Holder M.T., Nelesen S.M., Yu J., Stamatakis A.P. & Linder C.R. 2012: SATé-II: Very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. — *Syst. Biol.* **61**: 90–106.

Llangarí-Arizo L.M. & Rafael V. 2020: Five new species of *Drosophila* (Diptera, Drosophilidae) related with Araceae. — *Iheringia (Zool.)* **110**: e2020012, 13 pp.

Loaiza C. & Morrone J. 2011: Análisis panbiogeográfico de algunas Cactaceae del Ecuador. — *Gayana Bot.* **68**: 220–225.

Maddison W.P. & Maddison D.R. 2010: *Mesquite: A Modular System for Evolutionary Analysis, Version 2.74*. URL: http://mesquiteproject.org

Mafla M.A.B. & Romero E.G. 2009: The heterocromatin of *Drosophila inca*, *Drosophila yangana* and *Drosophila huancavilcae* of the *inca* subgroup, *repleta* group. — *Dros. Inf. Serv*. **92**: 10–15.

Matzke N.J. 2013: BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. *R Package, Version 0.2.1*. URL: https://github.com/nmatzke/BioGeoBEARS.

Montes C., Cardona A., McFadden R., Morón S.E., Silva C.A., Restrepo-Moreno S., Ramírez D.A., Hoyos N., Wilson J. & Farris D. 2012: Evidence for middle Eocene and younger land emergence in central Panama: Implications for Isthmus closure. — *Geol. Soc. Am. Bull.* **124**: 780–799.

Morales-Hojas R. & Vieira J. 2012: Phylogenetic patterns of geographical and ecological diversification in the subgenus *Drosophila*. — *PLoS ONE* **7**(11): e49552, 15 pp.

Moriyama E. & Powell J. 1997: Synonymous substitution rates in *Drosophila*: Mitochondrial versus nuclear genes. — *J. Mol. Evol.* **45**: 378–391.

Nabhan A.R. & Sarkar I.N. 2012: The impact of taxon sampling on phylogenetic inference: A review of two decades of controversy. — *Briefings Bioinform.* **13**: 122–134.

O'Grady P.M. & DeSalle R. 2018: Phylogeny of the genus *Drosophila*. — *Genetics* **209**: 1–25.

Obbard D.J., Maclennan J., Kim K.-W., Rambaut A., O'Grady P.M. & Jiggins F.M. 2012: Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. — *Mol. Biol. Evol*. **29**: 3459–3473.

Oliveira D.C.S.G., O'Grady P.M., Etges W.J., Heed W.B. & DeSalle R. 2005: Molecular systematics and geographical distribution of the *Drosophila longicornis* species complex (Diptera: Drosophilidae). — *Zootaxa* **1069**: 1–32.

Oliveira D.C.S.G., Leonidas M., Etges W.J., O'Grady P.M. & Desalle R. 2008: Species delimitation in the *Drosophila aldrichi* subcluster (Diptera: Drosophilidae) using DNA sequences. — *Zootaxa* **1725**: 37–47.

Oliveira D.C.S.G., Almeida F.C., O'Grady P.M., Armella M.A., DeSalle R. & Etges W.J. 2012: Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. — *Mol. Phylogen. Evol.* **64**: 533–544.

Patterson J.T. & Mainland G.B. 1945: *The Drosophilidae of Mexico*. University of Texas, Austin, TX, 101 pp.

Patterson J.T. & Stone W.S. 1952: *Evolution in the Genus* Drosophila. Macmillan, New York, 610 pp.

Pfeiler E. 2019: Genetic diversity and demographic history in the cactophilic *Drosophila repleta* species group (Diptera: Drosophilidae) in North America inferred from mitochondrial DNA barcodes. — *J. Heredity* **110**: 34–45.

Posada D. 2008: jModelTest: Phylogenetic model averaging. — *Mol. Biol. Evol*. **25**: 1253–1256.

R Development Core Team 2011: R: A language and environment for statistical computing. In R. D. C. Team (ed.): *R Foundation for Statistical Computing. Vol. 1, Issue 2.11.1.* R Foundation for Statistical Computing, p. 409.

Rafael V. & Arcos G. 1989: Subgrupo *inca*, un nuevo subgrupo del grupo *repleta*, con desacripción de *Drosophila huancavilcae* n. sp (Diptera, Drosophilidae). — *Evolución Biológica* **3**: 233–243.

Rafael V. & Vela D. 2003: *Drosophila yangana* sp. nov. un nuevo miembro del grupo *repleta*, subgrupo *inca* (Diptera: Drosophilidae). — *Revta Pontif. Univ. Catol. Ecuador* **71**: 129–139.

Ramos V.A. 2009: Anatomy and global context of the Andes: Main geologic features and the Andean orogenic cycle. — *Mem. Geol. Soc. Am.* **204**: 31–65.

Ree R.H. & Smith S.A. 2008: Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. — *Syst. Biol.* **57**: 4–14.

Reig O.A. 1986: Diversity patterns and differentiation of high Andean rodents. In Vuilleumier F. & Monasterio M. (eds): *High Altitude Tropical Biogeography.* Oxford University Press, Oxford, pp. 404–439.

Russo C.a.M., Mello B., Frazão A. & Voloch C.M. 2013: Phylogenetic analysis and a time tree for a large drosophilid data set (Diptera: Drosophilidae). — *Zool. J. Linn. Soc.* **169**: 765–775.

Soares A.E.R. & Schrago C.G. 2015: The influence of taxon sampling on Bayesian divergence time inference under scenarios of rate heterogeneity among lineages. — *J. Theor. Biol.* **364**: 31–39.

Stefanini M.I., Gottschalk M.S., Calvo N.S. & Soto I.M. 2021: Evolution of male genitalia in the *Drosophila repleta* species group (Diptera: Drosophilidae). — *J. Evol. Biol.* **34**: 1488–1502.

Sturtevant A.H. 1915: A sex-linked character in *Drosophila repleta*. — *Am. Nat.* **49**: 189–192.

Sturtevant A.H. 1921: *The North American Species of* Drosophila. Carnegie Institute, Washington, iv + 150 pp. + 3 pl.

Throckmorton L.H. 1975: The phylogeny, ecology and geography of *Drosophila*. In King R.C. (ed.): *Handbook of Genetics Vol. 3.* Plenum, New York, pp. 421–469.

Townsend J.P. & Leuenberger C. 2011: Taxon sampling and the optimal rates of evolution for phylogenetic inference. — *Syst. Biol.* **60**: 358–365.

Vilela C.R. 1983: A revision of the *Drosophila repleta* species group (Diptera, Drosophilidae). — *Revta Bras. Entomol.* **27**: 1–114.

Vilela C.R. & Bächli G. 1990: Taxonomic studies on Neotropical species of seven genera of Drosophilidae (Diptera). — *Mitt. Schweiz. Entomol. Ges. (Suppl.)* **63**: 332 pp.

Vuilleumier F. 1969: Pleistocene speciation in birds living in the High Andes. — *Nature* **223**: 1179–1180.

Wasserman M. 1992: Cytological evolution of the *Drosophila repleta* species group. In Powell J.R. & Krimbas C.B. (eds): *Inversion Polymorphism in Drosophila.* CRC Press, Boca Raton, FL, pp. 455–541.

Weir J.T. & Price M. 2011: Andean uplift promotes lowland speciation through vicariance and dispersal in *Dendrocincla* woodcreepers. — *Mol. Ecol.* **20**: 4550–4563.

Wharton L.H. 1942: Analysis of the *repleta* group of *Drosophila*. — *Univ. Texas Publ.* **4228**: 23–52.

Young K.R. & Reynel C. 1997: Huancabamba region, Peru and Ecuador. In Davis S.D, Heywood V.H., Herrera-MacBryde O., Villa-Lobos J. & Hamilton A.C. (eds): *Centres of Plant Diversity: A Guide and Strategy for their Conservation, 3: North America, Middle America, South America, Caribbean Islands.* IUCN, Geneva, pp. 465–469.

Received December 7, 2023; revised and accepted March 18, 2024
Published online March 26, 2024

Supplementary Table S1 and Fig. S1 follow on next pages.

Table S1. Complete list of the species of *Drosophila* used in this study and corresponding GenBank entries. Accession numbers for newly generated sequencing data are highlighted in bold type.

Drosophila species		ND2	COII	COI	Marf	SinA
1.	*acanthoptera*	DQ202090	DQ202010	DQ202050	EU3416293	EU341611
2.	*anceps*	JF736133	JF736093	JF736059	JF736208	JF736324
3.	*arizonae*	EU341707	JF736122	EU341676	EU341636	EU341620
4.	*bifurca*	JF736166	JF736130	JF736090	JF736256	JF736378
5.	*borborema*	JF736157	JF736121	JF736081	JF736241	JF736362
6.	*buzzatii*	DQ202091	DQ202011	DQ202051	EU341631	EU341621
7.	*canapalpa*	JF736162	JF736126	JF736086	JF736248	JF736369
8.	*ellisoni*	DQ202092	DQ202012	DQ202052	JF736235	JF736356
9.	*eohydei*	JF736159	JF736124	JF736083	JF736245	JF736366
10.	*eremophila*	DQ202093	DQ202013	DQ202053	JF736249	JF736370
11.	*fulvimacula*	JF736156	JF736120	JF736080	JF736240	JF736361
12.	*fulvimaculoides*	JF736134	JF736094	JF736060	JF736209	JF736325
13.	*guayllabambae*	JF736167	JF736131	JF736091	JF736258	JF736380
14.	***huancavilcae***	**KC011834**	**KC011824**	**KC011819**	**KC011829**	**KC011839**
15.	***huaylasi***	**KC011835**	**KC011825**	**KC011820**	**KC011830**	**KC011840**
16.	*huichole*	DQ202098	DQ202018	DQ202058	JF736257	JF736379
17.	*hydei*	DQ202100	DQ202020	DQ202060	JF736212	JF736328
18.	***inca***	**KC011836**	**KC011826**	**KC011821**	**KC011831**	**KC011841**
19.	*leonis*	JF736136	JF736096	JF736062	JF736214	JF736330
20.	*longicornis*	DQ202101	DQ202021	DQ202061	JF736232	JF736353
21.	***machalilla***	**KC011837**	**KC011827**	**KC011822**	**KC011832**	**KC011842**
22.	*martensis*	JF736160	JF736125	JF736084	JF736247	JF736368
23.	*mayaguana*	DQ202107	DQ202027	DQ202067	EU341634	EU341623
24.	*mercatorum*	JF736155	EU493737	EU493607	JF736239	JF736360
25.	*meridiana*	JF736153	JF736118	JF736078	JF736236	JF736357
26.	*meridionalis*	DQ202110	DQ202030	DQ202070	JF736250	JF736372
27.	*mettleri*	JF736137	JF736097	JF736063	JF736215	JF736331
28.	*micromettleri*	JF736138	JF736098	JF736064	JF736216	JF736332
29.	*mojavensis*	EU493497	EU493738	EU493608	AY437307	EU341624
30.	*mulleri*	DQ202112	DQ202032	EU341625	EU341638	EU341625
31.	*nannoptera*	JF736140	JF736100	JF736066	JF736218	JF736334
32.	*navojoa*	EU341709	EU493739	EU341678	EU341635	EU341626
33.	*neorepleta*	DQ202113	DQ202033	DQ202073	JF736219	JF736335
34.	*nigricruria*	JF736141	JF736101	JF736067	JF736220	JF736336
35.	*nigrodumosa*	EU341710	JF736102	EU341679	EU341633	EU341627
36.	*nigrospiracula*	DQ202114	DQ202034	DQ202074	JF736221	JF736337
37.	*pachuca*	DQ202118	DQ202038	DQ202078	JF736251	JF736373
38.	*paranaensis*	JF736164	JF736128	JF736088	JF736252	JF736374
39.	*parisiena*	JF736142	JF736103	JF736068	JF736222	JF736338
40.	*pavani*	EU493474	JF736115	EU4935832	JF736231	JF736350
41.	*repleta*	EU341711	JF736105	EU341680	EU341630	EU341628
42.	*richardsoni*	JF736144	JF736106	JF736070	JF736224	JF736340
43.	*ritae*	DQ202122	DQ202042	DQ202082	JF736233	JF736354
44.	*serido*	JF736165	JF736129	JF736089	JF736254	JF736376
45.	*sonorae*	DQ202124	DQ202044	DQ202084	JF736225	JF736341
46.	*spenceri*	DQ202127	DQ202047	DQ202087	JF736255	JF736377
47.	*stalkeri*	DQ202128	DQ202048	DQ202088	JF736226	JF736342
48.	*starmeri*	JF736145	JF736107	JF736071	JF736227	JF736343
49.	*straubae*	JF736146	JF736108	JF736072	JF736228	JF736344
50.	*venezolana*	DQ202129	DQ202049	DQ202089	JF736243	JF736364
51.	*virilis*	EU493510	EU493751	EU493622	JF736234	JF736355
52.	*wassermani*	JF736147	JF736109	JF736073	JF736229	JF736345
53.	*wheeleri*	EU341705	JF736110	EU341685	EU341656	EU341616
54.	***yangana***	**KC011838**	**KC011828**	**KC011823**	**KC011833**	**KC011843**

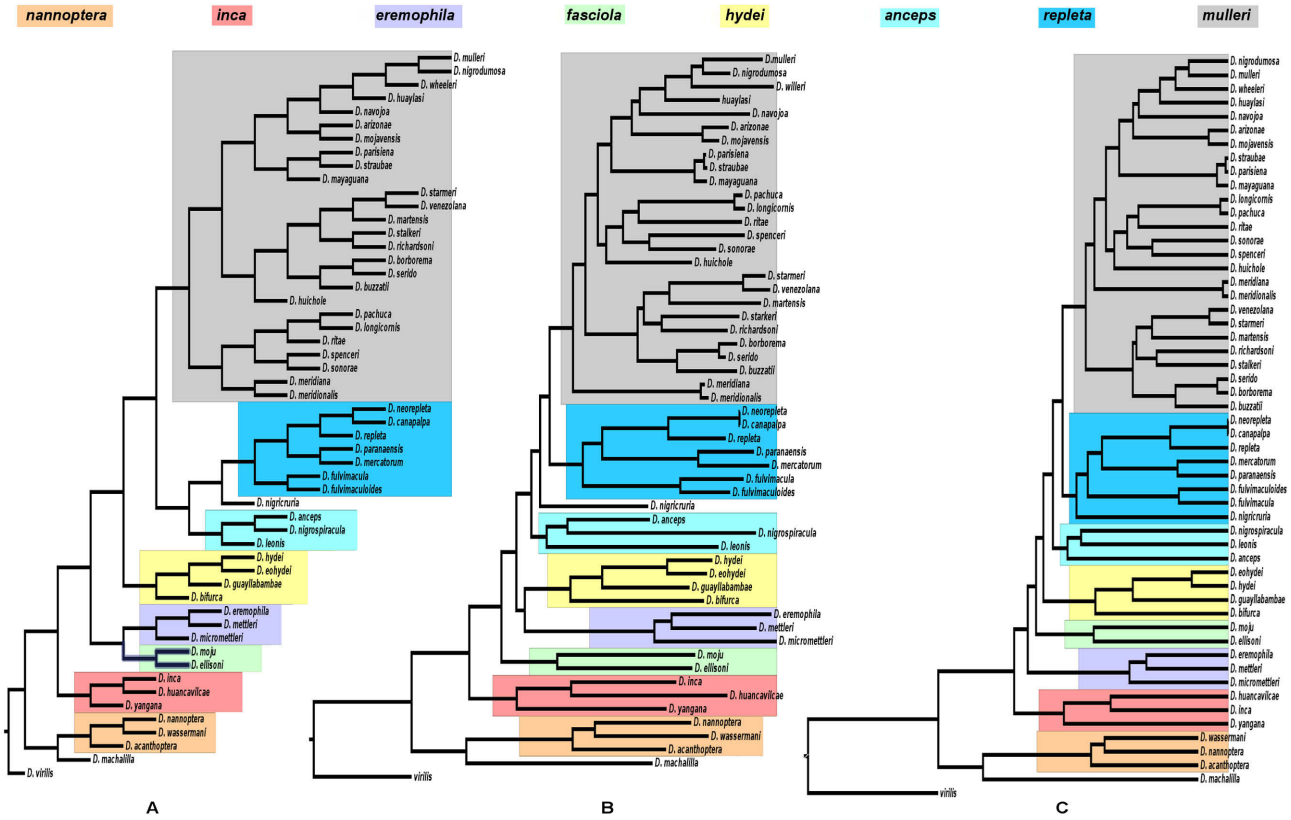

Fig. S1. The Phylogeny of the *Drosophila repleta* group: Phylogenetic trees showing topology based on the concatenated dataset of the 54 species of *Drosophila* analysed in this study. (A) MP analysis done using MESQUITE; (B) ML analysis done using SATé; (C) BS analysis done using BEAST. Colours denote nodes with: Bootstrap values > 0.75 for ML, Shimodaira-Hasegawa values > 0.90 for MP and Posterior Probabilities = 1 for BI.