ARTHROPOD SYSTEMATICS
& PHYLOGENY

77 (2): 239–250
2019

SENCKENBERG

Revised phylogenetic relationships within the *Drosophila buzzatii* species cluster (Diptera: Drosophilidae: *Drosophila repleta* group) using genomic data

Juan Hurtado*,[1,2,#], Francisca Almeida*,[1,2,#], Santiago Revale[3] & Esteban Hasson*,[1,2]

[1] Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina; Juan Hurtado [jhurtado@ege.fcen.uba.ar]; Francisca Almeida [falmeida@nyu.edu]; Esteban Hasson [ehasson@ege.fcen.uba.ar] — [2] Instituto de Ecología, Genética y Evolución de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas, Ciudad Autónoma de Buenos Aires, Argentina — [3] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK — * Corresponding authors; # Contributed equally to this work

Accepted on March 15, 2019.
Published online at www.senckenberg.de/arthropod-systematics on September 17, 2019.
Published in print on September 27, 2019.
Editors in charge: Brian Wiegmann & Klaus-Dieter Klass.

Abstract. The *Drosophila buzzatii* cluster is a South American clade that encompasses seven closely related cactophilic species and constitutes a valuable model system for evolutionary research. Though the monophyly of the cluster is strongly supported by molecular, cytological and morphological evidence, phylogenetic relationships within it are still controversial. The phylogeny of the *D. buzzatii* cluster has been addressed using limited sets of molecular markers, namely a few nuclear and mitochondrial genes, and the sharing of fixed chromosomal inversions. However, analyses based on these data revealed inconsistencies across markers and resulted in poorly resolved basal branches. Here, we revise the phylogeny of the *D. buzzatii* cluster based on a large transcriptomic dataset of 813 kb obtained from four members of this cluster: *D. antonietae*, *D. borborema*, *D. buzzatii* and *D. koepferae*, using the close relative *D. mojavensis* (also a member of the *repleta* group) as outgroup. Our phylogenomic analyses confirm that *D. buzzatii* is sister to the other six members of the cluster and, though incomplete lineage sorting likely obstructs phylogenetic resolution among these six species, allowed us to recover a novel topology. Divergence time estimates date the radiation of the cluster to the recent upper Pleistocene with most speciation events compressed to the last 500,000 years.

Key words. *buzzatii* phylogeny, cactophilic *Drosophila*, divergence time, Pleistocene speciation, phylogenomics, transcriptome assembling.

1. Introduction

The *Drosophila repleta* group is an array of about 100 cactophilic species endemic to the New World (Wasserman 1992; Oliveira et al. 2012). Most species within this group have the ability to utilize decaying cactus tissues as breeding and feeding substrates, a feature that allowed the successful colonization of the American deserts (Wasserman 1982; Ruiz & Heed 1988; Markow & O'Grady 2008). Due to the enormous number of species, specialists created categories, like complex and cluster, that lack formal taxonomic status, that are very helpful to systematize such diversity. Accordingly, the *D. repleta* group is composed by several subgroups, which themselves are further subdivided into complexes and lastly into clusters.

In Central and South America, one of the radiations of the *D. repleta* group gave rise to the *D. buzzatii* complex, which includes the *D. buzzatii*, *D. martensis* and *D. stalkeri* clusters (Ruiz & Wasserman 1993). The first (hereafter the *buzzatii* cluster) consists of the seven cactophilic species, *D. antonietae* Tidon-Sklorz & Sene, 2001, *D. borborema* Vilela & Sene, 1977, *D. buzzatii* Patterson & Wheeler, 1942, *D. gouveai* Tidon-Sklorz & Sene, 2001, *D. koepferae* Fontdevila et al., 1988, *D. serido* Vilela & Sene, 1977, and *D. seriema* Tidon-Sklorz & Sene, 1995. These species inhabit open areas of sub-Amazonian arid lands and semidesert and desert regions of southern South America, where flies use necrotic cac-

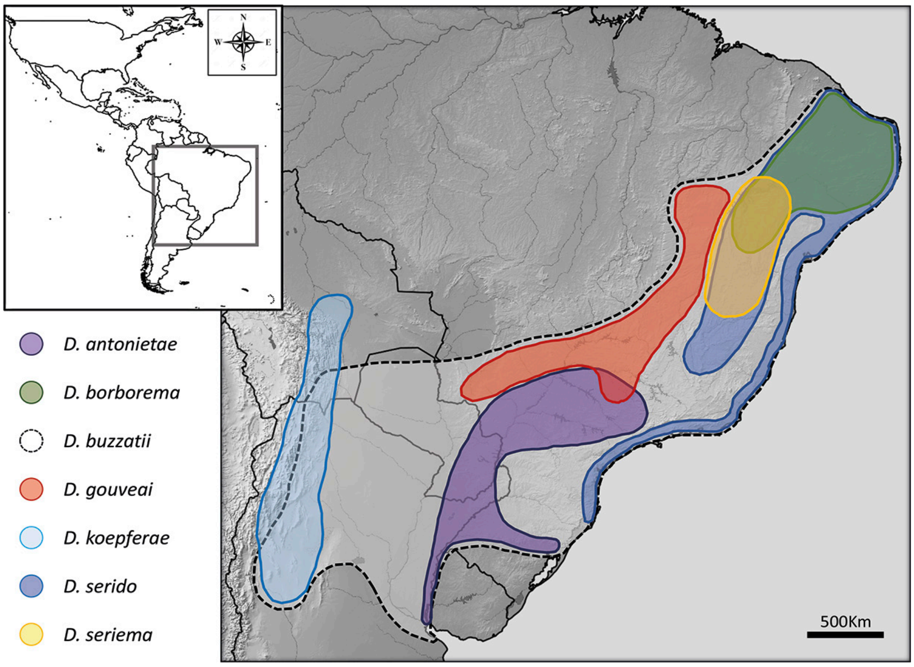


Fig. 1. Geographic distribution of the *buzzatii* cluster species. Sources: Cerda & Fontdevila (1998), Manfrin & Sene (2006), and Hurtado & Hasson (unpubl. field data).

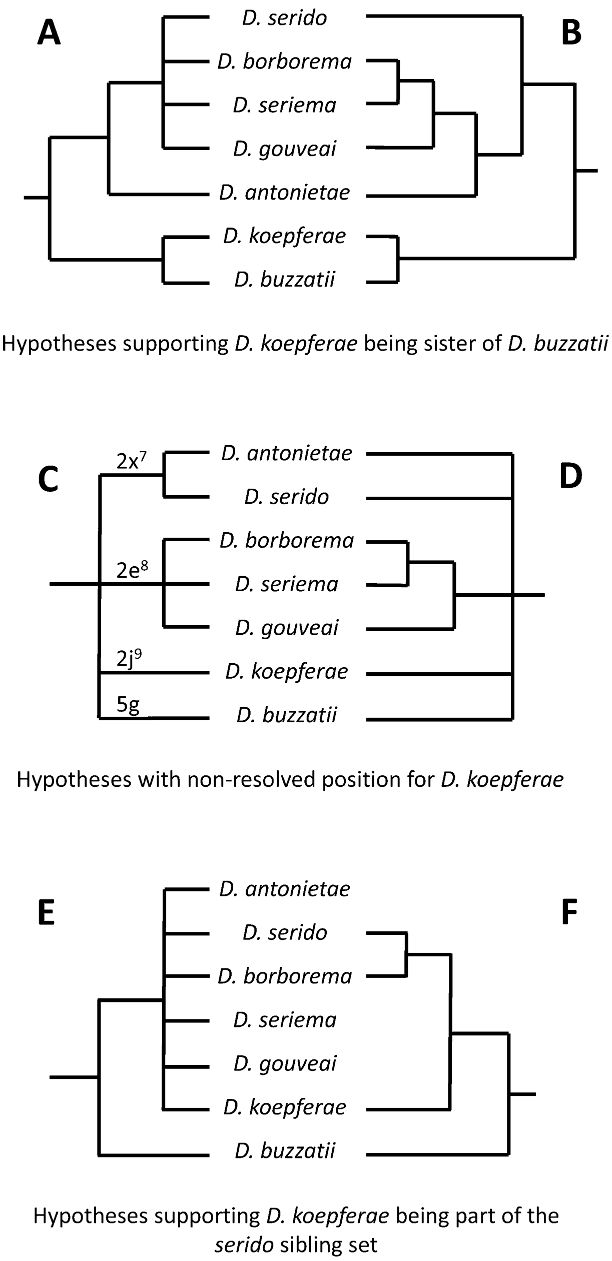
tus tissues as obligatory feeding and breeding resources (Sene et al. 1982; Pereira et al. 1983). Except for the semi-cosmopolitan *D. buzzatii*, which was introduced to Australia and Europe by human mediated dispersal of its main host cacti, prickly pears of the genus *Opuntia* (Barker et al. 1985; Fontdevila et al. 1989), the *buzzatii* cluster is endemic to South America (Fig. 1). Though *buzzatii* cluster species exhibit almost indistinct external morphology, differences in aedeagus (the external male intromittent organ) morphology and in banding patterns of salivary glands' polytene chromosomes allow precise species identification (reviewed in Manfrin & Sene 2006; Hasson et al. 2018).

The *buzzatii* species cluster has proven to be a valuable model system for evolutionary research (Manfrin & Sene 2006; Hasson et al. 2018). Studies in these species have offered unique insights into, for instance, the evolution of adaptations to arid conditions and host plant use (e.g. Hasson et al. 2009, 2018), chromosomal evolution (e.g. Wasserman 1992; Ruiz & Wasserman 1993), genome evolution (e.g. Guillén et al. 2015), fly–yeast–cactus coevolution (e.g. Sene et al. 1982), evolution of reproductive isolation (e.g. Machado et al. 2006), mechanisms of speciation (e.g. Ruiz et al. 2000), behavioral evolution (e.g. Iglesias & Hasson 2017), morphological evolution (e.g. Soto et al. 2007), sexual selection and sperm competition (e.g. Hurtado et al. 2013), and phylogeography (e.g. Franco et al. 2013). However, the significance of these findings is limited by the lack of a robust phylogenetic context.

Early phylogenetic appraisals in the *buzzatii* cluster based on the inspection of polytene chromosomes (Ruiz et al. 1982; Wasserman & Richardson 1987; Tosi & Sene 1989; Ruiz & Wasserman 1993; Ruiz et al. 2000) revealed four informative paracentric inversions – three in chromosome 2 and one in chromosome 5 – that support the monophyly of the *buzzatii* cluster within the *D. repleta* group. These inversions allow the identification of the main cytological lineages within the cluster: inversion *5g* fixed in the lineage leading to *D. buzzatii*, $2j^9$ in *D. koepferae*, $2x^7$ shared by *D. antonietae* and *D. serido*, and $2e^8$ shared by the triad *D. borborema*, *D. gouveai*, and *D. seriema* (Fig. 2C). The relationships within and between linages, however, cannot be completely defined on the basis of cytological characters because of the limited number of informative rearrangements within the cluster.

DNA sequence data is currently available for some species, but has yet to provide a robust phylogeny of the cluster. Seven nuclear and four mitochondrial genes have been sequenced so far in all or most of the seven *buzzatii* cluster species (Rodriguez-Trelles et al. 2000; Manfrin et al. 2001; Franco et al. 2010; Oliveira et al. 2012). However, many of the published phylogenetic trees based on these sequences are incongruent to each other or to the relationships inferred from cytological data, probably due to incomplete lineage sorting (ILS) and/or introgressive hybridization. Therefore, phylogenetic relationships in the *buzzatii* cluster remain unresolved since the initial studies by Fontdevila and coworkers (e.g. Rodriguez-Trelles et al. 2000) and Sene and coworkers (reviewed in Manfrin & Sene 2006).

Among unresolved taxonomic issues within the cluster, the phylogenetic position of *D. koepferae* is particularly controversial. On the one hand, *D. koepferae* appears to be more closely related to *D. buzzatii* according to general external morphology (Fontdevila et al. 1988), molecular evidence based on mtDNA (Manfrin et al. 2001), and chromosomal inversions and a X-linked marker considered together (Oliveira et al. 2011, 2013). On the other hand, it appears more closely related to the other five *buzzatii* cluster species as a member of



Hypotheses supporting *D. koepferae* being sister of *D. buzzatii*

Hypotheses with non-resolved position for *D. koepferae*

Hypotheses supporting *D. koepferae* being part of the *serido* sibling set

Fig. 2. Previous phylogenetic hypotheses for the *buzzatii* species cluster based on different datasets: **A**: Based on mitochondrial *COI* nucleotide sequences (Manfrin et al. 2001; Franco & Manfrin 2013). **B**: Based on fixed chromosomal inversions and X-linked *period* nucleotide sequences considered together (Oliveira et al. 2011, 2013). **C**: Based on fixed chromosomal inversions (Ruiz et al. 2000). **D**: Based on X-linked *period* nucleotide sequences (Franco et al. 2010). **E**: Based on *aedeagus* morphology (Tidon-Sklorz & Sene 2001), or on *pBuM* satellite DNA arrays (Kuhn & Sene 2005). **F**: Based on *Xdh* nucleotide sequences (Rodriguez-Trelles et al. 2000), or on the nucleotide sequences of 6 nuclear and 4 mitochondrial genes considered together (Oliveira et al. 2012).

the so-called *D. serido* sibling set on the basis of male genital morphology (Tidon-Sklorz & Sene 2001), the coding region of the *xanthine dehydrogenase* gene (*Xdh*) (Rodriguez-Trelles et al. 2000) and a few nuclear and mitochondrial markers considered together (Oliveira et al. 2012). In Fig. 2 we show which datasets considered so far support each hypothesis.

The aim of the present report is to contribute to the elucidation of the phylogenetic relationships within the *buzzatii* cluster using a large transcriptomic dataset. We generated RNA sequences that were used to assemble the adult male transcriptomes of *D. antonietae*, *D. borborema*, *D. buzzatii* and *D. koepferae*. We could not include the remaining three species of the cluster, which are endemic to Brazil, because of difficulties in obtaining collection and exportation permits from Brazilian authorities. From the assembled transcripts, we identified and selected fast-evolving protein-coding genes that were employed in phylogenetic searches to determine the relationships among members of the *buzzatii* cluster and to estimate divergence times.

2. Materials and methods

2.1. Fly strains

Single isofemale lines of *D. antonietae* (MG.2), *D. borborema* (BOR), *D. buzzatii* (M.11) and *D. koepferae* (7.1) were used throughout this study. MG.2 derives from a wild inseminated female collected in Isla Martín García (Buenos Aires Province, Argentina 34°10′42.12″S 58°15′23.15″W) in summer 2012. BOR was obtained from the *Drosophila* Species Stock Center (Stock Number: 15081 – 1281.01; University of California, San Diego, USA) and originally collected in Morro do Chapeu (Bahia State, Brazil 11°34′51.98″S 41°07′08.13″W). M.11 and 7.1 were stablished after eight generations of full-sib mating starting from the progeny of wild inseminated females collected in summer 2010 in Lavalle (Mendoza Province, Argentina 32°37′26.44″S 67°34′15.20″W) and Vipos (Tucumán Province, Argentina 26°29′34.19″S 65°19′34.01″W), respectively. Species were identified by means of the inspection of male genitalia and confirmed by examination of polytene chromosomes following Ruiz et al. (2000). All lines were maintained in the lab for > 60 generations on standard instant smashed potato medium under standard conditions (25 ± 1°C and a 12-h light : 12-h dark cycle) before RNA extraction.

2.2. Tissue dissection and RNA extraction

The data used herein was obtained with the aim of a comparative study on male reproductive genes among species. Thus, total RNA was extracted independently from reproductive accessory glands that were dissected from ~120 non-virgin males, and the bodies of ~20 males whose reproductive tracts were previously dissected. For each sample (combination of line and tissue type), collected tissues were pooled and stored in RNAlater at 8°C until RNA extraction. Total RNA was extracted using DirectZol RNA MiniPrep (Zymo) following manufacturers protocol. RNA preparations were quantified using

Qubit 2.0 (ThermoFisher Scientific), and RNA integrity was checked by agarose gel electrophoresis and further evaluated in a 2100 Bioanalyzer (Agilent Technologies). All flies were anesthetized using CO^2 before dissections. Dissections were performed using forceps under a 50 µL drop of cold RNAlater (Qiagen).

2.3. RNA library preparation and sequencing

RNA library preparation was performed following the Illumina TruSeq® RNA Sample Preparation Guide. We used Illumina paired-end (2 × 100 bp) libraries to generate the eight raw reads collections (one for each combination of line and tissue type). Library preparation and sequencing were performed at INDEAR (Argentina) for the accessory glands samples in a HiSeq 1500 platform and at MACROGEN (South Korea) for the samples of bodies devoid of reproductive tract in a HiSeq 2000 platform. Raw reads were quality-controlled using FastQC v. 0.11.5 (Andrews 2010) and in-house scripts (available upon request). The total number of reads/bases as well as some additional metrics for each library are summarized in the Electronic Supplement File 1 (Table S1). Reads were filtered and trimmed using Trimmomatic v. 0.36 (Bolger et al. 2014) applying the recommended parameters set by the transcriptome assembly tool Trinity v. 2.4.0 (Grabherr et al. 2011). Raw data are deposited at NCBI's Short Read Archive with accession no. PRJNA540063.

2.4. Transcriptome assembly

A whole transcriptome assembly comprising both kinds of samples (accessory glands and bodies without reproductive tract) was built for each species. Assembly was performed in a three-step process following the guidelines of the Trinity RNAseq assembler as well as PASA tool v. 2.1.0 (Haas et al. 2011; Rhind et al. 2011). First, *de novo* assembly was done using Trinity; second, reference-guided assembly was carried out using STAR aligner v. 2.5.2b (Dobin et al. 2013) to align the reads and Trinity for the assembly; and third, a hybrid approach for transcript reconstruction was carried out using reference-guided and *de novo* RNA-Seq assemblies to generate a comprehensive transcript database using PASA. We employed either the *D. buzzatii* st-1 genome (http://dbuz.uab.cat; Guillén et al. 2015) or the *D. koepferae* 7.1 genome (N. Moreyra et al. unpubl. data) as reference, choosing for each species the one against which more reads were successfully mapped (results not shown). Quality of transcriptome assemblies was assessed after each step by evaluating the read representation of the assemblies by aligning the raw data against them using bowtie2 aligner v. 2.3.0 (Langmead & Salzberg 2012), and by exploring completeness according to conserved ortholog content using the tool BUSCO (Simão et al. 2015) at two different lineage levels: Eukaryota and Diptera. Complete metrics can be found in the Electronic Supplement File 1 (Tables S2, S3).

2.5. Transcript selection and orthology criteria

Following the assembly of the transcriptomes, we applied a series of filters to select transcripts to be used in the phylogenetic analysis. First, we discarded sequences without a complete predicted protein. Predicted protein sequences were generated with the software TransDecoder v. 3.0.0 (Haas et al. 2013). Second, we identified orthologs present not only in the four *buzzatii* species, but also in the outgroup taxon *D. mojavensis* Patterson, 1940 (*mulleri* complex: *repleta* group). To this end, using pBLAST (Altschul et al. 1990), we searched for homologous *D. mojavensis* proteins in the dmoj_r1.04 genome (http://flybase.org/) using as query the translated transcripts of the four *buzzatii* species. Then we selected hits with an E-value < 1e-10 (considering the close relationships between species) that had a minimum coverage of 50% for each of the four queries. We then reversed the search using each selected protein of *D. mojavensis* as query to identify the best pBLAST hit among predicted proteins in each of the four *buzzatii* species. In this way, we established a first list of putative ortholog groups with sequenced homologs in all five species that we referred to as quintets. Third, to avoid the inclusion of paralogs or spurious alignments, we filtered the list including only those quintets for which each protein sequence, among all protein sequences of the same species, was the best pBLAST hit against each of the other four protein sequences of the quintet. Since the radiation of the *buzzatii* cluster is very recent (Oliveira et al. 2012, see below), we decided to remove highly conserved sequences that would provide little information in the phylogenetic analyses, enriching the list of quintets with fast-evolving sequences. Therefore, we filtered out from the second list those quintets in which one or more of the five sequences had significant similarity (E-value < 1e-10) to a *D. melanogaster* protein (flybase dmel_r6.14). After filtering, the final list included 922 quintets (i.e. ortholog groups) from which we recovered the parent transcript sequences for the phylogenetic searches.

2.6. Transcripts fine-tuning alignment

To construct a concatenated matrix with sequences from quintets, we performed several steps and filters. First, for each quintet, coding regions (CDSs) and available 3' and 5' untranslated regions (UTRs) were separated into different files and then aligned with the software MAFFT v. 7.299 (Katoh & Standley 2013) using the *globalpair* algorithm and 100 iterations. For each quintet, CDSs and available 3' and 5' untranslated regions (UTRs) were separated before alignment. We then filtered out the alignments with gaps in more than 30% of the sites (40% for 3' and 5'UTRs) and/or shorter than 300 bp (150 bp for 3' and 5'UTRs). Next, we filtered out alignments with outliers, i.e. sequences that were very divergent from the other sequences in the alignment and are thus likely not orthologous. To this end, we used the perl script EvalMSA (Chiner-Oms & González-Candelas 2014) that allows identification of outliers in multiple sequence alignments. In this way we ended up with 361 CDS alignments, 219 of 3'UTRs, and 181 of 5'UTRs. Finally, we proceeded to concatenate all alignments in an 812,976 bp long matrix. Transcript sequences of the selected genes are available in the Electronic Supplement File 2.

Additionally, we searched for nucleotide sequence data available in public databases for all species of the *buzzatii* cluster. Sequences of the X-linked *period* and the mitochondrial *cytochrome oxidase I* (*COI*) were available for all *buzzatii* cluster species in GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Sequences of the autosomal *alpha-esterase 5* (α*E5*) were available in GenBank for all species except for *D. antonietae*. Nevertheless, we recovered the α*E5* transcript sequence of the latter from the assembled transcriptome by means of BLAST searches. Sequences of the *cytochrome oxidase II* (*COII*) gene were available in GenBank for all *buzzatii* cluster species except for *D. antonietae* and *D. seriema*. For each species and gene locus we chose the longest available sequence. Sequences IDs or accession numbers are shown in the Electronic Supplement File 1 (Table S4). Next, we retrieved *D. mojavensis* sequences of the same four loci by means of BLAST searches in flybase and aligned the resulting dataset using the same procedure described above for the quintets. *COI* sequences were excluded since they did not pass our alignment filters. Then, in order to examine relationships among all seven species of the cluster, a new matrix was constructed including the transcriptomic dataset described above plus the aligned sequences of α*E5*, *period* and *COII*. We called this matrix the incomplete octet matrix, due to the lack of transcriptome data for *D. gouveai*, *D. serido*, and *D. seriema*.

2.7. Phylogenetic analyses

To apply probabilistic methods in the phylogenetic searches, we first determined the best partition scheme and substitution models for each partition using the software PartitionFinder (Lanfear et al. 2012). The searches were accomplished using a maximum likelihood (ML) approach with RAxML v. 8.2.4 (Stamatakis 2014) and a Bayesian inference (BI) approach with MrBayes v. 3.2.6 (Ronquist & Huelsenbeck 2003). The ML searches were done in 100 independent runs, using the rapid hill climb algorithm and the GTRCAT model with parameters estimated individually for each partition as determined by PartitionFinder. Support values for the nodes were obtained with 1000 bootstrap replicates and plotted onto the tree with the highest likelihood (see Results). In MrBayes searches, we unlinked both substitution models and parameter estimates for each partition, applying the models suggested by PartitionFinder. The search was done with 10 million generations, sampling every 5000. Convergence of chains was checked using Tracer (Rambaut et al. 2018), by making sure all parameters had effective sample sizes larger than 200. Trees were visualized with FigTree (Rambaut 2009).

It has been shown that the 'gene tree vs. species tree' phylogenetic approach is more reliable than the 'concatenated matrix' approach in groups with a significant amount of incomplete lineage sorting (ILS), because ILS may cause erroneous species tree results (Kubatko & Degnan 2007). We employed the gene tree vs. species tree method implemented in the program ASTRAL-III (Zhang et al. 2018), which seeks to find a species tree from gene trees, maximizing the number of shared quartet trees in the gene trees (Mirarab et al. 2014). ASTRAL-III takes as input ML trees obtained separately with each alignment. We obtained the gene trees with RAxML for the 761 quintet alignments (including CDS, 3'UTR and 5'UTR) using both the standard search algorithm (*-f d*) and the rapid bootstrap algorithm *(-f a -N 100)* to estimate clade support values. The coalescent-based species tree generated with ASTRAL-III has branch lengths in coalescent units and node support as local posterior probabilities that are computed based on the percentage of quartets in the gene trees that agree with that branch (Sayyari & Mirarab 2016).

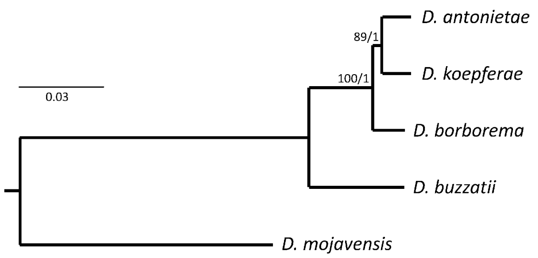
Besides ILS, interspecific hybridization may also cause differences in topology among gene trees. Thus, we employed the program PhyloNet version 3.6.8 (Than et al. 2008; Wen et al. 2018) that includes algorithms to evaluate the possibility that both ILS and hybridization were present in the evolution of the group. Again, the RAxML-generated gene trees were used as input. We employed both the parsimony (Minimizing Deep Coalescence = MDC) and the likelihood-based methods implemented in PhyloNet (*InferNetwork_MP* and *InferNework_ML* commands) to check whether a network would better describe the relationships within the *D. buzzatii* cluster (Yu et al. 2013). We also inferred the species tree using both statistical approaches and compared alternative relationships with AIC and the number of extra lineages (*CalGTProb* and *DeepCoalCount_Tree*). All analyses were repeated on gene trees with bootstrap support values taking into account only the nodes with bootstrap values larger than 50%. Given that UTR sequences come from the same genes as the CDSs, all analyses were repeated using only the 361 CDS alignments to avoid the use of non-independent loci.

Alternative topologies were also compared using the Shimodaira-Hasegawa test (Shimodaira & Hasegawa 1999) as implemented in RAxML and the approximately unbiased test (Shimodaira 2002) using CONSEL (Shimodaira & Hasegawa 2001).

2.8. Divergence time estimation

We estimated divergence times in the *buzzatii* cluster using transcriptomic data (quintets) and applying the molecular clock with the neutral mutation rate empirically estimated in *Drosophila melanogaster* (Keightley et al. 2009). Because the mutation rate is only applicable to

3



4

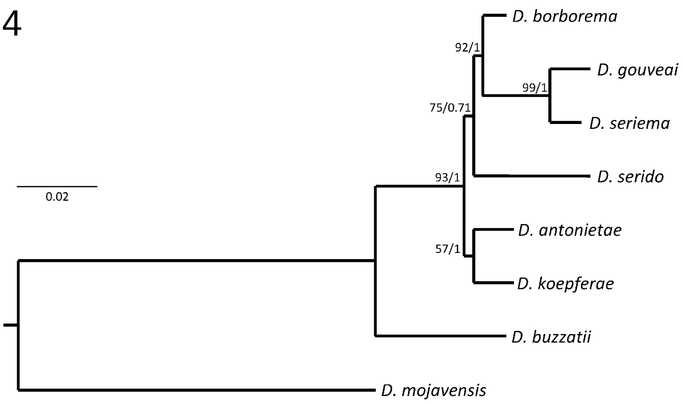


Fig. 3. Phylogenetic hypothesis for *D. antonietae*, *D. borborema*, *D. buzzatii* and *D. koepferae* based on transcriptomic data (concatenated quintet matrix). Bayesian posterior probabilities and ML bootstrap values are shown on nodes.

Fig. 4. Molecular phylogenetic hypothesis for the *buzzatii* species cluster based on the incomplete octet matrix. Bayesian posterior probabilities and ML bootstrap values are shown on nodes.

neutral sites, we selected only quintets with low codon bias, i.e. Codon Usage Bias index (CBU) < 0.375 (as in Obbard et al. 2012), for a total of 73 CDSs, and we kept only 4-fold degenerate third codon sites in the matrix. We then set a BEAST (Drummond et al. 2012) analysis applying a strict clock with a prior on the substitution rate of 3.46×10^{-9} (standard deviation 0.281), GTR + gamma substitution model, and a birth-death process with incomplete sampling as tree prior. The MCMC was run for five million generations with parameters logged every 2000 generations. Convergence of the chains was evaluated with Tracer (Rambaut et al. 2018), the first 20% trees were discarded as burn in, and the summarized, annotated trees were finally visualized with FigTree (Rambaut 2009).

3. Results

After filtering the assembled transcriptomes, a final dataset including 361 coding regions, 182 3'UTRs and 120 5'UTRs (for a combined matrix with more than 800,000 bp) was obtained to examine the phylogenetic relationships among the *buzzatii* species *D. antonietae*, *D. borborema*, *D. buzzatii* and *D. koepferae*. Both phylogenetic inference methods (ML and BI) based on our concatenated matrix recovered the same, well-supported relationships, allocating *D. koepferae* as phylogenetically closer to *D. antonietae* and *D. borborema* than to *D. buzzatii*. (Fig. 3). Though this is in fair agreement with phylogenetic relationships inferred on the basis of *aedeagus* morphology and most nuclear markers considered so far, in our tree *D. antonietae* is sister to *D. koepferae* (Dbuz,(Dbor,(Dant,Dkoe))) (Fig. 3), an arrangement that has never been reported (Fig. 2).

The gene tree vs. species tree analysis using ASTRAL-III resulted in a different topology: while *D. buzzatii*, in agreement with the topology obtained with the concatenated matrix and supported by 85% of the gene trees, was allocated as sister to all other species, *D. antonietae* and *D. borborema* appeared as sister species (Dbuz,(Dkoe,(Dant,Dbor))). The species tree analysis implemented in PhyloNet also favored Dbuz,(Dkoe,(Dant, Dbor)). However, relationships within the *antonietae-borborema-koepferae* clade were poorly resolved since the alternative arrangements were similarly supported (Table 1). Hybridization between *D. borborema* and *D. koepferae* was inferred with the MP criteria, but not with ML. According to the MDC method, this hybridization would reduce the number of extra lineages from the gene trees in more than 250. Similar outputs from ASTRAL-III or PhyloNet were obtained when we used bootstrapped trees and took into account only the clades with bootstrap support larger than 50 (results not shown).

In view of the different topologies obtained depending on the method, we tested whether the concatenated matrix statistically supported Dbor,(Dant,Dkoe) over the alternatives involving these three species. The AU test run on CONSEL found Dbor,(Dant,Dkoe) significantly better than both alternatives ($p < 0.006$ and $p < 0.001$), while the SH test run on RAxML found it to be better than Dbuz,(Dkoe,(Dbor,Dant)), but not better than Dbuz,(Dant,(Dkoe,Dbor)). Similar results were obtained in the CDS-only analysis (not shown).

Using the incomplete octet matrix, we also examined intra-cluster relationships among all seven *buzzatii* cluster species: the four for which we counted with assembled transcriptomes and the remaining three represented only by α*E5*, *period* and *COII* sequences. The recovered topology was the same with both ML and BI methods and concordant with the tree based on transcriptomic data alone (Fig. 4). Some nodes, however, were not very well supported. This is likely due to differences in the amount of sequence data across species and inconsistences across α*E5*, *period* and *COII*. Indeed, incongruences between these three loci were revealed by phylogenetic searches based on the sequences of each locus alone (Electronic Supplement File 3).

Table 1. Results of gene trees vs. species tree analysis (using PhyloNet) focusing on the three alternative topological arrangements involving *D. antonietae*, *D. borborema* and *D. koepferae*. First column shows the number of gene trees that agrees with each arrangement and the following columns show results of the PhyloNet analysis. — ***Superscripts***: ᵃ log likelihood of each topological arrangement; ᵇ Akaike information criteria of each arrangement; ᶜ Branch length of the sister clade; ᵈ Number of extra lineages obtained with each arrangement and the network as obtained under parsimony (MDC).

Arrangement	Number of supporting loci	lnL[a]	AIC[b]	Branch length[c]	Extra lineages[d]
Dbor,(Dant,Dkoe)	236 (31.4%)	-1387.38	2780.76	0.001	688
Dkoe,(Dant,Dbor)	265 (34.8%)	-1386.32	2778.64	0.034	651
Dant,(Dbor,Dkoe)	257 (33.8%)	-1387.15	2780.30	0.012	660
network	—	—	—	—	402

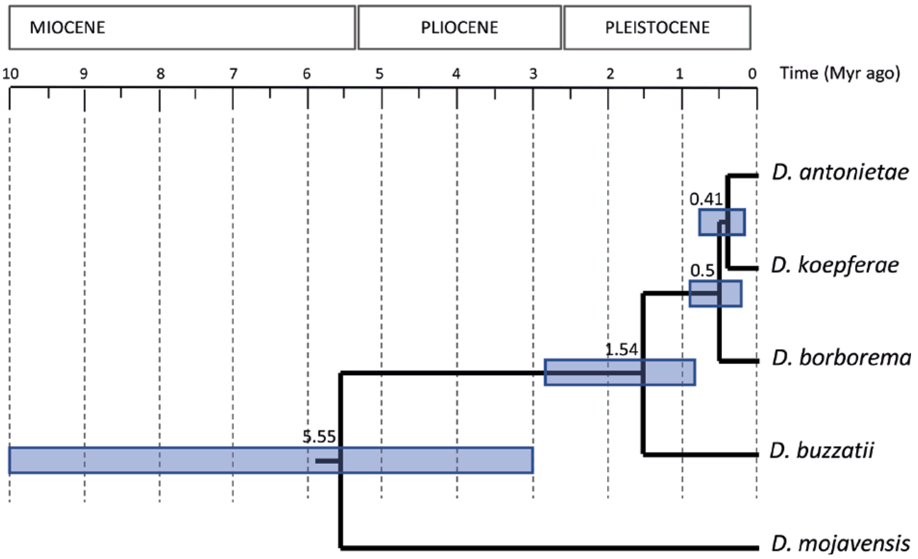


Fig. 5. Divergence times for four species of the *buzzatii* cluster plotted onto the ML/BI tree. Numbers at nodes are the time estimates, bars represent their 95% confidence intervals.

The divergence time estimates revealed that the *buzzatii* cluster is very young, with the first split that separates *D. buzzatii* occurring about 1.5 Myr ago and the remaining species diverging within the last ~ 0.5 Myr (Fig. 5). The dating analysis retrieved the same topology obtained in the ML analysis of the concatenated matrix.

4. Discussion

Past obstacles to a robust phylogeny of the *buzzatii* cluster are likely linked to the very recent radiation of the cluster (see below). At shallow divergence times, phylogenetic signal may be insufficient and phylogenetic challenges posed by incomplete lineage sorting (ILS) may prevail causing discordance across molecular markers. Additionally, as under laboratory conditions crosses between different *buzzatii* cluster species are known to produce partially fertile hybrid females (Fontdevila et al. 1988; Marin et al. 1993; Madi-Ravazzi et al. 1997; Machado et al. 2006), genetic introgression following interspecific gene flow may also account for the observed phylogenetic incongruences. Indeed, the analysis we performed with PhyloNet (MDC) suggests hybridization between *D. borborema* and *D. koepferae*. Also, previous studies explained the discordance between mitochondrial and nuclear markers as a consequence of unidirectional introgression in areas of sympatry between some species (Manfrin et al. 2001; Franco et al. 2010; Kokudai et al. 2011; but see Franco et al. 2015), and gene flow has been invoked to account for the presence of shared nucleotide polymorphisms in two nuclear gene loci in *D. buzzatii* and *D. koepferae* (Gómez & Hasson 2003; Piccinali et al. 2004). Phylogenomic analyses, based on a large number of rapidly-evolving loci, could potentially overcome these difficulties and help to untangle the phylogenetic relationships among these species. However, if the poor phylogenetic resolution reflects a hard polytomy resulting from an explosive radiation, even large genomic datasets will not fully resolve phylogenetic relationships within the cluster.

Our analyses based on the 813 kb concatenated matrix recovered a well-supported topology that fully resolve relationships among the four species for which we have transcriptomic data (Fig. 3), suggesting that speciation events splitting the lineages of at least these species were dichotomous. Moreover, despite the young age of the cluster and the possible occurrence of introgressive hybridization, our results confirm that *D. koepferae* is phylogenetically closer to *D. antonietae* and *D. borborema* than to *D. buzzatii*, solving the controversial position of *D. koepferae* within the cluster. Also, these results along with morphological and cytological evidence (Fig. 2C,E) give strong support to the monophyly of the so-called

D. serido sibling set, placing *D. buzzatii* alone as sister to the rest of the cluster. Our findings agree with previous reports based on *Xdh* sequences (Rodriguez-Trelles et al. 2000) and on the combined analysis of six nuclear and four mitochondrial loci (Oliveira et al. 2012). In contrast, the topologies recovered using *COI* sequences, placed *D. koepferae* as sister to *D. buzzatii* (Manfrin et al. 2001; Franco & Manfrin 2013).

However, the gene tree vs. species tree approach (implemented in ASTRAL-III and PhyloNet) resulted in a different topology for the *antonietae-borborema-koepferae* clade, with *D. antonietae* as sister to *D. borborema*, though the numbers of genes supporting the three alternative arrangements for this triad were very even (Table 1). This discordance across different gene trees may be due to ILS. In view of the inconsistence between methods, we cannot discard a hard polytomy for the *antonietae-borborema-koepferae* clade. A study involving a larger genomic dataset may help to conclusively resolve these relationships. Nevertheless, given that the arrangement recovered from the concatenated matrix was (according to CONSEL) statistically supported over the alternative ones, our dataset favors Dbuz,(Dbor,(Dant,Dkoe).

Although we have robustly assessed the relationships among the four species for which we have transcriptomic data, the relationships involving the other three species of the *buzzatii* cluster (*D. gouveai*, *D. serido*, and *D. seriema*) remain contentious. The topology we recovered from the incomplete octet matrix, including all *buzzatii* cluster species, supports the monophyly of a clade comprising *D. borborema*, *D. gouveai* and *D. seriema*, and resolves it, showing the latter two as sister species (Fig. 4). These species, however, are poorly represented in our matrix: *D. gouveai* is only represented by α*E5*, *period* and *COII* loci, while *D. seriema* by α*E5* and *period*. Moreover, the topologies obtained with α*E5*, *period* and *COII* are incongruent between each other and to our transcriptomic dataset (Electronic Supplement File 3). Oliveira et al. (2011), based on fixed chromosomal inversions and *period* sequences considered together, arrived at a different conclusion: though their analysis supported the monophyly of this triad, it places *D. seriema* as sister to *D. borborema* (Fig. 2D). Thus, a larger number of loci is required to draw precise conclusions regarding relationships within this triad of closely related species. The same is true in relation to *D. serido*, which was placed with poor statistical support as sister to the *borborema-gouveai-seriema* clade (Fig. 4).

Phylogenetic relationships among all *buzzatii* cluster species may also be inferred by jointly considering the cytological and molecular evidence. Polytene chromosome banding patterns are highly congruent with DNA sequence data (O'Grady et al. 2001) and have long been used to infer phylogenetic relationships among *Drosophila* species (e.g. Wasserman 1963, 1982). Furthermore, though relationships among the four *buzzatii* cytological lineages (that can be distinguished on the basis of fixed inversions) are unclear (Fig. 2C), all of them are represented in our quintet matrix (Fig. 3). Chromosome rearrangements do not favor any particular topology involving the *antonietae-borborema-koepferae* clade because these species have fixed different inversions. However, they provide strong evidence for the relationships of the three species for which transcriptomic data is not available: *D. gouveai* and *D. seriema* as part of a triad with *D. borborema*, based on the sharing of inversion *2e*^{8}, and *D. serido* as sister to *D. antonietae*, based on the sharing of inversion *2x*^{7}. Fig. 6 shows the tree that results from incorporating the cytological evidence to our concatenated quintet matrix-based inferences. The position of *D. serido* in this tree contradicts the incomplete octet matrix-based phylogeny shown in Fig. 4. Though more data are necessary to discern between these alternative hypotheses, we favor the arrangement suggested by the cytological data for *D. serido* since this species was represented only by three loci (α*E5*, *period* and *COII*) in our molecular matrix.

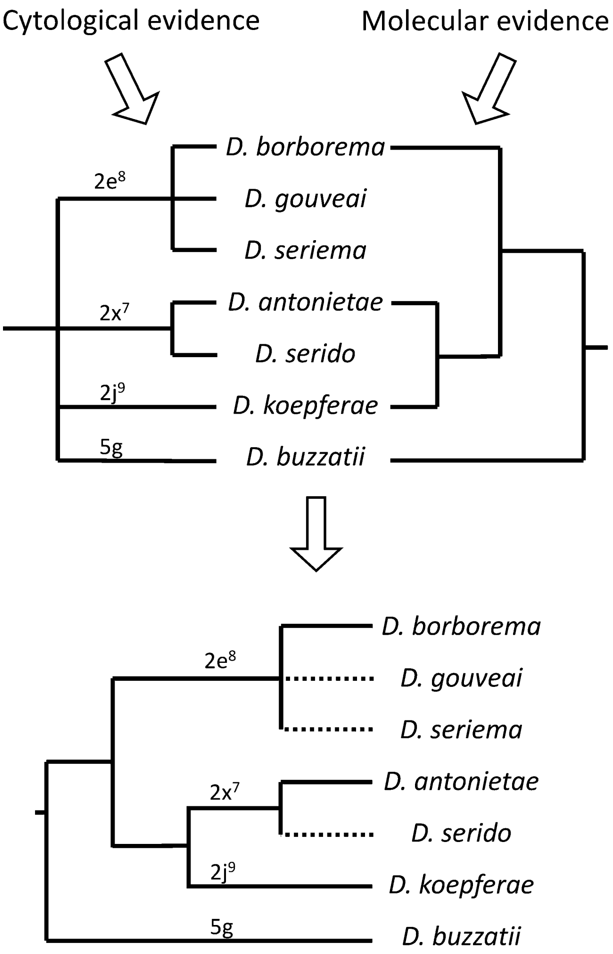


Fig. 6. Phylogenetic hypothesis for all seven *buzzatii* cluster species that results from combining the cytological evidence and our inferences based on the concatenated quintet matrix. Fixed inversions are shown on the branches.

Divergence times reported herein revealed that the *buzzatii* cluster radiation occurred entirely during the Pleistocene. According to our analyses, *D. buzzatii* split from the ancestor of the *D. serido* sibling set approximately 1.5 Myr ago. The radiation of the remaining six species seems to be extremely recent, < 0.5 – 1 Myr ago (Fig. 5). Previous estimations of divergence times among *buzzatii* species yielded larger figures. Gomez & Hasson (2003) and Oliveira et al. (2012) dated the split of *D. buzzatii* to ~4 or 4.6 Myr ago, respectively. Manfrin et al.'s (2001) estimates were even older, dating the radiation of the cluster (from the last to the first split) back to 3 – 12 Myr ago. Unlike the previous studies, which were based on one or a few loci and had to rely on the Hawaiian-calibrated substitution rate of the *alcohol dehydrogenase* gene (*Adh*) (Russo et al. 1995), we utilized a large number of loci, which enabled us to filter out genes with codon usage bias and select only four-fold degenerate sites while retaining enough information for divergence time estimates. This allowed us to use the empirical mutation rate, which is more precise than calibrations based on Hawaiian island ages. Obbard et al. (2012) reviewed the calibration strategies commonly used to date *Drosophila* species divergence, including the one employed in our analysis. They found that estimates based on the mutation rate most closely align with the fossil records and, in accordance with our results, are younger than most other estimates.

Considering the obligate ecological association between flies of the *buzzatii* cluster species and cacti, and the recent diversification revealed by our analysis, speciation events in the cluster may be explained by the Pleistocene refuge hypothesis. This hypothesis argues that Pleistocene glacial cycles successively generated isolated patches of similar habitats across which populations may have diverged into species (Haffer 1969; Endler 1982). The model originally refers to forest patches surrounded by dry habitats in dryer periods, but conversely, the expansion of humid habitats in wetter periods could result in fragmentation of the South American arid lands, contributing to the diversification of cacti-associated taxa. Different lines of evidence suggest that the distribution of arid vegetation (and associated fauna) in the Neotropics has been greatly affected during climatic cycles of the late Pleistocene (de Oliveira et al. 1999; Pennington et al. 2000; Costa 2003; Auler et al. 2004; Almeida et al. 2007; Pennington et al. 2009). Thus, speciation models built on expansion and retraction of vegetation during the Pleistocene climatic oscillations, like the Pleistocene refuge hypothesis or the similar canopy-density hypothesis (Cowling et al. 2001), offer a reliable scenario for the recent radiation of the *buzzatii* cluster.

In conclusion, we present a new revised phylogenetic hypothesis for the *buzzatii* cluster and show that the radiation of these closely related species may be more recent than previously thought. Although we resolved contentious issues concerning phylogenetic relationships within the cluster, others remain open due to lack of data, such as the position of *D. serido* and the internal arrangement of the *borborema-gouveai-seriema* clade. However, in view of the extremely shallow divergence times within the *D. serido* sibling set and the observed discordance across gene trees, we believe that a hard polytomy, product of an explosive radiation during the upper Pleistocene, is not unlikely for the *D. serido* sibling set. It also remains unclear the role and extent of interspecific hybridization in the evolution of the cluster. In this sense, the transcriptomic data analyzed herein, generated within the framework of ongoing genome projects involving *D. antonietae*, *D. borborema*, *D. buzzatii* and *D. koepferae*, provide the first step in the venture of unveiling phylogenetic relationships in the *buzzatii* cluster using massive genomic data. Nevertheless, it is necessary to generate similar datasets for the Brazilian *D. gouveai*, *D. serido* and *D. seriema* to achieve a deeper understanding of the evolutionary history of this species cluster that diversified in Southern South America.

5. Acknowledgments

The authors wish to thank Nicolás Moreyra for providing access to the *D. koepferae* 7.1 genome and to Eduardo Soto and Julián Mensch for comments that helped in different stages of the present study. We also would like to express our appreciation to Kyle Benowitz for the critical reading of previous versions of this paper. This work was supported with University of Buenos Aires, CONICET and Agencia Nacional de Promoción Científica y Técnica grants awarded to EH.

6. References

Almeida F.C., Bonvicino C.R., Cordeiro-Estrela P. 2007. Phylogeny and temporal diversification of *Calomys* (Rodentia, Sigmodontinae): Implications for the biogeography of an endemic genus of the open/dry biomes of South America. – Molecular Phylogenetics and Evolution **42**: 449 – 466.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. – Journal of Molecular Biology **215**: 403 – 410.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. – Available online from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Auler A.S., Wang X., Edwards R.L., Chencg H., Cristalli P.S., Smart P.L., Richards D.A. 2004. Quaternary ecological and geomorphic changes associated with rainfall events in presently semi-arid north- eastern, Brazil. – Journal of Quaternary Science **19**: 693 – 701.

Barker J.S.F., Sene F.M., East P.D., Pereira M.A.Q.R. 1985. Allozyme and chromosomal polymorphism of *Drosophila buzzatii* in Brazil and Argentina. – Genetica **67**: 161 – 170.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. – Bioinformatics btu170.

Cerda H., Fontdevilla A. 1998. Evolutionary divergence of *Drosophila venezolana* (*martensis* Cluster, *buzzatii* Complex) on Gran Roque Island, Venezuela. – Drosophila Information Service **81**: 144 – 147.

Chiner-Oms A., González-Candelas F. 2014. EvalMSA: A program to evaluate multiple sequence alignments and detect outliers. – Evolutionary Bioinformatics Online **12**: 277 – 284.

Costa L.P. 2003. The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals. – Journal of Biogeography **30**: 71 – 86.

Cowling S.A., Maslin M.A., Sykes M.T. 2001. Paleovegetation simulations of lowland Amazonia and implications for neotropical allopatry and speciation. – Quaternary Research **55**: 140 – 149.

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. 2013. STAR: ultrafast universal RNA-seq aligner. – Bioinformatics **29**: 15–21.

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

Endler J.A. 1982. Problems in distinguishing historical from ecological factors in biogeography. – American Zoologist **22**: 441–452.

Fontdevila A. 1989. Founder effects in colonizing populations: the case of *Drosophila buzzatii*. Pp. 74–95 in: Fontdevila A. (ed.), Evolutionary Biology of Transient Unstable Populations. – Springer Verlag, Berlin-Heidelberg.

Fontdevila A., Pla C., Hasson E., Wasserman M., Sanchez A., Naveira H., Ruiz A. 1988. *Drosophila koepferae*: a new member of the *Drosophila serido* (Diptera, Drosophilidae) superspecies taxon. – Annals of the Entomological Society of America **81**(3): 380–385.

Franco F.F., Lavagnini T.C., Sene F.M., Manfrin M.H. 2015. Mitonuclear discordance with evidence of shared ancestral polymorphism and selection in cactophilic species of *Drosophila*. – Biological Journal of the Linnean Society **116**: 197–210.

Franco F.F., Manfrin M.H. 2013. Recent demographic history of cactophilic *Drosophila* species can be related to Quaternary palaeoclimatic changes in South America. – Journal of Biogeography **40**: 142–154.

Franco F.F., Silva-Bernardi E.C.C., Sene F.M., Hasson E.R., Manfrin M.H. 2010. Intra- and interspecific divergence in the nuclear sequences of the clock gene period in species of the *Drosophila buzzatii* cluster. – Journal of Zoological Systematics and Evolutionary Research **48**: 322–331.

Gómez G.A., Hasson E. 2003. Transpecific polymorphisms in an inversion linked esterase locus in *Drosophila buzzatii*. – Molecular Biology and Evolution **20**: 410–423.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. – Nature Biotechnology **29**(7): 644–652.

Guillén Y., Rius N., Delprat A., Williford A., Muyas F., Puig M., Casillas S., Ramia M., Egea R., Negre B., Mir G., Camps J., Moncunill V., Ruiz-Ruano F.J., Cabrero J., de Lima L., Dias G.B., Ruiz J.C., Kapusta A., Garcia-Mas J., Gut M., Gut I.G., Torrents D., Camacho J.P., Kuhn G.C., Feschotte C., Clark A.G., Betrán E., Barbadilla A., Ruiz A. 2015. Genomics of ecological adaptation in cactophilic *Drosophila*. – Genome Biology Evolution **7**(1): 349–366.

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., MacManes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., LeDuc R.D., Friedman N., Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. – Nature Protocols **8**(8): 1494.

Haas B.J., Zeng Q., Pearson M.D., Cuomo C.A., Wortman J.R. 2011. Approaches to fungal genome annotation. – Mycology **2**(3): 118–141.

Haffer J. 1969. Speciation in Amazonian forest birds. – Science **165**: 131–137.

Hasson E., Soto I.M., Carreira V.P., Corio C., Soto E.M., Betti M.I.L. 2009. Host plants, fitness and developmental instability in a guild of cactophilic species of the genus *Drosophila*. Pp. 89–109 in: Santos E.B. (ed.), Ecotoxicology Research Developments. – Nova Science Publishers, Inc., New York.

Hasson E., De Panis D., Hurtado J., Mensch J. 2018. Host plants adaptation in cactophilic species of the *Drosophila buzzatii* cluster: fitness and transcriptomics. – Journal of Heredity esy043.

Hurtado J., Iglesias P.P., Lipko P., Hasson E. 2013. Multiple paternity and sperm competition in the sibling species *Drosophila buzzatii* and *Drosophila koepferae*. – Molecular Ecology **22**: 5016–5026.

Iglesias P.P., Hasson E. 2017. The role of courtship song in female mate choice in South American cactophilic *Drosophila*. – PLoS ONE **12**(5): e0176119.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. – Molecular Biology and Evolution **30**: 772–780.

Keightley P.D., Trivedi U., Thomson M., Oliver F., Kumar S., Blaxter M.L. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. – Genome Research **19**: 1195–1201.

Kokudai C.B.S., Sene F.M., Manfrin M.H. 2011. Sympatry and asymmetric introgression between the cactophilic species *Drosophila serido* and *Drosophila antonietae* (Diptera: Drosophilidae). – Annals of the Entomological Society of America **104**: 434–442.

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. – Systematic Biology **56**(1): 17–24.

Kuhn G.C.S., Sene F.M. 2005. Evolutionary turnover of two *pBuM* satellite DNA subfamilies in the *Drosophila buzzatii* species cluster (*repleta* group): from alpha to alpha/beta arrays. – Gene **349**: 77–85.

Lanfear R., Calcott B., Ho S.Y., Guindon S. 2012. Partition Finder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. – Molecular Biology and Evolution **29**: 1695–1701.

Langmead B., Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. – Nature Methods **9**: 357–359.

Machado C.A., Hey J. 2003. The causes of phylogenetic conflict in a classic *Drosophila* species group. – Proceedings of the Royal Society of London **270**(1520): 1193–1202.

Machado L.P., Madi-Ravazzi L., Tadei W.J. 2006. Reproductive relationships and degree of synapsis in the politene chromosomes of the *Drosophila buzzatii* species cluster. – Brazilian Journal of Biology **66**: 279–293.

Madi-Ravazzi L., Bicudo H.E., Manzato J.A. 1997. Reproductive compatibility and chromosome pairing in the *Drosophila buzzatii* complex. – Cytobios **89**: 21–30.

Manfrin M.H., de Brito R.O.A., Sene F.M. 2001. Systematics and evolution of the *Drosophila buzzatii* (Diptera: Drosophilidae) cluster using mtDNA. – Annals of the Entomological Society of America **94**: 333–346.

Manfrin M.H., Sene F.M. 2006. Cactophilic *Drosophila* in South America: A model for evolutionary studies. – Genetica **126**: 57–75.

Marin I., Ruiz A., Pla C., Fontdevila A. 1993. Reproductive relationships among ten species of the *Drosophila repleta* group from South America and the West Indies. – Evolution **47**: 1616–1624.

Markow T.A., O'Grady P.M. 2008. Reproductive ecology of *Drosophila*. – Functional Ecology **22**: 747–759.

Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree. – Bioinformatics (ECCB special issue) **30**(17): 541–548.

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. – Molecular Biology and Evolution **29**: 3459–3473.

O'Grady P.M., Baker R.H., Durando C.M., Etges W.J., DeSalle R. 2001. Polytene chromosomes as indicators of phylogeny in several species groups of *Drosophila*. – BMC Evolutionary Biology **1**: 6.

Oliveira C.C., Manfrin M.H., Sene F.M., Etges W.J. 2013. Evolution of male courtship songs in the *Drosophila buzzatii* species cluster. Pp. 137–164 in: Michalak P. (ed.), Speciation: Natural

Processes, Genetics and Biodiversity. – Nova Science Publishers, Inc., New York.

Oliveira C.C., Manfrin M.H., Sene F.M., Jackson L.L., Etges W.J. 2011. Variations on a theme: diversification of cuticular hydrocarbons in a clade of cactophilic *Drosophila*. – BMC Evolutionary Biology **11**: 179.

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. – Molecular Phylogenetics and Evolution **64**: 533–544.

de Oliveira P.E., Barreto A.M.F., Suguio K. 1999. Late Pleistocene/Holocene climatic and vegetational history of the Brazilian caatinga: the fossil dunes of the middle São Francisco River. – Palaeogeography, Palaeoclimatology, Palaeoecology **152**: 319–337.

Patterson J.T., Wheeler M.R. 1942. Description of new species of the subgenera *Hirtodrosophila* and *Drosophila*. – Austin: University of Texas Publication **4213**: 67–109.

Pennington R.T., Lavin M., Oliveira-Filho A. 2009. Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. – Annual Review of Ecology, Evolution, and Systematics **40**: 437–457.

Pennington R.T., Prado D.A., Pendry C. 2000. Neotropical seasonally dry forests and Pleistocene vegetation changes. – Journal of Biogeography **27**: 261–273.

Pereira M.A.Q.R., Vilela C.R., Sene F.M. 1983. Notes on breeding and feeding sites of some species of the *repleta* group of the genus *Drosophila* (Diptera, Drosophilidae). – Ciência e Cultura **35**(9): 1313–1319.

Piccinali R., Aguadé M., Hasson E. 2004. Comparative molecular population genetics of the *Xdh* locus in the cactophilic sibling species *Drosophila buzzatii* and *D. koepferae*. – Molecular Biology and Evolution **21**(1): 141–152.

Rambaut A. 2009. FigTree: Tree Figure Drawing Tool. – Available online from: http://tree.bio.ed.ac.uk/software/figtree.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Tracer: MCMC Trace Analysis Package. – Available online from: http://beast.community/tracer.

Rhind N., Chen Z., Yassour M., Thompson D.A., Haas B.J., Habib N., Wapinski I., Roy S., Lin M.F., Heiman D.I., Young S.K., Furuya K., Guo Y., Pidoux A., Chen H.M., Robbertse B., Goldberg J.M., Aoki K., Bayne E.H., Berlin A.M., Desjardins C.A., Dobbs E., Dukaj L., Fan L., FitzGerald M.G., French C., Gujja S., Hansen K., Keifenheim D., Levin J.Z., Mosher R.A., Müller C.A., Pfiffner J., Priest M., Russ C., Smialowska A., Swoboda P., Sykes S.M., Vaughn M., Vengrova S., Yoder R., Zeng Q., Allshire R., Baulcombe D., Birren B.W., Brown W., Ekwall K., Kellis M., Leatherwood J., Levin H., Margalit H., Martienssen R., Nieduszynski C.A., Spatafora J.W., Friedman N., Dalgaard J.Z., Baumann P., Niki H., Regev A., Nusbaum C. 2011. Comparative functional genomics of the fission yeasts. – Science **332**(6032): 930–936.

Rodriguez-Trelles F., Alarcón L., Fontdevila A. 2000. Molecular evolution and phylogeny of the *buzzatii* complex (*Drosophila repleta* group): A maximum-likelihood approach. – Molecular Biology and Evolution **17**: 1112–1122.

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. – Bioinformatics **19**: 1572–1574.

Ruiz A., Cansian A.M., Kuhn G.C.S., Alves M.A.R., Sene F.M. 2000. The *Drosophila serido* speciation puzzle: putting new pieces together. – Genetica **108**: 217–227.

Ruiz A., Fontdevila A., Wasserman M. 1982. The evolutionary history of *Drosophila buzzatii*. III: Cytogenetic relationship between two sibling species of the *buzzatii* cluster. – Genetics **101**: 503–518.

Ruiz A., Heed W.B. 1988. Host-plant specificity in the cactophilic *Drosophila mulleri* species complex. – The Journal of Animal Ecology **57**: 237–249.

Ruiz A., Wasserman M. 1993. Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. – Heredity **70**: 582–596.

Russo C.A., Takezaki N., Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. – Molecular Biology and Evolution **12**: 391–404.

Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. – Molecular Biology and Evolution **33**(7): 1654–1668.

Sene F.M., Pereira M.A.Q.R., Vilela C.R. 1982. Evolutionary aspects of cactus breeding *Drosophila* in South America. Pp. 97–106 in: Barker J.S.F., Starmer W.T. (eds), Ecological Genetics and Evolution. The Cactus-Yeast-*Drosophila* Model System. – Sydney: Academic Press.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. – Systematic Biology **51**: 492–508.

Shimodaira H., Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. – Molecular Biology and Evolution **16**(8): 1114–1116.

Shimodaira H., Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. – Bioinformatics **17**: 1246–1247.

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. – Bioinformatics **31**(19): 3210–3212.

Soto I.M., Carreira V.P., Fanara J.J., Hasson E. 2007. Evolution of male genitalia: environmental and genetic factors affect genital morphology in two *Drosophila* sibling species and their hybrids. – BMC Evolutionary Biology **7**: 77.

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

Than C., Ruths D., Nakhleh L. 2008. PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary histories. – BMC Bioinformatics **9**: 322.

Tidon-Sklorz R., Sene F.M. 1995. *Drosophila seriema*: a new member of the *Drosophila serido* (Diptera, Drosophilidae) superspecies taxon. – Annals of the Entomological Society of America **88**: 139–142.

Tidon-Sklorz R., Sene F.M. 2001. Two new species of the *Drosophila serido* sibling set (Diptera, Drosophilidae). – Iheringia Série Zoologia **90**: 141–146.

Tosi D., Sene F.M. 1989. Further studies on chromosomal variability in *Drosophila serido* (Diptera, Drosophilidae). – Revista Brasileira de Genética **1**(4): 729–746.

Vilela C.R., Sene F.M. 1977. Two new neotropical species of the repleta group of the genus *Drosophila* (Diptera, Drosophilidae). – Papéis Avulsos de Zoologia **30**(20): 295–299.

Wasserman M. 1963. Cytology and phylogeny of *Drosophila*. – American Naturalist **97**: 333–352.

Wasserman M. 1982. Evolution of the *repleta* group. Pp. 61–139 in: Ashburner M., Carson H.L., Thompson J.N. (eds), The Genetics and Biology of *Drosophila*, Vol. 3b. – Academic Press, London.

Wasserman M. 1992. Cytological evolution of the *Drosophila repleta* species group. Pp. 455–552 in: Krimbas C.B., Powell J.R. (eds), *Drosophila* Inversion Polymorphism. – CRC Press, Boca Raton, Florida.

Wasserman M., Richardson R.H. 1987. Evolution of Brazilian *Drosophila mulleri* complex species. – Journal of Heredity **78**: 282–286.

Wen D., Yu Y., Zhu J., Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. – Systematic Biology **67**(4): 735–740.

Yu Y., Barnett R.M., Nakhleh L. 2013. Parsimonious inference of hybridization in the presence of incomplete lineage sorting. – Systematic Biology **62**(5): 738–751.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. – BMC Bioinformatics **19**(6): 153.

Electronic Supplement Files

at http://www.senckenberg.de/arthropod-systematics

File 1: hurtado&al-drosophilabuzzatii-asp2019-electronicsupplement-1.xlsx — **Table S1.** Reads length, total number of reads / bases, quality score and %GC for each RNA-seq library. — **Table S2.** Metrics for transcript contigs length for each transcriptome assembly. — **Table S3.** Metrics for read representation and completeness for each transcriptome assembly. — **Table S4.** GenBank accession numbers / Flybase IDs for the additional sequences of the incomplete octet matrix. — DOI: 10.26049/ASP77-2-2019-03/1

File 2: hurtado&al-drosophilabuzzatii-asp2019-electronicsupplement-2.fasta — Transcript sequences of the 361 selected genes that were employed in the phylogenetic searches. The first four alphabetic characters of each name stand for the species while the last numeric characters stand for the ortholog group. — DOI: 10.26049/ASP77-2-2019-03/2

File 3: hurtado&al-drosophilabuzzatii-asp2019-electronicsupplement-3.pdf — **Fig. S1.** Molecular-based phylogenetic hypothesis for the *buzzatii* species cluster: **A**: Based on the autosomal *E5* sequences. **B**: Based on mitochondrial *COII* sequences. **C**: Based on the X-linked *period* sequences. **D**: Based on *E5*, *period* and *COII* sequences considered together. Red branches stand for nodes incongruent to the phylogenetic hypothesis based on transcriptomic data. — DOI: 10.26049/ASP77-2-2019-03/3