

Radiation of the *Drosophila nanoptera* species group in Mexico

M. LANG*, M. POLIHONAKIS RICHMOND†, A. E. ACURIO‡, T. A. MARKOW†§
& V. ORGOGOZO*

*CNRS UMR7592, Institut Jacques Monod, Université Paris Diderot, Paris, France

†Section of Cell and Developmental Biology, Division of Biological Sciences, University of California, La Jolla, CA, USA

‡Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Spain

§Laboratorio Nacional de Genómica de la Biodiversidad, CINVESTAV, Irapuato, Mexico

Keywords:

asymmetric male genitalia;
molecular phylogeny;
multilocus analysis;
nanoptera species group;
reproductive characters;
species divergence estimates.

Abstract

The *Drosophila nanoptera* species group, a taxon of Mexican cactophilic flies, is an excellent model system to study the influence of abiotic and biotic factors on speciation, the genetic causes of ecological specialization and the evolution of unusual reproductive characters. However, the phylogenetic relationships in the nanoptera species group and its position within the virilis-repleta phylogeny have not been thoroughly investigated. Using a multilocus data set of gene coding regions of eight nuclear and three mitochondrial genes, we found that the four described nanoptera group species diverged rapidly, with very short internodes between divergence events. Phylogenetic analysis of repleta group lineages revealed that *D. inca* and *D. canalinea* are sister to all other repleta group species, whereas the annulimana species *D. aracataca* and *D. pseudotalamancana* are sister to the nanoptera and bromeliae species groups. Our divergence time estimates suggest that the nanoptera species group radiated following important geological events in Central America. Our results indicate that a single evolutionary transition to asymmetric genitalia and to unusual sperm storage may have occurred during evolution of the nanoptera group.

Introduction

Species of the genus *Drosophila*, because of their well-defined phylogenetic relationships and diverse ecologies and life histories, provide an attractive group of model organisms for the study of evolution (Markow & O'Grady, 2007). A few taxa in Drosophilidae have evolved the ability to feed and breed in necrotic cactus, predominantly in the repleta and nanoptera species groups (Markow & O'Grady, 2007). The nanoptera species group consists of only four described species: *Drosophila nanoptera* (Wheeler, 1949), *D. acanthoptera* (Wheeler, 1949), *D. wassermani* (Pitnick & Heed, 1994) and *D. pachea* (Patterson & Wheeler, 1942). Even though all species of the nanoptera group live on columnar cacti, they exhibit diverse degrees of ecological specialization. Whereas *D. nanoptera* can live on a variety of host plants of the

genera *Stenocereus*, *Pachycereus*, *Escontria* and *Myrtillocactus* (Heed, 1982), *D. acanthoptera* and *D. wassermani* are restricted to species in the genus *Stenocereus* (Heed, 1982). An even more tight ecological specialization links *D. pachea* to a single host plant, the senita cactus *Lophocereus schottii* (Engelmann, 1852), which is toxic to the other three species of the nanoptera group (Heed & Kircher, 1965; Etges *et al.*, 1999) and which provides a particular sterol (lathosterol) absolutely required for *D. pachea* survival (Heed & Kircher, 1965; Lang *et al.*, 2012).

The geographic distribution of *D. pachea* coincides with the distribution of senita cactus, which is restricted to the Sonoran desert in north-west mainland Mexico and to the Baja California peninsula (Fig. 1; Lindsay, 1963; Hastings *et al.*, 1972). The Gulf of California and the Sierra Madre Occidental mountain range on the mainland separate the distribution of *D. pachea* from *D. wassermani* (Fig. 1b). The other nanoptera species are found in an overlapping region in southern Mexico (Heed, 1982; Markow & O'Grady, 2005). *Drosophila nanoptera* generally localizes in highlands, whereas *D. wassermani* is primarily found in lowlands (Heed,

Correspondence: Michael Lang, CNRS UMR7592, Institut Jacques Monod, Université Paris Diderot, Bâtiment Buffon 416B, 15 rue Hélène Brion, 75205 Paris Cedex 13, France. Tel.: +33 157278099; fax: +33 157278087; e-mail: lang@ijm.univ-paris-diderot.fr

1982). Specimens of *D. acanthoptera* were also reported from Venezuela (Hunter, 1970), but this single sampling record, based on morphological characterization, remains dubious.

The nannoptera species group inhabits a zone of important geological history. About 15 million years ago (Ma), seismic activity along a volcanic arc formed the Isthmus of Panama that connected Central and South America (Montes *et al.*, 2012a,b). The formation of the isthmus had a huge biological and climatic impact and provided a means for terrestrial fauna to move between the two continents (Webb, 1976; Leigh *et al.*, 2014). More recently, about 6–3 Ma, the Baja California peninsula formed as a result of a series of complex geological events that caused the successive separation of landmasses from mainland Mexico (Lizarralde *et al.*, 2007; Umhoefer, 2011). Increased desertification of North and South America in the past 10 Ma due to a global climate change and an uplift period of the Andes during the late Miocene-Pliocene (Gregory-Wodzicki, 2000; Capitanio *et al.*, 2011) was accompanied by radiations of major suc-

culent plant lineages (Arakaki *et al.*, 2011). Whether these geological events might have influenced speciation within the nannoptera species group is unknown.

In addition to the ability to utilize cactus tissue as a resource, some unusual reproductive characters have evolved in the nannoptera species group. For instance, sperm gigantism was observed in *D. pachea* and *D. nannoptera*, but not in the other two members of the group, *D. acanthoptera* and *D. wassermani* (Pitnick *et al.*, 1995). Two additional very curious reproductive characters, genital asymmetry (Vilela & Baechli, 1990; Pitnick & Heed, 1994; Lang & Orgogozo, 2012) and site of sperm storage in females (Pitnick *et al.*, 1999), have also been reported in a few species of the nannoptera group. Whereas *D. nannoptera*, like most other Drosophilidae, has fully symmetric genitalia (Vilela & Baechli, 1990; Huber *et al.*, 2007), the other three species of the nannoptera group possess diverse genital organs with conspicuous left-right asymmetric morphologies. *Drosophila pachea* displays an epandrial lobe size asymmetry (Pitnick & Heed, 1994; Lang & Orgogozo, 2012),

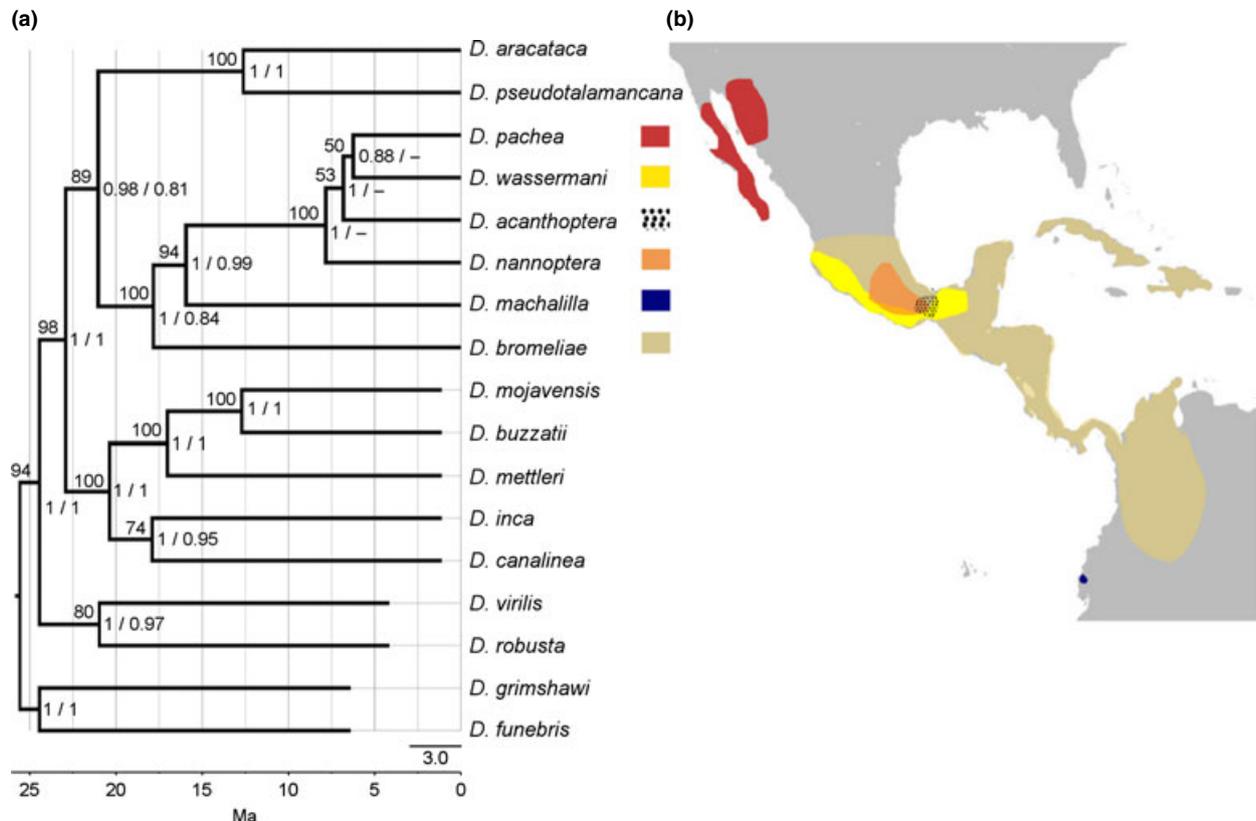


Fig. 1 Phylogenetic relationships of the nannoptera group and related species. (a) Phylogenetic tree generated in BEAST based on the concatenated data set with nine partitions. Bootstrap support from maximum likelihood (PhyML) analysis is presented on the left side of each node. Bayesian posterior probabilities are presented on the right side of each node for the BEAST analysis/and *BEAST analysis, respectively. The time scale was calculated according to estimates B in Table 1. (b) Distributions of the species of the nannoptera group and of *Drosophila machalilla* and *D. bromeliae*, reproduced from Heed (1982) and Markow & O'Grady (2005).

D. acanthoptera possesses an asymmetric phallus (Vilela & Baechli, 1990), and *D. wassermani* has a left–right concave–convex-shaped cercus (Pitnick & Heed, 1994; Fig. 2). Furthermore, *D. nanoptera*, like most other *Drosophila* species, uses two types of organs for post-copulatory storage of sperm in females: the paired spermathecae and the single seminal receptacle. However, *D. acanthoptera*, *D. wassermani* and *D. pachea* use only the spermathecae (Pitnick *et al.*, 1999). Therefore, with regard to the evolution of asymmetric genitalia and unusual sperm storage, the most parsimonious scenario would be that *D. nanoptera* is an out group relative to the other three species of the nanoptera group.

For the reasons mentioned previously, the nanoptera group thus represents an interesting model system to tackle a variety of important questions in evolutionary biology, such as the influence of abiotic and biotic factors on speciation, the genetic causes of ecological specialization and the evolution of reproductive characters. To address these questions and to trace back the

evolution of different characters across the nanoptera species group, a reliable phylogeny of the four species and related taxa is required. Whereas relationships within the repleta group have been characterized to a great extent (Van der Linde *et al.*, 2010; Oliveira *et al.*, 2012), previous phylogenetic studies of the nanoptera group have led to equivocal and conflicting results (Pitnick *et al.*, 1995, 1997, 1999; Van der Linde *et al.*, 2010; Oliveira *et al.*, 2012; Yassin, 2013). Some of these analyses were based on relatively few genetic loci (Pitnick *et al.*, 1995, 1997, 1999) and others that included more loci either lacked appropriate out groups (Oliveira *et al.*, 2012) or did not examine all members of the nanoptera species group (Van der Linde *et al.*, 2010; Yassin, 2013). In a morphological analysis based on internal reproductive organ morphology, Heed (1982) proposed a phylogeny of the nanoptera group, with *D. wassermani* and *D. pachea* forming two sister species, which are in turn sister to *D. acanthoptera*, and with *D. nanoptera* being out group relative to the other

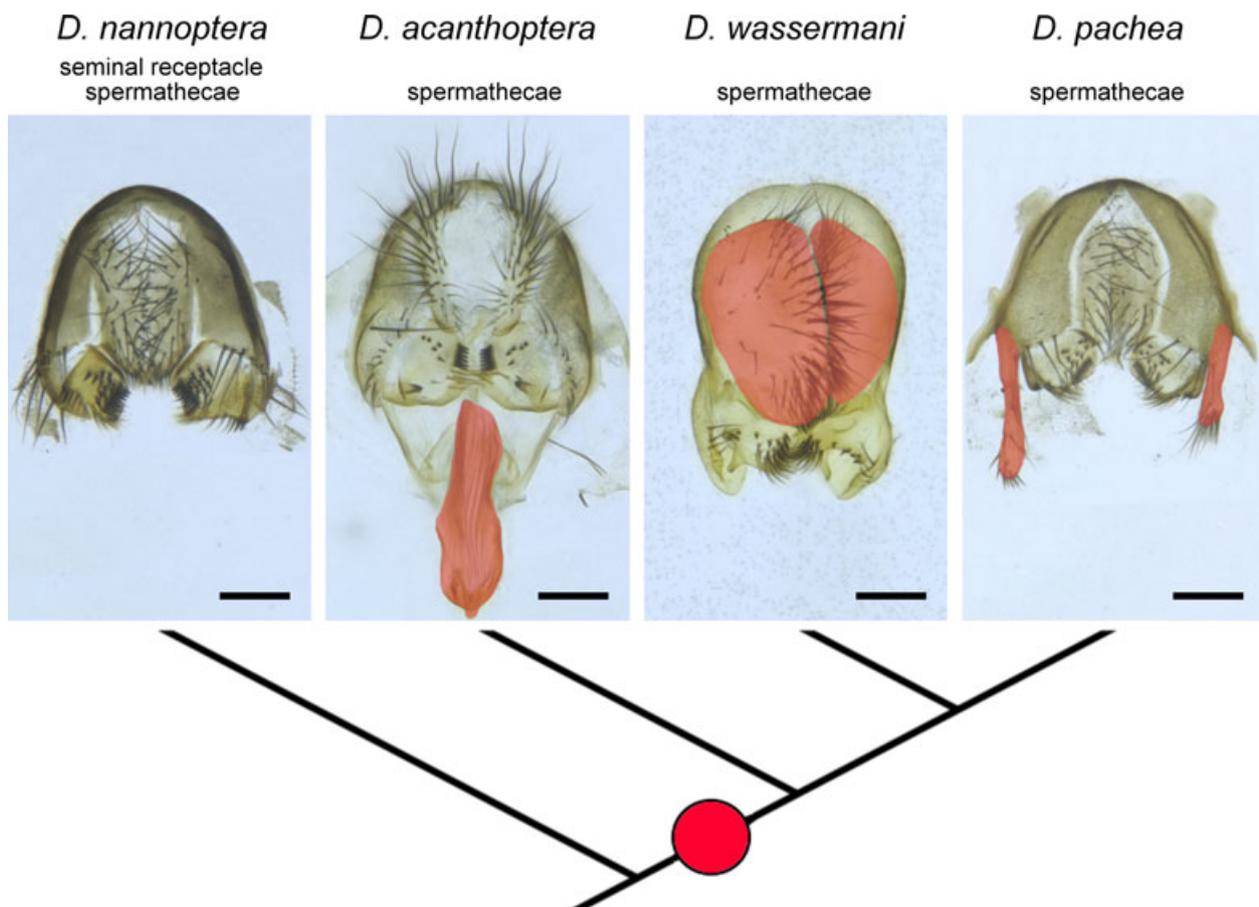


Fig. 2 Hypothetical character evolution of asymmetric male genitalia in the nanoptera species group. The red dot indicates the putative origin of both left–right asymmetric male genitalia and spermathecae-restricted sperm storage. Images below each species names illustrate male external genitalia of each species. Asymmetric parts were artificially coloured in red. Sperm storage organs are indicated below each species name. The scale bar is 100 μ m.

three species. According to this topology, the evolution of asymmetric genitalia and the evolutionary change in sperm storage would have occurred only once, whereas sperm gigantism would have evolved twice within the nanoptera group.

The closest relatives of the nanoptera group are thought to be the bromeliae species group (Van der Linde *et al.*, 2010; Yassin, 2013) and a newly described species, *D. machalilla* (Acurio *et al.*, 2013; A. Acurio, K. Goodman, D.C. Oliveira, V. Rafael & A. Ruiz, unpublished) that was proposed to belong to a new species group, the atalaia group (Acurio *et al.*, 2013). Whereas the nanoptera and bromeliae species groups are part of the virilis-repleta group radiation (Throckmorton, 1975), the particular branching order of lineages leading to the nanoptera group has never been fully resolved (Van der Linde *et al.*, 2010; Oliveira *et al.*, 2012; Yassin, 2013).

Here, we address the phylogenetic relationships of the four species of the nanoptera group and related taxa. Using a multilocus data set of gene coding regions of eight nuclear and three mitochondrial genes, we found that the four described nanoptera group species diverged rapidly. We discuss the nanoptera group radiation with respect to important geological events in Central America. Furthermore, our results allow us to propose a scenario for the evolution of reproductive traits in the nanoptera group.

Materials and methods

Taxon sampling

In addition to the four described species of the nanoptera group, we sequenced two species, *D. bromeliae* and *D. machalilla*, that are hypothesized to be close relatives of the nanoptera group based on morphology and previous phylogenetic analyses (Van der Linde *et al.*, 2010; Oliveira *et al.*, 2012; Yassin, 2013; A. Acurio, K. Goodman, D.C. Oliveira, V. Rafael & A. Ruiz, unpublished). We also included the annulimana group species *D. aracataca* and *D. pseudotalamancana*, and *D. canalinea* of the canalinea group to represent sister lineages of the repleta group. From the repleta group, we included four representative species, *D. inca* (inca subgroup), *D. mettleri* (mulleri subgroup), *D. buzzatii* (mulleri subgroup, buzzatii species complex) and *D. mojavenensis* (mulleri subgroup, mojavenensis species complex). *Drosophila virilis* and *D. robusta* were chosen as distant lineages of the repleta-virilis radiation (Van der Linde *et al.*, 2010; Yassin, 2013) and *D. funebris* and *D. grimshawi* were used to root the phylogeny. The latter two species belong to different species radiations of the *Drosophila* subgenus (Throckmorton, 1975; Van der Linde *et al.*, 2010; Yassin, 2013). Flies were obtained from the *Drosophila* Species Stock Center (Table S1), except for *D. machalilla* and *D. inca* (both collected by A. Acurio) and *D. buzzatii* (provided by Jean David).

DNA sequencing

Genomic DNA was obtained in a single extraction per species including 2–5 adults using the DNeasy blood and tissue kit (QIAGEN, Hilden, Germany). Partial genomic regions of eight nuclear genes (*amy*, *amyrel*, *boss*, *fkh*, *marf*, *sinA*, *snf*, *wee*) and three mitochondrial genes (*ND2*, *COI*, *COII*) were amplified by PCR with gene-specific or degenerate primers (Liu & Beckenbach, 1992; Bonacum *et al.*, 2001; Wang *et al.*, 2006; Da Lage *et al.*, 2007; Table S2). Degenerate oligonucleotides optionally contained T7 or SP6 universal primer sequences at their 5' end (Table S2), following Bonacum *et al.* (2001). For PCR amplifications, we used 0.4 μ M oligonucleotides, 1 u GoTaq[®] DNA Polymerase (Promega, Fitchburg, WI, USA) per 35 μ L reaction volume, 2 mM MgCl₂ and 200 μ M dNTP, and reactions were carried out using standard thermocycle conditions. PCR products were purified and Sanger-sequenced with gene-specific primers or with T7, SP6 universal primers at Cogenics (www.cogenics.com, Beckman Coulter, Brea, California, USA). Sequence data (GenBank accession numbers KF632591–KF632711; Table S3) were examined and aligned with Geneious 6.1.3 (Biomatters, Auckland, New Zealand). Additional sequence data were retrieved from GenBank (Table S3). We generated a data set that contained all the selected homologous genomic regions of all species. DNA sequences were aligned using MAFFT (Katoh *et al.*, 2002). Nuclear loci were tested for recombination with the pairwise homoplasy index (Φ_w) statistics using Phi-Pack (Bruen *et al.*, 2006). No evidence of recombination was detected (Table S4). A total of 121 polymorphic sites were detected within single sequences based on the presence of double peaks in sequencing chromatograms. Thirty-eight of these polymorphic sites were found in the nanoptera species group sequences. All polymorphic sites were excluded from the analysis. Furthermore, noncoding DNA sequences were removed from the data set, as well as a short region of the *wee* locus (sequences homologous to positions 267–306 in *D. pachea wee*, accession number KF632622), which was difficult to align and that did not contain any parsimony informative sites in the nanoptera species group. The extremities of each locus-specific alignment were also trimmed to be in codon frame. Alignments were realigned with the Geneious translation alignment program and either used separately or concatenated in the following order: amy-amyrel-boss-fkh-marf-sinA-snf-wee-ND2-COI-COII. The number of informative sites was calculated using MEGA4 (Tamura *et al.*, 2007).

Phylogenetic analysis

Phylogenetic analysis was performed using both maximum likelihood and Bayesian approaches. For all analyses, models of nucleotide substitution were selected using the Akaike Information Criterion as cal-

culated in jModelTest 2.1.3 (Posada, 2008; Darriba *et al.*, 2012). Maximum likelihood inference was carried out on the concatenated data set in PhyML (Guindon & Gascuel, 2003) using a GTR+I+G model of nucleotide substitution. Node support was determined by performing 100 bootstrap replicates.

Two types of Bayesian analyses were carried out in BEAST v1.7.5 (Drummond *et al.*, 2012). The first analysis was carried out on the concatenated data set using nine partitions with individual and unlinked models of nucleotide substitution (Table S5). The partitions corresponded to the eight nuclear loci (*amy*, *amyrel*, *boss*, *fkh*, *marf*, *sinA*, *snf* and *wee*) plus a ninth partition for mitochondrial sequences (*ND2*, *COI*, *COII*). Mitochondrial genes were combined into one partition because they are located in the same order and orientation in the mitochondrial genome, and largely evolve as a single unit with little to no recombination (Ballard, 2000). Clock models were linked, and a common strict clock rate was assumed for all partitions using the Yule birth process tree prior. We also estimated a species tree using *BEAST (Heled & Drummond, 2010). For the *BEAST analysis, tree and clock models were unlinked for each partition and a relaxed exponential clock model was assumed.

For species divergence time estimates, we set priors for most recent common ancestors (MRCA) using estimates from Obbard *et al.* (2012) for the splits *D. grimshawi* – *D. virilis*: 13 ± 2.5 Ma and *D. mojavensis* – *D. virilis*: 10 ± 2.5 Ma (estimates A). Alternatively, calibration dates for the divergence of *D. grimshawi* – *D. virilis*: 42.9 Ma \pm 8.7 (Tamura *et al.*, 2004), *D. mojavensis* – *D. virilis*: 26 ± 6 Ma (Russo *et al.*, 1995; Spicer & Bell, 2002; Oliveira *et al.*, 2012), *D. mojavensis* – *D. buzzatii*: 11.3 ± 2 Ma and *D. mojavensis* – *D. mettleri*: 16.3 ± 2 Ma (Oliveira *et al.*, 2012) were used (estimates B). Priors were assumed to follow a normal distribution with the mean and a standard deviation according to the literature estimates. Markov-Chain Monte-Carlo (MCMC) runs were performed with a chain length of 10^8 generations and were recorded every 1000 generations. Estimates were computed with Tracer version 1.5 (Rambaut & Drummond, 2009), and MCMC output analysis was carried out using TreeAnnotator (Drummond *et al.*, 2012). The first 2000 sampled trees were discarded as the burn-in. Phylogenies were visualized and annotated with Figtree version 1.4 (Rambaut & Drummond, 2012).

Results

Phylogenetic analysis

To analyse the phylogenetic relationships of the nanoptera group, we gathered DNA sequences of partial coding gene regions of eight nuclear genes (*amy*, *amyrel*, *boss*, *fkh*, *marf*, *sinA*, *snf*, *wee*) and three mitochondrial

genes (*ND2*, *COI*, *COII*) from 17 species. The entire data set comprised 6810 aligned positions, including 4208 constant positions, 695 single variable positions and 1907 parsimony informative (28%) positions (Table S3).

Phylogenetic analysis of the concatenated, nine-partition data set was performed using a maximum likelihood approach in PhyML (Guindon & Gascuel, 2003) and by Bayesian inference in BEAST (Drummond *et al.*, 2012). Phylogenetic relationships inferred from these analyses resulted in identical tree topologies, but with varying node support values (Fig. 1). The resulting phylogeny supports Acurio *et al.*'s findings (A. Acurio, K. Goodman, D.C. Oliveira, V. Rafael & A. Ruiz, unpublished) that the nanoptera species group is a sister clade of the atalaia species group, with *D. machalilla* being more closely related to the nanoptera clade than to *D. bromeliae* (bromeliae group). In our phylogeny, *D. inca* and *D. canalinea* form a lineage sister to the repleta group. Furthermore, both the Bayesian and the maximum likelihood phylogeny provided, for the first time, strong support for the monophyly of annulimana species, *D. aracataca* and *D. pseudotalamanca*, which we found to be more closely related to the nanoptera group than to the repleta group. Phylogenetic relationships within the nanoptera group were relatively well-resolved in the Bayesian analysis, but not in the maximum likelihood analysis. Our results from the Bayesian analysis were congruent with the phylogenetic relationships previously suggested based on morphological data (Heed, 1982) (Fig. 1a).

Analysis of concatenated multilocus data has recently been criticized as it poorly integrates locus-specific phylogenetic signals and can lead to false phylogenetic inferences with high statistical support (Song *et al.*, 2012). Therefore, we also analysed our data set with *BEAST (Heled & Drummond, 2010), an extension of the BEAST package that incorporates coalescence models to estimate a species tree from multiple gene-specific phylogenies. The topology of the species tree inferred in *BEAST was similar to the phylogeny obtained with the concatenated data set (Fig. 1a, Fig. S1), except that relationships within the nanoptera group differed, with *D. pachea* being sister to the clade containing *D. acanthoptera* and the sister species pair *D. nanoptera* and *D. wassermani* (Fig. S1). However, the posterior probability for the corresponding nodes were low, suggesting that our data set might not contain enough information for species tree estimation using *BEAST. Within the nanoptera group, we observed only 120 parsimony informative sites across all genes in the data set (Table S3). The mean length of DNA sequence per nuclear locus was 477 ± 178 bp (SD), which, on average, included only 5 ± 3 (SD) parsimony informative sites among the nanoptera species group (Table S3). We wondered whether the number of informative sites per locus was too low for a

coalescent multilocus phylogenetic approach to produce a well-resolved species tree. To estimate the number of loci that would be necessary to reliably establish nanoptera group relationships using *BEAST, we produced partial data sets containing 2, 4, 6 or 8 loci and inferred the phylogeny of each data set. The average node support was calculated and the number of necessary loci was approximated by a logarithmic regression (Fig. S2). Whereas we are aware that the number of parsimony informative sites per locus is low and that the data sets are partially redundant, this analysis showed that for the entire phylogeny, the average posterior probability was 0.90 when six and eight loci were used. Support for the nodes within the nanoptera group also increased with the number of loci, but at a much lower rate. We estimated that approximately 60 loci would be required to obtain a posterior probability of 0.90 for the internal nodes within the nanoptera group.

Divergence time estimates

We estimated the divergence times of the nanoptera group radiation and the splits of *D. machalilla* and *D. bromelia* from the branch leading to the nanoptera species group. There is conflicting information about species divergence times in *Drosophila*. Most estimates are based on the phylogeny of Hawaiian Drosophilidae where species divergence times can be approximated based on the ages of the islands they inhabit (Price & Clague, 2002). Recently, Obbard *et al.* (2012) proposed a refinement of this approach to take lineage-specific variation of mutation rates into account (Obbard *et al.*, 2012). This new approach suggested a younger age for the virilis-repleta radiation, of about 10 Ma compared to the previous estimates of 20 Ma (see Material and methods). We computed species divergence times either based on Obbard *et al.* (2012) (dates A) or based on previous species divergence estimates (dates B) (Table 1).

As the calibration estimates in A were about half the ages in B, divergence time estimates of dates A were expectedly also half the age compared to estimates B.

Table 1 Divergence time estimates.

| Dated nodes | Divergence time estimates* | |
|---|----------------------------|---------------------|
| | A | B |
| <i>D. mojavensis</i> – <i>D. buzzatii</i> | 5.5 Ma (3.5–7.5) | 12.2 Ma (10.0–14.5) |
| <i>D. mojavensis</i> – <i>D. mettleri</i> | 7.5 Ma (4.8–10.1) | 16.7 Ma (13.7–19.6) |
| <i>D. pachea</i> – <i>D. wassermani</i> | 3.0 Ma (1.9–4.1) | 6.7 Ma (5.3–8.1) |
| <i>D. pachea</i> – <i>D. nanoptera</i> | 3.7 Ma (2.4–5.0) | 8.3 Ma (6.7–10.0) |
| <i>D. pachea</i> – <i>D. machalilla</i> | 7.5 Ma (4.9–10.2) | 16.9 Ma (13.7–20.1) |
| <i>D. pachea</i> – <i>D. bromeliae</i> | 8.4 Ma (5.4–11.3) | 17.9 Ma (15.8–20.0) |
| <i>D. virilis</i> – <i>D. mojavensis</i> | 10.6 Ma (7.0–14.3) | 18.9 Ma (15.3–22.4) |

*Estimates are the posterior means with 95% highest posterior density intervals.

We estimated that the nanoptera group lineage diverged about 3.7 Ma (dates A) or 8.3 Ma (dates B). Furthermore, the most recent split of *D. pachea* and *D. wassermani* was estimated to have occurred shortly thereafter, about 3.0 Ma (dates A) or 6.7 Ma (dates B). The *D. machalilla* lineage separated from the nanoptera group about 7.5 Ma (dates A) or 16.9 Ma (dates B) and the bromeliae group separated from the nanoptera group lineage about 8.4 Ma (dates A) or 17.9 Ma (dates B). Based on the conflicting calibration, these dates do not precisely estimate speciation events, but they put the nanoptera group radiation into an approximate time frame.

Discussion

Origin of the nanoptera group

The four species of the nanoptera group are endemic to distinct regions of Mexico (Fig. 1b). Our phylogenetic analysis uncovered three closely related outgroups to the nanoptera group: *D. machalilla*, the *bromeliae* species group (comprising five species including *D. bromeliae*) and the members of the annulimana group. Members of these species groups are primarily found in South America but also in Central America (Fig. 1b) (Sturtevant, 1916; Duda, 1927; Pavan & da Cunha, 1947; Do Val & Marques, 1996; Da Silva *et al.*, 2004; Markow & O'Grady, 2005; Acurio *et al.*, 2013). These species distributions thus suggest that the ancestor of the four nanoptera group species may have originated from South America. Interestingly, our dating analysis estimates that the nanoptera group diverged from *D. machalilla* around 16.9 Ma (B) – 7.5 Ma (A). This time period corresponds to the closure of the Isthmus of Panama, about 15–9 Ma (Montes *et al.*, 2012b), suggesting that the ancestor of the nanoptera group may have migrated over the newly formed isthmus from South America. Most species were found to migrate across the isthmus much later, at about 3–2 Ma (Leigh *et al.*, 2014). However, exceptions are known such as the extinct carnivora *Cyonasua* and ground sloths, which migrated about 9 Ma from north to south and south to north, respectively (Webb, 1976). Furthermore, recent data suggest that the isthmus was already passable for stingless bees at late Eocene and early Miocene times (20–15 Ma), which migrated from South to Central America (Roubik & de Camargo, 2012). The isthmus might have faced multiple events of temporary land bridge formations and disconnections, allowing a few species to cross continents before a permanent land bridge formed about 4–3 Ma (Webb, 1976; Roubik & de Camargo, 2012; Stone, 2013).

Drosophila species have been extensively sampled in Mexico and Central America (Patterson & Stone, 1952), but multiple areas known as biodiversity hotspots in South America are still unexplored. An origin of the

nannoptera group in South America would suggest that yet undescribed close out group species of the nannoptera group might be present in these geographic areas.

Radiation of the nannoptera species group

Within the nannoptera species group, we observed only 120 parsimony informative sites in our entire data set (Table S6), whereas 850 nucleotide changes were lineage-specific. The distribution of mutations indicates that the four species of the nannoptera group diverged within a relatively short time period. Phylogenetic inference can be particularly difficult in these cases especially if node support remains low with increasing data. One interpretation of the current data is that the ancestral lineage diverged nearly simultaneously into the four described extant lineages (Walsh *et al.*, 1999; Humphries & Winker, 2010). Such scenario might be expected from species with a large geographic range where peripatric speciation can occur in different regions. One well-studied example is the *D. simulans* species complex, where *D. mauritiana* and *D. sechellia* diverged independently on islands that were geographically separated from the cosmopolitan species *D. simulans* (Garrigan *et al.*, 2012).

Alternatively, very short internode distances could result from a rapid succession of divergence events that could be inferred with increasing amounts of data (soft polytomy). Our current sequence data are insufficient to distinguish between a soft and a hard polytomy in the nannoptera species group. The rapid and ongoing decrease in high-throughput sequencing costs now makes it more practical to sequence and to compare whole genomes for future studies aiming at a better resolution of the nannoptera group phylogeny.

Phylogenetic relationships and evolution of reproductive traits within the nannoptera group

Under the hypothesis of a soft polytomy in the nannoptera species group, the phylogeny of the nannoptera group that we inferred using the concatenated data (Fig. 1a) appears to propose the most plausible scenario, despite a low node support in one of our analyses (PhyML maximum likelihood analysis). Indeed, several lines of evidence corroborate this topology. First, compared to the previous molecular phylogenetic analysis of Oliveira *et al.* (2012), which hypothesized different relationships for the nannoptera group, our phylogeny is based on a higher number of loci and on an increased number of relevant out group species close to the nannoptera group.

Second, our inferred topology recapitulates the species relationships presented by Heed (1982) based on internal reproductive organ anatomy and by Pitnick *et al.* (1999) independently based on cytochrome oxidase data. Third, it is congruent with chromosome

inversions. Comparisons of polytene chromosome banding patterns revealed that *D. nanoptera* and *D. wassermani* have a homosequential 'ancestral-like' chromosomal organization, whereas *D. acanthoptera* and *D. pachea* are derived with three and one inversion, respectively (Ward & Heed, 1970). A fourth, polymorphic, inversion is also found in *D. pachea* and is not detected in the other species of the nannoptera group (Etges *et al.*, 1999).

Fourth, our inferred topology is consistent with a parsimonious scenario of the evolution of the unusual reproductive characters within the nannoptera group. Genital asymmetry is found in *D. acanthoptera* (Vilela & Baechli, 1990), *D. wassermani* (Pitnick & Heed, 1994) and *D. pachea* (Pitnick & Heed, 1994; Lang & Orgogozo, 2012) (Fig. 2), whereas *D. nanoptera* (Vilela & Baechli, 1990), as well as the species *D. bromeliae*, *D. speciosa* and *D. aguape* of the bromeliae group and *D. machalilla* (atalaia group) have symmetric genitalia (Do Val & Marques, 1996; Da Silva *et al.*, 2004; Acurio *et al.*, 2013). Therefore, a single evolutionary transition to asymmetric genitalia might have occurred in the nannoptera group. Even though the asymmetry involves different male genitalia organs in each species, a common genetic and developmental process may underlie these distinct morphological asymmetries. We currently are trying to unravel the genetic factors that determine the asymmetric development of male genitalia in the three nannoptera species. In particular, we are testing whether genitalia clockwise rotation (Feuerborn, 1922; Suzanne *et al.*, 2010) during pupal development could be the signal that triggers differential growth between the left and right parts of various organs in distinct species. Furthermore, our inferred topology is consistent with a single evolutionary change in sperm storage in the nannoptera group. After copulation, females of *D. acanthoptera*, *D. wassermani* and *D. pachea* exceptionally use only the spermathecae to store the sperm and not the seminal receptacle as is typical for *Drosophilidae*, including *D. nanoptera* (Pitnick *et al.*, 1999). Future efforts are required to examine how copulation position might affect sperm transfer in the nannoptera species group and to determine whether asymmetric male genitalia and unusual sperm storage are functionally linked. Finally, according to our inferred phylogeny, sperm gigantism would have evolved twice independently in the nannoptera group, which is consistent with other reported instances of rapid evolution of sperm size in *Drosophila* (Pitnick *et al.*, 1995).

Evolutionary history of the nannoptera species group

Species divergence estimates for the virilis-repleta radiation vary greatly, from 30–20 Ma (Russo *et al.*, 1995; Spicer & Bell, 2002; Tamura *et al.*, 2004; Oliveira *et al.*, 2012) to 10 Ma (Obbard *et al.*, 2012) when adjusting

for lineage-specific mutation rates. We performed two separate estimations (Table 1) to account for two incompatible calibrations of species divergence estimates. The two sets of estimates (A and B) for the radiation of the nanoptera group approximate the lower and upper bounds of the geological time approximation of sea floor spreading of the southern Gulf of California. Formation of the Baja California peninsula started approximately 12 Ma due to changes in continental plate tectonics (Umhoefer, 2011). The peninsula itself formed along an almost north-south-directed rift, now partially covered by the Gulf of California. Landmasses separated from the continent as a result of complex geological events about 6–2.5 Ma and successively formed the peninsula (Lizarralde *et al.*, 2007; Umhoefer, 2011). Thus, the divergence of *D. wassermani* and *D. pachea* might have been influenced by the formation of the Baja California peninsula and by the separation of these landmasses from the continent (Heed, 1982). Whether senita cacti were already present in the forming Baja peninsula and whether *D. pachea* or its predecessors were already feeding on senita cactus when landmasses disconnected from the continent is unknown. A phylogenetic analysis of the senita cactus and its closely related species, together with estimations of divergence times, would be helpful to try to infer the evolutionary history of the close ecological relationship between *D. pachea* and its host cactus. The distribution area of *D. wassermani* is limited to the north by the Trans-Mexican Volcanic Belt and by the Sierra Madre Occidental mountain range (Fig. S3), which originated about 17–7 and 38–25 Ma, respectively (Ferrari *et al.*, 1999). A plausible scenario is that these mountains formed an obstacle for the ancestor of *D. wassermani* and *D. pachea*, which colonized further northern regions through the coastal lowlands of north-west Mexico (Heed, 1982). As this region successively re-arranged into the Baja California peninsula, the Gulf of California created a natural barrier and could have led to the isolation of *D. pachea* in the north and *D. wassermani* in the south (Fig. S3).

Drosophila machalilla, the most closely related out-group of the nanoptera group, is a recently described species that was collected in traps containing *Opuntia* cactus, and the columnar cactus *Armatocereus cartwrightianus* (Britton & Rose, 1920) was proposed to be their native host plant (Acurio *et al.*, 2013). As all nanoptera group species also feed on columnar cacti, the MRCA of *D. machalilla* and the nanoptera species group was likely to be already cactophilic. Our results suggest that the major radiation of succulent plants, which occurred in the past 10 million years in North and South America (Arakaki *et al.*, 2011), could have then contributed to shifts in cactus hosts and to speciation in the nanoptera group.

In summary, our results indicate that the four species of the nanoptera group originated within a short

time period. Our approximations of species divergence times suggest that the emergence of the southern Gulf of California might have been involved in the split between *D. pachea* and *D. wassermani*. The branching order of basal repleta lineages reveals that the annulimana species *D. aracataca* and *D. pseudotalamancana* are the most closely related taxa to the nanoptera and bromeliae species groups. Our phylogenetic analysis suggests that evolution of asymmetric genital and unusual sperm storage have evolved only once within the nanoptera group, and that the ancestor of the nanoptera group was already feeding on columnar cacti.

Acknowledgments

We are especially grateful to Amir Yassin for helpful discussions on the experimental design and on the manuscript. In addition, we thank Jean David for *D. buzzatii* specimen. ML is supported by the Fondation pour la Recherche Médicale (FRM) postdoctoral fellowship SPF20121226328. This work was also supported by a CNRS ATIP-AVENIR grant, given to VO. Sampling of Ecuadorian *Drosophila* specimen was carried out by AEA with the Scientific Research Permission 0016-071C-FAU-DNBAPVS/MA.

References

- Acurio, A., Rafael, V., Cespedes, 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.
- Arakaki, M., Christin, P.-A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R.M. *et al.* 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci.* **108**: 8379–8384.
- Ballard, J.W.O. 2000. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* Subgroup. *J. Mol. Evol.* **51**: 48–63.
- Bonacum, J., DeSalle, R., O'Grady, P., Olivera, D., Wintermute, J. & Zilversmit, M. 2001. New nuclear and mitochondrial primers for systematics and comparative genomics in Drosophilidae. *Drosoph. Inf. Serv.* **84**: 201–204.
- Britton, N. & Rose, J. 1920. *The Cactaceae: Descriptions and Illustrations of Plants in the Cactus Family*. Vol. 2, pp. 100. The Carnegie Institution of Washington, Washington.
- Bruen, T.C., Philippe, H. & Bryant, D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681.
- 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.
- Da Lage, J.-L., Kergoat, G.J., Maczkowiak, F., Silvain, J.-F., Cariou, M.-L. & Lachaise, D. 2007. A phylogeny of Drosophilidae using the Amyrel gene: questioning the *Drosophila melanogaster* species group boundaries. *J. Zool. Syst. Evol. Res.* **45**: 47–63.

- Da Silva, A., De, A.R. & Martins, M.B. 2004. A new anthophilic species of *Drosophila* Fallen belonging to the bromeliae group of species (Diptera, Drosophilidae). *Rev. Bras. Zool. Soc. Bras. Zool.* **21**: 435–437.
- Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**: 772.
- Do Val, F.C. & Marques, M.D. 1996. Drosophilidae (Diptera) from the Pantanal of Mato Grosso (Brazil), with the description of a new species belonging to the bromeliae group of the genus *Drosophila*. *Pap. Avulsos Zool.* **39**: 223–230.
- 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.
- Duda, O. 1927. Die sudamerikanischen Drosophiliden (Dipteren) unter Berücksichtigung auch der anderen neotropischen sowie der nearktischen Arten. *Arch. Für Naturgeschichte* **91**: 1–228.
- Engelmann, G. 1852. Four hundred and twenty-eight meeting, May 27, 1856. Annual Meeting; Synopsis of the Cactaceae of the Territory of the United States and Adjacent Regions. *Proc. Am. Acad. Arts Sci.* **3**: 256–314.
- Etges, W.J., Johnson, W.R., Duncan, G.A., Huckins, G. & Heed, W.B. 1999. Ecological genetics of cactophilic *Drosophila*. In: *Ecology of Sonoran Desert Plants and Plant Communities* (R. Robichaux, ed.), pp. 164–214. University of Arizona Press, Tucson, Arizona.
- Ferrari, L., López-Martínez, M., Aguirre-Díaz, G. & Carrasco-Núñez, G. 1999. Space-time patterns of Cenozoic arc volcanism in central Mexico: from the Sierra Madre Occidental to the Mexican Volcanic Belt. *Geology* **27**: 303–306.
- Feuerborn, H.J. 1922. Das Hypopygium “inversum” und “circumversum” der Dipteren. *Zool. Anz.* **55**: 89–213.
- Garrigan, D., Kingan, S.B., Geneva, A.J., Andolfatto, P., Clark, A.G., Thornton, K.R. *et al.* 2012. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Res.* **22**: 1499–1511.
- Gregory-Wodzicki, K.M. 2000. Uplift history of the Central and Northern Andes: a review. *Geol. Soc. Am. Bull.* **112**: 1091–1105.
- Guindon, S. & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**: 696–704.
- Hastings, J.R., Turner, R.M. & Warren, D.K. 1972. An atlas of some plant distributions in the Sonoran Desert. University of Arizona, Institute of Atmospheric Physics.
- Heed, W.B. 1982. The origin of *Drosophila* in the Sonoran Desert. In: *Ecological Genetics and Evolution: The Cactus-Yeast- Drosophila Model System* (J.S.F. Barker & W.T. Starmer, eds), pp. 65–80. Academic Press, Sydney, Australia.
- Heed, W.B. & Kircher, H.W. 1965. Unique sterol in the ecology and nutrition of *Drosophila pachea*. *Science* **149**: 758–761.
- Heled, J. & Drummond, A.J. 2010. Bayesian inference of species trees from Multilocus data. *Mol. Biol. Evol.* **27**: 570–580.
- Huber, B.A., Sinclair, B.J. & Schmitt, M. 2007. The evolution of asymmetric genitalia in spiders and insects. *Biol. Rev.* **82**: 647–698.
- Humphries, E.M. & Winker, K. 2010. Working through polytomies: Auklets revisited. *Mol. Phylogenet. Evol.* **54**: 88–96.
- Hunter, A.S. 1970. *Drosophila* of Venezuela. *Drosoph. Inf. Serv.* **45**: 124.
- Katoh, K., Misawa, K., Kuma, K. & Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**: 3059–3066.
- Lang, M. & Orgogozo, V. 2012. Distinct copulation positions in *Drosophila pachea* males with symmetric or asymmetric external genitalia. *Contr. Zool.* **81**: 87–94.
- Lang, M., Murat, S., Clark, A.G., Gouppil, G., Blais, C., Matzkin, L.M. *et al.* 2012. Mutations in the neverland gene turned *Drosophila pachea* into an obligate specialist species. *Science* **337**: 1658–1661.
- Leigh, E.G., O’Dea, A. & Vermeij, G.J. 2014. Historical biogeography of the Isthmus of Panama. *Biol. Rev.* **89**: 148–172.
- Lindsay, G. 1963. The genus *Lophocereus*. *Cactus Succul. J.* **35**: 176–192.
- Liu, H. & Beckenbach, A.T. 1992. Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Mol. Phylogenet. Evol.* **1**: 41–52.
- Lizarralde, D., Axen, G.J., Brown, H.E., Fletcher, J.M., Gonzalez-Fernandez, A., Harding, A.J. *et al.* 2007. Variation in styles of rifting in the Gulf of California. *Nature* **448**: 466–469.
- Markow, T. & O’Grady, P. 2005. *Drosophila*, A Guide to Species Identification and Use. Academic Press, Elsevier, Amsterdam.
- Markow, T.A. & O’Grady, P.M. 2007. *Drosophila* biology in the genomic age. *Genetics* **177**: 1269–1276.
- Montes, C., Bayona, G., Cardona, A., Buchs, D.M., Silva, C.A., Morón, S. *et al.* 2012a. Arc-continent collision and orocline formation: closing of the Central American seaway. *J. Geophys. Res.* **117**: 1–25.
- Montes, C., Cardona, A., McFadden, R., Morón, S.E., Silva, C.A., Restrepo-Moreno, S. *et al.* 2012b. Evidence for middle Eocene and younger land emergence in central Panama: implications for isthmus closure. *Geol. Soc. Am. Bull.* **124**: 780–799.
- 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., 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. Phylogenet. Evol.* **64**: 533–544.
- Patterson, J.T. & Stone, W.S. 1952. *Evolution in the Genus Drosophila*. Macmillan, New York.
- Patterson, J. & Wheeler, M. 1942. Description of new species of the subgenera *Hirtodrosophila* and *Drosophila*. Univ Tex Publ. Austin.
- Pavan, C. & da Cunha, A.B. 1947. *Espécies Brasileiras de Drosophila*. Universidade de São Paulo, São Paulo.
- Pitnick, S. & Heed, W. 1994. New Species of Cactus-Breeding *Drosophila* (Diptera: Drosophilidae) in the *Nanoptera* Species Group. *Ann. Entomol. Soc. Am.* **87**: 307–310.
- Pitnick, S., Markow, T.A. & Spicer, G.S. 1995. Delayed male maturity is a cost of producing large sperm in *Drosophila*. *Proc. Natl Acad. Sci.* **92**: 10614–10618.
- Pitnick, S., Spicer, G.S. & Markow, T. 1997. Phylogenetic examination of female incorporation of ejaculate in *Drosophila*. *Evolution* **51**: 833–845.
- Pitnick, S., Markow, T. & Spicer, G.S. 1999. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. *Evolution* **53**: 1804–1822.

- Posada, D. 2008. jModelTest: Phylogenetic Model Averaging. *Mol. Biol. Evol.* **25**: 1253–1256.
- Price, J.P. & Clague, D.A. 2002. How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proc. R. Soc. Lond. B Biol. Sci.* **269**: 2429–2435.
- Rambaut, A. & Drummond, A. 2009. Tracer v1.5.
- Rambaut, A. & Drummond, A. 2012. Figtree v1.4.
- Roubik, D.W. & de Camargo, J.M.F. 2012. The Panama microplate, island studies and relictual species of *Melipona* (Melikerria) (Hymenoptera: Apidae: Meliponini). *Syst. Entomol.* **37**: 189–199.
- Russo, C.A., Takezaki, N. & Nei, M. 1995. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* **12**: 391–404.
- Song, S., Liu, L., Edwards, S.V. & Wu, S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci.* **109**: 14942–14947.
- Spicer, G.S. & Bell, C.D. 2002. Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. *Ann. Entomol. Soc. Am.* **95**: 156–161.
- Stone, R. 2013. Battle for the Americas. *Science* **341**: 230–233.
- Sturtevant, A.H. 1916. Notes on North American Drosophilidae with descriptions of twenty-three new species. *Ann. Entomol. Soc. Am.* **9**: 323–343.
- Suzanne, M., Petzoldt, A.G., Spéder, P., Coutelis, J.-B., Steller, H. & Noselli, S. 2010. Coupling of apoptosis and l/r patterning controls stepwise organ looping. *Curr. Biol.* **20**: 1773–1778.
- Tamura, K., Subramanian, S. & Kumar, S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol. Biol. Evol.* **21**: 36–44.
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol. Biol. Evol.* **24**: 1596–1599.
- Throckmorton, L.H. 1975. The phylogeny, ecology, and geography of *Drosophila*. In: *Handbook of Genetics* (R.C. King, ed.), pp. 421–469. Plenum Publishing Corporation, New York.
- Umhoefer, P.J. 2011. Why did the Southern Gulf of California rupture so rapidly? – Oblique divergence across hot, weak lithosphere along a tectonically active margin. *GSA Today* **21**: 4–10.
- Van der Linde, K., Houle, D., Spicer, G.S. & Stepan, S.J. 2010. A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genet. Res.* **92**: 25–38.
- Vilela, C. & Baechli, G. 1990. Taxonomic studies on Neotropical species of seven genera of Drosophilidae (Diptera). *Mitt. Schweiz Ent. Ges.* **63**: 1–332.
- Walsh, H.E., Kidd, M.G., Moum, T. & Friesen, V.L. 1999. Polytomies and the Power of Phylogenetic Inference. *Evolution* **53**: 932–937.
- Wang, B., Park, J., Watabe, H., Gao, J., Xiangyu, J., Aotsuka, T. et al. 2006. Molecular phylogeny of the *Drosophila virilis* section (Diptera: Drosophilidae) based on mitochondrial and nuclear sequences. *Mol. Phylogenet. Evol.* **40**: 484–500.
- Ward, B.L. & Heed, W.B. 1970. Chromosome phylogeny of *Drosophila pachea* and related species. *J. Hered.* **61**: 248–258.
- Webb, S.D. 1976. Mammalian faunal dynamics of the great American interchange. *Paleobiology* **2**: 220–234.
- Wheeler, M. 1949. Taxonomic studies on the Drosophilidae. In: *Studies in the Genetics of Drosophila* (J.T. Patterson, ed.), pp. 157–195. University of Texas Publications No 4920, Austin.
- Yassin, A. 2013. Phylogenetic classification of the Drosophilidae Rondani (Diptera): the role of morphology in the post-genomic era. *Syst. Entomol.* **38**: 349–364.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Species resources.

Table S2 Oligonucleotides used in this study.

Table S3 GenBank accession numbers of the data set.

Table S4 Partition specific models of nucleotide evolution and parameters.

Table S5 Test of recombination.

Table S6 Parsimony informative sites.

Figure S1 *BEAST analysis and gene-specific tree topologies.

Figure S2 Approximation of the number of loci required for a *BEAST multilocus analysis of the nanno-pteran group with reliable node support.

Figure S3 Hypothetical speciation scenario of *Drosophila wassermani* and *D. pachea*.

Received 12 November 2013; revised 20 December 2013; accepted 24 December 2013