Saltational Episodes of Reticulate Evolution in the *Drosophila saltans* Species Group

Carolina Prediger [1,2] Erina A. Ferreira [2] Samara Videira Zorzato [1,3] Aurélie Hua-Van [2] Lisa Klasson [4] Wolfgang J. Miller [5] Amir Yassin [2,3,†*] Lilian Madi-Ravazzi [1,†*]

[1]Department of Biology, UNESP—São Paulo State University, São José do Rio Preto, São Paulo, Brazil

[2]Laboratoire Évolution, Génomes, Comportement et Écologie, CNRS, IRD, Université Paris-Saclay, Gif-sur-Yvette, France

[3]Institut de Systématique, Évolution, Biodiversité (ISYEB), CNRS, MNHN, EPHE, Sorbonne Université, Univ. des Antilles, Paris, France

[4]Molecular Evolution, Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

[5]Lab Genome Dynamics, Department Cell & Developmental Biology, Center for Anatomy and Cell Biology, Medical University of Vienna, Vienna, Austria

[†]These authors contributed equally.

***Corresponding authors:** E-mails: amir.yassin@universite-paris-saclay.fr; lilian.madi@unesp.br.

Associate editor: Greg Gibson

Abstract

Phylogenomics reveals reticulate evolution to be widespread across taxa, but whether reticulation is due to low statistical power or it is a true evolutionary pattern remains a field of study. Here, we investigate the phylogeny and quantify reticulation in the *Drosophila saltans* species group, a Neotropical clade of the subgenus *Sophophora* comprising 23 species whose relationships have long been problematic. Phylogenetic analyses revealed conflicting topologies between the X chromosome, autosomes and the mitochondria. We extended the ABBA-BABA test of asymmetry in phylogenetic discordance to cases where no "true" species tree could be inferred, and applied our new test (called 2A2B) to whole genome data and to individual loci. We used four strategies, two based on our new assemblies using either conserved genes or ≥50 kb-long syntenic blocks with conserved collinearity across Neotropical *Sophophora*, and two consisted of windows from pseudo-reference genomes aligned to either an ingroup or outgroup species. Evidence for reticulation varied among the strategies, being lowest in the synteny-based approach, where it did not exceed ∼7% of the blocks in the most conflicting species quartets. High incidences of reticulation were restricted to three nodes on the tree that coincided with major paleogeographical events in South America. Our results identify possible technical biases in quantifying reticulate evolution and indicate that episodic rapid radiations have played a major role in the evolution of a largely understudied Neotropical clade.

Key words: **phylogenomic discordance, genome assembly, historical biogeography, introgression, cyto-nuclear conflicts, Neotropical speciation, *Sophophora*.**

Introduction

Knowledge of phylogenetic relationships among species is a requirement for many evolutionary studies. However, it is often difficult to reconstruct well-resolved bifurcating trees for some clades. This could either be due to the lack of signal in the evaluated data, a condition known as "soft polytomy", or due to persistent phylogenetic conflicts among datasets leading to "hard polytomies" and reticulate patterns of interspecific relationships. A plethora of biological processes could cause such conflicts, including incomplete lineage sorting (Maddison 1989, 1997; Walsh et al. 1999; Townsend et al. 2012), horizontal gene transfer, introgression and hybridization (Schrempf and Szöllősi 2020), and adaptive radiations (Glor 2010). Phylogenetic conflict may also be caused by technical errors, such as, sequencing error, contamination, wrong model selection and general lack of quality control (Philippe et al. 2011). Recent advances in genomic analyses have significantly reduced such errors and, in a wide range of taxa, increased the number of analyzed genes hence helping to resolve early conflicting topologies. However, in many other cases, whole-genome analyses demonstrated persistent phylogenetic conflicts (e.g. in plants (Wickett et al. 2014; Gagnon et al. 2022), birds (Suh 2016), sponges and ctenophores (Philippe et al. 2009; Pick et al. 2010; Chang et al. 2015; Whelan

Received: March 09, 2024. **Revised:** November 26, 2024. **Accepted:** December 02, 2024

et al. 2015; Simion et al. 2017), mammals (Morgan et al. 2013; Romiguier et al. 2013; Doronina et al. 2015), amphibians (Hime et al. 2021), and insects (Owen and Miller 2022)).

Of the different processes that can reduce phylogenetic resolution, introgression has attracted much attention, and led to the development of a number of bioinformatic tools and tests that quantify its extent across the genome (Durand et al. 2011; Pease and Hahn 2015; Malinsky et al. 2021). Site-based methods usually count the number of bi-allelic sites supporting each of three possible topologies in a species triplet with an outgroup (Fig. 1a). Comparisons between the proportions of the three topologies can yield one of four possible outcomes (Fig. 1b): (i) complete trifurcation, all topologies are equally encountered; (ii) incomplete trifurcation, such as in the case of hybrid speciation wherein two topologies significantly exceed the third one but do not significantly differ from each other; (iii) incomplete bifurcation, such as in the case of asymmetric introgression wherein the proportion of all topologies significantly differ; and (iv) complete bifurcation, one topology significantly exceeds the two others, which in their turn have nearly equal proportions. Categories ii and iii are often considered to evidence reticulate evolution. The earliest of introgression tests, ABBA-BABA or the Patterson's *D* statistic (Green et al. 2010; Durand et al. 2011), compared the two later cases (iii and iv), i.e. it presumed that a "true" species tree exists. A later test, HyDe (Blischak et al. 2018), quantifies admixture (γ) from the ratio of shared alleles with the test going from 0 (full isolation) to 0.5 (full hybridization) and therefore it can also cover case ii. The two tests differ in how they measure significance, using jackknifing or bootstrapping in Patterson's *D* and normal approximation in HyDe. Of late, another site-based test was developed using χ^2 to test for deviation of parity between the three topologies as in case i (Sayyari and Mirarab 2018). A unified test that can test the prevalence of each of the four categories across the genome is still lacking.

Reminiscent to the problem of soft and hard polytomies, wherein the statistical power of a phylogenetic analysis largely depends on the size of a locus, interpreting reticulation is largely influenced by how a locus is defined. For example, at the species level, an analysis on a whole concatenated alignment of a quartet should have a greater power (i.e. a total evidence approach). However, sites are often physically linked and evolutionarily-dependent. Consequently, whole-genome analyses may be biased toward genomic regions with more variable sites (Martin et al. 2015). Therefore, phylogenetic analyses are usually conducted at individual loci under the multi-species coalescent model, which often outperforms concatenation in multiple groups of animals (Jiang et al. 2020). In theory, a window size of ~100-kb long is appropriate to overcome the non-independence of sites under realistic recombination rates (Pease and Hahn 2015), but the definition of such windows would largely depend on the taxonomic scale of the analysis. For example, at shallow phylogenetic depths, it is customary to use a pseudo-reference genome approach, whereas reads from multiple species are aligned to a reference genome (Sarver et al. 2017). Here, the distance from the reference genome species can greatly influence the quality of mapping and subsequent analyses due to potential gene duplications and rearrangements (Rick et al. 2023). Alternatively, single-copy orthologous genes, such as those that have been popularized by the BUSCO program (Manni et al. 2021), can be used (Suvorov, Kim, et al. 2022; Suvorov, Scornavacca, et al. 2022). However, such gene are presumably under strong selection, with the relaxation of selection in some lineages may conflate introgression estimates (Frankel and Ané 2023). Longer syntenic blocks may be a better choice to maintain the advantages for methods based on conserved orthologous genes while adding information from less conserved, longer non-coding regions in these blocks. However, the use of such blocks in introgression analyses has not, to the best of our knowledge, been explicitly investigated.

Phylogenetic conflicts have been reported for the jumping fly *D. saltans* species group, a clade of the subgenus *Sophophora* with 23 Neotropical species (de Magalhães 1962; Bächli 2022). The group retains its name from the peculiar "jumping" habit of its larvae; "*the larva seizes its posterior end with its mouthhooks, and stretches. The hooks pull loose suddenly, the larva straightens with considerable force, and as a result is thrown several inches into the air*" (Sturtevant 1942). The group was divided into five species subgroups, namely, *saltans*, *parasaltans*, *cordata*, *elliptica* and *sturtevanti* subgroups, mostly on the basis of male genitalia (de Magalhães and Björnberg 1957). Although the monophyly of the subgroups has been confirmed by different phylogenetic methods, the relationships among and within them are not. Hypotheses for their evolutionary relationships have been proposed using different methods and different morphological characters (de Magalhães and Björnberg 1957; Throckmorton 1962; Throckmorton and de Magalhães 1962; O'Grady et al. 1998; Yassin 2009; Souza et al. 2014; Roman et al. 2022), chromosome polymorphism (de Campos Bicudo 1973a), reproductive isolation (de Campos Bicudo 1973b; de Campos Bicudo and Prioli 1978; de Campos Bicudo 1979), protein polymorphism (Nascimento and de Campos Bicudo 2002), and gene sequences (Pélandakis and Solignac 1993; O'Grady et al. 1998; Rodríguez-Trelles et al. 1999; de Castro and Carareto 2004; de Setta et al. 2007; Roman et al. 2022). The evolutionary relationships proposed are summarized in supplementary table S1, Supplementary Material online.

Unlike other species groups in the subgenus *Sophophora*, such as the *melanogaster*, *obscura* and *willistoni* groups, genomic resources and genetic investigations in the *saltans* species group are scarce. Indeed, only four genomes have been sequenced and assembled to date (Kim et al. 2021). To bridge this gap and to test for the extent of phylogenetic conflicts, we sequenced and assembled genomes for 15 species with representatives from the five subgroups. Phylogenetic analyses using well-conserved genes resolved the evolutionary relationships among the subgroups but also highlighted conflicts

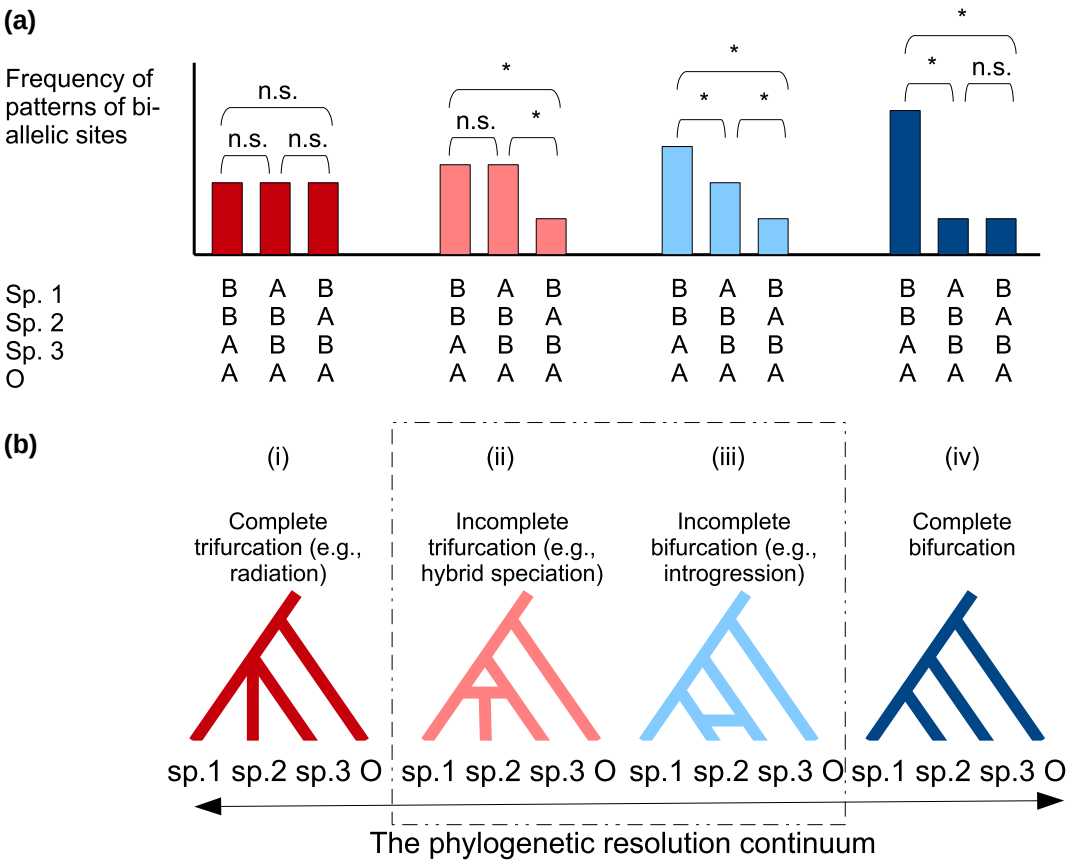

Fig. 1. The distribution of bi-allelic sites proportions among three ingroup species with an outgroup and the rationale of the 2A2B test. a) Statistical pairwise comparisons of the three possible bi-allelic patterns generate four categories. b) Each category corresponds to a possible evolutionary scenario along a phylogenetic resolution continuum: (i) complete trifurcation, (ii) incomplete trifurcation, (iii) incomplete bifurcation, and (iv) complete bifurcation. Categories ii and iii correspond to cases of reticulate evolution. The 2A2B test estimates significant deviation from parity in pairwise comparisons of the topologies. *: significant deviation, n.s. = non-significant.

between X-linked, autosomal and mitochondrial loci. To test how each of the four incongruence categories prevails across the species, we devised a new χ^2-based test that we called 2A2B. The test uses pairwise comparisons of the three topologies proportions shown in Fig. 1. We applied this test to sets of conserved orthologous genes and to long syntenic blocks with conserved collinearity across the Neotropical *Sophophora*, as well as to 100-kb-long windows from pseudo-reference genomes aligned to either an ingroup (*Drosophila sturtevanti*) or outgroup (*Drosophila willistoni*) species. We found reticulation levels to differ among the datasets and the subgroups, and to coincide with major historical biogeographical events in the Neotropics.

Results

Short-read Assembly of 17 Genomes Recovered 90% of BUSCO Genes

We sequenced using short-read Illumina approach 17 whole genomes from 15 species collected across various locations in the Neotropical region. Genome size, estimated from 21-k-mer frequency spectrum using GenomeScope 2 (Ranallo-Benavidez et al. 2020), ranged from 154.0 to 356.8 Mb. Our de novo assemblies using MaSuRCA (Zimin et al. 2013) resulted in genome lengths ranging from 177.5 to 287.7 Mb, with N50 values ranging from 2 to 92 Kb (supplementary table S2, Supplementary Material online). To evaluate the completeness of our assembled genomes, we searched for single-copy genes (SCGs) using BUSCO (Simão et al. 2015). We found that over 90% of the searched genes were complete for all of the genomes (supplementary table S2, Supplementary Material online). Kim et al. (2021) assembled using both short Illumina and long Nanopore reads the genomes of four *saltans* group species, all of which we have independently sequenced. Whereas their assemblies' contigs were much longer, with N50 ranging from 2 to 6 Mb, the BUSCO score for the same set of species did not largely differ (98% vs. 95–96% in our study; supplementary table S2, Supplementary Material online). However, Kim et al. (2021) aimed at generating high quality reference which is not the purpose of this study.

Muller Elements Analysis Resolves Relationships Between the Subgroups and Unravels a Minor X-autosomal Conflict in the *sturtevanti* Subgroup

Phylogenomic analyses were performed using 2,159 SCG shared across all species. Gene trees, inferred for each

SCG using maximum-likelihood (ML) in IQ-TREE (Minh et al. 2020) produced 1,263 distinct topologies, with 206 of them found more than once. To test if SCG chromosomal position may underlie the discrepancies in gene trees, we localized each SCG to its corresponding Muller element according to the position of its *Drosophila melanogaster* ortholog identified by Blast (Camacho et al. 2009). As a result, we generated five independent datasets, each corresponding to the Muller elements A, B, C, D, and E, comprising 337, 370, 425, 419, and 568 SCGs, respectively. These datasets were then used to reconstruct the species trees using the multi-species coalescent model, and the genes within them were concatenated for Bayesian and ML phylogenetic inferences.

The trees generated by the five data sets showed very similar topologies with well supported nodes either in the Bayesian Inference implemented in BEAST (Bouckaert et al. 2019), the ML implemented in IQ-TREE, or for the multi-species coalescent model analysis implemented in ASTRAL-III (Zhang et al. 2018) (supplementary figs. S1 and S2 and S3, Supplementary Material online). The *parasaltans* subgroup was placed as sister to all other subgroups, followed by the emergence of the *sturtevanti* subgroup. The *cordata* and *elliptica* subgroups showed a close relationship, and were sister to the *saltans* subgroup. The only discrepancy between the topologies was the placement of *Drosophila lehrmanae*, a newly discovered species in the *sturtevanti* subgroup (Madi-Ravazzi et al. 2021). For this species, while ML and multi-species coalescent analyses reported lack of branch support for multiple trees (supplementary figs. S2 and S3, Supplementary Material online), Bayesian inference recovered well supported branches and two topologies (Fig. 2a and supplementary fig. S1, Supplementary Material online). The two distinct topologies distinguished the Muller elements forming the X chromosome (elements A and D who are fused in a fusion shared by the Neotropical *Sophophora*, i.e. the *saltans* and *willistoni* groups (Sturtevant and Novitski 1941; Dobzhansky and Pavan 1943; Cavalcanti 1948), and the Muller elements representing autosomal chromosomes (elements B, C and E).

The published genome of *Drosophila prosaltans* (Kim et al. 2021) did not group with the genome of this species sequenced by us, instead it grouped with *D. saltans*. The genome previously published comes from a line collected in El Salvador in 1957. According to de Magalhães (1962) detailed morphological revision of multiple geographical specimens of the *saltans* group, the sampling site of this particular strain is outside the geographical range of *D. prosaltans*, but within the expected range of *D. saltans*. Furthermore, the *D. saltans* and *D. prosaltans* lines used in our study underwent thorough morphological analyzes (Souza et al. 2014; Roman and Madi-Ravazzi 2021), indicating that the lines we used were accurately identified. Therefore, it is likely that the previously sequenced *D. prosaltans* strain from El Salvador was misidentified and we consider it here to belong to *D. saltans*.

Mitogenomes Show Cytonuclear Conflicts in the *sturtevanti* and *saltans* subgroups

We assembled mitochondrial genomes for the 15 *saltans* group species using MitoZ (Meng et al. 2019). We did not use the previously assembled four strains since several mitochondrial scaffolds were likely removed in those assemblies (Kim et al. 2021). We conducted phylogenetic analysis on the aligned mitogenomes genes using both IQ-TREE and BEAST. Overall, the mitochondrial trees matched the topology of the nuclear gene trees regarding the inter-subgroup relationships. However, three major discrepancies were identified (Fig. 2b). First, the position of *D. lehrmanae* within the *sturtevanti* subgroup did not agree with either the X or autosomal SCG topologies, proposing a topology wherein *D. lehrmanae* is a sister species of *D. sturtevanti* (a topology recovered once in multi-species coalescent analysis (Muller element C, supplementary fig. S3, Supplementary Material online) and ML (Muller element B, supplementary fig. S2, Supplementary Material online)). Second, whereas the mitochondrial tree recovered the monophyletic relationship between the *elliptica*, *cordata* and *saltans* subgroups, the position of *Drosophila neocordata* (*cordata* subgroup) differed, being sister to the three species of the *elliptica* subgroup in the nuclear trees and to the six species of the *saltans* subgroup in the mitochondrial tree. Third, whereas nuclear trees recovered three lineages within the *saltans* subgroup, namely, *austrosaltans*, *nigrosaltans-pseudosaltans*, and *septentriosaltans-prosaltans-saltans*, only two lineages are revealed by the mitochondrial tree. Each of the mitochondrial clades involved one species from otherwise sister species in the nuclear trees, i.e. *Drosophila nigrosaltans* and *D. saltans* in one clade and their respective closely-related species *Drosophila pseudosaltans* and *D. prosaltans* in the other clade, suggesting that multiple cytoplasmic introgression events might have occurred in this subgroup (Fig. 2b). Because *D. saltans* and *D. prosaltans* are reported as sister species in the nuclear trees and are separated in the two mitochondrial clades, the two mitoclades were called *S* and *P*, respectively. Remarkably, the branching order in the *P* mitoclade is congruent with the nuclear tree, unlike the order in the *S* mitoclade. The *S* mitoclade may therefore be of reticulate origin, probably reflecting an older introgression from *Drosophila austrosaltans* to *D. nigrosaltans* and a more recent introgression between *D. austrosaltans* and *D. saltans*. However, due to the particularities of mtDNA compared to nuclear genomes (see Discussion below), the incomplete lineage sorting of two major haplotypes in the ancestor of the *saltans* subgroup may have resulted in the evolution of the two mitoclades.

The Extent of Reticulation Differs Between Assembly- and Pseudo-reference-based Approaches

To quantify the extent of introgression at a genome-wide scale, we analyzed allele distribution of bi-allelic phylogenetically informative sites in 28 species quartets using

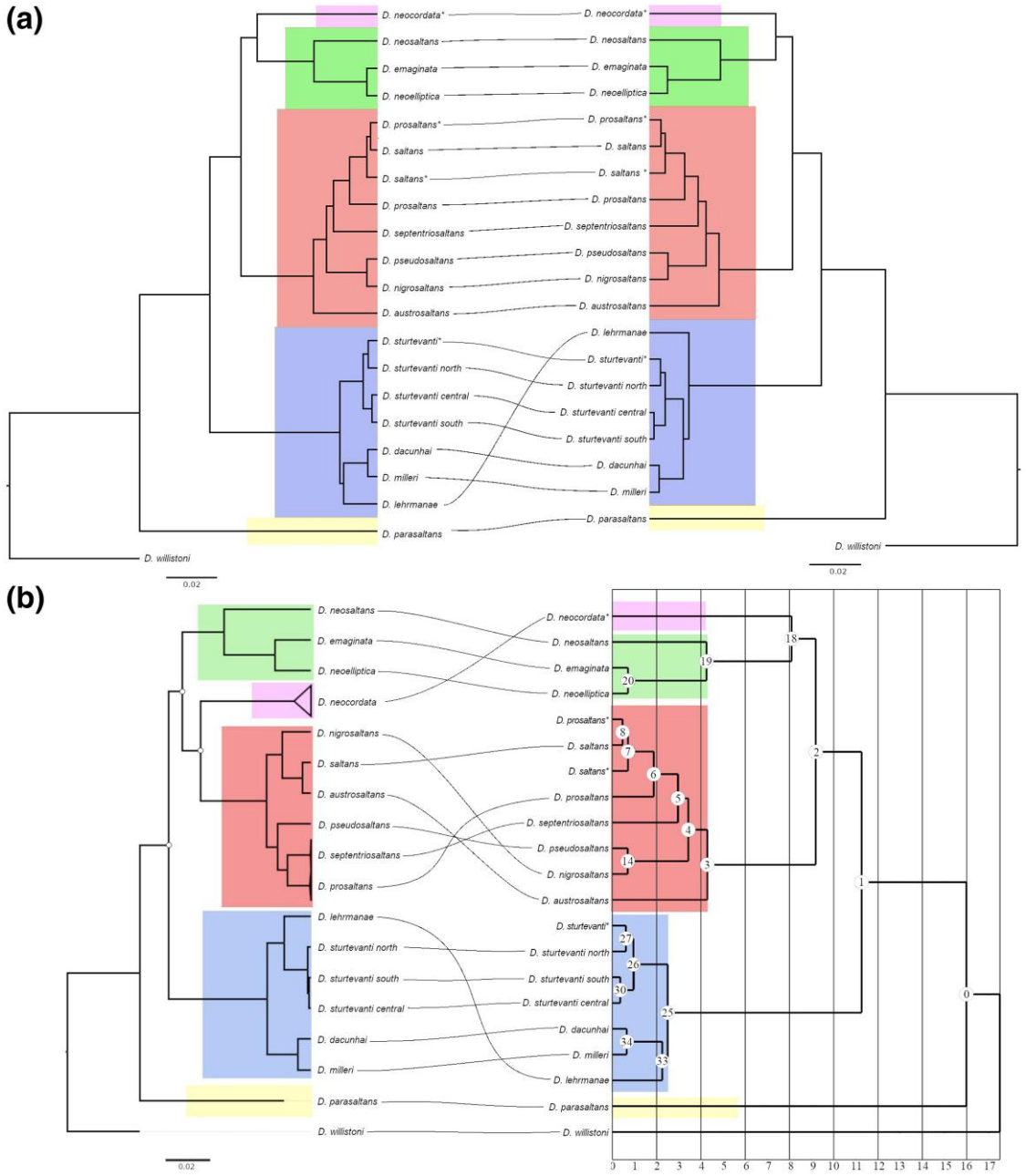

Fig. 2. Phylogenomic conflict between the X chromosome, autosomes, and the mitochondria. a) Comparison of autosomal topology (left, represented by Muller element b) and X chromosome topology (right, represented by Muller element a) demonstrates overall agreement with minor incongruence. b) Mitochondrial–nuclear disagreement highlights stronger incongruence between mitochondrial topology (left) and X chromosome topology (right, same as above but after midpoint rooting). All trees were inferred through Bayesian analysis. Divergence time estimation (in million years ago, myr) for the X chromosome topology is provided (left). Nodes are numbered for subsequent analyses. Maximum-likelihood and ASTRAL-III trees are provided in supplementary figs. S2 and S3, Supplementary Material online, respectively. Posterior probabilities (PP) of all nodes were equal to 1, except for nodes indicated with a blank circle where PP was equal to 0.87.

Patterson's *D* estimate of the standard ABBA-BABA test. We used four distinct locus-defining strategies to explore how strategy choice can influence estimates of *D*. Two strategies relied on our de novo genome assemblies. The first strategy included the use of SCGs defined by BUSCO (2,159 loci). The second strategy identified syntenic blocks equal to or greater than 50 kb-long with collinearity across the 15 *saltans* assemblies and *D. willistoni*, i.e. across Neotropical *Sophophora* (1,797 loci). The third and fourth strategies used a pseudo-reference approach, wherein the reads of each *saltans* group species were mapped to an ingroup (*D. sturtevanti*, 2,328 loci) or an outgroup species (*D. willistoni*, 1,863 loci) and a pseudo-reference was inferred by replacing variant sites in the reference genome by the ones in the reads (as in Mai et al. 2020). In these pseudo-reference strategies, a

locus was defined as a 100 kb-long window of the reference genome. To avoid the influence of linked loci from highly-variable regions on genome-wide estimates, we randomly sampled 20 sites from each locus with ≥20 evaluated sites, and repeated resampling for 100 reiterations for each locus-defining strategy.

For a bi-allelic site, the standard ABBA-BABA (Patterson's *D*) test considers that a true species tree exists, with site distribution BBAA, and then evaluate the deviation from parity of the two discordant configurations ABBA and BABA. We found that genome-wide absolute Patterson's *D* estimates, i.e. $(\sum ABBA - \sum BABA)/(\sum ABBA + \sum BABA)$, greatly differed among the four strategies (Fig. 3a). Synteny-based strategy had the lowest $|D|$ across the 28 quartets ($|D| = 0.025 \pm 0.004$) compared to BUSCO ($|D| = 0.066 \pm 0.014$, Student's t $P < 0.01$), *D. sturtevanti* pseudo-reference ($|D| = 0.176 \pm 0.025$, Student's t $P < 4.77 \times 10^{-6}$), and *D. willistoni* pseudo-reference ($|D| = 0.329 \pm 0.053$, Student's t $P < 5.97 \times 10^{-6}$). So, of the four strategies, the use of syntenic blocks is the most conservative.

To gain further insights on the reasons of this discrepancy between the four locus-defining strategies, we separately investigated assembly- and pseudo-reference-based approaches. The BUSCO approach mainly differs from the synteny-based approach in having, on average, shorter sequence since many loci are ≤50-kb long. Short sequences usually show faster lineage-specific evolutionary rates (Lipman et al. 2002), which can increase variance and also induce biases on Patterson's *D* due to the violation of the molecular clock assumption (Frankel and Ané 2023). Besides, the BUSCO analysis involves an additional step, which is the annotation of the conserved protein-coding genes (Manni et al. 2021). To account for these possible sources of errors, we counted for each lineage the number of singletons distinguishing the outgroup from the three ingroup species (i.e. BBBA sites), as well as the singletons distinguishing between the two most closely-related species in the quartet as defined from the phylogenetic analyses above (i.e. BAAA and ABAA). we called this test 3A1B, and it is reminiscent to the classical Tajima's (1993) Relative Rate Test. Under the assumption of a constant molecular clock or the absence of misannotation of a BUSCO gene in one of the two sister species both BBBA − BAAA and BBBA—ABAA should be > 0. We focused on the *elliptica* subgroup with *D. saltans* as an outgroup (i.e. (((*emarginata*,*neoelliptica*), *neosaltans*), *saltans*)) quartet, since this quartet showed a strong discrepancy between the two assembly-based approaches, with genome-wide *D* being 0.25 for BUSCO and −0.08 for synteny-based strategy. By applying the 3A1B test, we found that 37 BUSCO loci violated its assumption, all in the negative direction for *Drosophila emarginata*, i.e. BAAA > BBBA. By excluding these loci, Patterson's *D* dropped from 0.25 to 0.04. By contrast, only five loci violated the assumption in the synteny-based strategy. Their exclusion did not change Patterson's *D* estimate which remained −0.08. We then applied the 3A1B test for all locus-defining approaches, excluding loci that violated its basic assumption in each quartet. The genome-wide absolute Patterson's *D* estimates dropped from 0.066 ± 0.014 to 0.030 ± 0.004 for BUSCO genes (Student's t $P < 0.05$) and from 0.025 ± 0.004 to only 0.023 ± 0.004 for synteny-based loci (Student's t $P = 0.76$). After correction, genome-wide estimates from both BUSCO and synteny loci did not significantly differ (Student's t $P = 0.28$), even if synteny estimates remained slightly lower (Fig. 3b).

For pseudo-reference approaches, excluding rapidly evolving loci through the 3A1B test may not be enough, since quality and depth of mapping as well as reference genome biases constitute additional sources of errors. We therefore reconducted the analyses using more stringent parameters, only retaining loci with a mapping quality >20 and a depth >10 and <100 and excluding all within-species polymorphic sites. These parameters dramatically reduced the number of evaluated loci to only 17% and 4% in the ingroup (*D. sturtevanti*) and outgroup (*D. willistoni*) reference genome approaches. Estimation of genome-wide absolute Patterson's *D* dropped to 0.101 ± 0.023 (Student's t $P < 0.05$) and 0.102 ± 0.017 (Student's t $P < 2.9 \times 10^{-4}$) for the ingroup and outgroup references, respectively. This means that whereas ingroup and outgroup approaches were greatly different before the application of the stringent criteria (Student's t $P < 0.05$), absolute Patterson's *D* did not significantly differ between the two approaches after corrections (Student's t $P = 0.95$). For the abovementioned *elliptica* subgroup example, whereas before the stringent criteria Patterson's *D* were 0.02 and −0.71 for ingroup and outgroup approaches, these values dropped to −0.005 and −0.27 in the two approaches, respectively. However, correlation across quartets was strong (Pearson's $r = 0.56$, $P = 0.002$), indicating that distinct criteria may be applied to improve estimates depending on the distance from the reference genome species, with *D. willistoni* being more distant and hence prone to more variance due to the lower number of conserved loci that can be retained.

The 2A2B Test Unravels Lower Signal for Introgression When Syntenic Blocks are Used

To consider cases where a true species tree cannot be inferred, we designed a test that we call 2A2B. The test is an extension of the standard ABBA-BABA test in that it does not only test the significance of deviation from parity of the ABBA and BABA patterns, but also test the deviation from parity of each configuration to the BBAA pattern. Therefore, the test allows classifying each species triplet with an outgroup into one of the four categories along the trifurcation–bifurcation continuum given in Fig. 1. We applied the test at both the genome-wide level after corrections using the 3A1B test and mapping criteria (see above).

We found that for all locus-defining strategies, bifurcation categories (iii and iv) predominated (Fig. 4a). Only five

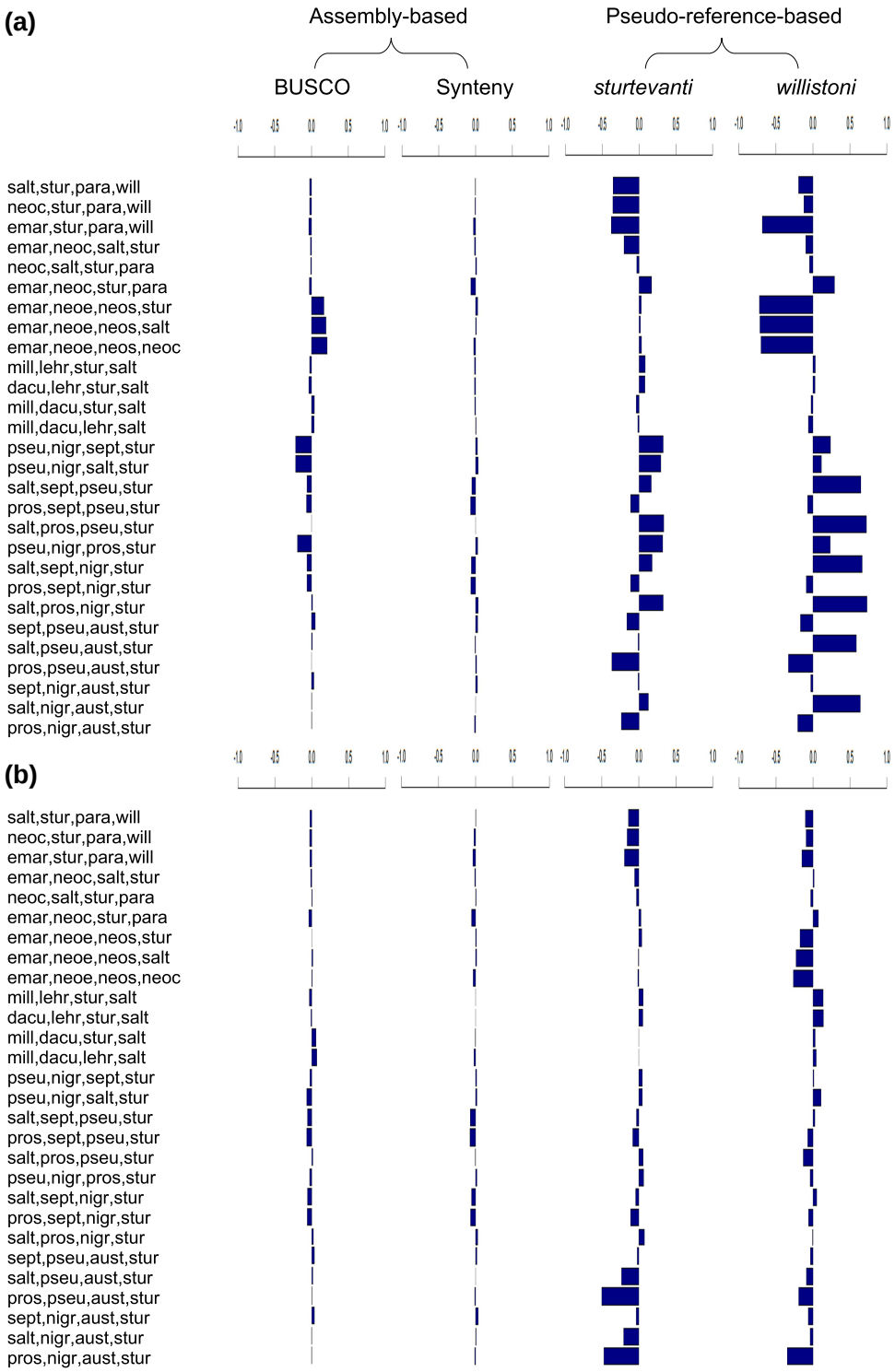

Fig. 3. Extent of phylogenomic discordance using four locus-definition strategies (see text) as measured by Patterson's *D* on a concatenated sequence of 20 randomly selected bi-allelic sites per locus, across 28 species quartets a) before and b) after correction for molecular clock (assembly-based approaches) and mapping quality (pseudo-reference approaches).

exceptions were observed. One incidence of category ii (incomplete trifurcation) was observed in the BUSCO strategy concerning the position of *D. lhermanae* relative to both *D. sturtevanti* and *Drosophila dacunhai*. Indeed, the position of *D. lhermanae* was the only conflicting result in the phylogenetic analyses based on BUSCO genes and it also showed a conflict with mitochondrial analysis (Fig. 2). Four incidences of category i (complete trifurcation) were observed in outgroup pseudo-reference strategy. Of which, two concerned the position of *D. lehrmanae* to both *D. sturtevanti* on the one hand and to each of the two sister species *D. dacunhai* and

Fig. 4. The 2A2B test using four locus-definition strategies (see text) across 28 species quartets as measured on a concatenated sequence of 20 randomly selected bi-allelic sites per locus after correction for molecular clock (assembly-based approaches) and mapping quality (pseudo-reference approaches). Colors of bars correspond to the four categories presented in Fig. 1.

Drosophila milleri on the other hand. The other two complete trifurcation cases concerned the relationships between the three lineages of the *saltans* subgroups, i.e. *austrosaltans*, *nigrosaltans-pseudosaltans*, *septentriosaltans-prosaltans-saltans*, in cases where species of the different mitoclades were analyzed together, i.e. *D. pseudosaltans* and *D. saltans* or *D. nigrosaltans* and *Drosophila septentriosaltans* (Fig. 2).

Category iv (complete bifurcation), which is the most consistent with a true species tree, was the most frequent in synteny-based strategy (Fig. 4). This was followed by both the BUSCO and the outgroup pseudo-reference strategies but the difference in category iv frequency was not significant (Fisher's exact test of category iv vs. all other categories). However, with only four exceptions, ingroup pseudo-reference approach identified category iii (incomplete bifurcation), which mostly indicates cases of introgression, in all analyzed quartets (Fisher's exact test between synteny-based and ingroup pseudo-reference strategy $P < 1 \times 10^{-4}$). The total number of bi-allelic sites (20 per locus) that were used in the genome-wide analysis differed among the four strategies, averaging 22,893 (BUSCO) and 21,306 (synteny) in assembly-based approaches, and 35,747 and 5,308 sites for the ingroup *D. sturtevanti* and the outgroup *D. willistoni* pseudo-reference approaches, respectively. The low number for the *D. willistoni* pseudo-reference is expected since increasing mapping quality criteria on the genome of this distant species would only favor strongly conserved loci. For the *D. sturtevanti* pseudo-reference, applying the same mapping quality criteria has likely maintained less conserved loci. Although we have excluded heterozygous sites from every species to reduce mapping biases of the reference allele, such biases may even extend to homozygous sites strongly linked to the excluded heterozygous site in the same read. Universal mapping criteria that can reduce mapping biases independent from the reference genome do not exist and mapping biases remain difficult to detect

(Lin et al. 2024). However, these biases are the most likely explanation of the extensive signal of introgression in the ingroup *D. sturtevanti* pseudo-reference. To test the effect of mapping quality on introgression estimates, we applied the 2A2B test at the locus level before and after applying mapping quality thresholds. In agreement with our assumptions and the findings on Patterson's *D* (Fig. 3), the proportion of loci supporting categories ii and iii, i.e. categories that are the most associated with reticulate evolution, did not exceed ~7% in assembly-based strategies, but they dropped after application of mapping quality criteria from 46% to 11% and from 34% to 4% for *D. sturtevanti* and *D. willistoni* pseudo-reference genomes strategies, respectively (supplementary fig. S4, Supplementary Material online). Ingroup pseudo-reference approaches, while providing more data, may not reflect the true phylogenetic relationships between the species, hence inflating the estimation of the extent of introgression.

Historical Biogeographical Events Coincided With Episodes of low Phylogenetic Resolution

Phylogenetic analyses and assembly-based 2A2B test identified three parts on the species tree with evidence for incomplete bifurcation (i.e. category iii): at the inter-subgroup level in any combination that involved representatives from at least two subgroups of the *cordata*, *elliptica* and *saltans* subgroups, at the base of the *sturtevanti* subgroup, and in the *saltans* subgroup in quartets involving species belonging to distinct mitoclades. Whereas category iii is usually interpreted as evidence of introgression following secondary contact, other biological processes such as ancestral population structure at the time of phyletic radiation or asymmetry of evolutionary rates due to responses to distinct adaptive pressures (Hibbins and Hahn 2022; Frankel and Ané 2023) can generate similar signals. Because all these processes can associate with major geographical changes affecting species distribution, ecology and frequency of secondary contacts, a knowledge of the historical biogeography of the clade under study is crucial for the understanding of the processes that have likely led to phylogenetic conflicts.

Like phenotypic traits, ancestral geographical distribution can be inferred on phylogenetic trees from present day geographical data (Quintero et al. 2015). Therefore, we mapped current distribution of the studied species on the Bayesian X tree that was dated in reference to recent family-wide estimates (Suvorov, Kim, et al. 2022) (see Methods). The historical biogeography supported an origin of the *saltans* group around 16 million years (myr) ago in the north-west of South America, which currently corresponds to the Amazonian region (Fig. 5). At that time, this region was dominated by a vast wetland environment known as the "Pebas System." The first radiation episode, i.e. the split between the ancestors of the *cordata*, *elliptica*, and *saltans* subgroups occurred circa 10 myr ago at this region, coinciding with the formation of the Amazon River and its huge fluvial networks that led to the transformation of the Pebas wetland into the dense and diverse Amazonian forests (Hoorn et al. 2022). The second episode occurred nearly 3 myr ago, which correlates with the geological formation of the isthmus of Panama (O'Dea et al. 2016), and the time of a connection between the Western Amazonian and the Southern Atlantic rainforests during the Pliocene (Batalha-Filho et al. 2013; Ledo and Colli 2017). This episode involved the ancestors of the *nigrosaltans-pseudosaltans* and *septentriosaltans-prosaltans-saltans* clades. Indeed, the first clade has a northern to central representative (*D. nigrosaltans*) and a southern species (*D. pseudosaltans*), whereas the second clade has a northern representative (*D. saltans*) and two central and central to southern species (*D. septentriosaltans* and *D. prosaltans*, respectively). The third radiation episode involves the ancestor of the *sturtevanti* subgroup and also occurred in the late Pliocene, ca. 2.5 myr ago, at the likely time of a connection between the Western Amazonian and the Southern Atlantic rainforests.

Discussion

Toward a Comprehensive Phylogeny of the *saltans* Species Group

A large number of *Drosophila* genomes have been sequenced and used in phylogenetic analyses (Khallaf et al. 2021; Kim et al. 2021; Li et al. 2022; Suvorov, Kim, et al. 2022), but studies with comprehensive sampling of nearly all species in a group remain relatively uncommon (Mai et al. 2020; Conner et al. 2021; Yusuf et al. 2022; Moreyra et al. 2023). Despite minor inconsistencies, our phylogenomic analysis of 15 species of the *Drosophila saltans* species group produced a consistent picture of the relationships between the five subgroups of this clade. All X, autosomal and mitochondrial phylogenies showed the *parasaltans* subgroup as the first to diverge, followed by the *sturtevanti* subgroup, and later by a clade comprising the *cordata*, *elliptica* and *saltans* subgroups, in which the position of the *cordata* subgroup differed between nuclear and mitochondrial trees. This general picture has not been previously proposed despite the tremendous number of phylogenetic investigations of this group (de Magalhães 1962; Throckmorton 1962; Throckmorton and de Magalhães 1962; O'Grady et al. 1998; Rodríguez-Trelles et al. 1999; de Castro and Carareto 2004; de Setta et al. 2007; Yassin 2009; Souza et al. 2014; Roman et al. 2022; see supplementary table S1, Supplementary Material online).

While our analysis has shed light on the intricate evolutionary dynamics within this clade, further sampling holds the potential to provide a more comprehensive understanding into this complex evolutionary history. For example, the inclusion of *Drosophila subsaltans*, *Drosophila lusaltans*, *Drosophila cordata*, and *Drosophila rectangularis* through whole-genome sequencing promises to provide insight onto unresolved phylogenetic questions raised from previously published observations on

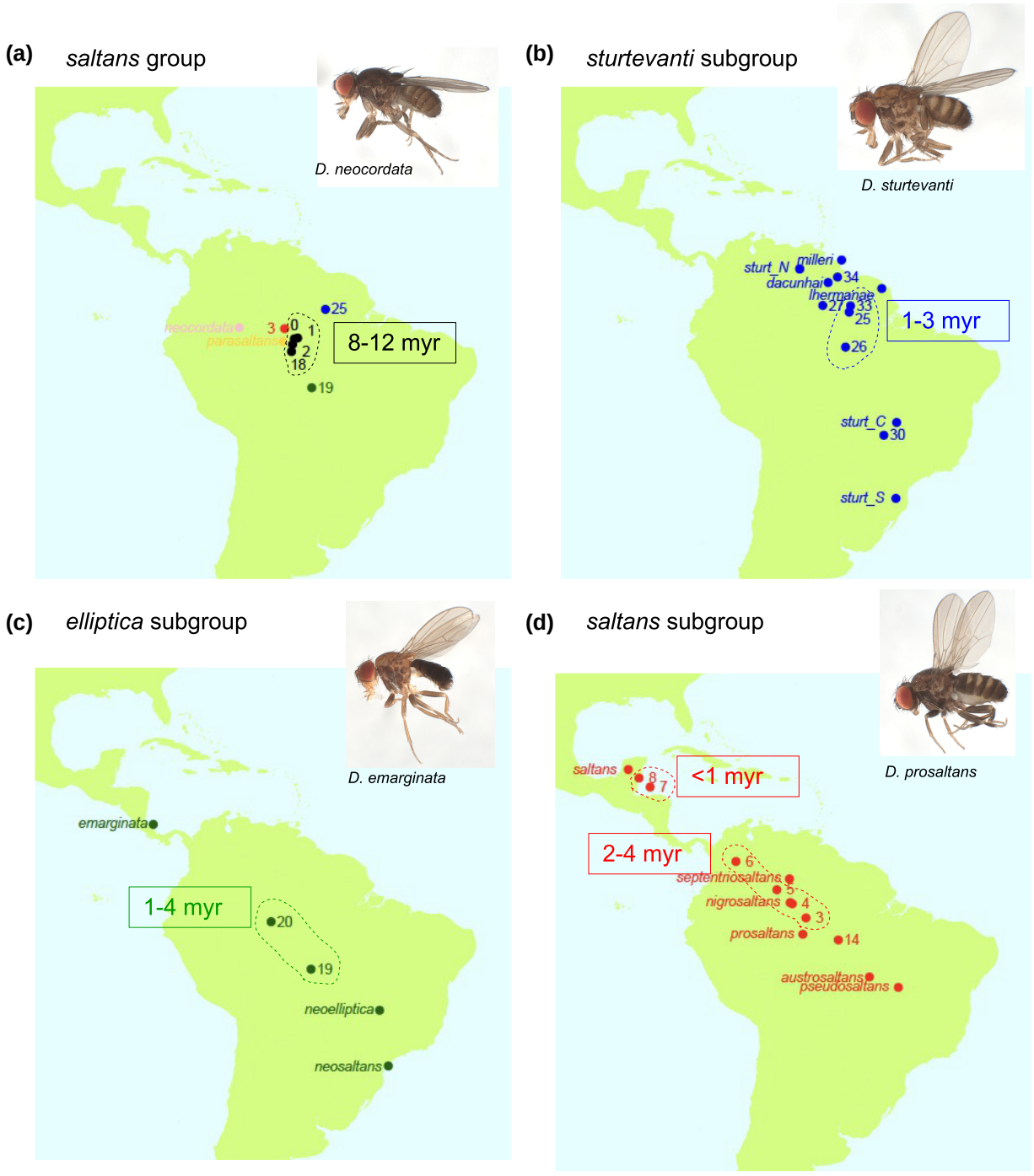

Fig. 5. Historical biogeography of the *saltans* species group showing correspondence between episodes of rapid radiation and paleogeographical events. Centroid geographical coordinates for each species were inferred and ancestral coordinates were estimated using BayesTraits on the Bayesian X chromosome chronogram given in Fig. 2b. Terminal and internal nodes coordinates are projected on the current map of South and Central America as dots with their labels and colors similar to Fig. 2b. Age ranges are given for internal nodes grouped by dotted contours showing the association between the split of a) the five subgroups and the formation of the Amazonian river around 10 myr ago, b) of the *sturtevanti* subgroup species and the connection between the Amazonian and Atlantic rainforests around 2.5 myr ago, and c) of the *elliptica* and d) the *saltans* subgroups species and the formation of the isthmus of Panama around 3 myr ago. Photos of adult males of four representative species are shown (©Amir Yassin).

reproductive isolation and morphology (de Magalhães 1962; de Campos Bicudo and Prioli 1978). These questions include the monophyly and positioning of the *parasaltans* and *cordata* subgroups. Additionally, the inclusion of the insular species *D. lusaltans* which presents low reproductive isolation (de Campos Bicudo 1973b), can bring new insights into the reticulate evolution in the *saltans* subgroup. The *saltans* subgroup showed the most dramatic signal of cyto-nuclear discordance and reticulate evolution. de Campos Bicudo (1973b) investigated

reproductive isolation among the seven then described species of this subgroup, and in a remarkably partial agreement with our nuclear phylogenomic trees, she concluded that *D. pseudosaltans*, *D. nigrosaltans* and *D. austrosaltans* showed strong reproductive isolation with *D. lusaltans*, *D. septentriosaltans*, *D. prosaltans* and *D. saltans*, i.e. the latter species show less reproductive isolation. Indeed, nearly all crosses among the last four species produce fertile females with some even producing fertile females and males (de Campos Bicudo 1973b). This porosity largely agrees with the high incidence of reticulate evolution we report here for this subgroup.

Two widespread species of the *saltans* subgroup, *D. saltans* and *D. prosaltans*, show a peculiar geographical disjunction, with the former species being restricted to Central America whereas *D. prosaltans* is widespread in South America south of Costa Rica. The discrimination between strains belonging to each species has long been erroneous (Dobzhansky 1944; Mayr and Dobzhansky 1945; Spassky 1957; de Magalhães 1962) and we showed here that their misidentification persists even in the genomic era (Kim et al. 2021; Suvorov, Kim, et al. 2022). Interestingly, de Campos Bicudo (1973b) provided evidence for reproductive reinforcement between these two sister species; sympatric populations in their junction zone in Costa Rica demonstrated stronger reproductive isolation than allopatric populations of both species. We have only included one to a few geographical lines from each species and a broader sampling to investigate the extent of their reproductive isolation and genome porosity is strongly needed.

Intra- and Inter-genomic Conflicts Impact the Inference of Phylogenetic Patterns

Concatenation helped recovering a sex chromosome versus autosome conflict, similar to the one described by Mai et al. (2020) for the *Drosophila nasuta* subgroup. Like these authors, this conflict was limited to a single part of the tree, i.e. the relationship of *Drosophila pulau* to *D. sulfurigaster sulfuricaster* and *D. s. bilimbata* in the *nasuta* subgroup and the placement of *D. lehrmanae* in the *sturtevanti* subgroup. The peculiarities of sex chromosomes such as their lower effective population size, different recombination and mutation rates, and greater exposition to natural selection when found in hemizygosity, lead to higher rates of adaptive evolution of their genes compared with autosomal genes (i.e. faster-X evolution) and also to the disproportional accumulation of genes related to reproductive isolation and Dobzhansky-Muller hybrid incompatibilities (i.e. Haldane's rule) (Charlesworth et al. 2018). Besides, low recombination rate underlies stronger linked selection effects against introgressed ancestry. Altogether, those characteristics are thought to be responsible for the resistance to hybridization in the sex chromosomes (Ellegren 2009; Qvarnström and Bailey 2009; Sankararaman et al. 2016; Charlesworth et al. 2018; Seixas et al. 2018; Mai et al. 2020; Matute et al. 2020; Moran et al. 2021; Reilly et al. 2022; Skov et al. 2023; but see David et al. 2022).

The second conflict regards a significant disagreement between mitochondrial (mtDNA) and nuclear data. Discordance between nuclear and mitochondrial genomes is a well-documented phenomenon in the tree of life as highlighted by Toews and Brelsford (2012). Several characteristics of mtDNA, such as being haploid and uniparentally inherited, resulting in a fourfold reduction in effective population size when compared with autosomal loci, affect its evolution and may lead to high incidences of incomplete lineage sorting (Wang et al. 2018). An alternative explanation can be cytoplasmic introgression, which has long been recognized in *Drosophila* (Solignac et al. 1986; Ballard 2000; Llopart et al. 2014). In a recent population study within the *willistoni* group, multiple mitochondrial introgressions were observed in *Drosophila paulistorum* populations (Baião et al. 2023). These included an ancient introgression with a highly divergent mitochondrial type, followed by more recent events. Rapid fixation of a mitochondrial type may be directly caused by a selective advantage provided by the mitochondrial type itself, or indirectly facilitated by a non-selective factor, such as *Wolbachia*, a bacterium known to modify the reproduction of its host (Baião et al. 2023). Although, interesting results have been reported from population approaches, conflicts between nuclear and mitochondrial genomes have not been addressed in recent phylogenomic analyses in the Drosophilidae (Mai et al. 2020; Khallaf et al. 2021; Suvorov, Kim, et al. 2022; Yusuf et al. 2022). The disagreement was particularly evident for the *saltans* subgroup, where it was most likely of recent origins, separating species that have diverged only 0.7 myr ago, i.e. *D. nigrosaltans* and *D. pseudosaltans*. Remarkably, the two mitoclades *P* and *S* do not correlate with the degree of reproductive isolation inferred by de Campos Bicudo (1973b), contrary to nuclear tree, indicating that mito-nuclear conflicts in the *saltans* subgroup did not contribute to the evolution of reproductive isolation in this clade.

Of the three subgroups for which multiple species were sequenced, the *saltans* subgroup had the highest incidence of incongruences inferred by the 2A2B test. In this subgroup, multiple large chromosomal inversions are known to be shared among closely-related species (Dobzhansky and Pavan 1943; Cavalcanti 1948; de Campos Bicudo 1973a; de Campos Bicudo and Prioli 1978) and evidence for balancing selection on ancestral inversion polymorphism has been demonstrated in a number of cases (de Campos Bicudo 1973a). For example, one autosomal inversion is polymorphic in *D. austrosaltans*, with one arrangement being homozygous in *D. septentriosaltans* and *prosaltans*, and the alternative arrangement homozygous in the remaining species including *D. saltans*. Such an inversion has likely been inherited from a polymorphic ancestor of the *saltans* subgroup, and was maintained polymorphic in the ancestor of the *septentriosaltans-prosaltans-saltans* clade before different

arrangements became fixed in each of these species, with one arrangement being already fixed in the ancestor of the *nigrosaltans-pseudosaltans* clade. Alternatively, the inversion might have originated in the ancestor of the *septentriosaltans-prosaltans-saltans* clade, then it was introgressed in *D. austrosaltans*, whereas the ancestral non-inverted configuration being introgressed on its own in *D. saltans*. However, this introgression scenario is less parsimonious, especially given the geographical isolation of *D. saltans* from the remaining species of the subgroup. Whether the high degree of incongruences in the *saltans* subgroup is associated with large ancestral inversions potentially absent in other bifurcating clades would require the future generation of chromosome-level assemblies for multiple *saltans* group species.

The 2A2B Test on Syntenic Blocks Helps Distinguishing "Soft" and "Hard" Introgressions

Our analyses reveal that the choice of locus-defining scheme can affect the degree of phylogenomic uncertainty. Indeed, most studies that have provided strong evidence for introgression using site-specific patterns comparisons have usually aligned reads from multiple species to a single reference assembly (cf. studies compiled by Dagilis et al. 2022). Whereas such an approach can increase the power (Martin et al. 2015; Pease and Hahn 2015), it also introduces biases due to paralogy, misalignments or absence of collinearity among species, that are now known to bias phylogenetic inference itself (Valiente-Mullor et al. 2021; Rick et al. 2023). Therefore, the most common practice now is to use sets of clade-specific single-copy orthologs, such as the BUSCO genes, in inferring phylogenies (Manni et al. 2021). Consequently, some recent introgression studies have focused on such genes. For example, in a study of 155 genomes covering a wide range of drosophilid lineages, Suvorov, Kim, et al. (2022) analyzed these conserved genes. While finding evidence for widespread introgression, they also showed that their discordance estimates were highly sensitive to the length of the analyzed genes as well as to the slightest relaxation of selective pressures. We attempted here to overcome this limitation by including larger syntenic blocks, which in addition to conserved protein-coding sequences should also include presumably neutral intronic and intergenic sites. Whereas, this approach showed the lowest $|D|$, indicating that patterns predicted by introgression do not exceed 7% of the genome, it led to only 1.5-fold increase in the number of analyzed sites, most likely because our assemblies were based on only short reads. With the continuing progress in whole-chromosome sequencing techniques and assembly-to-assembly alignment tools, a better understanding of the extent of probably true or "hard" introgression events from artificial or "soft" ones may be attained.

Phyletic radiations occur when isolation barriers simultaneously and rapidly evolve between multiple populations of a species with a broad geographical or ecological breadth. Consequently, genetic lineages will be incompletely sorted in the emerging species leading to a lack of phylogenetic signals. This situation usually occurs following a rapid environmental change or expansion into a new range with new empty niches. Indeed, our historical biogeography approach shows that such situation might have favored the episodic signals of low phylogenetic signals in the *saltans* group since the episodes coincided with either the transition from wetlands into Amazonia forests in the Late Miocene or the connection of the Amazonian forest with either Central American or the Southern Atlantic forests in the Pliocene. These events have been proposed to be responsible for a wide range of rapid radiations in Neotropical taxa, such as butterflies, amphibians and birds (Batalha-Filho et al. 2013; O'Dea et al. 2016; Ledo and Colli 2017; Hoorn et al. 2022). The dating and the analyses of the extent of phylogenetic discordance in the genomes of such clades are strongly needed to identify the impact of these events on Neotropical diversification.

Materials and Methods

Sample Collection, Whole-Genome Sequencing and Assembly

We performed whole genome pool sequencing on female flies from 15 different species from the *saltans* group, as well as three populations of *D. sturtevanti*. The specimens used for sequencing were obtained from one or multiple strains, and detailed information regarding the number of individuals and their collection locations can be found in supplementary table S2, Supplementary Material online. For all species, DNA extraction, PE library preparation using NEBNext Protocol, and 75× coverage double-paired Illumina Novoseq sequencing were conducted by End2End Genomics LLC, Davis, California. For all species except *D. neocordata*, we assembled the genome using the Maryland Super Read Cabog Assembler (MaSuRCA) (Zimin et al. 2013), which utilizes both the de Bruijn graph and overlap–layout–consensus (OLC) methods to generate super-reads. k-mer-based genome size was estimated using KMC3 (Kokot et al. 2017) and GenomeScope v.2.0 (Ranallo-Benavidez et al. 2020). To assess the assembly's completeness, we searched for single-copy-genes (SCG) using default parameters in Busco5 (Simão et al. 2015) with the diptera_odb10 database (Kuznetsov et al. 2023). In addition to the sequenced flies, we also assembled Illumina reads the genomes of *D. saltans*, *D. neocordata*, *D. prosaltans* and *D. sturtevanti* published by Kim et al. (2021) (assembly numbers ASM1890357v1, ASM1890361v1, ASM1815127v1 and ASM1815037v1, respectively), and we used the *D. neocordata* assembly published by Baião et al. (2023) in all analyses.

Phylogenomics: Nuclear Single-copy-genes

For phylogenomics analysis, SCG searches were carried out using 3,285 SCG from diptera_odb10 database inferred by Busco5 (Simão et al. 2015). SCG present in all species were

kept and aligned using the L-INS-i method implemented on MAFFT (Katoh and Standley 2013) (mafft –localpair –maxiterate 1000 –adjustdirection).

Genomic data of different species of *Drosophila* support that genes tend to be situated within the same Muller element across multiple species (see Schaeffer 2018). Taking this gene linkage into account, we reconstructed five independent datasets (Muller elements A-E), each comprising all SCG found in the respective Muller element. To achieve this, we performed a tBlastn search against the *D. melanogaster* reference genome v.6, and subsequently we concatenated genes found within the same Muller element. Phylogenetic trees were then constructed using ML and Bayesian methods implemented in the softwares IQ-TREE (Nguyen et al. 2015) and BEAST (Bouckaert et al. 2019), respectively. A GTR + G + I substitution model was used for all analyses. An MCMC chain of 10,000,000 generations under strict clock and a Yule's model was run in BEAST followed by a 25% burn-in sampling. Additionally, ML trees were generated for each gene, and for each Muller element, trees of corresponding genes used to reconstruct a species tree, using multi-species coalesce model implemented in ASTRAL-III (Zhang et al. 2018). Branch support was assigned using 1,000 ultrafast bootstrap in IQ-TREE (Hoang et al. 2018), posterior probability in BEAST and local posterior probability in ASTRAL-III.

Phylogenomics: Mitochondrial Genome

Mitochondrial genomes were assembled and annotated with MitoZ (Meng et al. 2019) using the Megahit assembler (Li et al. 2015). In order to ensure the exclusion of nuclear-embedded mitochondrial DNA sequences within the assembly, a strategic approach was taken. Considering that mitochondrial reads are found in higher frequency than nuclear-mitochondrial DNA sequences, the read subsampling were set to 0.5 gigabases (–data_size_for_mt_assembly 0.5). The genes obtained from the mitochondrial genome were aligned using the MAFFT alignment tool with the –auto parameter due to the close similarity between sequences. Subsequently, the aligned genes were concatenated into a dataset for phylogenetic analysis. The concatenated dataset served as the basis for reconstructing phylogenetic trees using both ML implemented in IQ-TREE and Bayesian Inference (BI) in BEAST. A GTR + G + I + F substitution model, inferred by IQ-TREE, was used for all analyses. An MCMC chain of 10,000,000 generations under relaxed clock and a Yule's model was run in BEAST followed by a 25% burn-in sampling. Branch support was assigned using 1,000 ultrafast bootstrap in IQ-TREE and posterior probability in BEAST.

Identification of Syntenic Blocks

The 20 genomes of the *saltans* group, the 15 sequenced here and the five published by Kim et al. (2021) and Baião et al. (2023) were preliminary annotated with Miniprot (miniprot -lut16, Li 2023). The primary objective of this annotation was to accurately map proteins from the robust and reliable genome annotation of *D. willistoni* (Clark et al. 2007). After protein mapping, the predicted gene loci were assessed to identify syntenic blocks present in the Neotropical *Sophophora* (comprising *D. willistoni* and *saltans* group). The identification of these blocks was based on gene order and orientation, achieved using an in-house Perl script. First, this script compares the scaffolds' genes order and orientation between the reference assemblies of *D. saltans* and *D. sturtevanti*, the synteny blocks were defined when all the genes were found in the same order and orientation in both species. The identified collinear blocks were then searched for the remaining genomes including the *D. willistoni* genome. Blocks with missing data, i.e. missing genes for one or more species were subsequently removed, and the remaining blocks were subjected to a size-based filtering with a threshold of ≥50 kb. The selected blocks were aligned using the Mafft software (Katoh and Standley 2013).

Pseudo-reference Inference

Illumina reads were aligned on either the *D. willistoni* reference genome (Clark et al. 2007: 12) or the recently published *D. sturtevanti* assembly (Kim et al. 2021) using minimap2 (Li 2018). Minimap2 generated BAM files were assembled into a pseudo-reference for each species using the BCFtools software (Danecek et al. 2021). We did not first apply any mapping quality criteria. For subsequent analyses, we only retained homozygous sites with mapping quality >20 and a depth >10 and <100. Each scaffold was then split into 100 kb-long windows.

Quantifying Reticulation Using the 2A2B Test

Combinations of 28 species quartets were evaluated to test reticulation, bi-allelic sited shared by pairs were searched in every locus, whether using BUSCO, syntenic blocks or pseudo-reference 100 kb-long windows. Loci with at least 20 informative sites were retained. To reduce the influence of loci with higher number of evaluated sites, we concatenated randomly drawn 20 sites from each locus. We then estimated Patterson's *D*, i.e. $(\sum ABBA - \sum BABA)/(\sum ABBA + \sum BABA)$, as an average of 100 iterations of this random sampling for each quartet. To account for cases where a true species may not be present, we devised a new test that we called 2A2B. For each locus or to the concatenated sequences constructed as above, the occurrences BBAA, ABBA or BABA patterns were counted, and three pairwise χ^2-based tests were conducted, denoted *D1*, *D2*, and *D3* as follows (D1 is similar to Patterson's D):

$$D1 = \left(\sum ABBA - \sum BABA\right)^2 / \left(\sum ABBA + \sum BABA\right),$$

$$D2 = \left(\sum BBAA - \sum ABBA\right)^2 / \left(\sum BBAA + \sum ABBA\right),$$

and

$$D3 = \left(\sum BBAA - \sum BABA\right)^2 / \left(\sum BBAA + \sum BABA\right).$$

Each of these estimates is reminiscent to the classical Tajima's (1993) Relative Rate Test (RTT), which tests the molecular clock hypothesis as difference in branch lengths from two ingroup taxa to a third outgroup taxon, with lengths estimated as counts of different sites. However, unlike RRT, branch lengths in our analyses are counted in terms of shared differences, i.e. singletons are not considered. Similar to RRT, we applied a chi-squared analysis to test for deviation from parity for each parameter. According to the significance levels of the three estimates, a locus can be classified under each of the four categories presented in Fig. 1. The test is run by the 2A2B perl script associated with this manuscript. To test molecular clock in our analyses, which can bias incongruence inference (Frankel and Ané 2023), we also applied a quartet-adapted version of the RRT that we called 3A1B test, wherein sites with singletons (BBBA, ABAA, and BAAA) were also counted and loci violating the assumption of BBBA—ABAA and BBBA—BAAA > 0 were excluded from the analyses. For pseudo-reference approaches, only sites that were mapped on the reference genomes were evaluated, i.e. any missing site of any species was not evaluated in the quartet including that species.

Historical Biogeography

To determine the sampling locations of the evaluated species, we parsed coordinates of current distribution of the studied species obtained from TaxoDros (https://www.taxodros.uzh.ch/search/class.php). Additionally, we incorporated sampling sites that we ourselves had conducted. The accuracy of our species identification was previously confirmed through DNA barcoding. After inspection of the geographical points and manual correction, we identified the most northern, southern, western, and eastern points for each species. We used those points to reconstruct the ancestral geographical range by conducting a BayesTraits (Meade and Pagel 2022) analysis for each cartesian point separately. The analyses were carried out using the GEO model with the phylogenetic tree generated reconstructed with the Muller element A (XL chromosome arm) after 1,000,000 MCMC iterations and 25% burn-in. The divergence times were estimated using this tree under Bayesian inference. The calibration point used was the split between *D. willistoni* and *D. sturtevanti* at 17.5 myr, as estimated by Suvorov, Kim, et al. (2022). To visualize the biogeographical history, we plotted the inferred centroid coordinates, i.e. averaged latitudes and longitudes of the four extreme points, for each terminal and internal node on the map of Central and South Americas using the maps v.3.4.2.1 software package (Deckmyn 2024) in R (R Core Team 2023).

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

This paper is dedicated to Prof. Hermione E. M. C. Bicudo for her significant contributions to the study of the evolutionary genetics of species of the *Drosophila saltans* species group. We would like to thank David Ogereau for his assistance and insights in genome assembly for this study.

Funding

We thank Fundação de Amparo à Pesquisa do Estado de São Paulo for the financial support to L.M.R. (Number processes: FAPESP 95/06165-1, 2014/14059-0 and 2016/11994-5) enabling us to collect and sequence many strains used here. We extend our thanks to the Conselho Nacional de Desenvolvimento Científico e Tecnológico for their support, specifically grant number CNPq 141545/2020-8, as well as the France Excellence Eiffel scholarship program for funding C.P.'s PhD scholarship. *Illumina* sequencing of *D. neocordata* was performed at the SNP&SEQ Technology Platform in Uppsala, Sweden, which is part of the Swedish National Genomics Infrastructure and Science for Life Laboratory. SNP&SEQ is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. This project was partly supported by a grant from The Swedish research council VR (2014-4353) to L.K. We want to thank the Nouragues research field station (managed by CNRS), which benefits from "Investissement d'Avenir" grants managed by Agence Nationale de la Recherche (AnaEE France ANR-11-INBS-0001; Labex CEBA ANR-10-LABX-25-01). The project was partly funded by the Austrian Science Fund FWF grant P28255-B22 to W.J.M. and by the Richard Lounsbery Foundation and the Agence Nationale de la Recherche (ANR-18-CE02-0008 and ANR-19-CE34-0011) grants to A.Y.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

Scripts used in this study, assemblies and intermediate files are publicly available on Figshare (https://doi.org/10.6084/m9.figshare.c.7100635.v1.). Genome sequences generated for this study are available on NCBI BioProject: PRJNA1078835.

References

Bächli G. TaxoDros: the database on Taxonomy of Drosophilidae. [accessed 2022 May 02]. https://www.taxodros.uzh.ch/

Baião GC, Schneider DI, Miller WJ, Klasson L. Multiple introgressions shape mitochondrial evolutionary history in Drosophila paulistorum and the Drosophila willistoni group. Mol Phylogenet Evol. 2023:**180**:107683. https://doi.org/10.1016/j.ympev.2022.107683.

Ballard JWO. Comparative genomics of mitochondrial DNA in members of the Drosophila melanogaster subgroup. J Mol Evol. 2000:**51**(1):48–63. https://doi.org/10.1007/s002390010066.

Batalha-Filho H, Fjeldså J, Fabre P-H, Miyaki CY. Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. J Ornithol. 2013:**154**(1):41–50. https://doi.org/10.1007/s10336-012-0866-7.

Blischak PD, Chifman J, Wolfe AD, Kubatko LS. Hyde: a python package for genome-scale hybridization detection. Syst Biol. 2018:**67**(5):821–829. https://doi.org/10.1093/sysbio/syy023.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019:**15**(4):e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009:**10**(1):421. https://doi.org/10.1186/1471-2105-10-421.

Cavalcanti AG. Geographic variation of chromosome structure in Drosophila prosaltans. Genetics. 1948:**33**(6):529–536. https://doi.org/10.1093/genetics/33.6.529.

Chang ES, Neuhof M, Rubinstein ND, Diamant A, Philippe H, Huchon D, Cartwright P. Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proc Natl Acad Sci U S A. 2015:**112**(48):14912–14917. https://doi.org/10.1073/pnas.1511468112.

Charlesworth B, Campos JL, Jackson BC. Faster-X evolution: theory and evidence from Drosophila. Mol Ecol. 2018:**27**(19):3753–3771. https://doi.org/10.1111/mec.14534.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. Evolution of genes and genomes on the Drosophila phylogeny. Nature. 2007:**450**(7167):203–218. https://doi.org/10.1038/nature06341.

Conner WR, Delaney EK, Bronski MJ, Ginsberg PS, Wheeler TB, Richardson KM, Peckenpaugh B, Kim KJ, Watada M, Hoffmann AA, et al. A phylogeny for the Drosophila montium species group: a model clade for comparative analyses. Mol Phylogenet Evol. 2021:**158**:107061. https://doi.org/10.1016/j.ympev.2020.107061.

Dagilis AJ, Peede D, Coughlan JM, Jofre GI, D'Agostino ERR, Mavengere H, Tate AD, Matute DR. A need for standardized reporting of introgression: insights from studies across eukaryotes. Evol Lett. 2022:**6**(5):344–357. https://doi.org/10.1002/evl3.294.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021:**10**(2):giab008. https://doi.org/10.1093/gigascience/giab008.

David JR, Ferreira EA, Jabaud L, Ogereau D, Bastide H, Yassin A. Evolution of assortative mating following selective introgression of pigmentation genes between two Drosophila species. Ecol Evol. 2022:**12**(4):e8821. https://doi.org/10.1002/ece3.8821.

de Campos Bicudo HM. Chromosomal polymorphism in the saltans group of Drosophila I. The saltans subgroup. Genetica. 1973a:**44**(4):520–552. https://doi.org/10.1007/BF00116808.

de Campos Bicudo HM. Reproductive isolation in the saltans group of Drosophila. I. The saltans subgroup. Genetica. 1973b:**44**(3):313–329. https://doi.org/10.1007/BF00161311.

de Campos Bicudo HM. Reproductive isolation of the saltans group of Drosophila. IV. The sturtevanti subgroup. Rev Brasil Genet. 1979:**2**:247–258.

de Campos Bicudo HM, Prioli AJ. Reproductive isolation in the saltans group of Drosophila II. The parasaltans subgroup. Genetica. 1978:**48**(1):17–22. https://doi.org/10.1007/BF00125282.

de Castro JP, Cararetto CMA. P elements in the saltans group of Drosophila: a new evaluation of their distribution and number of genomic insertion sites. Mol Phylogenet Evol. 2004:**32**(1):383–387. https://doi.org/10.1016/j.ympev.2004.01.005.

Deckmyn OS code by RAB and ARWR version by RBE by TPM and A. 2024. maps: Draw Geographical Maps. Available from: https://cran.r-project.org/web/packages/maps/index.html

de Magalhães LE. Notes on the taxonomy, morphology, and distribution of the saltans group of Drosophila, with descriptions of four new species. Univ Texas Publ. 1962:**2**:135–154.

de Magalhães LE, Björnberg AJS. Estudo da genitália masculina de "Drosophila" do grupo "saltans" (Diptera). Rev Bras Biol. 1957:**17**:435–450.

de Setta N, Loreto ELS, Cararetto CMA. Is the evolutionary history of the O-type P element in the saltans and willistoni groups of Drosophila similar to that of the canonical P element? J Mol Evol. 2007:**65**(6):715–724. https://doi.org/10.1007/s00239-007-9051-7.

Dobzhansky T. Experiments on sexual isolation in Drosophila. Proc Natl Acad Sci U S A. 1944:**30**(11):335–339. https://doi.org/10.1073/pnas.30.11.335.

Dobzhansky TG, Pavan C. Studies on Brazilian species of Drosophila. Boletim da Faculdade de Filosofia, Ciências e Letras da Universidade de São Paulo. Biol Geral Separata. 1943:**36**:7–72.

Doronina L, Churakov G, Shi J, Brosius J, Baertsch R, Clawson H, Schmitz J. Exploring massive incomplete lineage sorting in arctoids (laurasiatheria, carnivora). Mol Biol Evol. 2015:**32**(12):3194–3204. https://doi.org/10.1093/molbev/msv188.

Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Mol Biol Evol. 2011:**28**(8):2239–2252. https://doi.org/10.1093/molbev/msr048.

Ellegren H. The different levels of genetic diversity in sex chromosomes and autosomes. Trends Genet. 2009:**25**(6):278–284. https://doi.org/10.1016/j.tig.2009.04.005.

Frankel LE, Ané C. Summary tests of introgression are highly sensitive to rate variation across lineages. Syst Biol. 2023:**72**(6):1357–1369. https://doi.org/10.1093/sysbio/syad056.

Gagnon E, Hilgenhof R, Orejuela A, McDonnell A, Sablok G, Aubriot X, Giacomin L, Gouvêa Y, Bragionis T, Stehmann JR, et al. Phylogenomic discordance suggests polytomies along the backbone of the large genus Solanum. Am J Bot. 2022:**109**(4):580–601. https://doi.org/10.1002/ajb2.1827.

Glor RE. Phylogenetic insights on adaptive radiation. Annu Rev Ecol Evol Syst. 2010:**41**(1):251–270. https://doi.org/10.1146/annurev.ecolsys.39.110707.173447.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. A draft sequence of the neandertal genome. Science. 2010:**328**(5979):710–722. https://doi.org/10.1126/science.1188021.

Hibbins MS, Hahn MW. Phylogenomic approaches to detecting and characterizing introgression. Genetics. 2022:**220**(2):iyab173. https://doi.org/10.1093/genetics/iyab173.

Hime PM, Lemmon AR, Lemmon ECM, Prendini E, Brown JM, Thomson RC, Kratovil JD, Noonan BP, Pyron RA, Peloso PLV, et al. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. Syst Biol. 2021:**70**(1):49–66. https://doi.org/10.1093/sysbio/syaa034.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018:**35**(2):518–522. https://doi.org/10.1093/molbev/msx281.

Hoorn C, Boschman LM, Kukla T, Sciumbata M, Val P. The miocene wetland of western Amazonia and its role in neotropical biogeography. Bot J Linn Soc. 2022:**199**(1):25–35. https://doi.org/10.1093/botlinnean/boab098.

Jiang X, Edwards SV, Liu L. The multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. Syst Biol. 2020:**69**(4):795–812. https://doi.org/10.1093/sysbio/syaa008.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:**30**(4):772–780. https://doi.org/10.1093/molbev/mst010.

Khallaf MA, Cui R, Weißflog J, Erdogmus M, Svatoš A, Dweck HKM, Valenzano DR, Hansson BS, Knaden M. Large-scale characterization of sex pheromone communication systems in Drosophila.

Nat Commun. 2021:**12**(1):4165. https://doi.org/10.1038/s41467-021-24395-z.
Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ER, Pelaez J, et al. Highly contiguous assemblies of 101 drosophilid genomes. Elife. 2021:**10**:e66405. https://doi.org/10.7554/eLife.66405.
Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. Bioinformatics. 2017:**33**(17):2759–2761. https://doi.org/10.1093/bioinformatics/btx304.
Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 2023:**51**(D1):D445–D451. https://doi.org/10.1093/nar/gkac998.
Ledo RMD, Colli GR. The historical connections between the Amazon and the Atlantic Forest revisited. J Biogeogr. 2017:**44**(11):2551–2563. https://doi.org/10.1111/jbi.13049.
Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015:**31**(10):1674–1676. https://doi.org/10.1093/bioinformatics/btv033.
Li F, Rane RV, Luria V, Xiong Z, Chen J, Li Z, Catullo RA, Griffin PC, Schiffer M, Pearce S, et al. Phylogenomic analyses of the genus Drosophila reveals genomic signals of climate adaptation. Mol Ecol Resour. 2022:**22**(4):1559–1581. https://doi.org/10.1111/1755-0998.13561.
Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:**34**(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
Li H. Protein-to-genome alignment with miniprot. Bioinformatics. 2023:**39**(1):btad014. https://doi.org/10.1093/bioinformatics/btad014.
Lin M-J, Iyer S, Chen N-C, Langmead B. Measuring, visualizing, and diagnosing reference bias with biastools. Genome Biol. 2024:**25**(1):101. https://doi.org/10.1186/s13059-024-03240-8.
Lipman DJ, Souvorov A, Koonin EV, Panchenko AR, Tatusova TA. The relationship of protein conservation and sequence length. BMC Evol Biol. 2002:**2**:20. https://doi.org/10.1186/1471-2148-2-20.
Llopart A, Herrig D, Brud E, Stecklein Z. Sequential adaptive introgression of the mitochondrial genome in Drosophila yakuba and Drosophila santomea. Mol Ecol. 2014:**23**(5):1124–1136. https://doi.org/10.1111/mec.12678.
Maddison W. Reconstructing character evolution on polytomous cladograms. Cladistics. 1989:**5**(4):365–377. https://doi.org/10.1111/j.1096-0031.1989.tb00569.x.
Maddison WP. Gene trees in Species trees. Syst Biol. 1997:**46**(3):523–536. https://doi.org/10.1093/sysbio/46.3.523.
Madi-Ravazzi L, Segala LF, Roman BE, Alevi KCC, Prediger C, Yassin A, Hua-VAN AL, Miller WJ. Integrative taxonomy and a new species description in the sturtevanti subgroup of the Drosophila saltans group (Diptera: Drosophilidae). Zootaxa. 2021:**4980**(2):269292. https://doi.org/10.11646/zootaxa.4980.2.3.
Mai D, Nalley MJ, Bachtrog D. Patterns of genomic differentiation in the Drosophila nasuta species complex. Mol Biol Evol. 2020:**37**(1):208–220. https://doi.org/10.1093/molbev/msz215.
Malinsky M, Matschiner M, Svardal H. Dsuite—fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 2021:**21**(2):584–595. https://doi.org/10.1111/1755-0998.13265.
Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. Curr Protoc. 2021:**1**(12):e323. https://doi.org/10.1002/cpz1.323.
Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol Biol Evol. 2015:**32**(1):244–257. https://doi.org/10.1093/molbev/msu269.
Matute DR, Comeault AA, Earley E, Serrato-Capuchina A, Peede D, Monroy-Eklund A, Huang W, Jones CD, Mackay TFC, Coyne JA. Rapid and predictable evolution of admixed populations between two Drosophila species pairs. Genetics. 2020:**214**(1):211–230. https://doi.org/10.1534/genetics.119.302685.
Mayr E, Dobzhansky T. Experiments on sexual isolation in Drosophila. Proc Natl Acad Sci U S A. 1945:**31**(2):75–82. https://doi.org/10.1073/pnas.31.2.75.
Meade A, Pagel M. Ancestral state reconstruction using BayesTraits. In: Luo H, editors. Environmental microbial evolution: methods and protocols. Methods in molecular biology. New York, NY: Springer US; 2022. p. 255–266.
Meng G, Li Y, Yang C, Liu S. Mitoz: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 2019:**47**(11):e63. https://doi.org/10.1093/nar/gkz173.
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020:**37**(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.
Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. The genomic consequences of hybridization. Elife. 2021:**10**:e69016. https://doi.org/10.7554/eLife.69016.
Moreyra NN, Almeida FC, Allan C, Frankel N, Matzkin LM, Hasson E. Phylogenomics provides insights into the evolution of cactophily and host plant shifts in Drosophila. Mol Phylogenet Evol. 2023:**178**:107653. https://doi.org/10.1016/j.ympev.2022.107653.
Morgan CC, Foster PG, Webb AE, Pisani D, McInerney JO, O'Connell MJ. Heterogeneous models place the root of the placental mammal phylogeny. Mol Biol Evol. 2013:**30**(9):2145–2156. https://doi.org/10.1093/molbev/mst117.
Nascimento AP, de Campos Bicudo HE. Esterase patterns and phylogenetic relationships of Drosophila Species in the Saltans subgroup (Saltans group). Genetica. 2002:**114**(1):41–51. https://doi.org/10.1023/A:1014672502359.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. Mol Biol Evol. 2015:**32**(1):268–274. https://doi.org/10.1093/molbev/msu300.
O'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, Collins LS, de Queiroz A, Farris DW, Norris RD, et al. Formation of the Isthmus of Panama. Sci Adv. 2016:**2**(8):e1600883. https://doi.org/10.1126/sciadv.1600883.
O'Grady PM, Clark JB, Kidwell MG. Phylogeny of the Drosophila saltans species group based on combined analysis of nuclear and mitochondrial DNA sequences. Mol Biol Evol. 1998:**15**(6):656–664. https://doi.org/10.1093/oxfordjournals.molbev.a025969.
Owen CL, Miller GL. Phylogenomics of the Aphididae: deep relationships between subfamilies clouded by gene tree discordance, introgression and the gene tree anomaly zone. Syst Entomol. 2022:**47**(3):470–486. https://doi.org/10.1111/syen.12542.
Pease JB, Hahn MW. Detection and polarization of introgression in a five-taxon phylogeny. Syst Biol. 2015:**64**(4):651–662. https://doi.org/10.1093/sysbio/syv023.
Pélandakis M, Solignac M. Molecular phylogeny of Drosophila based on ribosomal RNA sequences. J Mol Evol. 1993:**37**(5):525–543. https://doi.org/10.1007/BF00160433.
Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol.2011:**9**(3):e1000602. https://doi.org/10.1371/journal.pbio.1000602.
Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Quéinnec E, et al. Phylogenomics revives traditional views on deep animal relationships. Curr Biol. 2009:**19**(8):706–712. https://doi.org/10.1016/j.cub.2009.02.052.
Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, Wiens M, Alié A, Morgenstern B, Manuel M, et al. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. Mol Biol Evol. 2010:**27**(9):1983–1987. https://doi.org/10.1093/molbev/msq089.

Quintero I, Keil P, Jetz W, Crawford FW. Historical biogeography using Species geographical ranges. Syst Biol. 2015:**64**(6): 1059–1073. https://doi.org/10.1093/sysbio/syv057.

Qvarnström A, Bailey RI. Speciation through evolution of sex-linked genes. Heredity (Edinb). 2009:**102**(1):4–15. https://doi.org/10.1038/hdy.2008.93.

R Core Team. 2023. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. https://www.R-project.org/.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:**11**(1):1432. https://doi.org/10.1038/s41467-020-14998-3.

Reilly PF, Tjahjadi A, Miller SL, Akey JM, Tucci S. The contribution of Neanderthal introgression to modern human traits. Curr Biol. 2022:**32**(18):R970–R983. https://doi.org/10.1016/j.cub.2022.08.027.

Rick JA, Brock CD, Lewanski AL, Golcher-Benavides J, Wagner CE. Reference genome choice and filtering thresholds jointly influence phylogenomic analyses. Syst Biol. 2023. **73**(1):76–101. https://doi.org/10.1093/sysbio/syad065.

Rodríguez-Trelles F, Tarrío R, Ayala FJ. Molecular evolution and phylogeny of the Drosophila saltans Species group inferred from the Xdh gene. Mol Phylogenet Evol. 1999:**13**(1):110–121. https://doi.org/10.1006/mpev.1999.0631.

Roman BE, Madi-Ravazzi L. Male terminalia morphology of sixteen species of the Drosophila saltans group Sturtevant (Diptera, Drosophilidae). Zootaxa. 2021:**5061**(3):523–544. https://doi.org/10.11646/zootaxa.5061.3.7.

Roman BE, Santana DJ, Prediger C, Madi-Ravazzi L. Phylogeny of Drosophila saltans group (Diptera: Drosophilidae) based on morphological and molecular evidence. PLoS One. 2022:**17**(4): e0266710. https://doi.org/10.1371/journal.pone.0266710.

Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. Less is more in Mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental Mammals. Mol Biol Evol. 2013:**30**(9):2134–2144. https://doi.org/10.1093/molbev/mst116.

Sankararaman S, Mallick S, Patterson N, Reich D. The combined landscape of denisovan and neanderthal ancestry in present-day humans. Curr Biol. 2016:**26**(9):1241–1247. https://doi.org/10.1016/j.cub.2016.03.037.

Sarver BAJ, Keeble S, Cosart T, Tucker PK, Dean MD, Good JM. Phylogenomic insights into mouse evolution using a pseudoreference approach. Genome Biol Evol. 2017:**9**(3):726–739. https://doi.org/10.1093/gbe/evx034.

Sayyari E, Mirarab S. Testing for polytomies in phylogenetic Species trees using quartet frequencies. Genes (Basel). 2018:**9**(3):132. https://doi.org/10.3390/genes9030132.

Schaeffer SW. Muller "elements" in Drosophila: how the search for the genetic basis for speciation led to the birth of comparative genomics. Genetics. 2018:**210**(1):3–13. https://doi.org/10.1534/genetics.118.301084.

Schrempf D, Szöllősi G. The sources of phylogenetic conflicts. In: Scornavacca C, Delsuc F, Galtier N, editors. Phylogenetics in the genomic era. No commercial publisher | Authors open access book; 2020. p. 3.1:1–3.1:23.

Seixas FA, Boursot P, Melo-Ferreira J. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. Genome Biol. 2018:**19**(1):91. https://doi.org/10.1186/s13059-018-1471-8.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:**31**(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N, Quéinnec É, Ereskovsky A, et al. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Curr Biol. 2017:**27**(7):958–967. https://doi.org/10.1016/j.cub.2017.02.031.

Skov L, Coll Macià M, Lucotte EA, Cavassim MIA, Castellano D, Schierup MH, Munch K. Extraordinary selection on the human X chromosome associated with archaic admixture. Cell Genom. 2023:**3**(3):100274. https://doi.org/10.1016/j.xgen.2023.100274.

Solignac M, Monnerot M, Mounolou J-C. Mitochondrial DNA evolution in themelanogaster species subgroup ofDrosophila. J Mol Evol. 1986:**23**(1):31–40. https://doi.org/10.1007/BF02100996.

Souza TAJ, Noll FB, de Campos Bicudo HE, Madi-Ravazzi L. Scanning electron microscopy of male Terminalia and its application to Species recognition and phylogenetic reconstruction in the Drosophila saltans group. PLoS One. 2014:**9**(6):e97156. https://doi.org/10.1371/journal.pone.0097156.

Spassky B. Morphological differences between sibling species of Drosophila. Univ Texas Publ. 1957:**5721**:48–61.

Sturtevant AH. The classification of the genus Drosophila, with descriptions of nine species. Unive Texas Publ. 1942:**4213**:5–51.

Sturtevant AH, Novitski E. THE HOMOLOGIES OF THE CHROMOSOME ELEMENTS IN THE GENUS DROSOPHILA. Genetics. 1941:**26**(5): 517–541. https://doi.org/10.1093/genetics/26.5.517.

Suh A. The phylogenomic forest of bird trees contains a hard polytomy at the root of neoaves. Zoologica Scripta. 2016:**45**(S1): 50–62. https://doi.org/10.1111/zsc.12213.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ, Lang M, Courtier-Orgogozo V, et al. Widespread introgression across a phylogeny of 155 Drosophila genomes. Curr Biol. 2022:**32**(1):111–123.e5. https://doi.org/10.1016/j.cub.2021.10.052.

Suvorov A, Scornavacca C, Fujimoto MS, Bodily P, Clement M, Crandall KA, Whiting MF, Schrider DR, Bybee SM. Deep ancestral introgression shapes evolutionary history of dragonflies and damselflies. Syst Biol. 2022b:**71**(3):526–546. https://doi.org/10.1093/sysbio/syab063.

Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics. 1993:**135**(2):599–607. https://doi.org/10.1093/genetics/135.2.599.

Throckmorton LH. The problem of phylogeny in the genus Drosophila.studies in genetics.II. Univ Texas Publ. 1962:**6205**: 207–343.

Throckmorton LH, de Magalhães LE. Changes with evolution of pteridine accumulations in species of the saltans group of the genus Drosophila. Univ Texas Publ. 1962:**6205**:489–505.

Toews DPL, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol. 2012:**21**(16):3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x.

Townsend JP, Su Z, Tekle YI. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. Syst Biol. 2012:**61**(5):835. https://doi.org/10.1093/sysbio/sys036.

Valiente-Mullor C, Beamud B, Ansari I, Francés-Cuesta C, García-González N, Mejía L, Ruiz-Hueso P, González-Candelas F. One is not enough: on the effects of reference genome for the mapping and subsequent analyses of short-reads. PLoS Comput Biol. 2021:**17**(1):e1008678. https://doi.org/10.1371/journal.pcbi.1008678.

Walsh HE, Kidd MG, Moum T, Friesen VL. Polytomies and the power of phylogenetic inference. Evolution. 1999:**53**(3):932–937. https://doi.org/10.2307/2640732.

Wang K, Lenstra JA, Liu L, Hu Q, Ma T, Qiu Q, Liu J. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. Commun Biol. 2018:**1**(1):169. https://doi.org/10.1038/s42003-018-0176-6.

Whelan NV, Kocot KM, Moroz LL, Halanych KM. Error, signal, and the placement of Ctenophora sister to all other animals. Proc Natl Acad Sci U S A. 2015:**112**(18):5773–5778. https://doi.org/10.1073/pnas.1503453112.

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci U S A.2014:**111**(45): E4859–E4868. https://doi.org/10.1073/pnas.1323926111.

Yassin A. Phylogenetic relationships among species subgroups in the Drosophila saltans group (Diptera: Drosophilidae): can morphology solve a molecular conflict. Zool Res. 2009:**30**(3):225–232. https://doi.org/10.3724/SP.J.1141.2009.03225.

Yusuf LH, Tyukmaeva V, Hoikkala A, Ritchie MG. Divergence and introgression among the virilis group of Drosophila. Evol Lett. 2022:**6**(6):537–551. https://doi.org/10.1002/evl3.301.

Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018:**19**(S6):153. https://doi.org/10.1186/s12859-018-2129-y.

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013:**29**(21):2669–2677. https://doi.org/10.1093/bioinformatics/btt476.