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*] and 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, Sweden.

[5]Center for Anatomy and Cell Biology, Department of Cell and Developmental Biology, Medical University of Vienna, Austria.

[*] These authors contributed equally.

Correspondence: amir.yassin@universite-paris-saclay.fr; lilian.madi@unesp.br.

Running title: Phylogenomics of the *Drosophila saltans* group

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

Scripts used in this study, assemblies and intermediate files are publicly available on Figshare (https://figshare.com/account/home#/collections/7100635; private link for review: https://figshare.com/s/04b0b0f46808a8cdeefa) and will be made public upon acceptance.

Genome sequences generated for this study are available on NCBI BioProject: PRJNA1078835.

Abstract

Phylogenomics revealed reticulate evolution to be widespread across taxa, but whether reticulation is due to low statistical power (soft polytomy) or it is a true evolutionary pattern (hard polytomy) remains a field of investigation. Here, we investigate the phylogeny and quantify reticulation in the *Drosophila saltans* species group, a Neotropical clade of the subgenus *Sophophora* comprising 23 species arranged in five subgroups, namely *cordata*, *elliptica*, *parasaltans*, *saltans* and *sturtevanti.* The relationships within and between the subgroups have long been problematic. We sequenced and assembled the genomes of 15 species. 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 ≥50 kb-long 1,797 syntenic blocks with conserved collinearity across Neotropical *Sophophora*. High incidences of reticulation (sometimes up to 90% of the blocks) were restricted to three nodes on the tree, at the split between the *cordata-elliptica-saltans* subgroups and at the origin of the *sturtevanti* and *saltans* subgroups. By contrast, cases with asymmetric discordances, which are often interpreted as evidence for interspecific introgression, did not exceed ~7% of the blocks. Historical biogeography analysis revealed that short inter-speciation times and greater overlap of ancestral geographical ranges can explain cases with predominant reticulation. Therefore, episodic rapid radiations have played a major role in the evolution of this largely understudied Neotropical clade.

Keywords: 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; Maddison 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 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 lead to reticulated evolution, introgression and hybridization have attracted much attention, first because they challenged long-held concept of reproductive isolation between species, and second due to the development of a number of bioinformatic tools and tests that quantify phylogenetic discordance 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. Comparisons between the proportions of the three topologies can yield one of four possible outcomes (Figure 1B): (i) complete reticulation, all topologies are equally encountered; (ii) incomplete reticulation, such as in the case of full hybridization 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. The earliest of introgression tests, Patterson's *D* or the ABBA-BABA 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 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.

Polytomies and incongruencies have been reported for the jumping fly *Drosophila saltans* species group, a clade of the subgenus *Sophophora* with 23 Neotropical species (Magalhães 1962; Bächli 2024). 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 (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. Hypothesis for their evolutionary relationships have been proposed using different methods and different morphological characters (Magalhães and Björnberg 1957; Throckmorton 1962; Throckmorton and Magalhães 1962; O'Grady et al. 1998; Yassin 2009; Souza et al. 2014; Roman et al. 2022), chromosome polymorphism (Bicudo 1973a), reproductive isolation (Bicudo 1973b; Bicudo and Prioli 1978; Bicudo 1979), protein polymorphism (Nascimento and Bicudo 2002) and gene sequences (Pélandakis and Solignac 1993; O'Grady et al. 1998; Rodríguez-Trelles et al. 1999; Castro and Carareto 2004; de Setta et al. 2007; Roman et al. 2022). The evolutionary relationships proposed are summarized in Supplementary Table S1.

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 between X-linked, autosomal and mitochondrial loci. To test how each of the four incongruence categories prevails across the genome, we devised a new χ^2-based test that we called 2A2B. The test uses pairwise comparisons of the three topologies proportions shown in Figure 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 (*D. sturtevanti*) or outgroup (*D. willistoni*) species. We found reticulation levels to differ among the datasets and the subgroups, and to correlate with rate of speciation and historical biogeography in the *saltans* group.

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-kmer 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). To evaluate the completeness of our assembled genomes, we searched for single-copy genes (SCG) 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). 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).

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 in IqTree2 (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 *D. 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,

150 and 568 SCG, respectively. These datasets were then used to reconstruct the species trees
151 using the multi-species coalescent model, and the genes within them were concatenated for
152 Bayesian and Maximum Likelihood phylogenetic inferences.
153 The trees generated by the 5 data sets showed very similar topologies with well supported
154 nodes either for the multi-species coalescent model analysis implemented in ASTRAL-III
155 (Zhang et al. 2018), the maximum-likelihood implemented in IqTree2, or Bayesian Inference
156 implemented in BEAST (Bouckaert et al. 2019) (Supplementary Figures S1, S2 and S3). The
157 *parasaltans* subgroup was placed as sister to all other subgroups, followed by the emergence
158 of the *sturtevanti* subgroup. The *cordata* and *elliptica* subgroups showed a close relationship,
159 and were sister to the *saltans* subgroup. The only discrepancy between the topologies was the
160 placement of *D. lehrmanae*, a newly discovered species in the *sturtevanti* subgroup (Madi-
161 Ravazzi et al. 2021). For *D. lehrmanae*, while maximum-likelihood and multi-species
162 coalescent analyses reported lack of branch support for multiple trees (Supplementary Figures
163 S2 and S3), Bayesian inference recovered well supported branches and two topologies (Figure
164 2A and Supplementary Figure S1). The two distinct topologies distinguished the Muller
165 elements forming the X chromosome (elements A and D who are fused in a fusion shared by
166 the Neotropical *Sophophora*, *i.e.* the *saltans* and *willistoni* groups (Sturtevant and Novitski
167 1941; Dobzhansky and Pavan 1943; Cavalcanti 1948), and the Muller elements representing
168 autosomal chromosomes (elements B, C and E).
169 The published genome of *D. prosaltans* (Kim et al. 2021) did not group with the
170 genome of this species sequenced by us, instead it grouped with *D. saltans*. The genome
171 previously published comes from a line collected in El Salvador in 1957. According to
172 Magalhães' (1962) detailed morphological revision of multiple geographical specimens of the
173 *saltans* group, the sampling site of this particular strain is outside the geographical range of *D.*
174 *prosaltans*, but within the expected range of *D. saltans*. Furthermore, the *D. saltans* and *D.*

175 *prosaltans* lines used in our study underwent thorough morphological analyzes (Souza et al.
176 2014; Roman and Madi-Ravazzi 2021), indicating that the lines we used were accurately
177 identified. Therefore, it is most likely that the previously sequenced *D. prosaltans* strain from
178 El Salvador was misidentified and we consider it here to belong to *D. saltans*.
179
180 *Mitogenomes show cytonuclear conflicts in the* sturtevanti *and* saltans *subgroups*
181 We assembled mitochondrial genomes for the 15 *saltans* group species using MitoZ
182 (Meng et al. 2019). We did not use the previously assembled four strains since several
183 mitochondrial scaffolds were likely removed in those assemblies (Kim et al. 2021). We
184 conducted phylogenetic analysis on the aligned mitogenomes genes using both IqTree2 and
185 MrBayes (Ronquist et al. 2012). Overall, the mitochondrial trees matched the topology of the
186 nuclear gene trees regarding the inter-subgroup relationships. However, three major
187 discrepancies were identified (Figure 2B). First, the position of *D. lehrmanae* within the
188 *sturtevanti* subgroup did not agree with either the X or autosomal SCG topologies, proposing
189 a topology wherein *D. lehrmanae* is a sister species of *D. sturtevanti* (a topology recovered
190 once in multi-species coalescent analysis (Muller element C, Supplementary Figure S3) and
191 maximum likelihood (Muller element B, Supplementary Figure S2)). Second, whereas the
192 mitochondrial tree recovered the monophyletic relationship between the *elliptica*, *cordata* and
193 *saltans* subgroups, the position of *D. neocordata* (*cordata* subgroup) differed, being sister to
194 the three species of the *elliptica* subgroup in the nuclear trees and to the six species of the
195 *saltans* subgroup in the mitochondrial tree. Third, whereas nuclear trees recovered three
196 lineages within the *saltans* subgroup, namely*, austrosaltans*, *nigrosaltans-pseudosaltans*, and
197 *septentriosaltans-prosaltans-saltans*, only two lineages are revealed by the mitochondrial tree.
198 Intriguingly, each of the mitochondrial clades involved one species from otherwise sister
199 species in the nuclear trees, *i.e. D. nigrosaltans* and *D. saltans* in one clade and their

respective closely-related species *D. pseudosaltans* and *D. prosaltans* in the other clade. Because *D. saltans* and *D. prosaltans* are reported as closely related species in the nuclear trees and are separated in the two mitochondrial clades, the two mitotypes were called S and P, respectively. The distribution of closely-related species between distinct mitotypic groups suggest that multiple cytoplasmic introgression events might have occurred in this subgroup (Figure 2B).

The extent of reticulation differs between assembly- and pseudo-reference-based approaches

To quantify phylogenomic incongruences at a finer locus-scale, we analyzed allele distribution of bi-allelic phylogenetically informative sites in 28 species quartets using four distinct locus-defining strategies. Two strategies relied on our *de novo* genome assemblies. The first strategy included the use of single-copy genes (SCG) 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 strategies, a locus was defined as a 100 kb-long window of the reference genome.

For a bi-allelic site, the standard ABBA-BABA (Patterson's *D*) test considers a true species tree exists, with site distribution AABB, 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 (Figure 3A). Synteny-based strategy had the lowest $|D|$ across the

28 quartets ($|D| = 0.027 \pm 0.005$) compared to BUSCO ($|D| = 0.090 \pm 0.022$, Student's t $P <$ 0.008), *D. sturtevanti* pseudo-reference ($|D| = 0.176 \pm 0.026$, Student's t $P < 4.77 \times 10^{-6}$), and *D. willistoni* pseudo-reference ($|D| = 0.325 \pm 0.053$, Student's t $P < 5.97 \times 10^{-6}$). So, of the four strategies, the use of syntenic blocks is the most conservative.

To 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 AABB pattern. Therefore, the test allows classifying each locus into one of the four categories along the reticulation-bifurcation continuum given in Figure 1. Only loci with ≥20 evaluated sites were considered for this test. We found assembly and pseudo-reference strategies significantly differed in multiple aspects (Figure 3B). First, the extent of non-informative loci, *i.e.* loci with <20 evaluated sites, was higher for assembly strategies. Similarly, the number of total evaluated sites ranged from 16,279-185,163 and 21,838-281,749 in BUSCO and syntenic strategies, respectively, to 364,472-1,177,118 and 153,238-452,562 in *D. sturtevanti* and *D. willistoni* pseudo-reference genomes strategies, respectively. Second, assembly strategies had higher proportions of complete reticulation (*i.e.* category i), averaging ~40 and 39% in BUSCO and syntenic strategies, respectively, compared to pseudo-reference strategies, where it averaged ~9 and 20% in *D. sturtevanti* and *D. willistoni* pseudo-reference genomes strategies, respectively. This effect was independent from the number of the evaluated sites ($r = -0.17$, $P = 0.40$, in the synteny strategy), *i.e.* the high incidence of reticulation is not the result of low phylogenetic signal leading to a soft polytomy. Third, partial reticulation and partial bifurcation (*i.e.* categories ii and iii, respectively) predominated in the pseudo-reference strategies. Whereas the proportion of these categories did not exceed ~7% in assembly-based strategies, it reached ~46% and 34% in *D. sturtevanti* and *D. willistoni* pseudo-reference genomes strategies,

250 respectively (Student's *t* for synteny vs. *sturtevanti* pseudo-reference strategies $P < 7.24 \times 10^{-9}$). Because these two categories are related to hybridization and introgression, pseudo-reference strategies therefore seem to provide stronger evidence for introgression, in agreement to their higher |*D*| shown above. However, the phylogeny inferred from the pseudo-reference alignments significantly differed from each other and from other analyses, mainly in the position of the earliest branching subgroup, being the *sturtevanti* subgroup for the *sturtevanti*-based pseudo-references and the *cordata* subgroup for the *willistoni*-based pseudo-reference (Supplementary Figure S4). 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.

260

261 *Saltational episodes of reticulated evolution in the* saltans *group*

262 Assembly-based 2A2B test identified three parts on the species tree with reticulation indices (*i.e.* the proportion of blocks in categories iii and iv) exceeding 70% (Figure 3). At the inter-subgroup level, high incidences of reticulation were observed in any combination that involved representatives from at least two subgroups of the *cordata*, *elliptica* and *saltans* subgroups. For the *sturtevanti* subgroup, ~90% of the blocks could not resolve the relationships between *D. sturtevanti*, *D. lehrmanae* and the *dacunhai-milleri* clade, in agreement with the conflicting topologies between the X, autosomes and mitochondrial loci shown above. For the *saltans* subgroup, reticulation dominated (60-75%) in all comparisons involving *D. austrosaltans*, and representatives of the *nigrosaltans-pseudosaltans* and the *septentriosaltans-prosaltans-saltans* clades. However, not every subgroup with multiple representatives showed excess reticulation, since for the *elliptica* subgroup, almost no evidence for reticulate evolution was found whether *D. sturtevanti*, *D. neocordata* or any species of the *saltans* subgroup were used as an outgroup.

To test if historical biogeography can explain the present incongruences, we mapped current distribution of the studied species on the Bayesian X tree (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 (Figure 4A). 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

Towards 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 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 (Magalhães 1962; Throckmorton 1962; Throckmorton and Magalhães 1962; O'Grady et al. 1998; Rodríguez-Trelles et al. 1999; Castro and Carareto 2004; de Setta et al. 2007; Yassin 2009; Souza et al. 2014; Roman et al. 2022; see Supplementary Table S1).

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 *D. subsaltans*, *D. lusaltans*, *D. cordata*, and *D. rectangularis* through whole-genome sequencing promises to provide insight onto unresolved phylogenetic questions raised from previously published observations on reproductive isolation and morphology (Magalhães 1962; 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

323 reproductive isolation (Bicudo 1973b), can bring new insights into the reticulation evolution
324 in the *saltans* subgroup. The *saltans* subgroup showed the most dramatic signal of cyto-
325 nuclear discordance and reticulated evolution. Bicudo (1973b) investigated reproductive
326 isolation among the seven then described species of this subgroup, and in a remarkably partial
327 agreement with our nuclear phylogenomic trees, she concluded that *D. pseudosaltans*, *D.*
328 *nigrosaltans* and *D. austrosaltans* showed more basal relationships compared to *D. lusaltans*,
329 *D. septentriosaltans*, *D. prosaltans* and *D. saltans*. Indeed, nearly all crosses among the last
330 four species produce fertile females with some even producing fertile females and males
331 (Bicudo 1973b). This behavioral porosity largely agrees with the high incidence of reticulate
332 evolution we report here for this subgroup.
333 Two widespread species of the *saltans* subgroup, *D. saltans* and *D. prosaltans*, show a
334 peculiar geographical disjunction. The discrimination between strains belonging to each
335 species has long been erroneous (Dobzhansky 1944; Mayr and Dobzhansky 1945; Spassky
336 1957; Magalhães 1962) and we showed here that their misidentification persists even in the
337 genomic era (Kim et al. 2021; Suvorov et al. 2022). Interestingly, Bicudo (1973b) provided
338 evidence for reproductive reinforcement between these two sister species; sympatric
339 populations in their junction zone in Costa Rica demonstrated stronger reproductive isolation
340 than allopatric populations of both species. We have only included one to a few geographical
341 lines from each species and a broader sampling to investigate the extent of their reproductive
342 isolation and genome porosity is strongly needed.
343
344 *Intra- and inter-genomic conflicts impact the inference of phylogenetic patterns*
345 Concatenation helped recovering a sex chromosome versus autosome conflict, similar
346 to the one described by Mai et al. (2020) for the *Drosophila nasuta* subgroup. Like these
347 authors, this conflict was limited to a single part of the tree, *i.e.* the relationship of *D. pulau* to

348 *D. sulfurigaster sulfuricaster* and *D. s. bilimbata* in the *nasuta* group and the placement of *D.*
349 *lehrmanae* in the *sturtevanti* subgroup. The peculiarities of sexual chromosomes such as their
350 lower effective number, different recombination and mutation rates, and greater exposition to
351 natural selection when found in hemizygosity, lead to higher rates of adaptive evolution of
352 sexual-linked genes compared with autosomal genes (*i.e.* faster-X evolution) and also to the
353 disproportional accumulation of genes related to reproductive isolation and Dobzhansky-
354 Muller hybrid incompatibilities (*i.e.* Haldane's rule) (Charlesworth et al. 2018). Altogether,
355 those characteristics are thought to be responsible for the resistance to hybridization in the
356 sexual chromosomes (Ellegren 2009; Qvarnström and Bailey 2009; Sankararaman et al. 2016;
357 Charlesworth et al. 2018; Seixas et al. 2018; Mai et al. 2020; Matute et al. 2020; Moran et al.
358 2021; Reilly et al. 2022; Skov et al. 2023; but see David et al. 2022).
359 The second conflict regards a significant disagreement between mitochondrial
360 (mtDNA) and nuclear data. Discordance between nuclear and mitochondrial genomes is a
361 well-documented phenomenon in the tree of life as highlighted by Toews and Brelsford
362 (2012). Several characteristics of mtDNA, such as being haploid and uniparentally inherited,
363 resulting in a fourfold reduction in effective population size when compared with autosomal
364 chromosome loci, affect its evolution. Cytoplasmic introgression has long been recognized in
365 *Drosophila* (Solignac et al. 1986; Ballard 2000; Llopart et al. 2014). In a recent population
366 study within the *willistoni* group, multiple mitochondrial introgressions were observed in *D.*
367 *paulistorum* populations (Baião et al. 2023). These included an ancient introgression with a
368 highly divergent mitochondrial type, followed by more recent events. While nuclear-
369 mitochondrial incompatibilities likely posed challenges, the study also suggested two possible
370 alternatives to overcome these challenges: a selective advantage provided by the
371 mitochondrial type itself. Or a non-selective factor, such as *Wolbachia,* a bacteria known to
372 modify the reproduction of its host, could facilitate a mitochondrial type fixation (Baião et al.

373 2023). Although, interesting results have been reported from population approaches, conflicts
374 between nuclear and mitochondrial genomes have not been addressed in recent phylogenomic
375 analyses in the Drosophilidae (Mai et al. 2020; Khallaf et al. 2021; Suvorov et al. 2022; Yusuf
376 et al. 2022). The disagreement was particularly evident for the *saltans* subgroup, where it was
377 most likely of recent origins, separating species that have diverged only 0.7 myr ago, *i.e. D.*
378 *nigrosaltans* and *D. pseudosaltans*. Remarkably, the two mitotypes P and S do not correlate
379 with the degree of reproductive isolation inferred by Bicudo (1973b), contrary to nuclear tree,
380 indicating that cytoplasmic introgression in the *saltans* subgroup did not contribute to the
381 evolution of reproductive isolation in this clade.

382 Of the three subgroups for which multiple species were sequenced, the *saltans*
383 subgroup had the highest incidence of reticulation inferred by the 2A2B test. In this subgroup,
384 multiple large chromosomal inversions are known to be shared among closely-related species
385 (Dobzhansky and Pavan 1943; Cavalcanti 1948; Bicudo 1973a; Bicudo and Prioli 1978) and
386 evidence for balancing on ancestral inversion has been demonstrated in a number of cases
387 (Bicudo 1973a). Whether the high degree of reticulation in the *saltans* subgroup is associated
388 with large ancestral inversions potentially absent in other bifurcating clades would require the
389 future generation of chromosome-level assemblies for multiple *saltans* group species.

390

391 *The 2A2B test on syntenic blocks helps distinguishing "soft" and "hard" introgressions*

392 Our analyses test revealed that failure to account for a lack of phylogenetic signal in
393 the tested loci and alignment on a reference genome both significantly inflate the estimation
394 of the extent of interspecific introgression. Indeed, most studies that have provided strong
395 evidence for introgression using site-specific patterns comparisons have usually aligned reads
396 from multiple species to a single reference assembly (cf. studies compiled by Dagilis et al.
397 2021). Whereas such an approach would increase the power (Martin et al. 2015; Pease and

398 Hahn 2015), it also introduces biases due to paralogy, misalignments or absence of
399 collinearity among species, that are now known to bias phylogenetic inference itself
400 (Valiente-Mullor et al. 2021; Rick et al. 2023). Therefore, the most common practice now is
401 to use sets of clade-specific single-copy orthologs, such as the BUSCO genes, in inferring
402 phylogenies (Manni et al. 2021). Consequently, some recent introgression studies have
403 focused on such genes. For example, in a study of 155 genomes covering a wide range of
404 drosophilid lineages, Suvorov et al. (2022) analyzed these conserved genes. While finding
405 evidence for widespread introgression, they also showed that their discordance estimates were
406 highly sensitive to the length of the analyzed genes as well as by the slightest relaxation of
407 selective pressures. We attempted here to overcome this limitation by including larger
408 syntenic blocks, which in addition to conserved protein-coding sequences should also include
409 presumably neutral intronic and intergenic sites. Whereas, this approach showed the lowest *D*,
410 indicating that patterns predicted by introgression do not exceed 7% of the genome, it led to
411 only 1.5-fold increase in the number of analyzed sites, most likely because our assemblies
412 were based on only short reads. With the continuing progress in whole-chromosome
413 sequencing techniques and assembly-to-assembly alignment tools, a better understanding of
414 the extent of probably true or "hard" introgression events from artificial or "soft" ones may be
415 attained.

416 Phyletic radiations occur when isolation barriers simultaneously and rapidly evolve
417 between multiple populations of a species with a broad geographical or ecological breadth.
418 Consequently, genetic lineages will be incompletely sorted in the emerging species leading to
419 a lack of phylogenetic signals. This situation usually occurs following a rapid environmental
420 change or expansion into a new range with new empty niches. Indeed, our historical
421 biogeography approach shows that such situation might have favored the episodic signals of
422 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 how much hard polytomies associate with these events may have occurred.

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. For all species, DNA extraction, PE library preparation using NEBNext Protocol, and 75x 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.*

448 *saltans*, *D. neocordata*, *D. prosaltans* and *D. sturtevanti* published by Kim et al. (2021)
449 (assembly numbers ASM1890357v1, ASM1890361v1, ASM1815127v1 and
450 ASM1815037v1, respectively), and we used the *D. neocordata* assembly published by Baião
451 et al. (2023) in all analyses.
452
453 *Phylogenomics: Nuclear single-copy-genes*
454 For phylogenomics analysis, SCG searches were carried out using 3,285 SCG from
455 diptera_odb10 database on Busco5 (Simão et al. 2015). SCG present in all species were kept
456 and aligned using the L-INS-i method implemented on MAFFT (Katoh and Standley 2013)
457 (mafft --localpair --maxiterate 1000 –adjustdirection).
458 Genomic data of different species of *Drosophila* support that genes tend to be situated
459 within the same Muller element across multiple species (see Schaeffer 2018). Taking this
460 gene linkage into account, we reconstructed five independent datasets (Muller elements A-E),
461 each comprising all SCG found in the respective Muller element. To achieve this, we
462 performed a tBlastn search against the *D. melanogaster*, and subsequently we concatenated
463 genes found within the same Muller element. Phylogenetic trees were then constructed using
464 maximum likelihood and Bayesian methods implemented in the softwares IQ-TREE (Nguyen
465 et al. 2015) and BEAST (Bouckaert et al. 2019), respectively. Additionally, maximum-
466 likelihood trees were generated for each gene, the output tree from each Muller element data-
467 set were used to reconstruct to species trees, using multi-species coalesce model implemented
468 in ASTRAL-III (Zhang et al. 2018).
469
470 *Phylogenomics: Mitochondrial Genome*
471 Mitochondrial genomes were assembled and annotated with MitoZ (Meng et al. 2019),
472 with the Megahit assembler (Li et al. 2015). In order to ensure the exclusion of nuclear-

473 embedded mitochondrial DNA sequences within the assembly, a strategic approach was
474 taken. Considering that mitochondrial reads are found in higher frequency than nuclear-
475 mitochondrial DNA sequences, the read subsampling were set to 0.5 gigabases (--
476 data_size_for_mt_assembly 0.5). The genes obtained from the mitochondrial genome were
477 aligned using the MAFFT alignment tool with the --auto parameter due to the close similarity
478 between sequences. Subsequently, the aligned genes were concatenated into a dataset for
479 phylogenetic analysis. The concatenated dataset served as the basis for reconstructing
480 phylogenetic trees using both Maximum Likelihood (ML) implemented in IQ-TREE (Nguyen
481 et al. 2015) and Bayesian Inference (BI) in BEAST (Bouckaert et al. 2019).
482
483 *Identification of syntenic blocks*
484 The 20 genomes of the *saltans* group, the 15 sequenced here and the 5 published by Kim et al.
485 (2021) and Baião et al. (2023) were preliminary annotated with Miniprot (miniprot -Iut16, Li
486 2023). The primary objective of this annotation was to accurately map proteins from the
487 robust and reliable genome annotation of *D. willistoni* (Clark et al. 2007). After the protein
488 mapping, the predicted gene loci were assessed to identify syntenic blocks present in the
489 Neotropical *Sophophora* (comprising *D. willistoni* and *saltans* group). The identification of
490 these blocks was based on gene order and orientation, achieved using an in-house Perl script.
491 First, this script compares the scaffolds' genes order and orientation between the reference
492 assemblies of *D. saltans* and *D. sturtevanti*, the synteny blocks were defined when all the
493 genes were found in the same order and orientation in both species. The identified collinear
494 blocks were then searched for the remaining genomes. Blocks with missing data, *i.e.* missing
495 genes for one or more species were subsequently removed, and the remaining blocks were
496 subjected to a size-based filtering with a threshold of ≥50kb. The selected blocks were aligned
497 using the Mafft software (Katoh and Standley 2013).

498

499 *Pseudo-reference inference*

500 Illumina reads were aligned on either the *D. willistoni* reference genome (Clark et al.
501 2007: 12) or the recently published *D. sturtevanti* assembly (Kim et al. 2021) using minimap2
502 (Li 2018). Minimap2 generated BAM files were assembled into a pseudo-reference for each
503 species using the BEDTools software (Quinlan and Hall 2010). Each scaffold was then split
504 into 100 kb-long windows, and phylogenetic analyses were conducted on each window
505 separately using IqTree. A supertree was then inferred for each pseudo-reference genome
506 approach using ASTRAL-III.

507

508 *Quantifying reticulation using the 2A2B test*

509 Combinations of 28 species quartets were evaluated to test reticulation, bi-allelic sited
510 shared by pairs were searched in every locus, whether using SCG, syntenic blocks or pseudo-
511 reference 100 kb-long windows. Loci with at least 20 informative sites were retained. For
512 each locus, the occurrences AABB, ABBA or BABA patterns were counted, and three
513 pairwise χ^2-based tests were conducted, denoted *D1*, *D2*, and *D3* as follows (D1 is
514 reminiscent to Patterson's D):

515 $$D1 = (\sum ABBA - \sum BABA)/(\sum ABBA + \sum BABA),$$

516 $$D2 = (\sum AABB - \sum ABBA)/(\sum AABB + \sum ABBA), \text{ and}$$

517 $$D3 = (\sum AABB - \sum BABA)/(\sum AABB + \sum BABA).$$

518 Each of these estimates is reminiscent to the classical Tajima's (1993) Relative Rate Test
519 (RTT), which tests the difference in branch lengths from two ingroup taxa to a third outgroup
520 taxon, with lengths estimated as counts of different sites. However, unlike RRT, branch
521 lengths in our analyses are counted in terms of shared differences, *i.e.* singletons are not
522 considered. Similar to RRT, we applied a chi-squared analysis to test for deviation from parity

523 for each parameter. According to the significance levels of the three estimates, a locus can be
524 classified under each of the four categories presented in Figure 1. The test is run by the 2A2B
525 perl script associated with this manuscript.
526
527 *Historical biogeography*
528 To determine the sampling locations of the evaluated species, we conducted searches
529 in TaxoDros (https://www.taxodros.uzh.ch/search/class.php). Additionally, we incorporated
530 sampling sites that we ourselves had conducted. The accuracy of our species identification
531 was confirmed through DNA barcoding. After inspection of the geographical points and
532 manual correction, we identified the most northern, southern, western, and eastern points for
533 each species. We used those points to reconstruct the ancestral geographical range by
534 conducting a BayesTraits (Meade and Pagel 2022) analysis for each cartesian point
535 separately. The analyses were carried out using the GEO model with the phylogenetic tree
536 generated reconstructed with the Muller element A (XL chromosome arm), 1,000,000 of
537 MCMC iterations and 25% burn-in. The divergence times were estimated using this tree
538 under Bayesian inference. The calibration point used was the split between *D. willistoni* and
539 *D. sturtevanti* at 17.5 myr, as estimated by Suvorov et al. (2022).
540
541
542 **Acknowledgments**
543 We would like to thank David Ogereau for his assistance and insights in genome assembly for
544 this study. We thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for
545 the financial support to L.M.R. (Number processes: 95/06165-1, 2014/14059-0 and 2016/
546 11994-5) enabling us to collect and sequence many strains used here. We extend our thanks to
547 the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for their

support, specifically grant number 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. .

References

Bächli G. 2024. TaxoDros: the database on Taxonomy of Drosophilidae. Available from: https://www.taxodros.uzh.ch/

Baião GC, Schneider DI, Miller WJ, Klasson L. 2023. Multiple introgressions shape mitochondrial evolutionary history in *Drosophila paulistorum* and the *Drosophila willistoni* group. *Molecular Phylogenetics and Evolution* 180:107683.

Ballard JWO. 2000. Comparative Genomics of Mitochondrial DNA in Members of the Drosophila melanogaster Subgroup. *J Mol Evol* 51:48–63.

Batalha-Filho H, Fjeldså J, Fabre P-H, Miyaki CY. 2013. Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. *J Ornithol* 154:41–50.

Bicudo HEMC. 1973a. Chromosomal polymorphism in the saltans group of Drosophila I. The saltans subgroup. *Genetica* 44:520–552.

Bicudo HEMC. 1973b. Reproductive isolation in the saltans group of Drosophila. I. The saltans subgroup. *Genetica* 44:313–329.

Bicudo HEMC. 1979. Reproductive isolation of the saltans group of Drosophila. IV. the sturtevanti subgroup. *Revista Brasileira de Genética* 2:247–258.

Bicudo HEMC, Prioli AJ. 1978. Reproductive isolation in the saltans group of Drosophila II. The parasaltans subgroup. *Genetica* 48:17–22.

Blischak PD, Chifman J, Wolfe AD, Kubatko LS. 2018. HyDe: A Python Package for Genome-Scale Hybridization Detection. *Systematic Biology* 67:821–829.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, Maio ND, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 15:e1006650.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

Castro JP de, Carareto CMA. 2004. *P* elements in the *saltans* group of *Drosophila*: a new evaluation of their distribution and number of genomic insertion sites. *Molecular Phylogenetics and Evolution* 32:383–387.

Cavalcanti AGL. 1948. Geographic Variation of Chromosome Structure in Drosophila Prosaltans. *Genetics* 33:529–536.

Chang ES, Neuhof M, Rubinstein ND, Diamant A, Philippe H, Huchon D, Cartwright P. 2015. Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. *Proceedings of the National Academy of Sciences* 112:14912–14917.

Charlesworth B, Campos JL, Jackson BC. 2018. Faster-X evolution: Theory and evidence from Drosophila. *Molecular Ecology* 27:3753–3771.

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

Conner WR, Delaney EK, Bronski MJ, Ginsberg PS, Wheeler TB, Richardson KM, Peckenpaugh B, Kim KJ, Watada M, Hoffmann AA, et al. 2021. A phylogeny for the *Drosophila montium* species group: A model clade for comparative analyses. *Molecular Phylogenetics and Evolution* 158:107061.

Dagilis AJ, Peede D, Coughlan JM, Jofre GI, D'Agostino ERR, Mavengere H, Tate AD, Matute DR. 2021. 15 years of introgression studies: quantifying gene flow across Eukaryotes. :2021.06.15.448399. Available from: https://www.biorxiv.org/content/10.1101/2021.06.15.448399v1

David JR, Ferreira EA, Jabaud L, Ogereau D, Bastide H, Yassin A. 2022. Evolution of assortative mating following selective introgression of pigmentation genes between two Drosophila species. *Ecology and Evolution* 12:e8821.

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

Dobzhansky Th. 1944. Experiments on Sexual Isolation in Drosophila. *Proceedings of the National Academy of Sciences* 30:335–339.

Doronina L, Churakov G, Shi J, Brosius J, Baertsch R, Clawson H, Schmitz J. 2015. Exploring Massive Incomplete Lineage Sorting in Arctoids (Laurasiatheria, Carnivora). *Molecular Biology and Evolution* 32:3194–3204.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution* 28:2239–2252.

621 Ellegren H. 2009. The different levels of genetic diversity in sex chromosomes and
622 autosomes. *Trends in Genetics* 25:278–284.
623 Gagnon E, Hilgenhof R, Orejuela A, McDonnell A, Sablok G, Aubriot X, Giacomin L,
624 Gouvêa Y, Bragionis T, Stehmann JR, et al. 2022. Phylogenomic discordance suggests
625 polytomies along the backbone of the large genus Solanum. *American Journal of*
626 *Botany* 109:580–601.
627 Glor RE. 2010. Phylogenetic Insights on Adaptive Radiation. *Annual Review of Ecology,*
628 *Evolution, and Systematics* 41:251–270.
629 Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai
630 W, Fritz MH-Y, et al. 2010. A Draft Sequence of the Neandertal Genome. *Science*
631 328:710–722.
632 Hime PM, Lemmon AR, Lemmon ECM, Prendini E, Brown JM, Thomson RC, Kratovil JD,
633 Noonan BP, Pyron RA, Peloso PLV, et al. 2021. Phylogenomics Reveals Ancient
634 Gene Tree Discordance in the Amphibian Tree of Life. *Systematic Biology* 70:49–66.
635 Hoorn C, Boschman LM, Kukla T, Sciumbata M, Val P. 2022. The Miocene wetland of
636 western Amazonia and its role in Neotropical biogeography. *Botanical Journal of the*
637 *Linnean Society* 199:25–35.
638 Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
639 Improvements in Performance and Usability. *Molecular Biology and Evolution*
640 30:772–780.
641 Khallaf MA, Cui R, Weißflog J, Erdogmus M, Svatoš A, Dweck HKM, Valenzano DR,
642 Hansson BS, Knaden M. 2021. Large-scale characterization of sex pheromone
643 communication systems in Drosophila. *Nat Commun* 12:4165.

Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ER, Pelaez J, et al. 2021. Highly contiguous assemblies of 101 drosophilid genomes.Coop G, Wittkopp PJ, Sackton TB, editors. *eLife* 10:e66405.

Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* 33:2759–2761.

Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. 2023. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research* 51:D445–D451.

Ledo RMD, Colli GR. 2017. The historical connections between the Amazon and the Atlantic Forest revisited. *Journal of Biogeography* 44:2551–2563.

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676.

Li F, Rane RV, Luria V, Xiong Z, Chen J, Li Z, Catullo RA, Griffin PC, Schiffer M, Pearce S, et al. 2022. Phylogenomic analyses of the genus Drosophila reveals genomic signals of climate adaptation. *Molecular Ecology Resources* 22:1559–1581.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Li H. 2023. Protein-to-genome alignment with miniprot. *Bioinformatics* 39:btad014.

Llopart A, Herrig D, Brud E, Stecklein Z. 2014. Sequential adaptive introgression of the mitochondrial genome in Drosophila yakuba and Drosophila santomea. *Molecular Ecology* 23:1124–1136.

Maddison W. 1989. Reconstructing Character Evolution on Polytomous Cladograms. *Cladistics* 5:365–377.

Maddison WP. 1997. Gene Trees in Species Trees. *Systematic Biology* 46:523–536.

Madi-Ravazzi L, Segala LF, Roman BE, Alevi KCC, Prediger C, Yassin A, Hua-VAN AL, Miller WJ. 2021. Integrative taxonomy and a new species description in the sturtevanti subgroup of the Drosophila saltans group (Diptera: Drosophilidae). *Zootaxa* 4980:269292.

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

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

Mai D, Nalley MJ, Bachtrog D. 2020. Patterns of Genomic Differentiation in the Drosophila nasuta Species Complex. *Molecular Biology and Evolution* 37:208–220.

Malinsky M, Matschiner M, Svardal H. 2021. Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* 21:584–595.

Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols* 1:e323.

Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution* 32:244–257.

Matute DR, Comeault AA, Earley E, Serrato-Capuchina A, Peede D, Monroy-Eklund A, Huang W, Jones CD, Mackay TFC, Coyne JA. 2020. Rapid and Predictable Evolution of Admixed Populations Between Two Drosophila Species Pairs. *Genetics* 214:211–230.

Mayr E, Dobzhansky Th. 1945. Experiments on Sexual Isolation in Drosophila. *Proceedings of the National Academy of Sciences* 31:75–82.

Meade A, Pagel M. 2022. Ancestral State Reconstruction Using BayesTraits. In: Luo H, editor. Environmental Microbial Evolution: Methods and Protocols. Methods in

Molecular Biology. New York, NY: Springer US. p. 255–266. Available from: https://doi.org/10.1007/978-1-0716-2691-7_12

Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research* 47:e63.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37:1530–1534.

Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. 2021. The genomic consequences of hybridization.Wittkopp PJ, editor. *eLife* 10:e69016.

Moreyra NN, Almeida FC, Allan C, Frankel N, Matzkin LM, Hasson E. 2023. Phylogenomics provides insights into the evolution of cactophily and host plant shifts in *Drosophila*. *Molecular Phylogenetics and Evolution* 178:107653.

Morgan CC, Foster PG, Webb AE, Pisani D, McInerney JO, O'Connell MJ. 2013. Heterogeneous Models Place the Root of the Placental Mammal Phylogeny. *Molecular Biology and Evolution* 30:2145–2156.

Nascimento AP, Bicudo HEMC. 2002. Esterase Patterns and Phylogenetic Relationships of Drosophila Species in the Saltans Subgroup (Saltans Group). *Genetica* 114:41–51.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32:268–274.

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. 2016. Formation of the Isthmus of Panama. *Science Advances* 2:e1600883.

O'Grady PM, Clark JB, Kidwell MG. 1998. Phylogeny of the Drosophila saltans species group based on combined analysis of nuclear and mitochondrial DNA sequences. *Molecular Biology and Evolution* 15:656–664.

Owen CL, Miller GL. 2022. Phylogenomics of the Aphididae: Deep relationships between subfamilies clouded by gene tree discordance, introgression and the gene tree anomaly zone. *Systematic Entomology* 47:470–486.

Pease JB, Hahn MW. 2015. Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Systematic Biology* 64:651–662.

Pélandakis M, Solignac M. 1993. Molecular phylogeny of Drosophila based on ribosomal RNA sequences. *J Mol Evol* 37:525–543.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. *PLOS Biology* 9:e1000602.

Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Quéinnec E, et al. 2009. Phylogenomics Revives Traditional Views on Deep Animal Relationships. *Current Biology* 19:706–712.

Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, Wiens M, Alié A, Morgenstern B, Manuel M, et al. 2010. Improved Phylogenomic Taxon Sampling Noticeably Affects Nonbilaterian Relationships. *Molecular Biology and Evolution* 27:1983–1987.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.

Qvarnström A, Bailey RI. 2009. Speciation through evolution of sex-linked genes. *Heredity* 102:4–15.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* 11:1432.

Reilly PF, Tjahjadi A, Miller SL, Akey JM, Tucci S. 2022. The contribution of Neanderthal introgression to modern human traits. *Current Biology* 32:R970–R983.

Rick JA, Brock CD, Lewanski AL, Golcher-Benavides J, Wagner CE. 2023. Reference Genome Choice and Filtering Thresholds Jointly Influence Phylogenomic Analyses. *Systematic Biology*:syad065.

Rodríguez-Trelles F, Tarrío R, Ayala FJ. 1999. Molecular Evolution and Phylogeny of the *Drosophila saltans* Species Group Inferred from the *Xdh* Gene. *Molecular Phylogenetics and Evolution* 13:110–121.

Roman BE, Madi-Ravazzi L. 2021. Male terminalia morphology of sixteen species of the Drosophila saltans group Sturtevant (Diptera, Drosophilidae) . *Zootaxa* 5061:523–544.

Roman BE, Santana DJ, Prediger C, Madi-Ravazzi L. 2022. Phylogeny of Drosophila saltans group (Diptera: Drosophilidae) based on morphological and molecular evidence. *PLOS ONE* 17:e0266710.

Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. 2013. Less Is More in Mammalian Phylogenomics: AT-Rich Genes Minimize Tree Conflicts and Unravel the Root of Placental Mammals. *Molecular Biology and Evolution* 30:2134–2144.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* 61:539–542.

765 Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The Combined Landscape of
766 Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology*
767 26:1241–1247.
768 Sayyari E, Mirarab S. 2018. Testing for Polytomies in Phylogenetic Species Trees Using
769 Quartet Frequencies. *Genes* 9:132.
770 Schaeffer SW. 2018. Muller "Elements" in Drosophila: How the Search for the Genetic Basis
771 for Speciation Led to the Birth of Comparative Genomics. *Genetics* 210:3–13.
772 Schrempf D, Szöllősi G. 2020. The Sources of Phylogenetic Conflicts. In: No commercial
773 publisher | Authors open access book. p. 3.1:1. Available from: https://hal.science/hal-
774 02535482
775 Seixas FA, Boursot P, Melo-Ferreira J. 2018. The genomic impact of historical hybridization
776 with massive mitochondrial DNA introgression. *Genome Biology* 19:91.
777 de Setta N, Loreto ELS, Carareto CMA. 2007. Is the Evolutionary History of the O-Type P
778 Element in the saltans and willistoni Groups of Drosophila Similar to That of the
779 Canonical P Element? *J Mol Evol* 65:715–724.
780 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO:
781 assessing genome assembly and annotation completeness with single-copy orthologs.
782 *Bioinformatics* 31:3210–3212.
783 Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N,
784 Quéinnec É, Ereskovsky A, et al. 2017. A Large and Consistent Phylogenomic Dataset
785 Supports Sponges as the Sister Group to All Other Animals. *Current Biology* 27:958–
786 967.
787 Skov L, Coll Macià M, Lucotte EA, Cavassim MIA, Castellano D, Schierup MH, Munch K.
788 2023. Extraordinary selection on the human X chromosome associated with archaic
789 admixture. *Cell Genomics* 3:100274.

Solignac M, Monnerot M, Mounolou J-C. 1986. Mitochondrial DNA evolution in themelanogaster species subgroup ofDrosophila. *J Mol Evol* 23:31–40.

Souza TAJ, Noll FB, Bicudo HEMC, Madi-Ravazzi L. 2014. Scanning Electron Microscopy of Male Terminalia and Its Application to Species Recognition and Phylogenetic Reconstruction in the Drosophila saltans Group. *PLOS ONE* 9:e97156.

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

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

Sturtevant AH, Novitski E. 1941. THE HOMOLOGIES OF THE CHROMOSOME ELEMENTS IN THE GENUS DROSOPHILA. *Genetics* 26:517–541.

Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zoologica Scripta* 45:50–62.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Current Biology* 32:111-123.e5.

Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135:599–607.

Throckmorton LH. 1962. The problem of phylogeny in the genus Drosophila.Studies in Genetics.II. *University of Texas Publication* 6205:207–343.

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

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* 21:3907–3930.

Townsend JP, Su Z, Tekle YI. 2012. Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny. *Systematic Biology* 61:835.

Valiente-Mullor C, Beamud B, Ansari I, Francés-Cuesta C, García-González N, Mejía L, Ruiz-Hueso P, González-Candelas F. 2021. One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLOS Computational Biology* 17:e1008678.

Walsh HE, Kidd MG, Moum T, Friesen VL. 1999. Polytomies and the Power of Phylogenetic Inference. *Evolution* 53:932–937.

Whelan NV, Kocot KM, Moroz LL, Halanych KM. 2015. Error, signal, and the placement of Ctenophora sister to all other animals. *Proceedings of the National Academy of Sciences* 112:5773–5778.

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences* 111:E4859–E4868.

Yassin A. 2009. Phylogenetic relationships among species subgroups in the Drosophila saltans group (Diptera: Drosophilidae): Can morphology solve a molecular conflict. *Zool Res* 30:225–232.

Yusuf LH, Tyukmaeva V, Hoikkala A, Ritchie MG. 2022. Divergence and introgression among the virilis group of Drosophila. *Evolution Letters* 6:537–551.

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

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677.

840 **Figure legends**

841 **Figure 1 –** Distribution of bi-allelic patterns along the reticulation-bifurcation continuum and
842 the proposed 2A2B test. A) The distribution of bi-allelic sites of four species can generate
843 three distinct topologies. B) Based of the frequency of these topologies in a genome fragment,
844 this fragment can be categorized in (i) complete reticulation, (ii) incomplete reticulation, (iii)
845 incomplete bifurcation, and (iv) complete bifurcation. The 2A2B test estimates significant
846 deviation from parity in pairwise comparisons of the topologies. *: significant deviation, ns =
847 non-significant.

848

849 **Figure 2 –** Phylogenomic conflict between the X chromosome, autosomes, and the
850 mitochondria. A) Comparison of autosomal topology (left, represented by Muller element B)
851 and X chromosome topology (right, represented by Muller element A) demonstrates overall
852 agreement with minor incongruence. B) Mitochondrial-nuclear disagreement highlights
853 stronger incongruence between mitochondrial topology (left) and sexual chromosome
854 topology (right). All trees were inferred through Bayesian analysis. Divergence time
855 estimation (in million years ago, myr) for the sexual chromosome topology is provided (left).
856 Nodes are numbered for subsequent analyses. Maximum-likelihood and ASTRAL-III trees
857 are provided in Supplementary Figures S2 and S3, respectively. Posterior probabilities of all
858 nodes were equal to1.

859

860 **Figure 3 –** Extent of phylogenomic discordance across 28 species quartets as measured by A)
861 Patterson's *D* and B) the 2A2B test, using four locus-definition strategies (see text). For both
862 tests, assembly-based strategies have lower average *D* and provide stronger evidence for
863 reticulation at three episodes. Colors of bars in B) correspond to the four categories presented
864 in Figure 1, with grey bars corresponding to loci with less than 20 evaluated bi-allelic sites.

Figure 4 – Historical biogeography of the *saltans* species. Centroid geographical coordinates for each species were inferred and ancestral coordinates were inferred using BayesTraits on the Bayesian X chromosome chronogram given in Figure 2B. Internal nodes are projected on the current map of South and Central America. Three episodes of rapid radiation were identified by the 2A2B test (Figure 3B) and are identified here with their corresponding possible paleogeographical events, with photos of adult males of four representative species shown. Branch colors in the tree correspond to the subgroups colors given in Figure 2.

875 **List of Supplementary Figures**

876

877 **Figure S1 –** Phylogenetic relationships inferred using BEAST from BUSCO genes

878 concatenated according to their Muller's element.

879

880 **Figure S2 –** Phylogenetic relationships inferred using IqTree-2 from BUSCO genes

881 concatenated according to their Muller's element.

882

883 **Figure S3 –** Phylogenetic relationships inferred using ASTRAL-III from BUSCO genes

884 concatenated according to their Muller's element.

885

886 **Figure S4 –** Phylogenetic relationships inferred using ASTRAL-III from 100-kb long

887 windows for pseudo-references produced by aligning short reads either to A) the ingroup *D.*

888 *sturtevanti* or B) the outgroup *D. willistoni* reference assemblies.

889

890

891

892 **List of Supplementary Tables**

893

894 **Table S1 –** A synopsis of phylogenetic relationships previously published in the *Drosophila*

895 *saltans* species group. Species subgroups are represented by the first two letters in capital,

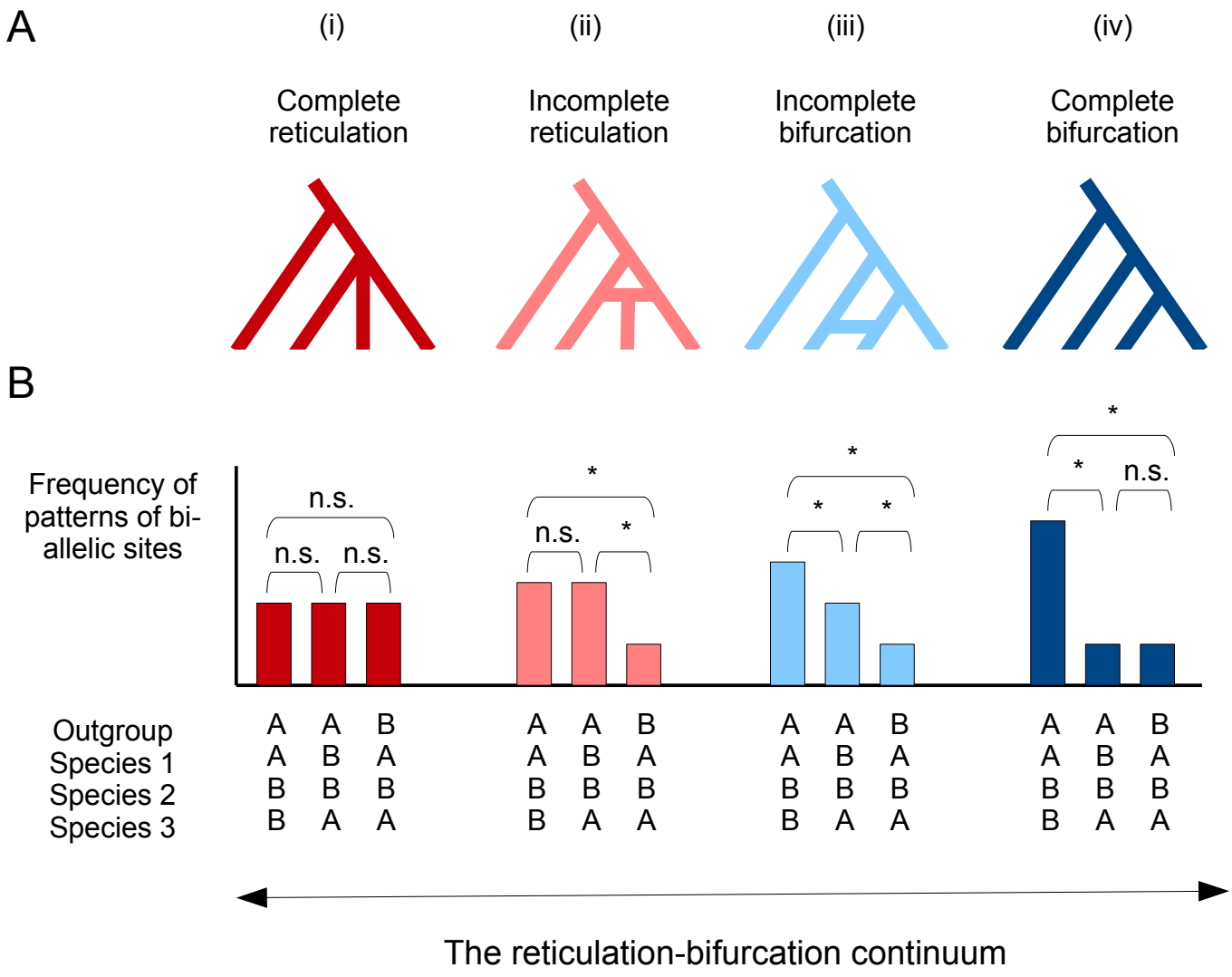
896 whereas species names are represented by the first three letter in small.

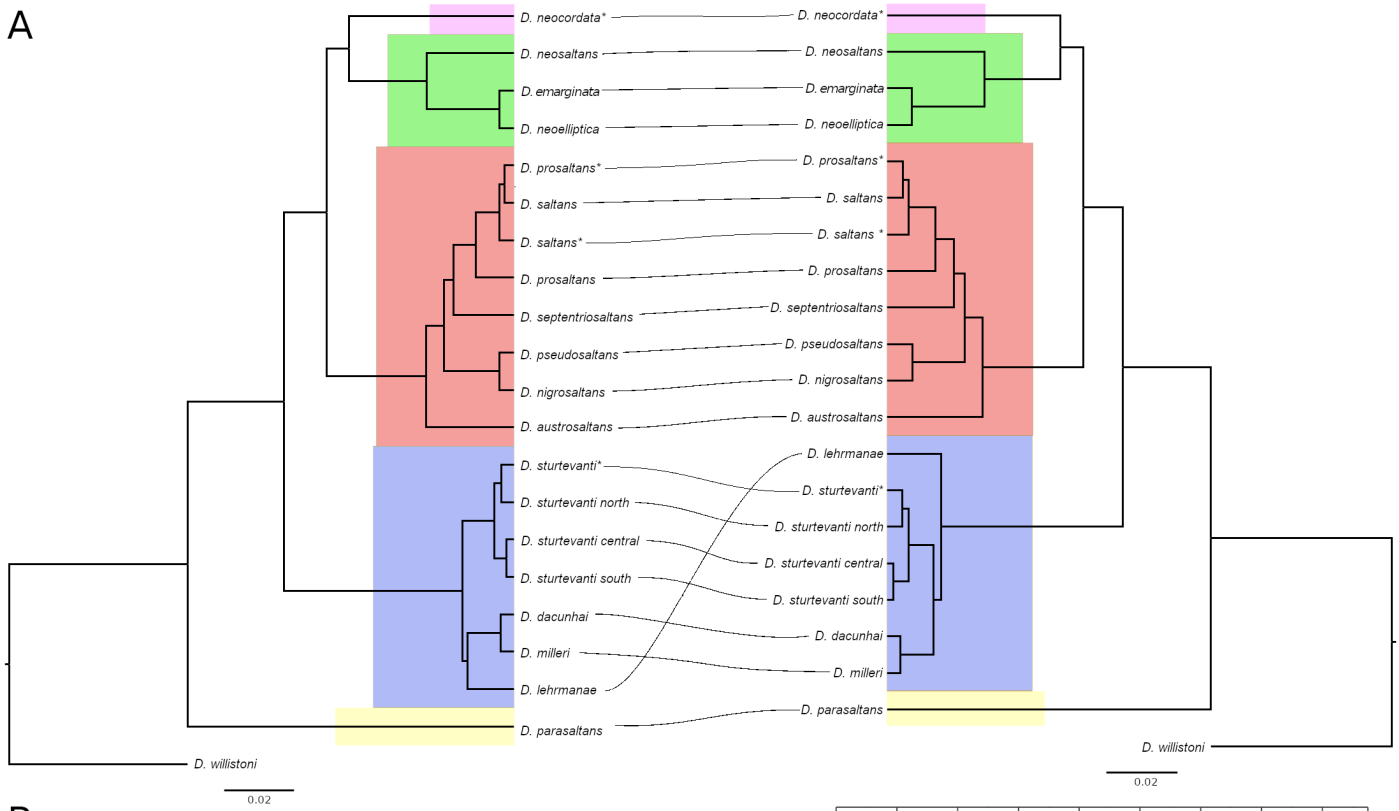
897

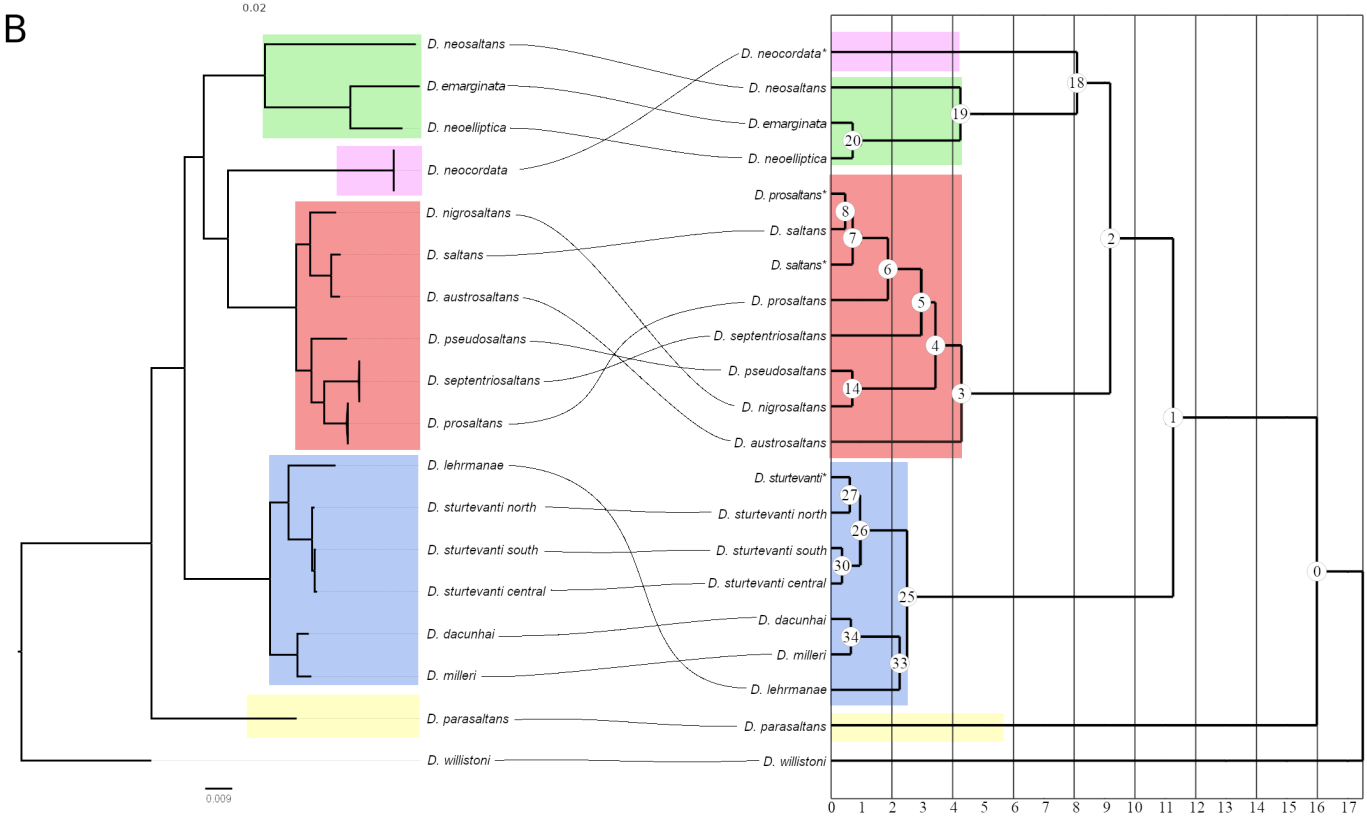
898 **Table S2 –** Source and genome assembly statistics including busco scores for the genomes

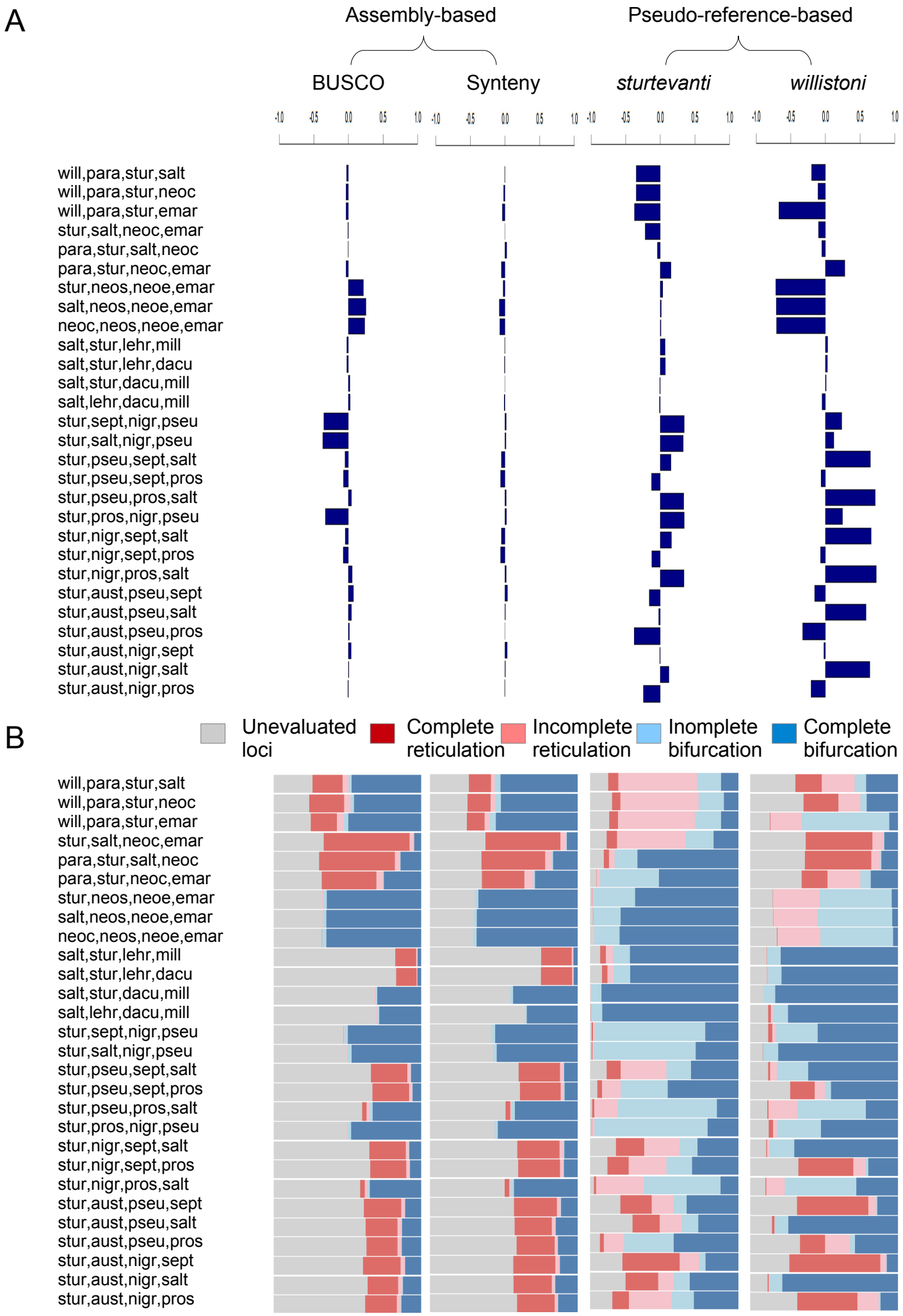
899 used in this study.

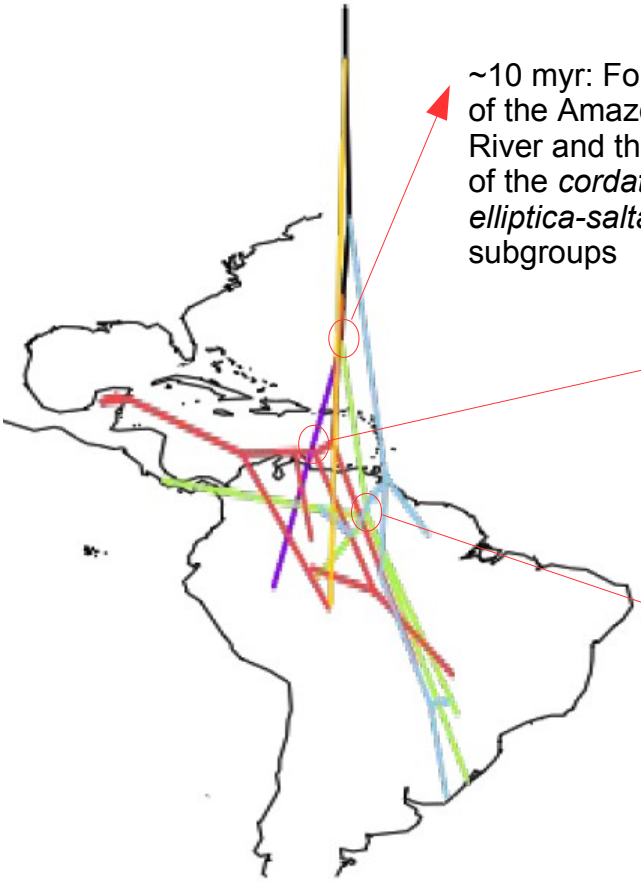
900


A
(i)
(ii)
(iii)
(iv)
Complete reticulation
Incomplete reticulation
Incomplete bifurcation
Complete bifurcation
B
Frequency of patterns of bi-allelic sites
n.s.
n.s.
n.s.
*
n.s.
*
*
*
*
*
*
n.s.
Outgroup
Species 1
Species 2
Species 3
A A B
A B A
B B B
B A A
A A B
A B A
B B B
B A A
A A B
A B A
B B B
B A A
A A B
A B A
B B B
B A A
The reticulation-bifurcation continuum


A
D. neocordata*
D. neosaltans
D. emarginata
D. neoelliptica
D. prosaltans*
D. saltans
D. saltans*
D. prosaltans
D. septentriosaltans
D. pseudosaltans
D. nigrosaltans
D. austrosaltans
D. sturtevanti*
D. sturtevanti north
D. sturtevanti central
D. sturtevanti south
D. dacunhai
D. milleri
D. lehrmanae
D. parasaltans
D. willistoni
0.02


B
D. neosaltans
D. emarginata
D. neoelliptica
D. neocordata
D. nigrosaltans
D. saltans
D. austrosaltans
D. pseudosaltans
D. septentriosaltans
D. prosaltans
D. lehrmanae
D. sturtevanti north
D. sturtevanti south
D. sturtevanti central
D. dacunhai
D. milleri
D. parasaltans
D. willistoni
0.009
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17


A
Assembly-based
Pseudo-reference-based
BUSCO
Synteny
sturtevanti
willistoni
-1.0
-0.5
0.0
0.5
1.0
will,para,stur,salt
will,para,stur,neoc
will,para,stur,emar
stur,salt,neoc,emar
para,stur,salt,neoc
para,stur,neoc,emar
stur,neos,neoe,emar
salt,neos,neoe,emar
neoc,neos,neoe,emar
salt,stur,lehr,mill
salt,stur,lehr,dacu
salt,stur,dacu,mill
salt,lehr,dacu,mill
stur,sept,nigr,pseu
stur,salt,nigr,pseu
stur,pseu,sept,salt
stur,pseu,sept,pros
stur,pseu,pros,salt
stur,pros,nigr,pseu
stur,nigr,sept,salt
stur,nigr,sept,pros
stur,nigr,pros,salt
stur,aust,pseu,sept
stur,aust,pseu,salt
stur,aust,pseu,pros
stur,aust,nigr,sept
stur,aust,nigr,salt
stur,aust,nigr,pros
B
Unevaluated loci
Complete reticulation
Incomplete reticulation
Inomplete bifurcation
Complete bifurcation



~10 myr: Formation of the Amazonian River and the split of the *cordata-elliptica-saltans* subgroups

~3 myr: Formation of Panama Isthmus and radiation of the *saltans* subgroup

~2.5 myr: Connection between the Amazonian and Atlantic Forests and radiation of the *sturtevanti* subgroup

D. neocordata

D. emarginata

D. prosaltans

D. sturtevanti