Saltational episodes of reticulate evolution in the jumping pomace fly *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 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

Abstract

Phylogenomics revealed reticulate evolution to be widespread across taxa, but whether reticulation is due to low statistical power (soft polytomy) or true evolutionary patterns (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*, whose relationships 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 ~5% of the blocks. Historical biogeography analysis revealed that short inter-speciational times and greater overlap of ancestral geographical ranges partly 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 also may 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; Whelan et al. 2015; Chang et al. 2015; Simion et al. 2017), mammals (Romiguier et al. 2013; Morgan 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 reticulate 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 (Figure 1A). 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*, compared the two later cases (iii and iv), *i.e.* it presumed that a “true” species tree exists. A later test, HyDe, 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 and a phylogenetic tree is still lacking.

Polytomies and incongruencies have been reported for the jumping pomace fly *Drosophila saltans* species group, a clade of the subgenus *Sophophora* with 23 Neotropical species (Magalhães 1962). 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; de 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 uses pairwise comparisons of the three topologies proportions in long syntenic blocks with conserved collinearity across the Neotropical *Sophophora* (Figure 1). We found reticulation levels to differ among the subgroups, in concordance with rate of speciation and historical biogeography.

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). Their genomes were also included in subsequent phylogenetic analyses, using the assembly of *D. willistoni* as an outgroup (Kim et al. 2021).

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 IqTree produced 1,263 distinct topologies, with 206 of them found more than once (Supplementary Data S1). 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, and 568 SCG, 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 Maximum Likelihood phylogenetic inferences.

The trees generated by the 5 data sets showed very similar topologies with well supported nodes either for the multi-species coalescent model analysis implemented in ASTRAL-III (Zhang et al. 2018), the maximum-likelihood implemented in IqTree (Nguyen et al. 2015) or Bayesian Inference implemented in BEAST (Bouckaert et al. 2019) (Supplementary Figures S1, S2 and S3). 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 *D. lehrmanae*, a newly discovered species in the *sturtevanti* subgroup (Madi-Ravazzi et al. 2021). For *D. lehrmanae*, while maximum-likelihood and multi species coalescent analyses reported lack of branch support for multiple trees, (Supplementary Figures S2 and S3), Bayesian inference recover well supported branches and two topologies (Figure 2A and Supplementary Figures S1). These two distinct topologies were identified among the Muller Elements forming the X chromosome (Muller elements A and D, a fusion shared by the Neotropical *Sophophora*, the *saltans* and *willistoni* groups, see Sturtevant and Novitski 1941; Dobzhansky and Pavan 1943; Cavalcanti 1948) and the Muller Elements representing autosomal chromosomes (Muller elements B, C, and E).

The published genome of *D. 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 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 most 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 IqTree and MrBayes. Overall, the mitochondrial trees matched the topology of the nuclear gene trees regarding the inter-subgroup relationships. However, three major discrepancies were identified (Figure 2B). First, the position of *D. lehrmanae* within the *sturtevanti* subgroup did not agree with either the X or autosomal SCG topologies, proposing topology wherein *D. lehrmanae* is a sister species of *D. sturtevanti* (topology recover once in Multi-Species Coalesce analysis (Muller element C, Supplementary Figure S3) and Maximum likelihood (Muller element B, Supplementary Figure S2)). Second, whereas the mitochondrial tree recovered the monophyletic relationship between the *elliptica*, *cordata* and *saltans* subgroups, the position of *D. 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. Intriguingly, each of the mitochondrial clades involved one species from otherwise sister 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 close 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.

Site-specific phylogenetic analysis of syntenic blocks quantifies the extent of reticulate evolution in the saltans group

Site-specific analyses of phylogenetic discordance are highly sensitive to locus size (Martin et al. 2015; Pease and Hahn 2015). To overcome this problem, we identified 1,797 syntenic blocks ≥50 kb-long with conserved collinearity across the 15 *saltans* assemblies and *D. willistoni* (see Methods). For a four-taxon species tree with an outgroup, three topologies can possibly be obtained for each site with two alleles (A and B), namely AABB, ABBA and BABA, with the AABB topology usually refers to the true species tree (Durand et al. 2011; Patterson et al. 2012). However, to consider cases where a true species tree cannot be inferred, we designed a test for reticulation, that we call 2A2B. The test consists of comparing each pair of the three topologies using a χ^2 test, and classify each block with ≥20 evaluated sites into one of the four categories along the reticulation-bifurcation continuum given in Figure 1B. We run this test for every possible quartet (Supplementary Table S3). Whereas blocks supporting bifurcating trees (categories iii and iv) predominated in most quartets, we identified three parts on the species tree with reticulation indices (*i.e.* the proportion of blocks in categories i and ii) 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.Remarkably, the proportion of categories supporting inter-specific hybridization (ii) or introgression (iii) rarely exceeded

5% of the 50-kb long syntenic blocks (Figure 3).

Historical biogeography partly explains excess reticulation

To test if historical biogeography could explain the present incongruences, we mapped current distribution of the studied species on the Bayesian X tree. For each species, a polygon connecting the four most extreme cartesian points was drawn and the ancestral position of each point was inferred using BayesTraits (Meade and Pagel 2022) (see Methods). This approach allowed us to infer an ancestral range at each internal node of the tree. The historical biogeography supported an Amazonian origin of the *saltans* group around 16 million years (myr) ago (Figure 2B, 4A, Supplementary Table S4). Internal nodes as old as or older than 4 myr ago had ranges confined to the central or northern parts of South America. These nodes included the ancestors of all species subgroups except *sturtevanti*. Northwestern dispersal into Panama and southern Central America occurred around 3 myr ago, which correlates with the geological formation of the isthmus of Panama (O'Dea et al. 2016), and involved the ancestors of the *nigrosaltans-pseudosaltans* and *septentriosaltans-prosaltans-saltans* clades. The *sturtevanti* subgroup diversified around 2.5 myr ago in the northern parts of South America.

We tested the effects of the successiveness of speciational times on the estimated proportion of syntenic blocks with reticulated evolution patterns (*i.e.* categories i and ii). For each quartet with an (outgroup,(sp.1,(sp.2,sp.3))) topology we tested the regression of the proportion of reticulation on the ratio of the divergence time between sp.2 and sp.3 and the divergence time of the three ingroup species (hereafter T2/T1 ratio). This ratio increased as the time between successive speciation events shortened. Reticulation positively correlated with this measurement, and the regression line followed an exponential pattern (R^2 = 0.98) (Figure 4C, Supplementary Table S5). We also tested the regression of reticulation on the degree of overlap between the ancestral ranges of sp.2 and sp.3 (i.e. node 1 and 2 in Figure 4D), and of all ingroup species (hereafter H2/H1 ratio). This ratio indicates the degree of conservation of ancestral habitat and possible connectivity. A strong exponential correlation was obtained for this ratio (R^2 = 0.89) (Figure 4C).

Discussion

Towards a comprehensive phylogeny of the saltans species group

A large number of *Drosophila* genomes have been sequenced and used in phylogenetic analyses (Suvorov et al. 2022; Khallaf et al. 2021; Kim et al. 2021; Li 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; de Castro and Carareto 2004; de Setta et al. 2007; Yassin 2009; Souza et al. 2014; Roman et al. 2022), see Supplementary Table S1 for previous suggested topologies).

After establishing a coherent phylogenetic picture for the *Drosophila saltans* species group and identifying the relationships among its subgroups, the next critical step lies in expanding our sampling efforts. 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 into 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 reproductive isolation (Bicudo 1973b), can bring new insights into the reticulation evolution. These prospects for heightened sampling efficacy and its potential to unlock further dimensions of the *saltans* subgroup's evolution are explored in greater detail in Supplementary Document S1.

The *saltans* subgroup showed the most dramatic signal of cyto-nuclear

discordance and reticulated evolution. Bicudo (1973a) 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 more basal relationships compared to *D. lusaltans*, *D. septentriosaltans*, *D. prosaltans* and *D. saltans*. Indeed, nearly all crosses among the last four species produce fertile females with some even producing fertile females and males (Bicudo 1973b). This behavioral 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. The discrimination between strains belonging to each species has long been erroneous (Dobzhansky 1944; Mayr and Dobzhansky 1945; Spassky 1957; Magalhães 1962) and we showed here that their misidentification persists even in the genomic era (Suvorov et al. 2022; Kim et al. 2021). Interestingly, Bicudo (1973a) 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 in the saltans *group*

Concatenation helped recovering a sexual versus autosome conflict, similar to the one described by Mai et al. (2020) for the *nasuta* subgroup. Like these authors, this conflict was limited to a single part of the tree, *i.e.* the relationship of *D. pulau* to *D. sulfurigaster sulfuricaster* and *D. s. bilimbata* in the *nasuta* group and the placement of *D. lehrmanae* in the *sturtevanti* subgroup. The peculiarities of sexual chromosomes such as the slower effective number, different recombination and mutation rates, the greater exposition to natural selection when found in hemizygosity, leads to higher rates of adaptive evolution of sexual-linked genes compared with autosomal genes (*i.e.* faster-X evolution) and also to the disproportional accumulation of genes related to reproductive isolation and Dobzhanski-Muller hybrid incompatibilities (*i.e.*

Haldane's rule). Altogether, those characteristics are thought to be responsible for the resistance to hybridization in the sexual 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 chromosome loci, affect its evolution. Cytoplasmic introgression 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 *D. paulistorum* populations. These included an ancient introgression with a highly divergent mitochondrial type, followed by more recent events. While nuclear-mitochondrial incompatibilities likely posed challenges, the study also suggested two possible alternatives to overcome these challenges: a selective advantage provided by the mitochondrial type it self. Or a non-selective factor, such as *Wolbachia, a* bacteria known to modify the reproduction of its host, could facilitate a mitochondrial type fixation (Baião et al. 2023). Although, interesting results have been report 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 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 mitotypes P and S do not correlate with the degree of reproductive isolation inferred by Bicudo (1973b), contrary to nuclear tree, indicating that cytoplasmic introgression in the *saltans* subgroup did not contribute to the evolution of reproductive isolation in this clade.

Syntenic blocks also allowed a quantification of the degree of reticulate evolution. Of the three subgroups for which multiple species were sequenced, the *saltans* subgroup had the highest incidence of reticulation. For all subgroups, the degree of reticulation correlated negatively with the time

between successive speciation events and positively with the degree of range conservatism. Indeed, reticulation is expected to increase with fast speciation increasing incomplete lineage sorting and/or range overlap promoting either gene flow or the selective retention of habitat-associated alleles (Avise and Robinson 2008; Degnan and Rosenberg 2009; Feng et al. 2022). In the *saltans* subgroups, multiple large chromosomal inversions are known to be shared among closely-related species (Dobzhansky and Pavan 1943; Cavalcanti 1948; Bicudo 1973a; Bicudo et al. 1978) and evidence for balancing on ancestral inversion has been demonstrated in a number of cases (Bicudo 1973a). Whether the high degree of reticulation in the *saltans* subgroup are 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.

Large syntenic blocks distinguish soft from hard polytomies in the saltans *group*

There is no consensus in current phylogenomic analysis between concatenating and partitioning approaches. Whereas the former approach increases the power, *i.e.* providing a total evidence, it also introduces bias due to the non-independence of linked loci and in some cases it cannot be computationally feasible to analyze whole genomes. Alternatively, multi-locus-coalescent (MLC) approaches that partition the data into presumably independent and neutral loci have been proposed. Those last approaches have broadly been applied in the investigation of phylogenetic discordances, mostly in studies inferring asymmetric introgressions. The definition of independent loci widely differs between studies with an impact on discordance estimate. For example, in a study of 155 genomes covering a wide range of drosophilid lineages, Suvorov et al. (2022) limited their MLC analyses on highly conserved single-copy protein-coding genes. However, their discordance estimates were highly sensitive to the length of the analyzed single genes as well as by the slightest relaxation of selective pressures, e.g., the exclusion of 5% of loci with the highest non-synonymous to synonymous ratio (dN/dS) led to a decrease of nearly 50% of discordance cases. An alternative approach is to align reads from multiple species to a well annotated genome, hence creating pseudo-references genome wherein different nucleotides replace their orthologous sites for each species. This approach was used by Mai et al. (2020) in the study

of the *D. nasuta* subgroup, a clade of 12 species that diverged ~3 myr ago (Suvorov et al. 2022). These authors defined loci in terms of 500-kb long windows for phylogenetic reconstruction and 50-kb long windows for discordance analyses. The 500-kb windows were either analyzed separately or concatenated according to chromosomal arm (Muller's element). Whereas such an approach would increase the signal, it also introduces biases due to paralogy, misalignments or absence of collinearity among species. Besides, this approach is highly sensitive to the choice of the reference genome (Valiente-Mullor et al. 2021; Rick et al. 2023).

We combined here both approaches. First, we based our phylogenetic analysis on conserved single-copy protein-coding genes like Suvorov et al. (2022), but like Mai et al. (2020) we concatenated those genes according to their Muller elements. Second, we inferred phylogenetic discordance using large ≥50 kb-long windows like Mai et al. (2020), but unlike these authors we did not infer pseudo-references and defined our windows on large syntenic blocks that conserved their collinearity across Neotropical *Sophophora*. Both approaches helped us to define signals of reticulate evolution that were not homogeneously distributed across the subgroups.(Mai et al. 2020; Khallaf et al. 2021; Suvorov et al. 2022; Yusuf et al. 2022)

Perhaps the most striking outcome of our synteny-based analysis is the low incidence of interspecific introgression compared to recent analyses across the genus *Drosophila* suggesting introgression to be widespread (Suvorov et al. 2022). Whether this discrepancy is due to the size of the analyzed loci or reflect genuine differences between the *saltans* group and other *Drosophila* clades would require the extension of the 2A2B test to these clades. Early phylogenetic studies in *Drosophila* suggested radiation episodes to be the most common evolutionary patterns in drosophilids (Throckmorton 1975). If hard polytomes are widespread, currently common introgression analyses based on the assumption of true bifurcating species trees may be misled. Given the ever growing evidence for introgression in other animal and plant clades, we strongly recommend the application of phylogenetic discordance tests in large syntenic blocks in these organisms as well to distinguish introgression from rapid radiation events.

Materials and Methods

Sample collection, whole genome sequencing and assembly

We performed whole genome sequencing on a group of female flies consisting of 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 S6. For all species except *D. neocordata*, which had its DNA extraction from ovaries and genome assembly described in BAIÃO et al. 2023, DNA was extracted following the manufacturer's instructions using the Promega DNeasy Kit. We conducted whole genome sequencing using the Illumina Hi-seq platform. The resulting genomes were then assembled using the Maryland Super Read Cabog Assembler (MaSuRCA) (Zimin et al. 2013), which utilizes both the Bruijn graph and overlap-layout-consensus (OLC) methods to generate super-reads. To assess the assembly's completeness, we searched for SCG using default parameters in Busco5 (Waterhouse et al. 2018) with the diptera_odb10 database (Kuznetsov et al. 2023).

Phylogenomics: Nuclear genes

In addition to the sequenced flies, we also utilized the reference 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) in our downstream analysis. For phylogenomics analysis, SCG searches were carried out using 3,285 SCG from diptera_odb10 database on Busco5 (Waterhouse et al. 2018). 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 the ancient proposition that genes tend to be situated within the same Muller element across multiple species, suggesting that natural selection has maintained a low rate of gene transposition between chromosomal arms (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*, and subsequently we concatenated genes found within the same Muller element. Phylogenetic trees were then constructed using maximum likelihood and Bayesian methods implemented in the softwares IQ-

TREE (Nguyen et al. 2015) and BEAST (Bouckaert et al. 2019), respectively. Additionaly, maximum-likelihood trees were generated for each gene, the output tree from each Muller element data-set were used to reconstruct to species trees, using multi-species coalesce model implemented in ASTRAL-III (Zhang et al. 2018).

Phylogenomics: Mitochondrial Genome

Mitochondrial genomes were assembled and annotated with MitoZ (Meng et al. 2019), with 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 Maximum Likelihood (ML) implemented in IQ-TREE (Nguyen et al. 2015) and Bayesian Inference (BI) in BEAST (Bouckaert et al. 2019).

Quantifying reticulation: 2A2B test

The 20 genomes of the *saltans* group, the 16 sequenced here and 4 published by Kim et al. (2021) 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*. After the 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 compare the scaffolds' genes order and orientation between the references genome *D. saltans* and *D. sturtevanti*, the synteny block were defined when all the genes were fond in same order and orientation for both species. The identified collinear blocks were than searched for the remaining genomes. Blocks with missing data, i.e. missing gene for one or more species were subsequently removed and subjected to a size-based filtering with

threshold of 50kb. The selected blocks were subjected to alignment using the Mafft (Katoh and Standley 2013). The resulting alignments were integral to the subsequent analysis, which aims to measure reticulation evolution, in the new 2A2B test.

All combination of quartets species were evaluated for test reticulation, bi-allelic non degenerated sited shared by pairs were searched in every synteny blocks. Collinear blocks that presented at least 20 informative sites between the evacuated quartets were keep. Bi-alellic sites shared between pairs of species quartets can generate 3 topologies, AABB (species 2 and 3 closely related), ABBA (species 1 and 2 closely related), and BABA (species 1 and 3 closely related) as shown in Figure 1A. For each synteny block, the occurrences of these three topologies were counted, and three different χ^{2}-based tests were conducted. First, the Patterson's D, this measure quantifies the difference in allele sharing between species pairs. It provides insights into whether a ABBA or BABA topology is more prevalent. It is calculated as the difference in allele sharing normalized by the total allele sharing as in Equation 1.

$$D1=\frac{\left(\sum ABBA-\sum BABA\right)^{2}}{\sum ABBA+\sum BABA} \quad (1),$$

The two other test were D2 and D3 (Equation 2 and 3) focus on discordant allele-sharing patterns (AABB vs. ABBA and AABB vs. BABA, respectively). They help identify cases where allele sharing between species pairs deviates from what's expected under a simple divergence model. Significant values for D1 or D2 might indicate that certain alleles are more shared between species pairs than expected.

$$D2\frac{\left(\sum AABB-\sum BABA\right)^{2}}{\sum AABB+\sum BABA} \quad (2), \text{ and}$$

$$D3=\frac{\left(\sum AABB-\sum ABBA\right)^{2}}{\sum AABB+\sum ABBA} \quad (3)$$

Afterward, based on how these three topologies were distributed within each synteny block and considering the significance of the three tests, the

collinear blocks were grouped into one of four categories (see figure 1B). Class I comprises complete reticulation, the frequencies of the three topologies do not deviate from neutral expectation, i.e. they are equal (D1, D2 and D3 are not significant), high frequencies of this class are caused by incomplete lineage sorting. Class II, comprises the cases in which two topologies do not significantly differentiate between them and are more frequent than the third topology (i.e. two D tests are significant), high frequencies of this class are expected in cases where hybridization had happen. Class iii, incomplete bifurcating, comprises the cases in with the frequency of all topologies are significantly different. Blocks classified in class iii show asymmetric introgression signals. Class iv comprise the cases in each one topology is much more frequent than the two alternatives ones, and the minor topologies are not significantly different from each other. High frequency of class iv is expected under the complete lineage sorting, and it is seen when one topology frequency greater outweighs two the alternative ones, which do not different between them. After each block classification, the overall genome porosity between the quartets were evaluated.

Historical biogeography

To determine the sampling locations of the evaluated species, we conducted searches in TaxoDros (https://www.taxodros.uzh.ch/search/class.php). Additionally, we incorporated sampling sites that we ourselves had conducted. It is important to note that the accuracy of our species identification was confirmed through BarCode verification. After inspection of the geographical points and manual correction of inverted coordinates, we identified the most northern, southern, western, and eastern points for each species. We used those points to reconstruct the ancestral geographical extremes in BayesTraits (Meade and Pagel 2022). This analysis was carried out using the GEO model with the phylogenetic tree generated reconstructed with the Miller element A (XL chromosome arm), 1.000.000 of MCMC 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* (17.5 myr), as estimated by Suvorov et al. (2022).

To assess the relationship between reticulation ratio and speciation ratio, we employed specific calculations. The reticulation ratio, indicating the frequency of syntenic blocks in class i and ii, was computed for groups of four

species. Similarly, the speciation (T2/T1) ratio was also calculated using quartets, it is determined by the divergence time between the ancestral of species 2 and species 3 in relation to the divergence time of the species 1. We also evaluated the relationship between reticulation and ancestral connectivity between the species. To do that, we utilized the ancestral geographical extremes and determined the predicted overlap area using the polygon R package. Finally, we computed a ratio according to Equation 4:

$$H2/H1\,ratio = \frac{2 \times H}{E_{A1} + E_{A2}} \quad (4)$$

Here, “H” represented the shared geographical area, “E” is the exclusive geographical area of ancestral 1 (A1) and 2 (A2). The fit for linear and exponential regressions between Reticulation and T2/T1 ratio and between Reticulation and H2/H1 ratio were calculated.

References

Avise JC, Robinson TJ. 2008. Hemiplasy: A New Term in the Lexicon of Phylogenetics.Kubatko L, editor. *Systematic Biology* 57:503–507.

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.

Ballard JWO. 2000. Comparative genomics of mitochondrial DNA in members of the Drosophila melanogaster subgroup. *Journal of Molecular Evolution* 51:48–63.

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* II:247–258.

Bicudo HEMC, Hosaki MK, Machado J, Marques MCN. 1978. Chromosomal polymorphism in the *saltans* group of *Drosophila* II. Further study on *D. prosaltans*. *Genetica* 48:5–15.

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

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15:1–28.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10:1–9.

de Castro JP, 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.

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.

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.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 24:332–340.

Dobzhansky T. 1944. Experiments on Sexual Isolation in *Drosophila*: III. Geographic Strains of *Drosophila Sturtevanti*. *Proceedings of the National Academy of Sciences* 30:335–339.

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.* 36:7–72.

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. *Mol Biol Evol* 28:2239–2252.

Ellegren H. 2009. The different levels of genetic diversity in sex chromosomes and autosomes. *Trends in Genetics* 25:278–284.

Feng S, Bai M, Rivas-González I, Li C, Liu S, Tong Y, Yang Haidong, Chen G, Xie D, Sears KE, et al. 2022. Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell* 185:1646-1660.e18.

Gagnon E, Hilgenhof R, Orejuela A, McDonnell A, Sablok G, Aubriot X, Giacomin L, Gouvêa Y, Bragionis T, Stehmann JR, et al. 2022. Phylogenomic discordance suggests polytomies along the backbone of the large genus *Solanum*. *Am J Bot*.

Glor RE. 2010. Phylogenetic Insights on Adaptive Radiation. *Annual Review of Ecology, Evolution, and Systematics* 41:251–270.

Guillín ER, Rafael V. 2017. Cinco especies nuevas del género *Drosophila* (Diptera, Drosophilidae) en la provincia de Napo, Ecuador. *Iheringia - Serie Zoologia* 107:1–12.

Hime PM, Lemmon AR, Lemmon ECM, Prendini E, Brown JM, Thomson RC, Kratovil JD, Noonan BP, Pyron RA, Peloso PLV, et al. 2021. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Systematic Biology* 70:49–66.

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

Khallaf MA, Cui R, Weißflog J, Erdogmus M, Svatoš A, Dweck HKM, Valenzano DR, Hansson BS, Knaden M. 2021. Large-scale characterization of sex pheromone 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 ERR, Pelaez J, et al. 2021. Highly contiguous assemblies of 101 drosophilid genomes. *eLife* 10:1–32.

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.

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. 2023. Protein-to-genome alignment with miniprot.Valencia A, editor. *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, Roman BE, Cesar K, Alevi C, Prediger C, Yassin A, Wolfgang J. 2021. Integrative taxonomy and a new species description in the *sturtevanti* subgroup of the *Drosophila saltans* group (Diptera: Drosophilidae). 4980:269–292.

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

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

Mai D, Nalley MJ, Bachtrog D, Wright S. 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.

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. *Methods Mol Biol* 2569:255–266.

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.

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. *Mol Biol Evol* 30:2145–2156.

Mourão CA, Bicudo HEMDC. 1967. Duas novas especies do grupo *saltans* (drosophilidae: diptera). *Papeis Avulsos de Zoologia* 20:123–134.

Nascimento AP, Bicudo HEMC. 2002. Esterase patterns and phylogenetic relationships of *Drosophila* species in the *saltans* subgroup (*saltans* group). *Genetica* 114:41–51.

Nguyen LT, 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.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient Admixture in Human History. *Genetics* 192:1065–1093.

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. *Journal of Molecular Evolution* 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.

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.

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. :2022.03.10.483737. Available from: https://www.biorxiv.org/content/10.1101/2022.03.10.483737v2

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.

Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology* 26:1241–1247.

Sayyari E, Mirarab S. 2018. Testing for Polytomies in Phylogenetic Species Trees Using Quartet Frequencies. *Genes* 9:132.

Schaeffer SW. 2018. Muller “Elements” in *Drosophila* : How the Search for the Genetic Basis for Speciation Led to the Birth of Comparative Genomics. *Genetics* 210:3–13.

Schrempf D, Szöllősi G. 2020. 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. p. chapter. 3.1, p. 3.1:1-3.1:23. Available from: https://hal.inria.fr/PGE

Seixas FA, Boursot P, Melo-Ferreira J. 2018. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. *Genome Biology* 19:91.

de Setta N, Loreto ELS, Carareto CMA. 2007. 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? *Journal of Molecular Evolution* 65:715–724.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.

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

Skov L, Coll Macià M, Lucotte EA, Cavassim MIA, Castellano D, Schierup MH, Munch K. 2023. Extraordinary selection on the human X chromosome associated with archaic admixture. *Cell Genomics* 3:100274.

Solignac M, Monnerot M, Mounolou JC. 1986. Mitochondrial DNA evolution in the melanogaster species subgroup of Drosophila. *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.

Spassky B. 1957. Morphological differences between sibling species of Drosophila. In: Genetics of Drosophila. Vol. 5721.

Sturtevant AH. 1942. The classification of the genus *Drosophila*, with descriptions of nine new species. *The University of Texas Publication* 4213:7–51.

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, Wadell P, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Current Biology* 32:111–123.

Tarrío R, Rodríguez-Trelles F, Ayala FJ. 1998. New *Drosophila* introns originate by duplication. *Proc Natl Acad Sci U S A* 95:1658–1662.

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

Throckmorton LH. 1975. The phylogeny, ecolopy, and geography of Drosophila. *Handbook of Genetics, Vol 3.*:421–469.

Throckmorton LH, Magalhães LE. 1962. Changes with evolution of pteridine accumulations in species of the *saltans* group of the genus *Drosophila*. *University of Texas Publications* 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* 61:835–849.

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.

Vilela CR, Bächli G. 1990. Taxonomic studies on Neotropical species of seven

genera of Drosophilidae (Diptera). *Mitteilungen der Schweizerischen Entomologischen Gesellschaft* 63:1–332.

Walsh HE, Kidd MG, Moum T, Friesen VL. 1999. Polytomies and the power of phylogenetic inference. *Evolution* 53:932–937.

Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* 35:543–548.

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. *Zoological Research* 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.

Figure 1. Distribution of bi-allelic patterns along the reticulation-bifurcation continuum and the 2A2B test. A) The distribution of bi-allelic sites of four species can generate three distinct topologies, BBAA with sp.3 and sp.4 as sister, ABBA with sp.1 and sp.2 and BABA with sp.1 and sp.3 are most closely related. B) Based of the frequency of these topologies in a genome fragment, this fragment can be categorized in (i) complete reticulation, T1=T2=T3, (ii) incomplete reticulation, T1=T2>T3, (iii) incomplete bifurcation, T1>T2>T3, and (iv) complete bifurcation, T1>t2=T3.

Figure 2. Phylogenomic Conflict of X Chromosome, Autosomal, and Mitochondria. A) Comparative Analysis of autosomal topology (left, represented

by Muller element B) and X-linked topology (right, represented by Muller element A) demonstrates overall agreement with minor Incongruence. B) Mitochondrial-Nuclear Disagreement highlight stronger incongruence between Mitochondrial Topology (left) and Sexual chromosome topology (right). Divergence time estimation (in million years ago, myr) for the Sexual Chromosome Topology is Provided.

Figure 3. The 2A2B test reveals a diminished introgression signal, while a prominent signal of reticulation evolution is evident within specific subgroups. The distribution of classes i-iv frequencies, spanning from symmetrical complete reticulation to asymmetrical bifurcation reticulation, is displayed for quartet species. A pronounced pattern of complete reticulation is apparent in the *saltans* and *sturtevanti* subgroups, whereas such a signal is absent in the *elliptica* subgroup.

Figure 4. Historical biogeography of the *saltans* group. A) the midpoint of the extreme geographical points for each ancestral node, reveals that the ancestral origins of all subgroups lie within the Amazonian forest, node numbers follows figure 2B. B) Blue line shows exponential relationship of reticulation in function of divergence time ratio of the three ingroup species (T2/T1 ratio) and reticulation (frequency of class i and ii). The black line depicts the exponential correlation between the overlap of ancestral ranges of the ingroup species (H2/H1 ratio) and reticulation. C) The exponential correlation between ancestral ranges of the ingroup species and reticulation. D) illustration of the method employed to calculate the overlap of ancestral ranges of the ingroup species (H2/H1 ratio). Specifically, the geographical ranges of the ancestors, nodes 1 and 2, were inferred using BayesTraits, enabling the determination of shared and unique proportions of geographical ranges.

List of Supplementary Material

Supplementary Data S1. 2,159 gene trees generated maximum likelihood estimation.

Supplementary Table S1. Summary of previous competing phylogenetic hypotheses in the *saltans* group. CO = *cordata* subgroup, EL = *elliptica*

subgroup, ST = *stutevanti* subgroup, PA = *parasaltans* subgroup, SA = *saltans* subgroup, aus= *D. austrosaltans*, nig = *D. nigrosaltans*, sal = *D. saltans*, pro = *D. prosaltans*, lus = *D. lusaltans*, sep = *D. septentriosaltans*, pse = *D. pseudosaltans*, stu= *D. sturtevanti*, leh = *D. lehrmanae*, mil = *D. milleri*, dac = *D. dacunhai*, nsa = *D. neosaltans*, nel = *D. neoelliptica*, ema= *D. emarginada*, OS = overall similarities, ai = measure of isolation for each interspecific cross, MP = maximum parsimonia, ML = maximum likelihood, BI =Bayesian inference.

Supplementary Table S2. Assembly quality, completeness of the genome of the saltans group. The total number of single copy genes used as baits is 3,285.

Supplementary Table S3. 2A2B Results for every quartets

SupplementaryTable S4. Node ages, ancestral area for the *saltans* group

Supplementary Table S5. T2/T1 and H2/H1ratio and reticulation estimated for the *saltans* group.

Supplementary Table S6. Location and number of individual used in the illumina PoolSeq.

Supplementary Figure S1. Bayesian Inference trees generated with 5 independent datasets, chromosome arms and respecteve Muller elements are indicated in each tree. Branch Posterior probabilities are shown for each node. The *parasaltans, sturtevanti, saltans, elliptica* and *cordata* subgroups are highlighted in yellow, blue, red, green and pink, respectively.

Supplementary Figure S2. Maximum likelihood trees generated with 5 independent datasets, each comprise the concatenate genes predicted to the Muller elements A-F. UltraFast Bootrap values are shown for each node. The *parasaltans, sturtevanti, saltans, elliptica* and *cordata* subgroups are highlighted in yellow, blue, red, green and pink, respectively.

Supplementary Figure S3. Species Tree generated under the multi-species coalecent model impremented in ASTRAL-III, from the 2,156 genes tree

available in Supplementary Data S1 and evaluated as 5 different data sets, acording to genes predicted to the Muller elements A-F. Branch support are shown for each node. The *parasaltans, sturtevanti, saltans, elliptica* and *cordata* subgroups are highlighted in yellow, blue, red, green and pink, respectively.

Supplementary Figure S4. Phylogenetic tree with inclusion of *D. lusaltans* and *D.* subsaltans. Mitochondrial tree reconstructed with inclusion of mitochondrial genes of *D. lusaltans* (A) and nuclear trees generated with the *Xdh* (B) and *Adh* (C) genes, which includes sequences of *D. subsaltans*. Branch supporter different than 1 are shown. The *parasaltans, sturtevanti, saltans, elliptica* and *cordata* subgroups are highlighted in yellow, blue, red, green and pink, respectively.

Figure 1

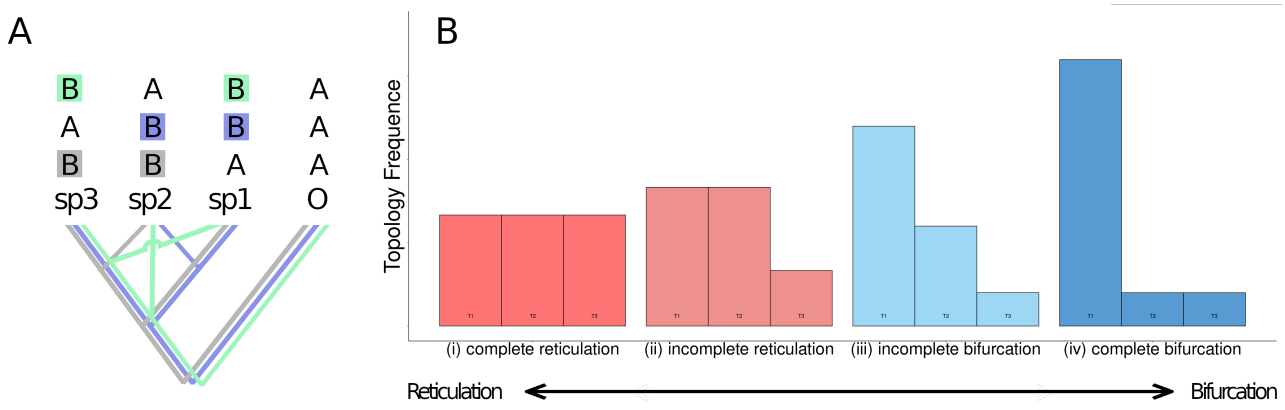
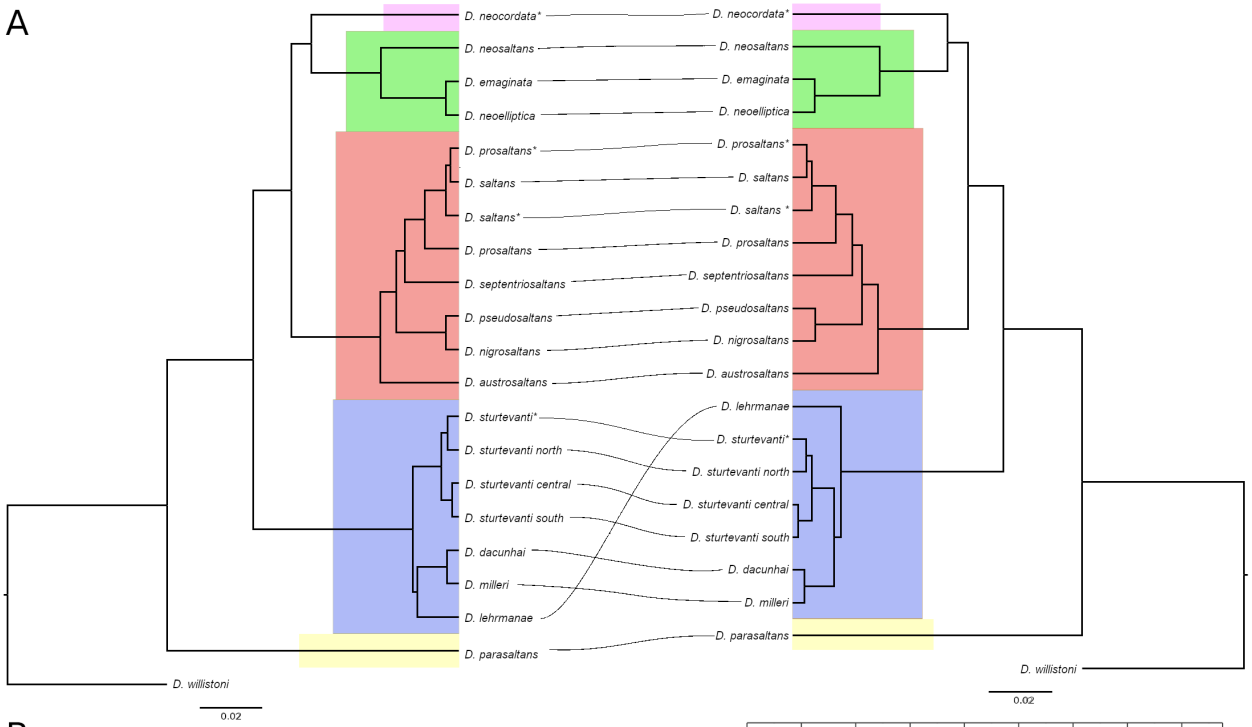
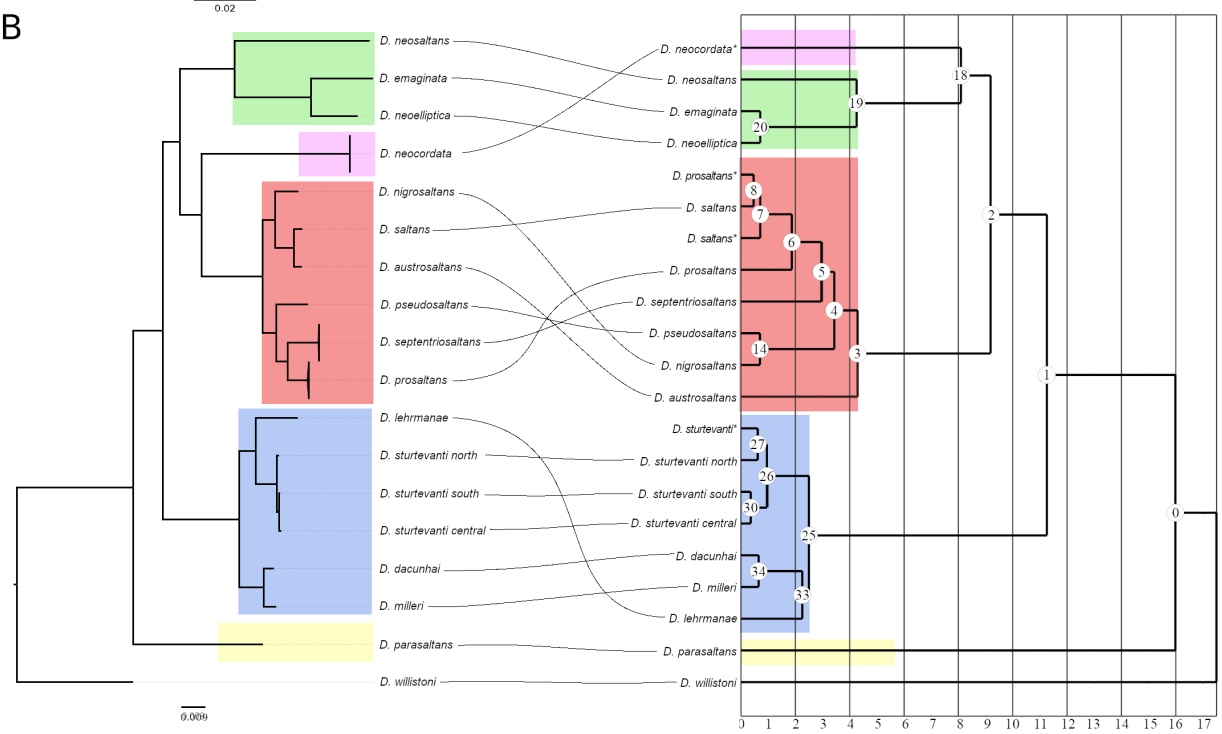

A
B A B A
A B B A
B B A A
sp3 sp2 sp1 O
B
Topology Frequence
(i) complete reticulation
(ii) incomplete reticulation
(iii) incomplete bifurcation
(iv) complete bifurcation
Reticulation
Bifurcation

Figure 2

A


D. neocordata*
D. neosaltans
D. emaginata
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. emaginata
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
0.069
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

1039 Figure 3
1040

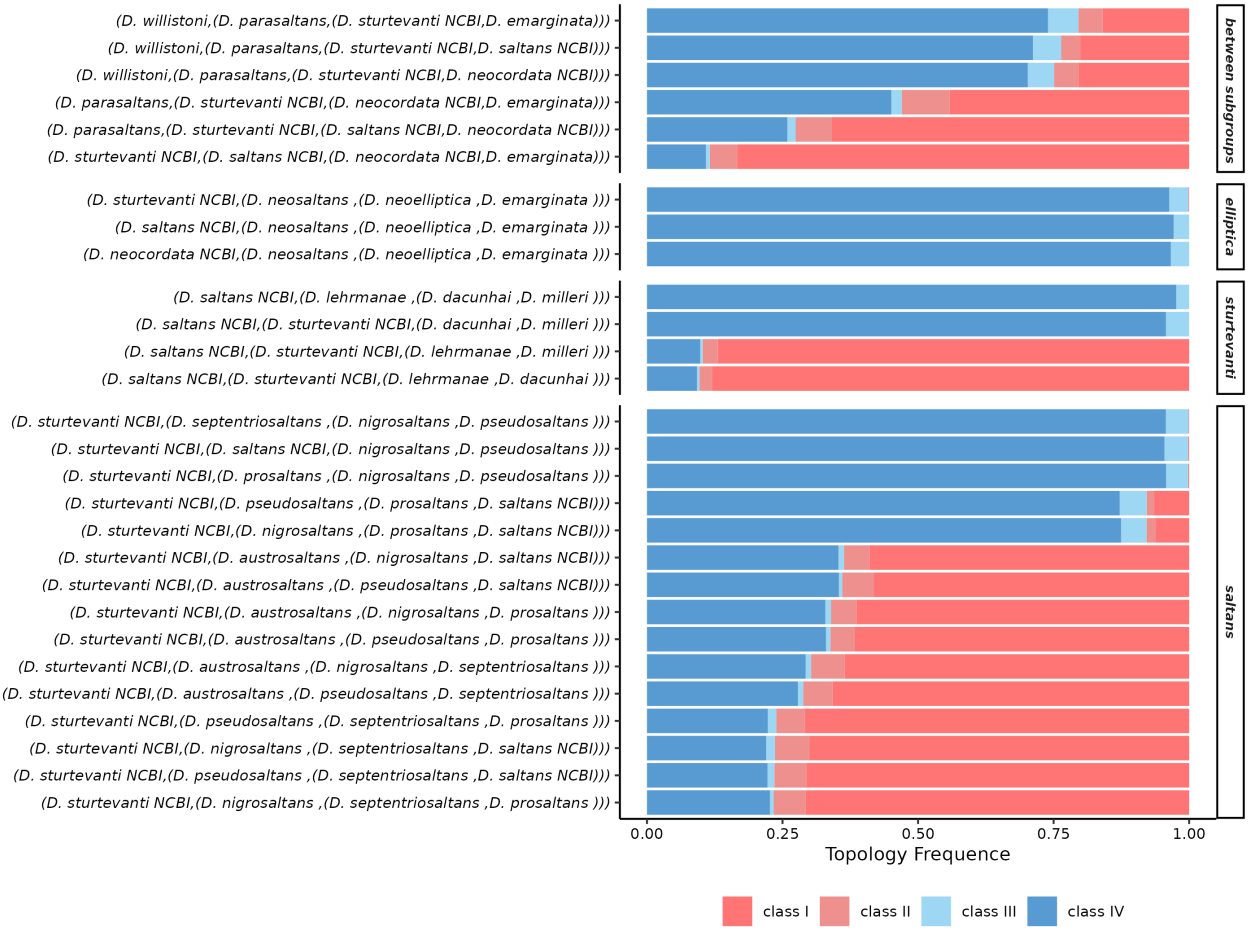
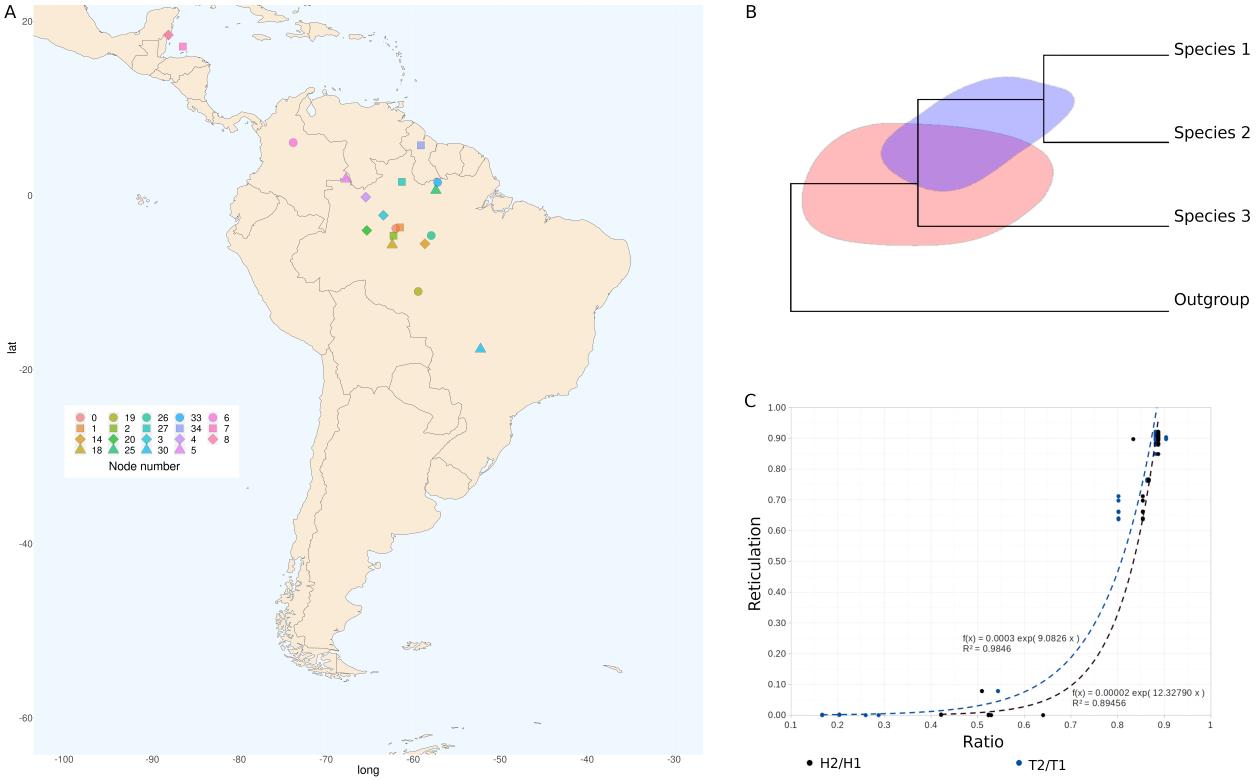


Figure 4


A
lat
long
Node number
0 1 14 18 19 2 20 25 26 27 3 30 33 34 4 5 6 7 8
B
Species 1
Species 2
Species 3
Outgroup
C
Reticulation
Ratio
f(x) = 0.0003 exp(9.0826 x)
R² = 0.9846
f(x) = 0.00002 exp(12.32790 x)
R² = 0.89456
H2/H1
T2/T1