Evolutionary persistence of a highly prevalent multicopy mitochondrial-derived nuclear insertion (Mega-NUMT) in Neotropical *Drosophila* flies

Authors:

Merce Montoliu-Nerin[1,4], Anton Strunov[2,5], Eleanor Heyworth[1], Daniela I. Schneider[2], Julia Thoma[2,6], Aurélie Hua-Van[3], Cécile Courret[3,7], Lisa Klasson[1], Wolfgang J. Miller[2]

Corresponding authors:

Lisa Klasson

Wolfgang J. Miller

Affiliations:

1. Molecular Evolution, Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
2. Lab Genome Dynamics, Department Cell & Developmental Biology, Center for Anatomy and Cell Biology, Medical University of Vienna, Vienna, Austria
3. Laboratoire Évolution, Génomes, Comportement et Écologie, CNRS, IRD, Université Paris-Saclay, Gif-sur-Yvette, France
4. Tree of Life Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom
5. Centre for Microbiology and Environmental Systems Science, University of Vienna, Djerassiplatz 1, 1030, Vienna, Austria
6. Department of Biosciences, Faculty of Medicine/Dentistry, Danube Private University, Förthofstraße 4, 3500 Krems, Austria
7. Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, 91198, Gif-sur-Yvette, France.

Merce Montoliu-Nerin: ORCID: 0000-0002-5200-0411
Anton Strunov: ORCID: 0000-0002-3574-6481
Eleanor Heyworth: ORCID: 0000-0001-8184-9556
Daniela I. Schneider: ORCID: 0000-0002-7740-8827
Julia Thoma: ORCID: 0000-0002-4222-7380
Aurélie Hua-Van: ORCID: 0000-0001-7859-2788
Cécile Courret: ORCID: 0000-0001-5849-8014
Lisa Klasson ORCID: 0000-0002-5874-7153
Wolfgang Miller: ORCID: 0000-0002-5044-6276

Abstract

Background: Although strict maternal transmission of mitochondria is a general feature of animals and humans for ensuring homogeneity in mitochondrial DNA (mtDNA) across generations, exceptions were reported in the recent past. For example, some extremely rare but spectacular cases of heteroplasmy and paternal transmission in humans have questioned the universal evolutionary principle. Hence, as an alternative, the Mega-NUMT concept was coined to explain this discovery and was thereafter partly proven to exist. This concept expands on the quite common transfer of mtDNA fragments to the nucleus (NUMTs) by considering the existence of multicopy mitochondrial nuclear insertions. Mega-NUMT reports are currently restricted to a few cases in animals, including humans. However, even in humans, their detailed genomic organization, natural prevalence, and potential biological functions remain unclear.

Methodology/Principal Findings: Here, we discovered that up to 60 full-sized mitochondrial genomes are integrated into the nuclear genome of the neotropical fruit fly *Drosophila paulistorum* using long-read sequencing and confirmed their presence by *in situ* hybridization. The copies are organized in one cluster on chromosome 3, which we, due to its similarity with the Mega-NUMT concept, designated the "*Dpau* Mega-NUMT". Contrary to the rarity in humans, this Mega-NUMT is found at high prevalence (40%) in both long-term laboratory lines and natural *D. paulistorum* populations of different semispecies. Additionally, the mitochondrial copies in the Mega-NUMT cluster are phylogenetically separated from the current mitotypes of *D. paulistorum*. Together, these observations suggest long-term maintenance of the Mega-NUMT in nature. Hence, we propose that the *Dpau* Mega-NUMT may have been transferred to the nuclear genome before *D. paulistorum* semispecies radiation and maintained at relatively high prevalence in nature by balancing selection due to yet undetermined functions.

Conclusions/Significance: To our knowledge, this is the first verified existence and detailed dissection of a Mega-NUMT outside cats and humans. We show that Mega-NUMTs can be persistent in nature, even at high prevalence, potentially due to balancing selection. Our findings strengthen the importance of high-quality long-read sequencing technologies for deciphering complex repeat-rich genomic regions to deepen our understanding of the dynamics of genome evolution within genomic "dark matter".

Introduction

Mitochondrial DNA (mtDNA) has been widely used as a genetic marker for molecular systematics of animals for more than 30 years (Harrison, 1989). The high copy number of mtDNA per cell, as well as its rapid mutation rate, makes it particularly informative for investigating phylogenetic relationships and evolutionary history (Brown et al., 1979). Mitochondrial markers have been extensively utilized to resolve taxonomic uncertainties and to identify cryptic species (Hebert et al., 2003; Elías-Gutiérrez et al., 2008). Additionally, the use of mitochondrial data allows the reconstruction of phylogeographic patterns, providing insights into historical population dynamics and speciation events (Avise et al., 1983; Baião et al., 2023).

Studies from the pre-genomic era using Southern blots and PCR have revealed the existence of nuclear mitochondrial DNA segments (NUMTs), described as fragments of mtDNA that have become integrated into the nuclear genome (Lopez et al., 1994; Hazkani-Covo & Graur, 2007). Although the mechanisms of integration are not yet fully described, a common explanation outlines the insertion of mitochondrial sequences during the repair of double-strand chromosomal breaks (Blanchard & Schmidt, 1995, 1996; Bensasson et al., 2001). NUMTs have primarily been found in genomic regions with high levels of repetitive elements (Behura, 2007; Tsuji et al., 2012; Dayama et al., 2014; Schiavo et al., 2017; Wang et al., 2020) and distinctive GC content (Tsuji et al., 2012; Dayama et al., 2014; Schiavo et al., 2017). After their integration, NUMTs often degrade quickly, as the different genetic codes in the nuclear and mitochondrial genomes prevent proper translation of the genes (Bensasson et al., 2001). Nonetheless, some functions associated with NUMTs have been found, including the formation of new exons (Noutsos et al., 2007) and regulatory elements (Vendrami et al., 2022).

The presence of NUMTs can lead to misinterpretations when studying mitochondrial genomes (Calvignac et al., 2011), as it can result in erroneous inferences of relationships between lineages (Kress et al., 2015; DeSalle & Goldstein, 2019). Furthermore, the presence of NUMTs may lead to wrongful inference of heteroplasmy (Parr et al., 2006; Balciuniene & Balciunas, 2019), which is the coexistence of more than one mtDNA sequence variant within an individual. The misinterpretations arise from the difficulty in distinguishing NUMTs from real heteroplasmic variants when using techniques such as PCR. However, the high copy number of mtDNA compared to nuclear loci containing NUMTs may commonly suffice to mask the NUMT signal. Even so, one must be cautious (Parfait et al., 1998), as several studies have reviewed investigations in which high copy numbers of NUMTs in the genome had been interpreted as heteroplasmy (Hirano et al.,

101 1997; Salas et al., 2005; Yao et al., 2006; Lutz-Bonengel et al., 2021). To describe such multicopy
102 NUMTs, the term Mega-NUMT was coined as a theoretical concept in response to a claim of
103 paternal transmission of mitochondria in seemingly heteroplasmic humans (Balciuniene &
104 Balciunas, 2019; Lutz-Bonengel et al., 2021). However, large multicopy NUMTs, covering
105 approximately half of the mt-genome and with an estimated 38-76 copies, had also previously been
106 found in the domestic cat and its close relatives (Lopez et al., 1994) and later also identified in
107 several cat sequences using long-read sequencing, albeit in somewhat fewer copies (Patterson et
108 al., 2023).

109 In this study, we focus on the Neotropical fly *Drosophila paulistorum* spp., commonly used
110 as a study system for the evolutionary process of sympatric speciation (Burla et al., 1949;
111 Dobzhansky & Spassky, 1959; Dobzhansky et al., 1964). The *D. paulistorum* superspecies belongs
112 to the D. willistoni group and is composed of six semispecies, named Amazonian (AM), Andean
113 Brazilian (AB), Centro-American (CA), Interior (IN), Orinocan (OR), and Transitional (TR),
114 presenting different levels of pre- and post-mating incompatibilities (Dobzhansky & Spassky,
115 1959). Phylogenetic incongruences between the nuclear and mitochondrial DNA of *D. paulistorum*
116 spp. have been described (Gleason et al., 1998; Robe et al., 2010; Baião et al., 2023), suggesting
117 past introgression events. Additionally, a recent phylogenomic analysis of *D. paulistorum* spp.
118 revealed that their mitochondrial genomes are polyphyletic and split into two major clades, α and
119 β (Baião et al., 2023).

120 Here, we identify a *D. paulistorum* line from the OR semispecies, O11, with an apparently
121 fixed presence of the two different mitotypes, α and β. However, our analysis revealed that in
122 addition to the presence of a β mitochondrion, O11 has a NUMT consisting of around 60 copies of
123 nearly complete mt-genomes with high similarity to the α mitotype on chromosome 3. We
124 designate this nuclear insertion as the *Dpau* Mega-NUMT. Furthermore, we found that this Mega-
125 NUMT is present in at least one line from three of the six described *D. paulistorum* semispecies
126 and has a relatively high prevalence of around 40% in fly populations from French Guiana. Taken
127 together, these data suggest that the *Dpau* Mega-NUMT was present before or just after the split
128 of the semispecies and may be maintained in nature by balancing selection.

Results

The two mitotypes, α and β, are fixed in individual flies of the *D. paulistorum* line O11

We sequenced PCR products of the mitochondrial *Cytochrome Oxidase I* (*COI*) gene from several *D. paulistorum* lines for mt-barcoding. We found multiple robust double peaks in the resulting chromatograms derived from single flies of the OR line O11, suggesting the presence of both α and β mitotypes within the same individual. The presence of a *Hin*dIII restriction site for the α-mitotype but not ß (**Fig. 1A**) was used to perform diagnostic RFLP/*Hin*dIII Southern blots on DNA from single O11 flies (Junakovic, 2004) probed with the *COI* cloned plasmid (**Fig. 1B** and **Table S1, Fig. S1A**). We observed that the O11 line again showed a clear heteroplasmic pattern, with both α and β mitotypes represented in individual flies at approximately similar abundances, suggesting almost equal copy numbers (**Fig. 1B**). In contrast, single flies of line C2 belonging to the CA semispecies and line A28 from the AM semispecies were homoplasmic and carried either the α or the β mitotype, respectively (**Fig. 1B**). Furthermore, we performed *CO1*-specific RFLP/*Hin*dIII-PCR on multiple single females and males (**Table S1, Fig. S1**) which showed a consistency on the presence of both α and β mitotypes in all tested O11 flies (**Fig. 1C**), which suggested fixation in this long-term lab line collected in the 1960s.

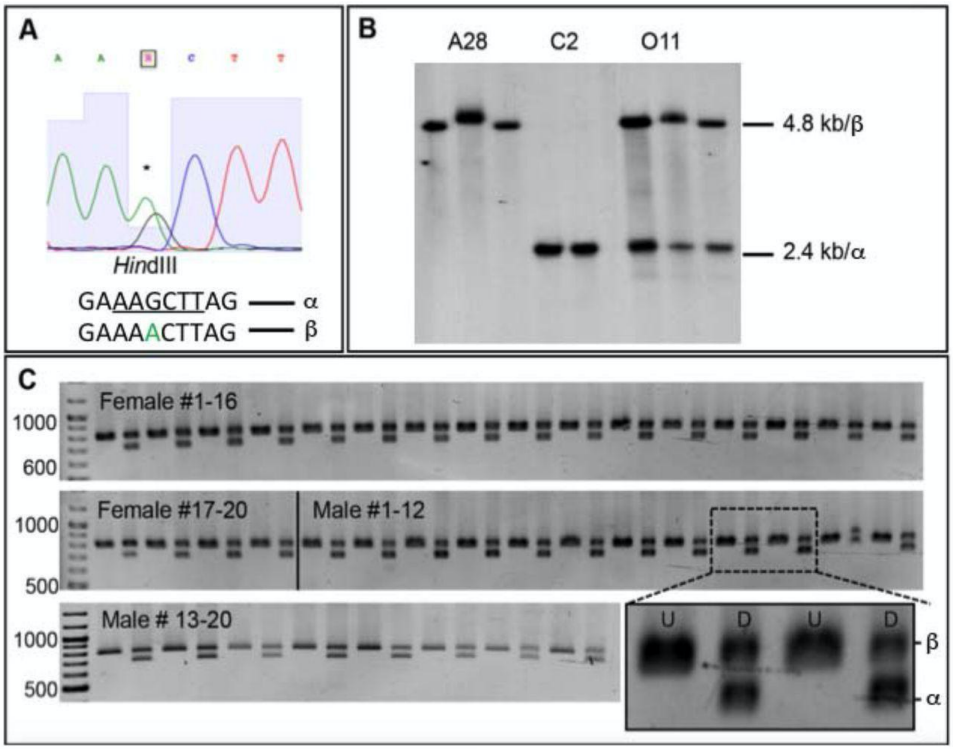


Figure 1. Presence of two mitotypes in the *D. paulistorum* Orinocan line O11. (**A**) Chromatogram derived from *COI* direct Sanger sequencing of one O11 female covering the α-diagnostic *Hin*dIII restriction

site (underlined). The asterisk highlights the double peak with Guanine (G) for the α- and Adenine (A) for the β-mitotype. (**B**) Single-fly *CO1* Southern blot digested with *Hin*dIII and probed with cloned *CO1*. While the Amazonian females of the A28 line (n=3) present only the β-mitotype (4.8kb) and the females of the Centro-American line C2 (n=2) only the α-mitotype (2.4kb), all randomly picked Orinocan O11 females (n=3) present both the α- and β-mitotypes (4.8 and 2.4kb). (**C**) α-diagnostic *CO1* RFLP/*Hin*dIII-PCR screen (**Fig. S1**) on O11 single flies (n=20 for each sex). Undigested (U) and digested (D) CO1-PCR fragments per fly are shown exemplary for two individual specimens (magnification at the bottom right in C). All tested single flies (40/40) exhibited clear double bands after *Hin*dIII digestion.

Subsequently, we measured the copy numbers of both α and β mitochondrial (mt) genomes, relative to the nuclear genome, using discriminant qPCR on the *CO3* gene (see Materials and Methods). The quantification of both the α and β mt genomes was in multiple tissues of both females and males and resulted in similar average values of 27.2±13.9 α and 31.6±22.7 β mt genome copies per nuclear genome (**Fig. 2**). The β mt-genome copy numbers were strongly affected by tissue (**Table S2**; F-value=27.694, p<0.001, F-test, **Fig. 2A**, **Table S2**) and a more subtly, yet significantly, by age and sex combined (**Table S2**; F-value=6.66, p<0.05, F-test), and by sex alone (**Table S2**; F-value=4.714, p<0.05, F-test), as expected based on mitochondrial function. In contrast, the number of copies of the α mt-genome was not significantly affected by any of the tested parameters (**Fig. 2B**, **Table S2**), suggesting that the α mitotype is not a typical mitochondrion.

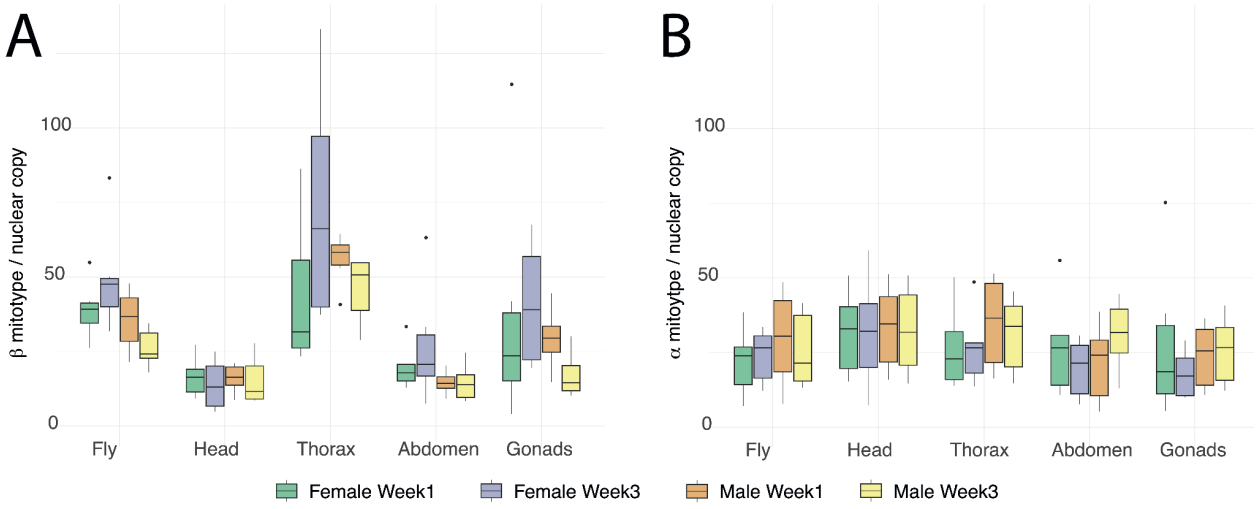


Figure 2. **qPCR quantification of α and β mitotypes in *D. paulistorum* O11 flies**. 120 flies were analyzed to assess the number of α and β mitotype copies relative to the nuclear genome copies using discriminant qPCR. Boxplots show the distribution of values, mean, and standard deviation for the (**A**) β mitotype and the (**B**) α mitotype.

182 **Discovery of a Mega-NUMT with long reads**

183 Given the seemingly fixed high abundance of two mitotypes in one individuum and unresponsive
184 behaviour of the α mitotype in different tissues of the *D. paulistorum* O11 line, we sequenced the
185 O11 genome using Oxford Nanopore technology (ONT). The assembly of the nuclear genome of
186 O11 was 231 Mb (111 contigs, N50 = 18.6 Mb) (**Table S3**), with a BUSCO completeness of 99%
187 (diptera_odb, **Table S4**). Using the mitochondrial genome from the O11 line (β mitotype, Baião et
188 al., 2023), we searched the nuclear assembly and found 12 mt-genome copies clustered at the end
189 of one contig (ctg000940). From here on, we refer to this cluster as the *Dpau* Mega-NUMT, given
190 its similarity to the theoretical Mega-NUMT concept. Further analysis of the 12 NUMT copies
191 revealed a conserved synteny of the mitochondrial genes across most copies (**Fig. 3**). Four of the
192 NUMT copies have lost genes compared to the mt-genome, and in all NUMT copies, many of the
193 genes have suffered frameshift mutations (**Fig. 3**, **Table S5**). When comparing the gene length to
194 the fraction of intact gene copies per protein-coding gene, we found a significant inverse
195 correlation, suggesting that the pseudogenization process is random (**Fig. S2**). However, given the
196 complexity of the Mega-NUMT, it is possible that the sequence of each gene copy in the assembly
197 is not 100% accurate. Nevertheless, at least one copy of each of the 13 mitochondrial protein-
198 coding genes is intact in the Mega-NUMT, albeit with the mitochondrial genetic code (**Table S5**).

199 A phylogeny based on the complete set of genes and pseudogenes in the *Dpau* Mega-
200 NUMT, together with the published mitochondrial genomes of *D. paulistorum* spp. and other
201 willistoni group species (Baião et al., 2023), shows that all individual NUMT copies in the Mega-
202 NUMT are found in one clade, which is sister to the α mitotype of the *D. paulistorum* C2 and MS
203 lines (**Fig. 3**) belonging to the CA and AB semispecies, respectively (**Table 1, Table S6**). Hence,
204 we refer to these NUMT copies as α-NUMTs. Importantly, though, the separation of the α mitotype
205 and α-NUMTs into different clades, and the monophyly of within these clades, suggests that none
206 of the α-NUMT copies in the *Dpau* Mega-NUMT are recent insertions from an extant α mt-
207 genome. This is further supported by the fact that the O11 line carries the β mitotype, as does the
208 majority of *D. paulistorum* semispecies and lines.

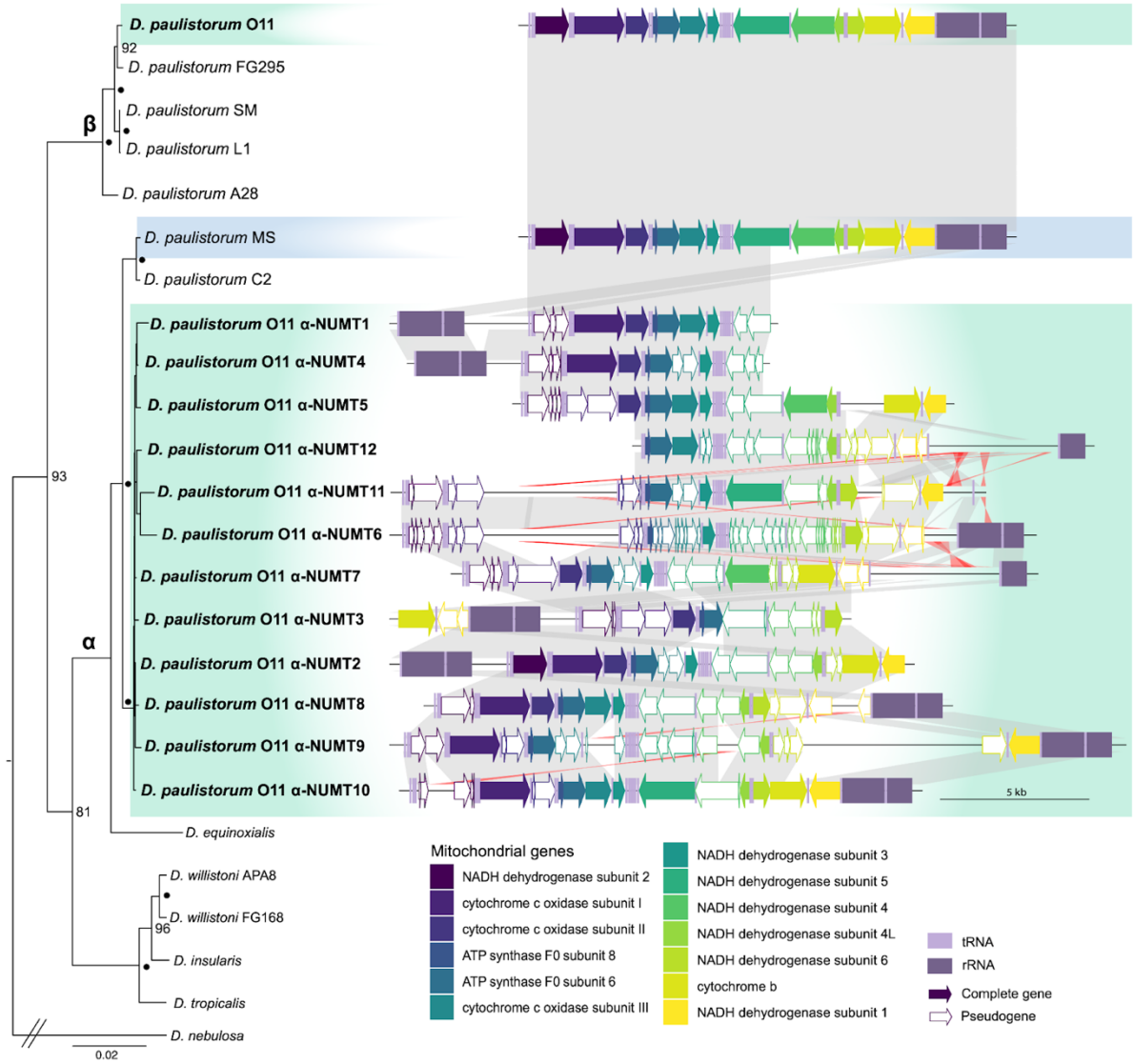
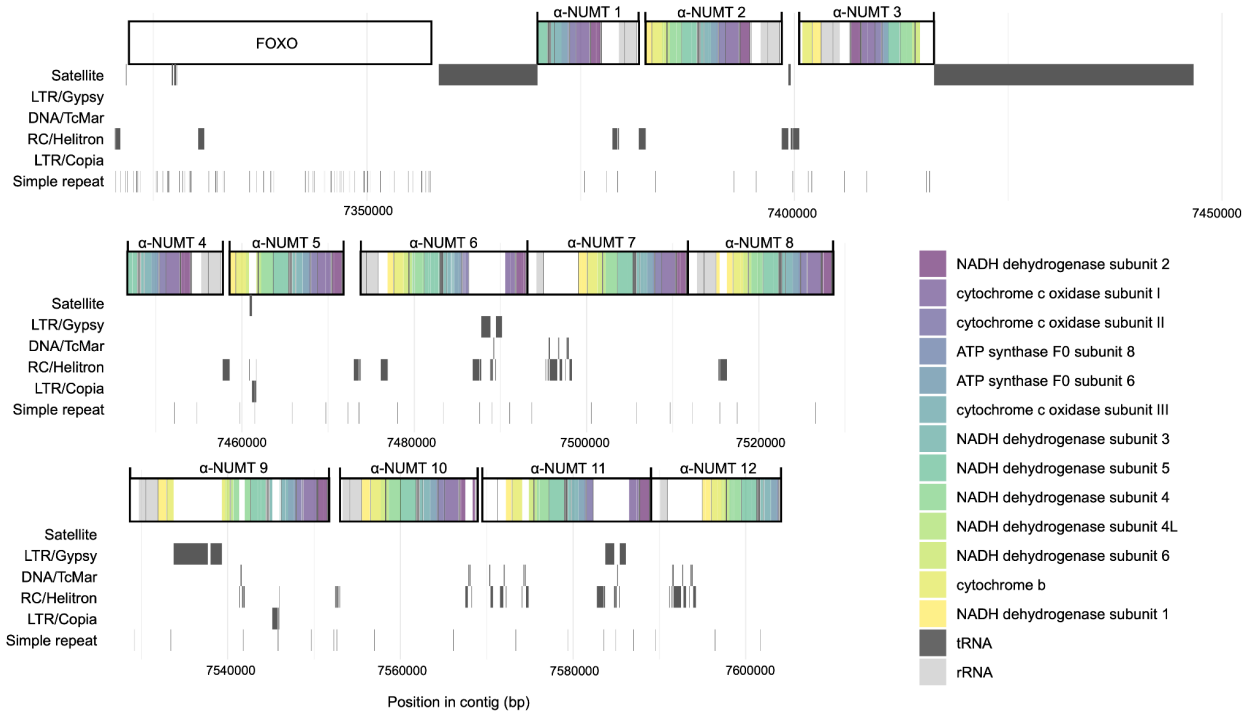


Figure 3. Phylogenetic position and synteny of NUMT copies of *D. paulistorum* O11 and closely related mitochondrial genomes. Maximum likelihood phylogeny based on the complete set of genes and pseudogenes of the NUMT copies found in the Mega-NUMT of *D. paulistorum* O11, together with the mitochondrial genes in *D. paulistorum* and closely related species. Black dots indicate nodes with 100 bootstrap supports. The α and β mitotype clades are annotated on the tree. Mitochondrial genomes of *D. paulistorum* O11 (green) and *D. paulistorum* MS (blue) were used as references for the β and α mitotypes. The grey ribbons between sequences represent homologous regions of the same direction, while red ribbons represent inversions. The 13 mitochondrial protein-coding genes are shown as colored arrows, and rRNA and tRNA genes as blocks. Intact genes are filled, and pseudogenes only have a colored outline.

Further, we analyzed the genomic region where the *Dpau* Mega-NUMT is located. The Mega-NUMT is directly flanked by an AT-rich 36 bp satellite sequence (**Fig. 4** and **Fig. S3**) and the gene

221 *FOXO* encoding the Forkhead box O transcription factor (Jünger et al., 2003), located in Muller
222 element E on chromosome 3, as deduced by comparison to *D. melanogaster*. The same 36 bp
223 satellite DNA is also found inside the Mega-NUMT, downstream of α-NUMT 3 (**Fig. 4** and **Fig.**
224 **S3**). The positions of *FOXO* and the Mega-NUMT next to each other were also confirmed by DNA-
225 hybridizations on polytene chromosomes of larval salivary glands of O11, and the bright
226 polytenized signal of the Mega-NUMT (**Fig. S4**) implies its euchromatic state since it is
227 polytenized and in close genomic vicinity to the essential host gene *FOXO*.



228

229 **Figure 4. Graphic representation of the end of contig000940 in the assembly of *D. paulistorum* O11.**
230 The region is divided into three rows to facilitate visualization. The position of the nuclear gene *FOXO* and
231 each α-NUMT is represented with a white box, which also contains colored lines that represent the position
232 of each protein-coding and RNA gene. The positions of the satellite and repeat elements are shown below
233 in the order: The AT-rich 36-bp satellite, *Gypsy*, *TcMariner*, *Helitron*, *Copia*-like transposons, and simple
234 repeats.

235 All α-NUMT copies except α-NUMT 3 have the same orientation in the *Dpau* Mega-NUMT (**Fig.**
236 **4**). The sequence of α-NUMT 3 also contains a rearrangement not seen in any of the other copies
237 of the contig (**Fig. 3** and **4**). Additionally, the 12 α-NUMT copies in the O11 nuclear genome
238 assembly are surrounded by repeat elements, and in some cases, insertions of transposable elements
239 are present within an α-NUMT sequence (**Fig. 4**). Specifically, we found remnants of the LTR

240 retrotransposon *Gypsy* within the α-NUMT copies 6, 9, and 11 and of the LTR element *Copia* in
241 α-NUMT copies 5 and 9 (**Fig. 4**). In addition, we found short regions with sequence similarity to
242 the DNA transposon Tc Mariner in six of the α-NUMTs and to the rolling-circle DNA transposable
243 element, *Helitron*, in nine of the α-NUMTs, as well as between α-NUMT copies when a non-
244 NUMT sequence was found between them (**Fig. 4**).

245 **Number of α-NUMT copies in the genome**

246 Since the discriminant qPCR results indicated that more than 12 α-NUMT copies could be present,
247 we investigated the coverage of ONT reads in the Mega-NUMT region of O11 to assess the true
248 number of α-NUMT copies. When mapping the reads to the assembly, we also included the β mt-
249 genome from Baião et al., (2023) and the genome of the bacterial symbiont *Wolbachia* of O11 from
250 Papachristos et al., (2025) to prevent mapping of reads from the real mitochondrion or *Wolbachia*
251 to the Mega-NUMT. Additionally, to be as stringent as possible in our assessment, only one
252 mapping per read was allowed. The read coverage over the *Dpau* Mega-NUMT shows a clear
253 discordance with the average coverage across the genome (**Fig. 5**). The *Dpau* Mega-NUMT has an
254 average coverage of 402±159, while the genome assembly as a whole has an average coverage of
255 80.8±77.4 (**Fig. 5**). Using these numbers, we thus infer that the genome contains around 60 whole
256 mt-genome α-NUMTs (**Fig. 5**). We note that the coverage over the different α-NUMT copies varies
257 substantially. Especially high coverage was found on α-NUMT 8, with 17 times higher coverage
258 than the genome average, followed by α-NUMTs 7, 9, and 10, with ten, seven-, and six-times
259 higher coverage than the genome average, respectively. These results indicate that the unassembled
260 α-NUMT copies resemble α-NUMT copies 7-10 the most. Specifically, we note that both α-NUMT
261 7 and 8 contain the full mt-genome and have no intervening sequence, suggesting that many of the
262 unresolved copies are also full mt-genomes with no intervening sequences (**Figs. 4** and **5**).

263 We note that the 60 α-NUMT copies estimated from long read mapping are more than the
264 ca. 30 we estimated by our qPCR results. As α-NUMT 3 lacks *CO3* completely, only 11 of the 12
265 copies in our assembly have a binding site for the primer CO3_SNP_1 used in qPCR for
266 determining the copy numbers (**Fig. 2**, **Table S1**). Hence, although we do not know exactly what
267 the remaining copies look like, this indicates that the discrepancy in total copy numbers of the α-
268 NUMT between the coverage and qPCR results might be explained by a lack of primer binding in
269 some of the α-NUMT copies.

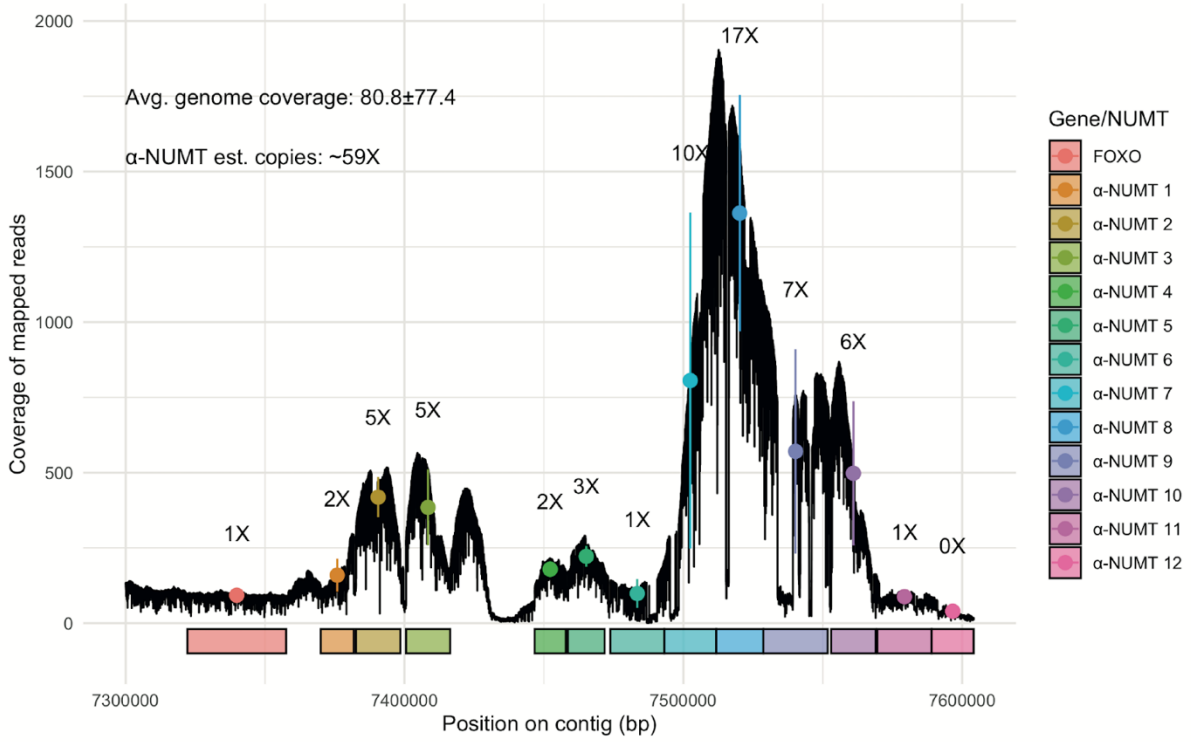


Figure 5. Coverage quantification of α-NUMTs in the nuclear genome of *D. paulistorum* O11. The x-axis represents the end region of contig000940 of the genome assembly of *D. paulistorum* O11, which includes the nuclear gene *FOXO*, and the 12 assembled α-NUMTs. Coverage was obtained from mapping Nanopore reads to the genome assembly. The average and standard deviation of the coverage over each α-NUMT and the *FOXO* gene are plotted in the same color as used for the box depicting each feature. The total number of α-NUMT copies in the Mega-NUMT is estimated from the average over all the assembled α-NUMT copies in relation to the average coverage of the whole genome assembly.

Mega-NUMT visualization on *D. paulistorum* chromosome.

To validate the genomic findings of the Mega-NUMT localization on the 3rd chromosome of *D. paulistorum* O11, we performed DNA fluorescent *in situ* hybridization (FISH) on mitotic chromosomes of larval brains using mitochondria-specific probes and a probe for the Mega-NUMT-linked satellite (see **Fig. 4** and the Materials and Methods section). In O11, we observed two satellite signals (red) on the 3rd chromosome—one in the (peri)centromeric region and a second in the subtelomeric region—as well as a Mega-NUMT signal (green) adjacent to the subtelomeric satellite signal (**Fig. 6A**), in agreement with the assembly data (**Fig. 4**). In contrast, the *D. paulistorum* line FG295 (**Table S6**), which also belongs to the OR semispecies (**Table 1**), shows only a single satellite signal in the (peri)centromeric region of the third chromosome, but no Mega-NUMT signal (**Figure 6B**).

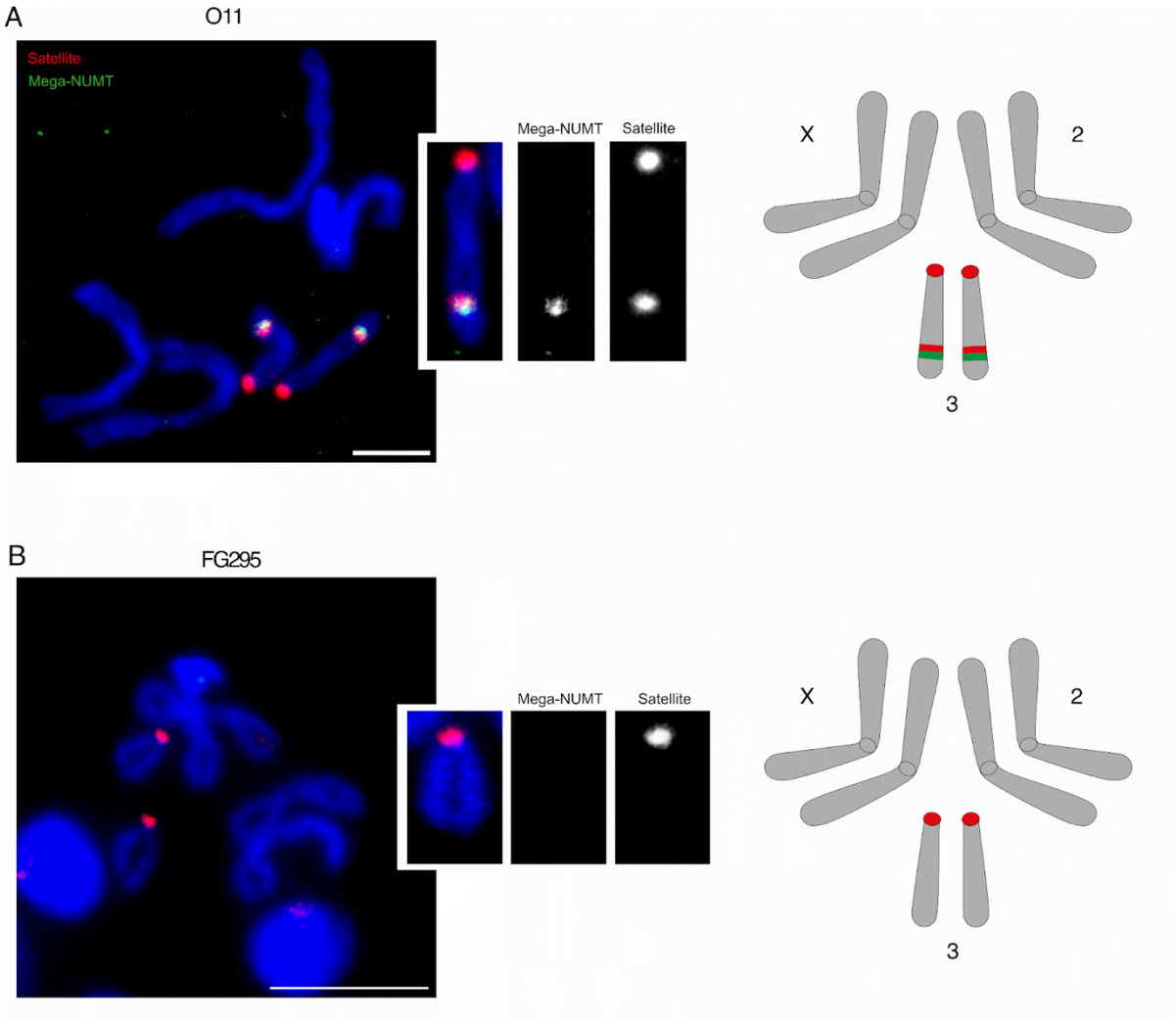


Figure 6. DNA-FISH on mitotic chromosomes from larval brains of the *D. paulistorum* O11 and FG295 lines. Probes targeting the Mega-NUMT (green) and the Mega-NUMT–linked satellite (red) were used. The scale bar represents 5 µm. The inset highlights the third chromosome, showing two satellite signals and one Mega-NUMT signal in O11 (**A**), but only a single satellite signal in FG295 (**B**). The schematic on the right summarizes the localization of the satellite (red) and the Mega-NUMT (green) in each strain.

Additionally, to follow Mega-NUMT dynamics in interphase and dividing cells, we investigated the localization of the *Dpau* Mega-NUMT in early embryonic mitosis by mt-DNA FISH (**Fig. S5**). Contrary to larval brains that were generated by squash fixation and thereby washing off the cytoplasm (**Fig. 6**), in our whole mount confocal mt-DNA-FISH assays on O11 embryos, we observed staining of mitochondria in the cytoplasm and the chromosome-associated Mega-NUMT signal as two dots, with some cases of a fused, larger dot (arrowheads and arrow, respectively, **Fig. S5A-D, F**). Each dot consists of two parts with a small bridge in between,

303 suggesting somatic pairing (inset of **Fig. S5B**). Based on our results on metaphase chromosomes
304 from 3rd instar larvae (**Fig. 6**), we conclude that each dot represents a Mega-NUMT signal located
305 on one of the duplicated arms of chromosome 3, with two smaller parts representing sister
306 chromatids. At the anaphase stage, we could clearly observe the separation of sister chromatid
307 signals, creating a 4-dot pattern (**Fig. S5E**). Altogether, systematic DNA-FISH at different stages
308 of early embryo mitosis reconfirmed the presence of only one Mega-NUMT signal on chromosome
309 3 of the O11 line.
310

311 **Prevalence of the Mega-NUMT across *D. paulistorum* semispecies and populations**

312 To test if the *Dpau* Mega-NUMT is restricted to only the O11 line, we analyzed α-NUMT
313 prevalence across different *D. paulistorum* lines. We used RFLP-PCR from adult flies to assess the
314 prevalence and abundance of α-NUMTs (**Table 1**). In addition to O11, another OR line, FG111,
315 presented α-NUMT, but it appeared not to be fixed in the population, as it was only found in 10
316 out of 40 flies that originated from an isofemale (**Table 1**). In yet another OR line, FG103, the
317 Mega-NUMT was lost between the first test (2014-15) and the last one (2018) (**Table 1**). Two AM
318 lines, FG16 and FG572, also presented a non-fixed Mega-NUMT (**Table 1**), while two lines from
319 the AB semispecies, MS and Yellow, both carried the Mega-NUMT in all tested flies (**Table 1**).
320 To sum up, we found the Mega-NUMT in several OR lines, as well as in the AM and AB
321 semispecies, but not in the few representatives of the other three semispecies of *D. paulistorum*
322 (CA, IN and TR). Possibly, this is caused by a sampling bias, since none of these three semispecies
323 were found in our collections from French Guiana between 2014 and 2018 (**Table 1** and **Table**
324 **S7**). In addition to O11, we only found fixation of the *Dpau* Mega-NUMT in the old long-term
325 stocks of MS and Yellow that were maintained at small population sizes in the laboratory since the
326 1960s or earlier. Our finding that the Mega-NUMT is not fixed in isofemale lines from our recent
327 collections in French Guiana suggests that it is frequently hemizygous in wild populations.

328 **Table 1. α-NUMT abundance and prevalence in *D. paulistorum*** lines based on *CO1* RFLP-PCR from adult flies.
329 Abundance is given in percentage of α-NUMTs compared to the mitochondrial copies per fly (**Fig S1**) and is presented
330 separately for females (left) and males (right). Prevalence of the Mega-NUMT among tested individuals is presented
331 both as a percentage and in absolute numbers (in parentheses). Abbreviations: Spp: Semispecies, AB: Andean-
332 Brazilian, AM: Amazonian, CA: Centro-American, IN: Interior, OR: Orinocan, TR: Transitional, Coll: Collection
333 year, ND: Not determined. Further details about the fly lines are provided in **Table S6**.

Line	Coll.	Spp.	Mitotype	α-NUMT detected	α-NUMT abundance (%)[#]	α-NUMT prevalence
O11	1957	OR	ß	Yes	30.7±3.7/36.8±3.6	100% (39/39)
FG103	2014	OR	ß	Yes	ND[§]	0% (0/40)[§]
FG111	2014	OR	ß	Yes	13.4±1.1/18.3±2.9	25% (10/40)
FG295	2014	OR	ß	-	0	0% (0/40)
TP37	2009	OR	ß	-	0	0% (0/20)
POA1	2003	OR	ß	-	0	0% (0/20)
RP	1995	OR	ß	-	0	0% (0/20)
A28	1952	AM	ß	-	0	0% (0/40)
FG16	2018	AM	ß	Yes	57.2±8.6/66.5±7.2	55% (22/40)
FG18	2018	AM	ß	-	0	0% (0/40)
FG572	2015	AM	ß	Yes	45.3±8.1/63.7±7.0	90% (36/40)
MS[*]	1962	AB	α	Yes	21.8±3.1/36.0±4.1	100% (40/40)
Yellow[*]	1960	AB	α	Yes	20.4±4.7/27.4±4.3	100% (40/40)
C2	1954	CA	α	-	0	0% (0/20)
White	1960	CA	α	-	0	0% (0/20)
L1	1958	IN	ß	-	0	0% (0/20)
SM	1956	TR	ß	-	0	0% (0/20)

[#]Abundance of the α-NUMT is given in percentage of females and males determined by *CO1* RFLP-PCR.
[§]The line FG103 presented α-NUMT in 2014/15 (determined by Sanger and Illumina sequencing) but has lost it afterwards (verified by *CO1* RFLP-PCR and Sanger sequencing in 2018/19).
[*]To distinguish αNUMTs from the α-mitotype of AB semispecies, the α-diagnostic *Eco*RV site of *CO1* was used.

To evaluate the prevalence of the *Dpau* Mega-NUMT in nature, we screened a total of 80 *D. paulistorum* spp. isofemale lines at generation F1 post-collection that were sampled between 2014 and 2024 from five different locations in French Guiana (**Fig. 7**; **Table S7**).

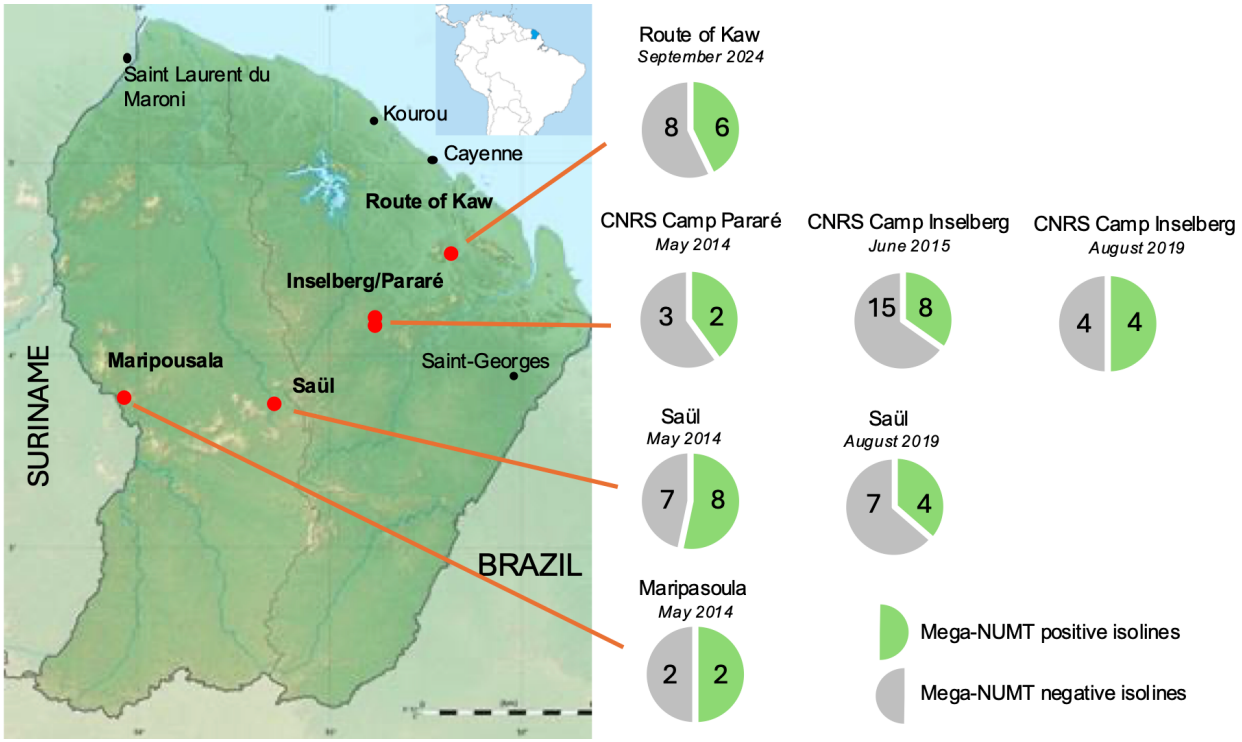


Fig. 7. Natural prevalence of Mega-NUMTs in *D. paulistorum* spp. populations in French Guiana. All 80 fly samples were collected between 2014 and 2024 at five different locations (red dots, and **Table S7**). The individual pie charts are in reference to locality and collection date and show the actual number of Mega-NUMT-positive (green) and negative (gray) specimens. Species designation and Mega-NUMT-typing per established *D. paulistorum* spp isoline (n=80) was performed on multi-fly DNA at generation F1 post-collection by *CO1* PCR and direct sequencing. The presence of robust double peaks in their chromatograms at α-NUMT-diagnostic positions was used in combination with cloning and sequencing.

Sampled flies were established as isofemale lines and barcoded at F1 post-collection as members of the *D. paulistorum* species complex by *CO1* PCR followed by direct Sanger sequencing and/or cloning of *CO1* (**Table S7**). Similar to O11 (**Fig 1A**), the presence of robust double peaks in the chromatograms from direct sequencing was used for Mega-NUMT typing, suggesting that 34 of the 80 isofemale lines harbor α-NUMTs (**Table S7**). The high prevalence of α-NUMT of 43% in our recent collections from *D. paulistorum* populations, together with its presence in at least three of the six semispecies (**Table 1**), strongly suggests their ancestral evolutionary state in this neotropical species complex and its stable maintenance both in nature and over long evolutionary time.

Discussion

In animal species, mtDNA is normally transmitted strictly uniparentally from mother to progeny (Ladoukakis & Zouros, 2017), with rare exceptions (Breton et al., 2007). However, reports of mitochondrial heteroplasmy and biparental transmission in humans challenged this view (Schwartz & Vissing, 2002; Luo et al., 2018) and prompted skepticism by the scientific community (McWilliams & Suomalainen, 2019; Salas et al., 2020). Consequently, the theoretical Mega-NUMT concept, proposing nuclear integration of multiple copies of whole mitochondrial genomes, was coined as an alternative explanation for such unorthodox findings (Balciuniene & Balciunas, 2019). Although this concept has been verified in humans (Wei et al., 2020; Lutz-Bonengel et al., 2021; Pagnamenta et al., 2021; Wei et al., 2022), similar phenomena have so far only been described in a limited number of animal systems where only partial regions of the mt-genome were found at multicopy (Lopez et al., 1994; Patterson et al., 2023; Biró et al., 2024), but not in Drosophila (Rogers & Griffiths-Jones, 2012), despite a relatively high prevalence of NUMTs in insects (Hebert et al., 2023). On the other hand, nuclear insertions of whole bacterial genomes larger than 1 Mb, like the α-proteobacterium *Wolbachia*, have been repeatedly detected in invertebrate genomes, including insects and nematodes (Kondo et al., 2002; Hotopp et al., 2007; Nikoh et al., 2008; Klasson et al., 2014; and recently reviewed in Keeling, 2024).

The Mega-NUMT of *D. paulistorum* spp.

In this study, we found that multiple whole mitochondrial genomes are integrated in the nuclear genome of the O11 line from the *D. paulistorum* species complex, designated α-NUMTs because of their close phylogenetic relationship with the cytoplasmic α-mitotype of the Centro-American and Andean-Brazilian semispecies of *D. paulistorum* (Baião et al., 2023). These α-NUMTs localize in a nuclear cluster in the Müller element E on chromosome 3 with up to 60 nearly complete 15kb mitochondrial genomes, potentially with a total size of 900 kb. Given the agreement between this finding and the Mega-NUMT concept, we named this cluster the "*Dpau* Mega-NUMT". Similar to earlier NUMT reports (Behura, 2007; Tsuji et al., 2012; Dayama et al., 2014; Schiavo et al., 2017; Wang et al., 2020; Biró et al., 2024), the *Dpau* Mega-NUMT of the O11 line is located in a genomic region containing other repetitive elements, *i.e.,* a chromosome 3-specific satellite and several different TEs (**Fig. 6A, B**), which suggests it might have served as a "safe haven" for such sequences (Werren, 2011). Such repetitive sequences may have enlarged the *Dpau* Mega-NUMT beyond 1 Mb in size, which is larger than the 641 kb NUMT recently discovered in the genome of

392 *Arabidopsis thaliana,* consisting of a partially duplicated large mitochondrial genome (Fields et
393 al., 2022) and currently one of the largest described fully sequenced NUMTs. Additionally,
394 contrary to currently described human Mega-NUMTs with prevalence below 0,5% (Wei et al.,
395 2020; Bai et al., 2021; Lutz-Bonengel et al., 2021), the *Dpau* Mega-NUMT is found at high
396 prevalence, around 40% in long-term laboratory lines (3/8; **Table 1**) and 42.5% (34/80) in *D.*
397 *paulistorum* spp. flies collected between 2014 and 2024 at different locations in French Guiana
398 (**Fig. 7**; **Table S7**). Furthermore, the *Dpau* Mega-NUMT is present in members of at least three of
399 the six semispecies of *D. paulistorum*, either in homo- or hemizygosity (**Table 1**). Finally, the α-
400 NUMTs form a separate monophyletic clade to the exclusion of the current α mitotype (**Fig. 3**),
401 indicating that the insertion did not involve any of the extant mitotypes but possibly a more
402 ancestral one. Therefore, we proposed that the *Dpau* Mega-NUMT might have been inserted in the
403 nuclear genome before or just after the semispecies radiation, estimated to have occurred between
404 1.2-2.1 MYs ago (Zanini et al., 2018; Kim et al., 2022).

405 **Integration and chromosomal position**

406 The overall well-conserved head-to-tail structure of the *D. paulistorum* α-NUMT cluster suggests
407 that the *Dpau* Mega-NUMT might have integrated as a concatemer of multiple mt-genomes. As
408 earlier proposed by Balciuniene & Balciunas, 2019, the mechanism of generating Mega-NUMTs
409 of complete mt-genomes may be explained by replication via a rolling circle intermediate. Such
410 mt-rolling circles have, for example, been reported in yeast, nematodes, and humans (Pohjoismäki
411 et al., 2009; Ling et al., 2016; Ling & Yoshida, 2020), but not in Drosophila until now. The
412 switching between conventional circular to rolling circle replication might be quite feasible because
413 key components of the mitochondrial transcription and replication apparatus are derived from the
414 T-odd lineage of bacteriophage rather than from an α-Proteobacterium, as the endosymbiont
415 hypothesis would predict (Filée et al., 2002; Shutt & Gray, 2006). In addition, rolling circle
416 replication (RCR) has been described in Drosophila for *histone* and the *Stellate* and *Suppressor of*
417 *Stellate (Su(Ste))* genes (Cohen et al., 2005; Cohen & Segal, 2009). Moreover, *Helitron* transposons
418 like *DINE-1* elements that can reach extremely high copy numbers in *D. melanogaster* (Thomas et
419 al., 2014) also propagate via RCR (reviewed in Barro-Trastoy & Köhler, 2024).

420 The nuclear integration site of the mt-concatemer in the chromosome 3 of *D. paulistorum*
421 spp. is interesting, since it is a fusion between Muller elements E and F, orthologous to
422 chromosome arm 3R and chromosome 4 of the melanogaster group, respectively (Muller, 1940;
423 Spassky & Dobzhansky, 1950). The location of the *Dpau* Mega-NUMT in the O11 line next to the

424 satellite DNA (**Fig 6A**), also present in the acrocenter of the fusion autosome E/F (**Fig. 6B**),
425 potentially suggests a chromosomal inversion event splitting the satellite in two, but not affecting
426 the upstream gene *FOXO*. Such paracentric inversions are quite commonly found in the *D.*
427 *paulistorum* species complex (Kastritsis, 1967) and can even play adaptive roles (Krimbas &
428 Powell, 1992), like for the successful recent colonization of urban environments (Valiati & Valente,
429 1997). In addition, inversions are enriched on chromosome 3 and often maintained in
430 heterozygosity, potentially under balancing selection (Gromko & Richmond, 1978). These earlier
431 findings suggest that chromosome 3 of *D. paulistorum* spp. may be an evolutionary hotspot for
432 chromosomal rearrangements like inversions and translocations, but also mitochondrial-derived
433 insertions like the *Dpau* Mega-NUMT as reported here.

434 **Why is the *Dpau* Mega-NUMT maintained at high prevalence and over evolutionary time?**
435 Contrary to currently described human Mega-NUMTs, which are extremely rare in prevalence, the
436 *Dpau* Mega-NUMT appears to be evolutionarily persistent since it is present in lines from at least
437 three different *D. paulistorum* semispecies, *i.e.*, Amazonian, Andean Brazilian and Orinocan.
438 Additionally, it is present at a relatively high prevalence of potentially more than 40% in natural
439 populations (**Figure 7**; **Table S7**). Even so, the fact that the Mega-NUMT is not fixed in *D.*
440 *paulistorum* except for old laboratory stocks (**Fig. 1**; **Table 1**), and that natural populations are
441 mainly hemizygous, suggests balancing selection rather than an essential function.

442 Although almost half of the mitochondrial genes in Mega-NUMT present in the 12 resolved
443 α-NUMT copies in our assembly (**Table S5**) still appear to encode intact ORFs (albeit with the mt
444 genetic code), we believe that they are unlikely to be under selection for having the same function
445 as when they were present in the mt-genome. Mainly because the mt genetic code will cause
446 problems during translation, but also since the pseudogenizing mutations appear to have occurred
447 randomly throughout the genes (**Fig. S2**), suggesting neutrality. However, we can't exclude that
448 some of the DNA in the Mega-NUMT might be transcribed and serve yet undetermined functions.
449 Potentially, the TEs within and between α-NUMT units, especially remnants of LTR
450 retrotransposons like *Copia* and *Gypsy* (**Fig. 4**), may contribute to expression, as LTRs are well
451 known to harbor cell- and tissue-specific *cis*-regulatory elements that affect expression of
452 neighboring host genes (McDonald et al., 1997). Alternatively, TEs can also create *de novo*
453 regulatory regions acting *in trans*, so-called piRNA clusters (recently reviewed in Pritam and
454 Signor, 2025, which, if expressed in the antisense orientation of the transposon, can serve as
455 efficient epigenetic silencers of homologous TEs via the piRNA pathway (Brennecke et al., 2008).

Indeed, *de novo* creation of such a repressive piRNA cluster has recently been demonstrated in *D. melanogaster* (Rafanel et al., 2025).

Another possibility is that the Mega-NUMT provides a structural function. Other studies have seen that NUMTs are located close to centromeric or telomeric regions (Puertas & González-Sánchez, 2020), suggesting that they might be involved in their function. However, the *Dpau* Mega-NUMT of O11 is not found in either the centromeric or telomeric chromosomal region of chromosome 3, according to our FISH. Even so, a yet undetermined structural role for the *Dpau* Mega-NUMT is still possible.

Further experiments will be necessary to elucidate any potential function and fitness benefits of the Mega-NUMT in the *D. paulistorum* species complex.

Conclusions

To our knowledge, this is the first verified existence and detailed dissection of a Mega-NUMT outside cats and humans. Moreover, we show that Mega-NUMTs can persist at high prevalence in nature and over relatively long time periods, suggesting balancing selection. Our findings strengthen the importance of high-quality long-read sequencing technologies for deciphering complex repeat-rich genomic regions and mobile DNAs to deepen our understanding of genomic "dark matter". Finally, it is expected that with the rapidly increasing number of high-quality genomes, the *Dpau* Mega-NUMTs will not remain a unicorn. Instead, similar Mega-NUMTs will very likely be found in further eukaryotic genomes.

Data availability

All data will be available upon publication.

Acknowledgements

We thank Tom Martin for performing the DNA extraction for Nanopore sequencing of *D. paulistorum* O11. ONT sequencing of O11 was performed by the Uppsala Genome Center (UGC) in Uppsala, Sweden. The facility is part of the National Genomics Infrastructure (NGI) Sweden and the Science for Life Laboratory. Some of the data handling was enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX. UGC and UPPMAX are supported by the Swedish Research Council. This work was supported by the Swedish Research

487 Council VR grant 2014–4353 to LK and by the Austrian Science Fund FWF grants P28255B22,
488 FW613A0501 and FW613A0502 to WJM. We also thank the Nouragues research field station
489 (managed by CNRS), which benefits from "Investissement d'Avenir" grants managed by Agence
490 Nationale de la Recherhe (AnaEE France ANR-11-INBS-0001; Labex CEBA ANR-10-LABX-25-
491 01) to WJM.
492

493 Contributions

494 LK and WJM conceived and designed the study. MMN performed assemblies, annotations and
495 phylogenetic analyses. EH performed qPCRs. JT and DIS performed line screening and RFLP
496 analyses. AS and CC performed FISH. AH-V and WJM collected flies. LK performed
497 bioinformatic analyses. MMN, WJM and LK wrote the paper with contributions from CC and AS.
498 All authors read and approved the manuscript.
499

500 Material and Methods

501 Fly strains

502 All fly lines used in the present study are listed in **Table S6**, including their collection date and
503 semispecies designation. All flies were kept at 25±1°C on Formula 4-24 Drosophila instant food
504 (Carolina, USA) with a 12 hrs. light-dark cycle.

505 Drosophila collections from French Guiana 2014-2024

506 Field samples of neotropical Drosophila specimens were collected in French Guiana at five
507 different geographic locations between 2014 and 2024 (**Table S7**) by trapping live flies with
508 banana/yeast baits in plastic PET bottles over a 48 hrs. time period. Taxonomic grouping - at least
509 at the species-group level - was performed by eye under the stereo loupe by using FlyNap. Trapped
510 inseminated single females were isolated into individual small single fly culture vials supplemented
511 with Carolina Instant Fly-Food, shipped back to the Medical University of Vienna to establish
512 consecutive isofemale lines in WM´s lab. Upon egg deposition and the emergence of F1 larvae,
513 DNA was isolated from the G0 mother for single fly DNA extraction with the Gentra Puregene
514 Tissue kit (Qiagen, Germany) and mitochondrial *CO1*-barcoding. PCR was performed with the
515 primer pair CO1-F and CO1-univR (**Table S1**), followed by direct Sanger sequencing and cloning
516 (for further details, see Madi-Ravazzi et al., 2021).

517 This study was conducted under ABS Permits No. ABSCH-IRCC-FR-245930-1 and ABSCH-
518 IRCC-FR-257164-1 issued by the French Ministry of Ecological Transition on 25 Jun 2018 and 16
519 Sep 2021 respectively, in compliance with the Nagoya Protocol and French Biodiversity Law
520 (2016).

521 **Single fly Southern blots**

522 Single fly Southern blots were performed following the protocol by Junakovic, 2004. In brief, DNA
523 was extracted from single individuals, digested with the restriction enzyme *Hin*dIII and processed
524 for electrophoresis on a vertical 0.8% agarose gel. After vacuum-blotting onto a positively charged
525 nylon membrane, fragments were hybridized with a P^{32}-labelled mitochondrial *CO1* probe derived
526 from a plasmid. Detection was performed by exposing the membranes to X-ray films at -80°C.

527 **Restriction Fragment Length Polymorphism-PCR (RFLP-PCR)**

528 Total DNA was extracted from single or multiple flies using the Gentra Puregene Tissue kit
529 (Qiagen, Germany). Flies were homogenized using a TissueLyser LT (Qiagen, Germany) at 50 Hz
530 for 50 sec and DNA was consequently isolated following the manufacturer's protocol. DNA
531 concentration was determined with a NanoDrop OneC Spectrophotometer (Thermo Scientific,
532 USA). RFLP-PCR for the mitochondrial Cytochrome c oxidase genes *CO1* and *CO2* (**Figure S1**)
533 was set up in 20 µl reactions containing 1x Promega reaction buffer, 2.5 mM $MgCl_2$, 0.5 µM of
534 each primer (**Table S1**), 200 µM of each dNTP, and 0.025 U of GoTaq G2 DNA Polymerase.
535 *CO1*/*CO2* reactions were performed using the following thermal profile: initial denaturation for 2
536 min at 95°C followed by 30 cycles consisting of 45 sec denaturation at 95°C, 45 sec annealing at
537 the appropriate primer conditions T_m°C (**Table S1**) and 30 sec extension at 72°C. Final extension
538 was carried out at 72°C for 10 min. PCR products were consequently used for RFLP (digest plus
539 undigested control) using the restriction enzymes *Hin*dIII and *Eco*RV-HF for *CO1*, and *Msp*I for
540 *CO2* (**Figure S1**). Reactions were performed in 20 µl volume containing 10 or 20 U of enzyme, 1x
541 reaction buffer and 10 µl PCR product. After overnight incubation at 37°C, enzymes were
542 inactivated by either heat (80°C for 20 min) or by adding EDTA-containing gel loading dye. 10 µl
543 of undigested and digested samples were then run on a 1.5% agarose gel and post-stained with 10
544 mg/ml ethidium bromide. Imaging was performed using a Molecular Imager ChemiDocTM XRS
545 Imaging System and the intensities of obtained PCR bands were quantified using Image Lab 5.2.1
546 software (Bio-Rad).

547 **Quantifying Mega-NUMT abundance within a fly**

548 Undigested samples representing the sum of both mitotypes were set as a reference band with
549 relative quantity 1. The relative levels of mitotypes were then quantified by summing up the two
550 digested fragments plus the undigested fragment. Due to technical differences in band intensity,
551 however, the sum of all fragments did not result in a relative quantity of 1. Hence, the actual sum
552 value was set to 100% to determine the percentage of mitotype levels. Smaller fragments with a
553 length up to approximately 100 bp were not detectable after ethidium bromide staining and we
554 hence used the correction factor C*f* = undigested [bp] / digested [bp] to estimate their relative
555 quantities.

556 **Quantitative PCR (qPCR)**

557 DNA was extracted by homogenizing whole flies in 200µl Chelex solution at 5% concentration in
558 distilled water. A total of 10µl of proteinase K (10mg/ml) was added per sample and the vials were
559 incubated for 6 hours at 56°C. After digestion, vials were centrifuged at 16,200 x g for 3 minutes,
560 and the supernatant was transferred to a new vial. DNA was stored at 4°C and freeze-thaw cycles
561 were avoided when possible. Samples were run in duplicate using SYBR Green reagent on a
562 StepOnePlus Real Time PCR machine (Applied Biosystems, USA). Technical repeats with >1 Ct
563 difference were rerun or discarded. Each reaction contained 10µl Maxima SYBR Green/ROX
564 qPCR Master Mix (2x), 1µl of forward primer, 1µl of reverse primer, 6µl nuclease-free water and
565 2µl of the DNA sample. Discriminant mitochondrial primers cycling conditions were 95°C for 20
566 seconds, followed by 40 cycles of: 95°C for 3 seconds and 55°C for 30 seconds (**Table S1**).
567 Discriminant primers used linked base technology to increase specificity (Exiqon, Denmark). Melt
568 curves and the fly control gene *rps17* were run on each plate to control for successful single-peak
569 amplification and to control for DNA quality. All primers had >90% efficiency. For absolute
570 quantification, each 96-well qPCR plate was analyzed using StepOne Software v2.2.2 (Applied
571 Biosystems, USA) and Ct values were obtained by comparing each primer sample to a single
572 standard curve of known concentration and using identical threshold and baseline levels for each
573 primer target across plates. Standard curves were created by amplifying positive control samples
574 using PCR, calculating DNA concentrations using QBIT, and then serially diluting the sample 1:10
575 with distilled water to create a 5-sample curve comprising known concentrations decreasing from
576 10pmol/ ml. Samples with Ct values over 30 were classed as negative, confirmed by our negative
577 controls. As a result, the detection limit was 50 copies per sample for *rps17*, 115 for *CO3*
578 (coxSNP_1) and 735 for *CO3* (coxSNP_2). DNA concentration was calculated for each sample by
579 comparing the sample Ct values back to the equation of the standard curve. DNA copy numbers

580 were then calculated by using the DNA product length to determine DNA concentration. To control
581 for fly size and extraction efficiency, copy numbers for each sample were normalized using the fly
582 housekeeping gene as a control, giving a final value in terms of mitochondrial genome copies to
583 fly nuclear genome copy ratio. Since *rps17* is located on Muller element D, which in *D.*
584 *paulistorum* is fused to Muller element A and part of the X chromosome, the copy numbers differ
585 between males and females. Hence, male *rps17* numbers were doubled. We used ANOVA in R
586 version 3.3.2 (R Core Team, 2016) when testing for significant changes in copy numbers of the α-
587 NUMT and the real β mitochondrial genome between ages and sexes of different tissues. The data
588 was log-transformed to increase normality.

589 **DNA extraction, genome sequencing and assembly**

590 *DNA extractions and Nanopore sequencing*

591 Total DNA for Oxford Nanopore (ONT) sequencing of O11 was extracted from many adult flies
592 of both sexes using the MagAttract HMW DNA Kit (Qiagen, Germany) with only minor
593 modifications of the manufacturer's instructions. Several similar independent DNA extractions
594 were then pooled, cleaned and concentrated with the Ampure XP beads clean-up protocol
595 (Beckman Coulter, USA). The Circulomics SRE XS kit (Pacific Biosciences, USA) was then used
596 to eliminate short DNA fragments before sequencing. A DNA library was prepared using the
597 Ligation Sequencing kit (Oxford Nanopore Technologies, UK) and sequenced on an R9.4
598 PromethION flowcell (FLO-PRO002) at the Uppsala Genome Center, Uppsala, Sweden.
599 Basecalling was performed with Guppy 4.3.4 and the HAC model. A total of 1.4M reads were
600 generated, with 19.54 Gb passed bases (**Table S3**).

601 *D. paulistorum O11 genome assembly*

602 Reads were subsampled with Filtlong v.0.2.1 (https://github.com/rrwick/Filtlong, last visited 18th
603 March 2022) to lower the overall coverage of the reads to facilitate the assembly process. After
604 testing different settings (**Table S3**), the O11 Nanopore reads were filtered to a coverage of 50X,
605 giving priority to the quality of the reads (--target_bases 12500000000 –min_length 1000 –
606 mean_q_weight 10). The whole genome assembly was built from the dataset of subsampled reads
607 using NextDenovo v.2.5.0 (Hu et al., 2024; https://github.com/Nextomics/NextDenovo, last visited
608 18th March 2022) with a read cutoff of 1 Kb, and setting the genome size to 250 Mb. The genome
609 assembly was then polished with both the same Nanopore reads and Illumina reads from Baião et
610 al., 2023 (available via NCBI Bioproject PRJNA643793). First, Nanopore reads were mapped to

611 the genome assembly with Minimap2 (Li, 2018) using the map-ont mode and bam files were sorted
612 and indexed with samtools (Danecek et al., 2021). Then, the pipeline PEPPER-Margin-
613 DeepVariant r0.4 (Shafin et al., 2021) was used to call variants. The variants were filtered with
614 bcftools v1.15.1-21 (Danecek et al., 2021) using the following thresholds: QUAL>=30 &&
615 FMT/DP>=10 && FMT/GQ>=30 && FMT/VAF>=0.8, and remaining SNPs were corrected in
616 the genome with bcftools v1.15.1-21 (Danecek et al., 2021). Illumina reads were aligned to the
617 genome assembly using BWA mem v0.7.17-r1198-dirty (Li, 2013) and sorted with samtools v1.14
618 (Danecek et al., 2021), and two runs of Pilon v1.24 (Walker et al., 2014) were used for correction.
619 BUSCO v5.2.2 with diptera_odb10 and augustus species fly was run to check genome assembly
620 completeness, and Quast v5.0.2 (Mikheenko et al., 2018) was run with default settings to
621 quantitatively assess the genome assembly stats. Scripts can be found in
622 https://github.com/mmontonerin/Dpaulistorum_Mega-NUMT.

623 **Detection and annotation of α-NUMT copies in the genome assembly**

624 The O11 β mt-genome (Baião et al., 2023) was used as a query when running BLASTn v2.12.0+
625 (Camacho et al., 2009) against the genome assembly. Hits over 500 bp long and with 90% identity
626 to the mitochondrial sequence were inspected, and the 12 mt-genome copies in the Mega-NUMT
627 region were identified on contig ctg000940. The position of each gene in the α-NUMT copy in the
628 Mega-NUMT was obtained by using gene sequences from the *D. willistoni* mt-genome (GD-H4-1
629 line) as queries in a BLASTn search. Each inferred gene was subsequently checked manually in
630 Artemis v.16 (Rutherford et al., 2000) to identify pseudogenes and adjust the position if needed.
631 Scripts can be found in https://github.com/mmontonerin/Dpaulistorum_Mega-NUMT.

632 **Phylogenomic analysis of the α-NUMT copies**

633 The 13 mitochondrial protein-coding genes from all α-NUMT copies and the genes from the real
634 mt-genomes from several species of the willistoni group (see **Fig. 3**), previously published in Baião
635 et al., 2023 (available via NCBI Bioproject PRJNA643793), were aligned using MAFFT v.7.407
636 (Katoh & Standley, 2013) and subsequently trimmed with trimAl v1.4.1 (Capella-Gutiérrez et al.,
637 2009) using the setting -gt 0.1. The python script gene_stitcher.py
638 (https://github.com/ballesterus/Utensils/blob/master/geneStitcher.py, last visited 12 March 2026)
639 was used to concatenate the trimmed alignments. The concatenated alignment was used to infer a
640 maximum likelihood phylogeny in IQTree2 v2.2.0 (Minh et al., 2020) with one partition per gene

641 and the parameter settings -m MFP -b 100 -T AUTO. Scripts can be found in
642 https://github.com/mmontonerin/Dpaulistorum_Mega-NUMT.

643 **Analyzing insertion sequences in α-NUMTs**

644 Finally, we used a snakemake pipeline for repeat discovery and annotation in the genome assembly,
645 which includes prediction of repeat content and transposable elements with RepeatModeler v1.0.8
646 (https://github.com/Dfam-consortium/RepeatModeler, last visited 12 March 2026). The pipeline
647 also creates a database based on UniProt/Swissprot (The UniProt Consortium, 2025) protein dataset
648 without transposable elements that initial predictions can be blasted against and removed if
649 matching. By this, protein-coding genes in multiple copies are avoided in our final repeat library.
650 The final repeat library was then used to mask the genome assembly with RepeatMasker v4.0.7
651 (https://github.com/Dfam-consortium/RepeatModeler, last visited 12 March 2026). Scripts can be
652 found in https://github.com/mmontonerin/Dpaulistorum_Mega-NUMT.

653 **Collection of staged fly embryos for DNA extraction and DNA FISH**

654 *D. paulistorum* embryos at early (0-30 min AEL), mid (3-4 hrs. AEL) and late (18-20 hrs. AEL)
655 stages of development were collected on homemade medium (agar, molasses, cornmeal and yeast
656 medium) in plastic Petri dishes (3 x 3 cm) using collection chambers at 25°C. Prior to collecting
657 embryos, 1–2-week-old flies were kept in these chambers for 2-3 days to adapt to the new
658 environment. Embryos were then collected by gently washing them off the plate onto a mesh with
659 water. A subset of embryos was further processed for DNA FISH (see below). The rest was
660 manually picked with a dissection needle and transferred immediately to an ice-cold lysis buffer
661 (Qiagen, Germany) and processed for DNA extraction or stored at -20°C for no more than one
662 week. In total, we collected 20 embryos for each sample with at least five biological replicates.

663 **Fluorescent in situ hybridization (FISH) for Mega-NUMT visualization**

664 We performed DNA FISH on various fly tissues using both satellite and Mega-NUMT probes. The
665 Mega-NUMT probe was generated by direct labeling via PCR with modified dNTPs (Alexa Fluor
666 488 dUTP; Jena Bioscience, Germany). We used a mixture of five mitochondrial loci (*CO1*, *CO2*,
667 *ND1*, *ND2* and *ND4*) of *D. paulistorum*, cloned beforehand in pTZ57R/5 vector (Thermo
668 Scientific, USA). For targeting the satellite sequence adjacent to the Mega-NUMT, we used a
669 primary oligo probe coupled with a sec6 adaptor (satO11-
670 CACACGCTCTCCGTCTTGGCCGTGGTCGATCAttttttttttTTTATAAAAATTAATACTGAG
671 GACAAACTGAGGACT) and the secondary probe Sec6 coupled with Cy3 (Sec6-

672 /5Cy3/aTGATCGACCACGGCCAAGACGGAGAGCGTGTGaa). The probe for the nuclear
673 single copy gene *FOXO* gene was generated in the same way as the Mega-NUMT probe, but by
674 only using three probes covering the *D. paulistorum FOXO* gene and labelled with Alexa Fluor
675 594 dUTP (Jena Bioscience, Germany).

676 To visualize Mega-NUMT in mitotically active neuroblasts of third instar larvae, we
677 dissected their brains in PBS and incubated for 15 min in 0.5% sodium citrate. Then we fixed the
678 brains for 15 min in 4% formaldehyde and 45% acetic acid before squashing them between the
679 slide and coverslip. Squashed samples were immediately immersed in liquid nitrogen and
680 transferred afterwards in 100% ethanol for 5 min. Then we air dried slides for at least 1 hour before
681 proceeding to the hybridization. For the hybridization, we used 20 pmol of satellite probes coupled
682 with 80 pmol of the sec6 probe and approximately 150 ng of Mega-NUMT probe in 50 µL of
683 hybridization buffer (50% formamide, 10% dextran sulfate, 2xSSC). We heated slides for 5 min at
684 95°C to denature and incubated them overnight at 37°C in a humid chamber. We then washed the
685 slides 3 times for 5 min with 4xSSCT and 3 times for 5 min with 0.1xSSC before mounting in
686 SlowFade with DAPI and sealing with nail polish.

687 To visualize Mega-NUMT in mitotically active early embryos (0-3 h), staged embryos were
688 collected as stated above, dechorionated in 50% bleach and fixed for 20 min in 3.7%
689 paraformaldehyde in PBX (PBS and 0.15% Triton X-100) mixed with heptane (1:1 v/v). Then the
690 embryos were postfixed in ice-cold methanol and further processed for FISH following a protocol
691 by Gemkow et al., 1996 with slight modifications. Additionally, salivary glands and polytene
692 chromosomes were used for Mega-NUMT visualization. The glands were dissected in PBS and
693 fixed for 20 min in 3.7% paraformaldehyde in PBX (PBS and 0.15% Triton X-100). After washing
694 three times in PBX, the tissues were kept in 70% ethanol until being used for DNA FISH. Polytene
695 chromosomes were prepared according to (Pimpinelli et al., 2010) and processed for DNA FISH
696 as described above.

697 Images were taken with an Olympus FW3000 laser scanning confocal microscope
698 (Olympus Corporation, Japan) using a 60x objective and further processed with Photoshop CS6
699 (Adobe CS, USA), adjusting levels for each channel.

700

701 **Bibliography**

702 Avise, J. C., Shapira, J. F., Daniel, S. W., Aquadro, C. F., & Lansman, R. A. (1983).

703 Mitochondrial DNA differentiation during the speciation process in Peromyscus.

704 *Molecular Biology and Evolution*, *1*(1), 38–56.
705 https://doi.org/10.1093/oxfordjournals.molbev.a040301
706 Bai, R., Cui, H., Devaney, J. M., Allis, K. M., Balog, A. M., Liu, X., Schnur, R. E.,
707 Shapiro, F. L., Brautbar, A., Estrada-Veras, J. I., Hochstetler, L., McConkie-Rosell, A.,
708 McDonald, M. T., Solomon, B. D., Hofherr, S., Richard, G., & Suchy, S. F. (2021).
709 Interference of nuclear mitochondrial DNA segments in mitochondrial DNA testing
710 resembles biparental transmission of mitochondrial DNA in humans. *Genetics in*
711 *Medicine*, *23*(8), 1514–1521. https://doi.org/10.1038/s41436-021-01166-1
712 Baião, G. C., Schneider, D. I., Miller, W. J., & Klasson, L. (2023). Multiple
713 introgressions shape mitochondrial evolutionary history in Drosophila paulistorum and
714 the Drosophila willistoni group. *Molecular Phylogenetics and Evolution*, *180*, 107683.
715 https://doi.org/10.1016/j.ympev.2022.107683
716 Balciuniene, J., & Balciunas, D. (2019). A Nuclear mtDNA Concatemer (Mega-NUMT)
717 Could Mimic Paternal Inheritance of Mitochondrial Genome. *Frontiers in Genetics*, *10*.
718 https://www.frontiersin.org/articles/10.3389/fgene.2019.00518
719 Barro-Trastoy, D., & Köhler, C. (2024). Helitrons: Genomic parasites that generate
720 developmental novelties. *Trends in Genetics*, *40*(5), 437–448.
721 https://doi.org/10.1016/j.tig.2024.02.002
722 Behura, S. K. (2007). Analysis of Nuclear Copies of Mitochondrial Sequences in
723 Honeybee (Apis mellifera) Genome. *Molecular Biology and Evolution*, *24*(7), 1492–
724 1505. https://doi.org/10.1093/molbev/msm068
725 Bensasson, D., Zhang, D.-X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial
726 pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology & Evolution*, *16*(6),
727 314–321. https://doi.org/10.1016/S0169-5347(01)02151-6

Biró, B., Gál, Z., Fekete, Z., Klecska, E., & Hoffmann, O. I. (2024). Mitochondrial genome plasticity of mammalian species. *BMC Genomics*, *25*(1), 278. https://doi.org/10.1186/s12864-024-10201-9

Blanchard, J. L., & Schmidt, G. W. (1995). Pervasive migration of organellar DNA to the nucleus in plants. *Journal of Molecular Evolution*, *41*(4), 397–406. https://doi.org/10.1007/BF00160310

Blanchard, J. L., & Schmidt, G. W. (1996). Mitochondrial DNA migration events in yeast and humans: Integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Molecular Biology and Evolution*, *13*(3), 537–548. https://doi.org/10.1093/oxfordjournals.molbev.a025614

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., & Hannon, G. J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science*, *322*(5906), 1387–1392. https://doi.org/10.1126/science.1165171

Breton, S., Beaupré, H. D., Stewart, D. T., Hoeh, W. R., & Blier, P. U. (2007). The unusual system of doubly uniparental inheritance of mtDNA: Isn't one enough? *Trends in Genetics*, *23*(9), 465–474. https://doi.org/10.1016/j.tig.2007.05.011

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, *76*(4), 1967–1971. https://doi.org/10.1073/pnas.76.4.1967

Burla, H., da Cunha, A. B., Cordeiro, A. R., Dobzhansky, Th., Malogolowkin, C., & Pavan, C. (1949). The Willistoni Group of Sibling Species of Drosophila. *Evolution*, *3*(4), 300–314. https://doi.org/10.2307/2405716

750 Calvignac, S., Konecny, L., Malard, F., & Douady, C. J. (2011). Preventing the pollution
751 of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion*,
752 *11*(2), 246–254. https://doi.org/10.1016/j.mito.2010.10.004
753 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., &
754 Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*,
755 *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421
756 Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for
757 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*,
758 *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348
759 Cohen, S., Agmon, N., Yacobi, K., Mislovati, M., & Segal, D. (2005). Evidence for
760 rolling circle replication of tandem genes in Drosophila. *Nucleic Acids Research*, *33*(14),
761 4519–4526. https://doi.org/10.1093/nar/gki764
762 Cohen, S., & Segal, D. (2009). Extrachromosomal Circular DNA in Eukaryotes: Possible
763 Involvement in the Plasticity of Tandem Repeats. *Cytogenetic and Genome Research*,
764 *124*(3–4), 327–338. https://doi.org/10.1159/000218136
765 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O.,
766 Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years
767 of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008.
768 https://doi.org/10.1093/gigascience/giab008
769 Dayama, G., Emery, S. B., Kidd, J. M., & Mills, R. E. (2014). The genomic landscape of
770 polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Research*, *42*(20),
771 12640–12649. https://doi.org/10.1093/nar/gku1038

DeSalle, R., & Goldstein, P. (2019). Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution*, *7*. https://www.frontiersin.org/articles/10.3389/fevo.2019.00302

Dobzhansky, T., & Spassky, B. (1959). Drosophila paulistorum, a cluster of species in statu nascendi*†. *Proceedings of the National Academy of Sciences*, *45*(3), 419–428. https://doi.org/10.1073/pnas.45.3.419

Dobzhansky, Th., Ehrman, L., Pavlovsky, O., & Spassky, B. (1964). The superspecies drosophila paulistorum*. *Proceedings of the National Academy of Sciences*, *51*(1), 3–9. https://doi.org/10.1073/pnas.51.1.3

Elías-Gutiérrez, M., Jerónimo, F. M., Ivanova, N. V., Valdez-Moreno, M., & Hebert, P. D. N. (2008). DNA barcodes for Cladocera and Copepoda from Mexico and Guatemala, highlights and new discoveries. *Zootaxa*, *1839*(1), Article 1. https://doi.org/10.11646/zootaxa.1839.1.1

Fields, P. D., Waneka, G., Naish, M., Schatz, M. C., Henderson, I. R., & Sloan, D. B. (2022). Complete Sequence of a 641-kb Insertion of Mitochondrial DNA in the Arabidopsis thaliana Nuclear Genome. *Genome Biology and Evolution*, *14*(5), evac059. https://doi.org/10.1093/gbe/evac059

Filée, J., Forterre, P., Sen-Lin, T., & Laurent, J. (2002). Evolution of DNA Polymerase Families: Evidences for Multiple Gene Exchange Between Cellular and Viral Proteins. *Journal of Molecular Evolution*, *54*(6), 763–773. https://doi.org/10.1007/s00239-001-0078-x

Gemkow, M. J., Buchenau, P., & Arndt-Jovin, D. J. (1996). FISH in whole-mount Drosophila embryos. RNA: Activation of a transcriptional locus, DNA: gene architecture

and expression. *Bioimaging*, *4*(2), 107–120. https://doi.org/10.1002/1361-6374(199606)4:2%3C107::AID-BIO8%3E3.0.CO;2-Q

Gleason, J. M., Griffith, E. C., & Powell, J. R. (1998). A MOLECULAR PHYLOGENY OF THE DROSOPHILA WILLISTONI GROUP: CONFLICTS BETWEEN SPECIES CONCEPTS? *Evolution*, *52*(4), 1093–1103. https://doi.org/10.1111/j.1558-5646.1998.tb01836.x

Gromko, M. H., & Richmond, R. C. (1978). MODES OF SELECTION MAINTAINING AN INVERSION POLYMORPHISM IN DROSOPHILA PAULISTORUM. *Genetics*, *88*(2), 357–366. https://doi.org/10.1093/genetics/88.2.357

Harrison, R. G. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution*, *4*(1), 6–11. https://doi.org/10.1016/0169-5347(89)90006-2

Hazkani-Covo, E., & Graur, D. (2007). A Comparative Analysis of numt Evolution in Human and Chimpanzee. *Molecular Biology and Evolution*, *24*(1), 13–18. https://doi.org/10.1093/molbev/msl149

Hebert, P. D. N., Bock, D. G., & Prosser, S. W. J. (2023). Interrogating 1000 insect genomes for NUMTs: A risk assessment for estimates of species richness. *PLOS ONE*, *18*(6), e0286620. https://doi.org/10.1371/journal.pone.0286620

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hirano, M., Shtilbans, A., Mayeux, R., Davidson, M. M., DiMauro, S., Knowles, J. A., & Schon, E. A. (1997). Apparent mtDNA heteroplasmy in Alzheimer's disease patients and

819 in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes.
820 *Proceedings of the National Academy of Sciences*, *94*(26), 14894–14899.
821 https://doi.org/10.1073/pnas.94.26.14894
822 Hotopp, J. C. D., Clark, M. E., Oliveira, D. C. S. G., Foster, J. M., Fischer, P., Torres, M.
823 C. M., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard,
824 J., Tomkins, J., Richards, S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H., &
825 Werren, J. H. (2007). Widespread Lateral Gene Transfer from Intracellular Bacteria to
826 Multicellular Eukaryotes. *Science*, *317*(5845), 1753–1756.
827 https://doi.org/10.1126/science.1142490
828 Hu, J., Wang, Z., Sun, Z., Hu, B., Ayoola, A. O., Liang, F., Li, J., Sandoval, J. R.,
829 Cooper, D. N., Ye, K., Ruan, J., Xiao, C.-L., Wang, D., Wu, D.-D., & Wang, S. (2024).
830 NextDenovo: An efficient error correction and accurate assembly tool for noisy long
831 reads. *Genome Biology*, *25*(1), 107. https://doi.org/10.1186/s13059-024-03252-4
832 Junakovic, N. (2004). Southern Blot Analysis of Individual Drosophila Flies. In W. J.
833 Miller & P. Capy (Eds), *Mobile Genetic Elements: Protocols and Genomic Applications*
834 (pp. 41–57). Humana Press. https://doi.org/10.1385/1-59259-755-6:041
835 Jünger, M. A., Rintelen, F., Stocker, H., Wasserman, J. D., Végh, M., Radimerski, T.,
836 Greenberg, M. E., & Hafen, E. (2003). The Drosophila Forkhead transcription factor
837 FOXO mediates the reduction in cell number associated with reduced insulin signaling.
838 *Journal of Biology*, *2*(3), 20. https://doi.org/10.1186/1475-4924-2-20
839 Kastritsis, C. D. (1967). A comparative study of the chromosomal polymorphs in the
840 incipient species of the Drosophila paulistorum complex. *Chromosoma*, *23*(2), 180–202.
841 https://doi.org/10.1007/BF00331112

842 Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software
843 Version 7: Improvements in Performance and Usability. *Molecular Biology and*
844 *Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010
845 Keeling, P. J. (2024). Horizontal gene transfer in eukaryotes: Aligning theory with data.
846 *Nature Reviews Genetics*, *25*(6), 416–430. https://doi.org/10.1038/s41576-023-00688-5
847 Kim, B. Y., Wang, J. R., Miller, D. E., Barmina, O., Delaney, E., Thompson, A.,
848 Comeault, A. A., Peede, D., D'Agostino, E. R., Pelaez, J., Aguilar, J. M., Haji, D.,
849 Matsunaga, T., Armstrong, E., Zych, M., Ogawa, Y., Stamenković-Radak, M., Jelić, M.,
850 Veselinović, M. S., … Petrov, D. A. (2022). Correction: Highly contiguous assemblies of
851 101 drosophilid genomes. *eLife*, *11*, e78579. https://doi.org/10.7554/eLife.78579
852 Klasson, L., Kumar, N., Bromley, R., Sieber, K., Flowers, M., Ott, S. H., Tallon, L. J.,
853 Andersson, S. G. E., & Dunning Hotopp, J. C. (2014). Extensive duplication of the
854 Wolbachia DNA in chromosome four of Drosophila ananassae. *BMC Genomics*, *15*(1),
855 1097. https://doi.org/10.1186/1471-2164-15-1097
856 Kondo, N., Nikoh, N., Ijichi, N., Shimada, M., & Fukatsu, T. (2002). Genome fragment
857 of Wolbachia endosymbiont transferred to X chromosome of host insect. *Proceedings of*
858 *the National Academy of Sciences*, *99*(22), 14280–14285.
859 https://doi.org/10.1073/pnas.222228199
860 Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes
861 for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, *30*(1), 25–35.
862 https://doi.org/10.1016/j.tree.2014.10.008
863 Krimbas, C. B., & Powell, J. R. (1992). *Drosophila Inversion Polymorphism*. CRC Press.

Ladoukakis, E. D., & Zouros, E. (2017). Evolution and inheritance of animal mitochondrial DNA: Rules and exceptions. *Journal of Biological Research-Thessaloniki*, *24*(1), 2. https://doi.org/10.1186/s40709-017-0060-4

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM* (arXiv:1303.3997). arXiv. https://doi.org/10.48550/arXiv.1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Ling, F., Niu, R., Hatakeyama, H., Goto, Y., Shibata, T., & Yoshida, M. (2016). Reactive oxygen species stimulate mitochondrial allele segregation toward homoplasmy in human cells. *Molecular Biology of the Cell*, *27*(10), 1684–1693. https://doi.org/10.1091/mbc.E15-10-0690

Ling, F., & Yoshida, M. (2020). Rolling-Circle Replication in Mitochondrial DNA Inheritance: Scientific Evidence and Significance from Yeast to Human Cells. *Genes*, *11*(5). https://doi.org/10.3390/genes11050514

Lopez, J. V., Yuhki, N., Masuda, R., Modi, W., & O'Brien, S. J. (1994). Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, *39*(2), 174–190. https://doi.org/10.1007/BF00163806

Luo, S., Valencia, C. A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J., Chen, S. M., Chien, Y.-H., Hwu, W.-L., Fan, P.-C., Wong, L.-J., Atwal, P. S., & Huang, T. (2018). Biparental Inheritance of Mitochondrial DNA in Humans. *Proceedings of the National Academy of Sciences*, *115*(51), 13039–13044. https://doi.org/10.1073/pnas.1810946115

Lutz-Bonengel, S., Niederstätter, H., Naue, J., Koziel, R., Yang, F., Sänger, T., Huber, G., Berger, C., Pflugradt, R., Strobl, C., Xavier, C., Volleth, M., Weiß, S. C., Irwin, J. A., Romsos, E. L., Vallone, P. M., Ratzinger, G., Schmuth, M., Jansen-Dürr, P., … Parson, W. (2021). Evidence for multi-copy Mega-NUMTs in the human genome. *Nucleic Acids Research*, *49*(3), 1517–1531. https://doi.org/10.1093/nar/gkaa1271

McDonald, J. F., Matyunina, L. V., Wilson, S., King Jordan, I., Bowen, N. J., & Miller, W. J. (1997). Ltr retrotransposons and the evolution of eukaryotic enhancers. *Genetica*, *100*(1), 3–13. https://doi.org/10.1023/A:1018392117410

McWilliams, T. G., & Suomalainen, A. (2019). Mitochondrial DNA can be inherited from fathers, not just mothers. *Nature*, *565*(7739), 296–297. https://doi.org/10.1038/d41586-019-00093-1

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., 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*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Muller, H. J. (1940). An analysis of the process of structural change in chromosomes ofDrosophila. *Journal of Genetics*, *40*(1), 1–66. https://doi.org/10.1007/BF02982481

Nikoh, N., Tanaka, K., Shibata, F., Kondo, N., Hizume, M., Shimada, M., & Fukatsu, T. (2008). Wolbachia genome integrated in an insect chromosome: Evolution and fate of laterally transferred endosymbiont genes. *Genome Research*, *18*(2), 272–280. https://doi.org/10.1101/gr.7144908

Noutsos, C., Kleine, T., Armbruster, U., DalCorso, G., & Leister, D. (2007). Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. *Trends in Genetics*, *23*(12), 597–601. https://doi.org/10.1016/j.tig.2007.08.016

Pagnamenta, A. T., Wei, W., Rahman, S., & Chinnery, P. F. (2021). Biparental inheritance of mitochondrial DNA revisited. *Nature Reviews Genetics*, *22*(8), 477–478. https://doi.org/10.1038/s41576-021-00380-6

Papachristos, K., Miller, W. J., & Klasson, L. (2025). A co-speciation dilemma and a lifestyle transition with genomic consequences in Wolbachia of Neotropical Drosophila. *BMC Genomics*, *27*(1), 41. https://doi.org/10.1186/s12864-025-12340-z

Parfait, B., Rustin, P., Munnich, A., & Rötig, A. (1998). Coamplification of Nuclear Pseudogenes and Assessment of Heteroplasmy of Mitochondrial DNA Mutations. *Biochemical and Biophysical Research Communications*, *247*(1), 57–59. https://doi.org/10.1006/bbrc.1998.8666

Parr, R. L., Maki, J., Reguly, B., Dakubo, G. D., Aguirre, A., Wittock, R., Robinson, K., Jakupciak, J. P., & Thayer, R. E. (2006). The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. *BMC Genomics*, *7*(1), 185. https://doi.org/10.1186/1471-2164-7-185

Patterson, E. C., Lall, G. M., Neumann, R., Ottolini, B., Sacchini, F., Foster, A. P., Jobling, M. A., & Wetton, J. H. (2023). Defining cat mitogenome variation and accounting for numts via multiplex amplification and Nanopore sequencing. *Forensic Science International: Genetics*, *67*, 102944. https://doi.org/10.1016/j.fsigen.2023.102944

Pimpinelli, S., Bonaccorsi, S., Fanti, L., & Gatti, M. (2010). Fluorescent In Situ Hybridization (FISH) of Mitotic Chromosomes from Drosophila Larval Brain. *Cold Spring Harbor Protocols*, *2010*(3), pdb.prot5391. https://doi.org/10.1101/pdb.prot5391

Pohjoismäki, J. L. O., Goffart, S., Tyynismaa, H., Willcox, S., Ide, T., Kang, D., Suomalainen, A., Karhunen, P. J., Griffith, J. D., Holt, I. J., & Jacobs, H. T. (2009). Human Heart Mitochondrial DNA Is Organized in Complex Catenated Networks Containing Abundant Four-way Junctions and Replication Forks*. *Journal of Biological Chemistry*, *284*(32), 21446–21457. https://doi.org/10.1074/jbc.M109.016600

Pritam, S., & Signor, S. (2025). Evolution of piRNA-guided defense against transposable elements. *Trends in Genetics*, *41*(5), 390–401. https://doi.org/10.1016/j.tig.2024.11.011

Puertas, M. J., & González-Sánchez, M. (2020). Insertions of mitochondrial DNA into the nucleus—Effects and role in cell evolution. *Genome*, *63*(8), 365–374. https://doi.org/10.1139/gen-2019-0151

R Core Team. (2016). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rafanel, B., Protsenko, L., Handler, D., Brennecke, J., & Senti, K.-A. (2025). *Antisense transposon insertions into host genes trigger piRNA mediated immunity* (p. 2025.07.28.667215). bioRxiv. https://doi.org/10.1101/2025.07.28.667215

Robe, L. J., Cordeiro, J., Loreto, E. L. S., & Valente, V. L. S. (2010). Taxonomic boundaries, phylogenetic relationships and biogeography of the Drosophila willistoni subgroup (Diptera: Drosophilidae). *Genetica*, *138*(6), 601–617. https://doi.org/10.1007/s10709-009-9432-5

Rogers, H. H., & Griffiths-Jones, S. (2012). Mitochondrial Pseudogenes in the Nuclear Genomes of Drosophila. *PLOS ONE*, *7*(3), e32593. https://doi.org/10.1371/journal.pone.0032593

958 Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., &
959 Barrell, B. (2000). Artemis: Sequence visualization and annotation. *Bioinformatics*,
960 *16*(10), 944–945. https://doi.org/10.1093/bioinformatics/16.10.944
961 Salas, A., Schönherr, S., Bandelt, H.-J., Gómez-Carballa, A., & Weissensteiner, H.
962 (2020). Extraordinary claims require extraordinary evidence in asserted mtDNA
963 biparental inheritance. *Forensic Science International: Genetics*, *47*, 102274.
964 https://doi.org/10.1016/j.fsigen.2020.102274
965 Salas, A., Yao, Y.-G., Macaulay, V., Vega, A., Carracedo, Á., & Bandelt, H.-J. (2005). A
966 Critical Reassessment of the Role of Mitochondria in Tumorigenesis. *PLOS Medicine*,
967 *2*(11), e296. https://doi.org/10.1371/journal.pmed.0020296
968 Schiavo, G., Hoffmann, O. I., Ribani, A., Utzeri, V. J., Ghionda, M. C., Bertolini, F.,
969 Geraci, C., Bovo, S., & Fontanesi, L. (2017). A genomic landscape of mitochondrial
970 DNA insertions in the pig nuclear genome provides evolutionary signatures of
971 interspecies admixture. *DNA Research*, *24*(5), 487–498.
972 https://doi.org/10.1093/dnares/dsx019
973 Schwartz, M., & Vissing, J. (2002). Paternal Inheritance of Mitochondrial DNA. *New*
974 *England Journal of Medicine*, *347*(8), 576–580. https://doi.org/10.1056/NEJMoa020350
975 Shafin, K., Pesout, T., Chang, P.-C., Nattestad, M., Kolesnikov, A., Goel, S., Baid, G.,
976 Kolmogorov, M., Eizenga, J. M., Miga, K. H., Carnevali, P., Jain, M., Carroll, A., &
977 Paten, B. (2021). Haplotype-aware variant calling with PEPPER-Margin-DeepVariant
978 enables high accuracy in nanopore long-reads. *Nature Methods*, *18*(11), 1322–1332.
979 https://doi.org/10.1038/s41592-021-01299-w

980 Shutt, T. E., & Gray, M. W. (2006). Bacteriophage origins of mitochondrial replication
981 and transcription proteins. *Trends in Genetics*, *22*(2), 90–95.
982 https://doi.org/10.1016/j.tig.2005.11.007
983 Spassky, B., & Dobzhansky, T. (1950). Comparative genetics of Drosophila willistoni.
984 *Heredity*, *4*(2), 201–215. https://doi.org/10.1038/hdy.1950.15
985 The UniProt Consortium. (2025). UniProt: The Universal Protein Knowledgebase in
986 2025. *Nucleic Acids Research*, *53*(D1), D609–D617.
987 https://doi.org/10.1093/nar/gkae1010
988 Thomas, J., Vadnagara, K., & Pritham, E. J. (2014). DINE-1, the highest copy number
989 repeats in Drosophila melanogaster are non-autonomous endonuclease-encoding rolling-
990 circle transposable elements (Helentrons). *Mobile DNA*, *5*(1), 18.
991 https://doi.org/10.1186/1759-8753-5-18
992 Tsuji, J., Frith, M. C., Tomii, K., & Horton, P. (2012). Mammalian NUMT insertion is
993 non-random. *Nucleic Acids Research*, *40*(18), 9073–9088.
994 https://doi.org/10.1093/nar/gks424
995 Valiati, V. H., & Valente, V. L. S. (1997). Chromosomal polymorphism in urban
996 populations of Drosophila paulistorum. *Brazilian Journal of Genetics*, *20*, 567–581.
997 https://doi.org/https://doi.org/10.1590/S0100-84551997000400004
998 Vendrami, D. L. J., Gossmann, T. I., Chakarov, N., Paijmans, A. J., Litzke, V., Eyre-
999 Walker, A., Forcada, J., & Hoffman, J. I. (2022). Signatures of Selection on Mitonuclear
1000 Integrated Genes Uncover Hidden Mitogenomic Variation in Fur Seals. *Genome Biology*
1001 *and Evolution*, *14*(7), evac104. https://doi.org/10.1093/gbe/evac104
1002 Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.
1003 A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool

1004 for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.
1005 *PLOS ONE*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963
1006 Wang, J.-X., Liu, J., Miao, Y.-H., Huang, D.-W., & Xiao, J.-H. (2020). Tracking the
1007 Distribution and Burst of Nuclear Mitochondrial DNA Sequences (NUMTs) in Fig Wasp
1008 Genomes. *Insects*, *11*(10), Article 10. https://doi.org/10.3390/insects11100680
1009 Wei, W., Pagnamenta, A. T., Gleadall, N., Sanchis-Juan, A., Stephens, J., Broxholme, J.,
1010 Tuna, S., Odhams, C. A., Fratter, C., Turro, E., Caulfield, M. J., Taylor, J. C., Rahman, S.,
1011 & Chinnery, P. F. (2020). Nuclear-mitochondrial DNA segments resemble paternally
1012 inherited mitochondrial DNA in humans. *Nature Communications*, *11*(1), 1740.
1013 https://doi.org/10.1038/s41467-020-15336-3
1014 Wei, W., Schon, K. R., Elgar, G., Orioli, A., Tanguy, M., Giess, A., Tischkowitz, M.,
1015 Caulfield, M. J., & Chinnery, P. F. (2022). Nuclear-embedded mitochondrial DNA
1016 sequences in 66,083 human genomes. *Nature*, *611*(7934), 105–114.
1017 https://doi.org/10.1038/s41586-022-05288-7
1018 Werren, J. H. (2011). Selfish genetic elements, genetic conflict, and evolutionary
1019 innovation. *Proceedings of the National Academy of Sciences*, *108*(supplement_2),
1020 10863–10870. https://doi.org/10.1073/pnas.1102343108
1021 Yao, Y.-G., Salas, A., Bravi, C. M., & Bandelt, H.-J. (2006). A reappraisal of complete
1022 mtDNA variation in East Asian families with hearing impairment. *Human Genetics*,
1023 *119*(5), 505–515. https://doi.org/10.1007/s00439-006-0154-9
1024 Zanini, R., Müller, M. J., Vieira, G. C., Valiati, V. H., Deprá, M., & Valente, V. L. da S.
1025 (2018). Combining morphology and molecular data to improve Drosophila paulistorum
1026 (Diptera, Drosophilidae) taxonomic status. *Fly*, *12*(2), 81–94.
1027 https://doi.org/10.1080/19336934.2018.1429859