Title:

Persistence and invasiveness of high-level heteroplasmy through biparental transmission of a selfish mitochondrial genome in *Drosophila*

Running title: Biparental transmission of a selfish mitotype in *Drosophila*

Authors: Guilherme C. Baião[1,§], Anton Strunov[2,§], Eleanor Heyworth[1,§], Daniela I. Schneider[2,*], Julia Thoma[2], Lisa Klasson[1,#], Wolfgang J. Miller[2,#]

Author affiliations: [1] Molecular Evolution, Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Husargatan 3, 751 24, Uppsala, Sweden; [2] Lab Genome Dynamics, Department Cell & Developmental Biology, Center for Anatomy and Cell Biology, Medical University of Vienna, Schwarzspanierstraße 17, 1090, Vienna, Austria.

§ The authors contributed equally

# Corresponding authors, and equal contribution

* Current address: Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, USA.

Emails:

guilherme.baiao@icm.uu.se
anton.strunov@meduniwien.ac.at
heyworth.eleanor@gmail.com
daniela.schneider@moffitt.org
thoma.julia@yahoo.de
lisa.klasson@icm.uu.se
wolfgang.miller@meduniwien.ac.at

Keywords: Heteroplasmy, mitochondria, biparental & paternal transmission, selfish genetic elements, Drosophila, evolution, genomics, introgression, development.

ABSTRACT

Heteroplasmy is the coexistence of more than one type of mitochondrial genomes in an organism. Although widespread sequencing has identified several cases of transient or low-level heteroplasmy that primarily occur through mutation or paternal leakage, stable, high-titer heteroplasmy remains rare in animals. In this study we present a unique, stable and high-level heteroplasmy in male and female flies belonging to the neotropical *Drosophila paulistorum* species complex. We show that mitochondria of *D. paulistorum* are polyphyletic and form two clades, α and β, with two subclades each. Mitochondrial genomes of the α2 subclade appear functional based on their genomic integrity but are exclusively found in heteroplasmic flies and never in homoplasmy, suggesting that they are a secondary mitotype with distinct functionality from the primary one. Using qPCR, we show that α2 titers do not respond to energetic demands of the cell and are generally higher in males than females. By crossing hetero- and homoplasmic flies, we find that α2 can be transmitted to their offspring via both parents and that levels are dependent on nuclear background. Following α2 mitotype levels during embryogenesis, we demonstrate that this secondary mitotype replicates rapidly just after fertilization of the egg in a period when primary mitochondria are dormant. This so-called "Replication precox" phenotype likely prevents the α2 mitotype from being outcompeted by the primary mitotype – and thereby secures its persistence and further spread as a selfish complete mt-genome, we hereby designate "Spartacus". Finally, we reconstruct the evolutionary history of mitochondrial genomes in the willistoni subgroup uncovering signs of multiple mitotype losses and introgressions. Our data indicate an α-like mitochondrial ancestor in the willistoni subgroup, with the β mitotype likely acquired via introgression from an unidentified donor. We hypothesize that the selfish characteristics of α2 might have emerged as a response to competition for inheritance with the introgressed β mitotype.

INTRODUCTION

It is generally assumed that individuals of most animal species carry a single, maternally inherited mitochondrial haplotype (a mitotype) [1, 2]. However, recent advances in sequencing technology and genomic analyses show that mitochondrial heteroplasmy, the presence of multiple mitotypes in one individual, is found in a number of vertebrate and invertebrate species [3, 4].

Most often, heteroplasmy is the result of either mitochondrial mutations that accumulate over an individual's lifetime, or paternal leakage. In both cases, the additional mitotype is usually transient and persists at low levels within an individual [5]. Most host-mitochondria relationships are ancient, with both parties coevolving over extended periods of time, and disturbances caused by the introduction and persistence of new mitotypes may reduce fitness or cause disease [6, 7]. Stable, long-term heteroplasmy is therefore relatively uncommon and may be evolutionarily disadvantageous (Reviewed in [8]). Accordingly, selection against heteroplasmy has been proposed as an explanation for the evolution of uniparental inheritance in mitochondria [9]. Mechanistically, transmission of male mitochondria and its genome in animals is prevented by both pre- and postmating defense and destruction mechanisms of the host ([5, 10-15] and references therein), and if these fail, the rare paternal mitotype is usually outcompeted by maternal mitochondria during early development or within subsequent generations [16].

The best-known case of animal heteroplasmy is that of freshwater mussels of the family *Unionidae*, which display double uniparental mitochondrial inheritance (DUI). In these species, two different mitochondrial haplotypes, M and F, are transmitted exclusively by males and females, respectively [17, 18]. Stable heteroplasmy is also found in several isopods in which mitochondria encode different tRNA variants [19-21]. Among insects, prevalence (the frequency of heteroplasmic individuals in a population) at various levels (the proportions of different mitotypes within an individual) has been described in natural populations of the bed bug *Cimex lectularius* L. [22], the neotropical ants of the *Ectatomma ruidum* complex [23], the brown dog tick *Rhipicephalus* spp. [24], and the leaf beetles *Gonioctena intermedia* [25]. For the genetic model system *Drosophila*, natural heteroplasmy has been detected in *D. mauritiana* [26], *D. simulans* [27-29], *D. melanogaster* [30], and *D. subobscura* [31, 32]. Slightly higher prevalence and levels have been found in interspecies F_1 hybrids of different closely related *Drosophila* species ([33-35] and references therein). In spite of a growing number of reports on heteroplasmy in animals including humans (recently reviewed in [36]), little is known as to how heteroplasmic systems have evolved and what the functional differences between mitotypes could be (recently reviewed in [37]).

However, when discussing causes and consequences of mitochondrial heteroplasmy, it might be useful to consider mitotypes as primary or secondary endosymbionts, similarly to

91 what is done for bacterial endosymbionts of insects where multiple infections and obligate
92 mutualism are common phenomena. Hence, primary bacterial endosymbionts as well as
93 mitochondria are normally maternally inherited, long-term co-evolved with the host and have
94 a well-defined function that make them obligate for the eukaryotic host cell. Secondary
95 bacterial endosymbionts on the other hand are non-fixed recently acquired or evolved
96 mitotypes that can have variable, often condition-specific and sometimes harmful effects on
97 their hosts [38-40]. While primary endosymbionts might evolve from secondary
98 endosymbionts as a way for the host to secure higher fitness over long evolutionary time,
99 secondary symbionts can be maintained either because they provide a benefit under specific
100 conditions or because they have the ability to cheat the host and be transmitted selfishly
101 without any positive effect [40]. Similarly, secondary mitochondria and/or their genomes might
102 thus be maintained in the host for the same reason and result in heteroplasmy (reviewed in
103 [37]). Additionally, obligate dependencies on multiple endosymbionts evolved in several insect.
104 In such cases, the genome of one symbiont has often lost genes whose function has to be
105 complemented by genes in the other symbiont genome, thus making the host dependent on
106 both symbionts for a function that was previously performed by one [41, 42]. Hence, similarly,
107 stable heteroplasmy might also be maintained if mitotypes with complementary gene sets are
108 present in the same host.

109 The Neotropical *Drosophila willistoni* group has been a model for evolutionary studies
110 since the 1950s [43]. Within it, species belonging to the willistoni subgroup have received the
111 most attention, with *Drosophila paulistorum,* in particular, being the focus of many studies on
112 speciation and hybridization [44, 45]. *D. paulistorum* is a complex of six semispecies under
113 incipient speciation, namely Amazonian (AM), Andean Brazilian (AB), Centro-American (CA),
114 Interior (IN), Orinocan (OR) and Transitional (TR) [46], each displaying partially overlapping
115 geographical distribution and varying degrees of pre- and postmating reproductive
116 incompatibilities with the others ([44], and references therein). All semispecies also carry the
117 bacterial endosymbiont *Wolbachia*, which provides mutualistic functions and causes
118 premating isolation between members of different semispecies pairs [47-51]. The whole
119 willistoni subgroup is a recent radiation and phylogenetic incongruences between *D.*
120 *paulistorum* nuclear and mitochondrial evolutionary histories [52-54] suggest that hybridization
121 and introgression events between semispecies may have occurred in the past [53].

122 In the present study, we discovered and investigated the unusual biparental
123 inheritance of mitochondria resulting in stable and high-level heteroplasmy in *Drosophila*
124 *paulistorum*. We assembled the mitochondrial genomes of six species of the willistoni
125 subgroup and three representatives of the closely related saltans species group. We
126 discovered that *D. paulistorum* lines carry distinct mitotypes from two non-sister clades - α
127 and β - and identified several lines exhibiting stable heteroplasmy. We further determined that
128 both mitotypes can be divided into two subclades, called 1 and 2, where the α2 mitotype is so

129 far observed exclusively in heteroplasmic lines. This finding implies that α2 is unable to
130 provide full mitochondrial functionality, and might therefore be regarded as a secondary
131 mitotype, in contrast to the fully functional primary mitotypes (α1, β1 and β2), which can exist
132 in homoplasmy.

133 In addition, the α2 mitotype was found in 41% of the tested *D. paulistorum* lines, where
134 it is stably maintained at variable but significant levels with up to 66% depending on host
135 background. Unexpectedly, this secondary mitotype, unlike the primary one we investigated,
136 was not distributed within the fly in a way that correlated with the expected OXPHOS-mediated
137 ATP demands of different tissues. Yet, the α2 genome showed no signs of deterioration,
138 suggesting it may perform some unknown function for the host. Additionally, we demonstrated
139 that the α2 mitotype is not integrated as a Mega-NUMT in the chromosomes but a cytoplasmic
140 compound, which is biparentally transmitted and starts replicating rapidly after egg fertilization,
141 while the primary mitotype is still quiescent. This replication advantage, which we name the
142 "Replication precox" phenotype, allows α2 to reach high levels during early embryogenesis
143 and thus avoid being outcompeted by the much larger number of maternally inherited primary
144 mitotypes. Hence, we propose that α2 is a newly discovered selfish mitotype that can persist
145 and spread in and between populations, and possibly even across species by introgression,
146 via the combination of biparental transmission and the Replication precox phenotype.

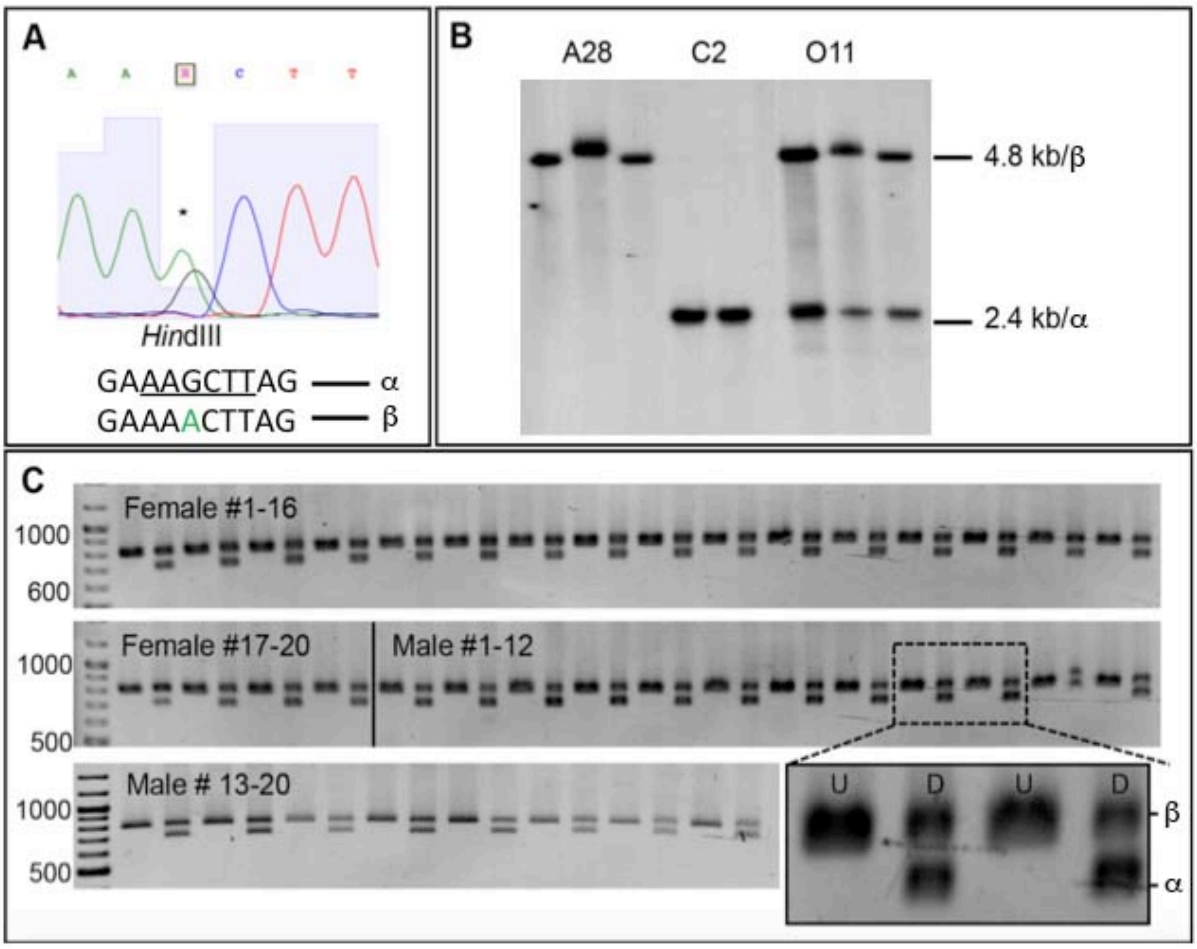
147 Using phylogenetic reconstruction, we observed a complex evolutionary history of
148 mitochondria in *D. paulistorum*, with both ancient and recent introgressions. We suggest that
149 an α-like mitochondrion was the ancestral mitotype of the species and that heteroplasmy might
150 have originated through introgression of the β mitotype from an unsampled or extinct donor,
151 a concept known as 'ghost' introgression [55-57]. We speculate that the unusual biparental
152 transmission mode of α2 might have developed through competition with β mitochondria for
153 maternal transmission.
154

155 # RESULTS

156 **High level heteroplasmy is common in the *D. paulistorum* species complex**

157 In the course of our phylogenetic survey, we sequenced PCR products of the mitochondrial
158 *Cytochrome Oxidase 1* (*CO1*) gene from several *D. paulistorum* lines for typing and found
159 multiple robust double peaks in the resulting chromatograms derived from single flies of the
160 Orinocan O11 line, suggesting heteroplasmy. The double peak shown in **Fig 1A** covers a
161 diagnostic *Hin*dIII restriction site present in one mitotype a but absent in the other. Thus, to
162 independently confirm heteroplasmy, we performed diagnostic RFLP/*Hin*dIII Southern blots
163 on single fly DNA probed with *CO1*. We found that the O11 line is truly heteroplasmic, with
164 individual flies carrying two mitotypes, hereafter called α and β, at approximately similar levels

165 (**Fig 1B**). In contrast, single flies of the Centro-American C2 and Amazonian A28 lines were
166 homoplasmic and carried either the α or β mitotype, respectively (**Fig 1B)**. By using *CO1*-
167 specific RFLP/*Hin*dIII-PCR we further showed that all tested O11 flies (20 females, 20 males)
168 exhibited both the α and β specific bands (**Fig 1C**), which suggested fixation of the high-level
169 heteroplasmic state in this *D. paulistorum* long-term lab line.



170
171 **Fig 1**. High-level heteroplasmy in the *D. paulistorum* Orinocan line O11. **(A)** Chromatogram
172 derived from *CO1* direct Sanger sequencing of one O11 female covering the α-diagnostic
173 *Hin*dIII restriction site (underlined). The asterisk highlights the double peak with Cytosine (C)
174 for α and Thymine (T) for β-mitotype. **(B)** Single-fly *CO1* Southern blot digested with *Hin*dIII.
175 While the Amazonian females of the A28 line (n=3) are homoplasmic for the β-mitotype (4.8kb)
176 and the ones of the Centro-American line C2 (n=2) are homoplasmic for α (2.4kb), all randomly
177 picked Orinocan O11 females (n=3) show high-level heteroplasmy of the α- and β-mitotypes
178 (4.8 and 2.4kb). **(C)** α-diagnostic *CO1* RFLP/*Hin*dIII-PCR screen on O11 single flies (n=20
179 for each sex). Undigested (U) and digested (D) *CO1*-PCR fragments per fly are shown (bottom
180 right in **C**). As all tested single flies (40/40) exhibited clear double bands after *Hin*dIII digestion,
181 hence fixation of the α/β heteroplasmy in the O11 line was assumed. RFLP strategies are
182 presented schematically in **Fig S1**.

183 *Whole-genome sequencing shows widespread heteroplasmy forming two distinct*
184 *mitochondrial clades and four subclades in D. paulistorum*

185 Next, we used Illumina sequencing to screen for heteroplasmy on a genome-wide level in *D.*
186 *paulistorum* and several related species. We sequenced a total of 23 *Drosophila* lines

representing six species of the willistoni group and three species of the closely related saltans group (**Fig 2A, Table S1**). Thirteen *D. paulistorum* lines were sequenced, including representatives of the six classical semispecies and seven recently collected isofemale lines from Brazil and French Guiana (**Table S1**). Five of the sequenced *D. paulistorum* lines (38.5%) were found to be heteroplasmic, i.e., the Orinocan FG103, FG111 and O11 lines, the Amazonian FG572 line and the Andean-Brazilian MS line (**Fig 2B**). Importantly, heteroplasmy was not detected in any other tested neotropical species apart from *D. paulistorum*.

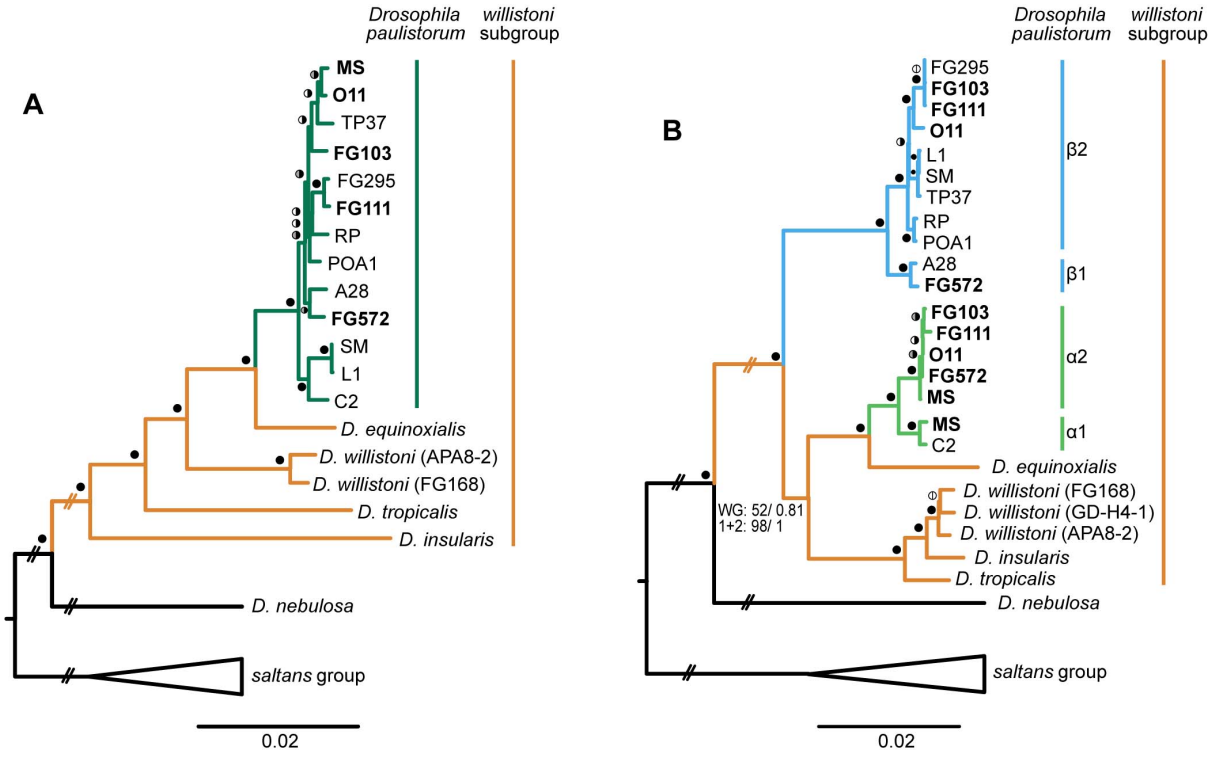


Fig 2. Phylogenies of the sequenced *Drosophila* lines. Trees based on **(A)** the nuclear genes *28S*, *Adh*, *CAD/r*, *Ddc*, *Hb* and *Per*; and **(B)** whole mitochondrial genomes. Fly lines in bold are heteroplasmic. *D. paulistorum* mitochondria split into two distinct non-sister clades: α (green) and β (blue). Support for the branch defining α as sister clade to other species of the willistoni subgroup is low for whole genome (WG) phylogeny, but high when only first and second codon positions of protein coding genes are analyzed (1+2). The same topologies are obtained by maximum likelihood (GTRGAMMA model) and Bayesian analyses. Colored circles next to the nodes indicate ML bootstrap support (left half) and Bayesian posterior probability (right half). Black corresponds to 99-100% bootstrap support or 0.99-1.0 posterior probability, grey represents 70-98% bootstrap support or 0.7-0.98 posterior probability, white indicates support below 70%. Branches leading to the outgroups and marked by two transverse lines were shortened and don't follow the presented scale.

Among all sequenced lines of willistoni and saltans group species, a total of 28 mitochondrial genomes were assembled, one for each homoplasmic *Drosophila* line and two for each of the heteroplasmic *D. paulistorum* lines (see Methods). The genomes are approximately 16.000 bp long, have a GC content around 21% (**Table S2**), and encode the same 13 proteins, 22 tRNAs and two rRNAs as the mitochondria of *D. melanogaster* and most other animals.

214 Importantly, no loss-of-function mutations were found, indicating that all genomes are likely to
215 encode fully functional mitochondria. The assembled mitochondrial genomes also show
216 considerably higher read coverage than nuclear contigs from the same lines, which
217 distinguishes them from potential nuclear mitochondrial DNA sequences (NUMTs) (**Table S3**).
218 All of the assembled mitochondrial genomes have short control regions (CR), between 1.1 kb
219 in the willistoni group and 1.3 kb in the saltans group, compared to the 4.6 kb CR in *D.*
220 *melanogaster*. A few substitutions occur between *D. paulistorum* lines in or near the putative
221 origins of replication, but no major differences in length are observed between CRs (**Table S2**,
222 **Fig S2**).
223 Phylogenetic analysis of the assembled genomes revealed that the mitochondria of *D.*
224 *paulistorum* are polyphyletic and split into two major clades (**Fig 2B**), the α and β mitotypes
225 that were previously detected in single flies of O11 (**Fig 1**). The mitochondrial genomes of the
226 α and β clades are relatively distantly related to each other with an average nucleotide
227 divergence of 2.94%. A majority of the substitutions that occur between them in protein coding
228 genes are synonymous (**Tables S4, S5**), indicating that purifying selection is acting on the
229 genomes of both clades. Based on the phylogeny, the α and β clades can be further divided
230 into two subclades, named 1 and 2 **(Fig 2B**). Interestingly, while the α1, β1 and β2 mitotypes
231 occur both in homoplasmic and heteroplasmic lines, the α2 mitotype is never found in
232 homoplasmy and is present in all heteroplasmic lines, where it is partnered with a mitotype
233 from another subclade.
234
235 *Level and prevalence of heteroplasmy in D. paulistorum lines*
236 To further investigate the heteroplasmy of *D. paulistorum*, we assessed its prevalence (the
237 percentage of heteroplasmic flies within a line) and levels (percentage of each mitotype within
238 a sample) using mitotype-specific RFLP-PCR. Seventeen lines belonging to the six *D.*
239 *paulistorum* semispecies were tested and seven (41%) displayed heteroplasmy (**Table 1**).
240 Apart from the lines already identified as heteroplasmic through genome sequencing (**Fig 2B)**,
241 we further identified heteroplasmy in the two Amazonian lines FG16 and FG572 as well as in
242 the Andean-Brazilian Yellow lines. Interestingly, the Orinocan FG103 line was clearly
243 heteroplasmic at the time of sequencing, shortly after its field collection and the establishment
244 of the isofemale line in 2014/2015 (**Table 1**), but thereafter lost the α2 mitotype in consecutive
245 generations. Heteroplasmic lines were thus observed in the Amazonian, Andean-Brazilian and
246 Orinocan semispecies and originate from both long-term lab lines and recently collected field
247 samples (**Table S1**).
248
249
250

Table 1. Levels of α2 and prevalence of heteroplasmy in *D. paulistorum* lines based on Illumina sequencing of ovary samples (α2 level Illumina) and RFLP-PCR from adult flies (α2 level RFLP**)**. A dash (-) in the α2 level from Illumina sequencing means that we are unable to determine that the α2 mitotype existed. Levels of α2 obtained through RFLP-PCR are presented separately for females (left) and males (right). Prevalence of heteroplasmy among tested individuals (Het. prevalence) is presented both as a percentage and in absolute numbers (in parenthesis). Abbreviations: AB: Andean-Brazilian, AM: Amazonian, CA: Centro-American, Coll: Collection year, Het: heteroplasmy, IN: Interior, ND: Not determined, OR: Orinocan, Spp: Semispecies, TR: Transitional. Further details about the fly lines are provided in **Table S1**.

Line	Coll.	Spp.	Mitotype	α2 level Illumina	α2 level RFLP	Het. prevalence
O11	1957	OR	α2/ß2	19.4%	30.7±3.7/36.8±3.6	100% (39/39)
FG103	2014	OR	α2/ß2	6.5%	ND§	0% (0/40)§
FG111	2014	OR	α2/ß2	2.3%	13.4±1.1/18.3±2.9	25% (10/40)
FG295	2014	OR	ß2	-	0	0% (0/40)
TP37	2009	OR	ß2	-	0	0% (0/20)
POA1	2003	OR	ß2	-	0	0% (0/20)
RP	1995	OR	ß2	-	0	0% (0/20)
A28	1952	AM	ß1	-	0	0% (0/40)
FG16	2018	AM	α2/ß1	ND	57.2±8.6/66.5±7.2	55% (22/40)
FG18	2018	AM	ß1	ND	0	0% (0/40)
FG572	2015	AM	α2/ß1	49.5%	45.3±8.1/63.7±7.0	90% (36/40)
MS	1962	AB	α1/α2	19.1%	21.8±3.1/36.0±4.1	100% (40/40)
Yellow	1960	AB	α1/α2	ND	20.4±4.7/27.4±4.3	100% (40/40)
C2	1954	CA	α1	-	0	0% (0/20)
White	1960	CA	α1	ND	0	0% (0/20)
L1	1958	IN	ß2	-	0	0% (0/20)
SM	1956	TR	ß2	-	0	0% (0/20)

§The line FG103 was heteroplasmic in 2014/15 (determined by Sanger and Illumina sequencing), but has lost the α2 mitotype afterwards (verified by RFLP-PCR and Sanger sequencing in 2018/19).

Whereas heteroplasmy is fixed in the three independent long-term lab lines O11, MS and Yellow, α2 prevalence ranges between 25% and 90% in the recently collected isofemale lines FG111, FG16 and FG572 (**Table 1**). Levels of α2 were high within individual flies of most lines (**Table 1, Fig S3)** and appear to be associated with the nuclear background of the host. The three heteroplasmic lines of the Orinocan semispecies, FG103, FG111 and O11 displayed different α2 levels (**Table 1**). However, the two lab lines of the Andean-Brazilian semispecies (MS, Yellow) showed levels of α2 that were similar to each other at 20%. Finally, the highest α2 levels (45-60%) were detected in the heteroplasmic lines of the Amazonian semispecies, FG16 and FG572 (**Table 1, Fig S3**). In addition, we found that levels of α2 were significantly higher in male flies compared to females in all tested lines except for FG111 ($p<0.001$ for FG572, MS, $p<0.01$ for Yellow and FG16, $p<0.05$ for O11, Tukey HSD test, **Fig S3; Supplemental data file**). However, since α2 levels were measured as a ratio between mitotypes, it is not clear if the higher level in males results from fluctuations in copy numbers of α2, of the primary mitotype (α1, β1 or β2), or of both.

280

281 **Titers of α2 and β2 mitotypes in *D. paulistorum* tissues**

282 To explore the dynamics and potential function of α2 and β2 mitochondria, we measured the
283 genome copy numbers of each mitotype relative to nuclear genome copies (from here on
284 referred to as titer) in the heteroplasmic O11 line of the Orinocan semispecies. Since
285 mitochondrial DNA-copy numbers are variable in different tissues and under different
286 conditions, their location and titer can provide clues as to their function. Hence, besides whole
287 flies we also investigated α2 and β2 titers in the head, thorax, abdominal carcasses without
288 gonads and gonads of male and female O11 flies using discriminant qPCR. We also explored
289 the potential influence of sex, age (**Fig 3**; **Fig S4**), and also infection with the endosymbiont
290 *Wolbachia* (**Fig S5**; **Supplemental data file**) on titers of both mitotypes. We found that titers
291 of the primary β2 mitotype (**Fig 3A**) were significantly influenced by tissue ($F_{4,80}$=25.44,
292 p<0.001). As expected, β2 is found in higher titers in energy-demanding tissues where ATP is
293 provided by oxidative phosphorylation (OXPHOS) such as in the flight muscle of the thorax
294 ([5], and references therein). Further, we found a weak significant interaction between age
295 and sex ($F_{1,80}$=6.12, p<0.05) potentially driven by younger male flies having slightly more of
296 the β2 mitotype than older males, in combination with different tissue location patterns
297 between the sexes. Females had overall lower numbers of the primary β2 mitotype per nuclear
298 genome copy than males ($F_{1,80}$=28.68, p<0.001).

299 In contrast to β2, we found no significant effect of fly tissue ($F_{4,80}$=1.426, p=0.23) or
300 age ($F_{1,80}$=0, p=0.99) on the titer of α2. Additionally, there was no significant effect of
301 *Wolbachia* infection on the titer of either the β2 ($F_{1,80}$=1.41, p=0.24) or the α2 ($F_{1,80}$=0, p=0.97)
302 mitotypes (**Fig S5**). The copy number of α2 is, however, significantly affected by fly sex (**Fig**
303 **3B**, $F_{1,80}$=51.34, p<0.001), with males carrying approximately twice as many α2 copies as
304 females per nuclear genome. Significant higher titers of α2 in O11 males corroborate our
305 finding of generally higher α2 levels in heteroplasmic males compared to heteroplasmic
306 female flies (**Table 1; Fig S3** and **Supplemental Data file**), confirming that the significant
307 differences between the sexes in the percentage data are due to more copies of α2, and not
308 for example, lower β2 numbers.

309 Although not fully conclusive, the fact that only the titers of the β2 mitotype varied
310 depending on fly tissue, but not the ones of α2, is in line with our other results suggesting that
311 β2 is the functional and hence the primary mitotype and that α2 might be dysfunctional and/or
312 not able to generate sufficient energy for its host.

313

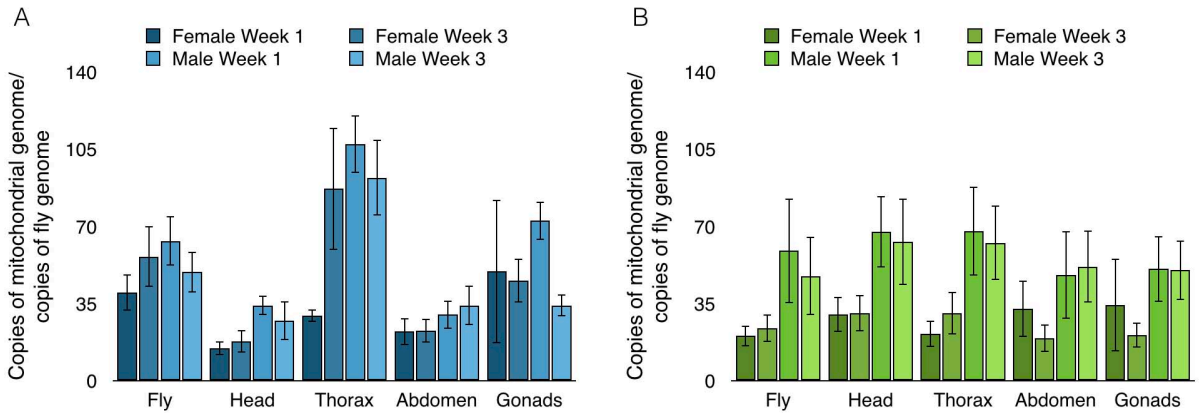


Fig 3. Absolute copies of mtDNA from β2 and α2 mitotypes in O11 *D. paulistorum* flies, relative to fly nuclear genome copy. Bar charts present (**A**) β2 and (**B**) α2 mitotype in male and female flies, dissected tissues and across two time points (1 and 3 weeks). Three biological repeats are plotted with error bars of 1 standard error. (**C**) Shows the level of β2 (blue) and α2 (green) mitotype DNA at 1 week. Relative ratios of both mitotypes are shown in **Fig S4** and data of absolute mt-copies including *Wolbachia*-knockdown specimens in **Fig S5** and **Supplemental data file**.

Biparental transmission and levels of the α2 mitotype in intra- and inter-semispecies hybrids of *D. paulistorum*

We used *CO1* RFLP-PCR to investigate if paternal transmission plays a role in the maintenance of high level heteroplasmy in *D. paulistorum*. Relative levels of α2 in F_1 offspring were measured in various crosses involving homoplasmic females and heteroplasmic males in intra- and inter-semispecies combinations. Semispecies designation of the tester lines was determined by compatibility crosses (**Table S6**). By convention, the maternal line is designated first in the cross. By this, we uncovered unusual high levels of paternal transmission of α2. In intra-semispecies crosses, all tested F_1 flies in the FG295 (β2) x O11 (β2 and α2) cross (40/40, **Fig 4A**) and 57.5% flies in the A28 x FG572 cross (23/40, **Fig 4B**) were heteroplasmic (Supplemental data for Fig 4). F_1 female hybrids from FG295 x O11 cross had significantly lower α2 levels than O11 females (parental line) ($p<0.001$, Tukey HSD test), whereas the F_1 male hybrids from the same cross as well as both F_1 sexes from the A28 x FG572 cross did not significantly differ in levels from their parental ones of the same sex ($p=0.99$ for former and $p=0.06$ and $p=0.46$ for the latter two, respectively). As heteroplasmy is not fixed in the FG572 line (**Table 1**) paternal introgression in A28 was also incomplete (**Fig 4B).**

In inter-semispecies crosses between Amazonian and Orinocan flies all tested F_1 flies of A28 (β1) x O11 (40/40, **Fig 4C**) and 45% flies of FG18 (β1) x O11 (18/40, **Fig 4D**) were heteroplasmic. In both combinations, the α2 mitotype levels of F_1 flies of both sexes were significantly lower than in the paternal O11 line ($p<0.001$ for both crosses, Tukey HSD test).

343 Combining data from all four crosses (**Fig 4 A-D**), paternal transmission of α2 was found in
344 76% of F_1 hybrids (121/160). This finding clearly demonstrates that the α2 mitotype of *D.*
345 *paulistorum* are not only sporadically leaking but are highly transmissible via males. To
346 independently confirm these results, we tested only the heteroplasmic F_1 hybrids, not the
347 homoplasmic ones, for their α2 levels using a second mitochondrial locus (*CO2*) and observed
348 patterns similar to *CO1* (**Fig S6**).

349 Additionally, we investigated whether inheritance of the paternal mitotype is α2-
350 specific. For this, we used mitotype-specific RFLP-PCR or Sanger sequencing on F_1 hybrid
351 flies generated by eight different cross combinations aimed at tracing paternal transmission of
352 either α1 (number of F_1 flies assayed = 59), β1 (31), or the β2 (76) mitochondria. Paternal
353 transmission was not detected in any of the 166 tested F_1 flies (**Table S7**). Hence, a high level
354 of paternal transmission is unique for α2 mitotype.

355 We verified our findings using highly sensitive discriminant qPCR on A28 x O11 inter-
356 semispecies crosses. We assessed quantities of the β1, β2 and α2 mitotypes in O11 males
357 (α2 and β2), A28 females (β1) and F_1 hybrid offspring. As expected, the pure parental lines
358 possessed only the expected mitotypes (**Fig 4E, F**). Despite robust paternal inheritance of α2
359 in all tested F_1 hybrids (16/16), the hybrid offspring do not show any significant differences in
360 titers for any mitotype when compared to their homoplasmic mothers, which is likely due to
361 low α2 titers in the hybrids and relatively few biological replicates. However, significantly
362 different titers for all mitotypes were seen when F_1 hybrids are compared to their heteroplasmic
363 fathers **(Fig 4E, F, G**), i.e., higher β1, lower β2 and lower α2 titers are seen in F_1 hybrids
364 compared to the paternal line. In contrast to the less sensitive RFLP-PCR method, our
365 discriminant qPCR method allowed us to detect very low levels (~0.1-0.3%) of paternal β2
366 leakage in 11 out of 16 hybrid offspring (**Supplemental data file**).

367 As expected for an inter-semispecies cross [44, 45], all F_1 hybrids were sterile and did
368 not produce any live larvae when crossed either to siblings or backcrossed to parental lines
369 (**Table S8**). This prevented exploration of the mitotype dynamics in further generations from
370 this combination of parental lines. We were also unable to explore the dynamics of the reverse
371 cross (female O11 with male A28), and thereby test if the β1 mitotype could be paternally
372 transmitted, as none of the crosses we set up (0/200 attempts) produced larvae.

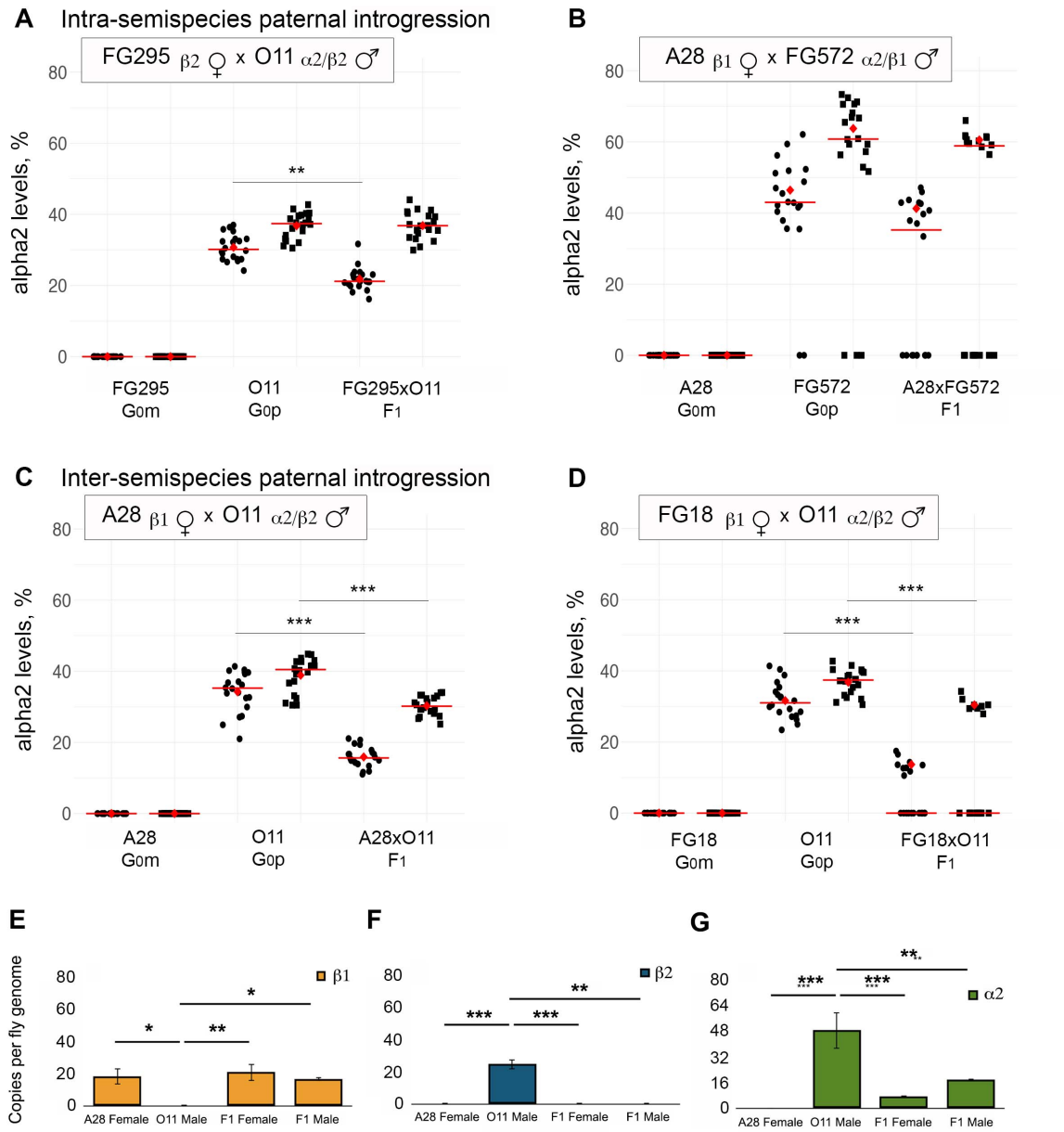


Fig. 4. Levels of the paternal α2 mitotype in intra- and inter-semispecies hybrids of *D. paulistorum*. Paternal inheritance of the α2 mitotype in different *D. paulistorum* hybrids determined by RFLP-PCR **(A-D)**. Scatter plots present relative α2 levels in single fly F_1 hybrids derived from homoplasmic females (circle) and heteroplasmic males (square) of **(A, B)** intra-semispecies (FG295 x O11; A28 x FG572) and **(C, D)** inter-semispecies crosses (A28 x O11; FG18 x O11). Relative levels in F_1 hybrids (F_1 circle, F_1 square) were calculated based on *Hind*III-RFLP using *CO1* locus, and compared to the maternal (G_0m fem, G_0m mal) and paternal (G_0p fem, G_0p mal) lines. Single values are plotted including median (red crossbar) of all individuals and mean (red rhombus) of heteroplasmic flies (n=20 per sex, flies were 7-10 days old). Significance of difference between α2 levels was estimated by ANOVA (p<0.05*, p<0.001***). Paternal α2-inheritance determined by discriminant qPCR (**E-G**). Number of copies of β1 (**E**), β2 (**F**) and α2 (**G**) mitochondrial genomes in 3-weeks old O11, A28 parents and F_1 hybrids relative to fly nuclear genome copy. Data was generated from crosses from single virgin males and single virgin female flies. N=8 in all cases and error bars are 1 standard error. Post-hoc significances are shown as *** = p<0.001, ** = p<0.01, * = p<0.05.

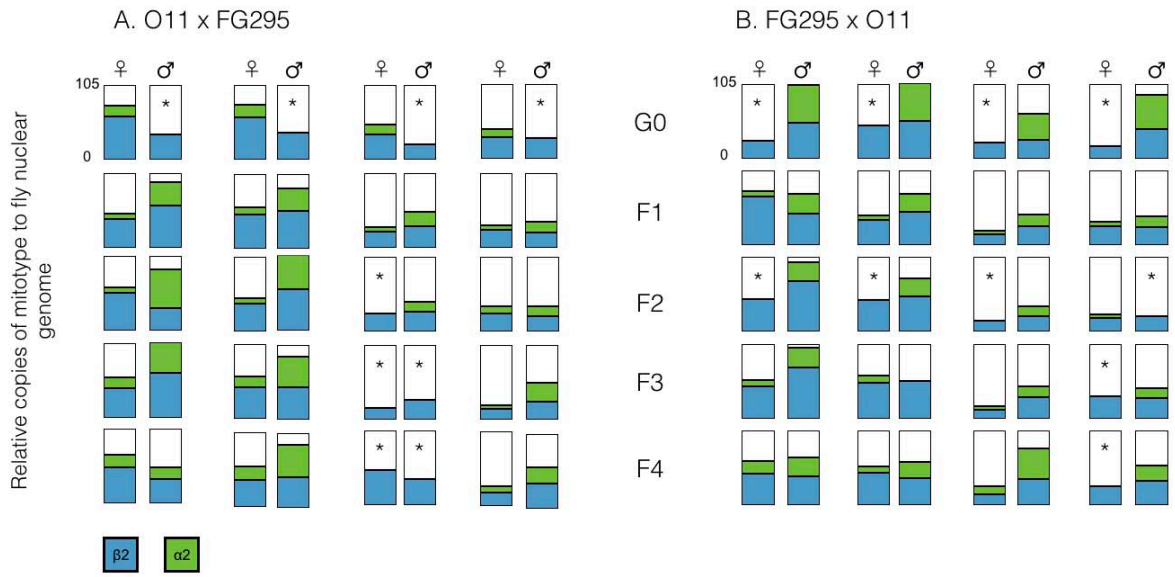
389 **High invasiveness and stable persistence of the α2 mitotype over multiple**
390 **generations**

391 After identifying that α2 was the only mitotype paternally transmitted in high frequency to F_1
392 flies, we investigated persistence of heteroplasmy over further generations. We used
393 discriminant qPCR to follow the inheritance patterns of the α2 mitotype during four generations
394 of fertile intra-semispecies FG295 (homoplasmic β2) and O11 (heteroplasmic α2 and β2)
395 hybrid crosses. We followed four single-fly G_0 crosses in each direction (FG295 x O11 and
396 O11 x FG295), and mated a randomly picked pair of virgin siblings to each other in each
397 consecutive generation (F_1-F_4) to explore single-fly inheritance patterns of the α2 mitotype
398 (**Fig 5**). The results showed both maternal (**Fig 5A**) and paternal (**Fig 5B**) transmission of the
399 α2 mitotype, as well as high α2 invasiveness (ability to invade a homoplasmic line) and
400 persistence (ability to be maintained over generations) in F_4 flies (13/16).

401 In both crossing directions, all F_1 hybrids (16/16) were heteroplasmic, hence
402 demonstrating not only paternal but biparental transmission of α2. However, in the following
403 generations (F_2-F_4), transmission success was more varied. All tested offspring from three of
404 the four crosses starting with a heteroplasmic mother carried both α2 and β2, and one cross
405 lost the heteroplasmy at F_3 (**Fig 5A**), whereas the crosses starting with a heteroplasmic father
406 showed more stochasticity in transmission, with the α2 mitotype not necessarily transferred to
407 both tested offspring in these later generations (**Fig 5B**). However, more individuals need to
408 be tested in order to confirm this pattern.

409 In agreement with earlier results, we found that there were differences in mitotype titers
410 between heteroplasmic and homoplasmic flies, which are dependent on sex. While titers of
411 the β2 mitotype were not significantly different between hybrid males and females across the
412 experiment (Wilcoxon rank sum test W=866, p=0.29), male hybrids had higher titers of the α2
413 mitotype compared to hybrid females (Wilcoxon rank sum test W=1208, p<0.001). Overall,
414 heteroplasmic flies also carried a higher total mitochondrial DNA load than homoplasmic flies
415 (Wilcoxon rank sum test W=183, p<0.001). After separating the data by the direction of the
416 cross and the sex of the fly, we found that heteroplasmic sons had significantly higher total
417 mitotype titers than their homoplasmic parents (Wilcoxon rank sum test W=5, p<0.05 for
418 female parents; W=2, p<0.01 for male parents), while heteroplasmic daughters do not
419 (Wilcoxon rank sum test W=10, p=0.138; W=12, p=0.13 for male parents), demonstrating the
420 noticeably higher titers in male heteroplasmic flies specifically. In contrast, heteroplasmic
421 fathers have significantly higher total mt-DNA titers than their heteroplasmic sons and
422 daughters (sons W=47, p<0.05; daughters W=43, p<0.05), whereas the difference in titer
423 between heteroplasmic mothers and their heteroplasmic offspring is non-significant for either
424 sex (sons W=19, p=0.382; daughters W= 38, p=0.202).

425 Given that we only observe significant variability in the α2 titer between sexes as well
426 as between parents and their offspring (no significant difference was observed for any test on
427 β2), the higher titer of mt-DNA in heteroplasmic flies overall suggest that the presence of α2
428 increases the number of total mitochondrial DNA copies per cell and is added on top of the
429 primary mitotype, i.e., β2. The reduction in total mt-DNA titer between heteroplasmic fathers
430 of the O11 line and their heteroplasmic offspring of both sexes cannot be explained by the
431 significantly different α2 titers observed between male and female flies. Instead, this result
432 indicates an interaction between mitotype titer and host nuclear background, which is more
433 pronounced in males.



434

435 **Fig 5. Presence and titer of α2 and β2 in (A) O11 x FG295 and (B) FG295 x O11 hybrids.**
436 Relative copies of α2 to number of fly nuclear copies is shown in green. Relative copies of β2
437 to number of fly nuclear copies is shown in blue. Values were rounded up if the decimal was
438 over 0.5 and rounded down if under 0.5. The figure depicts four biological repeats of each
439 direction of the hybrid cross, each representing a single-pair mating of a male and female fly
440 for each line. A mated pair derived from randomly picked virgin flies of each generation was
441 tested for four generations after the initial mating and copy numbers of α2 and β2 mitotypes
442 were quantified through discriminant qPCR. ∗ means that α2 was below the cutoff threshold
443 on qPCR and the fly is considered homoplasmic.
444

445 To further investigate whether fly nuclear background affects the persistence of the α2
446 mitotype we backcrossed virgin F_1 hybrids from O11 x FG295 (by convention, the maternal
447 line is designated first) to the homoplasmic FG295 line (**Fig 6A)**, thereby increasing the
448 nuclear background that coevolved with the homoplasmic β2 mitotype in each generation. As
449 before, all F_1 flies acquired the α2 mitotype when transmitted maternally (**Fig 6A**, left column),
450 but their respective levels were lower than in the G_0 parent flies (p<0.001, Tukey HSD test).
451 Backcrossing these F_1 flies to the homoplasmic FG295 line (BC1, maternal introgression)
452 showed a decrease in α2 prevalence in the offspring (**Table S9**). As in F_1 flies, the levels of

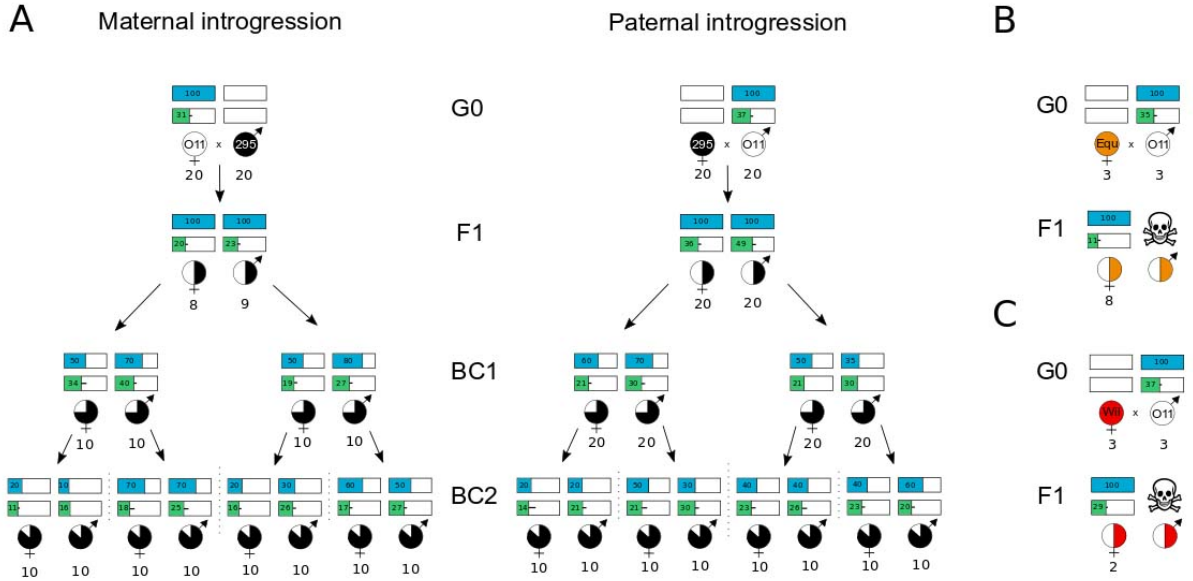
453 α2 in offspring from the hybrid parents were significantly lower than in G_0 flies (p<0.001, Tukey
454 HSD test). A second backcross of the BC1 offspring to the homoplasmic FG295 line revealed
455 that α2 prevalence decreased in the BC2 offspring from the BC1 hybrid mothers, whereas
456 their paternal counterparts kept a similar prevalence as that observed in BC1. Overall α2
457 levels in BC2 flies were significantly lower in both sexes in all backcrossing directions (p<0.001,
458 Tukey HSD test) compared to their parental levels (G_0 females - Mean=44.5, SD=4.5; G_0
459 males - Mean=52.4, SD=4.8).

460 Again, all F_1 flies also inherited the α2 mitotype when initially transmitted paternally
461 (**Fig 6A**, right column). Compared to G_0 flies of the same sex, the levels of α2 were significantly
462 lower in F_1 females but not significantly different in F_1 males (p<0.01 and p=0.56, respectively,
463 Tukey HSD test). Backcrossing of these F_1 flies to the FG295 line (BC1, paternal introgression)
464 showed a decrease in α2 prevalence in the offspring when transmitted both maternally and
465 paternally (**Table S9**). Compared to the F_1 flies of the same sex, the levels of α2 in all BC
466 offspring from both hybrid parents decreased to approximately 20% in females and 30% in
467 males (p<0.001, Tukey HSD test).

468 A second backcross of the BC1 offspring to the FG295 line revealed a pattern similar
469 to the one of the reciprocal O11 x FG295 cross. The prevalence of α2 was low in the
470 backcrossed offspring derived from hybrid mothers whereas the flies from hybrid fathers had
471 a prevalence similar to that observed in BC1. The only exception was the offspring obtained
472 from mothers derived from the F_1 hybrid fathers (40% for females and 40% for males). By
473 increasing the FG295 nuclear background, all BC2 offspring showed significantly lower α2
474 levels (p<0.001, Tukey HSD test; see **Table S9**) compared to the levels in their heteroplasmic
475 O11 parents of the same sex. In summary, we observed that backcrossing heteroplasmic F_1
476 hybrid flies from both crossing directions to a homoplasmic β2 line not only lowers relative α2
477 levels but also results in a decrease in heteroplasmy prevalence in further generations. These
478 data indicate that the α2 mitotype can successfully invade previously homoplasmic lines,
479 although its transmission success is affected by the host nuclear background.

480 Additionally, we investigated the capacity of α2 to invade other species by performing
481 crosses between females of *D. equinoxialis* and *D. willistoni* and males of *D. paulistorum* O11
482 line (**Fig 6B, 6C**). Since hybrid flies from these interspecies crosses are fully sterile, we could
483 not assess α2 prevalence, persistence and levels beyond F_1 flies. Both cross combinations
484 showed 100% prevalence of the α2 mitotype in F_1 offspring, although due to the high
485 embryonic lethality of hybrids we were only able to analyze eight F_1 flies in the *D. equinoxialis*
486 x O11 cross (**Fig 6B**) and two F_1 flies in the *D. willistoni* x O11 cross (**Fig 6C**). In agreement
487 with earlier experiments [58], no F_1 males were obtained in either interspecies crosses, thus,
488 only females were analyzed for mitotype levels. Similar to results derived from *D. paulistorum*
489 inter-semispecies crosses (**Fig 4**) F_1 females of *D. equinoxialis* x O11 cross (**Fig 6B**) had

490 significantly lower α2 levels in comparison to G_0 flies (11%, $p<0.001$, Tukey HSD test). α2
491 levels in the two F_1 females of *D. willistoni* x O11 cross were also decreased, but due to the
492 extremely low number of viable F_1 hybrid flies (**Fig 6C**) we were not able to analyze
493 significance of this result.
494



495

496 **Figure 6. Prevalence and levels of the α2 mitotype introgressed in different genetic**
497 **backgrounds. (A)** Dynamics of α2 prevalence (blue) and levels (green) determined by α-
498 diagnostic *CO1*/*Hin*dIII RFLP in females and males by backcrossing F_1 flies derived from O11
499 x FG295 (left: maternal introgression; white x black) and FG295 x O11 (right: paternal
500 introgression; black x white) parents to FG295 flies. Arrows in this figure represent a backcross
501 to homoplasmic (β2) FG295 flies. (**B, C)** Dynamics of α2 prevalence (blue) and levels (green)
502 determined by the α2-diagnostic CO2/*Msp*I RFLP in F_1 flies from interspecies crosses of **(B)**
503 *D. equinoxialis* (orange) and **(C)** *D. willistoni* (red) females to heteroplasmic *D. paulistorum*
504 O11 males (white). The skull symbolizes complete absence of F_1 males in both crosses. Lines
505 inside the green bars show standard deviation. Values below sex symbols indicate the number
506 of single flies used in RFLP-PCR analysis and listed in **Table S9**.
507

508 **α2 dynamics during embryogenesis in heteroplasmic lines and F_1 hybrids uncover their**
509 **biparental transmission mode**

510 As demonstrated by our crossing experiments, the biparentally inherited α2 mitotype of *D.*
511 *paulistorum* is highly invasive and maintained at consistently high levels in the adult fly. Since
512 the number of mitochondrial DNA copies per sperm cell is much smaller than in a mature egg
513 [59], and paternally leaked mitochondria and their genome are usually both passively diluted
514 and actively destroyed [15], we asked how paternally inherited α2 mitotype avoids depletion.
515 By using RFLP-PCR, we analyzed α2 levels in unfertilized eggs and embryos from the O11,
516 FG572 and MS lines (**Fig 7A-C**) as well as F_1 embryos from reciprocal crosses between O11
517 and FG295 flies (**Fig 7D, E**). The α2 levels at different stages of embryogenesis were

518 compared to those in unfertilized eggs (E0) and parental flies (G_0f - females, G_0m - males).
519 Strikingly, none of the five lines analyzed showed traces of α2 in unfertilized eggs. The finding
520 is not surprising for the F_1 flies from FG295 x O11 cross (**Fig 7E**), where the eggs come from
521 homoplasmic FG295 mothers with only the β2 mitotype, but is unexpected for mothers of the
522 four heteroplasmic lines (**Fig 7A-D**) that carry high α2 levels (**Table 1**). We infer that α2 levels
523 are very low in unfertilized eggs of heteroplasmic mothers and below the detection level of our
524 RFLP-PCR method. Immediately after fertilization (E1; 0-30 minutes after egg laying - AEL),
525 however, α2 reached relatively high levels (similar or close to parental) in F_1 embryos of all of
526 the lines except for the ones obtained from the maternal O11 x FG295 cross (**Fig 7D**). For
527 O11 and FG572 E1 embryos (**Fig 7A, B**) median α2 levels were not significantly different from
528 the levels of either parent (p>0.05; here and further on ANOVA with post-hoc Tukey HSD test),
529 except for FG572 males (p<0.01; Supplemental data of **Fig 7**). Median α2 levels of E1
530 embryos from MS, FG295 x O11 and O11 x FG295 (**Fig 7C-E**) were significantly lower than
531 in parents (p<0.05). In mid-embryogenesis (E2; 3-4 hours AEL) α2 levels of O11 were only
532 significantly decreased compared to E1 (p<0.001; **Fig 7A**), whereas in the four other lines (**Fig**
533 **7B-E**) α2 levels continued to increase (statistically non-significant in FG572, MS embryos and
534 F_1 of O11 x FG295, p>0.05; but significant in F_1 of FG295 x O11 crosses, p<0.001). In late
535 embryogenesis (E3; 18-20 hours AEL) all lines showed a significant increase in α2 levels
536 compared to E2 (p<0.001) except for FG572 and F_1 embryos from the maternal O11 x FG295
537 cross shown in **Fig 7B** and **D** (p>0.05). Only in F_1 embryos from the maternal O11 x FG295
538 cross α2 levels at E3 were significantly lower in comparison to parents (9.7%; p<0.001). In all
539 other cases the level of α2 in E3 embryos reached similar or even significantly higher levels
540 compared to their respective heteroplasmic parents (p<0.001 for G_0f in O11 and MS as well
541 as G_0m in MS; p<0.05 for FG295 x O11; p>0.05 for G_0f in FG572, FG295 x O11 and G_0m in
542 O11 and FG572).

543 Additionally, we compared α2 levels of embryos to the levels in emerged flies (F_1f and
544 F_1m, **Fig 7**). There were no significant differences in α2 levels between G_0 and F_1 flies of the
545 same sex except for F_1 flies from the intra-semispecies crosses O11 x FG295 and FG295 x
546 O11 (p<0.01). In comparison between F_1 flies and embryos of heteroplasmic lines we found
547 similar statistically significant differences as between embryos and G_0 flies (see
548 **Supplemental data** of **Fig 7**). F_1 flies from both O11 x FG295 and FG295 x O11 crosses (**Fig**
549 **7D, E**) had significantly higher α2 levels than at any embryonic stages (p<0.05), except in late
550 embryos (E3) from the paternal FG295 x O11 cross with no significant difference to F_1 males
551 (p=0.99), but significantly higher α2 levels than F_1 females (p=0.002). Strikingly, a comparison
552 of the maternal (**Fig 7D**) and paternal (**Fig 7E**) α2 dynamics between embryonic stages (E1-
553 E3) and F_1 adult flies (**Fig 7D, E**) uncovered significant differences in α2 levels depending on
554 the mode of transmission (p<0.001 for all embryonic stages and F_1 flies). Hence during *D.*

paulistorum embryogenesis, the paternally introgressed α2 mitotype is replicating faster and possibly earlier than α2 that were transmitted by the female.

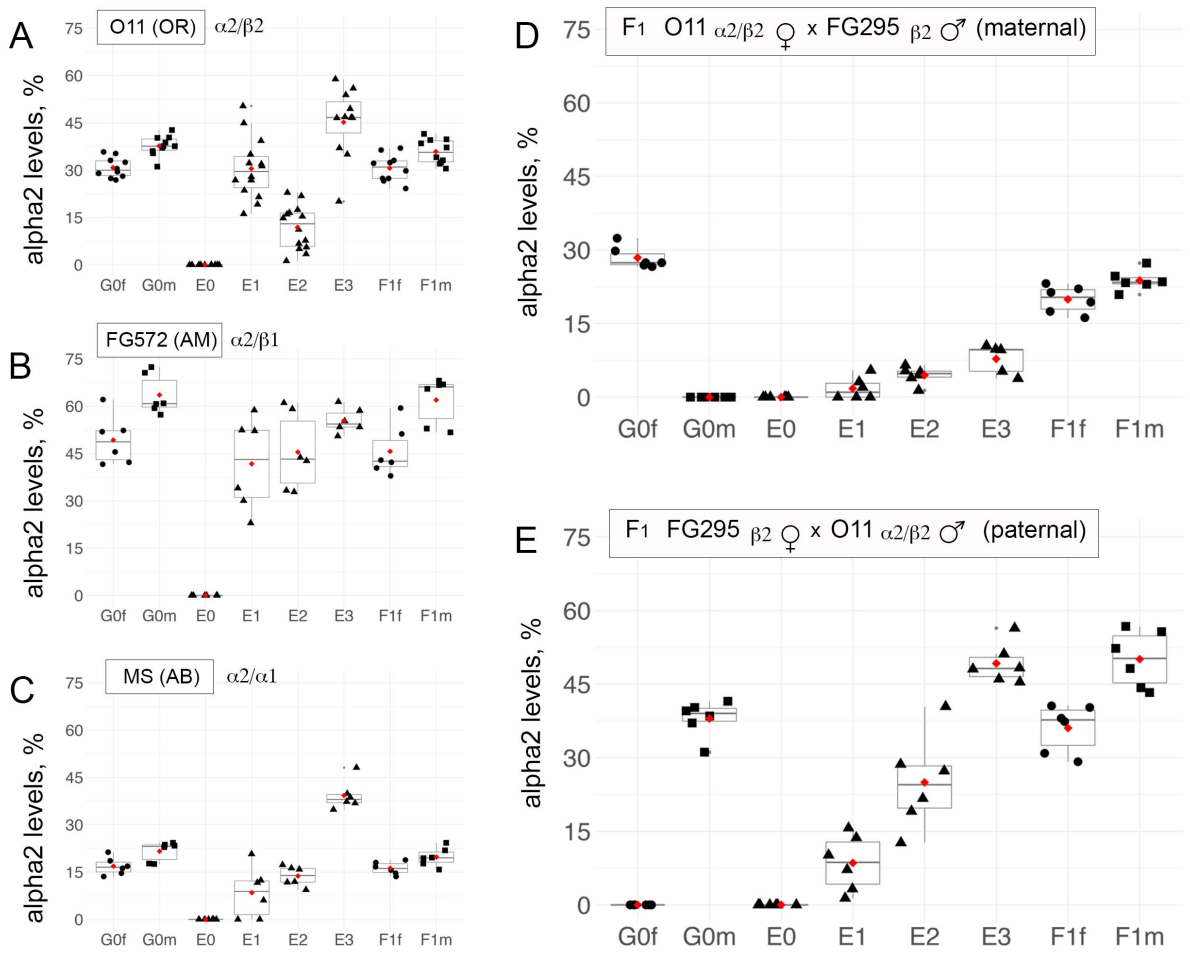


Fig 7. Relative α2 levels during *D. paulistorum* embryogenesis determined by *CO1*/*Hind*III (A, B, D, E) and CO2/*Msp*I (C) RFLP-PCR. Scatter dot plots combined with box plots show levels of α2 (%) in the heteroplasmic lines O11 **(A)**, FG572 **(B)**, and MS **(C)** and α2 levels in F_1 hybrids from maternal **(D)** and paternal **(E)** introgressions obtained from reciprocal crosses between parental heteroplasmic O11 and homoplasmic FG295 flies. Dynamics of α2 levels are presented in adult parental females (G_0f) and males (G_0m), unfertilized eggs (E0), embryos of 0-30 min (E1), 3-4 hrs. (E2), and 18-20 hrs. (E3) of development at 25°C and F_1 adult females (F_1f) and males (F_1m). Each dot represents the α2 level in percent determined from a pool of 20 staged embryos collected for each sample. Females are shown as circles, males as squares and embryos as triangles. Mean values for every stage and sex are depicted with red rhombus.

Collectively, these data strongly suggest that α2 arrives in very low copy numbers from both sexes to the mature egg and, in order to avoid out-dilution by the primary mitotype, starts replicating immediately after fertilization. This strategy works regardless of its transmission mode and thereby ensures that α2 already reaches high levels in mature embryos (E3), equivalent to levels of the adult fly. As the strict maternally-inherited primary mitochondria of

577 *D. melanogaster* and other animals are quiescent and do not replicate during early
578 embryogenesis [13, 60], this early replication behavior is unique for α2 and we hence name it
579 "Replication precox" (see Discussion).
580
581 **Genomic features unique to the α2 mitotype**
582 The unusual properties of the α2 mitotype prompted us to search for unique genomic features
583 that distinguish this secondary mitotype from the other *D. paulistorum* mitotypes (**Table 2, 3**).
584 However, only three non-synonymous substitutions were found to be exclusive to the α2
585 mitotype (**Table 2**).
586
587 **Table 2. Mutations unique to the α2 clade among protein-coding genes of the sampled**
588 ***D. paulistorum* lines**. Three non-synonymous substitutions are unique to α2. Among these,
589 only the H -> N substitution in COIII results in a change in amino acid property. Amino acid
590 and nucleotide changes are presented from other mitotypes to α2. Nr = number.
591

Gene	Nr non-synonymous	Nr synonymous	Amino-acid change	Chemical properties
ATP6	0	3		
ATP8	0	1		
COI	0	5		
COII	0	1		
COIII	1	0	H -> N	Positive charge to neutral
CYTB	0	2		
ND1	0	2		
NADH2	0	3		
ND3	1	0	V -> M	No change
ND4	1	3	V -> I	No change
ND4L	0	0		
ND5	0	4		
ND6	0	4		

592
593 One of the substitutions is found in *ND3* and causes a replacement of Valine with Methionine
594 (aa position 18 in *D. paulistorum*, 21 in humans) and another substitution in *ND4* causes a
595 replacement of Valine with Isoleucine (aa position 379 in *D. paulistorum*, 391 in humans). Both
596 of these replacements are between chemically equivalent amino acids and likely have little
597 functional impact on the proteins. However, the substitution in *CO3* causes the replacement
598 of a positively charged Histidine by a neutral Asparagine in α2 (aa position 7 in *D. paulistorum*,
599 6 in humans) and is predicted to have medium functional impact according to the
600 MutationAssessor server (http://mutationassessor.org/r3/). An inspection of multi-species
601 alignments in Uniprot and Pfam revealed that Asparagine is rarely found at this position of the
602 *CO3* protein in other animals. The only organisms in the database with Asparagine in the
603 same position are a few species of mites, representatives of some invertebrate marine phyla,
604 and a wide variety of mollusks. Interestingly, among the mollusk sequences, many are from
605 the F mitotype of heteroplasmic unionid mussels with doubly uniparental inheritance (DUI).

606 Additionally, the selfish α2 mitotype differs from the primary mitotypes of *D. paulistorum* by
607 four A to G transitions in the large rRNA gene and a T to A transversion in tRNA-Ala and a 3
608 bp insertion in tRNA-Leu2 (**Table 3**), both of which are located in the T-loop.
609

610 **Table 3. Mutations unique to the α2 subclade among non-protein-coding genes in the**
611 **sampled *D. paulistorum* lines.** Substitutions are found in the large rRNA and tRNA-Ala while
612 an insertion occurs in the tRNA-Leu2. In both tRNAs, the mutations are located in the T-loop.

Gene	Number of mutations	Nucleotide changes
lrRNA	4	A -> G (4x)
tRNA-Ala	1	T -> A
tRNA-Leu2	1	- (β) / AT (α1) -> ATATT

613

614 Since the nuclear-encoded protein Elongation factor Tu (EFTu) is known to interact specifically
615 with the tRNA T-loop during translation [61, 62], we took a closer look at this protein in *D.*
616 *paulistorum*. Two divergent copies of EFTu are encoded in the nuclear genome of *D.*
617 *melanogaster*, *mEFTu-1* and *mEFTu-2*. The two genes have largely sex-specific expression
618 in adults [63] and the proteins differ somewhat in binding specificities to tRNAs with and
619 without T- and D-arms [64]. Using our previously assembled transcriptomes of A28, C2 and
620 O11 [50] we identified orthologs of the two *D. melanogaster mEFTu* genes in *D. paulistorum*.
621 An alignment of the two proteins together with orthologous proteins from all *Drosophila*
622 species found in Flybase (release 2019_04) showed that mEFTu-2, the gene with testis-
623 specific expression, has unique amino acids in two positions in *D. paulistorum*. Interestingly,
624 both positions are close to the C-terminal end of the protein, which is the region that interacts
625 with the T-stem of the tRNA [64], suggesting a potential mito-nuclear coadaptation. However,
626 since only one of the three sequenced *D. paulistorum* lines currently carries the α2 mitotype,
627 and we have no data to say whether the other two, A28 and C2, can carry or have ever carried
628 the α2 mitotype, no firm conclusions can be drawn.
629

630 **The evolutionary history of mitochondria in *D. paulistorum* and the *willistoni***
631 **subgroup**

632 Well-supported phylogenies can provide useful insights into the evolutionary history of genetic
633 elements. Although our whole-genome mitochondrial phylogeny showed very high support in
634 almost all internal nodes, the node defining the relationship between α, β and the clade
635 containing *D. willistoni*, *D. insularis* and *D. tropicalis* (hereafter referred to as the willistoni
636 clade) has low support (**Fig 2B**). Hence, in order to confirm the topology of our mitochondrial
637 phylogeny, we investigated the cause of the poorly supported node by testing whether
638 recombination between clades or bias in nucleotide composition could be affecting the
639 phylogenetic signal.

640 *Nucleotide composition creates conflicting phylogenetic signals between the α, β and the*
641 *willistoni mitochondrial clades*
642 We used several methods to search for signs of recombination between the clades. First, we
643 created single gene trees for the 13 mitochondrial protein coding genes and tested for conflicts
644 between the resulting topologies. For eleven of the 13 trees a definite topology could not be
645 determined as the bootstrap support was less than 80% on the node defining the relationships
646 of the α, β and willistoni clades. The remaining two trees had different topologies, as the *ND2*
647 gene featured the α clade as sister either to the willistoni clade (BS = 81%) and the *ND5* gene
648 placed α as sister to the β clade (*ND5* gene, BS = 90%). Second, we ran tests for intragenic
649 recombination on all protein coding genes. Only two genes, *CO2* and *ND3*, were detected with
650 marginally significant signals (p<0.05) using Phipack [65]. Finally, we estimated the
651 recombination over mutation rate along the whole mitochondrial genome to be low (0.01).
652 These results indicate that the recombination rate is low and that recombination signals are
653 weak and inconsistent. Thus, even if some recombination occurred between mitochondria, it
654 is unlikely to be the cause of the low support on the internal node of the whole-genome
655 mitochondrial tree.
656 Nucleotide composition bias and codon usage can also influence phylogenies and are
657 particularly likely to affect the more neutral third codon position. Hence, we built separate trees
658 using either only the third or a combination of the first and second codon positions of the
659 protein coding genes. Interestingly, we observed significantly increased support for α as sister
660 to the willistoni clade when the third codon position was removed (**Fig 2B**). In contrast, the
661 tree based only on the third codon position showed high (Bayesian posterior probability 0.99)
662 or fairly good (ML bootstrap 88%) support for the α and β clades being sisters (**Fig S7**). The
663 conflicting phylogenetic signal between the two first codon positions and the third codon
664 position suggests that nucleotide composition bias and perhaps codon usage are indeed the
665 cause of low support on the node defining the relationship between the α, β and willistoni
666 clades in the whole genome mitochondrial tree.
667 Since the third codon position is more likely to be affected by such biases, we conclude
668 that the topology supported by the first and second codon positions of protein coding genes
669 **(Fig 2B)** is most likely the accurate representation of the evolutionary relationships between
670 mitochondrial clades. Thus, *D. paulistorum* mitochondria are polyphyletic, and α is sister to
671 the willistoni rather than to the β clade. The resulting phylogeny **(Fig 2B)** implies an α-like
672 mitochondrial ancestor in *D. paulistorum,* and suggests that β was most likely acquired from
673 a so far unidentified lineage within the *willistoni* subgroup.

674 *Mitochondrial and nuclear phylogenies show a complex evolutionary history of mitochondria*
675 *in the willistoni subgroup*

676 A comparison between nuclear and mitochondrial phylogenies shows that neither mitotype
677 relationship nor presence of heteroplasmy follow the topology of the nuclear phylogeny.
678 Mitotypes are shared by relatively distantly related lines belonging to different semispecies,
679 as in the case of α1 in C2 (CA) and MS (AB) and β2 in O11 (OR), L1 (IN) and SM (TR),
680 whereas the primary β1 mitotype is exclusively found in AM semispecies (**Fig 2B**).
681 Heteroplasmic lines are also spread across the nuclear phylogeny, and the mitotype that
682 accompanies α2 in heteroplasmy varies according to what semispecies the line belongs. This
683 pattern suggests that the mitochondria of *D. paulistorum* were not only vertically transmitted
684 but have also been introgressed and lost multiple times during evolution. Further introgression
685 events are indicated by the positioning of *D. paulistorum* and *D. equinoxialis* in relation to other
686 species of the willistoni subgroup on either tree. In the nuclear phylogeny, *D. paulistorum* and
687 *D. equinoxialis* are the most derived species in the willistoni group, but the mitochondrial
688 phylogeny places the α and β clades in a basal position with regards to the willistoni clade
689 (**Fig 2A, B)**.

690 *NUMTs provide insights into the evolutionary history of mitochondria in the willistoni subgroup*

691 Insertions of mitochondrial DNA in the nuclear genome (NUMTs) can provide valuable
692 information about mitochondrial evolution. Using Illumina reads we generated whole genome
693 assemblies for all willistoni subgroup lines (**Fig 2A**) and searched for contigs containing
694 sequences with at least 90% identity to our mitochondrial genomes. Potential NUMTs were
695 identified based on (*i*) coverage similar to that of nuclear genes, (*ii*) low identity towards the α
696 and β genomes, and (*iii*) presence of mixed nuclear-mitochondrial sequences (see Methods).
697 Sequences classified as NUMTs were found in assemblies of all fly lines (**Tables S3, S8**).
698 Interestingly, eleven of the 13 *D. paulistorum* lines, including all lines carrying only β
699 mitochondria, shared a putative NUMT that spans a genomic region containing the *ND2* and
700 *CO1* genes. Contigs containing this sequence had similar coverage to nuclear genes in all
701 lines except FG295 and L1, where the NUMT had lower coverage than the nuclear average,
702 and FG572, in which the NUMT coverage was slightly higher than the nuclear average. In the
703 L1 and SM lines, the contigs also included non-mitochondrial sequences which confirm their
704 nuclear location (**Table S3**, **S10**). A phylogenetic analysis of this genomic region suggested
705 that this putative NUMT forms a clade together with the α mitotype and the mitochondria of
706 other willistoni subgroup species (**Fig 8**).

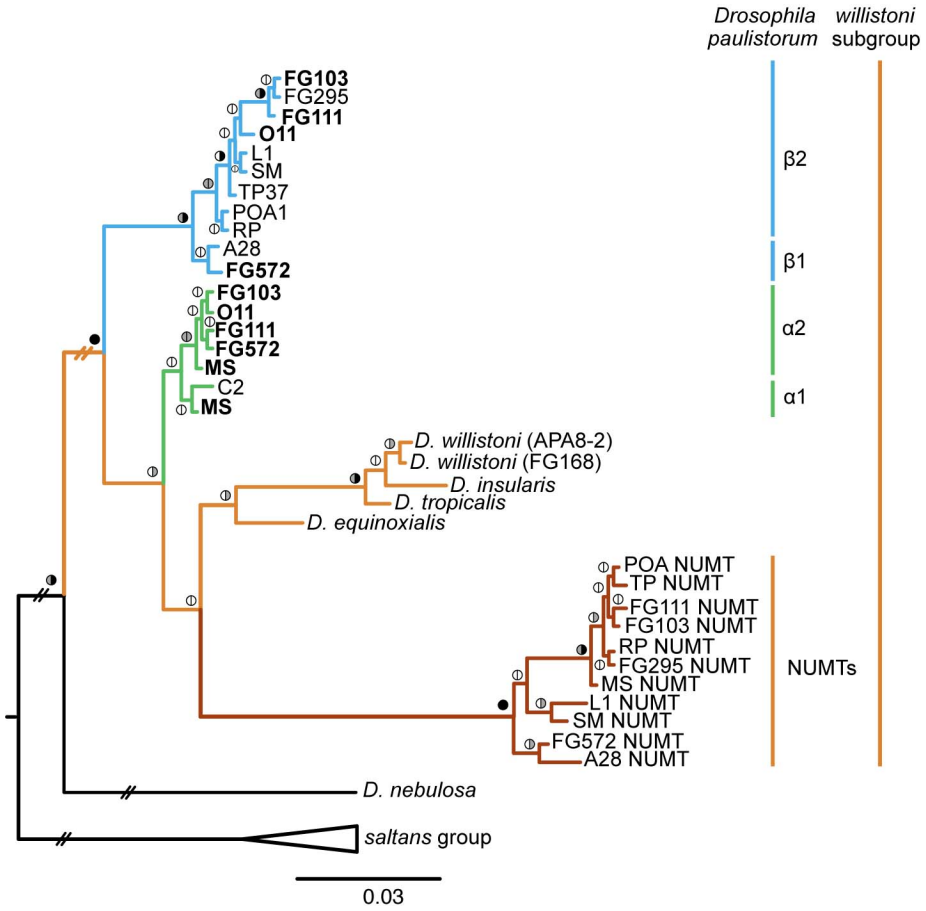


Fig 8. Phylogeny of the mitochondrial region spanning the *ND2-CO1* genes in the willistoni subgroup. NUMT sequences recovered from genomic assemblies of *D. paulistorum* lines branch together with the α and the willistoni clades. Fly lines in bold are heteroplasmic. The same topologies are obtained by maximum likelihood (GTRGAMMA model) and Bayesian analyses. Colored circles next to the nodes indicate ML bootstrap support (left half) and Bayesian posterior probability (right half). Black corresponds to 99-100% bootstrap support or 0.99-1.0 posterior probability, grey represents 70-98% bootstrap support or 0.7-0.98 posterior probability, white indicates support below 70%. Branches leading to the outgroups and marked by two transverse lines were shortened and don't follow the presented scale.

The presence of the sequence in nearly all *D. paulistorum* lines together with the nuclear-like topology suggest that this NUMT stems from a single nuclear insertion that occurred prior to the diversification of *D. paulistorum* semispecies. The ancient origin together with the phylogenetic position of this NUMT close to α and the willistoni clades further support the hypothesis of an α-like ancestral mitotype of *D. paulistorum*.

Finally, we observed that the genome assembly of *D. equinoxialis* contains several β–related contigs, which were not found in species of the willistoni clade (**Tables S3, S8**). The coverage of these contigs is lower than the nuclear average for the line and they do not contain matches to non-mitochondrial sequences. Hence, while these *D. equinoxialis* sequences might stem from NUMTs, we cannot fully exclude that they represent low-level contamination. Nevertheless, it is interesting to note that the presence of β sequences in *D. equinoxialis* might

730 suggest that the last common ancestor of *D. paulistorum* and *D. equinoxialis* carried a β-like
731 mitochondrion.
732
733 **Could α2 be a Mega-NUMT and not a secondary mitotype?**
734 Besides the ancient NUMTs found in the willistoni subgroup species (**Fig 8**) it seems likely
735 that the α2 mitotype could be a so-called Mega-NUMT, i.e., a multicopy nuclear locus stably
736 integrated in the *D. paulistorum* genome of heteroplasmic flies. This concept was recently
737 coined by [66] in response to a publication by [67] who reported on the spectacular discovery
738 of biparental mitochondrial transmission in some human families. Since then, additional
739 publications have emerged pointing towards the true existence of Mega-NUMTs in humans
740 [68, 69].
741 In order to experimentally test this alternative hypothesis for α2 in *D. paulistorum*, we
742 performed DNA-Fluorescent in Situ Hybridizations (FISH) with two differentially labelled
743 probes (**Fig 9**); one targeting the mitochondrial genomes covering five loci (red), and the other
744 targeting the rRNA locus (green) of *D. paulistorum*. Hybridization and washings were
745 performed under relatively stringent conditions (see Material and Method) in order to avoid
746 cross-hybridization of the mt-probes with the fixed ancient NUMT shown in **Figure 8**. As
747 numbers of α2 mt-genomes per O11 nuclear genome are estimated to at least 20 copies in
748 females and more than 50 in the males (**Fig 3B**), we expect to detect either one or two
749 chromosomal Mega-NUMT clusters, dependent on the sex, with a minimal size of at least 300
750 kb each. We further envision that one mt-signal should be located on the Y chromosome of
751 males and one on an autosome in both sexes. Therefore, we used the rRNA locus as a control,
752 as it is expected to be of similar or bigger-size (>300 kb) than the α2 Mega-NUMT.
753 In our DNA-FISH experiments on whole-mount early embryos (**Fig 9A-F** and
754 **Supplemental data file**), we detected exclusively in the heteroplasmic strains (O11 and MS)
755 two bright and compact mt-signals in all interphase cells (**Fig 9A-C**), which are closely
756 associated with the nuclear periphery, and in mitotic cells 1-2 less compacted mt-aggregates
757 (red) associating with the distal chromosomal ends (**Fig 9D, E**). As expected, no mt-signals,
758 besides cytoplasmic ones, were traceable in any of the tested homoplasmic strains (FG295),
759 which are devoid of α2 (**Fig 9C, F**).
760 As fixed cells of whole-mount specimens maintain cytoplasmic components like
761 mitochondria, we performed DNA-FISH on mitotically-active neuroblasts of squashed 3rd
762 instar larval brains after hypotonic treatment, which washes out the cytoplasm and leaves
763 behind the nucleus. Although the rRNA control probe worked on all metaphase chromosomes
764 (**Fig 9 G-I**), mt-signals were untraceable even under enhanced laser intensity in specimens of
765 the two heteroplasmic strains.

766 Furthermore, DNA-FISH with the same probe set (mt-probe and rRNA control) was
767 performed on squashed polytene chromosomes from 3rd larval salivary glands of homo- and
768 heteroplasmic lines under similar stringent conditions (**Fig 9 J-L**). Although the rRNA clusters
769 on the X and the Y chromosomes were brightly stained depending on the sex of the larvae,
770 no chromosomal signals of the mt-probe were detected in any of the 16 polytene interphases
771 derived from the two heteroplasmic strains (**Fig 9 J, K** and **Supplemental data file**). Even
772 though the α-heterochromatin is under-replicated in polytene chromosomes of *Drosophila*, a
773 potential Mega-NUMT (>300 kb) should not escape detection in telomers or the chromocenter
774 by Confocal Laser Microscopy at maximal laser intensity and overexposure (**Fig S9** and
775 **Supplemental data file**). Thus, the clear absence of mt-FISH signals in any cytoplasmic-free
776 samples (**Fig 9G-L**) allowed us to reject the Mega-NUMT hypothesis, *i.e.*, the physical
777 integration of α2 into the host chromosome, and, as an alternative hypothesis, its integration
778 in B-chromosomes, which usually are defined by the presence of their own telomeres and a
779 centromere during mitosis [70].
780 We interpret our findings obtained by DNA-FISH as follows: The α2 mitotype is neither
781 a classic Mega-NUMT, nor integrated in an accessory B-chromosome. Instead, it resembles
782 a facultative, invasive and selfish genetic element, which is cytoplasmic and still maintains a
783 complete and potentially functional genome evolving under purifying selection. This mitotype
784 associates closely with the host nucleus in interphase and mitotic chromosomes. During early
785 development α2 probably hooks up with the replication machinery for increasing their own
786 copy numbers. Just after fertilization, embryonic nuclei duplicate rapidly, primary mitotypes
787 are transcriptionally and replicatively quiescent, but, as shown in **Fig. 7**, levels of α2 increase
788 quickly from relatively very low to significantly high copy numbers. We speculate of a spatial
789 and temporal hitchhiking of α2 with early nuclear replication just after fertilization and also later
790 in other dividing cells, like neuroblasts in larval brains. This "Replication precox" phenotype is
791 assisted by a yet undetermined mechanism that safeguards the selective replicational
792 advantage of this selfish mitotype by counteracting out-dilution by the primary mitotype.
793

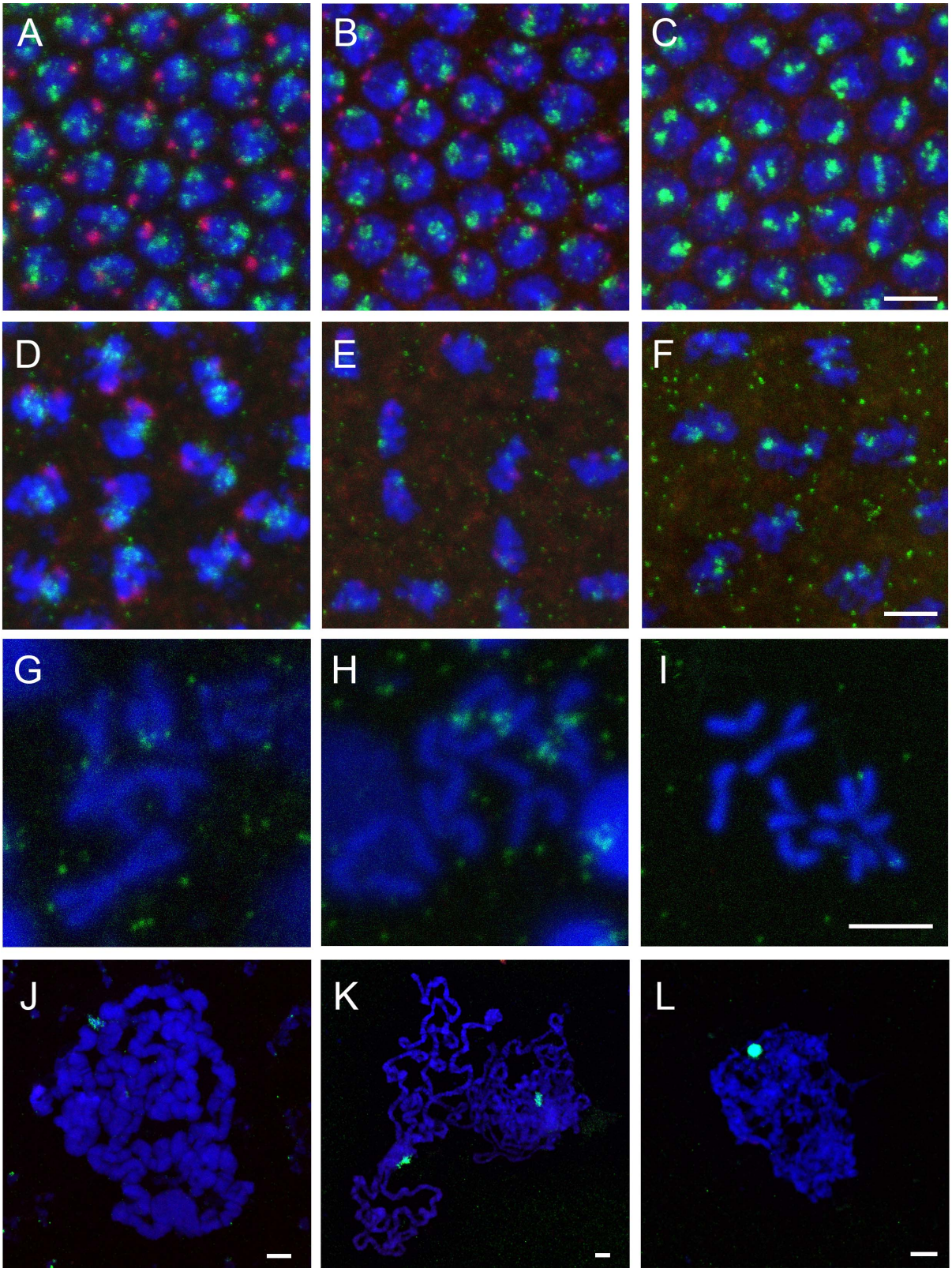


Fig 9. Fluorescent in situ hybridization on *D. pauilstorum* tissues with mitochondrial and ribosomal DNA-probes. Applied probes are a labelled mixture of CO1, CO2, ND1, ND2 and ND4 derived from cloned plasmids for mitochondria (red) and as nuclear control the 28S rRNA (green), and DNA is stained with DAPI (blue). Interphase nuclei of heteroplasmic O11, MS (**A, B**) and homoplasmic FG295 (**C**) lines from early embryos (0-3h): Note dense red signals (mt-probe) associated with the nuclei (blue) only in heteroplasmic (**A**, **B**) embryos (N=8 for each), whereas in the homoplasmic control (**C**) mt-signals are found exclusively in the

802 cytoplasm (N=3), and the hybridization control rRNA probe (green) binds in all three lines.
803 Mitotic chromosomes around prometaphase of heteroplasmic O11, MS (**D**, **E**) and
804 homoplasmic FG295 (**F**) lines from early embryos (0-3h): Note loose red signals (mt-probe)
805 are associated loosely with the distal chromatin (blue) in heteroplasmic embryos (N=3).
806 Squashes of mitotic chromosomes of heteroplasmic O11 (**G, H**) and homoplasmic FG295 (**I**)
807 3rd instar larval brains: Note the absence of red signal (mt-probe) in both lines (N=10 for O11
808 and N=4 for FG295), whereas the green probe (28S rRNA) is detected in all of them. Polytene
809 chromosomes of heteroplasmic O11, MS (**J-K**) and homoplasmic FG295 (**L**) 3rd instar larval
810 salivary glands: Note the absence of mt-probe (red) in all three lines, whereas rRNA control
811 is detected (N=8 for each line). Scale bar: 5 μm.
812
813

814 # DISCUSSION

815 We report on a unique system of stable mitochondrial heteroplasmy in the neotropical *D.*
816 *paulistorum* species complex that might exist due to biparental transmission. Using
817 sequencing, RFLP-PCR and bioinformatic analyses we discovered high-level heteroplasmy in
818 41% of the tested fly lines including representatives from three of the six *Drosophila*
819 *paulistorum* semispecies (Andean-Brazilian, Amazonian and Orinocan). However, given the
820 limited number of lines currently available from the Centro-American, Interior and Transitional
821 semispecies, it is possible that heteroplasmy may be even more broadly distributed among
822 the species. We also found that heteroplasmy is fixed with balanced levels of each mitotype
823 in individuals of three long-term laboratory lines (O11, MS and Yellow), but not in the more
824 recently collected lines, in which α2 levels are also tightly balanced but not fixed (**Table 1**).
825 These results suggest that a line has to be genetically homogeneous for heteroplasmy to be
826 fixed and that levels are controlled by the nuclear genetic background. In the course of our
827 current study we observed that the Orinocan FG103 line collected in 2014 lost the secondary
828 mitotype within the first two years in the lab, suggesting that long-term maintenance of fly lines
829 under lab conditions might also contribute to an underestimation of prevalence. Further
830 sampling will be necessary to determine the prevalence of heteroplasmy in natural
831 populations.
832

833 **Function of the α2 mitotype in *D. paulistorum***

834 Using phylogenetic analysis, we found that the mitochondria of *D. paulistorum* are polyphyletic
835 and form two phylogenetically distinct clades - α and β - each further divided into two
836 subclades, 1 and 2. Heteroplasmy is always associated with the α2 mitotype, and this mitotype
837 is in turn never observed in homoplasmy. Such a pattern may occur because α2 alone does
838 not perform all of the functions required for sustaining a fly, or perhaps because homoplasmic
839 α2 lines are rare and difficult to sample in the wild, or both. Although systematic field sampling
840 would be necessary to determine which hypothesis is correct, we found additional evidence

841 suggesting that α2 might not provide full mitochondrial functionality. Mitochondria are usually
842 enriched in energy-demanding tissues [71, 72], and by studying the heteroplasmic O11 line
843 we found that only the β2 mitotype follows this pattern. The α2 mitotype, on the other hand,
844 has unchanging titers across fly tissues. Although not conclusive, this suggests that β2 is the
845 mitotype that provides energy via oxidative phosphorylation, while α2 does not respond to the
846 energy demands of the host. Despite this, the genes encoded on the α2 genome appear to
847 be evolving under purifying selection, as indicated by low dN/dS ratios and the absence of
848 pseudogenes. In addition, very few genomic differences are seen between α2 and the closely
849 related α1 mitotype (0,67% divergence), which since it is the sole mitotype found in two
850 Centro-American *D. paulistorum* lines (C1 and Yellow) must be fully functional. The low
851 sequence divergence between α1 and α2 strongly suggests that the transition from a primary
852 α-like ancestor to a selfish mitotype took place relatively recently, but is found consistently in
853 different and reproductively-isolated semispecies. Considering these facts, we surmise that
854 α2 is likely functional based on its genome but that its steady titer throughout tissues and
855 absence in homoplasmy suggest that it does not provide sufficient energy to sustain a fly, at
856 least under the standard laboratory conditions of our experiments.
857
858 **Biparental inheritance of the α2 mitotype in *D. paulistorum***
859 The question remains as to how and why α2 is maintained in *D. paulistorum* flies. It may serve
860 a yet undefined essential or condition-specific function for the fly, or alternatively it may be
861 cheating the system and perpetuate selfishly while providing no or few benefits to the host.
862 One possible way of accomplishing the latter is for the mitotype to increase its own vertical
863 transmission, for example through biparental inheritance.
864 Using RFLP-PCR and discriminant qPCR, we detected high level paternal
865 transmission of the α2 mitotype in offspring from both intra- and inter-semispecies crosses of
866 *D. paulistorum* (**Fig 4, 5 and 6A, B**), and even in offspring from crosses between *D.*
867 *paulistorum* and two closely related species of the willistoni subgroup (**Fig 6C**). None of the
868 other mitotypes found in *D. paulistorum* were paternally transmitted in high levels. Generally,
869 active pre- and post-fertilization recognition and elimination mechanisms prevent paternal
870 transmission of mitochondria and their DNA in animals (recently reviewed in [15]). In *D.*
871 *melanogaster,* the mechanism is not yet fully understood, but two genes, *EndoG* and *tamas,*
872 are known to be involved in eliminating mitochondrial DNA from developing spermatids [13,
873 14, 73]. After fertilization, paternally leaked mitochondria are eliminated by endolytic and
874 autophagic degradation systems [74] while the rare paternal ones that escape are passively
875 outcompeted by the larger number of maternally transmitted mitochondria and their genomes
876 already present in the egg [75]. In order for a mitotype to be transmitted paternally and – even
877 more importantly – to stably persist in adult flies in large numbers, at least one of these

878 mechanisms need to be bypassed. We did not specifically determine if the α2 mitotype evades
879 the active mechanisms that eliminate paternal mitochondrial DNA (reviewed in [15]), or
880 associates with the sperm nucleus, but the observation that none of the other mitotypes of *D.*
881 *paulistorum* were paternally transmitted in large numbers (**Table S7**) implies that there is no
882 general loss of such active defense functions in *D. paulistorum*. The spurious low level (~0.1-
883 0.3%) paternal leakage of β2 in some hybrid crosses could be a result of a partial breakdown
884 of such mechanisms due to the mixed nuclear hybrid background ([23] and references
885 therein). However, low levels of paternally transmitted β2 is also in line with previous reports
886 of paternal leakage within *Drosophila* species [26, 29, 30], indicating that similar levels of
887 leakage can occur even without interspecies hybridization and breakdown of defenses. Thus,
888 the mechanism for paternal transmission is specific to the invasive α2 mitotype.
889 In *D. melanogaster* and *D. simulans,* the spermatozoon is estimated to carry
890 approximately 20 copies of mitochondrial DNA [59], which is five to six orders of magnitude
891 less than the number of copies found in an egg [59]. We expected that similarly small numbers
892 of the α2 copies are present in *D. paulistorum* sperm and at the onset of embryo development
893 when α2 is transmitted by the male. However, using relative RFLP-PCR, we found that rapid
894 and premature replication in early embryos resulted in α2 levels that were similar to those of
895 adult flies just a few hours after fecundation (**Fig 7**). Previous studies of *D. melanogaster*
896 showed that maternally transmitted mitochondria are quiescent and do not replicate for up to
897 10 hours after fertilization [13, 60], which seems to be an evolutionary conserved state
898 (reviewed in [76]). Hence, the premature and unique temporal replication of the α2 mitotype,
899 that we name "Replication precox", might allow this mitotype to escape passive dilution by the
900 still dormant primary mitotype and thereby guarantee its presence in high levels in adult flies.
901 We also found that the α2 mitotype level increased earlier and faster in hybrid embryos when
902 it was transmitted paternally rather than maternally (**Fig 7D, E**). This could be due to more α2
903 mitotype copies being deposited by the sperm than by the egg, or to a higher rate or earlier
904 onset of α2 replication when it is paternally inherited. Since we found that O11 males have
905 significantly higher levels of α2 than females, also more copies in testes than in ovaries, and
906 DNA-FISH showed a close spatial association with nuclear replication, our data suggest that
907 the former hypothesis is possibly more likely. One tempting mechanistic hypothesis for the
908 Replication precox phenotype is that, because of their close vicinity to nuclei, α2 replication
909 might profit by parasitizing on their replication machinery. The elucidation of the mechanism
910 for diagnostic recognition and/or selective replication of α2 is currently underway in our
911 laboratory.

912 **Variability in mitotype titer and levels between sexes**

913 We found that α2 titers were consistently and significantly higher in males than in females.
914 Since α2 does not seem to be a fully functional mitotype, the higher α2 titer in males cannot
915 be interpreted purely based on different energy needs between sexes. So why would males
916 have a higher titer of α2 than females? One possibility is that high titer in males has evolved
917 specifically to facilitate paternal transmission of the invasive α2 mitotype, since a higher titer
918 increases the chance of transmission. Alternatively, higher male titer and the Replication
919 precox phenotype of α2 could have co-evolved in order to compensate for a reduced
920 functionality of the maternally transmitted primary mitotype in males. The mother's curse
921 hypothesis suggests that DNA of strictly maternally inherited mitochondria co-evolve with
922 females and might therefore accumulate mutations that are harmful to male functions [77]. If
923 such mutations have accumulated in the maternally transmitted primary mitotype of *D.*
924 *paulistorum*, the biparentally transmitted α2 mitotype might have co-evolved with males and
925 express a still undetermined function that is more important to males than females. In the O11
926 line, higher relative titer of α2 compared to the primary mitotype β2 are found in heads and
927 gonads (**Fig 3B**), suggesting that perhaps behavior and fertility might be important further
928 areas of investigation for α2 specific functions in males.
929 In addition, our data suggest that the total mitochondrial DNA titer is significantly
930 affected by the presence of α2, as this secondary mitotype appears in addition to the primary
931 mitochondrial DNA copies in a cell. A similar quantitative effect on mitotype copy numbers
932 was previously observed in heteroplasmic *D. subobscura* flies carrying a selfish mitotype with
933 a large deletion that results in a replicative advantage [78].
934 Our crossing data also suggests that the nuclear genetic background influences the
935 total mitochondrial DNA titer, particularly in males. This is indicated by a significant reduction
936 in overall mt-DNA titer between pure O11 heteroplasmic males and heteroplasmic hybrid
937 offspring of both intra- and interspecific crosses, but not between pure O11 heteroplasmic
938 females and hybrid offspring. This observation is in line with long-term co-evolution between
939 the heteroplasmic cytoplasm and the nuclear genome in males. In this context it is interesting
940 to note that two of the genomic differences, which we found in the α2 genomes, are located
941 in the T-loops of tRNAs and the nuclear encoded gene *mEFTU-2* that is specifically expressed
942 in testis of *D. melanogaster* has mutations in a region where the protein is predicted to
943 interacts with the tRNA T-loop [63, 64]. Hence, if assuming a similar male-specific expression
944 of *mEFTU-2* in *D. paulistorum*, this observation might strengthen the idea of a co-adaptation
945 between the α2 mitotype and male functions.

946 **Evolutionary history of the α and β mitotypes in the neotropical willistoni group**

947 The ancestrality of α-like mitochondria in *D. paulistorum* is strongly supported by our
948 phylogenetic analysis after we accounted for the nucleotide composition bias in our

949 mitochondrial phylogeny, since the α clade is clearly sister to the other willistoni group species.
950 Furthermore, the presence of an α-related NUMT, which is shared by most sequenced *D.*
951 *paulistorum* lines (**Table S10**) and shows a topology that is similar to our nuclear tree (**Fig 8**),
952 indicates that this sequence was transferred from an α-like ancestral mitotype to the nuclear
953 genome before semispecies diversification.

954 Since the β mitotype does not share an immediate common ancestor with α or any
955 other lineage in our phylogeny, its presence in *D. paulistorum* is likely the result of a "ghost"
956 introgression from an unsampled or possibly even extinct species [55-57]. The phylogenetic
957 position of β suggests an origin within the willistoni group but outside the willistoni subgroup.
958 Thus, possible donors include species from the other two subgroups within the willistoni group, the
959 bocainensis and the alagitans. Unfortunately, sequence data is not available for all species in
960 the *bocainensis* subgroup, but *D. nebulosa,* which is included in our phylogeny, branches
961 basally to β. A similar position was observed for three other representatives of the bocainensis
962 subgroup, *D. capricorni, D. fumipennis* and *D. sucinea* in previously published single gene
963 trees [54]. No sequence data is available for any alagitans species. Alternatively, the β
964 mitotype may be the genomic remnant of an extinct species of the willistoni group which once
965 hybridized with *D. paulistorum*. Cases of introgressions followed by extinction of the donor
966 species have been reported in a variety of organisms and are often associated with deep
967 mitochondrial divergence in clades with otherwise high nuclear similarity [57, 79]. To our
968 knowledge, the *D. paulistorum* species complex is the first reported case of stable
969 heteroplasmy originating from mitochondrial DNA introgression.

970 Our analyses allow us to further infer that the introgression of β likely happened before
971 the diversification of the semispecies in *D. paulistorum*. The presence of the β mitotype in 11
972 of 17 sampled lines from four different semispecies of *D. paulistorum* makes it more
973 parsimonious that the mitotype was inherited once from a common ancestor rather than
974 acquired multiple times. The ancestral introgression of β is further suggested by the observed
975 convergence in third codon position between the α and β mitotypes (**Table S4**). Regardless
976 of whether the underlying mechanism of convergence is exposure of the β genome to the
977 mutational bias (amelioration) or the selective pressure (codon usage) of *D. paulistorum*, it
978 likely required significant time to act. A third indication that β introgression predates *D.*
979 *paulistorum* diversification is the presence of potential β-NUMTs in *D. equinoxialis*. Although
980 we could not conclusively determine the nuclear origin of these sequences, their presence
981 might indicate that *D. equinoxialis* at some point carried a β-like mitotype. If so, β could have
982 been introgressed already in the common ancestor of *D. paulistorum* and *D. equinoxialis*.

983 A comparison between the phylogenetic trees of nuclear and mitochondrial DNA
984 suggests that several other introgression events might have occurred in the willistoni subgroup
985 apart from the acquisition of β. We observed that the α clade branches basally to the willistoni

986 clade even though nuclear data positions *D. paulistorum* as one of the most derived species
987 in the willistoni subgroup. This suggests that perhaps the extant α mitotype does not directly
988 descend from the ancestral mitochondria of *D. paulistorum*, even if the ancestral
989 mitochondrion was also an α-like mitotype. Instead, the phylogenetic placement of all extant
990 α mitotypes indicates that an introgression in the ancestor of *D. paulistorum* and *D.*
991 *equinoxialis* from an unsampled deep diverging α-carrying lineage within the willlistoni group
992 might have given rise to the current α mitotypes in a similar mode to what we propose for β.
993 One possibility is that the two extant mitotypes were introgressed at the same time, either
994 together or separately.

995 Incongruences between the nuclear and mitochondrial trees also indicate more recent
996 introgressions between *D. paulistorum* semispecies and lines, as mitotype distribution does
997 not follow the nuclear phylogeny. Since *D. paulistorum* is a recent radiation, i.e., a species
998 cluster *in statu nascendi* ([44] and references therein), such pattern could be explained by
999 hybridization in a more recent past, when the pre- and postmating incompatibilities that exist
1000 today between semispecies (reviewed in [80], [44], and this study) were still incomplete.
1001 Although both F_1 hybrid sexes from our inter-semispecies crosses were sterile, which would
1002 have prevented mitotype introgression, possibly some rare fertile F_1 female flies could be
1003 found with extended tests and/or other line combinations. As discussed in early literature [81],
1004 it is possible that sporadic hybridization occurs between less discriminating and partially
1005 compatible populations or semispecies, and those could perhaps lead to the serendipitous
1006 events of mitochondrial introgression. As shown earlier, the Transitional semispecies is
1007 particularly permissive to inter-semispecies crossings and has the potential to bridge borders
1008 [82]. Specific lines of other semispecies have also been shown to be less isolated from others,
1009 potentially allowing at least some rare cases of genetic leakage and exchange [83]. For the
1010 one Transitional line included in our study, SM, we see a strong discrepancy between the
1011 nuclear and mitochondrial trees, with the mitochondrial genome nested within the Orinocan
1012 mitochondrial clade and the nuclear tree positioning it with Centro-American. Both our
1013 phylogenetic analysis and experimental data from crosses suggest that α2 mitotypes are
1014 particularly likely to be transferred in crosses between semispecies, and even when crossed
1015 with sister species.

1016 Another possible driver of mitotype introgression after a hybridization event is the
1017 maternally inherited endosymbiont *Wolbachia*. *Wolbachia* strains with strong reproductive
1018 manipulation phenotypes or which provide enhanced host fitness can spread fast in the
1019 population and are known to cause sweeps of the mitotype associated with infected females
1020 [84, 85]. *D. paulistorum* is known to carry obligate *Wolbachia* symbionts, but it is not yet known
1021 if these cause any reproductive manipulation besides premating isolation [47, 49]. Thus, we
1022 currently have no indication that *Wolbachia* has any part in the introgression, past or on-going.

1023 **Introgression and heteroplasmy as drivers of the evolution of selfish mitochondrial**
1024 **traits**
1025 Our inferences on mitochondrial evolution within *D. paulistorum* provide us with a hypothesis
1026 about how features specific to α2 may have evolved. We speculate that the introgression of β
1027 into an α-carrying *D. paulistorum* likely created competition between mitotypes for a chance
1028 to be transferred to the next host generations. If the ancestrally introgressed β mitotype was
1029 more efficient at maternal transmission or perhaps just a better energy producer than the
1030 ancestral α, selection could have made it the more dominant mitotype in the *D. paulistorum*
1031 species complex. Over time, this could have led to a reduction in prevalence and possibly
1032 resulted in extinction of the α mitotype from the species. In this scenario, the evolution of
1033 biparental transmission and the Replication precox phenotype may have allowed the α2 lineage
1034 to persist selfishly next to the primary mitotype. Such traits would perhaps guarantee the
1035 transmission of α2 to the next host generation independently of its ability to meet the energetic
1036 needs of the host or to compete with the maternal mitotype for presence in the egg.

1037 **Can we be sure that α2 is not a Mega-NUMT?**
1038 Finally, we are aware of a potential alternative hypothesis which would suggest that α2 is not
1039 a cytoplasmic mitotype but might stem from a so-called "Mega-NUMT". It was recently shown
1040 that Mega-NUMTs can mimic heteroplasmy and extremely rare cases of paternal
1041 mitochondrial DNA leakage in humans and other vertebrates (recently reviewed in [68]). In
1042 this case, mitochondrial DNA-concatemers possibly produced by extrachromosomal rolling
1043 circle replication were found as large-sized chromosomal insertions in human autosomes and
1044 wrongly assumed to be unusual biparental transmission of an additional mitotype [67].
1045 However, in *D. paulistorum*, we reject this alternative hypothesis based on the following
1046 reasons: Contrary to Mega-NUMTs, which rapidly accumulate degenerative mutations and
1047 structural variation, the genome of the α2 mitotype resembles a fully functional mitochondrial
1048 DNA in size and content. Furthermore, the α2 genome is highly conserved within and between
1049 heteroplasmic *D. paulistorum* semispecies as well as between long-term lab lines and recently
1050 collected lines, which suggests that it evolves under purifying and not neutral selection. In
1051 addition, Mega-NUMTs are cryptic and if detected they are transmitted at very low frequencies
1052 and thereby do not follow strict Mendelian inheritance. These findings are in sharp contrast to
1053 α2 of *D. paulistorum* flies that shows almost perfect biparental transmission within a line, and
1054 as shown by our introgression experiments, this selfish mitotype has the capacity for invasive
1055 spreading. A recent publication by [69] demonstrated the physical association of a Mega-
1056 NUMT with around 45 to 56 mt-genome copies in chromosome 14 of healthy humans. In our
1057 case however, we failed to detect clear mt-signals of any large sized nuclear-integrations of
1058 α2 by means of DNA-FISH on fixed whole-mounted early embryos, cytoplasm-free neuroblast

1059 metaphaes and polytene chromosomes from heteroplasmic lines, even though a potential α2
1060 Mega-NUMT is estimated to be bigger than 300 kb. Finally, our backcross experiments clearly
1061 demonstrated the persistance and spreading capacity of the α2 mitotype in a homoplasmic
1062 nuclear backgrounds, showing that α2 is self-sufficient and hence not associated with any
1063 Mega-NUMTs.

1064 CONCLUSION

1065 In this paper we present *D. paulistorum* as the first example of an animal with high level, stable
1066 heteroplasmy, which is derived from introgression and is robustly maintained in both sexes in
1067 under laboratory conditions and in nature. The system is remarkable for involving a secondary
1068 mitotype with a seemingly functional genome, which appears to not respond to the energy
1069 demands of the host and instead relies on biparental inheritance and rapid replication in the
1070 early embryo as potentially selfish strategies for spreading and persisting. Additionally, its
1071 higher titre and tighter co-evolution with the nuclear background in males might hint at a
1072 specific male related function.

1073 Given its intact putatively functional genome and recently evolved selfish phenotype,
1074 the α2 mitotype can potentially be referred to as a sexually transmitted, parasitic infection
1075 rather than a primary maternally inherited mitotype. Hence, since α2 appears to have a more
1076 "rebellious" lifestyle with less host control than typical primary mitotypes, but more similar to a
1077 parasitic hitchhiker on nuclear replication, we propose to name it "*Spartacus*" in reference to
1078 the famous ancient Roman slave, gladiator and then commander who escaped from imperial
1079 oppression by rebellion, for at least a short period of his lifetime.

1080 Our results contribute to the increasing realization, driven by recent large-scale
1081 sequencing efforts, that animal mitochondrial genomes often do not follow the dominant model
1082 of homoplasmic, maternally inherited organelles, neither on ecological or evolutionary times
1083 scales. Sometimes even organisms which have been studied for decades, such as *D.*
1084 *paulistorum,* harbor unexpected complexity.

1085 MATERIALS & METHODS

1086 Fly strains, semispecies typing, and *Wolbachia* knockdown

1087 Information regarding the *Drosophila* lines used in this study is summarized in **Table S1**. All
1088 flies were kept at 25±1°C on Formula 4-24 *Drosophila* instant food (Carolina, USA) with a 12
1089 hr light-dark cycle. Typing of the biological semispecies of the *D. paulistorum* isofemale lines,
1090 which were recently collected from French Guiana (FG16, FG103, FG111, FG295 and FG572),
1091 was done by analyzing their reproductive compatibility towards reference lab strains from each
1092 of the six semispecies (A28 (AM), O11 (OR), C2 (CA), L1 (IN), SM (TR) and MS (AB)**).** Each
1093 new line was assigned to the semispecies with which it produced fertile F1 offspring. Crosses

1094 were performed in 2-3 biological and three technical replicas (in total 30 crosses in 2-3
1095 biological replicas). Each biological replicate consisted of 3-6 females and 3-6 males, which
1096 were kept in the same vial for 3-5 days before transfer into a new vial for generating one
1097 technical replicate. Egg-to-adult development was monitored and information on all crosses
1098 including the number of F_1 flies hatched is summarized in **Table S6**.
1099
1100 *Wolbachia knockdown*
1101 In order to test the potential effect of the *Wolbachia* endosymbiont of *D. paulistorum* [47] on
1102 mitotype persistence and levels, we generated lines with lowered *Wolbachia* infection titer.
1103 *Wolbachia* was knocked down by keeping flies on food containing 0.05% tetracycline [w/v] for
1104 four generations. Flies were repeatedly tested using PCR in order to confirm the absence of
1105 the symbiont. After four generations, knock-down flies were moved to tubes previously
1106 occupied by males of the same line, so that they could feed on feces of flies with wild-type gut
1107 microbiota. This aimed to restore any gut flora affected by the antibiotic and keep the
1108 knockdown and wild-type fly lines as similar as possible with the exception of *Wolbachia*
1109 presence [49].
1110
1111 **Single fly Southern blots**
1112 Single fly Southern blots were performed following the protocol by [86]. In brief, DNA was
1113 extracted from single individuals, digested with the restriction enzyme *Hin*dIII and processed
1114 for electrophoresis on a vertical 0.8% agarose gel. After vacuum-blotting onto a positively
1115 charged nylon membrane, fragments were hybridized with P_{32}-labelled mitochondrial *CO1*
1116 probe derived from a plasmid. Detection was performed by exposing the membranes to X-ray
1117 films at -80 degrees.
1118
1119 **Restriction Fragment Length Polymorphism-PCR (RFLP-PCR)**
1120 Total DNA was extracted from single or multiple flies using the Gentra Puregene Tissue kit
1121 (Qiagen, Germany). Flies were homogenized using a TissueLyser LT (Qiagen) at 50 Hz for
1122 50 sec and DNA was consequently isolated following the manufacturer's protocol. DNA
1123 concentration was determined with a NanoDrop OneC Spectrophotometer (Thermo Scientific).
1124 RFLP-PCR for the mitochondrial Cytochrome c oxidase genes *CO1* and *CO2* was set up in
1125 20 µl reactions containing 1x Promega reaction buffer, 2.5 mM $MgCl_2$, 0.5 µM of each primer
1126 (**Table S11**), 200 µM of each dNTP, and 0.025 U of GoTaq G2 DNA Polymerase. *CO1*/*CO2*
1127 reactions were performed using the following thermal profile: initial denaturation for 2 min at
1128 95°C followed by 30 cycles consisting of 45 sec denaturation at 95°C, 45 sec annealing at the
1129 appropriate primer conditions Tm°C (**Table S11**) and 30 sec extension at 72°C. Final
1130 extension was carried out at 72°C for 10 min. PCR products were consequently used for RFLP
1131 (digest plus undigested control) using the restriction enzymes *Hin*dIII and *Eco*RV-HF for *CO1*,

1132 and *Msp*I for *CO2*. Reactions were performed in 20 µl volume containing 10 or 20 U of enzyme,
1133 1x reaction buffer and 10 µl PCR product. After overnight incubation at 37°C, enzymes were
1134 inactivated by either heat (80°C for 20 min) or by adding EDTA-containing gel loading dye. 10
1135 µl of undigested and digested samples were then run on a 1.5% agarose gel and post-stained
1136 with 10 mg/ml ethidium bromide. Imaging was performed using a Molecular Imager
1137 ChemiDocTM XRS Imaging System and mitochondrial heteroplasmy levels were quantified
1138 using Image Lab 5.2.1 software.
1139
1140 *Quantifying levels of heteroplasmy*
1141 Undigested samples representing the sum of both mitotypes were set as reference band with
1142 relative quantity 1. The relative levels of mitotypes were then quantified by summing up the
1143 two digested fragments plus the undigested fragment. Due to technical differences in band
1144 intensity, however, the sum of all fragments did not result in a relative quantity of 1. Hence the
1145 actual sum value was set to 100% to determine the percentage of mitotype levels. Smaller
1146 fragments with length up to approximately 100 bp were not detectable after ethidium bromide
1147 staining and we hence used the correction factor $Cf = \frac{undigested[bp]}{digested[bp]}$ to estimate their relative
1148 quantities.
1149
1150 **DNA extraction, genome sequencing and assembly**
1151 For each line 20 ovaries were dissected from 3-day old adult females and DNA extraction was
1152 performed using the Gentra Puregene Blood and Tissue kit (Qiagen). DNA samples used for
1153 PacBio sequencing were produced using the protocol described in [87]. Briefly, *Drosophila*
1154 flies were transferred to apple juice agar plates and allowed to oviposit for 2 hours, after which
1155 the eggs were collected, washed and dechorionated in 50% bleach and manually
1156 homogenized using a plastic pestle. Following centrifugation and filtration of the homogenate,
1157 the resulting cell pellet was subjected to whole genome amplification using the Repli-g midi kit
1158 (Qiagen) after which the DNA was purified using QIAamp DNA mini kit (Qiagen) according to
1159 the manufacturer's recommendations.
1160
1161 *Sequencing*
1162 *Drosophila paulistorum* lines from the classic six semispecies plus lines from more recent
1163 collections were used for genome sequencing (**Table 1 and S1)**. DNA extracted from ovaries
1164 was used to construct 350 bp fragment TruSeq libraries, multiplexed and sequenced at the
1165 Uppsala SNP and Seq platform in three different runs on an Illumina 2500 HiSeq machine
1166 generating 2x125 bp reads. The Illumina reads were quality filtered and trimmed using
1167 Trimmomatic-0.30 and error corrected using SPAdes-3.5.0. Amplified DNA extracted from
1168 early embryos from three lines (O11, POA-1 and TP37) was used to create 5 kb fragment

1169 SMRTbell libraries. Each library was run using P6-C4 chemistry in one SMRT cell on a RSII
1170 PacBio instrument. PacBio libraries were produced and sequencing was performed at the
1171 Uppsala Genome Center, Uppsala, Sweden.
1172
1173 *Mitochondrial genome assembly*
1174 The two different mitochondrial genomes α and β of the O11 line were generated by mapping
1175 all Illumina reads to the published mitochondrial genome of *D. willistoni* (NCBI accession
1176 number BK006338.1) using BWA-MEM [88] and thereafter extracting all aligned reads and
1177 their mates using SAMtools [89]. The extracted Illumina reads were then subsampled to
1178 achieve an approximate coverage of 100 and assembled together with PacBio filtered
1179 subreads using SPAdes-3.10.1 [90] with multiple different k-mers settings. The β mitotype
1180 consistently came out as a single full length contig, so the best assembly was chosen based
1181 on the contiguity of the putative α mitotype. From this assembly the contig representing the β
1182 mitotype and the contigs with a homogenous coverage likely representing the α mitotype was
1183 taken out and used as trusted contigs together with the Illumina data to generate a better
1184 assembly of the α mitotype. Finally, the Illumina and PacBio data were aligned against the
1185 resulting contigs using BWA, MEM and SW respectively. The resulting bam-file was converted
1186 to an ace-file that was used for manual curation of the sequences in Consed [91]. The
1187 mitochondrial genomes of all other *Drosophila* lines were generated by mapping Illumina
1188 reads to the two O11 mitotypes with BWA-MEM, extracting all aligned reads and their mates
1189 using SAMtools and assembling the extracted reads with SPAdes-3.10.1 with multiple
1190 different k-mers settings. Since PacBio data were only generated from the O11, TP37 and
1191 POA-1 lines, the assemblies of all other lines were run using only Illumina reads. For all lines,
1192 this assembly procedure generated one contig that represented a complete (O11) or nearly
1193 complete circular mitochondrial genome.
1194 To test for heteroplasmy and to correct possible errors in contigs, the Illumina reads
1195 for each line were mapped back to the contig containing a complete or nearly complete
1196 mitotype for that line using BWA-MEM and indels were realigned with the IndelRealigner from
1197 GATK [92]. The resulting bam-file was then used to call variants using Freebayes with a range
1198 of parameters and converted to ace-format for manual curation in Consed. A line was
1199 determined to be heteroplasmic if 1) at least 80 sites (ca 0,5%) had a variant and 2) the
1200 variants had a consistent alternative allele frequency over the whole or a significant part of the
1201 mitochondrial genome. These settings were chosen to avoid calling variant genomes based
1202 on sequence errors and to make sure that only complete mitochondrial genomes (not NUMTs)
1203 were used to decide heteroplasmy status. If a line was seen to be heteroplasmic using these
1204 criteria, two different alternatives were used to resolve the assembly of the second mitotype.
1205 1) The Illumina reads mapping to the O11 α and O11 β mitotype were split into two separate

1206 data sets and assembled separately, 2) in the case of MS that has two α mitotypes, variant
1207 calling was performed with lofreq [93], variants with a frequency between 8-30% were selected
1208 and used to make a new consensus with bcftools and the reads were mapped to the two
1209 variants of the mitochondrial genome. The same procedure was done in multiple rounds until
1210 no more variant sites within the correct frequency were found. The final bam-file was converted
1211 to ace-format and manually inspected in Consed.
1212
1213 **Alignments and phylogenetic analyses**
1214 Datasets of either whole mitochondrial genomes, nuclear markers or mitochondrial genes
1215 were aligned with MAFFT v6.850b (mafft-linsi algorithm) [94] and trimmed with TrimAl
1216 v1.2rev59 [95] using the automated1 for removal of poorly aligned regions. Concatenated sets
1217 of either mitochondrial or nuclear genes were created by combining trimmed alignments of
1218 single genes. Datasets of mitochondrial protein coding genes containing either only the 3rd or
1219 a combination of 1st and 2nd codon positions were created by parsing the alignments with an
1220 in-house script.
1221 Maximum likelihood trees were built with RAxML version 7.2.6 [96] using -f a -m GTRGAMMA
1222 -x 12345 and -# 1000. Bayesian trees were made with MrBayes version 3.2.6 [97, 98] using
1223 "nst= mixed" for estimating evolutionary models. Analyses were run for 10.000.000
1224 generations with "nruns= 2", "nchains= 4" and "samplefreq= 1000". A burn-in of 25% was
1225 applied. Trees were visualized in FigTree v. 1.4.3 [99] and rooted on the species of the *saltans*
1226 group.
1227
1228 **Identification of synonymous and non-synonymous substitutions**
1229 Synonymous and non-synonymous substitutions in pairwise comparisons between lines of the
1230 α x β and α1 x α2 clades were identified in single gene alignments using codeml (PAML
1231 package v. 4.8) [100]. Values obtained from pairwise comparisons between lines were then
1232 used to estimate an average for the clades. The same method was used for calculating the
1233 number of nonsynonymous substitutions per non-synonymous site (dN), synonymous
1234 substitutions per synonymous site (dS) and the dN/dS ratios. Synonymous and non-
1235 synonymous mutations unique to α1 and α2 were counted through visual inspection of single
1236 gene alignments using Aliview v. 2019 [101].
1237
1238 **Assessing rarity and functional impact of α2-unique mutations**
1239 We used the MutationAssessor server (www.mutationsassessor.org) to estimate the potential
1240 impact on protein function for non-synonymous substitutions exclusive to α2. Since the server
1241 only analyses human sequences, we used the Uniprot website (www.uniprot.org) to align *D.*
1242 *melanogaster* and human *CO3*, *ND3* and *ND4* amino acid sequences and identify the

1243 positions in the human genes which correspond to the non-synonymous substitution sites
1244 unique to α2. With that information we generated the following input strings which were then
1245 used in the server: COX3_HUMAN H6N, NU3M_HUMAN T21M and NU4M_HUMAN I391V.
1246 We used Uniprot to assess whether the α2-unique non-synonymous mutation in *CO3* was
1247 found in other species. This was done by searching for "COX3" in the UniProtKB database,
1248 aligning all entries of the first results page (250 entries) and visually checking for the presence
1249 of an N residue in the position of interest. The process was repeated for the subsequent result
1250 pages until 2000 entries had been aligned and checked.
1251
1252 **Quantitative Real-Time PCR (qPCR)**
1253 DNA was extracted by homogenizing whole flies in 200µl Chelex solution at 5% concentration
1254 in distilled water. A total of 10µl of proteinase K (10mg/ml) was added per sample and vials
1255 were incubated for 6 hours at 56°C. After digestion, vials were centrifuged at 16,200 x g for 3
1256 minutes, and the supernatant was transferred to a new vial. DNA was stored at 4°C and
1257 freeze-thaw cycles were avoided when possible.
1258 Samples were run in duplicate using SYBR Green reagent on a StepOnePlus Real Time PCR
1259 machine (Applied Biosystems). Technical repeats with >1 Ct difference were rerun or
1260 discarded. Each reaction contained 10µl Maxima SYBR Green/ROX qPCR Master Mix (2x),
1261 1µl of forward primer, 1µl of reverse primer, 6µl nuclease-free water and 2µl of the DNA
1262 sample. Cycling conditions were 95°C for 20 seconds, followed by 40 cycles of: 95°C for 3
1263 seconds and 60°C for 30 seconds for rps, and *Wolbachia* primers (**Table S11**). Discriminant
1264 mitochondrial primers cycling conditions were 95°C for 20 seconds, followed by 40 cycles of:
1265 95°C for 3 seconds and 55°C for 30 seconds (**Table S11**). Discriminant primers used linked
1266 base technology to increase specificity (Exiqon). Melt curves and the fly control gene *rps17*
1267 were run on each plate to control for successful single-peak amplification and to control for
1268 DNA quality. All primers had >90% efficiency. For absolute quantification, each 96-well qPCR
1269 plate was analyzed using StepOne Software v2.2.2 (Applied Biosystems) and Ct values were
1270 obtained by comparing each primer sample to a single standard curve of known concentration
1271 and using identical threshold and baseline levels for each primer target across plates.
1272 Standard curves were created by amplifying positive control samples using PCR, calculating
1273 DNA concentrations using QBIT, and then serially diluting the sample 1:10 with distilled water
1274 to create a 5-sample curve comprising known concentrations decreasing from 10pmol/ ml.
1275 Samples with Ct values over 30 were classed as negative, confirmed by our negative controls.
1276 As a result, the detection limit was 50 copies per sample for *rps17*, 115 for *CO3* (SNP 1), 735
1277 for *CO3* (SNP 2) and 1627 for *CO3* (SNP 3). This varied for the FG295 qPCR, where samples
1278 with Ct values over 20 were classed as negative for the *CO3* (SNP 2) primer due to very high
1279 amplification of some DNA segments. This was confirmed by negative flies, but led to a high
1280 threshold of detection of 6150 for *CO3* (SNP 2). DNA concentration was calculated for each

1281 sample by comparing sample Ct values back to the equation of the standard curve. DNA copy
1282 numbers were then calculated by using DNA product length to determine DNA concentration.
1283 To control for fly size and extraction efficiency, copy numbers for each sample were
1284 normalized using the fly housekeeping gene as a control, giving a final value in terms of
1285 mitochondrial genome copies to fly nuclear genome copy ratio.
1286 Since we used three independent experimental approaches to determine levels of
1287 heteroplasmy in *D. paulistorum*, i.e., NGS, RFLP-PCR, and qPCR, we compared the
1288 robustness of these. We looked at the levels of α2 in O11 females (**Fig S8**), and found that all
1289 three techniques determined similar levels, hence confirming the robustness of our
1290 experimental approach.
1291
1292 **Cross-mating and embryo collection**
1293 For mass-mating crosses (pools), virgin males and females from different *D. paulistorum* lines
1294 were obtained by collecting newly emerged flies within 2-3 h intervals to avoid any possible
1295 sibmating. Collected flies were kept separated for 5-7 days to ensure their virgin status.
1296 Crosses were set up in pools using 10-20 females and 20-30 males per each vial and in 2-3
1297 replicates. Flies were mated for one week and then transferred to new vials or used for
1298 consequent embryo collecting. F_1 flies emerging from the first vials were collected for DNA
1299 extraction. Mass-mating A28 x O11 hybrids were generated by mating 10 five-day-old virgin
1300 A28 females to 10 O11 males. F_1 individuals were used for assays. For generating
1301 interspecies hybrids *D. paulistorum* O11 males were mated either with *D. equinoxialis* or *D.*
1302 *willistoni* virgin females in pools of 20 females and 60 males per cross. F_1 flies emerging from
1303 the bottles were collected for DNA extraction. Single fly crosses were generated by collecting
1304 virgin flies every 6-8 hours to avoid sibling mating. Collected flies were kept separated for 5-7
1305 days to ensure their virgin status before use in crossings. Single pairs of male and female flies
1306 were mated for two weeks in fresh vials. Emerging F_1 flies were collected, aged to 8-10 days
1307 and frozen for subsequent DNA extraction. For FG295 x O11 and O11 x FG295 crosses,
1308 single-pair sibling mating were continued for four consequent generations.
1309
1310 *Backcrossing*
1311 For backcrossing experiments, heteroplasmic F1 flies obtained from reciprocal crosses
1312 between O11 and FG295 parents (G0) were crossed in pools to the homoplasmic FG295 line.
1313 Virgin F_1 males and females were obtained from the G0 crosses by collecting newly emerged
1314 flies within 2-3 h intervals to avoid sib mating. Collected flies were kept separated for 5-7 days
1315 to ensure their virgin status. Backcross 1 (BC1) was set up in pools using 5-10 females and
1316 5-10 males per each vial. The first 10 F_1 flies of each sex emerging from the vials were
1317 collected for further backcrossing with FG295 line and the remaining F_1 flies were frozen in

1318 liquid nitrogen for future DNA extraction. Backcross 2 (BC2) was set up similarly to BC1.
1319 Between 10 and 20 flies were analyzed for each backcross.
1320
1321 *Collection of staged embryos*
1322 *D. paulistorum* embryos at early (0-30 min AEL), mid (3-4 hrs AEL) and late (18-20 hrs AEL)
1323 stages of development were collected on homemade medium (agar, molasses, cornmeal and
1324 yeast medium) in plastic Petri dishes (3 x 3 cm) using collection chambers at 25°C. Prior to
1325 collecting embryos, 1-2 weeks old flies were kept in these chambers for 2-3 days to adapt to
1326 the new environment. Embryos were then manually picked with a dissection needle and
1327 transferred immediately to ice-cold lysis buffer (Qiagen). We collected 20 embryos for each
1328 sample with at least five biological replicates. Embryos were immediately processed for DNA
1329 extraction or stored at -20°C for no more than one week.
1330
1331 **Recombination analysis**
1332 We checked for intergenic recombination between lines by analyzing single gene trees built
1333 as described in the "Alignments and phylogenetic analyses" methods section. Tests for
1334 intragenic recombination on single-gene and whole-genome alignments were done using
1335 Geneconv v.1.81 and PhiPack with default parameters. Recombination over mutation rates
1336 were calculated for single-gene and whole-genome alignments using ClonalFrame v.1.2 with
1337 default settings.
1338
1339 **Nuclear markers**
1340 Illumina reads from each fly line were separately assembled with both Megahit v.1.1.2 [102]
1341 and SPAdes v. 3.13.0 using default parameters. We then blasted the *D. willistoni* CDSs of the
1342 selected nuclear markers (**Table S12**) to each assembly and retrieved sequences with an
1343 alignment length of over 200 bp and above 70% identity. The obtained contigs were aligned
1344 to the corresponding *D. willistoni* gene with MAFFT v6.850b. The alignments were visualized
1345 with Aliview v. 2019 and inspected for assembly issues and, if necessary, multiple exons were
1346 joined into a single sequence using Aliviews' "merge" function. For each fly line and marker
1347 gene, results obtained from both assemblies were compared and only the one with highest
1348 quality was used in further analyses.
1349
1350 **Identification of nuclear mitochondrial DNA (NUMTs)**
1351 Illumina reads of each line were separately mapped with BWA-MEM v.0.7.10 to the genome
1352 assemblies generated with Megahit v. 1.1.2. The average per-base coverage of each
1353 assembled contig was estimated with SAMtools mpileup and contigs with average coverage
1354 lower than 2 were excluded from the analysis.

1355 Each genomic assembly was blasted to all previously assembled *D. paulistorum*
1356 mitochondrial genomes. Sequences corresponding to high scoring pairs (HSPs) with at least
1357 200 bp length and 90% similarity to mitochondrial DNA were retrieved. Recovered sequences
1358 with more than 97% identity to either α or β genomes were deemed "conserved" and
1359 tentatively assigned to the clade they were most similar to, while those with 90-97% identity
1360 were considered divergent. Sequences retrieved from each assembly were mapped back to
1361 an α and a β mitochondrial genome (C2, O11 beta) with BWA-MEM and the number of bases
1362 covered by them were counted with Samtools mpileup. This was done in order to determine
1363 how much of the mitochondrial genome was covered by contigs belonging to each category
1364 (α, β, diverged). Genes present in the retrieved sequences were identified by blasting each
1365 sequence to a database containing all genes from the A28 mitochondrial genome (CDS, tRNA
1366 and rRNA). All contigs containing mitochondrial sequences were blasted against NCBI's nt
1367 database for identification of potential matches to non-mitochondrial references.
1368 Sequences retrieved from each genomic assembly were considered NUMTs if the
1369 contigs in which they were found a) also contained matches to non-mitochondrial references;
1370 or b) had average coverage similar (± 30%) to the reference nuclear coverage of its line. The
1371 reference nuclear coverage for each line was calculated as the average coverage of the
1372 contigs containing the genes *CAD/r*, *Beta Spectrin* and *FASN3*. These were identified by
1373 blasting *D. willistoni* CDSs of these genes to each assembly. The reference mitochondrial
1374 coverage of each line was defined as the average per base coverage of the longest contig
1375 which had 100% identity to the line's assembled mitochondrial genome(s).
1376
1377 **Statistics**
1378 Statistical analyses varied depending on the data source and distribution. For the qPCR tissue
1379 distribution of α and β mitotypes between different ages and sexes we used ANOVA on log-
1380 transformed data to increase normality. In the qPCR analysis of A28 x O11 hybrids, we used
1381 ANOVA with post-hoc Tukey tests of significance. Due to low numbers of repeats and power
1382 in the intraspecies FG295 x O11 hybrids, Wilcoxon Rank Sum tests were used for significance
1383 testing. Assumptions and variance were checked before use. These analyses were
1384 completed using R version 3.3.2 [103].
1385 In the RFLP-PCR experiments, the α2-levels in embryos from different *D. paulistorum*
1386 semispecies and crosses at different developmental stages were compared using One-way
1387 ANOVA with post-hoc Tukey HSD tests. Quantifications were made with the help of an online
1388 calculator (www.astatsa.com). The raw data showing levels of α2 in percent were obtained
1389 with Image Lab 5.2.1 software and transformed from percentage to arcsine for normal
1390 distribution before applying the statistics. RFLP-data were statistically analyzed using ANOVA
1391 and Student's t tests.

1392

1393 **Fluorescent in situ hybridization (FISH) on DNA**

1394 DNA-FISH was performed on three different tissues: early embryos, mitotic chromosomes
1395 from 3rd instar larval brains and polytene chromosomes from salivary glands. Early embryos
1396 (0-3 h) were collected as stated above, dechorionated in 50% bleach and fixed for 20 min in
1397 3.7% paraformaldehyde in PBX (PBS and 0.15% Triton X-100) mixed with heptane (1:1 v/v).
1398 Then the embryos were postfixed in ice-cold methanol and further processed for FISH
1399 following a protocol by [104] with slight modifications. In situ hybridization was performed at
1400 37°C overnight in hybridization buffer containing approximately 150 ng of probe generated by
1401 direct labeling via PCR with modified dNTPs (Alexa Fluor 488 dUTP and Alexa Fluor 594
1402 dUTP; Jena Bioscience, Germany). The mito-probe was used as a mixture of 5 mitochondrial
1403 loci (CO1, CO2, ND1, ND2 and ND4) of *D. paulistorum*, cloned beforehand in pTZ57R/5 vector
1404 (Thermo Scientific, USA) and isolated plasmids were labeled with the red AF594 fluorophore.
1405 As a hybridization control, we used a cloned probe targeting a 28S rRNA locus of *D.*
1406 *paulistorum* labeled it with the green AF488 fluorophore.

1407 Mitotic chromosomes from 3rd instar larval brains were processed for DNA-FISH according to
1408 the protocol from [105] after hypotonic treatment washing off cytoplasmic compounds and
1409 RNase treatment. DNA-FISH on polytene chromosomes was performed according to [106]. In
1410 both cases we were using the same probes and washing conditions as described above for
1411 fixed whole-mount embryos.

1412 Images were taken with Olympus FW3000 laser scanning confocal microscope using 60x
1413 objective and further processed with Photoshop CS6 adjusting levels for each channel.

1414 ACKNOWLEDGEMENTS

1415 We thank Matthias Schaefer, Lee Ehrman, DeEtte Walker, Martin Kapun and Silvia Bulgheresi
1416 for critical and constructive discussions during the whole period of the project and Wolfgang
1417 Däuble for valuable contributions in the early stage of the project, Jonathan Haars for valuable
1418 contributions to preliminary data for embryos and both Julia Mras, Christian Binnyei and Uliana
1419 Smirnova for excellent technical support. Sequencing was performed at the SNP&SEQ
1420 Technology Platform and Uppsala Genome Center in Uppsala, Sweden, which is part of the
1421 Swedish National Genomics Infrastructure. Finally, we want to thank Aurélie Hua-Van for fly
1422 sampling and the Nouragues research field station (managed by CNRS), which benefits from
1423 "Investissement d'Avenir" grants managed by Agence Nationale de la Recherche (AnaEE
1424 France ANR-11-INBS-0001; Labex CEBA ANR-10-LABX-25-01). The project was funded by
1425 the Swedish research council VR grant 2014-4353 to LK, and by the Austrian Science Fund
1426 FWF grant P28255-B22 to WJM.

1427

1428 # DATA AVAILABILITY

1429 Raw sequence data used in this project as well as assembled whole mitochondrial genomes
1430 and nuclear markers are available in NCBI under the Bioproject PRJNA643793. Accession
1431 numbers of individual mitochondrial genomes are listed in **Table S2**.

1432

1433 # REFERENCES

1434 1. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human
1435 mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of
1436 America. 1980;77(11):6715-9. doi: 10.1073/pnas.77.11.6715. PubMed PMID: 6256757.
1437 2. Birky CW. Relaxed cellular controls and organelle heredity. Science.
1438 1983;222(4623):468. doi: 10.1126/science.6353578.
1439 3. Stefano G, B., Bjenning C, Wang F, Wang N, Kream R, M. . Mitochondrial
1440 Heteroplasmy. In: Santulli G, editor. Mitochondrial Dynamics in Cardiovascular Medicine
1441 Advances in Experimental Medicine and Biology. 982. Cham: Springer; 2017.
1442 4. Burr SP, Pezet M, Chinnery PF. Mitochondrial DNA Heteroplasmy and Purifying
1443 Selection in the Mammalian Female Germ Line. Development, Growth & Differentiation.
1444 2018;60(1):21-32. doi: 10.1111/dgd.12420.
1445 5. White DJ, Wolff JN, Pierson M, Gemmell NJ. Revealing the hidden complexities of
1446 mtDNA inheritance. Molecular Ecology. 2008;17(23):4925-42. doi: 10.1111/j.1365-
1447 294X.2008.03982.x.
1448 6. Sharpley MS, Marciniak C, Eckel-Mahan K, McManus M, Crimi M, Waymire K, et al.
1449 Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and
1450 cognition. Cell. 2012;151(2):333-43. doi: 10.1016/j.cell.2012.09.004. PubMed PMID:
1451 23063123.
1452 7. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy:
1453 implications for human health and disease. Nature Reviews Genetics. 2015;16(9):530-42. doi:
1454 10.1038/nrg3966.
1455 8. Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, et al.
1456 Cytonuclear integration and co-evolution. Nature Reviews Genetics. 2018;19(10):635-48. doi:
1457 10.1038/s41576-018-0035-9.
1458 9. Christie JR, Schaerf TM, Beekman M. Selection against Heteroplasmy Explains the
1459 Evolution of Uniparental Inheritance of Mitochondria. PLOS Genetics. 2015;11(4):e1005112.
1460 doi: 10.1371/journal.pgen.1005112.
1461 10. Birky CW. Uniparental inheritance of mitochondrial and chloroplast genes:
1462 mechanisms and evolution. Proceedings of the National Academy of Sciences.
1463 1995;92(25):11331. doi: 10.1073/pnas.92.25.11331.
1464 11. Pitnick S, Karr TL. Paternal products and by-products in *Drosophila* development.
1465 Proceedings of the Royal Society B: Biological Sciences. 1998;265(1398):821-6. doi:
1466 10.1098/rspb.1998.0366. PubMed PMID: 9628039.
1467 12. Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G.
1468 Ubiquitin tag for sperm mitochondria. Nature. 1999;402(6760):371-2. doi: 10.1038/46466.
1469 13. DeLuca Steven Z, O'Farrell Patrick H. Barriers to Male Transmission of Mitochondrial
1470 DNA in Sperm Development. Developmental Cell. 2012;22(3):660-8. doi:
1471 https://doi.org/10.1016/j.devcel.2011.12.021.
1472 14. Yu Z, O'Farrell PH, Yakubovich N, DeLuca SZ. The Mitochondrial DNA Polymerase
1473 Promotes Elimination of Paternal Mitochondrial Genomes. Current Biology. 2017;27(7):1033-
1474 9. doi: 10.1016/j.cub.2017.02.014.
1475 15. Sato K, Sato M. Multiple ways to prevent transmission of paternal mitochondrial DNA
1476 for maternal inheritance in animals. The Journal of Biochemistry. 2017;162(4):247-53. doi:
1477 10.1093/jb/mvx052.
1478 16. Wolff JN, Gemmell NJ. Lost in the zygote: the dilution of paternal mtDNA upon
1479 fertilization. Heredity. 2008;101(5):429-34. doi: 10.1038/hdy.2008.74.

17. Passamonti M, Ghiselli F. Doubly Uniparental Inheritance: Two Mitochondrial Genomes, One Precious Model for Organelle DNA Inheritance and Evolution. DNA and Cell Biology. 2009;28(2):79-89. doi: 10.1089/dna.2008.0807.
18. Zouros E, Ball AO, Saavedra C, Freeman KR. Mitochondrial DNA inheritance. Nature. 1994;368(6474):818-. doi: 10.1038/368818a0.
19. Doublet V, Souty-Grosset C, Bouchon D, Cordaux R, Marcadé I. A thirty million year-old inherited heteroplasmy. PLoS One. 2008;3(8):e2938-e. doi: 10.1371/journal.pone.0002938. PubMed PMID: 18698356.
20. Doublet V, Raimond R, Grandjean F, Lafitte A, Souty-Grosset C, Marcadé I. Widespread atypical mitochondrial DNA structure in isopods (Crustacea, Peracarida) related to a constitutive heteroplasmy in terrestrial species. Genome. 2012;55(3):234-44. doi: 10.1139/g2012-008.
21. Peccoud J, Chebbi MA, Cormier A, Moumen B, Gilbert C, Marcadé I, et al. Untangling Heteroplasmy, Structure, and Evolution of an Atypical Mitochondrial Genome by PacBio Sequencing. Genetics. 2017;207(1):269. doi: 10.1534/genetics.117.203380.
22. Robison GA, Balvin O, Schal C, Vargo EL, Booth W. Extensive Mitochondrial Heteroplasmy in Natural Populations of a Resurging Human Pest, the Bed Bug (Hemiptera: Cimicidae). Journal of medical entomology. 2015;52(4):734-8. Epub 2015/05/23. doi: 10.1093/jme/tjv055. PubMed PMID: 26335484.
23. Meza-Lázaro RN, Poteaux C, Bayona-Vásquez NJ, Branstetter MG, Zaldívar-Riverón A. Extensive mitochondrial heteroplasmy in the neotropical ants of the *Ectatomma ruidum* complex (Formicidae: Ectatomminae). Mitochondrial DNA Part A. 2018;29(8):1203-14. doi: 10.1080/24701394.2018.1431228.
24. Mastrantonio V, Latrofa MS, Porretta D, Lia RP, Parisi A, Iatta R, et al. Paternal leakage and mtDNA heteroplasmy in *Rhipicephalus spp* ticks. Scientific reports. 2019;9(1):1460. doi: 10.1038/s41598-018-38001-8.
25. Kastally C, Mardulyn P. Widespread co-occurrence of two distantly related mitochondrial genomes in individuals of the leaf beetle *Gonioctena intermedia*. Biology Letters. 2017;13(11):20170570. doi: 10.1098/rsbl.2017.0570.
26. Solignac M, Monnerot M, Mounolou JC. Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*. Proceedings of the National Academy of Sciences of the United States of America. 1983;80(22):6942-6. doi: 10.1073/pnas.80.22.6942. PubMed PMID: 6316335.
27. Matsuura ET, Fukuda H, Chigusa SI. Mitochondrial DNA heteroplasmy maintained in natural populations of *Drosophila simulans* in Réunion. Genetical Research. 1991;57(2):123-6. Epub 2009/04/14. doi: 10.1017/S0016672300029189.
28. Dean MD, Ballard KJ, Glass A, Ballard JW. Influence of two *Wolbachia* strains on population structure of East African *Drosophila* simulans. Genetics. 2003;165(4):1959-69. Epub 2004/01/06. PubMed PMID: 14704179; PubMed Central PMCID: PMC1462921.
29. Wolff JN, Nafisinia M, Sutovsky P, Ballard JWO. Paternal transmission of mitochondrial DNA as an integral part of mitochondrial inheritance in metapopulations of *Drosophila simulans*. Heredity. 2013;110(1):57-62. Epub 2012/09/26. doi: 10.1038/hdy.2012.60. PubMed PMID: 23010820.
30. Nunes MDS, Dolezal M, Schlötterer C. Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*. Molecular ecology. 2013;22(8):2106-17. Epub 2013/03/04. doi: 10.1111/mec.12256. PubMed PMID: 23452233.
31. Volz-Lingenhöhl A, Solignac M, Sperlich D. Stable heteroplasmy for a large-scale deletion in the coding region of Drosophila subobscura mitochondrial DNA. Proceedings of the National Academy of Sciences. 1992;89(23):11528. doi: 10.1073/pnas.89.23.11528.
32. Lecher P, Béziat F, Alziari S. Tissular distribution of heteroplasmy and ultrastructural studies of mitochondria from a *Drosophila subobscura* mitochondrial deletion mutant. Biology of the Cell. 1994;80(1):25-33. doi: 10.1016/0248-4900(94)90013-2.
33. Kondo R, Matsuura ET, Chigusa SI. Further observation of paternal transmission of *Drosophila* mitochondrial DNA by PCR selective amplification method. Genetical Research. 1992;59(2):81-4. Epub 2009/04/14. doi: 10.1017/S0016672300030287.

34. Sherengul W, Kondo R, Matsuura ET. Analysis of paternal transmission of mitochondrial DNA in *Drosophila*. Genes & Genetic Systems. 2006;81(6):399-404. doi: 10.1266/ggs.81.399.
35. Dokianakis E, Ladoukakis ED. Different degree of paternal mtDNA leakage between male and female progeny in interspecific *Drosophila* crosses. Ecology and evolution. 2014;4(13):2633-41. Epub 2014/06/19. doi: 10.1002/ece3.1069. PubMed PMID: 25077015.
36. Hahn A, Zuryn S. The Cellular Mitochondrial Genome Landscape in Disease. Trends in Cell Biology. 2019;29(3):227-40. doi: 10.1016/j.tcb.2018.11.004.
37. Havird JC, Forsythe ES, Williams AM, Werren JH, Dowling DK, Sloan DB. Selfish mitonuclear conflict. Current Biology. 2019;29(11):R496-R511. doi: https://doi.org/10.1016/j.cub.2019.03.020.
38. Oliver KM, Degnan PH, Burke GR, Moran NA. Facultative Symbionts in Aphids and the Horizontal Transfer of Ecologically Important Traits. Annual Review of Entomology. 2009;55(1):247-66. doi: 10.1146/annurev-ento-112408-085305.
39. Sloan DB, Moran NA. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Molecular biology and evolution. 2012;29(12):3781-92. Epub 2012/07/19. doi: 10.1093/molbev/mss180. PubMed PMID: 22821013.
40. Ferrari J, Vavre F. Bacterial symbionts in insects or the story of communities affecting communities. Philosophical Transactions of the Royal Society B: Biological Sciences. 2011;366(1569):1389-400. doi: 10.1098/rstb.2010.0226.
41. Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A. *Serratia symbiotica* from the Aphid *Cinara cedri*: A Missing Link from Facultative to Obligate Insect Endosymbiont. PLOS Genetics. 2011;7(11):e1002357. doi: 10.1371/journal.pgen.1002357.
42. McCutcheon JP, McDonald BR, Moran NA. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proceedings of the National Academy of Sciences. 2009. Epub 2009/08/27. doi: 0906424106 [pii]
10.1073/pnas.0906424106. PubMed PMID: 19706397.
43. Burla H, Da Cunha AB, Cordeiro AR, Dobzhansky T, Malogolowkin C, Pavan C. The *willistoni* group of sibling species of *Drosophila*. Evolution. 1949;3:300-14. doi: 10.1111/j.1558-5646.1949.tb00031.x.
44. Ehrman L, Powell JR. The *Drosophila willistoni* species group. In: Ashburner M, Carson HL, Thompson Jr. JN, editors. The Genetics and Biology of *Drosophila*. 3b. New York: Academic Press Inc; 1982. p. 193-220.
45. Dobzhansky T, Ehrman L, Pavlovsky O, Spassky B. The superspecies *Drosophila paulistorum*. Proceedings of the National Academy of Sciences. 1964;51(1):3. doi: 10.1073/pnas.51.1.3.
46. Dobzhansky T, Spassky B. *Drosophila paulistorum*, a cluster of species in *statu nascendi*. Proceedings of the national Academy of Sciences. 1959;45(3):419-28.
47. Miller WJ, Ehrman L, Schneider D. Infectious speciation revisited: impact of symbiont-depletion on female fitness and mating behavior of *Drosophila paulistorum*. PLoS Pathogens. 2010;6(12):e1001214. Epub 2010/12/15. doi: 10.1371/journal.ppat.1001214. PubMed PMID: 21151959; PubMed Central PMCID: PMCPMC2996333.
48. Miller W, J, , Schneider D. Endosymbiotic microbes as adaptive manipulators of arthropod behaviour and natural driving sources of host speciation. In: Hughes D, Brodeur J, Thomas F, editors. Host manipulation by parasites. Oxford: University Press; 2012. p. 119-39.
49. Schneider DI, Ehrman L, Engl T, Kaltenpoth M, Hua-Van A, Le Rouzic A, et al. Symbiont-driven male mating success in the Neotropical *Drosophila paulistorum* superspecies. Behavior Genetics. 2019;49(1):83-98. doi: 10.1007/s10519-018-9937-8.
50. Baião GC, Schneider DI, Miller WJ, Klasson L. The effect of *Wolbachia* on gene expression in *Drosophila paulistorum* and its implications for symbiont-induced host speciation. BMC Genomics. 2019;20(1):465. doi: 10.1186/s12864-019-5816-9.
51. Strunov A, Schneider DI, Albertson R, Miller WJ. Restricted distribution and lateralization of mutualistic *Wolbachia* in the *Drosophila* brain. Cellular Microbiology. 2017;19(1). doi: 10.1111/cmi.12639. PubMed PMID: 27353950.

52. Gleason JM, Griffith EC, Powell JR. A molecular phylogeny of the *Drosophila willistoni* group: Conflicts between species concepts? Evolution. 1998;52(4):1093-103. doi: 10.2307/2411239.
53. Robe LJ, Cordeiro J, Loreto ELS, Valente VLS. Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). Genetica. 2010;138(6):601-17. doi: 10.1007/s10709-009-9432-5.
54. Zanini R, Müller MJ, Vieira GC, Valiati VH, Deprá M, Valente VLdS. Combining morphology and molecular data to improve *Drosophila paulistorum* (Diptera, Drosophilidae) taxonomic status. Fly. 2018:1-14. doi: 10.1080/19336934.2018.1429859.
55. Ross CL, Harrison RG. A fine scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus* Evolution. 2002;56(11):2296-312. doi: 10.1111/j.0014-3820.2002.tb00153.x.
56. Toews DPL, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology. 2012;21(16):3907-30. doi: 10.1111/j.1365-294X.2012.05664.x.
57. Zhang D, Tang L, Cheng Y, Hao Y, Xiong Y, Song G, et al. “Ghost Introgression” As a Cause of Deep Mitochondrial Divergence in a Bird Species Complex. Molecular Biology and Evolution. 2019;36(11):2375-86. doi: 10.1093/molbev/msz170.
58. Winge H. Interspecific hybridisation between the six cryptic species of the *Drosophila willistoni* group. Heredity. 1965;20:9-19.
59. Wolff JN, Sutovsky P, Ballard JWO. Mitochondrial DNA content of mature spermatozoa and oocytes in the genetic model *Drosophila*. Cell and Tissue Research. 2013;353(1):195-200. doi: 10.1007/s00441-013-1628-4.
60. Rubenstein JLR, Brutlag D, Clayton DA. The mitochondrial DNA of *Drosophila melanogaster* exists in two distinct and stable superhelical forms. Cell. 1977;12(2):471-82. doi: 10.1016/0092-8674(77)90123-4.
61. Yamane T, Miller DI, Hopfield JJ. Interaction of elongation factor Tu with the aminoacyl transfer ribonucleic acid dimer Phe-tRNA-Glu-tRNA. Biochemistry. 1981;20(20):449-52.
62. Moazed D, Robertson JM, Noller HF. Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. Nature. 1988;334(6180):362-4. doi: 10.1038/334362a0.
63. Marygold SJ, Attrill H, Lasko P. The translation factors of *Drosophila melanogaster*. Fly. 2017;11(1):65-74. Epub 2016/08/05. doi: 10.1080/19336934.2016.1220464. PubMed PMID: 27494710.
64. Sato A, Suematsu T, Aihara K-k, Kita K, Suzuki T, Watanabe K, et al. Duplication of *Drosophila melanogaster* mitochondrial EF-Tu: pre-adaptation to T-arm truncation and exclusion of bulky aminoacyl residues. Biochemical Journal. 2017;474(6):957-69. doi: 10.1042/BCJ20160929.
65. Bruen TC, Philippe H, Bryant D. A Simple and Robust Statistical Test for Detecting the Presence of Recombination. Genetics. 2006;172(4):2665. doi: 10.1534/genetics.105.048975.
66. Balciuniene J, Balciunas D. A Nuclear mtDNA Concatemer (Mega-NUMT) Could Mimic Paternal Inheritance of Mitochondrial Genome. Frontiers in Genetics. 2019;10:518.
67. Luo S, Valencia CA, Zhang J, Lee N-C, Slone J, Gui B, et al. Biparental Inheritance of Mitochondrial DNA in Humans. Proceedings of the National Academy of Sciences. 2018;115(51):13039. doi: 10.1073/pnas.1810946115.
68. Wei W, Pagnamenta AT, Gleadall N, Sanchis-Juan A, Stephens J, Broxholme J, et al. Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans. Nature communications. 2020;11(1):1740. doi: 10.1038/s41467-020-15336-3.
69. Lutz-Bonengel S, Niederstätter H, Naue J, Koziel R, Yang F, Sänger T, et al. Evidence for multi-copy Mega-NUMTs in the human genome. Nucleic Acids Research. 2021;49(3):1517-31. doi: 10.1093/nar/gkaa1271.
70. Hanlon SL, Hawley RS. B Chromosomes in the *Drosophila* Genus. Genes (Basel). 2018;9(10). doi: 10.3390/genes9100470.
71. Werner M, Simmons LW. Insect Sperm Motility. Biological Reviews. 2008;83(2):191-208. doi: 10.1111/j.1469-185X.2008.00039.x.
72. du Plessis SS, Agarwal A, Mohanty G, van der Linde M. Oxidative phosphorylation versus glycolysis: what fuel do spermatozoa use? Asian J Androl. 2015;17(2):230-5. doi: 10.4103/1008-682X.135123. PubMed PMID: 25475660.

73. Chiang ACY, McCartney E, O'Farrell PH, Ma H. A Genome-wide Screen Reveals that Reducing Mitochondrial DNA Polymerase Can Promote Elimination of Deleterious Mitochondrial Mutations. Current Biology. 2019;29(24):4330-6.e3. doi: https://doi.org/10.1016/j.cub.2019.10.060.
74. Politi Y, Gal L, Kalifa Y, Ravid L, Elazar Z, Arama E. Paternal Mitochondrial Destruction after Fertilization Is Mediated by a Common Endocytic and Autophagic Pathway in *Drosophila*. Developmental Cell. 2014;29(3):305-20. doi: https://doi.org/10.1016/j.devcel.2014.04.005.
75. Birky CW. The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models. Annual Review of Genetics. 2001;35(1):125-48. doi: 10.1146/annurev.genet.35.102401.090231.
76. St. John J. The control of mtDNA replication during differentiation and development. Biochimica et Biophysica Acta. 2014;1840(4):1345-54. doi: https://doi.org/10.1016/j.bbagen.2013.10.036.
77. Gemmell NJ, Metcalf VJ, Allendorf FW. Mother's curse: the effect of mtDNA on individual fitness and population viability. Trends in Ecology & Evolution. 2004;19(5):238-44. doi: 10.1016/j.tree.2004.02.002.
78. Petit N, Touraille S, Debise R, Morel F, Renoux M, Lécher P, et al. Developmental changes in heteroplasmy level and mitochondrial gene expression in a *Drosophila subobscura* mitochondrial deletion mutant. Current genetics. 1998;33(5):330-9.
79. Huang DI, Hefer CA, Kolosova N, Douglas CJ, Cronk QCB. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). New Phytologist. 2014;204(3):693-703. doi: 10.1111/nph.12956.
80. Ehrman L, Kernaghan RP. Infectious Heredity in *Drosophila paulistorum*. Ciba Foundation Symposium 6 - Pathogenic Mycoplasmas. 1972:227-50. doi: doi:10.1002/9780470719893.ch11
10.1002/9780470719893.ch11.
81. Dobzhansky T, Pavlovsky O. Experiments on the Incipient Species of the *Drosophila paulistorum* complex. Genetics. 1967;55(1):141-56. PubMed PMID: PMC1211364.
82. Ehrman L. The transitional races of *Drosophila paulistorum*: a study of hybrid sterility Proceedings of the National Academy of Sciences. 1962;48(2):157. doi: 10.1073/pnas.48.2.157.
83. Malogolowkin C, Simmons AS, Levene H. On the nature of the transitional strains in *Drosophila paulistorum*. Proceedings of the National Academy of Sciences. 1964;51(2):260. doi: 10.1073/pnas.51.2.260.
84. Turelli M, Hoffmann AA, McKechnie SW. Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. Genetics. 1992;132(3):713.
85. Hurst GD, Jiggins FM. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B: Biological Sciences. 2005;272(1572):1525-34. Epub 2005/07/29. doi: 10.1098/rspb.2005.3056. PubMed PMID: 16048766; PubMed Central PMCID: PMCPMC1559843.
86. Junakovic N. Southern Blot Analysis of Individual *Drosophila* Flies. In: Miller WJ, Capy P, editors. Mobile Genetic Elements Methods in Molecular Biology. 260: Humana Press; 2004.
87. Ellegaard KM, Klasson L, Naslund K, Bourtzis K, Andersson SG. Comparative genomics of *Wolbachia* and the bacterial species concept. PLoS Genetics. 2013;9(4):e1003381. Epub 2013/04/18. doi: 10.1371/journal.pgen.1003381. PubMed PMID: 23593012; PubMed Central PMCID: PMCPMC3616963.
88. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997v2. 2013.
89. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/08. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943.
90. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of

Computational Biology. 2012;19(5):455-77. doi: 10.1089/cmb.2012.0021. PubMed PMID: 22506599; PubMed Central PMCID: PMC3342519.
91. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. Genome Research. 1998;8(3):195-202. PubMed PMID: 9521923.
92. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010;20(9):1297-303. doi: 10.1101/gr.107524.110.
93. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic acids research. 2012;40(22):11189-201. Epub 2012/10/12. doi: 10.1093/nar/gks918. PubMed PMID: 23066108.
94. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution. 2013;30(4):772-80. doi: 10.1093/molbev/mst010.
95. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3. doi: 10.1093/bioinformatics/btp348. PubMed PMID: 19505945; PubMed Central PMCID: PMC2712344.
96. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. Epub 2014/01/21. doi: 10.1093/bioinformatics/btu033. PubMed PMID: 24451623.
97. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17(8):754-5. doi: 10.1093/bioinformatics/17.8.754.
98. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572-4. doi: 10.1093/bioinformatics/btg180.
99. Rambaut A. FigTree v1.4: Tree Figure Drawing Tool. Available at: http://tree.bio.ed.ac.uk/software/figtree/2009.
100. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 2007;24(8):1586-91. doi: 10.1093/molbev/msm088.
101. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30(22):3276-8. Epub 2014/08/05. doi: 10.1093/bioinformatics/btu531. PubMed PMID: 25095880.
102. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6. doi: 10.1093/bioinformatics/btv033.
103. R-Core-Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical computing; 2015.
104. Gemkow MJ, Buchenau P, Arndt-Jovin DJ. FISH in whole-mount *Drosophila* embryos. RNA: activation of a transcriptional locus, DNA: gene architecture and expression. Bioimaging. 1996;4(2):107-20. doi: https://doi.org/10.1002/1361-6374(199606)4:2<107::AID-BIO8>3.0.CO;2-Q.
105. Bewley GC. *Drosophila*: A practical approach. Edited by D.B. Roberts Oxford-Washington, DC: IRL Press, 1986. 295 pp. Developmental Genetics. 1987;8(1):59-60. doi: https://doi.org/10.1002/dvg.1020080108.
106. Biémont C, Monti-Dedieu L, Lemeunier F. Detection of transposable elements in *Drosophila* salivary gland polytene chromosomes by in situ hybridization. Methods in molecular biology 2004;260:21-8. doi: 10.1385/1-59259-755-6:021. PubMed PMID: 15020799.