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 Thoma², 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, Schwarzschanerstraße 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 $\alpha 2$ 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 $\alpha 2$ 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 $\alpha 2$ can be transmitted to their offspring via both parents and that levels are dependent on nuclear background. Following $\alpha 2$ 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 precoc” phenotype likely prevents the $\alpha 2$ 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 $\alpha 2$ 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

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

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

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

far observed exclusively in heteroplasmic lines. This finding implies that $\alpha 2$ is unable to provide full mitochondrial functionality, and might therefore be regarded as a secondary mitotype, in contrast to the fully functional primary mitotypes ($\alpha 1$, $\beta 1$ and $\beta 2$), which can exist in homoplasmy.

In addition, the $\alpha 2$ mitotype was found in 41% of the tested *D. paulistorum* lines, where it is stably maintained at variable but significant levels with up to 66% depending on host background. Unexpectedly, this secondary mitotype, unlike the primary one we investigated, was not distributed within the fly in a way that correlated with the expected OXPHOS-mediated ATP demands of different tissues. Yet, the $\alpha 2$ genome showed no signs of deterioration, suggesting it may perform some unknown function for the host. Additionally, we demonstrated that the $\alpha 2$ mitotype is not integrated as a Mega-NUMT in the chromosomes but a cytoplasmic compound, which is biparentally transmitted and starts replicating rapidly after egg fertilization, while the primary mitotype is still quiescent. This replication advantage, which we name the “Replication precox” phenotype, allows $\alpha 2$ to reach high levels during early embryogenesis and thus avoid being outcompeted by the much larger number of maternally inherited primary mitotypes. Hence, we propose that $\alpha 2$ is a newly discovered selfish mitotype that can persist and spread in and between populations, and possibly even across species by introgression, via the combination of biparental transmission and the Replication precox phenotype.

Using phylogenetic reconstruction, we observed a complex evolutionary history of mitochondria in *D. paulistorum*, with both ancient and recent introgressions. We suggest that an α -like mitochondrion was the ancestral mitotype of the species and that heteroplasmy might have originated through introgression of the β mitotype from an unsampled or extinct donor, a concept known as ‘ghost’ introgression [55-57]. We speculate that the unusual biparental transmission mode of $\alpha 2$ might have developed through competition with β mitochondria for maternal transmission.

RESULTS

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

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

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

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

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

Next, we used Illumina sequencing to screen for heteroplasmy on a genome-wide level in *D. 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.

Importantly, no loss-of-function mutations were found, indicating that all genomes are likely to encode fully functional mitochondria. The assembled mitochondrial genomes also show considerably higher read coverage than nuclear contigs from the same lines, which distinguishes them from potential nuclear mitochondrial DNA sequences (NUMTs) (Table S3). All of the assembled mitochondrial genomes have short control regions (CR), between 1.1 kb in the willistoni group and 1.3 kb in the saltans group, compared to the 4.6 kb CR in *D. melanogaster*. A few substitutions occur between *D. paulistorum* lines in or near the putative origins of replication, but no major differences in length are observed between CRs (Table S2, Fig S2).

Phylogenetic analysis of the assembled genomes revealed that the mitochondria of *D. paulistorum* are polyphyletic and split into two major clades (Fig 2B), the α and β mitotypes that were previously detected in single flies of O11 (Fig 1). The mitochondrial genomes of the α and β clades are relatively distantly related to each other with an average nucleotide divergence of 2.94%. A majority of the substitutions that occur between them in protein coding genes are synonymous (Tables S4, S5), indicating that purifying selection is acting on the genomes of both clades. Based on the phylogeny, the α and β clades can be further divided into two subclades, named 1 and 2 (Fig 2B). Interestingly, while the α_1 , β_1 and β_2 mitotypes occur both in homoplasmic and heteroplasmic lines, the α_2 mitotype is never found in homoplasmy and is present in all heteroplasmic lines, where it is partnered with a mitotype from another subclade.

Level and prevalence of heteroplasmy in D. paulistorum lines

To further investigate the heteroplasmy of *D. paulistorum*, we assessed its prevalence (the percentage of heteroplasmic flies within a line) and levels (percentage of each mitotype within a sample) using mitotype-specific RFLP-PCR. Seventeen lines belonging to the six *D. paulistorum* semispecies were tested and seven (41%) displayed heteroplasmy (Table 1). Apart from the lines already identified as heteroplasmic through genome sequencing (Fig 2B), we further identified heteroplasmy in the two Amazonian lines FG16 and FG572 as well as in the Andean-Brazilian Yellow lines. Interestingly, the Orinocan FG103 line was clearly heteroplasmic at the time of sequencing, shortly after its field collection and the establishment of the isofemale line in 2014/2015 (Table 1), but thereafter lost the α_2 mitotype in consecutive generations. Heteroplasmic lines were thus observed in the Amazonian, Andean-Brazilian and Orinocan semispecies and originate from both long-term lab lines and recently collected field samples (Table S1).

Table 1. Levels of $\alpha 2$ and prevalence of heteroplasmy in *D. paulistorum* lines based on Illumina sequencing of ovary samples ($\alpha 2$ level Illumina) and RFLP-PCR from adult flies ($\alpha 2$ level RFLP). A dash (-) in the $\alpha 2$ level from Illumina sequencing means that we are unable to determine that the $\alpha 2$ mitotype existed. Levels of $\alpha 2$ 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	$\alpha 2$ level Illumina	$\alpha 2$ level RFLP	Het. prevalence
O11	1957	OR	$\alpha 2/\beta 2$	19.4%	30.7 $\pm$ 3.7/36.8 $\pm$ 3.6	100% (39/39)
FG103	2014	OR	$\alpha 2/\beta 2$	6.5%	ND [§]	0% (0/40) [§]
FG111	2014	OR	$\alpha 2/\beta 2$	2.3%	13.4 $\pm$ 1.1/18.3 $\pm$ 2.9	25% (10/40)
FG295	2014	OR	$\beta 2$	-	0	0% (0/40)
TP37	2009	OR	$\beta 2$	-	0	0% (0/20)
POA1	2003	OR	$\beta 2$	-	0	0% (0/20)
RP	1995	OR	$\beta 2$	-	0	0% (0/20)
A28	1952	AM	$\beta 1$	-	0	0% (0/40)
FG16	2018	AM	$\alpha 2/\beta 1$	ND	57.2 $\pm$ 8.6/66.5 $\pm$ 7.2	55% (22/40)
FG18	2018	AM	$\beta 1$	ND	0	0% (0/40)
FG572	2015	AM	$\alpha 2/\beta 1$	49.5%	45.3 $\pm$ 8.1/63.7 $\pm$ 7.0	90% (36/40)
MS	1962	AB	$\alpha 1/\alpha 2$	19.1%	21.8 $\pm$ 3.1/36.0 $\pm$ 4.1	100% (40/40)
Yellow	1960	AB	$\alpha 1/\alpha 2$	ND	20.4 $\pm$ 4.7/27.4 $\pm$ 4.3	100% (40/40)
C2	1954	CA	$\alpha 1$	-	0	0% (0/20)
White	1960	CA	$\alpha 1$	ND	0	0% (0/20)
L1	1958	IN	$\beta 2$	-	0	0% (0/20)
SM	1956	TR	$\beta 2$	-	0	0% (0/20)

[§]The line FG103 was heteroplasmic in 2014/15 (determined by Sanger and Illumina sequencing), but has lost the $\alpha 2$ 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, $\alpha 2$ prevalence ranges between 25% and 90% in the recently collected isofemale lines FG111, FG16 and FG572 (**Table 1**). Levels of $\alpha 2$ 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 $\alpha 2$ levels (**Table 1**). However, the two lab lines of the Andean-Brazilian semispecies (MS, Yellow) showed levels of $\alpha 2$ that were similar to each other at 20%. Finally, the highest $\alpha 2$ 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 $\alpha 2$ 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 $\alpha 2$ 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 $\alpha 2$, of the primary mitotype ($\alpha 1$, $\beta 1$ or $\beta 2$), or of both.

Titers of $\alpha 2$ and $\beta 2$ mitotypes in *D. paulistorum* tissues

To explore the dynamics and potential function of $\alpha 2$ and $\beta 2$ mitochondria, we measured the genome copy numbers of each mitotype relative to nuclear genome copies (from here on referred to as titer) in the heteroplasmic O11 line of the Orinocan semispecies. Since mitochondrial DNA-copy numbers are variable in different tissues and under different conditions, their location and titer can provide clues as to their function. Hence, besides whole flies we also investigated $\alpha 2$ and $\beta 2$ titers in the head, thorax, abdominal carcasses without gonads and gonads of male and female O11 flies using discriminant qPCR. We also explored the potential influence of sex, age (**Fig 3; Fig S4**), and also infection with the endosymbiont *Wolbachia* (**Fig S5; Supplemental data file**) on titers of both mitotypes. We found that titers of the primary $\beta 2$ mitotype (**Fig 3A**) were significantly influenced by tissue ($F_{4,80}=25.44$, $p<0.001$). As expected, $\beta 2$ is found in higher titers in energy-demanding tissues where ATP is provided by oxidative phosphorylation (OXPHOS) such as in the flight muscle of the thorax ([5], and references therein). Further, we found a weak significant interaction between age and sex ($F_{1,80}=6.12$, $p<0.05$) potentially driven by younger male flies having slightly more of the $\beta 2$ mitotype than older males, in combination with different tissue location patterns between the sexes. Females had overall lower numbers of the primary $\beta 2$ mitotype per nuclear genome copy than males ($F_{1,80}=28.68$, $p<0.001$).

In contrast to $\beta 2$, we found no significant effect of fly tissue ($F_{4,80}=1.426$, $p=0.23$) or age ($F_{1,80}=0$, $p=0.99$) on the titer of $\alpha 2$. Additionally, there was no significant effect of *Wolbachia* infection on the titer of either the $\beta 2$ ($F_{1,80}=1.41$, $p=0.24$) or the $\alpha 2$ ($F_{1,80}=0$, $p=0.97$) mitotypes (**Fig S5**). The copy number of $\alpha 2$ is, however, significantly affected by fly sex (**Fig 3B**, $F_{1,80}=51.34$, $p<0.001$), with males carrying approximately twice as many $\alpha 2$ copies as females per nuclear genome. Significant higher titers of $\alpha 2$ in O11 males corroborate our finding of generally higher $\alpha 2$ levels in heteroplasmic males compared to heteroplasmic female flies (**Table 1; Fig S3 and Supplemental Data file**), confirming that the significant differences between the sexes in the percentage data are due to more copies of $\alpha 2$, and not for example, lower $\beta 2$ numbers.

Although not fully conclusive, the fact that only the titers of the $\beta 2$ mitotype varied depending on fly tissue, but not the ones of $\alpha 2$, is in line with our other results suggesting that $\beta 2$ is the functional and hence the primary mitotype and that $\alpha 2$ might be dysfunctional and/or not able to generate sufficient energy for its host.

Fig 3. Absolute copies of mtDNA from $\beta 2$ and $\alpha 2$ mitotypes in O11 *D. paulistorum* flies, relative to fly nuclear genome copy. Bar charts present (A) $\beta 2$ and (B) $\alpha 2$ 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 $\beta 2$ (blue) and $\alpha 2$ (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 $\alpha 2$ 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 $\alpha 2$ in F₁ 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 $\alpha 2$. In intra-semispecies crosses, all tested F₁ flies in the FG295 ($\beta 2$) x O11 ($\beta 2$ and $\alpha 2$) 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). F₁ female hybrids from FG295 x O11 cross had significantly lower $\alpha 2$ levels than O11 females (parental line) ($p < 0.001$, Tukey HSD test), whereas the F₁ male hybrids from the same cross as well as both F₁ 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 F₁ flies of A28 ($\beta 1$) x O11 (40/40, Fig 4C) and 45% flies of FG18 ($\beta 1$) x O11 (18/40, Fig 4D) were heteroplasmic. In both combinations, the $\alpha 2$ mitotype levels of F₁ flies of both sexes were significantly lower than in the paternal O11 line ($p < 0.001$ for both crosses, Tukey HSD test).

Combining data from all four crosses (**Fig 4 A-D**), paternal transmission of $\alpha 2$ was found in 76% of F_1 hybrids (121/160). This finding clearly demonstrates that the $\alpha 2$ mitotype of *D. paulistorum* are not only sporadically leaking but are highly transmissible via males. To independently confirm these results, we tested only the heteroplasmic F_1 hybrids, not the homoplasmic ones, for their $\alpha 2$ levels using a second mitochondrial locus (*CO2*) and observed patterns similar to *CO1* (**Fig S6**).

Additionally, we investigated whether inheritance of the paternal mitotype is $\alpha 2$ -specific. For this, we used mitotype-specific RFLP-PCR or Sanger sequencing on F_1 hybrid flies generated by eight different cross combinations aimed at tracing paternal transmission of either $\alpha 1$ (number of F_1 flies assayed = 59), $\beta 1$ (31), or the $\beta 2$ (76) mitochondria. Paternal transmission was not detected in any of the 166 tested F_1 flies (**Table S7**). Hence, a high level of paternal transmission is unique for $\alpha 2$ mitotype.

We verified our findings using highly sensitive discriminant qPCR on A28 x O11 inter-semispecies crosses. We assessed quantities of the $\beta 1$, $\beta 2$ and $\alpha 2$ mitotypes in O11 males ($\alpha 2$ and $\beta 2$), A28 females ($\beta 1$) and F_1 hybrid offspring. As expected, the pure parental lines possessed only the expected mitotypes (**Fig 4E, F**). Despite robust paternal inheritance of $\alpha 2$ in all tested F_1 hybrids (16/16), the hybrid offspring do not show any significant differences in titers for any mitotype when compared to their homoplasmic mothers, which is likely due to low $\alpha 2$ titers in the hybrids and relatively few biological replicates. However, significantly different titers for all mitotypes were seen when F_1 hybrids are compared to their heteroplasmic fathers (**Fig 4E, F, G**), i.e., higher $\beta 1$, lower $\beta 2$ and lower $\alpha 2$ titers are seen in F_1 hybrids compared to the paternal line. In contrast to the less sensitive RFLP-PCR method, our discriminant qPCR method allowed us to detect very low levels (~0.1-0.3%) of paternal $\beta 2$ leakage in 11 out of 16 hybrid offspring (**Supplemental data file**).

As expected for an inter-semispecies cross [44, 45], all F_1 hybrids were sterile and did not produce any live larvae when crossed either to siblings or backcrossed to parental lines (**Table S8**). This prevented exploration of the mitotype dynamics in further generations from this combination of parental lines. We were also unable to explore the dynamics of the reverse cross (female O11 with male A28), and thereby test if the $\beta 1$ mitotype could be paternally transmitted, as none of the crosses we set up (0/200 attempts) produced larvae.

Fig. 4. Levels of the paternal α_2 mitotype in intra- and inter-semispecies hybrids of *D. paulistorum*. Paternal inheritance of the α_2 mitotype in different *D. paulistorum* hybrids determined by RFLP-PCR (A-D). Scatter plots present relative α_2 levels in single fly F₁ 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 F₁ hybrids (F₁ circle, F₁ 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 α_2 levels was estimated by ANOVA ($p < 0.05^*$, $p < 0.001^{***}$). Paternal α_2 -inheritance determined by discriminant qPCR (E-G). Number of copies of β_1 (E), β_2 (F) and α_2 (G) mitochondrial genomes in 3-weeks old O11, A28 parents and F₁ 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 $*** = p < 0.001$, $** = p < 0.01$, $* = p < 0.05$.

High invasiveness and stable persistence of the $\alpha 2$ mitotype over multiple generations

After identifying that $\alpha 2$ was the only mitotype paternally transmitted in high frequency to F_1 flies, we investigated persistence of heteroplasmy over further generations. We used discriminant qPCR to follow the inheritance patterns of the $\alpha 2$ mitotype during four generations of fertile intra-semispecies FG295 (homoplasmic $\beta 2$) and O11 (heteroplasmic $\alpha 2$ and $\beta 2$) hybrid crosses. We followed four single-fly G_0 crosses in each direction (FG295 x O11 and O11 x FG295), and mated a randomly picked pair of virgin siblings to each other in each consecutive generation (F_1 - F_4) to explore single-fly inheritance patterns of the $\alpha 2$ mitotype (**Fig 5**). The results showed both maternal (**Fig 5A**) and paternal (**Fig 5B**) transmission of the $\alpha 2$ mitotype, as well as high $\alpha 2$ invasiveness (ability to invade a homoplasmic line) and persistence (ability to be maintained over generations) in F_4 flies (13/16).

In both crossing directions, all F_1 hybrids (16/16) were heteroplasmic, hence demonstrating not only paternal but biparental transmission of $\alpha 2$. However, in the following generations (F_2 - F_4), transmission success was more varied. All tested offspring from three of the four crosses starting with a heteroplasmic mother carried both $\alpha 2$ and $\beta 2$, and one cross lost the heteroplasmy at F_3 (**Fig 5A**), whereas the crosses starting with a heteroplasmic father showed more stochasticity in transmission, with the $\alpha 2$ mitotype not necessarily transferred to both tested offspring in these later generations (**Fig 5B**). However, more individuals need to be tested in order to confirm this pattern.

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

Given that we only observe significant variability in the $\alpha 2$ titer between sexes as well as between parents and their offspring (no significant difference was observed for any test on $\beta 2$), the higher titer of mt-DNA in heteroplasmic flies overall suggest that the presence of $\alpha 2$ increases the number of total mitochondrial DNA copies per cell and is added on top of the primary mitotype, i.e., $\beta 2$. The reduction in total mt-DNA titer between heteroplasmic fathers of the O11 line and their heteroplasmic offspring of both sexes cannot be explained by the significantly different $\alpha 2$ titers observed between male and female flies. Instead, this result indicates an interaction between mitotype titer and host nuclear background, which is more pronounced in males.

Fig 5. Presence and titer of $\alpha 2$ and $\beta 2$ in (A) O11 x FG295 and (B) FG295 x O11 hybrids. Relative copies of $\alpha 2$ to number of fly nuclear copies is shown in green. Relative copies of $\beta 2$ to number of fly nuclear copies is shown in blue. Values were rounded up if the decimal was over 0.5 and rounded down if under 0.5. The figure depicts four biological repeats of each direction of the hybrid cross, each representing a single-pair mating of a male and female fly for each line. A mated pair derived from randomly picked virgin flies of each generation was tested for four generations after the initial mating and copy numbers of $\alpha 2$ and $\beta 2$ mitotypes were quantified through discriminant qPCR. * means that $\alpha 2$ was below the cutoff threshold on qPCR and the fly is considered homoplasmic.

To further investigate whether fly nuclear background affects the persistence of the $\alpha 2$ mitotype we backcrossed virgin F₁ hybrids from O11 x FG295 (by convention, the maternal line is designated first) to the homoplasmic FG295 line (**Fig 6A**), thereby increasing the nuclear background that coevolved with the homoplasmic $\beta 2$ mitotype in each generation. As before, all F₁ flies acquired the $\alpha 2$ mitotype when transmitted maternally (**Fig 6A**, left column), but their respective levels were lower than in the G₀ parent flies ($p < 0.001$, Tukey HSD test). Backcrossing these F₁ flies to the homoplasmic FG295 line (BC1, maternal introgression) showed a decrease in $\alpha 2$ prevalence in the offspring (**Table S9**). As in F₁ flies, the levels of

$\alpha 2$ in offspring from the hybrid parents were significantly lower than in G_0 flies ($p < 0.001$, Tukey HSD test). A second backcross of the BC1 offspring to the homoplasmic FG295 line revealed that $\alpha 2$ prevalence decreased in the BC2 offspring from the BC1 hybrid mothers, whereas their paternal counterparts kept a similar prevalence as that observed in BC1. Overall $\alpha 2$ levels in BC2 flies were significantly lower in both sexes in all backcrossing directions ($p < 0.001$, Tukey HSD test) compared to their parental levels (G_0 females - Mean=44.5, SD=4.5; G_0 males - Mean=52.4, SD=4.8).

Again, all F_1 flies also inherited the $\alpha 2$ mitotype when initially transmitted paternally (**Fig 6A**, right column). Compared to G_0 flies of the same sex, the levels of $\alpha 2$ were significantly lower in F_1 females but not significantly different in F_1 males ($p < 0.01$ and $p = 0.56$, respectively, Tukey HSD test). Backcrossing of these F_1 flies to the FG295 line (BC1, paternal introgression) showed a decrease in $\alpha 2$ prevalence in the offspring when transmitted both maternally and paternally (**Table S9**). Compared to the F_1 flies of the same sex, the levels of $\alpha 2$ in all BC offspring from both hybrid parents decreased to approximately 20% in females and 30% in males ($p < 0.001$, Tukey HSD test).

A second backcross of the BC1 offspring to the FG295 line revealed a pattern similar to the one of the reciprocal O11 x FG295 cross. The prevalence of $\alpha 2$ was low in the backcrossed offspring derived from hybrid mothers whereas the flies from hybrid fathers had a prevalence similar to that observed in BC1. The only exception was the offspring obtained from mothers derived from the F_1 hybrid fathers (40% for females and 40% for males). By increasing the FG295 nuclear background, all BC2 offspring showed significantly lower $\alpha 2$ levels ($p < 0.001$, Tukey HSD test; see **Table S9**) compared to the levels in their heteroplasmic O11 parents of the same sex. In summary, we observed that backcrossing heteroplasmic F_1 hybrid flies from both crossing directions to a homoplasmic $\beta 2$ line not only lowers relative $\alpha 2$ levels but also results in a decrease in heteroplasmy prevalence in further generations. These data indicate that the $\alpha 2$ mitotype can successfully invade previously homoplasmic lines, although its transmission success is affected by the host nuclear background.

Additionally, we investigated the capacity of $\alpha 2$ to invade other species by performing crosses between females of *D. equinoxialis* and *D. willistoni* and males of *D. paulistorum* O11 line (**Fig 6B, 6C**). Since hybrid flies from these interspecies crosses are fully sterile, we could not assess $\alpha 2$ prevalence, persistence and levels beyond F_1 flies. Both cross combinations showed 100% prevalence of the $\alpha 2$ mitotype in F_1 offspring, although due to the high embryonic lethality of hybrids we were only able to analyze eight F_1 flies in the *D. equinoxialis* x O11 cross (**Fig 6B**) and two F_1 flies in the *D. willistoni* x O11 cross (**Fig 6C**). In agreement with earlier experiments [58], no F_1 males were obtained in either interspecies crosses, thus, only females were analyzed for mitotype levels. Similar to results derived from *D. paulistorum* inter-semispecies crosses (**Fig 4**) F_1 females of *D. equinoxialis* x O11 cross (**Fig 6B**) had

significantly lower $\alpha 2$ levels in comparison to G_0 flies (11%, $p < 0.001$, Tukey HSD test). $\alpha 2$ levels in the two F_1 females of *D. willistoni* x O11 cross were also decreased, but due to the extremely low number of viable F_1 hybrid flies (**Fig 6C**) we were not able to analyze significance of this result.

Figure 6. Prevalence and levels of the $\alpha 2$ mitotype introgressed in different genetic backgrounds. (A) Dynamics of $\alpha 2$ prevalence (blue) and levels (green) determined by α -diagnostic *CO1/HindIII* RFLP in females and males by backcrossing F_1 flies derived from O11 x FG295 (left: maternal introgression; white x black) and FG295 x O11 (right: paternal introgression; black x white) parents to FG295 flies. Arrows in this figure represent a backcross to homoplasmic ($\beta 2$) FG295 flies. (B, C) Dynamics of $\alpha 2$ prevalence (blue) and levels (green) determined by the $\alpha 2$ -diagnostic *CO2/MspI* RFLP in F_1 flies from interspecies crosses of (B) *D. equinoxialis* (orange) and (C) *D. willistoni* (red) females to heteroplasmic *D. paulistorum* O11 males (white). The skull symbolizes complete absence of F_1 males in both crosses. Lines inside the green bars show standard deviation. Values below sex symbols indicate the number of single flies used in RFLP-PCR analysis and listed in Table S9.

$\alpha 2$ dynamics during embryogenesis in heteroplasmic lines and F_1 hybrids uncover their biparental transmission mode

As demonstrated by our crossing experiments, the biparentally inherited $\alpha 2$ mitotype of *D. paulistorum* is highly invasive and maintained at consistently high levels in the adult fly. Since the number of mitochondrial DNA copies per sperm cell is much smaller than in a mature egg [59], and paternally leaked mitochondria and their genome are usually both passively diluted and actively destroyed [15], we asked how paternally inherited $\alpha 2$ mitotype avoids depletion. By using RFLP-PCR, we analyzed $\alpha 2$ levels in unfertilized eggs and embryos from the O11, FG572 and MS lines (**Fig 7A-C**) as well as F_1 embryos from reciprocal crosses between O11 and FG295 flies (**Fig 7D, E**). The $\alpha 2$ levels at different stages of embryogenesis were

compared to those in unfertilized eggs (E0) and parental flies (G_{0f} - females, G_{0m} - males). Strikingly, none of the five lines analyzed showed traces of $\alpha 2$ in unfertilized eggs. The finding is not surprising for the F₁ flies from FG295 x O11 cross (**Fig 7E**), where the eggs come from homoplasmic FG295 mothers with only the $\beta 2$ mitotype, but is unexpected for mothers of the four heteroplasmic lines (**Fig 7A-D**) that carry high $\alpha 2$ levels (**Table 1**). We infer that $\alpha 2$ levels are very low in unfertilized eggs of heteroplasmic mothers and below the detection level of our RFLP-PCR method. Immediately after fertilization (E1; 0-30 minutes after egg laying - AEL), however, $\alpha 2$ reached relatively high levels (similar or close to parental) in F₁ embryos of all of the lines except for the ones obtained from the maternal O11 x FG295 cross (**Fig 7D**). For O11 and FG572 E1 embryos (**Fig 7A, B**) median $\alpha 2$ levels were not significantly different from the levels of either parent ($p > 0.05$; here and further on ANOVA with post-hoc Tukey HSD test), except for FG572 males ($p < 0.01$; Supplemental data of **Fig 7**). Median $\alpha 2$ levels of E1 embryos from MS, FG295 x O11 and O11 x FG295 (**Fig 7C-E**) were significantly lower than in parents ($p < 0.05$). In mid-embryogenesis (E2; 3-4 hours AEL) $\alpha 2$ levels of O11 were only significantly decreased compared to E1 ($p < 0.001$; **Fig 7A**), whereas in the four other lines (**Fig 7B-E**) $\alpha 2$ levels continued to increase (statistically non-significant in FG572, MS embryos and F₁ of O11 x FG295, $p > 0.05$; but significant in F₁ of FG295 x O11 crosses, $p < 0.001$). In late embryogenesis (E3; 18-20 hours AEL) all lines showed a significant increase in $\alpha 2$ levels compared to E2 ($p < 0.001$) except for FG572 and F₁ embryos from the maternal O11 x FG295 cross shown in **Fig 7B** and **D** ($p > 0.05$). Only in F₁ embryos from the maternal O11 x FG295 cross $\alpha 2$ levels at E3 were significantly lower in comparison to parents (9.7%; $p < 0.001$). In all other cases the level of $\alpha 2$ in E3 embryos reached similar or even significantly higher levels compared to their respective heteroplasmic parents ($p < 0.001$ for G_{0f} in O11 and MS as well 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 O11 and FG572).

Additionally, we compared $\alpha 2$ levels of embryos to the levels in emerged flies (F_{1f} and F_{1m}, **Fig 7**). There were no significant differences in $\alpha 2$ levels between G₀ and F₁ flies of the same sex except for F₁ flies from the intra-semispecies crosses O11 x FG295 and FG295 x O11 ($p < 0.01$). In comparison between F₁ flies and embryos of heteroplasmic lines we found similar statistically significant differences as between embryos and G₀ flies (see **Supplemental data of Fig 7**). F₁ flies from both O11 x FG295 and FG295 x O11 crosses (**Fig 7D, E**) had significantly higher $\alpha 2$ levels than at any embryonic stages ($p < 0.05$), except in late embryos (E3) from the paternal FG295 x O11 cross with no significant difference to F₁ males ($p = 0.99$), but significantly higher $\alpha 2$ levels than F₁ females ($p = 0.002$). Strikingly, a comparison of the maternal (**Fig 7D**) and paternal (**Fig 7E**) $\alpha 2$ dynamics between embryonic stages (E1-E3) and F₁ adult flies (**Fig 7D, E**) uncovered significant differences in $\alpha 2$ levels depending on the mode of transmission ($p < 0.001$ for all embryonic stages and F₁ flies). Hence during *D.*

paulistorum embryogenesis, the paternally introgressed $\alpha 2$ mitotype is replicating faster and possibly earlier than $\alpha 2$ that were transmitted by the female.

Fig 7. Relative $\alpha 2$ levels during *D. paulistorum* embryogenesis determined by *CO1/HindIII* (A, B, D, E) and *CO2/Mspl* (C) RFLP-PCR. Scatter dot plots combined with box plots show levels of $\alpha 2$ (%) in the heteroplasmic lines O11 (A), FG572 (B), and MS (C) and $\alpha 2$ levels in F₁ hybrids from maternal (D) and paternal (E) introgressions obtained from reciprocal crosses between parental heteroplasmic O11 and homoplasmic FG295 flies. Dynamics of $\alpha 2$ 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 F₁ adult females (F_{1f}) and males (F_{1m}). Each dot represents the $\alpha 2$ 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 $\alpha 2$ 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 $\alpha 2$ already reaches high levels in mature embryos (E₃), equivalent to levels of the adult fly. As the strict maternally-inherited primary mitochondria of

D. melanogaster and other animals are quiescent and do not replicate during early embryogenesis [13, 60], this early replication behavior is unique for $\alpha 2$ and we hence name it “Replication precox” (see Discussion).

Genomic features unique to the $\alpha 2$ mitotype

The unusual properties of the $\alpha 2$ mitotype prompted us to search for unique genomic features that distinguish this secondary mitotype from the other *D. paulistorum* mitotypes (Table 2, 3). However, only three non-synonymous substitutions were found to be exclusive to the $\alpha 2$ mitotype (Table 2).

Table 2. Mutations unique to the $\alpha 2$ clade among protein-coding genes of the sampled *D. paulistorum* lines. Three non-synonymous substitutions are unique to $\alpha 2$. Among these, only the H -> N substitution in COIII results in a change in amino acid property. Amino acid and nucleotide changes are presented from other mitotypes to $\alpha 2$. Nr = number.

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		

One of the substitutions is found in *ND3* and causes a replacement of Valine with Methionine (aa position 18 in *D. paulistorum*, 21 in humans) and another substitution in *ND4* causes a replacement of Valine with Isoleucine (aa position 379 in *D. paulistorum*, 391 in humans). Both of these replacements are between chemically equivalent amino acids and likely have little functional impact on the proteins. However, the substitution in *CO3* causes the replacement of a positively charged Histidine by a neutral Asparagine in $\alpha 2$ (aa position 7 in *D. paulistorum*, 6 in humans) and is predicted to have medium functional impact according to the MutationAssessor server (<http://mutationassessor.org/r3/>). An inspection of multi-species alignments in Uniprot and Pfam revealed that Asparagine is rarely found at this position of the *CO3* protein in other animals. The only organisms in the database with Asparagine in the same position are a few species of mites, representatives of some invertebrate marine phyla, and a wide variety of mollusks. Interestingly, among the mollusk sequences, many are from the F mitotype of heteroplasmic unionid mussels with doubly uniparental inheritance (DUI).

Additionally, the selfish $\alpha 2$ mitotype differs from the primary mitotypes of *D. paulistorum* by four A to G transitions in the large rRNA gene and a T to A transversion in tRNA-Ala and a 3 bp insertion in tRNA-Leu2 (**Table 3**), both of which are located in the T-loop.

Table 3. Mutations unique to the $\alpha 2$ subclade among non-protein-coding genes in the sampled *D. paulistorum* lines. Substitutions are found in the large rRNA and tRNA-Ala while 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 ($\alpha 1$) -> ATATT

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

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

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

Nucleotide composition creates conflicting phylogenetic signals between the α , β and the willistoni mitochondrial clades

We used several methods to search for signs of recombination between the clades. First, we created single gene trees for the 13 mitochondrial protein coding genes and tested for conflicts between the resulting topologies. For eleven of the 13 trees a definite topology could not be determined as the bootstrap support was less than 80% on the node defining the relationships of the α , β and willistoni clades. The remaining two trees had different topologies, as the *ND2* gene featured the α clade as sister either to the willistoni clade (BS = 81%) and the *ND5* gene placed α as sister to the β clade (*ND5* gene, BS = 90%). Second, we ran tests for intragenic recombination on all protein coding genes. Only two genes, *CO2* and *ND3*, were detected with marginally significant signals ($p < 0.05$) using Phipack [65]. Finally, we estimated the recombination over mutation rate along the whole mitochondrial genome to be low (0.01). These results indicate that the recombination rate is low and that recombination signals are weak and inconsistent. Thus, even if some recombination occurred between mitochondria, it is unlikely to be the cause of the low support on the internal node of the whole-genome mitochondrial tree.

Nucleotide composition bias and codon usage can also influence phylogenies and are particularly likely to affect the more neutral third codon position. Hence, we built separate trees using either only the third or a combination of the first and second codon positions of the protein coding genes. Interestingly, we observed significantly increased support for α as sister to the willistoni clade when the third codon position was removed (**Fig 2B**). In contrast, the tree based only on the third codon position showed high (Bayesian posterior probability 0.99) or fairly good (ML bootstrap 88%) support for the α and β clades being sisters (**Fig S7**). The conflicting phylogenetic signal between the two first codon positions and the third codon position suggests that nucleotide composition bias and perhaps codon usage are indeed the cause of low support on the node defining the relationship between the α , β and willistoni clades in the whole genome mitochondrial tree.

Since the third codon position is more likely to be affected by such biases, we conclude that the topology supported by the first and second codon positions of protein coding genes (**Fig 2B**) is most likely the accurate representation of the evolutionary relationships between mitochondrial clades. Thus, *D. paulistorum* mitochondria are polyphyletic, and α is sister to the willistoni rather than to the β clade. The resulting phylogeny (**Fig 2B**) implies an α -like mitochondrial ancestor in *D. paulistorum*, and suggests that β was most likely acquired from a so far unidentified lineage within the *willistoni* subgroup.

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

A comparison between nuclear and mitochondrial phylogenies shows that neither mitotype relationship nor presence of heteroplasmy follow the topology of the nuclear phylogeny. Mitotypes are shared by relatively distantly related lines belonging to different semispecies, as in the case of α 1 in C2 (CA) and MS (AB) and β 2 in O11 (OR), L1 (IN) and SM (TR), whereas the primary β 1 mitotype is exclusively found in AM semispecies (**Fig 2B**). Heteroplasmic lines are also spread across the nuclear phylogeny, and the mitotype that accompanies α 2 in heteroplasmy varies according to what semispecies the line belongs. This pattern suggests that the mitochondria of *D. paulistorum* were not only vertically transmitted but have also been introgressed and lost multiple times during evolution. Further introgression events are indicated by the positioning of *D. paulistorum* and *D. equinoxialis* in relation to other species of the willistoni subgroup on either tree. In the nuclear phylogeny, *D. paulistorum* and *D. equinoxialis* are the most derived species in the willistoni group, but the mitochondrial phylogeny places the α and β clades in a basal position with regards to the willistoni clade (**Fig 2A, B**).

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

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

suggest that the last common ancestor of *D. paulistorum* and *D. equinoxialis* carried a β -like mitochondrion.

Could $\alpha 2$ be a Mega-NUMT and not a secondary mitotype?

Besides the ancient NUMTs found in the willistoni subgroup species (**Fig 8**) it seems likely that the $\alpha 2$ mitotype could be a so-called Mega-NUMT, i.e., a multicopy nuclear locus stably integrated in the *D. paulistorum* genome of heteroplasmic flies. This concept was recently coined by [66] in response to a publication by [67] who reported on the spectacular discovery of biparental mitochondrial transmission in some human families. Since then, additional publications have emerged pointing towards the true existence of Mega-NUMTs in humans [68, 69].

In order to experimentally test this alternative hypothesis for $\alpha 2$ in *D. paulistorum*, we performed DNA-Fluorescent in Situ Hybridizations (FISH) with two differentially labelled probes (**Fig 9**); one targeting the mitochondrial genomes covering five loci (red), and the other targeting the rRNA locus (green) of *D. paulistorum*. Hybridization and washings were performed under relatively stringent conditions (see Material and Method) in order to avoid cross-hybridization of the mt-probes with the fixed ancient NUMT shown in **Figure 8**. As numbers of $\alpha 2$ mt-genomes per O11 nuclear genome are estimated to at least 20 copies in females and more than 50 in the males (**Fig 3B**), we expect to detect either one or two chromosomal Mega-NUMT clusters, dependent on the sex, with a minimal size of at least 300 kb each. We further envision that one mt-signal should be located on the Y chromosome of males and one on an autosome in both sexes. Therefore, we used the rRNA locus as a control, as it is expected to be of similar or bigger-size (>300 kb) than the $\alpha 2$ Mega-NUMT.

In our DNA-FISH experiments on whole-mount early embryos (**Fig 9A-F** and **Supplemental data file**), we detected exclusively in the heteroplasmic strains (O11 and MS) two bright and compact mt-signals in all interphase cells (**Fig 9A-C**), which are closely associated with the nuclear periphery, and in mitotic cells 1-2 less compacted mt-aggregates (red) associating with the distal chromosomal ends (**Fig 9D, E**). As expected, no mt-signals, besides cytoplasmic ones, were traceable in any of the tested homoplasmic strains (FG295), which are devoid of $\alpha 2$ (**Fig 9C, F**).

As fixed cells of whole-mount specimens maintain cytoplasmic components like mitochondria, we performed DNA-FISH on mitotically-active neuroblasts of squashed 3rd instar larval brains after hypotonic treatment, which washes out the cytoplasm and leaves behind the nucleus. Although the rRNA control probe worked on all metaphase chromosomes (**Fig 9 G-I**), mt-signals were untraceable even under enhanced laser intensity in specimens of the two heteroplasmic strains.

Furthermore, DNA-FISH with the same probe set (mt-probe and rRNA control) was performed on squashed polytene chromosomes from 3rd larval salivary glands of homo- and heteroplasmic lines under similar stringent conditions (**Fig 9 J-L**). Although the rRNA clusters on the X and the Y chromosomes were brightly stained depending on the sex of the larvae, no chromosomal signals of the mt-probe were detected in any of the 16 polytene interphases derived from the two heteroplasmic strains (**Fig 9 J, K** and **Supplemental data file**). Even though the α -heterochromatin is under-replicated in polytene chromosomes of *Drosophila*, a potential Mega-NUMT (>300 kb) should not escape detection in telomers or the chromocenter by Confocal Laser Microscopy at maximal laser intensity and overexposure (**Fig S9** and **Supplemental data file**). Thus, the clear absence of mt-FISH signals in any cytoplasmic-free samples (**Fig 9G-L**) allowed us to reject the Mega-NUMT hypothesis, *i.e.*, the physical integration of $\alpha 2$ into the host chromosome, and, as an alternative hypothesis, its integration in B-chromosomes, which usually are defined by the presence of their own telomeres and a centromere during mitosis [70].

We interpret our findings obtained by DNA-FISH as follows: The $\alpha 2$ mitotype is neither a classic Mega-NUMT, nor integrated in an accessory B-chromosome. Instead, it resembles a facultative, invasive and selfish genetic element, which is cytoplasmic and still maintains a complete and potentially functional genome evolving under purifying selection. This mitotype associates closely with the host nucleus in interphase and mitotic chromosomes. During early development $\alpha 2$ probably hooks up with the replication machinery for increasing their own copy numbers. Just after fertilization, embryonic nuclei duplicate rapidly, primary mitotypes are transcriptionally and replicatively quiescent, but, as shown in **Fig. 7**, levels of $\alpha 2$ increase quickly from relatively very low to significantly high copy numbers. We speculate of a spatial and temporal hitchhiking of $\alpha 2$ with early nuclear replication just after fertilization and also later in other dividing cells, like neuroblasts in larval brains. This “Replication precox” phenotype is assisted by a yet undetermined mechanism that safeguards the selective replicational advantage of this selfish mitotype by counteracting out-dilution by the primary mitotype.

Fig 9. Fluorescent in situ hybridization on *D. paulistorum* 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

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

DISCUSSION

We report on a unique system of stable mitochondrial heteroplasmy in the neotropical *D. paulistorum* species complex that might exist due to biparental transmission. Using sequencing, RFLP-PCR and bioinformatic analyses we discovered high-level heteroplasmy in 41% of the tested fly lines including representatives from three of the six *Drosophila paulistorum* semispecies (Andean-Brazilian, Amazonian and Orinocan). However, given the limited number of lines currently available from the Centro-American, Interior and Transitional semispecies, it is possible that heteroplasmy may be even more broadly distributed among the species. We also found that heteroplasmy is fixed with balanced levels of each mitotype in individuals of three long-term laboratory lines (O11, MS and Yellow), but not in the more recently collected lines, in which α 2 levels are also tightly balanced but not fixed (**Table 1**). These results suggest that a line has to be genetically homogeneous for heteroplasmy to be fixed and that levels are controlled by the nuclear genetic background. In the course of our current study we observed that the Orinocan FG103 line collected in 2014 lost the secondary mitotype within the first two years in the lab, suggesting that long-term maintenance of fly lines under lab conditions might also contribute to an underestimation of prevalence. Further sampling will be necessary to determine the prevalence of heteroplasmy in natural populations.

Function of the α 2 mitotype in *D. paulistorum*

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

suggesting that $\alpha 2$ might not provide full mitochondrial functionality. Mitochondria are usually enriched in energy-demanding tissues [71, 72], and by studying the heteroplasmic O11 line we found that only the $\beta 2$ mitotype follows this pattern. The $\alpha 2$ mitotype, on the other hand, has unchanging titers across fly tissues. Although not conclusive, this suggests that $\beta 2$ is the mitotype that provides energy via oxidative phosphorylation, while $\alpha 2$ does not respond to the energy demands of the host. Despite this, the genes encoded on the $\alpha 2$ genome appear to be evolving under purifying selection, as indicated by low dN/dS ratios and the absence of pseudogenes. In addition, very few genomic differences are seen between $\alpha 2$ and the closely related $\alpha 1$ mitotype (0,67% divergence), which since it is the sole mitotype found in two Centro-American *D. paulistorum* lines (C1 and Yellow) must be fully functional. The low sequence divergence between $\alpha 1$ and $\alpha 2$ strongly suggests that the transition from a primary α -like ancestor to a selfish mitotype took place relatively recently, but is found consistently in different and reproductively-isolated semispecies. Considering these facts, we surmise that $\alpha 2$ is likely functional based on its genome but that its steady titer throughout tissues and absence in homoplasmy suggest that it does not provide sufficient energy to sustain a fly, at least under the standard laboratory conditions of our experiments.

Biparental inheritance of the $\alpha 2$ mitotype in *D. paulistorum*

The question remains as to how and why $\alpha 2$ is maintained in *D. paulistorum* flies. It may serve a yet undefined essential or condition-specific function for the fly, or alternatively it may be cheating the system and perpetuate selfishly while providing no or few benefits to the host. One possible way of accomplishing the latter is for the mitotype to increase its own vertical transmission, for example through biparental inheritance.

Using RFLP-PCR and discriminant qPCR, we detected high level paternal transmission of the $\alpha 2$ mitotype in offspring from both intra- and inter-semispecies crosses of *D. paulistorum* (Fig 4, 5 and 6A, B), and even in offspring from crosses between *D. paulistorum* and two closely related species of the willistoni subgroup (Fig 6C). None of the other mitotypes found in *D. paulistorum* were paternally transmitted in high levels. Generally, active pre- and post-fertilization recognition and elimination mechanisms prevent paternal transmission of mitochondria and their DNA in animals (recently reviewed in [15]). In *D. melanogaster*, the mechanism is not yet fully understood, but two genes, *EndoG* and *tamas*, are known to be involved in eliminating mitochondrial DNA from developing spermatids [13, 14, 73]. After fertilization, paternally leaked mitochondria are eliminated by endolytic and autophagic degradation systems [74] while the rare paternal ones that escape are passively outcompeted by the larger number of maternally transmitted mitochondria and their genomes already present in the egg [75]. In order for a mitotype to be transmitted paternally and – even more importantly – to stably persist in adult flies in large numbers, at least one of these

mechanisms need to be bypassed. We did not specifically determine if the $\alpha 2$ mitotype evades the active mechanisms that eliminate paternal mitochondrial DNA (reviewed in [15]), or associates with the sperm nucleus, but the observation that none of the other mitotypes of *D. paulistorum* were paternally transmitted in large numbers (**Table S7**) implies that there is no general loss of such active defense functions in *D. paulistorum*. The spurious low level (~0.1-0.3%) paternal leakage of $\beta 2$ in some hybrid crosses could be a result of a partial breakdown of such mechanisms due to the mixed nuclear hybrid background ([23] and references therein). However, low levels of paternally transmitted $\beta 2$ is also in line with previous reports of paternal leakage within *Drosophila* species [26, 29, 30], indicating that similar levels of leakage can occur even without interspecies hybridization and breakdown of defenses. Thus, the mechanism for paternal transmission is specific to the invasive $\alpha 2$ mitotype.

In *D. melanogaster* and *D. simulans*, the spermatozoon is estimated to carry approximately 20 copies of mitochondrial DNA [59], which is five to six orders of magnitude less than the number of copies found in an egg [59]. We expected that similarly small numbers of the $\alpha 2$ copies are present in *D. paulistorum* sperm and at the onset of embryo development when $\alpha 2$ is transmitted by the male. However, using relative RFLP-PCR, we found that rapid and premature replication in early embryos resulted in $\alpha 2$ levels that were similar to those of adult flies just a few hours after fecundation (**Fig 7**). Previous studies of *D. melanogaster* showed that maternally transmitted mitochondria are quiescent and do not replicate for up to 10 hours after fertilization [13, 60], which seems to be an evolutionary conserved state (reviewed in [76]). Hence, the premature and unique temporal replication of the $\alpha 2$ mitotype, that we name “Replication precox”, might allow this mitotype to escape passive dilution by the still dormant primary mitotype and thereby guarantee its presence in high levels in adult flies. We also found that the $\alpha 2$ mitotype level increased earlier and faster in hybrid embryos when it was transmitted paternally rather than maternally (**Fig 7D, E**). This could be due to more $\alpha 2$ mitotype copies being deposited by the sperm than by the egg, or to a higher rate or earlier onset of $\alpha 2$ replication when it is paternally inherited. Since we found that O11 males have significantly higher levels of $\alpha 2$ than females, also more copies in testes than in ovaries, and DNA-FISH showed a close spatial association with nuclear replication, our data suggest that the former hypothesis is possibly more likely. One tempting mechanistic hypothesis for the Replication precox phenotype is that, because of their close vicinity to nuclei, $\alpha 2$ replication might profit by parasitizing on their replication machinery. The elucidation of the mechanism for diagnostic recognition and/or selective replication of $\alpha 2$ is currently underway in our laboratory.

Variability in mitotype titer and levels between sexes

We found that $\alpha 2$ titers were consistently and significantly higher in males than in females. Since $\alpha 2$ does not seem to be a fully functional mitotype, the higher $\alpha 2$ titer in males cannot be interpreted purely based on different energy needs between sexes. So why would males have a higher titer of $\alpha 2$ than females? One possibility is that high titer in males has evolved specifically to facilitate paternal transmission of the invasive $\alpha 2$ mitotype, since a higher titer increases the chance of transmission. Alternatively, higher male titer and the Replication precox phenotype of $\alpha 2$ could have co-evolved in order to compensate for a reduced functionality of the maternally transmitted primary mitotype in males. The mother's curse hypothesis suggests that DNA of strictly maternally inherited mitochondria co-evolve with females and might therefore accumulate mutations that are harmful to male functions [77]. If such mutations have accumulated in the maternally transmitted primary mitotype of *D. paulistorum*, the biparentally transmitted $\alpha 2$ mitotype might have co-evolved with males and express a still undetermined function that is more important to males than females. In the O11 line, higher relative titer of $\alpha 2$ compared to the primary mitotype $\beta 2$ are found in heads and gonads (**Fig 3B**), suggesting that perhaps behavior and fertility might be important further areas of investigation for $\alpha 2$ specific functions in males.

In addition, our data suggest that the total mitochondrial DNA titer is significantly affected by the presence of $\alpha 2$, as this secondary mitotype appears in addition to the primary mitochondrial DNA copies in a cell. A similar quantitative effect on mitotype copy numbers was previously observed in heteroplasmic *D. subobscura* flies carrying a selfish mitotype with a large deletion that results in a replicative advantage [78].

Our crossing data also suggests that the nuclear genetic background influences the total mitochondrial DNA titer, particularly in males. This is indicated by a significant reduction in overall mt-DNA titer between pure O11 heteroplasmic males and heteroplasmic hybrid offspring of both intra- and interspecific crosses, but not between pure O11 heteroplasmic females and hybrid offspring. This observation is in line with long-term co-evolution between the heteroplasmic cytoplasm and the nuclear genome in males. In this context it is interesting to note that two of the genomic differences, which we found in the $\alpha 2$ genomes, are located in the T-loops of tRNAs and the nuclear encoded gene *mEFTU-2* that is specifically expressed in testis of *D. melanogaster* has mutations in a region where the protein is predicted to interact with the tRNA T-loop [63, 64]. Hence, if assuming a similar male-specific expression of *mEFTU-2* in *D. paulistorum*, this observation might strengthen the idea of a co-adaptation between the $\alpha 2$ mitotype and male functions.

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

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

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

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

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

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

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

Incongruences between the nuclear and mitochondrial trees also indicate more recent introgressions between *D. paulistorum* semispecies and lines, as mitotype distribution does not follow the nuclear phylogeny. Since *D. paulistorum* is a recent radiation, i.e., a species cluster *in statu nascendi* ([44] and references therein), such pattern could be explained by hybridization in a more recent past, when the pre- and postmating incompatibilities that exist today between semispecies (reviewed in [80], [44], and this study) were still incomplete. Although both F_1 hybrid sexes from our inter-semispecies crosses were sterile, which would have prevented mitotype introgression, possibly some rare fertile F_1 female flies could be found with extended tests and/or other line combinations. As discussed in early literature [81], it is possible that sporadic hybridization occurs between less discriminating and partially compatible populations or semispecies, and those could perhaps lead to the serendipitous events of mitochondrial introgression. As shown earlier, the Transitional semispecies is particularly permissive to inter-semispecies crossings and has the potential to bridge borders [82]. Specific lines of other semispecies have also been shown to be less isolated from others, potentially allowing at least some rare cases of genetic leakage and exchange [83]. For the one Transitional line included in our study, SM, we see a strong discrepancy between the nuclear and mitochondrial trees, with the mitochondrial genome nested within the Orinocan mitochondrial clade and the nuclear tree positioning it with Centro-American. Both our phylogenetic analysis and experimental data from crosses suggest that α_2 mitotypes are particularly likely to be transferred in crosses between semispecies, and even when crossed with sister species.

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

Introgression and heteroplasmy as drivers of the evolution of selfish mitochondrial traits

Our inferences on mitochondrial evolution within *D. paulistorum* provide us with a hypothesis about how features specific to $\alpha 2$ may have evolved. We speculate that the introgression of β into an α -carrying *D. paulistorum* likely created competition between mitotypes for a chance to be transferred to the next host generations. If the ancestrally introgressed β mitotype was more efficient at maternal transmission or perhaps just a better energy producer than the ancestral α , selection could have made it the more dominant mitotype in the *D. paulistorum* species complex. Over time, this could have led to a reduction in prevalence and possibly resulted in extinction of the α mitotype from the species. In this scenario, the evolution of biparental transmission and the Replication precox phenotype may have allowed the $\alpha 2$ lineage to persist selfishly next to the primary mitotype. Such traits would perhaps guarantee the transmission of $\alpha 2$ to the next host generation independently of its ability to meet the energetic needs of the host or to compete with the maternal mitotype for presence in the egg.

Can we be sure that $\alpha 2$ is not a Mega-NUMT?

Finally, we are aware of a potential alternative hypothesis which would suggest that $\alpha 2$ is not a cytoplasmic mitotype but might stem from a so-called “Mega-NUMT”. It was recently shown that Mega-NUMTs can mimic heteroplasmy and extremely rare cases of paternal mitochondrial DNA leakage in humans and other vertebrates (recently reviewed in [68]). In this case, mitochondrial DNA-concatemers possibly produced by extrachromosomal rolling circle replication were found as large-sized chromosomal insertions in human autosomes and wrongly assumed to be unusual biparental transmission of an additional mitotype [67]. However, in *D. paulistorum*, we reject this alternative hypothesis based on the following reasons: Contrary to Mega-NUMTs, which rapidly accumulate degenerative mutations and structural variation, the genome of the $\alpha 2$ mitotype resembles a fully functional mitochondrial DNA in size and content. Furthermore, the $\alpha 2$ genome is highly conserved within and between heteroplasmic *D. paulistorum* semispecies as well as between long-term lab lines and recently collected lines, which suggests that it evolves under purifying and not neutral selection. In addition, Mega-NUMTs are cryptic and if detected they are transmitted at very low frequencies and thereby do not follow strict Mendelian inheritance. These findings are in sharp contrast to $\alpha 2$ of *D. paulistorum* flies that shows almost perfect biparental transmission within a line, and as shown by our introgression experiments, this selfish mitotype has the capacity for invasive spreading. A recent publication by [69] demonstrated the physical association of a Mega-NUMT with around 45 to 56 mt-genome copies in chromosome 14 of healthy humans. In our case however, we failed to detect clear mt-signals of any large sized nuclear-integrations of $\alpha 2$ by means of DNA-FISH on fixed whole-mounted early embryos, cytoplasm-free neuroblast

metaphases and polytene chromosomes from heteroplasmic lines, even though a potential $\alpha 2$ Mega-NUMT is estimated to be bigger than 300 kb. Finally, our backcross experiments clearly demonstrated the persistence and spreading capacity of the $\alpha 2$ mitotype in a homoplasmic nuclear backgrounds, showing that $\alpha 2$ is self-sufficient and hence not associated with any Mega-NUMTs.

CONCLUSION

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

Given its intact putatively functional genome and recently evolved selfish phenotype, the $\alpha 2$ mitotype can potentially be referred to as a sexually transmitted, parasitic infection rather than a primary maternally inherited mitotype. Hence, since $\alpha 2$ appears to have a more “rebellious” lifestyle with less host control than typical primary mitotypes, but more similar to a parasitic hitchhiker on nuclear replication, we propose to name it “*Spartacus*” in reference to the famous ancient Roman slave, gladiator and then commander who escaped from imperial oppression by rebellion, for at least a short period of his lifetime.

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

MATERIALS & METHODS

Fly strains, semispecies typing, and *Wolbachia* knockdown

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

were performed in 2-3 biological and three technical replicas (in total 30 crosses in 2-3 biological replicas). Each biological replicate consisted of 3-6 females and 3-6 males, which were kept in the same vial for 3-5 days before transfer into a new vial for generating one technical replicate. Egg-to-adult development was monitored and information on all crosses including the number of F₁ flies hatched is summarized in **Table S6**.

Wolbachia knockdown

In order to test the potential effect of the *Wolbachia* endosymbiont of *D. paulistorum* [47] on mitotype persistence and levels, we generated lines with lowered *Wolbachia* infection titer. *Wolbachia* was knocked down by keeping flies on food containing 0.05% tetracycline [w/v] for four generations. Flies were repeatedly tested using PCR in order to confirm the absence of the symbiont. After four generations, knock-down flies were moved to tubes previously occupied by males of the same line, so that they could feed on feces of flies with wild-type gut microbiota. This aimed to restore any gut flora affected by the antibiotic and keep the knockdown and wild-type fly lines as similar as possible with the exception of *Wolbachia* presence [49].

Single fly Southern blots

Single fly Southern blots were performed following the protocol by [86]. In brief, DNA was extracted from single individuals, digested with the restriction enzyme *Hind*III and processed for electrophoresis on a vertical 0.8% agarose gel. After vacuum-blotting onto a positively charged nylon membrane, fragments were hybridized with P₃₂-labelled mitochondrial *CO1* probe derived from a plasmid. Detection was performed by exposing the membranes to X-ray films at -80 degrees.

Restriction Fragment Length Polymorphism-PCR (RFLP-PCR)

Total DNA was extracted from single or multiple flies using the Gentra Puregene Tissue kit (Qiagen, Germany). Flies were homogenized using a TissueLyser LT (Qiagen) at 50 Hz for 50 sec and DNA was consequently isolated following the manufacturer's protocol. DNA concentration was determined with a NanoDrop OneC Spectrophotometer (Thermo Scientific). RFLP-PCR for the mitochondrial Cytochrome c oxidase genes *CO1* and *CO2* was set up in 20 µl reactions containing 1x Promega reaction buffer, 2.5 mM MgCl₂, 0.5 µM of each primer (**Table S11**), 200 µM of each dNTP, and 0.025 U of GoTaq G2 DNA Polymerase. *CO1/CO2* reactions were performed using the following thermal profile: initial denaturation for 2 min at 95°C followed by 30 cycles consisting of 45 sec denaturation at 95°C, 45 sec annealing at the appropriate primer conditions T_m°C (**Table S11**) and 30 sec extension at 72°C. Final extension was carried out at 72°C for 10 min. PCR products were consequently used for RFLP (digest plus undigested control) using the restriction enzymes *Hind*III and *Eco*RV-HF for *CO1*,

and *MspI* for *CO2*. Reactions were performed in 20 µl volume containing 10 or 20 U of enzyme, 1x reaction buffer and 10 µl PCR product. After overnight incubation at 37°C, enzymes were inactivated by either heat (80°C for 20 min) or by adding EDTA-containing gel loading dye. 10 µl of undigested and digested samples were then run on a 1.5% agarose gel and post-stained with 10 mg/ml ethidium bromide. Imaging was performed using a Molecular Imager ChemiDoc™ XRS Imaging System and mitochondrial heteroplasmy levels were quantified using Image Lab 5.2.1 software.

Quantifying levels of heteroplasmy

Undigested samples representing the sum of both mitotypes were set as reference band with relative quantity 1. The relative levels of mitotypes were then quantified by summing up the two digested fragments plus the undigested fragment. Due to technical differences in band intensity, however, the sum of all fragments did not result in a relative quantity of 1. Hence the actual sum value was set to 100% to determine the percentage of mitotype levels. Smaller fragments with length up to approximately 100 bp were not detectable after ethidium bromide staining and we hence used the correction factor $Cf = \frac{\text{undigested}[bp]}{\text{digested}[bp]}$ to estimate their relative quantities.

DNA extraction, genome sequencing and assembly

For each line 20 ovaries were dissected from 3-day old adult females and DNA extraction was performed using the Gentra Puregene Blood and Tissue kit (Qiagen). DNA samples used for PacBio sequencing were produced using the protocol described in [87]. Briefly, *Drosophila* flies were transferred to apple juice agar plates and allowed to oviposit for 2 hours, after which the eggs were collected, washed and dechorionated in 50% bleach and manually homogenized using a plastic pestle. Following centrifugation and filtration of the homogenate, the resulting cell pellet was subjected to whole genome amplification using the Repli-g midi kit (Qiagen) after which the DNA was purified using QIAamp DNA mini kit (Qiagen) according to the manufacturer's recommendations.

Sequencing

Drosophila paulistorum lines from the classic six semispecies plus lines from more recent collections were used for genome sequencing (**Table 1 and S1**). DNA extracted from ovaries was used to construct 350 bp fragment TruSeq libraries, multiplexed and sequenced at the Uppsala SNP and Seq platform in three different runs on an Illumina 2500 HiSeq machine generating 2x125 bp reads. The Illumina reads were quality filtered and trimmed using Trimmomatic-0.30 and error corrected using SPAdes-3.5.0. Amplified DNA extracted from early embryos from three lines (O11, POA-1 and TP37) was used to create 5 kb fragment

SMRTbell libraries. Each library was run using P6-C4 chemistry in one SMRT cell on a RSII PacBio instrument. PacBio libraries were produced and sequencing was performed at the Uppsala Genome Center, Uppsala, Sweden.

Mitochondrial genome assembly

The two different mitochondrial genomes α and β of the O11 line were generated by mapping all Illumina reads to the published mitochondrial genome of *D. willistoni* (NCBI accession number BK006338.1) using BWA-MEM [88] and thereafter extracting all aligned reads and their mates using SAMtools [89]. The extracted Illumina reads were then subsampled to achieve an approximate coverage of 100 and assembled together with PacBio filtered subreads using SPAdes-3.10.1 [90] with multiple different k-mers settings. The β mitotype consistently came out as a single full length contig, so the best assembly was chosen based on the contiguity of the putative α mitotype. From this assembly the contig representing the β mitotype and the contigs with a homogenous coverage likely representing the α mitotype was taken out and used as trusted contigs together with the Illumina data to generate a better assembly of the α mitotype. Finally, the Illumina and PacBio data were aligned against the resulting contigs using BWA, MEM and SW respectively. The resulting bam-file was converted to an ace-file that was used for manual curation of the sequences in Consed [91]. The mitochondrial genomes of all other *Drosophila* lines were generated by mapping Illumina reads to the two O11 mitotypes with BWA-MEM, extracting all aligned reads and their mates using SAMtools and assembling the extracted reads with SPAdes-3.10.1 with multiple different k-mers settings. Since PacBio data were only generated from the O11, TP37 and POA-1 lines, the assemblies of all other lines were run using only Illumina reads. For all lines, this assembly procedure generated one contig that represented a complete (O11) or nearly complete circular mitochondrial genome.

To test for heteroplasmy and to correct possible errors in contigs, the Illumina reads for each line were mapped back to the contig containing a complete or nearly complete mitotype for that line using BWA-MEM and indels were realigned with the IndelRealigner from GATK [92]. The resulting bam-file was then used to call variants using Freebayes with a range of parameters and converted to ace-format for manual curation in Consed. A line was determined to be heteroplasmic if 1) at least 80 sites (ca 0,5%) had a variant and 2) the variants had a consistent alternative allele frequency over the whole or a significant part of the mitochondrial genome. These settings were chosen to avoid calling variant genomes based on sequence errors and to make sure that only complete mitochondrial genomes (not NUMTs) were used to decide heteroplasmy status. If a line was seen to be heteroplasmic using these criteria, two different alternatives were used to resolve the assembly of the second mitotype.

- 1) The Illumina reads mapping to the O11 α and O11 β mitotype were split into two separate

data sets and assembled separately, 2) in the case of MS that has two α mitotypes, variant calling was performed with lofreq [93], variants with a frequency between 8-30% were selected and used to make a new consensus with bcftools and the reads were mapped to the two variants of the mitochondrial genome. The same procedure was done in multiple rounds until no more variant sites within the correct frequency were found. The final bam-file was converted to ace-format and manually inspected in Consed.

Alignments and phylogenetic analyses

Datasets of either whole mitochondrial genomes, nuclear markers or mitochondrial genes were aligned with MAFFT v6.850b (mafft-linsi algorithm) [94] and trimmed with TrimAl v1.2rev59 [95] using the automated1 for removal of poorly aligned regions. Concatenated sets of either mitochondrial or nuclear genes were created by combining trimmed alignments of single genes. Datasets of mitochondrial protein coding genes containing either only the 3rd or a combination of 1st and 2nd codon positions were created by parsing the alignments with an in-house script.

Maximum likelihood trees were built with RAxML version 7.2.6 [96] using -f a -m GTRGAMMA -x 12345 and -# 1000. Bayesian trees were made with MrBayes version 3.2.6 [97, 98] using “nst= mixed” for estimating evolutionary models. Analyses were run for 10.000.000 generations with “nrns= 2”, “nchains= 4” and “samplefreq= 1000”. A burn-in of 25% was applied. Trees were visualized in FigTree v. 1.4.3 [99] and rooted on the species of the *saltans* group.

Identification of synonymous and non-synonymous substitutions

Synonymous and non-synonymous substitutions in pairwise comparisons between lines of the $\alpha \times \beta$ and $\alpha 1 \times \alpha 2$ clades were identified in single gene alignments using codeml (PAML package v. 4.8) [100]. Values obtained from pairwise comparisons between lines were then used to estimate an average for the clades. The same method was used for calculating the number of nonsynonymous substitutions per non-synonymous site (dN), synonymous substitutions per synonymous site (dS) and the dN/dS ratios. Synonymous and non-synonymous mutations unique to $\alpha 1$ and $\alpha 2$ were counted through visual inspection of single gene alignments using Aliview v. 2019 [101].

Assessing rarity and functional impact of $\alpha 2$ -unique mutations

We used the MutationAssessor server (www.mutationsassessor.org) to estimate the potential impact on protein function for non-synonymous substitutions exclusive to $\alpha 2$. Since the server only analyses human sequences, we used the Uniprot website (www.uniprot.org) to align *D. melanogaster* and human CO3, ND3 and ND4 amino acid sequences and identify the

positions in the human genes which correspond to the non-synonymous substitution sites unique to $\alpha 2$. With that information we generated the following input strings which were then used in the server: COX3_HUMAN H6N, NU3M_HUMAN T21M and NU4M_HUMAN I391V. We used Uniprot to assess whether the $\alpha 2$ -unique non-synonymous mutation in CO3 was found in other species. This was done by searching for “COX3” in the UniProtKB database, aligning all entries of the first results page (250 entries) and visually checking for the presence of an N residue in the position of interest. The process was repeated for the subsequent result pages until 2000 entries had been aligned and checked.

Quantitative Real-Time PCR (qPCR)

DNA was extracted by homogenizing whole flies in 200 μ l Chelex solution at 5% concentration in distilled water. A total of 10 μ l of proteinase K (10mg/ml) was added per sample and vials were incubated for 6 hours at 56°C. After digestion, vials were centrifuged at 16,200 x g for 3 minutes, and the supernatant was transferred to a new vial. DNA was stored at 4°C and freeze-thaw cycles were avoided when possible.

Samples were run in duplicate using SYBR Green reagent on a StepOnePlus Real Time PCR machine (Applied Biosystems). Technical repeats with >1 Ct difference were rerun or discarded. Each reaction contained 10 μ l Maxima SYBR Green/ROX qPCR Master Mix (2x), 1 μ l of forward primer, 1 μ l of reverse primer, 6 μ l nuclease-free water and 2 μ l of the DNA sample. Cycling conditions were 95°C for 20 seconds, followed by 40 cycles of: 95°C for 3 seconds and 60°C for 30 seconds for *rps*, and *Wolbachia* primers (**Table S11**). Discriminant mitochondrial primers cycling conditions were 95°C for 20 seconds, followed by 40 cycles of: 95°C for 3 seconds and 55°C for 30 seconds (**Table S11**). Discriminant primers used linked base technology to increase specificity (Exiqon). Melt curves and the fly control gene *rps17* were run on each plate to control for successful single-peak amplification and to control for DNA quality. All primers had >90% efficiency. For absolute quantification, each 96-well qPCR plate was analyzed using StepOne Software v2.2.2 (Applied Biosystems) and Ct values were obtained by comparing each primer sample to a single standard curve of known concentration and using identical threshold and baseline levels for each primer target across plates. Standard curves were created by amplifying positive control samples using PCR, calculating DNA concentrations using QBIT, and then serially diluting the sample 1:10 with distilled water to create a 5-sample curve comprising known concentrations decreasing from 10pmol/ ml. Samples with Ct values over 30 were classed as negative, confirmed by our negative controls. As a result, the detection limit was 50 copies per sample for *rps17*, 115 for CO3 (SNP 1), 735 for CO3 (SNP 2) and 1627 for CO3 (SNP 3). This varied for the FG295 qPCR, where samples with Ct values over 20 were classed as negative for the CO3 (SNP 2) primer due to very high amplification of some DNA segments. This was confirmed by negative flies, but led to a high threshold of detection of 6150 for CO3 (SNP 2). DNA concentration was calculated for each

sample by comparing sample Ct values back to the equation of the standard curve. DNA copy numbers were then calculated by using DNA product length to determine DNA concentration. To control for fly size and extraction efficiency, copy numbers for each sample were normalized using the fly housekeeping gene as a control, giving a final value in terms of mitochondrial genome copies to fly nuclear genome copy ratio.

Since we used three independent experimental approaches to determine levels of heteroplasmy in *D. paulistorum*, i.e., NGS, RFLP-PCR, and qPCR, we compared the robustness of these. We looked at the levels of $\alpha 2$ in O11 females (**Fig S8**), and found that all three techniques determined similar levels, hence confirming the robustness of our experimental approach.

Cross-mating and embryo collection

For mass-mating crosses (pools), virgin males and females from different *D. paulistorum* lines were obtained by collecting newly emerged flies within 2-3 h intervals to avoid any possible sibmating. Collected flies were kept separated for 5-7 days to ensure their virgin status. Crosses were set up in pools using 10-20 females and 20-30 males per each vial and in 2-3 replicates. Flies were mated for one week and then transferred to new vials or used for consequent embryo collecting. F₁ flies emerging from the first vials were collected for DNA extraction. Mass-mating A28 x O11 hybrids were generated by mating 10 five-day-old virgin A28 females to 10 O11 males. F₁ individuals were used for assays. For generating interspecies hybrids *D. paulistorum* O11 males were mated either with *D. equinoxialis* or *D. willistoni* virgin females in pools of 20 females and 60 males per cross. F₁ flies emerging from the bottles were collected for DNA extraction. Single fly crosses were generated by collecting virgin flies every 6-8 hours to avoid sibling mating. Collected flies were kept separated for 5-7 days to ensure their virgin status before use in crossings. Single pairs of male and female flies were mated for two weeks in fresh vials. Emerging F₁ flies were collected, aged to 8-10 days and frozen for subsequent DNA extraction. For FG295 x O11 and O11 x FG295 crosses, single-pair sibling mating were continued for four consequent generations.

Backcrossing

For backcrossing experiments, heteroplasmic F₁ flies obtained from reciprocal crosses between O11 and FG295 parents (G₀) were crossed in pools to the homoplasmic FG295 line. Virgin F₁ males and females were obtained from the G₀ crosses by collecting newly emerged flies within 2-3 h intervals to avoid sib mating. Collected flies were kept separated for 5-7 days to ensure their virgin status. Backcross 1 (BC₁) was set up in pools using 5-10 females and 5-10 males per each vial. The first 10 F₁ flies of each sex emerging from the vials were collected for further backcrossing with FG295 line and the remaining F₁ flies were frozen in

liquid nitrogen for future DNA extraction. Backcross 2 (BC2) was set up similarly to BC1. Between 10 and 20 flies were analyzed for each backcross.

Collection of staged embryos

D. paulistorum embryos at early (0-30 min AEL), mid (3-4 hrs AEL) and late (18-20 hrs AEL) stages of development were collected on homemade medium (agar, molasses, cornmeal and yeast medium) in plastic Petri dishes (3 x 3 cm) using collection chambers at 25°C. Prior to collecting embryos, 1-2 weeks old flies were kept in these chambers for 2-3 days to adapt to the new environment. Embryos were then manually picked with a dissection needle and transferred immediately to ice-cold lysis buffer (Qiagen). We collected 20 embryos for each sample with at least five biological replicates. Embryos were immediately processed for DNA extraction or stored at -20°C for no more than one week.

Recombination analysis

We checked for intergenic recombination between lines by analyzing single gene trees built as described in the “Alignments and phylogenetic analyses” methods section. Tests for intragenic recombination on single-gene and whole-genome alignments were done using Geneconv v.1.81 and PhiPack with default parameters. Recombination over mutation rates were calculated for single-gene and whole-genome alignments using ClonalFrame v.1.2 with default settings.

Nuclear markers

Illumina reads from each fly line were separately assembled with both Megahit v.1.1.2 [102] and SPAdes v. 3.13.0 using default parameters. We then blasted the *D. willistoni* CDSs of the selected nuclear markers (**Table S12**) to each assembly and retrieved sequences with an alignment length of over 200 bp and above 70% identity. The obtained contigs were aligned to the corresponding *D. willistoni* gene with MAFFT v6.850b. The alignments were visualized with Aliview v. 2019 and inspected for assembly issues and, if necessary, multiple exons were joined into a single sequence using Aliview’s “merge” function. For each fly line and marker gene, results obtained from both assemblies were compared and only the one with highest quality was used in further analyses.

Identification of nuclear mitochondrial DNA (NUMTs)

Illumina reads of each line were separately mapped with BWA-MEM v.0.7.10 to the genome assemblies generated with Megahit v. 1.1.2. The average per-base coverage of each assembled contig was estimated with SAMtools mpileup and contigs with average coverage lower than 2 were excluded from the analysis.

Each genomic assembly was blasted to all previously assembled *D. paulistorum* mitochondrial genomes. Sequences corresponding to high scoring pairs (HSPs) with at least 200 bp length and 90% similarity to mitochondrial DNA were retrieved. Recovered sequences with more than 97% identity to either α or β genomes were deemed “conserved” and tentatively assigned to the clade they were most similar to, while those with 90-97% identity were considered divergent. Sequences retrieved from each assembly were mapped back to an α and a β mitochondrial genome (C2, O11 beta) with BWA-MEM and the number of bases covered by them were counted with Samtools mpileup. This was done in order to determine how much of the mitochondrial genome was covered by contigs belonging to each category (α , β , diverged). Genes present in the retrieved sequences were identified by blasting each sequence to a database containing all genes from the A28 mitochondrial genome (CDS, tRNA and rRNA). All contigs containing mitochondrial sequences were blasted against NCBI’s nt database for identification of potential matches to non-mitochondrial references.

Sequences retrieved from each genomic assembly were considered NUMTs if the contigs in which they were found a) also contained matches to non-mitochondrial references; or b) had average coverage similar ($\pm 30\%$) to the reference nuclear coverage of its line. The reference nuclear coverage for each line was calculated as the average coverage of the contigs containing the genes *CAD/r*, *Beta Spectrin* and *FASN3*. These were identified by blasting *D. willistoni* CDSs of these genes to each assembly. The reference mitochondrial coverage of each line was defined as the average per base coverage of the longest contig which had 100% identity to the line’s assembled mitochondrial genome(s).

Statistics

Statistical analyses varied depending on the data source and distribution. For the qPCR tissue distribution of α and β mitotypes between different ages and sexes we used ANOVA on log-transformed data to increase normality. In the qPCR analysis of A28 x O11 hybrids, we used ANOVA with post-hoc Tukey tests of significance. Due to low numbers of repeats and power in the intraspecies FG295 x O11 hybrids, Wilcoxon Rank Sum tests were used for significance testing. Assumptions and variance were checked before use. These analyses were completed using R version 3.3.2 [103].

In the RFLP-PCR experiments, the $\alpha 2$ -levels in embryos from different *D. paulistorum* semispecies and crosses at different developmental stages were compared using One-way ANOVA with post-hoc Tukey HSD tests. Quantifications were made with the help of an online calculator (www.astatsa.com). The raw data showing levels of $\alpha 2$ in percent were obtained with Image Lab 5.2.1 software and transformed from percentage to arcsine for normal distribution before applying the statistics. RFLP-data were statistically analyzed using ANOVA and Student’s t tests.

Fluorescent in situ hybridization (FISH) on DNA

DNA-FISH was performed on three different tissues: early embryos, mitotic chromosomes from 3rd instar larval brains and polytene chromosomes from salivary glands. Early embryos (0-3 h) were collected as stated above, dechorionated in 50% bleach and fixed for 20 min in 3.7% paraformaldehyde in PBX (PBS and 0.15% Triton X-100) mixed with heptane (1:1 v/v). Then the embryos were postfixed in ice-cold methanol and further processed for FISH following a protocol by [104] with slight modifications. In situ hybridization was performed at 37°C overnight in hybridization buffer containing approximately 150 ng of probe generated by direct labeling via PCR with modified dNTPs (Alexa Fluor 488 dUTP and Alexa Fluor 594 dUTP; Jena Bioscience, Germany). The mito-probe was used as a mixture of 5 mitochondrial loci (CO1, CO2, ND1, ND2 and ND4) of *D. paulistorum*, cloned beforehand in pTZ57R/5 vector (Thermo Scientific, USA) and isolated plasmids were labeled with the red AF594 fluorophore. As a hybridization control, we used a cloned probe targeting a 28S rRNA locus of *D. paulistorum* labeled it with the green AF488 fluorophore. Mitotic chromosomes from 3rd instar larval brains were processed for DNA-FISH according to the protocol from [105] after hypotonic treatment washing off cytoplasmic compounds and RNase treatment. DNA-FISH on polytene chromosomes was performed according to [106]. In both cases we were using the same probes and washing conditions as described above for fixed whole-mount embryos. Images were taken with Olympus FW3000 laser scanning confocal microscope using 60x objective and further processed with Photoshop CS6 adjusting levels for each channel.

ACKNOWLEDGEMENTS

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

DATA AVAILABILITY

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

REFERENCES

1. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 1980;77(11):6715-9. doi: 10.1073/pnas.77.11.6715. PubMed PMID: 6256757.
2. Birky CW. Relaxed cellular controls and organelle heredity. *Science*. 1983;222(4623):468. doi: 10.1126/science.6353578.
3. Stefano G, B., Bjenning C, Wang F, Wang N, Kream R, M. . Mitochondrial Heteroplasmy. In: Santulli G, editor. *Mitochondrial Dynamics in Cardiovascular Medicine Advances in Experimental Medicine and Biology*. 982. Cham: Springer; 2017.
4. Burr SP, Pezet M, Chinnery PF. Mitochondrial DNA Heteroplasmy and Purifying Selection in the Mammalian Female Germ Line. *Development, Growth & Differentiation*. 2018;60(1):21-32. doi: 10.1111/dgd.12420.
5. White DJ, Wolff JN, Pierson M, Gemmell NJ. Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*. 2008;17(23):4925-42. doi: 10.1111/j.1365-294X.2008.03982.x.
6. Sharpley MS, Marciniak C, Eckel-Mahan K, McManus M, Crimi M, Waymire K, et al. Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. *Cell*. 2012;151(2):333-43. doi: 10.1016/j.cell.2012.09.004. PubMed PMID: 23063123.
7. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature Reviews Genetics*. 2015;16(9):530-42. doi: 10.1038/nrg3966.
8. Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, et al. Cytonuclear integration and co-evolution. *Nature Reviews Genetics*. 2018;19(10):635-48. doi: 10.1038/s41576-018-0035-9.
9. Christie JR, Schaerf TM, Beekman M. Selection against Heteroplasmy Explains the Evolution of Uniparental Inheritance of Mitochondria. *PLOS Genetics*. 2015;11(4):e1005112. doi: 10.1371/journal.pgen.1005112.
10. Birky CW. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences*. 1995;92(25):11331. doi: 10.1073/pnas.92.25.11331.
11. Pitnick S, Karr TL. Paternal products and by-products in *Drosophila* development. *Proceedings of the Royal Society B: Biological Sciences*. 1998;265(1398):821-6. doi: 10.1098/rspb.1998.0366. PubMed PMID: 9628039.
12. Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G. Ubiquitin tag for sperm mitochondria. *Nature*. 1999;402(6760):371-2. doi: 10.1038/46466.
13. DeLuca Steven Z, O'Farrell Patrick H. Barriers to Male Transmission of Mitochondrial DNA in Sperm Development. *Developmental Cell*. 2012;22(3):660-8. doi: <https://doi.org/10.1016/j.devcel.2011.12.021>.
14. Yu Z, O'Farrell PH, Yakubovich N, DeLuca SZ. The Mitochondrial DNA Polymerase Promotes Elimination of Paternal Mitochondrial Genomes. *Current Biology*. 2017;27(7):1033-9. doi: 10.1016/j.cub.2017.02.014.
15. Sato K, Sato M. Multiple ways to prevent transmission of paternal mitochondrial DNA for maternal inheritance in animals. *The Journal of Biochemistry*. 2017;162(4):247-53. doi: 10.1093/jb/mvx052.
16. Wolff JN, Gemmell NJ. Lost in the zygote: the dilution of paternal mtDNA upon 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, Zury 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, Degan 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: PMC2996333.
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, Sängner 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
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: PMC1559843.
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: PMC3616963.
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, Kernysky 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](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.