TITLE : The evolutionary history of Drosophila simulans Y chromosomes reveals molecular signatures of resistance to sex ratio meiotic drive

AUTHORS : Courret C[1*], Ogereau D[2]., Gilbert C[2], Larracuente A.M[1]., Montchamp-Moreau C[2].

AFFILIATIONS :

1. Department of Biology, University of Rochester, Rochester, NY, USA
2. Laboratoire Evolution, Génome, Comportement et Ecologie, UMR UPSaclay, CNRS 9191, IRD 247, 91190 Gif-sur-Yvette, France

*corresponding author : cecile.courret@rochester.edu

ORCID:

0000-0001-5849-8014 (CC)

0000-0002-4496-3007 (DO)

0000-0002-2131-7467 (CG)

0000-0001-5944-5686 (AML)

0000-0002-5044-9709 (CMM)

ABSTRACT

The recent evolutionary history of the Y chromosome in *Drosophila simulans*, a worldwide species of Afrotropical origin, is closely linked to that of X-linked meiotic drivers (Paris system). The spread of the Paris drivers in natural populations has elicited the selection of drive resistant Y chromosomes. To infer the evolutionary history of the Y chromosome in relation to the Paris drive, we sequenced 21 iso-Y lines, each carrying a Y chromosome from a different location. Among them, 13 lines carry a Y chromosome that is able to counteract the effect of the drivers. Despite their very different geographical origins, all sensitive Y's are highly similar, suggesting that they share a recent common ancestor. The resistant Y chromosomes are more divergent and segregate in four distinct clusters. The phylogeny of the Y chromosome confirms that the resistant lineage predates the emergence of Paris drive. The ancestry of the resistant lineage is further supported by the examination of Y-linked sequences in the sister species of *D. simulans, D. sechellia* and *D. mauritiana*. We also characterized the variation in repeat content among Y chromosomes and identified two simple satellites (*(AAACAAT)n* and *(AATGG)n*) associated with resistance. Altogether, the molecular polymorphism allows us to infer the demographic and evolutionary history of the Y chromosome and provides new insights on the genetic basis of resistance.

39 **INTRODUCTION**

40 In *Drosophila melanogaster* and *D. simulans*, like in most Drosophila species, the Y
41 chromosome is entirely heterochromatic, composed of transposable elements and other
42 repetitive sequences such as satellite DNA (Bachtrog 2013). In both species the Y chromosome
43 represents 20% of the male genome (around 40 Mb) (Hoskins et al. 2002) but contains only a
44 handful of genes, expressed during spermatogenesis and necessary for male fertility (Kennison
45 1981; Bernardo Carvalho, Koerich, et Clark 2009). Indeed, XO males are usually viable but
46 sterile. Beyond its essential role in male fertility, the Y chromosome of *D. melanogaster* has
47 been shown to impact male fitness through a variety of phenotypic effects, on behavior or
48 thermo tolerance (Huttunen et Aspi, s. d.; Rohmer et al. 2004). In fact, thanks to its
49 heterochromatic nature, it appears able to modulate gene expression across the whole genome
50 by modifying the chromatin landscape (Brown, Nguyen, et Bachtrog 2020). Similar effects are
51 likely in *D. simulans* (Branco et al. 2013).

52 The large phenotypic variation caused by the Y chromosome of *D. melanogaster* is in sharp
53 contrast with its very low variation in gene sequence compared to other chromosomes
54 (Zurovcova et Eanes 1999; Larracuente et Clark 2013). In *D. simulans,* the sibling species of
55 *D. melanogaster*, the gene diversity of the Y chromosome is even lower (Zurovcova et Eanes
56 1999; Kopp, Frank, et Fu 2006; Helleu et al. 2019).

57 This poor genetic variation can be explained by multiple factors. First, the Y chromosome has
58 a reduced effective population size compared to the rest of the genome (one quarter of that of
59 the autosomes). Second, because of its hemizygous state, the fitness conditions allowing the
60 maintenance of stable, non-neutral polymorphism are very restrictive (Clark 1987; Clark 1990).
61 Finally, the Y chromosome does not recombine, which causes natural selection on a Y-linked
62 locus to affect the entire chromosome. It follows that beneficial mutations can lead to the
63 fixation of the carrier Y chromosome into the population (Charlesworth, Coyne, et Barton

64 1987), while a deleterious mutation can lead to its loss. Population genetics models thus suggest
65 that neutral variations within populations should be rarely maintained whereas inter-population
66 differences can rapidly accumulate (Clark 1987). Thus, the *Drosophila* Y chromosome is
67 expected to be particularly sensitive both to the demographic history of populations and to
68 selection events. Indeed, in *D. melanogaster*, the ancestral African populations show
69 significantly less Y-linked variation than the cosmopolitan ones, which could be explained by
70 the demographic history of the species and a recent natural selection event in Africa
71 (Larracuente et Clark 2013).
72 Y chromosome evolution can also be shaped by intragenomic conflict triggered by X-linked
73 meiotic drivers. These favor the transmission of X chromosomes bearing them (hereafter X^{SR}),
74 thus, increasing the proportion of females in progeny and inducing a deviation from the
75 balanced - also called Fisherian- sex ratio. *Sex Ratio* (*SR*) systems are widespread in the
76 *Drosophila* genera, with 18 *SR* systems known in 16 species (Courret et al. 2019). During
77 spermatogenesis, an X^{SR} typically prevents the formation of functional Y bearing sperm,
78 resulting in a strong female bias in the progeny of carrier males. Thus, the effective size of the
79 Y chromosome is drastically reduced in populations invaded by X^{SR}. Following the spread of
80 X^{SR}, natural selection strongly favors any variant that can counteract the effect of the driver and
81 restore a balanced sex ratio (review in Helleu, Gérard, et Montchamp-Moreau 2015). Such drive
82 suppressors are expected to evolve at unlinked loci, *i.e.*, on autosomes and the Y chromosome.
83 Resistant Y chromosomes have direct and strong selective advantage and thus are expected to
84 quickly reach fixation (Clark 1987; Hall 2004).
85 In the worldwide species *Drosophila simulans*, autosomal suppressors and resistant Y
86 chromosomes have evolved in response to the spread of X-linked drivers known as the Paris
87 system (Cazemajor, Landré, et Montchamp-Moreau 1997). The Paris system likely emerged
88 less than 500 years ago in East Africa (Fouvry et al. 2011; Bastide et al. 2011). It then spread

89 to islands of the Indian Ocean and in Sub-Saharan Africa (Bastide et al. 2013). Drive
90 suppression is complete over the whole area and the frequency of X^{SR} was found to decrease in
91 some locations, like in Madagascar, Kenya or Mayotte (Bastide et al. 2011). During the last
92 decade, invasion of X^{SR} was observed in North African populations, where sensitive Y
93 chromosomes were then replaced by resistant ones within a few years (Helleu et al. 2019). On
94 the whole, in *D. simulans*, the Y chromosomes appeared to be highly polymorphic for their
95 resistance ability with a continuum of phenotypes, from complete resistance to high sensitivity
96 (Montchamp-Moreau, Ginhoux, et Atlan 2001). In contrast, they exhibit a very low level of
97 gene sequence variation, with only three single nucleotide polymorphisms (SNPs) detected
98 along a total of 13 kb of gene fragments; those define three haplotypes - TTA, TTT, CAA - the
99 latter of which is associated with resistance to Paris *SR* drive (Helleu et al. 2019). This strongly
100 suggests that the phenotypic variation results from variation in repeated sequences and/or from
101 structural variations, both of which are extensive on the Y chromosomes of *D. simulans* (Helleu
102 et al. 2019; Chang et al. 2022).
103 Thanks to the recent advance in long read sequencing, about one third of the Y chromosomes
104 of *D. melanogaster* and *D. simulans* (13 Mb out of 40), are now assembled (Chang et
105 Larracuente 2019; Chang et al. 2022). Here we take advantage of those assemblies to further
106 investigate the molecular variation of the Y chromosome in *D. simulans* in relation to the
107 evolution of the Paris *SR* system. We sequenced 21 iso-Y lines (derived from a single male)
108 from different locations around the world, 8 from the sensitive lineage TTA and 13 from the
109 resistant lineage CAA (Figure 1A). Despite their very different geographical origins all
110 sensitive Y chromosomes appear to be highly similar. The resistant Y chromosomes are more
111 divergent and segregate in four distinct groups. By reconstructing the phylogeny of the Y
112 chromosome, we confirmed that the resistant lineage predates the emergence of the Paris X^{SR}.
113 We also examined the variation in repeat content among Y chromosomes. We observed almost

114 no variations in gene and TE copy number, neither of them showing an association with the
115 resistant phenotype. However, we identified two simple satellites ((AAACAAT)n and
116 (AATGG)n) associated with resistance. Altogether this new approach allowed us to better
117 understand the evolutionary history of the Y chromosome in relation with *SR* systems and also
118 give us insight on the genetic basis of the resistance.
119
120
121 **RESULTS**
122 ***1. Resistance ability of the Y chromosomes from the resistant CAA lineage***
123 Among the 21 *D. simulans* Y chromosomes included in the present study, 13 belong to
124 the resistant CAA lineage and were checked for their resistance ability. To do so we produced
125 males carrying the Y chromosome to be tested, a reference X^{SR} chromosome, and suppressor-
126 free autosomes from the reference ST8 line (crossing scheme in Montchamp-Moreau, Ginhoux,
127 et Atlan 2001). For each Y, the individual progeny of 15 males was examined (Figure 1B).
128 We observed significant variation in the resistance ability among the CAA Y chromosomes
129 (ANOVA test, F=12.36, df=13, p<2e-16). The mean sex-ratio in the progeny of carrier males
130 ranged from 60% to 83% depending on the Y chromosome, compared to 89% for the sensitive
131 Y^{ST8} (Figure 1B, SupTable1). Except for Sey81 (Seychelles 1981) and New91 (New Caledonia
132 1991), which exhibit a low resistance ability with a mean sex-ratio of 81% and 83%,
133 respectively, all CAA Y chromosomes appear significantly more resistant compared to Y^{ST8},
134 our reference sensitive Y chromosome (Figure 1B, SupTable1).
135
136
137
138

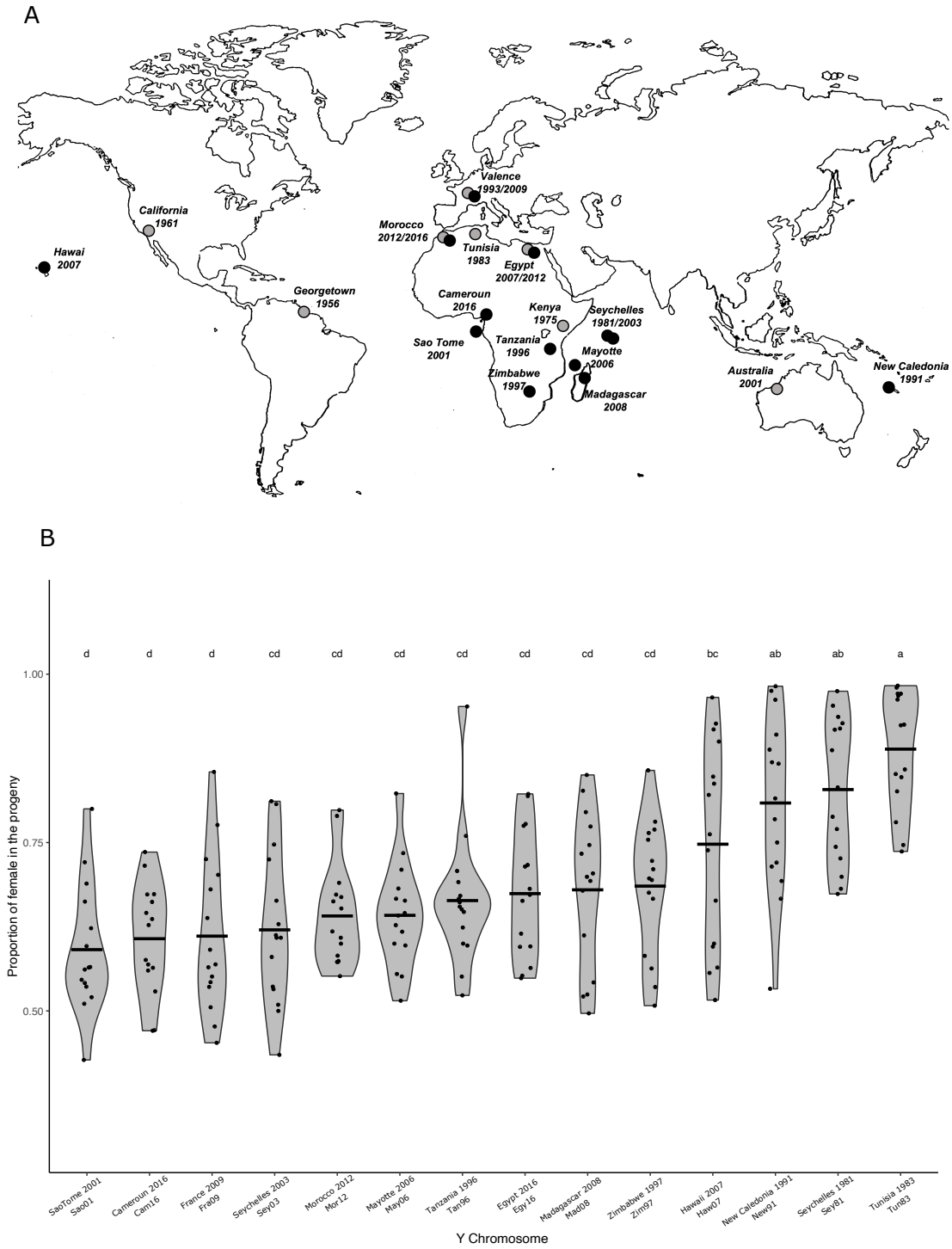

Figure 1. A. Geographical location and date of collection of the Y chromosomes studied (●) Lines carrying a Y chromosome with the CAA haplotype (resistant). (●) Lines carrying a Y chromosome with the TTA haplotype (sensitive). B. Test of the resistance ability of the Y chromosomes. Each violin plot represents the proportion of females in the progeny of males carrying the reference X^{SR} and the Y chromosome from each iso-Y line. Tun83 corresponds to our reference sensitive Y chromosome. Each dot corresponds to the progeny of an individual male. The letters above each violin plot correspond to the group determined by the Tukey HSD post hoc-test performed on an Anova model.

147 ***2. The molecular variation essentially occurs among resistant Y chromosomes.***

148 Using the Y chromosome from the XD1 strain (Chang et al. 2022) as a reference, we
149 identified 8190 Y-linked polymorphic sites among the 21 iso-Y lines sequenced. The global
150 nucleotide diversity is very low (1.808e-04) and most of this diversity comes from the resistant
151 Y chromosomes (SupFig1). On average, these share the same allele at 76.89% of the
152 polymorphic sites, while on average the sensitive Y chromosomes share the same alleles at
153 97.77 % of the detected polymorphic sites (Table1). Sensitive Y chromosomes are thus highly
154 similar despite originating from geographical regions far away from each other. They all mainly
155 share the same allele as the reference allele (SupFig1).

156 To examine the population structure of the Y chromosomes, we performed a PCA analysis
157 based on the 8190 SNPs identified. It revealed 5 different clusters. Unsurprisingly, based on
158 their high percentage of identity, all the sensitive Y chromosomes cluster together (Cluster I,
159 Figure 2). The resistant Y chromosomes segregate in 4 different clusters. Cluster II is composed
160 of Egy12, Tan96, Zim97, Mor16, Cam16 and Sey03, all originating from continental Africa or
161 nearby islands. Cluster III is composed of Sey81 and New91, from Seychelles and New
162 Caledonia, respectively. Cluster IV consists of May09 and Mad98, respectively sampled in
163 Mayotte and Madagascar, the area from which the *D. simulans* species is thought to have
164 originated (Lachaise et Silvain, 2004; Dean et Ballard 2004). Cluster V is composed of Fra09,
165 Haw07 and Sao01, which are from very far apart geographical areas (France, Hawaii and
166 SaoTome, respectively). The first axis of the PCA (29.8%) separates Cluster III from all the
167 other Y chromosomes (Figure 2). This means that the divergence between Sey81/New91 and
168 all the other Ys is higher than between all the other resistant Y chromosomes and the sensitive
169 Y chromosomes. Interestingly, those Y chromosomes from Seychelles and New Caledonia are
170 also very similar in resistance ability (Figure 1B).

171 None of the pairs of sensitive-resistant Y chromosomes originating from the same geographical
172 area cluster together. This confirms that the resistant chromosomes did not emerge locally but
173 rather arose from migration along with the driver. While the sensitive Y chromosomes (Egy07,
174 Mor12 and Fra93) are older than their resistant counterparts, *i.e.* were present respectively in
175 North Africa and Europe before the arrival of the Paris *SR* system, they appear to be very
176 similar. Indeed, they group together into Cluster I. By contrast, the resistant chromosomes
177 sampled more recently in those populations are more variable. The resistant Y chromosomes
178 from Morocco (Mor16) and Egypt (Egy12) group together in Cluster II, while the resistant Y
179 chromosome from France (Fra09) belongs to Cluster V. Both resistant Y chromosomes from
180 Morocco (Mor16) and Egypt (Egy12) appear to be closely related, whereas the resistant Y
181 chromosome from France (Fra09) likely originated from a different migration path.
182 Finally, the two resistant Y chromosomes collected in Seychelles in 1981 (Sey81) and 2003
183 (Sey03) do not cluster together. They have very different resistance ability (Figure 1B) and each
184 cluster with Y chromosomes showing a similar level of resistance (Figure 2). Because these Y
185 chromosomes were collected more than 20 years apart, we cannot conclude if there was a
186 polymorphism within the Seychelles islands where both Y chromosomes segregate in the
187 population or if there was a migration event from continental Africa which led to the
188 introduction of a new Y chromosome in Seychelles islands.
189

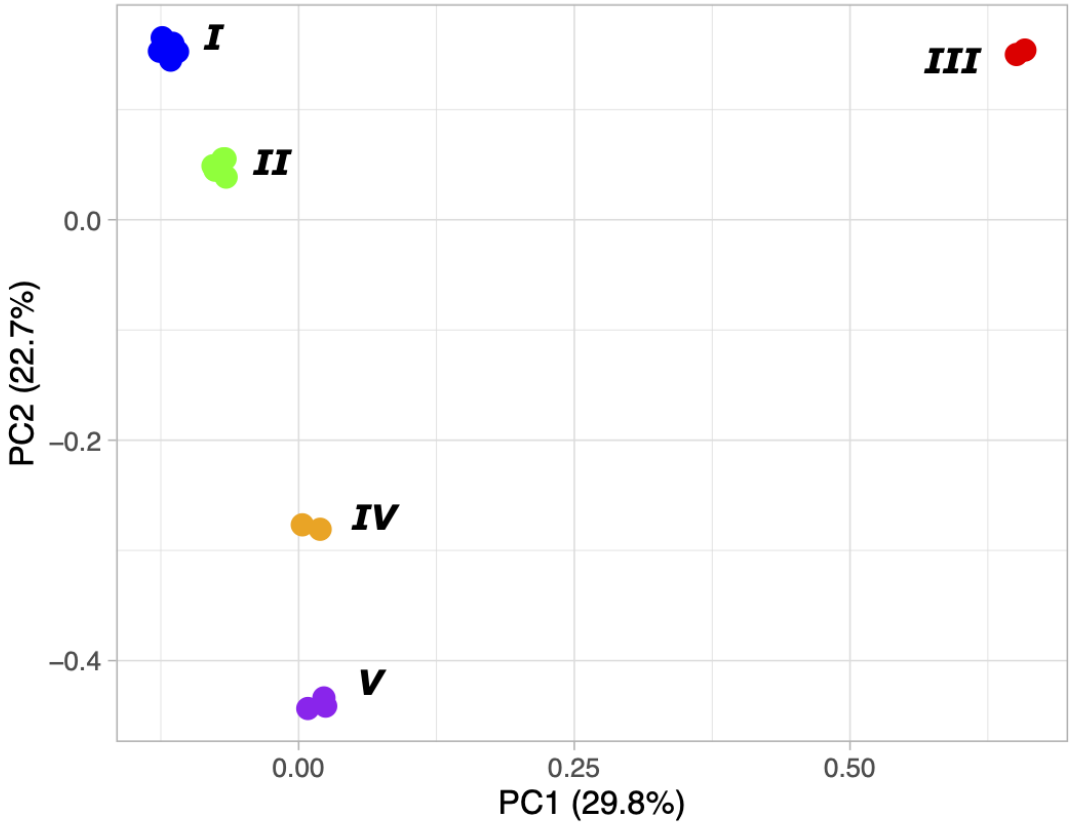

Figure 2. Principal component analysis based on nucleotide diversity. The 8190 polymorphic sites located on the Y chromosome are taken into account for this analysis. Y chromosomes have been colored based on the cluster. Cluster numbers are indicated by roman numbers.

Table 1: Summary of the number of shared SNPs per cluster.

Groups	*Sensitive (Cluster I)*	*Resistant (all)*	*Resistant (Cluster II)*	*Resistant (Cluster III)*	*Resistant (Cluster IV)*	*Resistant (Cluster V)*
N° shared SNP	6985	1123	6191	7488	7644	6704
Average % of identity	97.77	76.89	97.78	96.13	97.58	95.77

200 ### *3. Very little within-population variation of the Y chromosome*

201 To get an insight into the intra-population polymorphism of the Y chromosome, we
202 genotyped by PCR-sequencing 98 additional chromosomes from 22 populations. This first
203 allowed us to validate 11 SNPs detected by our approaches and to look at variation into the
204 population. We reconstructed a haplotype tree that recapitulates the clusters identified by the
205 PCA (Figure 2), except for cluster IV and V that grouped together as we do not have SNP
206 distinguishing each cluster (SupFig2). Most of the time all the individuals collected in the same
207 location at the same time carry the same haplotype (SupTable2). Only one population from
208 France (collected in 2014 and 2015) seems to be polymorphic, containing Y chromosomes from
209 cluster I, II and V. We postulate that recurrent annual migrations occur in western Europe, in
210 accordance with the observation that *D. simulans* cannot overwinter in temperate regions
211 (Boulétreau-Merle, Fouillet, et Varaldi 2003; Machado et al. 2016). In addition, migration in
212 *D. simulans* is likely largely human-aided and must have broadly amplified all over the world
213 in the recent past, as suggested by the rapid invasion of the P-transposable element within the
214 last decade (Hill, Schlötterer, et Betancourt 2016).

215

216 ### *4. The resistant lineage predates the emergence of the driver*

217 Previous work suggested that resistance to X^{SR} is ancestral to all *D. simulans* Y chromosomes
218 because the species sister to *D. simulans* (*D. mauritiana* and *D. sechellia*) carry the CAA
219 haplotype, which is associated with resistance in *D. simulans* (Helleu et al. 2019). We
220 reexamined this hypothesis in the light of our new sequencing data. We sought to identify the
221 corresponding allele carried by *D. melanogaster*, *D. mauritiana* and *D. sechellia* for the 8190
222 SNPs identified as polymorphic within *D. simulans*. We were able to recover 6138 SNPs in *D.*
223 *mauritiana*, 6126 SNPs in *D. sechellia* and 3909 SNPs in *D. melanogaster*. Based on the 8190
224 SNPs we reconstructed the phylogeny of the Y chromosomes using a Maximum likelihood

225 approach and *D. melanogaster* as the outgroup (Figure 3). The phylogeny defines 5 clades
226 corresponding to the 5 clusters identified by the PCA (Figure 2). According to this tree it
227 appears clear that the ancestral Y chromosome belongs to the resistant lineage. Indeed, the first
228 divergent clade is cluster III, composed of two resistant Y chromosomes Sey81 and New91.
229 Whereas the clade that includes all sensitive Y chromosomes (cluster I) appears to have a recent
230 common ancestor nested within clusters of resistant Y chromosomes. The sensitive haplotype
231 seems to diverge from the same ancestor as the resistant Y chromosome of cluster II.
232 In addition, we focused on the 590 SNPs that correspond to fixed differences between the four
233 clusters of resistant Y's and the cluster of sensitive Y's (Cluster I), in order to identify the allele
234 carried by the sister species. While we were not able to recover all the SNPs, 97.4% of the
235 recovered ones carry the resistant allele in *D. mauritiana*, 94.8% in *D. sechellia* and 93.7% in
236 *D. melanogaster*.
237 Altogether, we conclude that the common ancestor of the *melanogaster* clade was carrying a Y
238 chromosome closely related to the resistant lineage. Thus, while the sensitive lineage is the
239 most common Y chromosome lineage found worldwide, it is a divergent lineage that seems to
240 have a recent origin. This also means that the resistant Y chromosome lineage predates the
241 emergence of the Paris *Sex Ratio* driver in *D. simulans*.

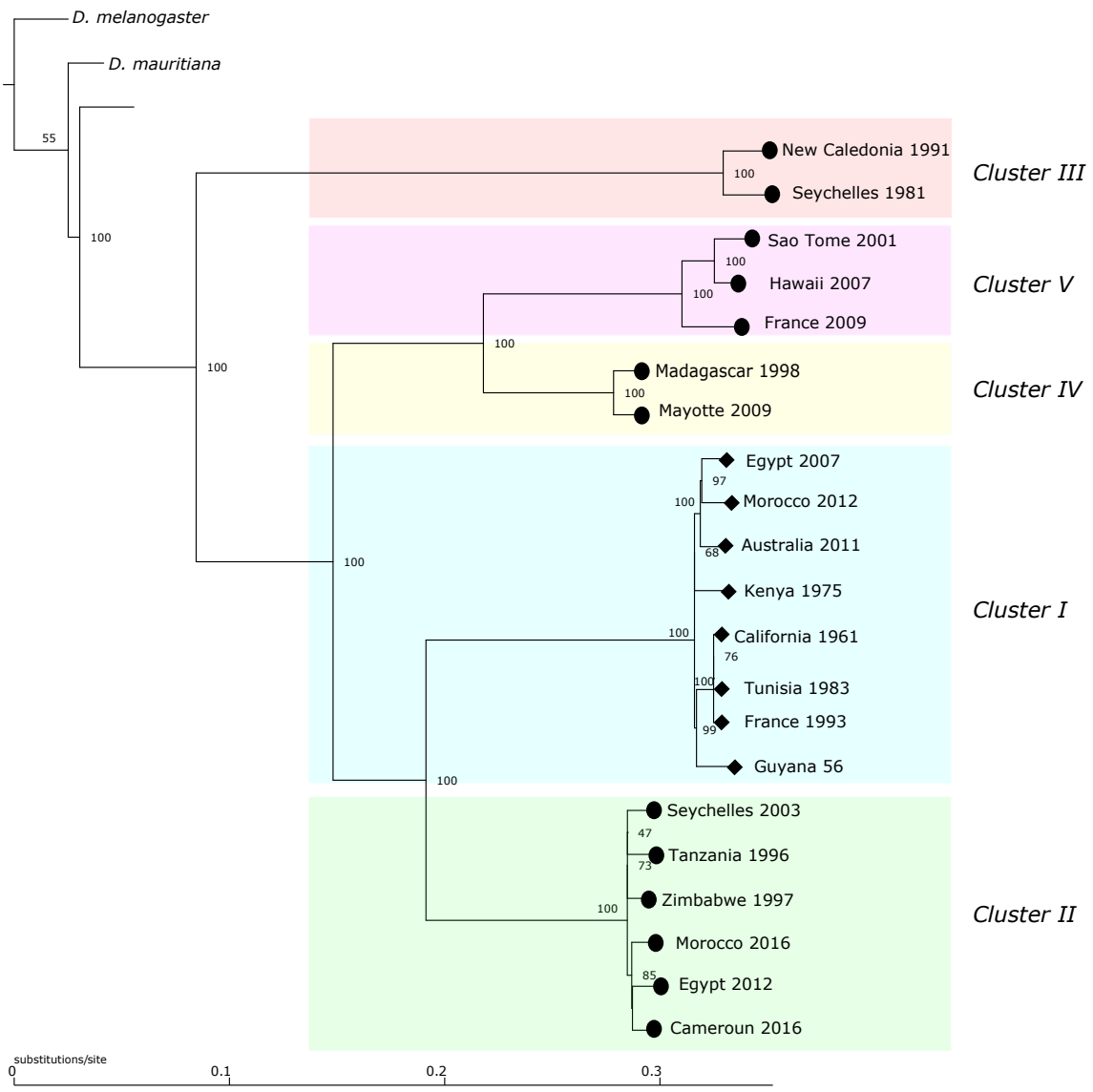

Figure 3. The inferred maximum likelihood phylogeny for the Y chromosome. *D.melanogaster* was used as the outgroup. Node labels indicate ultrafast bootstrap. The resistant Y chromosomes are indicated by a circle at the tip and sensitive Y chromosomes are indicated by diamond at the tip. Clusters are colored based on the cluster defined by the PCA analysis (Figure 2).

253 ***5. Gene copy numbers appear stable among Y chromosomes***

254 We first estimated the exon copy number for each of the 11 canonical Y-linked genes
255 (Figure 4A). Our estimates are in agreement with the previous estimates from the reference
256 genome (Chang et al. 2022). Most of the canonical genes are in single copy and do not exhibit
257 variation between Y chromosomes (Figure 4A). ARY was previously described as a multi copy
258 gene with 4 copies on the reference genome; here we estimated between 4 and 6 copies of ARY
259 among our 21 Y chromosomes. Similarly, *kl2* exons 8 to 12 are known to be duplicated, with
260 more than 15 copies for exon 10 for example. They are also the most variable among the 21 Y
261 chromosomes, with *kl2*-exon10 showing between 10 and 17 copies. *kl3*-exon16 shows an
262 interesting variation: while most of the Y chromosomes have 3 copies, the 3 Y chromosomes
263 that belong to Cluster V (Haw07, Sao01, Fra09) show a high duplication rate (between 7 and
264 11 copies) (Figure 4A).

265 In addition to the 11 canonical genes, two ampliconic genes have been described on the Y
266 chromosome of *D. simulans* : *CK2B* and *Lhk* (Chang et al. 2022). Accordingly, both exhibit
267 variation in exon copy number among our 21 Y chromosomes. We estimate between 83 and
268 170 copies of *CK2B*, between 91 and 156 copies of *Lhk*-exon1 and between 56 and 113 copies
269 of *Lhk*-exon2, depending on the Y (Figure 4B).

270 In general, only exons that are multi-copy appear to be variable in terms of copy number.
271 Importantly, we did not find any evidence for an association between the exon copy number
272 and the resistance phenotype (Figure 4 A and B). It thus appears unlikely that the variation in
273 resistance is due to a variation in exon copy number.

274

275 ***6. Very low Y-linked variation in transposable element abundance between populations***

276 Using our Illumina sequences we estimated the abundance of transposable elements and
277 complex satellites in 16 iso-Y lines (6 sensitive and 10 resistant; see materials and methods).
278 Because our Y chromosomes have been isolated in the same genetic background, we assumed

279 that variation in the abundance of repeats between iso-Y lines is associated with the Y
280 chromosome. To investigate whether variation in Y-linked complex repeat abundance shows
281 the same clustering as Y-linked nucleotide polymorphism, we performed a PCA analysis.
282 Globally the complex repeat abundance does not replicate the clusters defined by SNPs in
283 Figure 2. The sensitive Y chromosomes (Cluster I) show a diffuse distribution that partially
284 overlaps with some resistant Y chromosomes. Similarly, the distributions of resistant Y
285 chromosomes from clusters III, IV and V largely overlap. Only the Y resistant chromosomes
286 from cluster II fully separates from all the others (Figure 4C).
287 Next, we tested whether the phenotypic variations observed among the Y chromosomes were
288 associated with variations in gross repeat content. We thus performed a Wilcoxon test to assess
289 whether some transposable elements or complex satellites might be differentially enriched
290 between sensitive and resistant Y chromosomes. None of these repeats show significant
291 variation (SupTable3), indicating that the abundance in complex repeats is not associated with
292 the resistance phenotype. More generally, it appears that most complex repeats exhibit very low
293 to no variation in abundance between iso-Y lines (see SupTable3). We calculated the coefficient
294 of variation for each complex repeat and plotted the abundance for the Top25 most variable
295 ones (Figure 4D). The NTS-240 element is the only complex repeat that shows substantial
296 variation, with more than 10-fold variation in abundance between Y lines (Figure 4D). While
297 lower in abundance, the telomeric transposable elements (Het-A, TART and TAHRE) are also
298 among the most variable repeats (Figure 4D). This is in agreement with the existence of those
299 elements in various regions on the Y chromosome (Agudo et al. 1999). However overall, the
300 majority of these Top25 most variable repeats show relatively low variation between iso-Y
301 lines. It is important to note that the abundance represented here is the whole genome
302 abundance. Thus complex repeats that are not associated with the Y chromosome are expected
303 to show low to no variation, the level of which could reflect that of background noise. This is

the case for the *360bp* satellite, which is a complex satellite in the 1.688 family, highly abundant on chromosome II and III in *D. simulans* (Sproul et al. 2020; de Lima, Hanlon, et Gerton 2020) and not associated with the reference Y chromosome (Chang et al. 2022). The *360-bp* satellite shows 1.8-fold variation among iso-Y lines, and we therefore assume that repeats at or below this threshold likely represents background noise (Figure 4D).

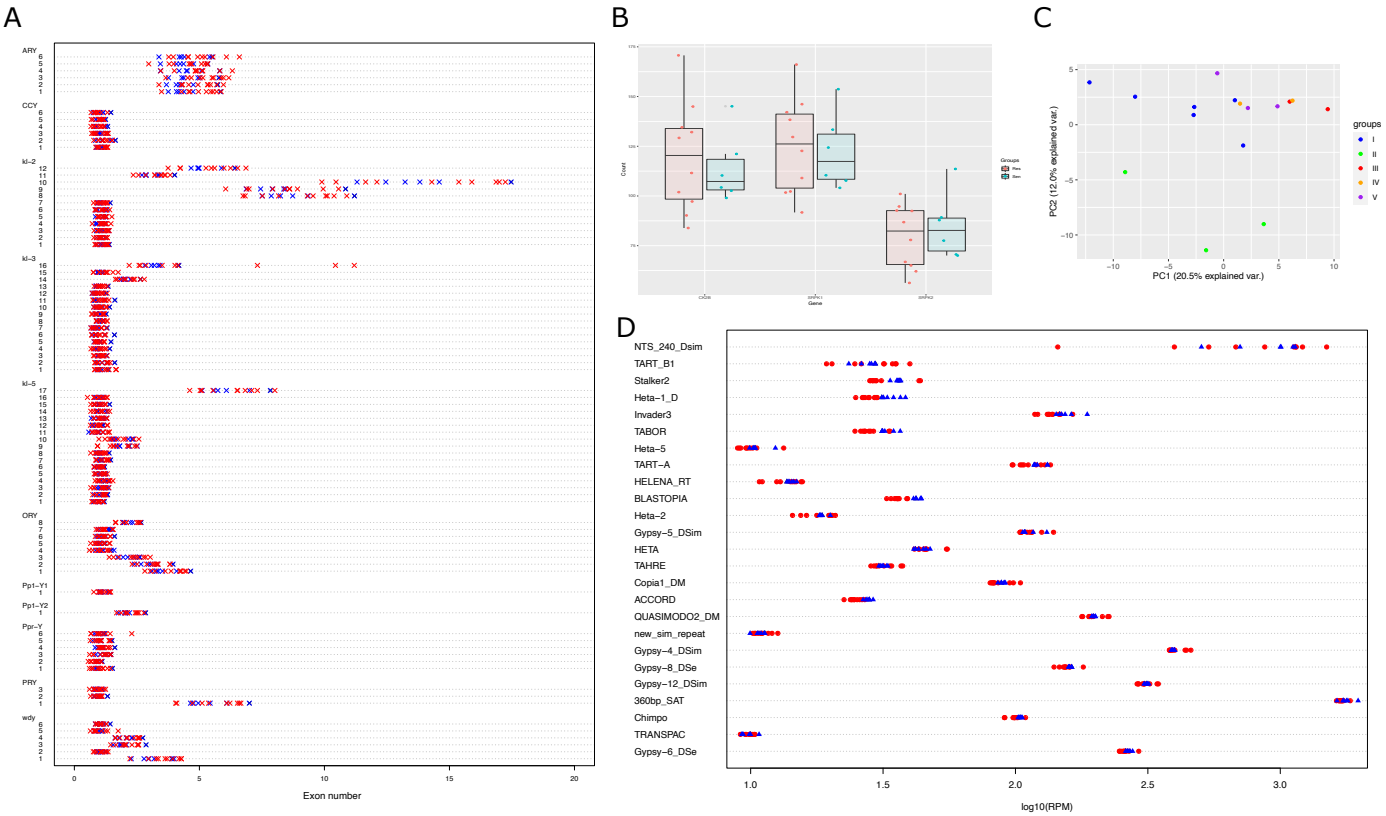

Figure 4: A. Dotplot representing the exon number for each Y associated genes. Each line corresponds to a specific exon of a specific gene. Red crosses correspond to resistant Y chromosomes and blue crosses to the sensitive Y chromosomes. B. Boxplot representing the exon number for the two Y-linked multicopy gene families: *CK2B* and *Lhk*. Resistant and sensitive Y. C. PCA based on the complex satellite abundance. Each Y chromosome is colored based on the groups defined previously using the nucleotide diversity (Figure 2). D. Dotplot representing the top 25 most variable complex satellites (based on coefficient of variation), among which the *360 bp* satellite is indicative of the background noise. To be able to distinguish the differences between strains, RPM counts have been transformed in log10. The Y resistant chromosomes are represented by a red circle while the sensitive Y chromosomes are shown with a blue triangle. Chromosomes have been separated in two boxplots, red and blue respectively. Each dot corresponds to a Y chromosome.

321

322 ***7. Simple satellite enrichment reveals two strong candidates for the resistance.***

323 Simple repeats represent over 5% of *D. simulans* genome and are a major component of
324 the Y chromosome (Wei et al. 2018). We compared the simple satellite composition between
325 the same 16 iso-Y lines (6 sensitives and 10 resistants). Using the k-seek pipeline (Wei et al.
326 2014) we identified 177 simple satellites, with the most abundant kmers being previously
327 described to be highly enriched in *D. simulans* (Wei et al. 2018) (SupTable4). Interestingly, we
328 found that simple satellite abundance recapitulates the structure between iso-Y lines defined by
329 the SNP calling (Figure 5A vs Figure 2).

330 Among the 177 satellites, 28 appear differently enriched on resistant versus sensitive Y
331 chromosomes (Figure 5B). Two satellites known to be Y-specific (Wei et al. 2018) are
332 particularly interesting. The first one, (AAACAAT)n is among the most abundant satellites and
333 is 2-fold more enriched in the sensitive lineages than in the resistant lineages (Figure 5C). While
334 it does not show a strong correlation with the resistance phenotype (SupFig3), its abundance
335 makes it a good candidate underpinning resistance to X^{SR}. Our second candidate, (AATGG)n,
336 is less abundant, but the sensitive Y chromosomes are also more enriched than the resistant
337 ones. More importantly, this satellite shows a very strong correlation with the resistance
338 phenotype (Spearman correlation test; pvalue<0.01, Figure 5D).

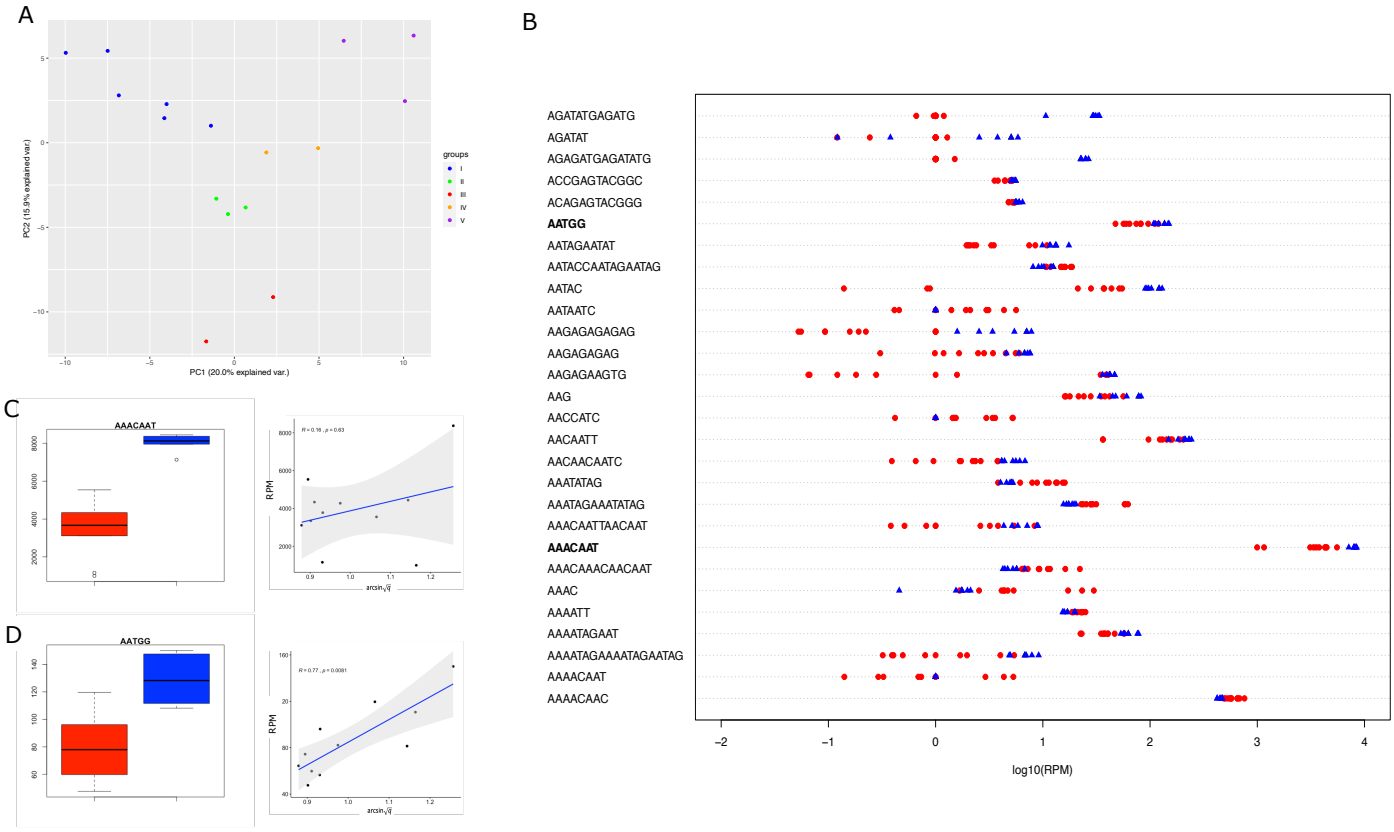

Figure 5. A. PCA based on the simple satellite abundance. Each Y chromosome is colored according to the cluster to which it belongs (defined by nucleotide diversity, Figure 2B). B. Dotplot of the 28 simple satellites differentially enriched between the resistant Y's (red circles) and the sensitive ones (blue triangle). RPM counts were transformed in log10. C and D. The box plots represent the abundance of (AAACAAT)n and (AATGG)n in each group (resistant in red and sensitive in blue). The dot plot represents the correlation between the satellite abundance (y axis) and the resistance ability (x-axis). Correlation was calculated based on a Spearman correlation test.

347 **DISCUSSION**

348

349 Y-linked nucleotide diversity in both *Drosophila melanogaster* and *Drosophila*
350 *simulans* is known to be very low compared to the rest of the genome (Larracuente et Clark
351 2013; Helleu et al. 2019). Those previous estimates were done on genic sequences, which limits
352 the estimation to a small portion of the Y chromosome. Here, taking into account the whole
353 assembled chromosome (around 13Mb) we show that the Y-linked nucleotide diversity in *D.*
354 *simulans* is about 15-fold lower than the estimate in *D. melanogaster* (π was estimated ranging
355 from 0.00266 to 0.01696 in Larracuente et Clark 2013, here $\pi = 0.0001808$). While the level of
356 nucleotide variation among the Y chromosomes examined here is very low, it reflects the
357 evolutionary history of the species.

358

359 ***Evolutionary history of the Y chromosome related to the demographic history of D.***
360 ***simulans***

361

362 The ancestral origin of the Resistance

363 *D. simulans* is thought to originate from Madagascar or the surrounding area (Lachaise
364 et Silvain, 2004.; Dean et Ballard 2004). We hypothesize that this ancestral population was
365 carrying a drive resistant Y chromosome (Figure 6A). Indeed, according to the phylogeny
366 (Figure 3), the sensitive Y chromosomes are derived recently in *D. simulans* from resistant Y
367 chromosomes, suggesting that the ancestral Y chromosome in *D. simulans* was of the resistant
368 type. The ancestry of the resistant lineage does not necessarily mean that the resistant phenotype
369 is ancestral. However, based on our phylogeny it is most parsimonious that the sensitive lineage
370 (Cluster I) corresponds to a secondary loss of resistance to Paris SR. Consistently, previous

371 work suggested that the Y chromosome of *D. sechellia* is resistant to the Paris drivers (Tao et
372 al. 2007; Rice 2014).
373
374 Spread of *D.simulans* in Seychelles and Continental Africa.
375 During a first expansion event, *D. simulans* is thought to have independently colonized
376 the Seychelles islands and the African continent (Lachaise et Silvain, 2004) (Figure 6A). Our
377 data well support this scenario. Indeed, most of the Y chromosomes from continental Africa
378 belong to Cluster II while the two chromosomes from Seychelles and New Caledonia belong to
379 Cluster III (Figure 2). The phylogeny is in agreement with the early divergence of these two
380 chromosomes. Not only do they appear genetically different but they also have a different
381 phenotype. Indeed the Y chromosomes from Seychelles (Sey81) and New Caledonia (New91)
382 appear less resistant than the other resistant Y chromosomes (Figure 1B).
383 This scenario for the demographic history of *D. simulans* was supported by the
384 distribution of the mitochondrial haplotype. In *Drosophila simulans* there are three different mt
385 haplotypes : *si*I, *si*II and *si*III (Solignac et Monnerot 1986). *si*II is the worldwide haplotype of
386 the species. In Madagascar and La Reunion, populations appear polymorphic for *si*III and *si*II.
387 The last haplotype, *si*I, is specific to the Indo-Pacific islands (Seychelles, New Caledonia and
388 Hawaii). We controlled for the mitochondrial haplotype of our strain. As expected most of them
389 carry the siII haplotype with the exception of the strain from Hawaii, New Caledonia and
390 Seychelles that carry the siI haplotype, and Mayotte that carry the siIII haplotype (SupTable5).
391
392 Emergence and spread of the Sensitive lineage
393 The phylogeny indicates that the sensitive Y chromosomes (Cluster I) and resistant Y
394 chromosomes from Continental Africa (Cluster II) have a recent common ancestor (Figure 3).
395 This suggests that the sensitive lineage emerged in Continental Africa and started to replace

396 resistant Y chromosomes (Figure 6B). This hypothesis is supported by the Ken75 Y
397 chromosome that comes from flies collected in Kenya in 1975. The Ken75 Y chromosome
398 appears similar to the sensitive Y chromosomes found across the world

399 The Y chromosome which has then spread out-of-Africa is the sensitive Y
400 chromosome, becoming the major Y chromosome across the species range (Helleu et al. 2019)
401 (Figure 1A, Figure 6C). The low nucleotide diversity (>97% similarity, Table 1) within the
402 sensitive Y lineage suggests that this event probably arose very recently. Indeed it has been
403 shown that the expansion of *D. simulans* is more recent than that of *D. melanogaster*, likely
404 6500-5000 years in Eurasia (Lachaise et Silvain, 2004) and 500 years in North America (Wall,
405 Andolfatto, et Przeworski 2002).

406 Why has the sensitive Y chromosome previously replaced the ancestral resistant Y
407 chromosome? Multiple non-exclusive factors could have favored the spread of Y sensitive
408 lineage. First, the emergence of an advantageous mutation may have led to a strong positive
409 selection on sensitive Y chromosomes. A selective advantage associated with the Y
410 chromosome can quickly lead to its fixation into the population (Clark 1987). The Y sensitive
411 chromosome might improve the fitness of the carrier over the resistant Y chromosome, they
412 can increase the fertility or favor the adaptation to a new environment. Second, the demographic
413 history of the species could also explain the low level of polymorphism of the sensitive Y
414 chromosome. Indeed, effects such as migration and bottleneck could have facilitated the spread
415 of the sensitive lineage and reduced its effective population size. Finally, the last hypothesis
416 involves recurrent genetic conflict in *D. simulans*. In *D. simulans,* there are at least 3
417 independent *SR* systems; Paris, Winter and Durham. The Winter and Durham SR systems are
418 cryptic and presumably older than the Paris system (Meiklejohn et al. 2018). *SR* systems appear
419 common in this species (Muirhead et Presgraves 2021; Vedanayagam, Lin, et Lai 2021). The
420 intense resulting genetic conflicts between X and Y chromosome have an important impact on

421 sex chromosome evolution. We hypothesize that the presence of an X-linked driver in *D.*
422 *simulans*, before the emergence of the Paris *SR* system, could have driven the spread of some
423 Y chromosomes belonging to the lineage sensitive to the Paris driver. Some of the Y
424 chromosomes sensitive to the Paris driver could have been resistant to this more ancient driver.
425 The quick replacement of sensitive by resistant Y chromosomes might have occurred in the past
426 but triggered by another driver. This ancestral driver would have favored the spread of the
427 sensitive Y chromosome lineage before becoming cryptic or disappearing.
428
429 Emergence of the Paris system and replacement of the sensitive by the resistance Y
430 lineage.
431 The Paris system likely emerged in East Africa and quickly spread in the Indian Ocean
432 Island and Sub-Saharan Africa (Bastide et al. 2011). During the last two decades we observed
433 the invasion of the system in North Africa, Europe and the Middle East (Bastide et al. 2013;
434 Helleu et al. 2019). Due to their transmission advantage the resistant Y chromosomes were
435 strongly selected and quickly spread along with the driver, replacing the Y sensitive
436 chromosomes (Helleu et al. 2019) (Figure 6D).
437 Our approach allows us to follow the migration path of the Paris SR system. For
438 example, the resistant Y chromosomes from Egypt, Morocco and France segregate in different
439 clusters indicating that the Paris SR system has migrated independently to North Africa and
440 Europe. Similarly, the two Y chromosomes from Seychelles are resistant but belong to two
441 different clusters: Sey81 and Sey03, which have been collected more than 20 years apart. The
442 2003 resistant lineage likely results from migration from continental Africa to the Seychelles
443 island at that time. In addition, Sey03 (cluster II) appears to be more resistant than Sey81
444 (cluster V), which could explain this replacement.
445

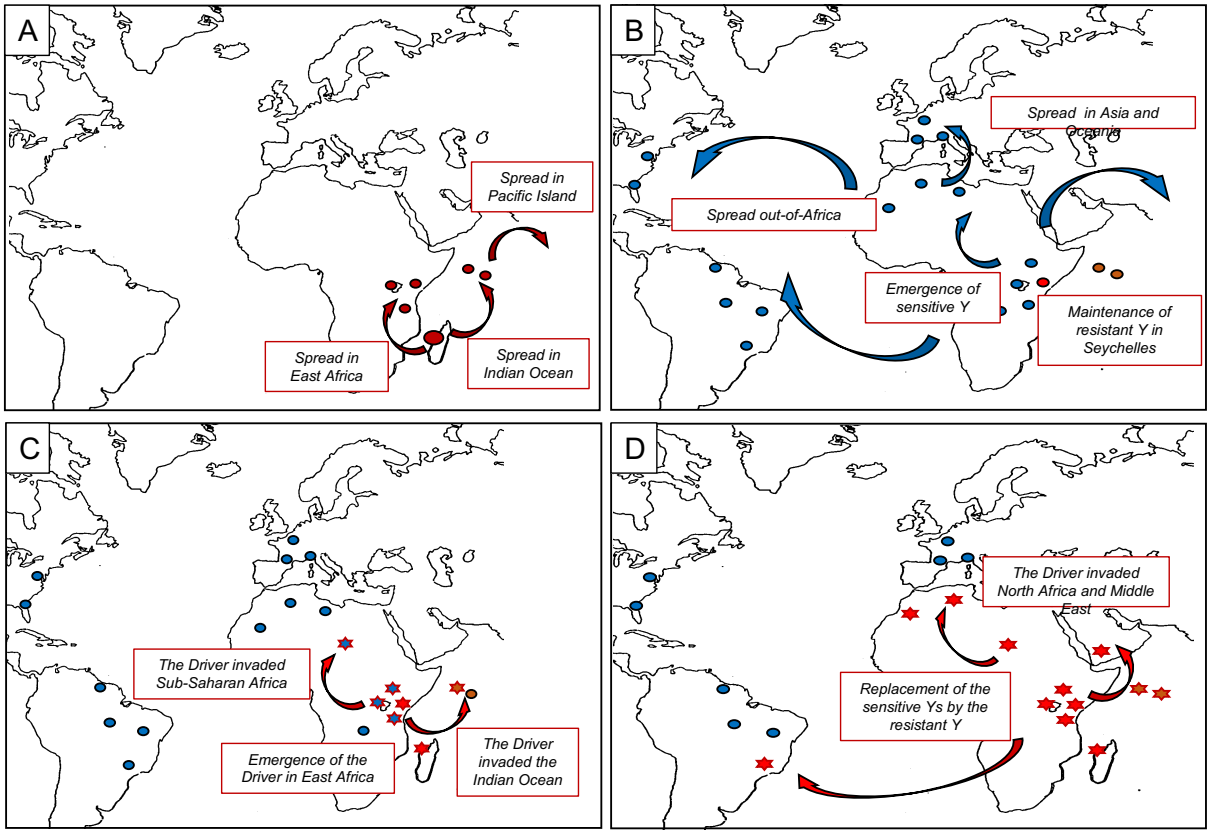

Figure 6. **Proposed scenario for the demographic history of *D. simulans* in relation to the Paris *SR* system**. A. Expansion of *D. simulans* out of its area of origin, toward the Seychelles and the African continent (Lachaise et Silvain, 2004.; Dean et Ballard 2004). The ancestral Y chromosome belongs to the resistant lineage, here represented by a red circle. B. Emergence of the sensitive lineage in subsaharan Africa and diffusion out-of-Africa. Here sensitive Y chromosomes are represented by a blue circle. C. Emergence of the Paris *SR* system in East-Africa (Bastide et al. 2011). Populations where the system is detected are represented by a star. D. Spread of the Paris *SR* system and replacement of the sensitive Y chromosome by resistant Y chromosome (Helleu et al. 2019).

461 ***Genetic basis of the resistance***

462 Beyond the evolutionary history of the Y chromosome, we also provide important
463 insight about the genetic basis of the resistance to Paris X^{SR}. Based on the large phenotypic
464 variation of the resistance we hypothesize that the resistance ability might be associated with
465 variation in the repeat content on the Y chromosome. We did not find any association between
466 the resistance and the gene or transposable element abundance on the Y chromosomes, they
467 even appear quite stable. However we do find a strong association with the simple satellite
468 composition (Figure 5A-B) and identified two strong candidates: (AAACAAT)n and
469 (AATGG)n (Figure 5C-D). Both are already known to be specific to the Y chromosome in *D.*
470 *simulans* (Wei et al. 2018) and exhibit a lower copy number on the resistant Y chromosome.
471 The (AATGG)n shows a strong negative correlation between the copy number and the
472 resistance ability. Y chromosomes with fewer copies appear to be more resistant. We did not
473 observe a similar correlation for the (AAACAAT)n satellite. However, it appears associated
474 with structural variation observed among resistant Y chromosomes (Helleu et al. 2019), which
475 could explain variation in the phenotype. One of the drivers of the Paris *SR* system has been
476 identified as *HP1D2*, a member of the heterochromatin protein gene family, known to be
477 involved in heterochromatin regulation. We have previously shown that HP1D2 is specifically
478 expressed in spermatogonia and is enriched on the Y chromosome (Helleu et al. 2016). In
479 addition, the drive is due to a loss of function of *HP1D2*. We thus hypothesize that the drive is
480 associated with a mis-regulation of the heterochromatin region of the Y chromosome. While
481 we do not know the precise target of HP1D2, those two simple satellites appear as good
482 candidates. Correlations between the target copy number and the strength of the drive have
483 already been shown in other meiotic drive systems. For example, in the *SD* system of *D.*
484 *melanogaster*, *Rsp* satellite is the target of the driver (Hartl 1974). Variation in *Rsp* copy number
485 is related to variation in the sensitivity to the driver (Lyttle 1989). Chromosomes with a higher

486 copy number of *Rsp* are more sensitive to the driver. Similarly, particular repeats may be
487 involved in the resistance to the Paris drivers and possibly be their target. Variation in resistance
488 could be due to variation in their abundance and/or in their localization. Considering that the
489 distortion is due to a misregulation of heterochromatic sequences, the localization of those
490 sequences among the Y chromosome could be more or less critical for chromosome
491 segregation. Also, we cannot exclude that the resistance phenotype is multifactorial, or that the
492 genetic basis of resistance is different between Y chromosome lineages.
493
494 To conclude, our study sheds light on the evolutionary history of the Y chromosome in
495 *D. simulans*. We show that Y resistant chromosomes likely predated the emergence of the Paris
496 *SR* system and that Y sensitive chromosomes have spread through the world in the recent past.
497 The Paris system is particularly young, at least in its present form, since the segmental
498 duplication which is involved in driving likely emerged less than 500 years ago (Bastide et al.
499 2013). While its impact on the recent evolutionary history of the Y chromosome is clearly
500 strong, it cannot explain on its own the whole pattern of variation we observed here. More
501 likely, the Y chromosome was recurrently involved in genetic conflicts, which resulted in
502 multiple selective sweeps. Furthermore, we also provide new insights on the genetic basis of
503 Paris drive resistance. Knowing that simple satellites are likely involved should help us to move
504 further in the understanding of the molecular mechanism underlying the Paris *SR* system.

505 **MATERIALS AND METHODS**

506

507 ***Fly stocks***

508 Except for Ken75, all iso-Y lines (derived from a single male) carry the same standard genetic
509 background (X^{ST} A^{ST}) from the ST8 reference strain (Montchamp-Moreau et Cazemajor 2002).
510 The crossing procedure used to isolate the Y chromosomes is presented in (Montchamp-
511 Moreau, Ginhoux, et Atlan 2001). The fourth chromosome was not controlled. The place and
512 date of collection of the strains from which the Y chromosomes studied originate are given in
513 SupTable5.

514 ST8 is the reference standard strain in the Paris *SR* system. It has been collected in Tunisia
515 (1983) and is free of driver and suppressor elements (Montchamp-Moreau et Cazemajor 2002).
516 The *ST8/C(1)RM,yw* line is derived from the reference standard line ST8. Females carry a
517 compound X chromosome *C(1)RM,yw* and males carry the X^{ST8} from the ST8 line. The Y
518 chromosome and the autosomes are from the ST8 line. `

519 The *ST8;Cy;Delta/Ubx* line carry the sex chromosomes from the ST8 line and dominant
520 markers on chromosomes II (^{Cy}) and III ($^{Ubx/Delta}$).

521 The X^{SR3} line is used to test the resistance ability of the Y chromosome. The males carry an X^{SR}
522 from the reference SR strain (collected in Seychelles in 1981, Cazemajor, Landré, et
523 Montchamp-Moreau 1997) and the females carry the compound X chromosome *C(1)RM,y,w*.
524 The Y chromosome and the autosomes come from the ST8 line.

525 For simplicity each iso-Y line name is coded as follows : the 3 first letters of their location +
526 last two digits of the collection year (for instance Morocco 2012 is Mor12).

527

528 ***Resistance test***

529 Each Y chromosome from the CAA lineage, and the standard Y^{ST8} have been tested for their
530 resistance ability against the same reference X^{SR3}. The crossing procedure is detailed in

531 Montchamp-Moreau, Ginhoux, et Atlan 2001. Briefly, males carrying the Y chromosome to be
532 tested (Y^{pop}) were mass-mated with *ST8/C(1)RM,yw* females. Three F1 females were
533 individually mated with X^{SR3} males, thus producing F2 males carrying both the X^{SR3} and Y^{pop}
534 and standard autosomal background. Five males from each F2 progeny were individually
535 mated, each with 2 ST8 females (so a total of 15 F2 males tested for each Y^{pop}). Only individual
536 F3 progenies consisting of at least 50 flies were considered.

537 Before statistical analyses, the data were transformed in $arcsin(\sqrt{q})$, where q corresponds to the
538 proportion of females in the progeny. To test for within-strain variation in Y resistance, we
539 performed a Kurskal Wallis test between groups of F2 progenies (from the same F1 cross). All
540 the iso-Y lines appear homogeneous (SupTable1). We performed a Wilcoxon test to determine
541 whether those lines were significantly different from ST8. They all appear significantly more
542 resistant than ST8 except Sey81 and New91 (SupTable1). Finally, we performed a Tukey HSD
543 post-hoc test on the anova model to determine groups (Figure 1B). All statistical analyses were
544 conducted in R base (R Core Team 2022). The R script used can be found on github:
545 Figure1/resistance.R.

546

547 ***DNA extraction and sequencing***

548 DNA extraction and sequencing was performed in two batches. In batch1 we sequenced: Tun83,
549 Egy07, Ken75, Zim97, Egy12, Tan96 and Sey81. In batch2 we sequenced: Tun83, Fra93,
550 Aus11, Mor12, Cal61, Guy56, Sey81, May09, Fra09, Mor16, Sao01, Cam16, New91, Haw07,
551 Mad98, Sey03. Tun83 was sequenced in both batches to control for potential technical bias.
552 Sey81 was sequenced in the first batch in its original genomic background (from Seychelles
553 1981) and then the Y chromosome was isolated in the same standard genomic background (from
554 ST8 strain) and re-sequenced in: the second batch. DNA was extracted from 50 adults J1 males

555 flies using the Nucleobond AXG20 column from Macherey-Nagel. For whole genome
556 sequencing, libraries were prepared using Illumina PCRfree kit and sequencing was done using
557 Illumina HiSeq3000 technology with pair-end 150bp reads. The sequencing depth was
558 determined in order to obtain a minimum autosomal coverage of 80x.

559

560 ***Mapping and Variant calling***

561 Paired-end reads were trimmed using trimgalore (*trim_galore --paired --nextera --length 75 --*
562 *phred33 --no_report_file –fastqc*) (Krueger et al. 2021). Read quality was assessed using
563 FASTQC.

564 Reads were mapped against the reference genome with bwa (v7.4) using the *BWA-MEM*
565 algorithm (default parameters). We used the heterochromatin-enriched assembly of *D. simulans*
566 (Chang et al. 2022) that include 15Mb of the Y chromosome as reference genome. Resulting
567 sam alignment files were converted into bam files and sorted using respectively samtools
568 (v1.11) *view* and *sort* command. PCR duplicates were removed using *Markduplicates* from
569 Picardtools (v2.12.0) (https://broadinstitute.github.io/picard/). For our SNP calling analyses we
570 want to keep only the uniquely mapping reads, whereas for the repeat quantification we need
571 to keep also the multi-mapping reads. We thus performed two different filtering based on
572 mapping quality using samtools *view* (Li et al. 2009). To keep multi-mapping reads we use the
573 following parameters *-b -h -f 3 -F 4 -F 8 -F 256 -F 2048*, to keep only the uniquely mapping
574 reads we use the following parameters *-b -h -f 3 -F 4 -F 8 -F 256 -F 2048 -q30.* Mean coverage
575 was determined using the *genomecov* tool of bedtools (Quinlan et Hall 2010). The pipeline used
576 (Mapping.sh) can be found on the github repository.

577 Variant calling was done on the uniquely mapping bam files using *mpileup* and *call* commands
578 from bcftools (v1.14) (Li 2011). We then performed different filtering steps. First isolate the
579 SNPs associated to the Y chromosome using *vcftools*. We use vcftools (Danecek et al. 2011) to

580 analyze the quality of the SNP calling on the Y chromosome based on variant mean depth,
581 variant quality, proportion of missing data per site, mean depth per individual, missing data per
582 individual. This allowed us to determine the threshold for the quality filtering : --remove-indels
583 --max-missing 0.9 --minQ 30 --min-meanDP 10 --max-meanDP 50 --minDP 10 --maxDP 50.
584 Because each iso-Y line originates from a single male and the Y chromosome is supposed to be
585 haploid, heterozygote sites are likely to be duplicated sequences and thus difficult to take into
586 account. We removed the heterozygotes sites using the *filter* command (-e 'GT="het"' option)
587 of bcftools. Note that because Tun83 and Sey81 were sequenced twice, after verification that
588 they were identical, we keep only the data from batch2 to avoid redundancy in the analyses.
589 The pipeline used (VariantCalling.sh) can be found on the github repository.
590 We calculate the nucleotide diversity along Y chromosomes per 10kb windows using the
591 vcftool--*window-pi 10000* option.
592 To investigate the population structure we performed a Principal Component Analysis (PCA)
593 on the filtered vcf, using the software plink (v1.9) (Chang et al. 2015). The pairwise distance
594 between strains was calculated using MEGA (Tamura, Stecher, et Kumar 2021).
595 All the following analyses on the vcf file were done using R scripts (VariantAnalyses.R on
596 github) using the following libraries: *vcfR, SNPrelate, gdsfmt, adegenet, ape, pheatmap,*
597 *maSigPro, stringr*.
598

599 ***Identification of corresponding SNPs in the sister species and phylogeny.***

600 To identify the corresponding allele in *D. melanogaster*, *D. sechellia* and *D. mauritiana* we first
601 needed to identify the corresponding location. To do so, we first created a consensus Y
602 chromosome where SNPs are coding in UPAC code, using bcftools *consensus* command (with
603 --iupac-codes option). We built a bed file containing the position of each SNP. Using bedtools
604 *getfasta* command we extracted 400bp sequences for each SNPs (200bp on each side of the

605 SNPs) in a fasta file. Then each sequence were blasted against the Y chromosome consensus of
606 *D. melanogaster*, *D. sechellia* and *D. mauritiana* using blastn (-evalue 1e-20 -qcov_hsp_perc
607 80 -perc_identity 85 -max_hsps1 -max_target_seqs 1). We output the blast result in a tab
608 delimited format (-outfmt "6 qseqid sseqid evalue length pident qstart qseq sseq") and extract
609 the corresponding allele using a R script.
610 To reconstruct the phylogenetic tree, we first created a fasta sequence with only the SNP for
611 each individual. The SNP for which we were not able to identify a corresponding allele in the
612 sister species were coded as gaps. We constructed the maximum likelihood tree using the GTR
613 model and ultrafast bootstrap (1000 replicates) using iqTREE web server (Nguyen et al. 2015).
614 *D. melanogaster* was used as the outgroup. The scripts used can be found in github :
615 Figure3/Ancestry.sh and Figure3/Ancestry.R.

616

617 ***Transposable element and complex satellite quantification***

618 Due to the batch effect observed we only used the strain from the second batch for the analysis.
619 To estimate the abundance of transposable elements and complex satellites in each iso-Y line
620 we first mapped the read against our reference genome. We applied the same pipeline as for the
621 SNP calling, with the exception that we kept multi-mapping reads (pipeline Mappingt.sh on
622 github). The reference genome was annotated using a custom repeat library (SupFile1.fasta)
623 using repeatmasker (options *-no_is -a -inv -pa 20 -div 20*) (Smit, AFA, Hubley, R, et Green, P.
624 2013). Using htseq-count (Putri et al. 2022)we counted the number of reads that map against
625 each repeats and normalize the reads count in RPM. We removed repeats for none of the strain
626 have a RPM greater than 10. All the subsequent analyses were done using Rstudio. We
627 performed a Wilcoxon test, followed by an “fdr” correction, to determine which repeats have a
628 different enrichment between sensitive and resistant. We calculated the coefficient of variation

629 for each repeat to determine which element should be the most variable, the top 25 is presented
630 in Figure 4D. The scripts used can be found in github : TEcount.sh and Figure4D/TE.R.
631

632 ***kmers quantification***

633 Due to the batch effect observed we only used the strain from the second batch for the analysis.
634 Barring the fourth chromosome (not controlled), only the Y chromosome differs; we thus
635 consider that variations observed between lines are due to the Y chromosome. To estimate the
636 abundance of simple satellites in each strain we used an assembly-free approach based on kmer
637 quantification. We used the software k-Seek which quantified the simple satellite tandemly
638 repeated using unassembled Illumina reads. We normalized the kmers count in RPM using the
639 total mapping reads. We removed the kmers for which none of the strain have an RPM greater
640 than 5. All the subsequent analyses were done using Rstudio. We performed a Wilcoxon test to
641 determine which kmers have a different enrichment between sensitive and resistant. Then we
642 performed a correlation test between the kmers count and the resistance ability for the 11 strains
643 that have been phenotype (Figure 1B, SupTable1). The scripts used can be found in github :
644 kmers.sh and Figure5/Kmers.R.

645

646 ***Gene copy number analysis.***

647 Due to the batch effect observed we only used the strain from the second batch for the analysis.
648 To estimate the exon copy number in each iso-Y line we first mapped the reads against our
649 reference genome. We applied the same pipeline as for the SNP calling, with the exception that
650 we kept multi-mapping reads (pipeline Mappingt.sh on github). We estimated the reads
651 coverage on each exon using *samtools depth*. For the 11 canonical genes we used a bed file
652 containing the coordinate of each exon on the reference genome based on Chang et al 2022. For

653 the two ampliconic genes, we blasted the consensus sequences against the reference genome
654 (option *-perc_identity 80*) to determine the coordinate of each insertion. For each iso-Y line we
655 calculated the mean coverage of all single copy exons and used it as a normalization factor. To
656 estimate the exon copy number we calculated the exon mean coverage divided by the
657 normalization factor and multiplied by the number of corresponding exons annotated on the
658 reference genome. The script used can be found in github : Figure4A/Gene_count.R.
659
660
661 ***Genotyping and Haplotype tree***
662
663 We designed primers around SNPs located on regions annotated as coding regions (Chang et al
664 2022), as they are the most likely to be unique. For each of them we performed a PCR with
665 DreamTaq™ Hot Start Green PCR (ThermoFisherScientific, ref K9021) following
666 manufacturer's instructions (hybridization temperature: 58°). Each PCR product was sequenced
667 using the Sanger method on a SeqStudio Genetic Analyzer (ThermoScientific). Only five
668 primers pairs were able to deliver a unique sequence (SupTable6), containing 11 SNPs in total
669 (SupTable2). We genotyped the sequenced iso-Y lines, to validate our SNP calling, and 86
670 additional iso-Y lines from 22 populations. Similarly, we genotyped the mitochondrial
671 haplotype of the 21 iso-Y lines using 2 pairs of primers in SupTable6.
672 We concatenated the 11 SNPs for each iso-Y strains in a nexus format and built the haplotype
673 tree by the Median Joining Method using POPart (Leigh et Bryant 2015).
674
675 ***Data availability***
676 All sequences are available from NCBI SRA under Bioproject accession PRJNA905841. All
677 the BASH pipelines and R scripts used in this study are available on github
678 https://github.com/CourretC/YchromosomesCourret2023.git

Acknowledgment

This research has been supported by the Centre National de la Recherche Scientifique (UMR9191) and by grants from the Agence Nationale de la Recherche (RESIST ANR-21-CE02-0004-02) and by National Institutes of Health NIH R35GM119515 to AML.

684 **SUPPLEMENTAL FIGURES**

685

686 SupFig1 : Y chromosome haplotypes. Each line corresponds to a different Y chromosome ordered based
687 on a clustering method. Each column corresponds to a SNP, the order of which is based on their physical
688 position on the Y chromosome. The blue alleles correspond to the reference allele and the red alleles
689 correspond to the alternative allele; white are missing data.

690

691 SupFig2 : Haplotype network of 98 Y chromosomes from 26 populations. The haplotypes are composed
692 of 11 SNPs. The name of each Y chromosome is indicated within each cluster (see SupTable2 for full
693 name). The clusters are colored according to the 4 clusters identified for the 21 iso-Y lines sequenced.

694

695 SupFig3 : Correlation plots between the satellite abundance (y axis) and the resistance ability (x-axis).
696 Correlation was calculated based on a Spearman correlation test

697

698

Reference

Agudo, M., A. Losada, J. P. Abad, S. Pimpinelli, P. Ripoll, et A. Villasante. 1999. « Centromeres from Telomeres? The Centromeric Region of the Y Chromosome of Drosophila Melanogaster Contains a Tandem Array of Telomeric HeT-A- and TART-Related Sequences ». *Nucleic Acids Research* 27 (16): 3318-24. https://doi.org/10.1093/nar/27.16.3318.

Bachtrog, Doris. 2013. « Y-Chromosome Evolution: Emerging Insights into Processes of Y-Chromosome Degeneration ». *Nature Reviews Genetics* 14 (2): 113-24. https://doi.org/10.1038/nrg3366.

Bastide, Héloïse, Michel Cazemajor, David Ogereau, Nicolas Derome, Frédéric Hospital, et Catherine Montchamp-Moreau. 2011. « Rapid Rise and Fall of Selfish Sex-Ratio X Chromosomes in Drosophila Simulans: Spatiotemporal Analysis of Phenotypic and Molecular Data ». *Molecular Biology and Evolution* 28 (9): 2461-70. https://doi.org/10.1093/molbev/msr074.

Bastide, Héloïse, Pierre R. Gérard, David Ogereau, Michel Cazemajor, et Catherine Montchamp-Moreau. 2013. « Local Dynamics of a Fast-Evolving *Sex-Ratio* System in *Drosophila Simulans* ». *Molecular Ecology* 22 (21): 5352-67. https://doi.org/10.1111/mec.12492.

Bernardo Carvalho, A., Leonardo B. Koerich, et Andrew G. Clark. 2009. « Origin and Evolution of Y Chromosomes: Drosophila Tales ». *Trends in Genetics* 25 (6): 270-77. https://doi.org/10.1016/j.tig.2009.04.002.

Boulétreau-Merle, Josselyne, Pierre Fouillet, et Julien Varaldi. 2003. « Divergent Strategies in Low Temperature Environment for the Sibling Species Drosophila Melanogaster and D. Simulans: Overwintering in Extension Border Areas of France and Comparison with African Populations ». *Evolutionary Ecology* 17 (5-6): 523-48. https://doi.org/10.1023/B:EVEC.0000005632.21186.21.

Branco, A T, Y Tao, D L Hartl, et B Lemos. 2013. « Natural Variation of the Y Chromosome Suppresses Sex Ratio Distortion and Modulates Testis-Specific Gene Expression in Drosophila Simulans ». *Heredity* 111 (1): 8-15. https://doi.org/10.1038/hdy.2013.5.

Broad Institute. 2019. « Picard Toolkit », 2019, Broad Institute, GitHub repository édition. https://broadinstitute.github.io/picard/.

Brown, Emily J, Alison H Nguyen, et Doris Bachtrog. 2020. « The Drosophila Y Chromosome Affects Heterochromatin Integrity Genome-Wide ». Édité par John Parsch. *Molecular Biology and Evolution* 37 (10): 2808-24. https://doi.org/10.1093/molbev/msaa082.

Cazemajor, Michel, Claudie Landré, et Catherine Montchamp-Moreau. 1997. « The *Sex-Ratio* Trait in *Drosophila Simulans:* Genetic Analysis of Distortion and Suppression ». *Genetics* 147 (2): 635-42. https://doi.org/10.1093/genetics/147.2.635.

Chang, Ching-Ho, Lauren E Gregory, Kathleen E Gordon, Colin D Meiklejohn, et Amanda M Larracuente. 2022. « Unique Structure and Positive Selection Promote the Rapid Divergence of Drosophila Y Chromosomes ». *ELife* 11 (janvier): e75795. https://doi.org/10.7554/eLife.75795.

Chang, Ching-Ho, et Amanda M. Larracuente. 2019. « Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the *Drosophila Melanogaster* Y Chromosome ». *Genetics* 211 (1): 333-48. https://doi.org/10.1534/genetics.118.301765.

Chang, Christopher C, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, et James J Lee. 2015. « Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets ». *GigaScience* 4 (1): 7. https://doi.org/10.1186/s13742-015-0047-8.

Charlesworth, B., J. A. Coyne, et N. H. Barton. 1987. « The Relative Rates of Evolution of Sex Chromosomes and Autosomes ». *The American Naturalist* 130 (1): 113-46. https://doi.org/10.1086/284701.

Clark, A G. 1990. « Two Tests of Y Chromosomal Variation in Male Fertility of Drosophila Melanogaster. » *Genetics* 125 (3): 527-34. https://doi.org/10.1093/genetics/125.3.527.

Clark, Andrew G. 1987. « Variation in *Y* Chromosome Segregation in Natural Populations of *Drosophila Melanogaster* ». *Genetics* 115 (1): 143-51. https://doi.org/10.1093/genetics/115.1.143.

Courret, Cécile, Ching-Ho Chang, Kevin H.-C. Wei, Catherine Montchamp-Moreau, et Amanda M. Larracuente. 2019. « Meiotic Drive Mechanisms: Lessons from *Drosophila* ». *Proceedings of the Royal Society B: Biological Sciences* 286 (1913): 20191430. https://doi.org/10.1098/rspb.2019.1430.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. « The Variant Call Format and VCFtools ». *Bioinformatics* 27 (15): 2156-58. https://doi.org/10.1093/bioinformatics/btr330.

Dean, Matthew D., et J.William O. Ballard. 2004. « Linking Phylogenetics with Population Genetics to Reconstruct the Geographic Origin of a Species ». *Molecular Phylogenetics and Evolution* 32 (3): 998-1009. https://doi.org/10.1016/j.ympev.2004.03.013.

Fouvry, Lucie, David Ogereau, Anne Berger, Frederick Gavory, et Catherine Montchamp-Moreau. 2011. « Sequence Analysis of the Segmental Duplication Responsible for Paris *Sex-Ratio* Drive in *Drosophila Simulans* ». *G3 Genes|Genomes|Genetics* 1 (5): 401-10. https://doi.org/10.1534/g3.111.000315.

Hall, David W. 2004. « Meiotic drive and sex chromosome cycling ». *Evolution* 58 (5): 925-31. https://doi.org/10.1111/j.0014-3820.2004.tb00426.x.

Hartl, Daniel L. 1974. «Genetic dissection of segregation distortion. i. suicide combinations of *sd* genes ». *Genetics* 76 (3): 477-86. https://doi.org/10.1093/genetics/76.3.477.

Helleu, Quentin, Cécile Courret, David Ogereau, Katie L Burnham, Nicole Chaminade, Mohamed Chakir, Sylvie Aulard, et Catherine Montchamp-Moreau. 2019. « Sex-Ratio Meiotic Drive Shapes the Evolution of the Y Chromosome in Drosophila Simulans ». *Molecular Biology and Evolution* 36 (12): 2668-81. https://doi.org/10.1093/molbev/msz160.

Helleu, Quentin, Pierre R. Gérard, Raphaëlle Dubruille, David Ogereau, Benjamin Prud'homme, Benjamin Loppin, et Catherine Montchamp-Moreau. 2016. « Rapid Evolution of a Y-Chromosome Heterochromatin Protein Underlies Sex Chromosome Meiotic Drive ». *Proceedings of the National Academy of Sciences* 113 (15): 4110-15. https://doi.org/10.1073/pnas.1519332113.

Helleu, Quentin, Pierre R. Gérard, et Catherine Montchamp-Moreau. 2015. « Sex Chromosome Drive ». *Cold Spring Harbor Perspectives in Biology* 7 (2): a017616. https://doi.org/10.1101/cshperspect.a017616.

Hill, Tom, Christian Schlötterer, et Andrea J. Betancourt. 2016. « Hybrid Dysgenesis in Drosophila Simulans Associated with a Rapid Invasion of the P-Element ». Édité par Harmit S. Malik. *PLOS Genetics* 12 (3): e1005920. https://doi.org/10.1371/journal.pgen.1005920.

Hoskins, Roger A, Christopher D Smith, Joseph W Carlson, A Bernardo Carvalho, Aaron Halpern, Joshua S Kaminker, Cameron Kennedy, et al. 2002. « Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. » *Genome Biology* 3 (12): research0085.1. https://doi.org/10.1186/gb-2002-3-12-research0085.

Huttunen, Susanna, et Jouni Aspi. s. d. « Complex Inheritance of Male Courtship Song Characters in Drosophila Virilis », 8.

Kennison, James A. 1981. « The genetic and cytological organization of the *y* chromosome of *drosophila melanogaster* ». *Genetics* 98 (3): 529-48. https://doi.org/10.1093/genetics/98.3.529.

Kopp, Artyom, Amanda Frank, et Jeffrey Fu. 2006. « Historical Biogeography of Drosophila Simulans Based on Y-Chromosomal Sequences ». *Molecular Phylogenetics and Evolution* 38 (2): 355-62. https://doi.org/10.1016/j.ympev.2005.06.006.

Krueger, Felix, Frankie James, Phil Ewels, Ebrahim Afyounian, et Benjamin Schuster-Boeckler. 2021. « FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo ». Zenodo. https://doi.org/10.5281/ZENODO.5127899.

Lachaise, Daniel, et Jean-François Silvain. s. d. « How Two Afrotropical Endemics Made Two Cosmopolitan Human Commensals: The Drosophila Melanogaster–D. Simulans Palaeogeographic Riddle », 23.

Larracuente, Amanda M, et Andrew G Clark. 2013. « Surprising Differences in the Variability of Y Chromosomes in African and Cosmopolitan Populations of *Drosophila Melanogaster* ». *Genetics* 193 (1): 201-14. https://doi.org/10.1534/genetics.112.146167.

Leigh, Jessica W., et David Bryant. 2015. « POPART : Full-feature Software for Haplotype Network Construction ». Édité par Shinichi Nakagawa. *Methods in Ecology and Evolution* 6 (9): 1110-16. https://doi.org/10.1111/2041-210X.12410.

Li, H. 2011. « A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data ». *Bioinformatics* 27 (21): 2987-93. https://doi.org/10.1093/bioinformatics/btr509.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, et 1000 Genome Project Data Processing Subgroup. 2009. « The Sequence Alignment/Map Format and SAMtools ». *Bioinformatics* 25 (16): 2078-79. https://doi.org/10.1093/bioinformatics/btp352.

Lima, Leonardo G de, Stacey L Hanlon, et Jennifer L Gerton. 2020. « Origins and Evolutionary Patterns of the *1.688* Satellite DNA Family in *Drosophila* Phylogeny ». *G3 Genes|Genomes|Genetics* 10 (11): 4129-46. https://doi.org/10.1534/g3.120.401727.

Lyttle, T W. 1989. « The Effect of Novel Chromosome Position and Variable Dose on the Genetic Behavior of the Responder (Rsp) Element of the Segregation Distorter (SD) System of Drosophila Melanogaster. » *Genetics* 121 (4): 751-63. https://doi.org/10.1093/genetics/121.4.751.

Machado, Heather E., Alan O. Bergland, Katherine R. O'Brien, Emily L. Behrman, Paul S. Schmidt, et Dmitri A. Petrov. 2016. « Comparative Population Genomics of Latitudinal Variation in *Drosophila Simulans* and *Drosophila Melanogaster* ». *Molecular Ecology* 25 (3): 723-40. https://doi.org/10.1111/mec.13446.

Meiklejohn, Colin D, Emily L Landeen, Kathleen E Gordon, Thomas Rzatkiewicz, Sarah B Kingan, Anthony J Geneva, Jeffrey P Vedanayagam, et al. 2018. « Gene Flow Mediates the Role of Sex Chromosome Meiotic Drive during Complex Speciation ». *ELife* 7 (décembre): e35468. https://doi.org/10.7554/eLife.35468.

Montchamp-Moreau, Catherine, et Michel Cazemajor. 2002. « *Sex-Ratio* Drive in *Drosophila Simulans* : Variation in Segregation Ratio of *X* Chromosomes From a Natural Population ». *Genetics* 162 (3): 1221-31. https://doi.org/10.1093/genetics/162.3.1221.

Montchamp-Moreau, Catherine, Valérie Ginhoux, et Anne Atlan. 2001. « The y chromosomes of drosophila simulans are highly polymorphic for their ability to suppress sex-ratio drive ». *Evolution* 55 (4): 728. https://doi.org/10.1554/0014-3820(2001)055[0728:TYCODS]2.0.CO;2.

Muirhead, Christina A., et Daven C. Presgraves. 2021. « Satellite DNA-Mediated Diversification of a Sex-Ratio Meiotic Drive Gene Family in Drosophila ». *Nature Ecology & Evolution* 5 (12): 1604-12. https://doi.org/10.1038/s41559-021-01543-8.

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, et Bui Quang Minh. 2015. « IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies ». *Molecular Biology and Evolution* 32 (1): 268-74. https://doi.org/10.1093/molbev/msu300.

Putri, Givanna H, Simon Anders, Paul Theodor Pyl, John E Pimanda, et Fabio Zanini. 2022. « Analysing High-Throughput Sequencing Data in Python with HTSeq 2.0 ». Édité par Valentina Boeva. *Bioinformatics* 38 (10): 2943-45. https://doi.org/10.1093/bioinformatics/btac166.

Quinlan, Aaron R., et Ira M. Hall. 2010. « BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features ». *Bioinformatics* 26 (6): 841-42. https://doi.org/10.1093/bioinformatics/btq033.

R Core Team. 2022. « R: A language and environment for statistical computing. », 2022. https://www.R-project.org/.

Rice, William R. 2014. « An X-Linked Sex Ratio Distorter in *Drosophila Simulans* That Kills or Incapacitates Both Noncarrier Sperm and Sons ». *G3 Genes|Genomes|Genetics* 4 (10): 1837-48. https://doi.org/10.1534/g3.114.013292.

Rohmer, Céline, Jean R. David, Brigitte Moreteau, et Dominique Joly. 2004. « Heat Induced Male Sterility in *Drosophila Melanogaster* : Adaptive Genetic Variations among Geographic Populations and Role of the Y Chromosome ». *Journal of Experimental Biology* 207 (16): 2735-43. https://doi.org/10.1242/jeb.01087.

Smit, AFA, Hubley, R, et Green, P. 2013. « RepeatMasker Open-4.0. », 2015 2013. http://www.repeatmasker.org.

Solignac, Michel, et Monique Monnerot. 1986. « Race formation, speciation, and introgression within *drosophila simulans, d. mauritiana* , and *d. sechellia* inferred from mitochondrial dna analysis ». *Evolution* 40 (3): 531-39. https://doi.org/10.1111/j.1558-5646.1986.tb00505.x.

Sproul, John S, Danielle E Khost, Danna G Eickbush, Sherif Negm, Xiaolu Wei, Isaac Wong, et Amanda M Larracuente. 2020. « Dynamic Evolution of Euchromatic Satellites on the X Chromosome in Drosophila Melanogaster and the Simulans Clade ». Édité par John Parsch. *Molecular Biology and Evolution* 37 (8): 2241-56. https://doi.org/10.1093/molbev/msaa078.

Tamura, Koichiro, Glen Stecher, et Sudhir Kumar. 2021. « MEGA11: Molecular Evolutionary Genetics Analysis Version 11 ». Édité par Fabia Ursula Battistuzzi. *Molecular Biology and Evolution* 38 (7): 3022-27. https://doi.org/10.1093/molbev/msab120.

Tao, Yun, Luciana Araripe, Sarah B Kingan, Yeyan Ke, Hailian Xiao, et Daniel L Hartl. 2007. « A Sex-Ratio Meiotic Drive System in Drosophila Simulans. II: An X-Linked Distorter ». Édité par Daniel Barbash. *PLoS Biology* 5 (11): e293. https://doi.org/10.1371/journal.pbio.0050293.

Vedanayagam, Jeffrey, Ching-Jung Lin, et Eric C. Lai. 2021. « Rapid Evolutionary Dynamics of an Expanding Family of Meiotic Drive Factors and Their HpRNA Suppressors ». *Nature Ecology & Evolution* 5 (12): 1613-23. https://doi.org/10.1038/s41559-021-01592-z.

Wall, Jeffrey D, Peter Andolfatto, et Molly Przeworski. 2002. « Testing Models of Selection and Demography in *Drosophila Simulans* ». *Genetics* 162 (1): 203-16. https://doi.org/10.1093/genetics/162.1.203.

Wei, Kevin H -C, Sarah E Lower, Ian V Caldas, Trevor J S Sless, Daniel A Barbash, et Andrew G Clark. 2018. « Variable Rates of Simple Satellite Gains across the Drosophila

Phylogeny ». *Molecular Biology and Evolution* 35 (4): 925-41. https://doi.org/10.1093/molbev/msy005.

Wei, Kevin H.-C., Jennifer K. Grenier, Daniel A. Barbash, et Andrew G. Clark. 2014. « Correlated Variation and Population Differentiation in Satellite DNA Abundance among Lines of *Drosophila Melanogaster* ». *Proceedings of the National Academy of Sciences* 111 (52): 18793-98. https://doi.org/10.1073/pnas.1421951112.

Zurovcova, Martina, et Walter F Eanes. 1999. « Lack of Nucleotide Polymorphism in the Y-Linked Sperm Flagellar Dynein Gene Dhc-Yh3 of Drosophila Melanogaster and D. Simulans ». *Genetics* 153 (4): 1709-15. https://doi.org/10.1093/genetics/153.4.1709.