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 multiple simple satellites 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 thermotolerance (Huttunen et Aspi, 2003.; Rohmer et al. 2004). In fact, variation in Y-linked
49 heterochromatin appears able to modulate gene expression across the whole genome by
50 modifying the chromatin landscape (Brown, Nguyen, et Bachtrog 2020; Alan T. Branco, Brito,
51 et Lemos 2017; Lemos et al. 2014; Francisco et Lemos 2014). Similar effects are likely in *D.*
52 *simulans* (Branco et al. 2013).

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

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

64 fixation of the carrier Y chromosome in the population (Charlesworth, Coyne, and Barton
65 1987), while a deleterious mutation can lead to its loss. Population genetics models thus suggest
66 that neutral variations within populations should be rarely maintained whereas inter-population
67 differences can rapidly accumulate (Clark 1987). Thus, the *Drosophila* Y chromosome is
68 expected to be particularly sensitive both to the demographic history of populations and to
69 selection events. Indeed, in *D. melanogaster*, the ancestral African populations show
70 significantly less Y-linked variation than the cosmopolitan ones, which could be explained by
71 the demographic history of the species and recent natural selection in Africa (Larracuente et
72 Clark 2013).
73 Y chromosome evolution can also be shaped by intragenomic conflict triggered by X-linked
74 meiotic drivers. These favor the transmission of X chromosomes bearing the driver (hereafter
75 X^{SR}), thus, increasing the proportion of females in progeny and inducing a deviation from the
76 balanced - also called Fisherian - sex ratio. *Sex Ratio* (*SR*) systems are widespread in the
77 *Drosophila* genus, with 19 *SR* systems known in 16 species, among which 3 systems have been
78 described in *Drosophila simulans* (Courret et al. 2019). During spermatogenesis, an X^{SR}
79 typically prevents the formation of functional Y bearing sperm, resulting in a strong female bias
80 in the progeny of carrier males. Thus, the effective size of the Y chromosome is drastically
81 reduced in populations invaded by X^{SR}. Following the spread of X^{SR}, natural selection strongly
82 favors any variant that can counteract the effect of the driver and restore a balanced sex ratio
83 (reviewed in Helleu, Gérard, et Montchamp-Moreau 2015). Such drive suppressors are
84 expected to evolve at unlinked loci, *i.e.*, on autosomes and the Y chromosome. Resistant Y
85 chromosomes have direct and strong selective advantage and thus are expected to quickly reach
86 fixation (Clark 1987; Hall 2004).
87 *Drosophila simulans* is particularly well suited to investigate the impact of SR drive on the
88 evolutionary history of the Y chromosome. In this worldwide species, autosomal suppressors

89 and resistant Y chromosomes have evolved in response to the spread of X-linked drivers known
90 as the Paris system (Cazemajor, Landré, et Montchamp-Moreau 1997). The Paris system likely
91 emerged less than 500 years ago in East Africa (Fouvry et al. 2011; Bastide et al. 2011). It then
92 spread to islands of the Indian Ocean and in Sub-Saharan Africa (Bastide et al. 2013). Drive
93 suppression is complete over its geographic range and the frequency of X^{SR} was found to
94 decrease in some locations, like in Madagascar, Kenya, or Mayotte (Bastide et al. 2011). During
95 the last decade, the invasion of X^{SR} was observed in North African populations, where sensitive
96 Y chromosomes were then replaced by resistant ones within a few years (Helleu et al. 2019).
97 On the whole, in *D. simulans*, the Y chromosomes appeared to be highly polymorphic for their
98 resistance ability with a continuum of phenotypes, from complete resistance to high sensitivity
99 (Montchamp-Moreau, Ginhoux, et Atlan 2001). In contrast, they exhibit a very low level of
100 gene sequence variation, with only three single nucleotide polymorphisms (SNPs) detected
101 along a total of 13 kb of gene fragments; those define three haplotypes - TTA, TTT, CAA - the
102 latter of which is associated with resistance to Paris *SR* drive (Helleu et al. 2019).This low
103 number of SNPs strongly suggests that the phenotypic variation results from variation in
104 repeated sequences and/or from structural variations, both of which are extensive on the Y
105 chromosomes of *D. simulans* (Helleu et al. 2019; Chang et al. 2022). Surprisingly, the CAA
106 haplotype appeared to be ancestral in the *simulans* clade (Helleu et al. 2019), which suggests
107 that resistant Y chromosomes have long predated the emergence of the Paris driver (Helleu et
108 al. 2019). Consistently with this scenario, the Y chromosome of its sibling, *D. sechellia,* is
109 resistant to Paris X^{SR} (Tao et al; 2007).
110
111 Thanks to the recent advance in long read sequencing, about one third of the Y chromosomes
112 of *D. melanogaster* and *D. simulans* (13 Mb out of 40), are now assembled (Chang et
113 Larracuente 2019; Chang et al. 2022). Here we take advantage of those assemblies to further

114 investigate the molecular variation of the Y chromosome in *D. simulans* in relation to the
115 evolution of the Paris *SR* system. We sequenced 21 iso-Y lines (each with a Y chromosome
116 originating from a single male) from different locations around the world; 8 from the sensitive
117 lineage TTA and 13 from the resistant lineage CAA (Figure 1A). Despite their very different
118 geographical origins all sensitive Y chromosomes appear to be highly similar. The resistant Y
119 chromosomes are more divergent and segregate in four distinct groups. By reconstructing the
120 phylogeny of the Y chromosome, we confirmed that the resistant lineage predates the
121 emergence of the Paris X^{SR}. We also examined the variation in repeat content among Y
122 chromosomes. We observed almost no variations in gene and TE copy number, neither of them
123 showing an association with the resistant phenotype. By contrast, extensive variation was found
124 among simple satellites, and we identified a set of satellites which presence and quantity are
125 associated with resistance to the SR system. Altogether this new approach allowed us to better
126 understand the evolutionary history of the Y chromosome in relation with *SR* systems and also
127 yielded new insights on the genetic basis of the resistance.
128
129
130

131 **RESULTS**

132

133 ***1. Resistance ability of the Y chromosomes from the resistant CAA lineage***

134 Among the 21 *D. simulans* Y chromosomes included in the present study, 13 belong to
135 the resistant CAA lineage and were checked for their resistance ability. To do so we produced
136 males carrying the Y chromosome to be tested, a reference X^{SR} chromosome, and suppressor-
137 free autosomes from the reference ST8 line (crossing scheme in Montchamp-Moreau, Ginhoux,
138 and Atlan 2001). For each Y, the individual progeny of 15 males was examined (Figure 1B).
139 We observed significant variation in the resistance ability among the CAA Y chromosomes
140 (ANOVA test, F=12.36, df=13, p<2e-16). The mean sex-ratio in the progeny of carrier males
141 ranged from 60% to 83% depending on the Y chromosome, compared to 89% for the sensitive
142 Y^{ST8} (Figure 1B, Supplementary Table1). Except for Sey81 (Seychelles 1981) and New91 (New
143 Caledonia 1991), which exhibit a low resistance ability with a mean sex-ratio of 81% and 83%,
144 respectively, all CAA Y chromosomes appear significantly more resistant compared to Y^{ST8},
145 our reference sensitive Y chromosome (Figure 1B, Supplementary Table1).

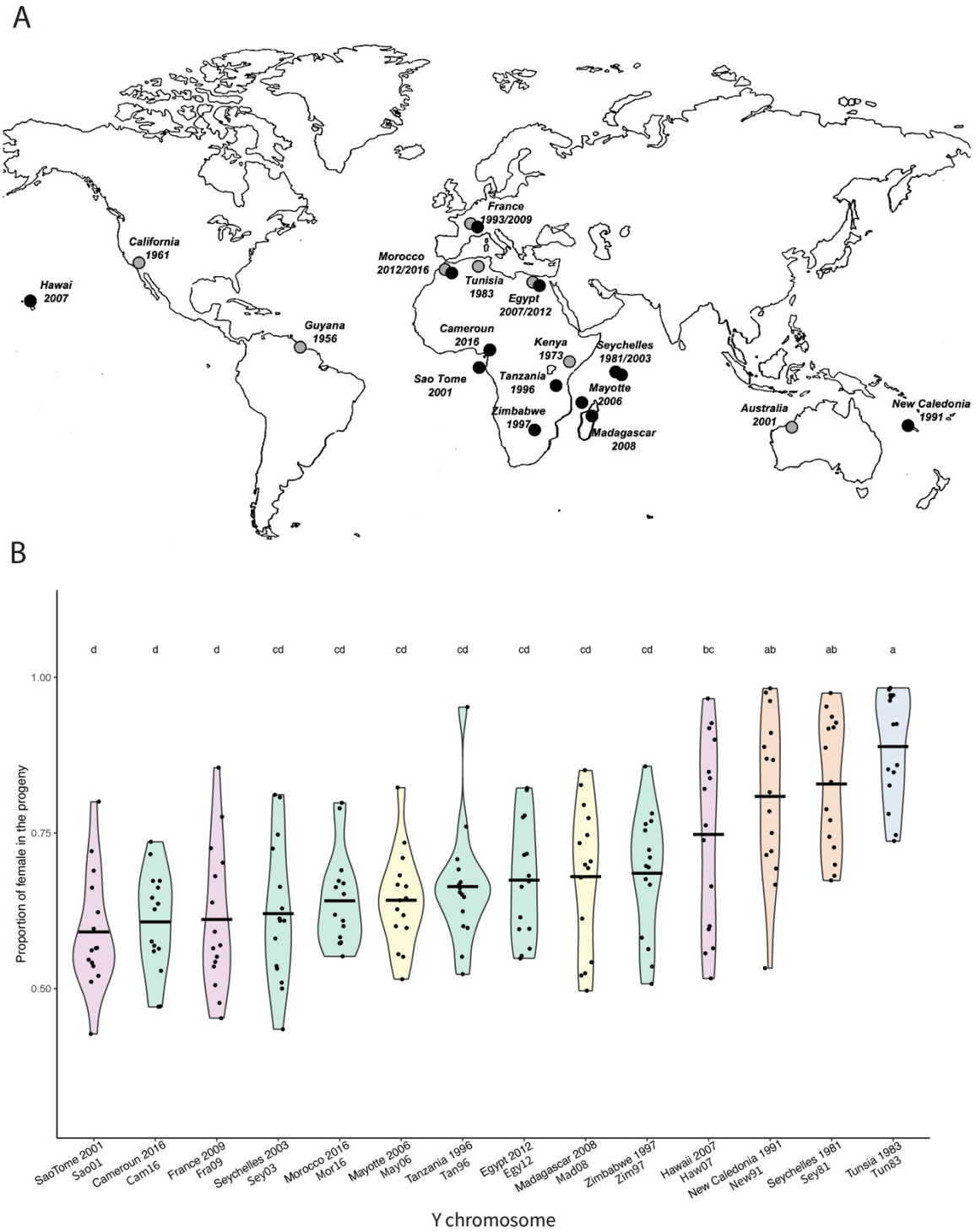


Figure 1. A. Geographical location and date of collection of the Y chromosomes studied (●) Lines carrying a Y chromosome with the CAA haplotype (resistant). (●) Lines carrying a Y chromosome with the TTA haplotype (sensitive). B. Test of the resistance ability of the Y chromosomes. Each violin plot represents 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. Violin plots are colored according to the clusters identified by PCA (same color code as in Figure 2).

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

155 Using the Y chromosome from the XD1 strain (Chang et al. 2022) as a reference, we
156 identified 8190 Y-linked polymorphic sites among the 21 iso-Y lines sequenced. The global
157 nucleotide diversity is very low (1.808e-04) and most of this diversity comes from the CAA
158 resistant Y chromosomes (Supplementary Figure 1). On average, these share the same allele at
159 76.89% of the polymorphic sites, while on average the sensitive TTA Y chromosomes share
160 the same alleles at 97.77 % of the detected polymorphic sites (Table1). Sensitive Y
161 chromosomes are thus highly similar despite originating from geographical regions far away
162 from each other. They all mainly share the same allele as the reference allele (Supplementary
163 Figure 1). It follows that the Y chromosome of the XD1 strain (from Winters, California, 1995)
164 belongs to the sensitive TTA lineage.

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

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

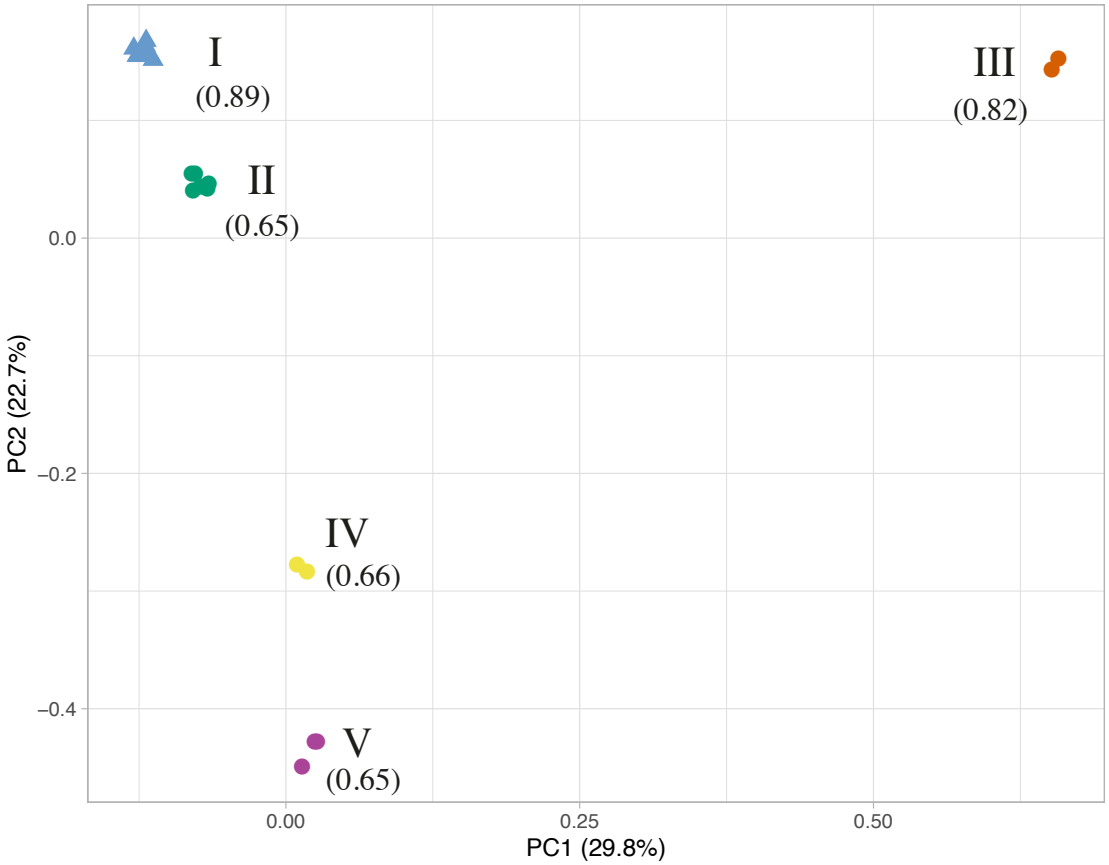


Figure 2. Principal component analysis based on nucleotide diversity among the 21 Y chromosomes (8190 polymorphic sites). The sensitive Y chromosomes are represented by a triangle while the resistant Ys are represented by a circle. Each Y chromosome is colored according to the cluster to which it belongs. Cluster numbers are indicated in roman numerals, with, in parentheses, the average female frequency produced over the Y chromosomes of the corresponding cluster in resistance tests.

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

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

213 To get an insight into the intra-population polymorphism of the Y chromosome, we
214 genotyped by PCR-sequencing 98 additional chromosomes from 22 populations. This first
215 allowed us to validate 11 SNPs detected by our approaches and to look at within-population
216 variation. We reconstructed a haplotype network (Supplementary Figure 2). It recapitulates the
217 clusters identified by the PCA (Figure 2), except for cluster IV and V that grouped together as
218 we do not have SNP distinguishing each cluster. Most of the time all the individuals collected
219 in the same location at the same time carry the same haplotype (Supplementary Table2). Only
220 one population from France (collected in 2014 and 2015) seems to be polymorphic, containing
221 Y chromosomes from clusters I, II and V. We postulate that recurrent annual migrations occur
222 in western Europe, in accordance with the observation that *D. simulans* cannot overwinter in
223 temperate regions (Boulétreau-Merle, Fouillet, and Varaldi 2003; Machado et al. 2016). In
224 addition, migration in *D. simulans* is likely largely human-aided and must have broadly
225 amplified all over the world in the recent past, as suggested by the rapid invasion of the P-
226 transposable element within the last decade (Hill, Schlötterer, et Betancourt 2016).

227

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

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

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

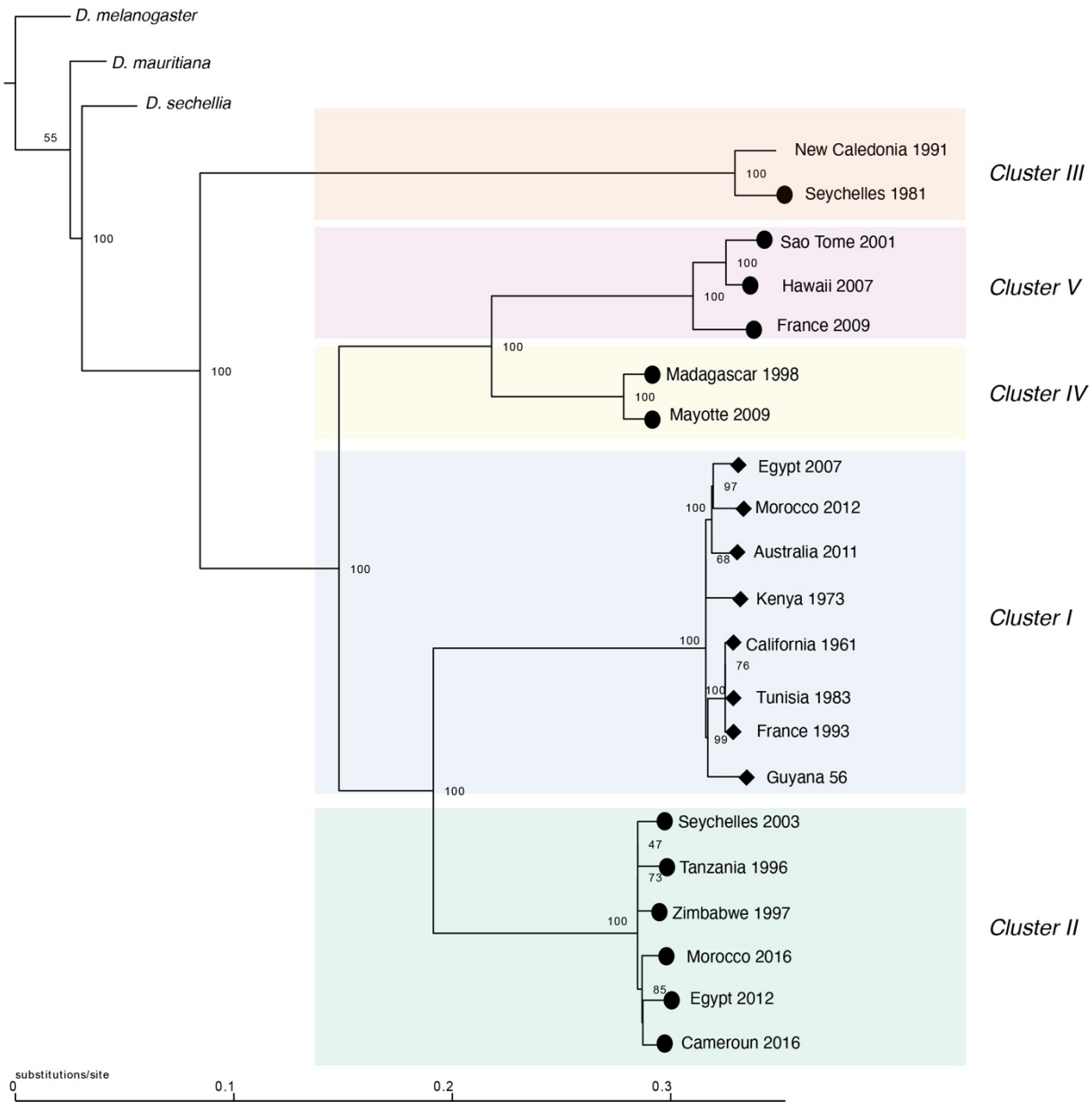


Figure 3. The inferred maximum likelihood phylogeny for the Y chromosome. *D. melanogaster* was used as the outgroup. Node labels indicate ultrafast bootstrap. The resistant Y chromosomes are indicated by a circle at the tip and sensitive Y chromosomes are indicated by diamond at the tip. Clusters are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

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

267 To assess whether some components of the Y chromosomes are associated with the
268 resistance phenotype, we first estimated the exon copy number for each of the 11 canonical Y-
269 linked genes (Figure 4A). Our estimates are in agreement with the previous estimates from the
270 reference genome (Chang et al. 2022). Most of the canonical genes are in single copy and do
271 not exhibit variation between Y chromosomes (Figure 4A). ARY was previously described as
272 a multi copy gene with 4 copies on the reference genome; here we estimated between 4 and 6
273 copies of ARY among our 21 Y chromosomes. Similarly, *kl2* exons 8 to 12 are known to be
274 duplicated, with more than 15 copies for exon 10 for example. They are also the most variable
275 among the 21 Y chromosomes, with *kl2*-exon10 showing between 10 and 17 copies. *kl3*-exon16
276 shows an interesting variation: while most of the Y chromosomes have 3 copies, the 3 Y
277 chromosomes that belong to Cluster V (Haw07, Sao01, Fra09) show a high duplication rate
278 (between 7 and 11 copies) (Figure 4A).

279 In addition to the 11 canonical genes, two ampliconic genes have been described on the Y
280 chromosome of *D. simulans* : *CK2βtes-Y* and *Lhk* (Chang et al. 2022). Accordingly, both exhibit
281 variation in exon copy number among our 21 Y chromosomes. We estimate between 83 and
282 170 copies of *CK2βtes-Y*, between 91 and 156 copies of *Lhk*-exon1 and between 56 and 113
283 copies of *Lhk*-exon2, depending on the Y (Figure 4B).

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

288

289

290

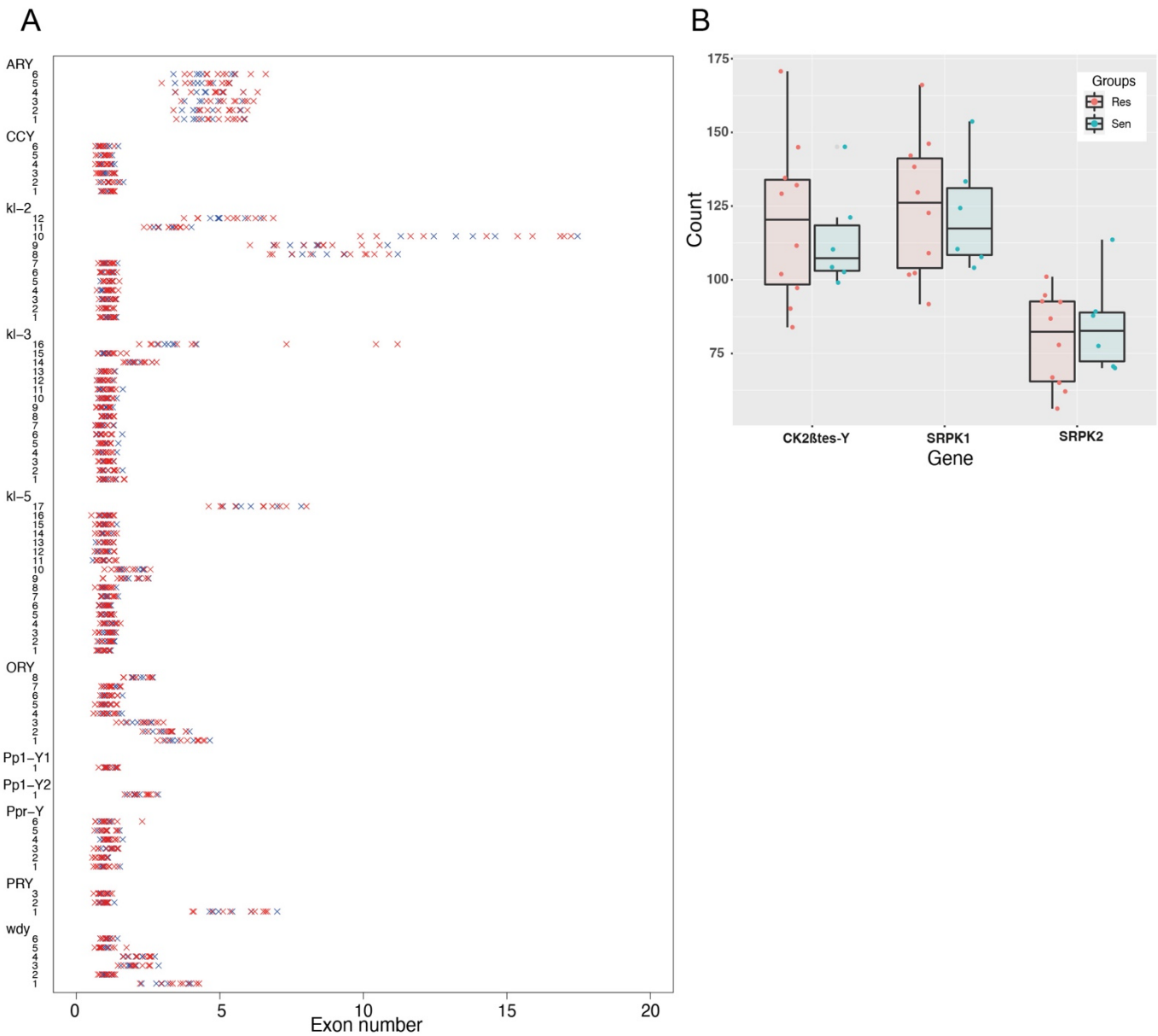


Figure 4: A. Strip chart representing the exon numbers for each Y-associated gene. Each line corresponds to a specific exon of a specific gene. Red crosses correspond to resistant Y chromosomes and blue crosses to the sensitive Y chromosomes. B. Boxplot representing the exon numbers for the two Y-linked multicopy gene families, *CK2βtes-Y* and *Lhk*, among resistant (Res) and sensitive (Sen) Y's.

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

301 Using the raw Illumina reads we then estimated the abundance of transposable elements and
302 complex satellites in 16 iso-Y lines (6 sensitive and 10 resistant; see materials and methods).
303 Because our Y chromosomes have been isolated in the same genetic background, we assumed
304 that variation in the abundance of repeats between iso-Y lines is associated with the Y
305 chromosome. To investigate whether variation in Y-linked complex repeat abundance shows
306 the same clustering as Y-linked nucleotide polymorphism, we performed a PCA analysis
307 (Figure 5A). Globally the complex repeat abundance does not replicate the clusters defined by
308 SNPs in Figure 2. The sensitive Y chromosomes (Cluster I) show a diffuse distribution that
309 partially overlaps with some resistant Y chromosomes. Similarly, the distributions of resistant
310 Y chromosomes from clusters III, IV and V largely overlap. Only the Y resistant chromosomes
311 from cluster II fully separates from all the others.

312 Next, we tested whether the phenotypic variations observed among the Y chromosomes were
313 associated with variations in gross repeat content. We thus performed a Wilcoxon test to assess
314 whether some transposable elements or complex satellites might be differentially enriched
315 between sensitive and resistant Y chromosomes. None of these repeats show significant
316 variation (Supplementary Table3), indicating that the abundance in complex repeats is not
317 associated with the resistance phenotype. More generally, it appears that most complex repeats
318 exhibit very low to no variation in abundance between iso-Y lines (Supplementary Table3). We
319 calculated the coefficient of variation for each complex repeat and plotted the abundance for
320 the Top25 most variable ones (Figure 5B). The NTS-240 element is the only complex repeat
321 that shows substantial variation, with more than 10-fold variation in abundance between Y lines
322 (Figure 5B). While lower in abundance, the telomeric transposable elements (Het-A, TART and
323 TAHRE) are also among the most variable repeats (Figure 5B).This is in agreement with the
324 existence of those elements in various regions on the Y chromosome (Agudo et al. 1999).

However overall, the majority of these Top 25 most variable repeats show relatively low variation between iso-Y lines. It is important to note that the abundance represented here is the whole genome abundance. Thus, complex repeats that are not associated with the Y chromosome are expected 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 repeat variations at or below this threshold likely represent background noise (Figure 5B).

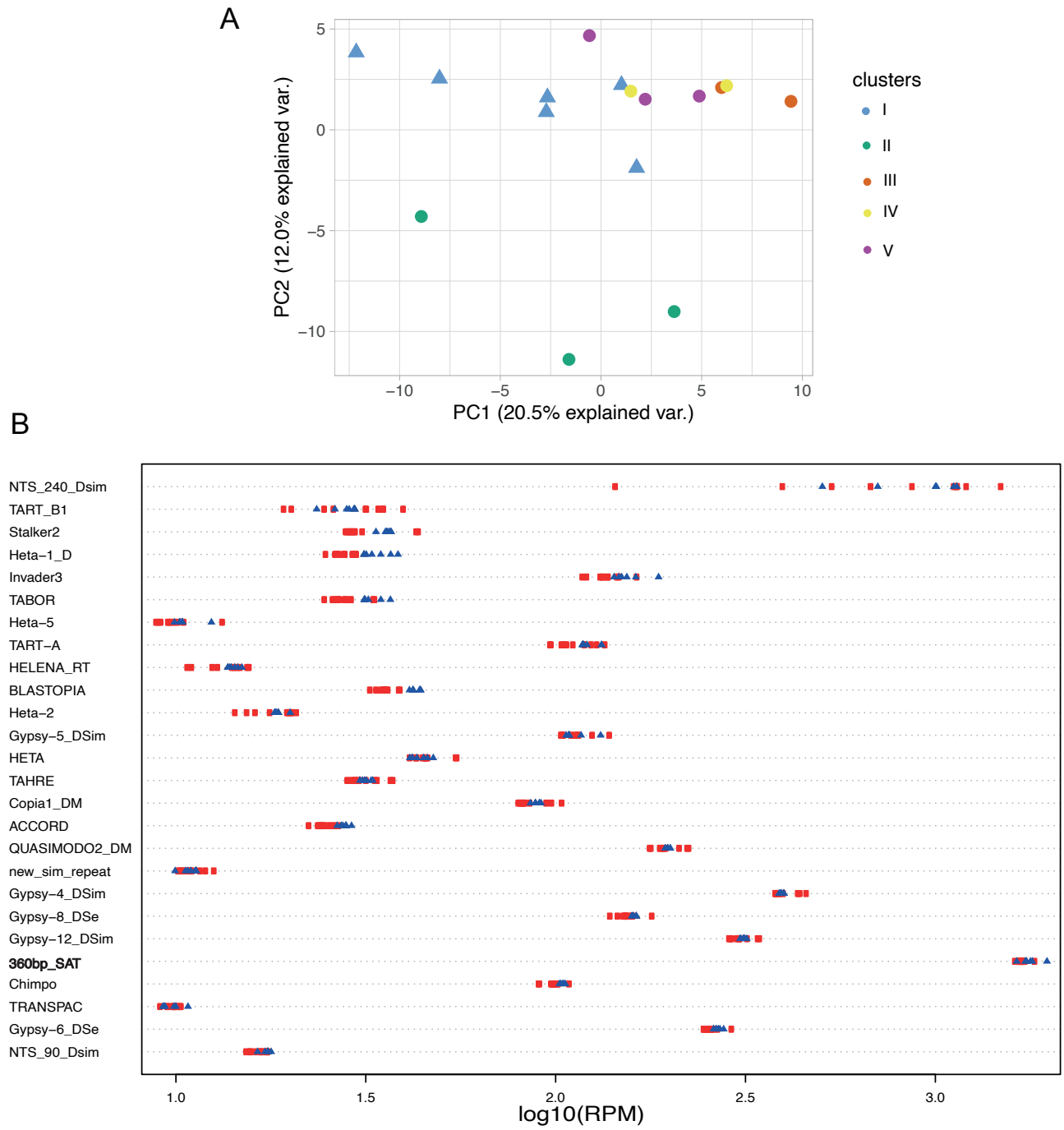


Figure 5: A. Principal Component Analyses (PCA) based on the complex satellite abundance. Sensitive Y chromosomes are represented by a triangle while the resistant Ys are represented by a circle. Each Y chromosome is colored based on the clusters defined previously using the nucleotide diversity (Figure 2). B. Strip chart representing the top 25 most variable complex satellites (based on coefficient of variation), among which the *360 bp* satellite (in bold) is indicative of the background noise. To be able to distinguish the differences between strains, Reads Per Million (RPM) counts have been transformed in log10. Each dot corresponds to a Y chromosome. The Y resistant chromosomes are represented by a red circle while the sensitive Y chromosomes are shown with a blue triangle.

344 ***7. Simple satellite enrichment reveals candidates for the resistance.***

345 Finally, we compared the simple satellite composition between the same 16 iso-Y lines
346 (6 sensitives and 10 resistants). Simple repeats represent over 5% of *D. simulans* genome and
347 are a major component of the Y chromosome (Wei et al. 2018). Using the k-seek pipeline (Wei
348 et al. 2014), we identified 177 simple satellites, including the most abundant kmers previously
349 described to be highly enriched in *D. simulans* (Wei et al. 2018) (Supplementary Table4).
350 Interestingly, we found that simple satellite abundance recapitulates the structure between iso-
351 Y lines defined by the SNP calling (Figure 6A vs Figure 2).

352 Among the 177 satellites, 43 appear differently enriched on resistant versus sensitive Y
353 chromosomes (Wilcoxon test, p.value<0.05) (Figure 6B). About two third of them (28/43) are
354 more enriched among sensitive Y chromosomes (Figure 6B, Supplementary Figure 3). Because
355 we observed variation in resistance among the former group (Figure 1B), we performed a
356 Spearman correlation test between the abundance of each of the 177 satellites and the sex ratio
357 observed in the resistance tests for the 10 resistant Ys plus the reference sensitive Y (Tunisia
358 1983), (Supplementary Figure 4). Based on their abundance and correlation with the phenotype
359 and on their location (enriched or specific to the Y chromosome), we identified a set of potential
360 candidate satellite that may be involved in the resistance phenotype.

361

362

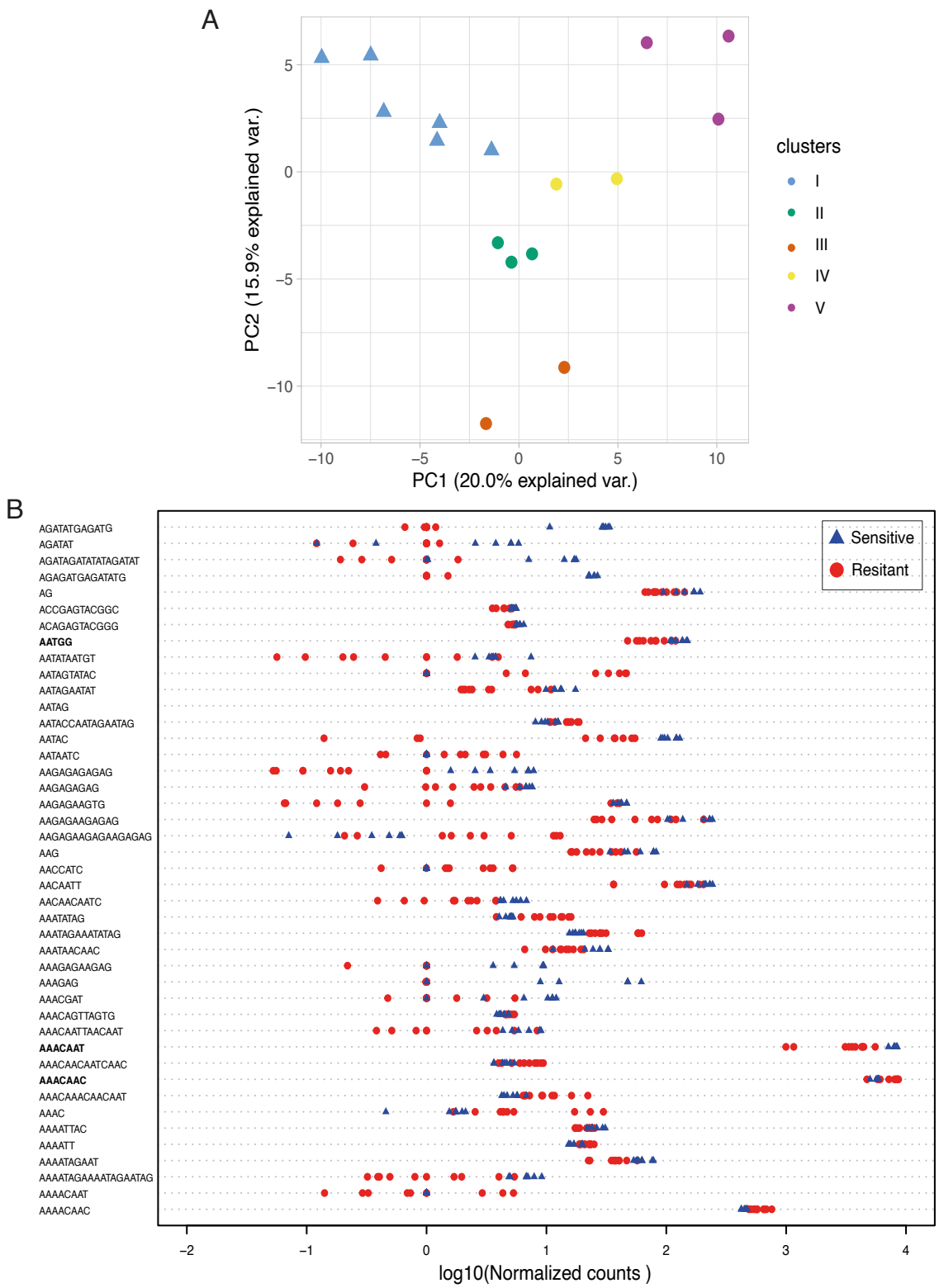


Figure 6. A. Principal component analyses (PCA) based on the simple satellite abundance. Each Y chromosome is colored according to the cluster to which it belongs (defined by nucleotide diversity, Figure 2B). Sensitive Y chromosomes are represented by triangles while the resistant ones are represented by circles. C. Strip chart of the 43 simple satellites differentially enriched (p.value<0.05) between the resistant Y's (red circles) and the sensitive Y's (blue triangles). Normalized counts were transformed in log10.

369 Among the satellites more abundant on sensitive Y chromosomes, four of them, known to be
370 associated to the Y chromosome (Wei et al. 2018), are particularly interesting. The first one,
371 (AAACAAT)n is among the most abundant satellites and is 2-fold more enriched in the
372 sensitive lineages than in the resistant lineages (Figure 7A, Wilcoxon test pvalue<2e-04). The
373 other three, (AATGG)n, (AAGAGAAGAGAG)n and (AACAATT)n, are less abundant, but the
374 sensitive Y chromosomes are also significantly more enriched than the resistant ones (Wilcoxon
375 test, pvalue<0.05). None of these satellites show significant correlation with the strength of the
376 resistance phenotype (Supplementary Figure 4). We performed FISH on mitotic chromosomes
377 for a subset of our iso-Y lines from each cluster, with probes for (AAACAAT)n, (AATGG)n
378 (Figure 7B, Supplementary Figure 5) and (AAGAGAAGAGAG)n and (AACAATT)n
379 (Supplementary Figure 6). We confirmed that (AAACAAT)n, (AATGG)n and (AACAATT)n
380 are specific to the Y chromosomes, at least at the level FISH can detect. With regard to
381 (AAGAGAAGAGAG)n, a signal was observed on the Y chromosomes and also on the X
382 chromosome (originating from the ST8 line, see Mat and Met) (Supplementary Figure 6). We
383 observed a strong signal with AAACAAT for all the Y chromosomes tested, which is consistent
384 with its high copy number. The signal for AATGG was clearly lower as expected from its lower
385 abundance (Figure 7A). While we detected a AATGG signal for all the sensitive Y
386 chromosomes, the signal was low or absent for the resistant Y chromosomes (Figure 7B,
387 Supplementary Figure 5). Once again, this is consistent with our estimates that there are fewer
388 copies on the resistant Y chromosomes. In addition to quantitative variation, we observed
389 structural variations associated with the four satellites (Figure 7B, Supplementary Figure 5-6).
390 In most of the cases they seem located at or close to the tip of the Y chromosomes long arm,
391 while they appear closer to the centromere on the resistant chromosomes from cluster II
392 (Figure7B, Supplementary Figure 5-6).
393

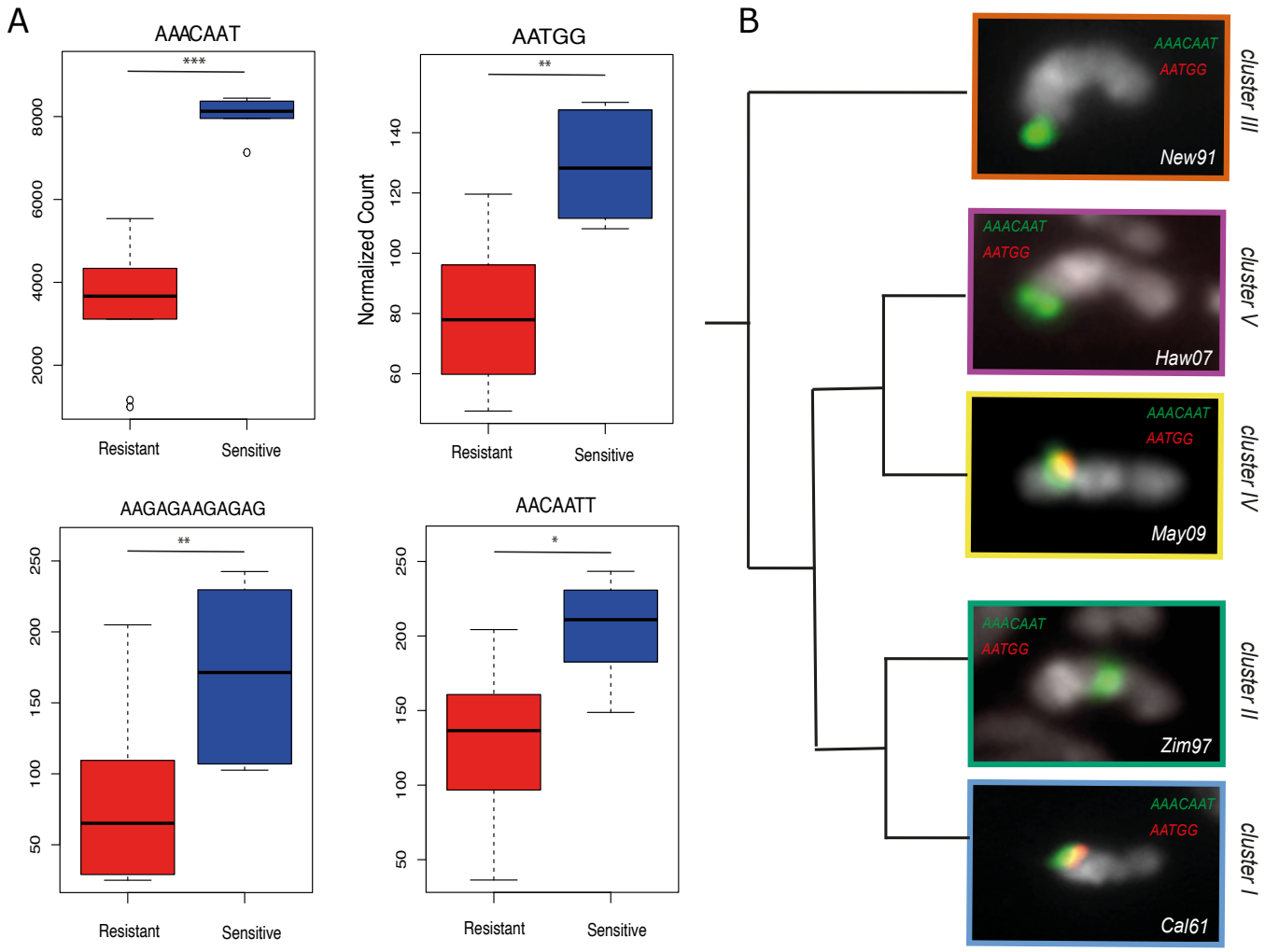
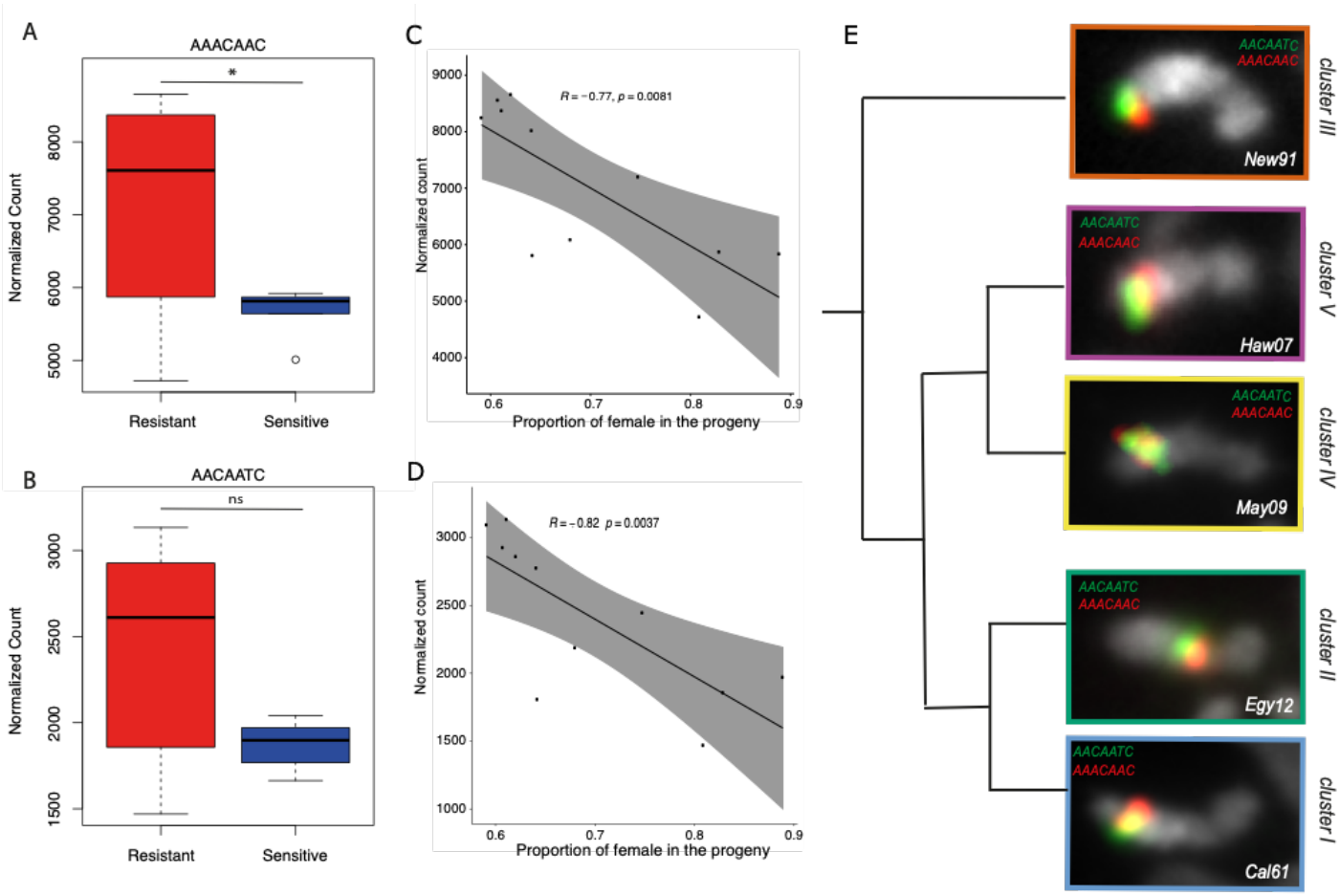


Figure 7. A. The box plots represent the abundance of (AATGG)n, (AAACAAT)n, (AAGAGAAGAGAG)n and (AATGG)n among resistant Ys (red) and sensitive Ys (blue). B. FISH on mitotic chromosomes from larval brains using $(AACAAT)_4$ (green) and $(AATGG)_6$ (red) probes. A single Y chromosome from each cluster is displayed. The whole karyotype FISH on the other iso-Y lines are shown in Supplementary Figure 5.

Finally, two additional satellites showing an opposite pattern came to our attention, (AAACAAC)n and (AACAATC)n. While they show no or low differential enrichment between the sensitive and resistant Y chromosomes (Figure 8A-B), they show a strong positive correlation between abundance and the resistance level (Figure 8C-D, Spearman test,

407 p.value<0.01). The most resistant Y chromosomes are also the most enriched for
408 (AAACAAC)n and (AACAATC)n. Both satellites are known to be associated with the Y
409 chromosome (Wei et al 2018), (AACAATC)n is even Y-specific. Using FISH on mitotic
410 chromosomes, we show that the signal of both satellites is specific to the Y chromosome (Figure
411 8E). And seems to have a localization similar to the previous candidates (Figure 7B and 8E).
412 (AAACAAC)n and (AACAATC)n are located close to the tip of the long arm of the Y
413 chromosome except for the Y chromosomes from cluster II, where they appear closer to the
414 centromere (Figure 8E , Supplementary Figure 7).



415

416 Figure 8: Abundance of (AAACAAC)n (A) and (AACAATC)n (B) satellites. The box plots represent
417 their abundance in each group of Y chromosomes (resistant in red and sensitive in blue). C and D. The
418 dot plots show the correlation between the satellite abundance (y axis) and the proportion of females in
419 the progeny (x-axis), inversely proportional to the resistance. Correlation was calculated based on a
420 Spearman correlation test. E. FISH on mitotic chromosomes from larval brains using $(AACAATC)_4$

421 (green) and $(AAACAAC)_4$ (red) probes. A single Y chromosome from each cluster is displayed. The
422 whole karyotype FISH on the other iso-Y lines are shown in Supplementary Figure 7.
423
424 To further validate the accuracy of our abundance estimates by k-seek, we quantified the
425 abundance of AAACAAT, AAACAAC and AACAATC using slot blot. We found a positive
426 correlation (Pearson correlation test , pvalue <0.05) between the abundance inferred by k-seek
427 and the blot quantifications (Supplementary Figure 8).
428

429 **DISCUSSION**

430

431 Y-linked nucleotide diversity in both *Drosophila melanogaster* and *Drosophila*
432 *simulans* is known to be very low compared to the rest of the genome (Larracuente et Clark
433 2013; Helleu et al. 2019). Those previous estimates were done using genic sequences, which
434 limits the estimation to a very small portion of the Y chromosome. Here, taking into account
435 the assembled part of the chromosome (around 13 Mb) we found that the Y-linked nucleotide
436 diversity in *D. simulans* is about 15-fold lower than the estimate in *D. melanogaster* (π was
437 estimated ranging from 0.00266 to 0.01696 in Larracuente et Clark 2013, here $\pi = 0.0001808$).
438 While the level of nucleotide variation among the Y chromosomes examined here is very low,
439 it reflects the evolutionary history of the species.

440

441 ***Evolutionary history of the Y chromosome related to the demographic history of D.***
442 ***simulans***

443

444 The ancestral origin of the Resistance

445 *D. simulans* is thought to originate from Madagascar or the surrounding area (Lachaise
446 et Silvain, 2004.; Dean et Ballard 2004). We hypothesize that this ancestral population was
447 carrying a drive resistant Y chromosome (Figure 6A). Indeed, according to the phylogeny
448 (Figure 3), the sensitive Y chromosomes are derived recently in *D. simulans* from resistant Y
449 chromosomes, suggesting that the ancestral Y chromosome in *D. simulans* was of the resistant
450 type. The ancestry of the resistant lineage does not necessarily mean that the resistant phenotype
451 is ancestral. However, based on our phylogeny it is most parsimonious that the sensitive lineage
452 (Cluster I) corresponds to a secondary loss of resistance to Paris SR. Consistent with this

453 scenario, previous work suggested that the Y chromosome of *D. sechellia* is resistant to the
454 Paris drivers (Tao et al. 2007; Rice 2014).
455
456 Spread of *D.simulans* in Seychelles and Continental Africa.
457 During a first expansion event, *D. simulans* is thought to have independently colonized
458 the Seychelles islands and the African continent (Lachaise et Silvain, 2004) (Figure 6A). Our
459 data support this scenario. Indeed, most of the Y chromosomes from continental Africa belong
460 to Cluster II while the two chromosomes from Seychelles and New Caledonia belong to Cluster
461 III (Figure 2). The phylogeny agrees with the early divergence of these two chromosomes. Not
462 only do they appear genetically different but they also have a different phenotype. Indeed, the
463 Y chromosomes from Seychelles (Sey81) and New Caledonia (New91) appear less resistant
464 than the other resistant Y chromosomes (Figure 1B).
465 This scenario for the demographic history of *D. simulans* was supported by the
466 distribution of the mitochondrial (mt) haplotype. In *Drosophila simulans* there are three
467 different mt haplotypes : *si*I, *si*II and *si*III (Solignac et Monnerot 1986). The *si*II haplotype is
468 found worldwide. In Madagascar and La Reunion, populations appear polymorphic for *si*III and
469 *si*II. The last haplotype, *si*I, is specific to the Indo-Pacific islands (Seychelles, New Caledonia
470 and Hawaii). We controlled for the mitochondrial haplotype of our strain. As expected most of
471 them carry the *si*II haplotype with the exception of the strains from Hawaii, New Caledonia and
472 Seychelles that carry the *si*I haplotype, and Mayotte that carries the *si*III haplotype
473 (Supplementary Table5).
474
475 Emergence and spread of the sensitive lineage
476 The phylogeny indicates that the sensitive Y chromosomes (Cluster I) and resistant Y
477 chromosomes from Continental Africa (Cluster II) have a recent common ancestor (Figure 3).

478 This suggests that the sensitive lineage emerged in Continental Africa and started to replace
479 resistant Y chromosomes (Figure 9B). This hypothesis is supported by the Ken73 Y
480 chromosome that comes from flies collected in Kenya in 1973. The Ken73 Y chromosome
481 belongs to Cluster I and appears similar to the sensitive Y chromosomes found across the world.

482 The Y chromosome that has then spread out-of-Africa is the sensitive Y chromosome,
483 which became the major Y chromosome across the species range (Helleu et al. 2019) (Figure
484 1A, Figure 6C). The low nucleotide diversity (>97% similarity, Table 1) within the sensitive Y
485 lineage suggests that this event probably arose very recently. Indeed it has been shown that the
486 expansion of *D. simulans* is more recent than that of *D. melanogaster*, likely 6500-5000 years
487 in Eurasia (Lachaise et Silvain, 2004) and 500 years in North America (Wall, Andolfatto, et
488 Przeworski 2002).

489 Why has the sensitive Y chromosome previously superseded the ancestral resistant Y
490 chromosome? Multiple non-exclusive factors could have favored the spread of Y sensitive
491 lineage. First, the emergence of an advantageous mutation may have led to a strong positive
492 selection on sensitive Y chromosomes. A selective advantage associated with a Y chromosome
493 can quickly lead to its fixation into the population (Clark 1987). The Y sensitive chromosomes
494 might improve the fitness of their carrier over the resistant Y chromosome, they can increase
495 the fertility or favor the adaptation to a new environment. Second, the demographic history of
496 the species could also explain the low level of polymorphism of the sensitive Y chromosomes
497 (Irvin et al. 1998; Hamblin et Veuille 1999). Indeed, effects such as migration and bottlenecks
498 could have facilitated the spread of the sensitive lineage and reduced its effective population
499 size. Finally, the last hypothesis involves recurrent genetic conflict. In *D. simulans,* there are at
500 least 3 independent *SR* systems: Paris, Winters and Durham. The Winters and Durham SR
501 systems are cryptic and presumably much older than the Paris system (Meiklejohn et al. 2018).
502 Thus, *SR* systems appear common in this species (Muirhead et Presgraves 2021; Vedanayagam,

503 Lin, et Lai 2021). The intense resulting genetic conflicts between X and Y chromosomes have
504 an important impact on sex chromosome evolution. We hypothesize that the presence of an X-
505 linked driver in *D. simulans*, before the emergence of the Paris *SR* system, could have driven
506 the spread of some Y chromosomes belonging to the lineage sensitive to the Paris driver. Some
507 of the Y chromosomes sensitive to the Paris driver could have been resistant to this more ancient
508 driver. A quick replacement of sensitive by resistant Y chromosomes might have occurred in
509 the past but triggered by another driver. This more ancient driver would have favored the spread
510 of the sensitive Y chromosome lineage before becoming cryptic or disappearing. Previous work
511 shows large variation in the resistance ability of the Y chromosomes to the Winters driver
512 (Branco et al. 2013).Testing the same set of Y chromosomes in the three *sex-ratio* systems
513 might help understanding the cause of the spread of the sensitive Y chromosomes.
514
515 Emergence of the Paris system and replacement of the sensitive by the resistant Y
516 lineage.
517 The Paris system likely emerged in East Africa less than 500 years ago and quickly
518 spread in the Indian Ocean Island and Sub-Saharan Africa (Bastide et al. 2011). During the last
519 two decades we observed the invasion of the system in North Africa, Europe and the Middle
520 East (Bastide et al. 2013; Helleu et al. 2019). Due to their transmission advantage the resistant
521 Y chromosomes were strongly selected and quickly spread along with the driver, replacing the
522 Y sensitive chromosomes (Helleu et al. 2019) (Figure 9D).
523 Our approach allows us to follow the migration path of the Paris SR system. For
524 example, the resistant Y chromosomes from Egypt, Morocco and France segregate in different
525 clusters, indicating that the Paris SR system has migrated independently to North Africa and
526 Europe. Similarly, the two resistant Y chromosomes from Seychelles - which have been
527 collected more than 20 years apart - belong to two different clusters. The Y chromosome

collected in 2003 (Sey03, cluster II) likely results from migration from continental Africa to the Seychelles island around that time. In addition, it appears to be more resistant than the one collected in 1981 (Sey81, cluster V), which could explain this replacement.

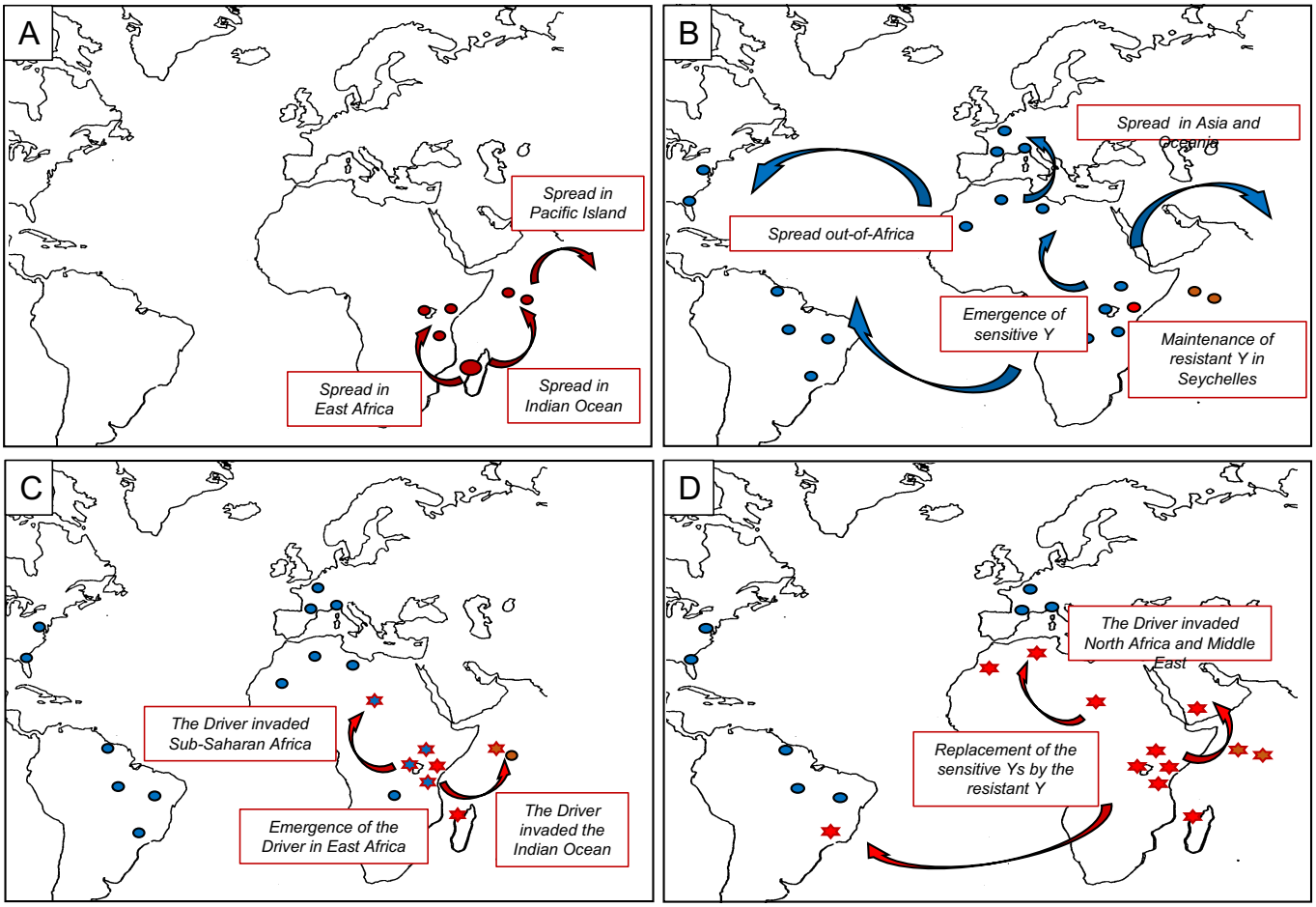


Figure 9. **Proposed scenario for the demographic history of *D. simulans* in relation to the Paris *SR* system**. A. Expansion of *D. simulans* out of its area of origin, toward the Seychelles and the African continent (Lachaise et Silvain, 2004.; Dean et Ballard 2004). The ancestral Y chromosome belongs 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).

541 ***Genetic basis of the resistance***

542 Beyond the evolutionary history of the Y chromosome, we also provide important
543 insight into the genetic basis of the resistance to Paris X^{SR}. Based on its large phenotypic
544 variation we hypothesize that the resistance ability might be associated with variation in the
545 repeat content on the Y chromosome. We did not find any association between the resistance
546 and the gene or transposable element abundance on the Y chromosomes, they even appear quite
547 stable. In contrast, we do find a strong association with the simple satellite composition (Figure
548 6A) and identified several candidates that could be involved in resistance: (AAACAAT)n,
549 (AATGG)n, (AAGAGAAGAGAG)n and (AACAATT)n, (Figure 7A) or (AAACAAC)n and
550 (AACAATC)n (Figure 8A-B). (AAACAAT)n, (AATGG)n, (AAGAGAAGAGAG)n and
551 (AACAATT)n are known to be enriched or specific to the Y chromosome in *D. simulans* (Wei
552 et al. 2018) (Supplementary Figure 5-6) and exhibit a lower copy number on the resistant Y
553 chromosome (Figure 7A). In addition to the copy number variation, (AAACAAT)n,
554 (AATGG)n appear to be associated with structural variation among resistant Y chromosomes
555 (Figure 7B, Supplementary Figure 5) (Helleu et al. 2019), which may also contribute to
556 variation in the phenotype. The two other candidates, (AAACAAC)n and (AACAATC)n, are
557 also enriched on the Y chromosome (Wei et al 2018) but, on the opposite, their abundance
558 positively correlates with their resistance ability (Figure 8). This would suggest that those
559 satellites could confer the ability to resist to the driver. One of the drivers of the Paris *SR* system
560 has been identified as *HP1D2*, a member of the heterochromatin protein gene family, known to
561 be involved in heterochromatin regulation. We have previously shown that HP1D2 is
562 specifically expressed in spermatogonia and is enriched on the Y chromosome (Helleu et al.
563 2016). In addition, the drive is due to a loss of function of *HP1D2*. We thus hypothesize that
564 the drive is associated with a mis-regulation of the heterochromatin region of the Y
565 chromosome. While we do not know the precise target of HP1D2, those simple satellites are

566 possible candidates. Correlations between the target copy number and the strength of the drive
567 has been shown in other meiotic drive systems. For example, in the *SD* system of *D.*
568 *melanogaster*, the *Rsp* satellite is the target of the driver (Hartl 1974). Variation in *Rsp* copy
569 number is related to variation in the sensitivity to the driver (Lyttle 1989). Chromosomes with
570 a higher copy number of *Rsp* are more sensitive to the driver. Similarly, particular repeats may
571 be involved in the resistance to the Paris drivers and possibly be their target. Four of the simple
572 satellite candidates we identified as associated with drive sensitivity of the Y chromosome,
573 follow this model (Figure 7A). Three of them, (AATGG)n, (AAACAAT)n and
574 (AAGAGGAAGAGAG)n, appear absent from the *D. sechellia* Y chromosome (Wei et al
575 2018). This is consistent with the finding that the *D. sechellia* Y chromosome is resistant to the
576 Paris driver. Variation in resistance could be due to variation in their abundance and/or in their
577 localization. Considering that the distortion is due to a misregulation of heterochromatic
578 sequences, the localization of those sequences on the Y chromosome could be more or less
579 critical for chromosome segregation. Also, we cannot exclude that the resistance phenotype is
580 multifactorial, or that the genetic basis of resistance is different between Y chromosome
581 lineages.
582
583 To conclude, our study sheds light on the evolutionary history of the Y chromosome in
584 *D. simulans*. We show that resistant Y chromosomes likely predated the emergence of the Paris
585 *SR* system and that sensitive Y chromosomes have spread through the world in the recent past.
586 The Paris system is particularly young, at least in its present form, since the segmental
587 duplication which is involved in driving likely emerged less than 500 years ago (Bastide et al.
588 2013). While its impact on the recent evolutionary history of the Y chromosome is clearly
589 strong, it cannot explain on its own the whole pattern of variation we observed here. More
590 likely, the Y chromosome was recurrently involved in genetic conflicts, which resulted in

multiple selective sweeps. Furthermore, we also provide new insights on the genetic basis of Paris drive resistance. Knowing that simple satellites are likely involved should help us to move further in the understanding of the molecular mechanism underlying the Paris *SR* system.

MATERIALS AND METHODS

Fly stocks

Except for Ken73, all iso-Y lines (derived from a single male) carry the same standard genetic background (X^{ST} A^{ST}) from the ST8 reference strain (Montchamp-Moreau et Cazemajor 2002). The crossing procedure used to isolate the Y chromosomes is presented in(Montchamp-Moreau, Ginhoux, et Atlan 2001). The fourth chromosome was not controlled. The place and date of collection of the strains from which the Y chromosomes studied originate are given in Supplementary Table5.

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

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

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

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

Resistance test

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

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

626 Before statistical analyses, the data were transformed in $arcsin(\sqrt{q})$, where q corresponds to the
627 proportion of females in the progeny. To testforwithin-strain variation in Y resistance, we
628 performed a Kurskal Wallis test between groups of F2 progenies (from the same F1 cross). All
629 the iso-Y lines appear homogeneous (Supplementary Table1). We performed a Wilcoxon test
630 to determine whether those lineswere significantly different from ST8. They all appear
631 significantly more resistant than ST8 except Sey81 and New91 (Supplementary Table1).
632 Finally, we performed a Tukey HSD post-hoc test on the anova model to determine groups
633 (Figure 1B). All statistical analyses were conducted in R base(R Core Team 2022). The R script
634 used can be found on github: Figure1/resistance.R.

635

636 ***DNA extraction and sequencing***

637 DNA extraction and sequencing was performed in two batches. In batch1 we sequenced: Tun83,
638 Egy07, Ken73, Zim97, Egy12, Tan96 and Sey81. In batch2 we sequenced: Tun83, Fra93,
639 Aus11, Mor12, Cal61, Guy56, Sey81, May09, Fra09, Mor16, Sao01, Cam16, New91, Haw07,
640 Mad98, Sey03. Tun83 was sequenced in both batches to control for potential technical bias.
641 Sey81 was sequenced in the first batch in its original genomic background (from Seychelles
642 1981) and then the Y chromosome was isolated in the same standard genomic background (from
643 ST8 strain) and re-sequenced in: the second batch. DNA was extracted from 50 adults J1

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

648

649 ***Mapping and Variant calling***

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

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

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

669 analyze the quality of the SNP calling on the Y chromosome based on variant mean depth,
670 variant quality, proportion of missing data per site, mean depth per individual, missing data per
671 individual. This allowed us to determine the threshold for the quality filtering : --remove-indels
672 --max-missing 0.9 --minQ 30 --min-meanDP 10 --max-meanDP 50 --minDP 10 --maxDP 50.
673 Because each iso-Y line originates from a single male and the Y chromosome is supposed to be
674 haploid, heterozygote sites are likely to be duplicated sequences and thus difficult to take into
675 account. We removed the heterozygotes sites using the *filter* command (-e 'GT="het"' option)
676 of bcftools. Note that because Tun83 and Sey81 were sequenced twice, after verification that
677 they were identical, we keep only the data from batch2 to avoid redundancy in the analyses.
678 The pipeline used (VariantCalling.sh) can be found on the github repository.
679 We calculate the nucleotide diversity along Y chromosomes per 10kb windows using the
680 vcftool--*window-pi 10000* option.
681 To investigate the population structure we performed a Principal Component Analysis (PCA)
682 on the filtered vcf, using the software plink (v1.9)(Chang et al. 2015). The pairwise distance
683 between strains was calculated using MEGA(Tamura, Stecher, et Kumar 2021).
684 All the following analyses on the vcf file were done using R scripts (VariantAnalyses.R on
685 github) using the following libraries: *vcfR, SNPrelate, gdsfmt, adegenet, ape, pheatmap,*
686 *maSigPro, stringr*.
687
688 ***Identification of corresponding SNPs in the sister species and phylogeny.***
689 To identify the corresponding allele in *D. melanogaster*, *D. sechellia* and *D. mauritiana* we first
690 needed to identify the corresponding location. To do so, we first created a consensus Y
691 chromosome where SNPs are coding in UPAC code, using bcftools*consensus* command (with
692 --iupac-codes option). We built a bed file containing the position of each SNP. Using
693 bedtools*getfasta* command we extracted 400bp sequences for each SNPs (200bp on each side

694 of the SNPs) in a fasta file. Then each sequence were blasted against the Y chromosome
695 consensus of *D. melanogaster*, *D. sechellia* and *D. mauritiana* using blastn (-evalue 1e-20 -
696 qcov_hsp_perc 80 -perc_identity 85 -max_hsps1 -max_target_seqs 1). We output the blast
697 result in a tab delimited format (-outfmt "6 qseqidsseqidevalue length pidentqstartqseqsseq")
698 and extract the corresponding allele using a R script.
699 To reconstruct the phylogenetic tree, we first created a fasta sequence with only the SNP for
700 each individual. The SNP for which we were not able to identify a corresponding allele in the
701 sister species were coded as gaps. We constructed the maximum likelihood tree using the GTR
702 model and ultrafast bootstrap (1000 replicates) using iqTREE web server(Nguyen et al. 2015).
703 *D. melanogaster* was used as the outgroup. The scripts used can be found in github :
704 Figure3/Ancestry.sh and Figure3/Ancestry.R.

705

706 ***Transposable element and complex satellite quantification***

707 Due to the batch effect observed we only used the strain from the second batch for the analysis.
708 To estimate the abundance of transposable elements and complex satellites in each iso-Y line
709 we first mapped the read against our reference genome. We applied the same pipeline as for the
710 SNP calling, with the exception that we kept multi-mapping reads (pipeline Mappingt.sh on
711 github). The reference genome was annotated using a custom repeat library (SupFile1.fasta)
712 using repeatmasker (options *-no_is -a -inv -pa 20 -div 20*)(Smit, AFA, Hubley, R, et Green, P.
713 2013). Using htseq-count (Putri et al. 2022)we counted the number of reads that map against
714 each repeats and normalize the reads count in RPM. We removed repeats for none of the strain
715 have a RPM greater than 10. All the subsequent analyses were done using Rstudio. We
716 performed a Wilcoxon test, followed by an “fdr” correction, to determine which repeats have a
717 different enrichment between sensitive and resistant. We calculated the coefficient of variation

718 for each repeat to determine which element should be the most variable, the top 25 is presented
719 in Figure 4D. The scripts used can be found in github : TEcount.sh and Figure4D/TE.R.
720

721 ***kmers quantification***

722 Due to the batch effect observed we only used the strain from the second batch for the analysis.
723 Barring the fourth chromosome (not controlled), only the Y chromosome differs; we thus
724 consider that variations observed between lines are due to the Y chromosome. To estimate the
725 abundance of simple satellites in each strain we used an assembly-free approach based on kmer
726 quantification. We used the software k-Seek which quantified the simple satellite tandemly
727 repeated using unassembled Illumina reads. We normalized the kmers count in using the total
728 mapping reads. We removed the kmers for which none of the strain have anormalized count
729 greater than 5. All the subsequent analyses were done using Rstudio. We performed a Wilcoxon
730 test to determine which kmers have a different enrichment between sensitive and resistant. Then
731 we performed a correlation test between the kmers count and the resistance ability for the 11
732 strains that have been phenotype (Figure 1B, Supplementary Table1). The scripts used can be
733 found in github : kmers.sh and Figure5/Kmers.R.

734

735 ***Gene copy number analysis.***

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

742 the two ampliconic genes, we blasted the consensus sequences against the reference genome
743 (option *-perc_identity 80*) to determine the coordinate of each insertion. For each iso-Y line we
744 calculated the mean coverage of all single copy exons and used it as a normalization factor. To
745 estimate the exon copy number we calculated the exon mean coverage divided by the
746 normalization factor and multiplied by the number of corresponding exons annotated on the
747 reference genome. The script used can be found in github : Figure4A/Gene_count.R.
748
749
750 ***Genotyping and Haplotype tree***
751
752 We designed primers around SNPs located on regions annotated as coding regions (Chang et al
753 2022), as they are the most likely to be unique. For each of them we performed a PCR with
754 DreamTaq™ Hot Start Green PCR (ThermoFisherScientific, ref K9021) following
755 manufacturer's instructions (hybridization temperature: 58°). Each PCR product was sequenced
756 using the Sanger method on a SeqStudio Genetic Analyzer (ThermoScientific). Only five
757 primers pairs were able to deliver a unique sequence (Supplementary Table6), containing 11
758 SNPs in total (Supplementary Table2). We genotyped the sequenced iso-Y lines, to validate our
759 SNP calling, and 86 additional iso-Y lines from 22 populations. Similarly, we genotyped the
760 mitochondrial haplotype of the 21 iso-Y lines using 2 pairs of primers in Supplementary Table6.
761 We concatenated the 11 SNPs for each iso-Y strains in a nexus format and built the haplotype
762 tree by the Median Joining Method using POPart(Leigh et Bryant 2015).
763
764 ***Fluorescence in situ hybridization***
765 The FISH was performed using primary probe for (AATGG)n and (AAACAAT)n coupled with
766 sec6 and sec5 adaptors. The sequences for each probe are in Supplementary Table7. Sec5 is
767 coupled with Cy5 while sec6 is coupled with Cy3. We dissected brains from third instar larvae
768 in PBS, incubate 8min in 0.5% Sodium citrate. We fix for 6 min in 4% formaldehyde, 45%

769 acetic acid before squashing. The brains were put between the slide and coverslip and squashed
770 hard before being immersed in liquid nitrogen. After 10 min in 100% ethanol, slides were air
771 dried for at least one hour before proceeding to the hybridization. For the hybridization we use
772 20pmol of primary probes and 80pmol of the secondary probes in 50 ul of hybridization buffer
773 (50% formamide, 10% dextran sulfate, 2xSSC). Slides were heated for 5 min at 95°C for
774 denaturation and incubated overnight at 37°C in humid chamber. Slides were then washed 3
775 times 5min with 4XSSCT and 3 times 5min with 0.1SSC before being mounted in slowfade
776 DAPI. We imaged using LEICA DM5500 microscope and edited the image using Fiji.
777
778 ***Slot blot.***
779 We performed the slot blot for 14 iso-Y strains. We extracted the DNA from 30 adult males
780 using the QIAGEN DNeasy tissue and blood kit. We use 200ng of DNA for the slot blot. First
781 the DNA was incubated 10 min in denaturing buffer (0.4 NaOH, 25 mM EDTA) at room
782 temperature then cooled down in ice cold loading buffer (0.1X SSC, 0.125 N NaOH). The blot
783 was performed using the slot blotter apparatus from BioRad. The membrane was UV cross-
784 linked for 3 min. Hybridization was performed overnight at 42°C using 15 ml of hybridization
785 buffer (ULTRAhyb, Thermo Scientific) and 100pmol of primary probes and 200pmol of
786 secondary fluorescent probes (Supplementary Table7). The next day we washed the membrane
787 3 times 15 min using the wash buffer (North2South™ Hybridization Stringency Wash Buffer,
788 Thermo Scientific). We imaged the membrane using Odyssey XL. For each strain, we made 3
789 replicates that were randomly placed on the blot. The intensity of the simple satellite signals
790 were normalized by the intensity of our control gene, *rp49*.
791
792 ***Data availability***

All sequences are available from NCBI SRA under Bioproject accession PRJNA905841. All the BASH pipelines and R scripts used in this study are available on github 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 AgenceNationale de la Recherche (RESIST ANR-21-CE02-0004-02) and by National Institutes of HealthNIH R35GM119515to AML.

801 **SUPPLEMENTAL FIGURES**

802

803 Supplementary Figure 1: Y chromosome haplotypes. Each line corresponds to a different Y chromosome
804 ordered based on a clustering method. Each column corresponds to a SNP, the order of which is based
805 on their physical position on the Y chromosome. The blue alleles correspond to the reference allele and
806 the red alleles correspond to the alternative allele; white are missing data.

807

808 Supplementary Figure 2 :Haplotype network of 98 Y chromosomes from 26 populations. The haplotypes
809 are composed of 11 SNPs. The name of each Y chromosome is indicated within each cluster (see
810 Supplementary Table2 for full name). The clusters are colored according to the 4 clusters identified for
811 the 21 iso-Y lines sequenced.

812

813 Supplementary Figure 3: The box plots represent the normalized abundance of satellite abundance (y
814 axis) in each group (resistant Y's in red and sensitive Y's in blue).

815

816 Supplementary Figure 4 : Correlation plots between the satellite abundance (y axis) and the resistance
817 ability (x-axis). Correlation was calculated based on a Spearman correlation test

818

819 Supplementary Figure 5 : FISH on mitotic chromosomes from larval brains using $(AACAAT)_4$ (green)
820 and $(AATGG)_6$ (red) probes.

821

822 Supplementary Figure 6 : FISH on mitotic chromosomes from larval brains using
823 $(AAGAGAAGAGAG)_4$ (green) and $(AACAATT)_{64}$ (red) probes.

824

825 Supplementary Figure 7 : FISH on mitotic chromosomes from larval brains using $(AACAATC)_4$
826 (green) and $(AAACAAC)_4$ (red) probes.

827

828 Supplementary Figure 8: Slot blot with DNA from 14 iso-Y strains, 3 replicates for each strain. The y-
829 axis correspond the signal from the hybridization using probes targeting AAACAAC (A), AAACAAT
830 (B) and AACAATC (C). The error bar correspond to the standard error. The signal intensity was
831 normalized by the rp49 signal and plotted against the relative abundance (in log10) inferred by k-seek.
832 The correlation was calculated based on Pearson correlation test, the red line corresponds to the
833 regression line.

834

835

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.

Branco, Alan T., Rute M Brito, et Bernardo Lemos. 2017. « Sex-Specific Adaptation and Genomic Responses to Y Chromosome Presence in Female Reproductive and Neural Tissues ». *Proceedings of the Royal Society B: Biological Sciences* 284 (1869): 20172062. https://doi.org/10.1098/rspb.2017.2062.

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.

Francisco, Flávio O., et Bernardo Lemos. 2014. « How Do Y-Chromosomes Modulate Genome-Wide Epigenetic States: Genome Folding, Chromatin Sinks, and Gene Expression ». *Journal of Genomics* 2: 94-103. https://doi.org/10.7150/jgen.8043.

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.

Hamblin, Martha T, et Michel Veuille. 1999. « Population Structure Among African and Derived Populations of Drosophila Simulans: Evidence for Ancient Subdivision and Recent Admixture ». *Genetics* 153 (1): 305-17. https://doi.org/10.1093/genetics/153.1.305.

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.

Irvin, Steven D, Kris A Wetterstrand, Carolyn M Hutter, et Charles F Aquadro. 1998. « Genetic Variation and Differentiation at Microsatellite Loci in Drosophila Simulans: Evidence for Founder Effects in New World Populations ». *Genetics* 150 (2): 777-90. https://doi.org/10.1093/genetics/150.2.777.

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.

Lemos, Bernardo, Alan T Branco, Pan-Pan Jiang, Daniel L Hartl, et Colin D Meiklejohn. 2014. « Genome-Wide Gene Expression Effects of Sex Chromosome Imprinting in *Drosophila* ». *G3 Genes|Genomes|Genetics* 4 (1): 1-10. https://doi.org/10.1534/g3.113.008029.

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.