

Reversed sex-biased mutation rates for indels and base substitutions in *Drosophila melanogaster*

Lauri Törmä^{1,2}, Claire Burny^{1,2}, Christian Schlötterer^{1,#}

1) Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria

2) Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria

#) corresponding author.

Keywords: Sex-biased mutations, *spellchecker1*, faster-X evolution, *Drosophila melanogaster*

Correspondence:

Christian Schlötterer

Institut für Populationsgenetik, Vetmeduni Vienna, 1210 Wien, Austria

Email: christian.schloetterer@vetmeduni.ac.at

1 **Abstract**

2 Sex biases in mutation rates may affect the rate of adaptive evolution. In many species, males
3 have higher mutation rates than females when single nucleotide variants (SNVs) are considered.
4 In contrast, indel mutations in humans and chimpanzees are female-biased. In *Drosophila*
5 *melanogaster*, direct estimates of mutation rates did not uncover sex differences, but a recent
6 analysis suggested the presence of male-biased SNVs mutations. Here we study the sex-specific
7 mutation processes using mutation accumulation data from mismatch-repair deficient *D.*
8 *melanogaster*. We find that sex differences in flies are similar to the ones observed in humans: a
9 higher mutation rate for SNVs in males and a higher indel rate in females. These results have
10 major implications for the study of neutral variation and adaptation in *Drosophila*.

11 **Introduction**

12 Mutations are the ultimate source of novelty in evolution. However, mutations do not occur at a
13 constant rate across the genome (1, 2), between taxa (3), or sexes (4). The neutral substitution
14 rate is determined by the primary mutation rate (5) and it needs to be considered when studying
15 the intensity of selection from divergence data. If mutation rates differ between sexes, genes are
16 expected to evolve at different rates depending on which chromosome they are located (6). For
17 example, a higher mutation rate in males would result in a lower substitution rate on the X
18 chromosome compared to the autosomes, while the Y chromosome has the highest substitution
19 rate.

20
21 Higher male mutation rates have been observed for SNVs in mammals (7–9), birds (10, 11),
22 salmonid fishes (12) and plants (13, 14). For indels, the observed sex-biases show no consistent
23 trend across species. Human and chimpanzee females have a higher indel rate than males (15–
24 17). In barn swallows, a sex-specific effect was shown for the mutation rate of a single
25 microsatellite locus (18, 19) while in mice and rats indels are male-biased (20).

26
27 While sex-specific mutation rates are frequently observed, the underlying processes are highly
28 taxon and mutation-type specific. In mammals, the higher rate of SNV mutations in males has
29 been explained by more germline cell divisions in testes (9, 21). This explanation was challenged
30 by the observation that even in young parents which have a similar number of germline cell
31 divisions, males have a three times higher mutation rate than females (22). Moreover, the male-
32 to-female mutation rate ratio barely increases with age, which is in conflict with the explanation
33 involving germline divisions (22). In plants, explanations for sex-specific mutation rates include
34 differences in per-replication mutation rates between the sex chromosomes (13), as well as
35 different contributions of somatic mutations in males and females (14). For indels, the quiescent
36 state of the oocyte may be responsible for the higher female mutation rate in humans (17). Similar

37 to yeast (23), the quiescent oocyte is expected to acquire mutations not during replication, but
38 from other sources, such as double-stranded breaks (17). In mice and rats, the male-bias of indel
39 mutation rates is consistent with the hypothesis that DNA replication errors are the major source
40 of mutations (20).

41
42 Consistent with a similar number of cell divisions for male and female germlines - on average 35.5
43 in males and 34.5 in females (24) - early sequencing studies in *Drosophila* found similar mutation
44 rates for both sexes (25–29). Whole-genome comparisons with other *Drosophila* species provided
45 conflicting results. These studies reported higher (30), equal (31), and lower (32) neutral
46 substitution rates for the X chromosome. A recent analysis comparing *D. melanogaster*, *D.*
47 *simulans*, and *D. yakuba* genomes challenged some of these findings and suggested that the X
48 to autosomal (X/A) ratio of mutation rates could range from 0.81 to 0.93, depending on the type
49 of neutral site used for the estimation (33). Further evidence for male-biased base substitution
50 rates comes from the comparison of neo-Y and neo-X chromosomes in *D. miranda*, which showed
51 higher substitution rates on the neo-Y chromosome (34). Since polymorphism and divergence
52 patterns are strongly affected by selection, it is particularly noteworthy that a trend for male-biased
53 mutations — although nonsignificant — was seen in mutation accumulation lines (35).

54
55 In this study, we estimate *de novo* mutation rates on the X chromosome and the autosomes using
56 data from a mutation accumulation study mismatch-repair (MMR) deficient *D. melanogaster* strain
57 consisting of 7,345 new single nucleotide variants (SNV) and 5,672 indels (36). We find opposing
58 patterns of sex-specific mutation rates for SNVs and indels. While males have a higher mutation
59 rate for SNVs (X/A mutation rates ratio: 0.876), indels are more common in females (X/A mutation
60 rates ratio: 1.959). We show that wild-type females also have a higher indel mutation rate and
61 conclude that our results are not an artifact from the mismatch repair deficiency, but reflect
62 genuine sex-specific differences in mutation processes. These results have major implications for

63 the interpretation of neutral variation and adaptation in *D. melanogaster*.

64

65 **Results and discussion**

66 We scrutinized sex-specific mutation rates in *D. melanogaster* by comparing the *de novo*
67 mutations on the X chromosome and the autosomes in a mismatch repair deficient genetic
68 background, lacking a functional copy of the MutS homolog *spellchecker1* (37). Because the X
69 chromosome spends more time in females than in males, a higher mutation rate in males leads
70 to a higher mutation rate on the autosomes compared to the X chromosome. Consistent with
71 previous results (33), we found that the X/A ratio of mutation rates was 0.876 (Poisson test, 95%
72 Confidence Interval (CI): 0.823-0.933, p-value= 2.923×10^{-5}) resulting from a X chromosomal
73 mutation rate of 8.71×10^{-7} (95% Poisson CI: 8.23×10^{-7} - 9.23×10^{-7}) and an autosomal rate of
74 9.94×10^{-7} (95% Poisson CI: 9.69×10^{-7} - 1.02×10^{-6}) (Figure 1a & Table 1). The difference between
75 chromosomes still persists after controlling for the base content between the chromosomes
76 (Figure 1b & Table 1).

77

78 With indel rates being sex-specific in multiple species (17–20), we tested for sex-specific indel
79 rates in *Drosophila*. The indel mutation rate on the X chromosome was 1.26×10^{-6} (95% Poisson
80 CI: 1.20×10^{-6} - 1.32×10^{-6}), which is almost twice as high as the autosomal rate of 6.42×10^{-7} (95%
81 Poisson CI: 6.22×10^{-7} - 6.62×10^{-7}), resulting in a significant rate ratio of 1.959 (95% Poisson CI:
82 1.850-2.074, p-value $<2.2 \times 10^{-16}$) (Figure 1c & Table 1).

83

84 Indels occur more frequently on stretches of short tandem repeats (microsatellites) due to DNA
85 replication slippage (38–40). Because microsatellite repeat number is an important factor in
86 determining the indel mutation rate (41–43), the inferred sex-specific indel mutation rate may be
87 the result of the heterogeneity in repeat length distribution between the X chromosome and the
88 autosomes (44). We accounted for possible differences in the distribution of repeat length on the

89 X and the autosomes by fitting a binomial generalized additive model (GAM, (45); see Methods)
90 separately for mono- and di-nucleotide repeats. The indel mutation rate changes significantly with
91 repeat length for both the homopolymers and dinucleotides (>5 degrees of freedom (dof), $\chi^2=$
92 5,046, $p<2\times 10^{-16}$; >6 dof, $\chi^2 = 1,676$, $p<2\times 10^{-16}$ respectively), but not in a monotonous way. The
93 mutation rate increases for homopolymer runs up to 10-12 bp for (Figure 2a) and for dinucleotides
94 up to 9-10 repeat units (Figure 2b). Microsatellites with more repeats were less likely to mutate
95 and also less common. We found that the indel rate is significantly higher on the X chromosome
96 for homopolymers (Odds Ratio X/A (OR)=1.163, 95% CI: 1.093-1.237, deviance analysis
97 $p=2.049\times 10^{-6}$) (Figure 2a) and dinucleotides (OR=1.291, 95% CI: 1.074-1.553, deviance analysis
98 $p=7.184\times 10^{-3}$) (Figure 2b). The highly consistent pattern across the two different repeat types
99 strongly suggests that female flies have a higher indel mutation rate than males.

100

101 To rule out that the sex-specific difference in indel rates is an artifact of the mismatch repair
102 system deficiency, we studied indel mutations in a natural *Drosophila* population. Using variant
103 calls from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) (46), we compared the
104 number of indels to the number of SNVs between the X chromosome and the autosomes. We
105 found an odds ratio significantly different from 1 (OR=1.320, 95% CI: 1.309-1.332, Fisher's exact
106 test, $p<2.2\times 10^{-16}$), indicating that more indels occur on the X chromosome which confirms the
107 female-biased indel mutation rate also in wild-type flies.

108

109 While we find the same sex-biased indel mutation rate as in humans and chimpanzees (17), we
110 consider it unlikely that the interpretation of this sex-specific difference can be applied to
111 *Drosophila*. For humans and chimpanzees, the quiescent state of the female oocyte has been
112 attributed to the higher indel rate in females (17). This is however unlikely to cause the
113 pronounced differences between the X chromosome and the autosomes in our experiment
114 because it was designed such that young females were fertilized, leaving little opportunity for

115 oocytes to enter a quiescent state. The mismatch repair targets replicating cells (47), which
116 implies that most of the observed mutations did not occur in the quiescent phase, but during
117 replication. Furthermore, the majority of indels (95.47% in mono- and di-nucleotides) occur on
118 microsatellites arguing in favor of replication errors. Hence, we conclude that, at least in
119 *Drosophila*, the higher indel rate in females cannot be explained by the quiescent phase of the
120 oocytes, but rather other sex-specific differences in replication and/or DNA repair must be
121 responsible for these differences.

122

123 One particularly interesting aspect of our study is that we detect sex-specific mutation rates for
124 two types of mutations, SNVs and indels, but the sex-bias is reversed between them (Figures 1a,
125 1c). It has been proposed that the mutation rate has evolved as a balance between the cost in
126 fitness due to accurate replication, repair, and deleterious mutations and the benefit of fitness
127 increase due to advantageous mutations (48). In the case of sex-biased mutation rates, this
128 implies that the cost/benefit balance differs between sexes (49). Since we show a different sex
129 bias for SNVs and indels, it is difficult to reconcile a simple cost/benefit balance with our data.
130 Rather, we consider it more likely that DNA damage and repair are sex-specific processes with
131 their own signatures. The same conclusion was reached recently in humans (22).

132

133 The impact of the different sex bias for SNVs and indels is nicely illustrated by the different length
134 distributions of AT-microsatellites on the X chromosome and autosomes in *D. melanogaster* (44).
135 Microsatellite length may be explained by an equilibrium process between replication slippage
136 generating longer repeats and base substitutions shortening the microsatellite (50). Hence, it was
137 previously proposed that the difference in length distribution in *D. melanogaster* may be either
138 explained by a higher slippage rate on the X chromosome or a higher base substitution rate in
139 males (44). Our study now demonstrates that both processes occur and their joint effects probably

140 explain the heterogeneous microsatellite length distribution between the X chromosome and
141 autosomes.

142

143 In this study, we demonstrated that increasing the mutation rate in mutation accumulation lines is
144 a powerful approach to study mutation processes, as it overcomes the typically encountered
145 limitation of too few observations. We anticipate that the analysis of mutation accumulation lines
146 with elevated mutation rates can provide a powerful method to study differences in the mutation
147 process, either between sexes, as in this study, or between different genomic regions.

148

149 **Materials and methods**

150 All statistical analyses were done with R (51) (version 3.5.0).

151

152 **Data and mutation rate estimations**

153 We used mutations generated in (36) to perform a direct estimation of mutation rates. The VCF
154 file for the DGRP Freeze 2.0 calls (46) was downloaded from:
155 <http://dgrp2.gnets.ncsu.edu/data.html> in October 2019. We used the GATK tool CallableLoci (52)
156 (version 4.0.12.0) to calculate the number of callable sites for each of the 7 lines from the same
157 BAM files that were used in (36). We estimated the SNVs - using all SNVs and separated for A/T
158 and G/C pairs - and indel mutation rates with the following formula: $m/(t \times c)$ where m is the total
159 number of mutations across the 7 lines, t is equal to 10 generations, c is the total number of
160 callable sites over each line. A 95% confidence interval was obtained by using an exact Poisson
161 test with the `poisson.test` R function, as suggested in (53).

162

163 **Microsatellites and the model**

164 We used a previously published R code (43) to search for homopolymer runs from
165 `BSgenome.Dmelanogaster.UCSC.dm6` (54). Dinucleotide repeats were annotated from the *D.*

166 *melanogaster* reference genome release 6.24 using the microsatellite finder software MISA (55)
167 (version 2.0). We included only microsatellites with i) at least 6 repeats and ii) a maximum length
168 of 18 repeats (homopolymers) or 17 repeat units (dinucleotides), as longer microsatellites did not
169 have any indel mutations. Following (43) we analyzed the occurrence of 5,415 (95.47%) indels
170 separately for homopolymers and dinucleotides using binomial generalized additive models (45)
171 with the non-linear effect of the repeat length. In addition we tested for differences between
172 chromosomes by using the chromosomal status (autosomes or X) as covariate. The
173 corresponding R function is `mgcv::gam(cbind(#indels, #repeats - #indels) ~ Chr + s(Length, fx =`
174 `FALSE, k=-1, bs = "cr"), family = binomial, data)` (45, 56) (mgcv R package version 1.8-31), where
175 `s` is a cubic spline, `"#indels"` is the number of indels per number of repeats, `"#repeats"` is the
176 number of repeats of a given length and `"Chr"`, the chromosomal status. Two models, with (M1)
177 and without (M0) the chromosome covariate `"Chr"`, were compared in a deviance analysis using
178 the `mgcv::anova.gam(M0, M1, test = "Chisq")` R function. Odds ratios (OR) and their 95%
179 confidence intervals are obtained using the Miettinen-Nurminen method (57); $OR = \exp(\beta_{Chr} \pm$
180 $1.96 \times SE)$, with β_{Chr} the slope associated with the chromosomal status (the autosomal status
181 being the reference level) and SE its corresponding standard error. The p-values associated with
182 the smoother of the length were reported from the summary R function. We reported the ratio of
183 indels in both types of repeats normalized to the genome-wide number of repeats (Figure 2, y-
184 axis).

185

186 **Code and data availability**

187 The code (R and bash scripts) and intermediate files such as indel counts per repeat type will be
188 accessible in the following github repository: `***`, available upon publication.

189

190 **Author contribution**

191 L. T., C. B. analyzed the data, L. T., C. B., C. S. wrote the paper, C.S. supervised the project.

192 **Acknowledgments**

193 This work was supported by the Austrian Science Fund (FWF, grant W1225). We thank A.
194 Futschik for statistical advice.

195

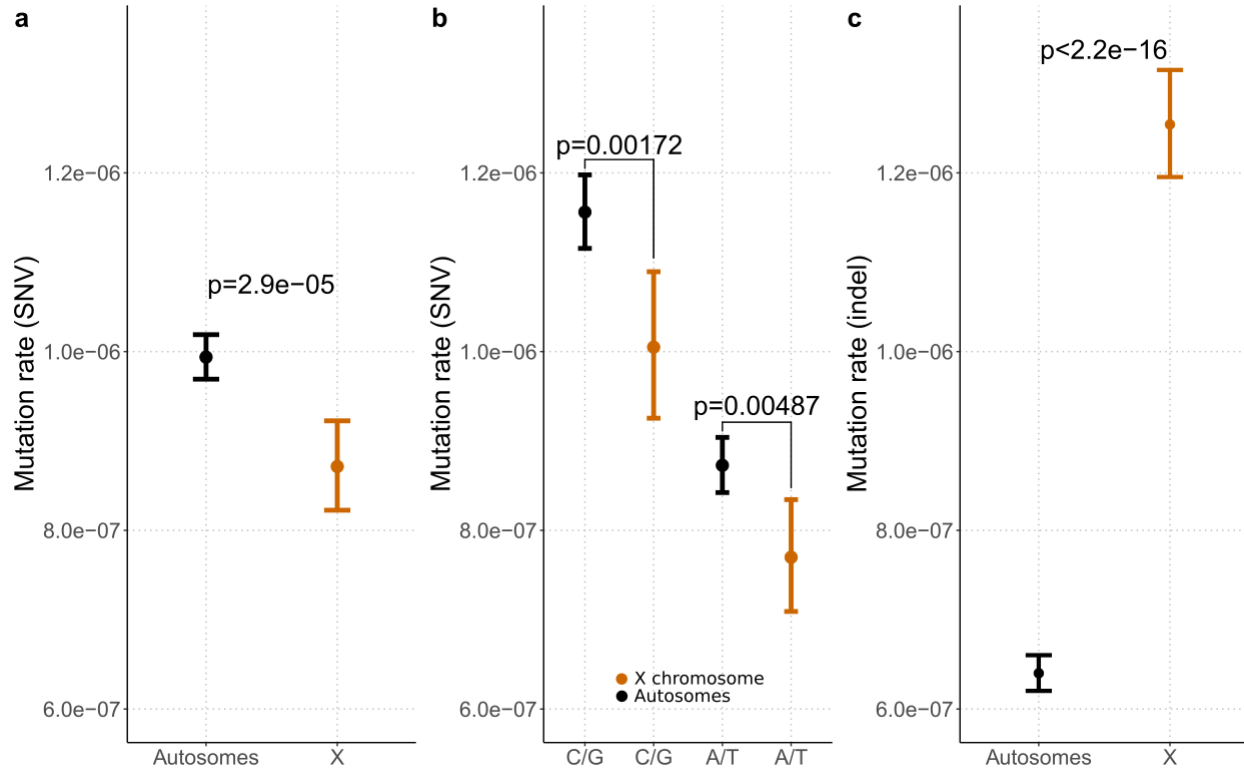
196 **References**

- 197 1. M. Wierdl, M. Dominska, T. D. Petes, Microsatellite instability in yeast: dependence on the
198 length of the microsatellite. *Genetics* **146**, 769–779 (1997).
- 199 2. Z. J. Assaf, S. Tilk, J. Park, M. L. Siegal, D. A. Petrov, Deep sequencing of natural and
200 experimental populations of *Drosophila melanogaster* reveals biases in the spectrum of
201 new mutations. *Genome Res.* **27**, 1988–2000 (2017).
- 202 3. W. Sung, M. S. Ackerman, S. F. Miller, T. G. Doak, M. Lynch, Drift-barrier hypothesis and
203 mutation-rate evolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18488–18492 (2012).
- 204 4. L. D. Hurst, H. Ellegren, Sex biases in the mutation rate. *Trends Genet.* **14**, 446–452
205 (1998).
- 206 5. M. Kimura, Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
- 207 6. M. Kirkpatrick, D. W. Hall, Male-biased mutation, sex linkage, and the rate of adaptive
208 evolution. *Evolution* **58**, 437–440 (2004).
- 209 7. R. A. Gibbs, *et al.*, Genome sequence of the Brown Norway rat yields insights into
210 mammalian evolution. *Nature* **428**, 493–521 (2004).
- 211 8. K. Lindblad-Toh, *et al.*, Genome sequence, comparative analysis and haplotype structure of
212 the domestic dog. *Nature* **438**, 803–819 (2005).
- 213 9. M. A. Wilson Sayres, C. Venditti, M. Pagel, K. D. Makova, Do variations in substitution rates
214 and male mutation bias correlate with life-history traits? A study of 32 mammalian
215 genomes. *Evolution* **65**, 2800–2815 (2011).
- 216 10. H. Ellegren, A. K. Fridolfsson, Male-driven evolution of DNA sequences in birds. *Nat.*
217 *Genet.* **17**, 182–184 (1997).
- 218 11. N. W. Kahn, T. W. Quinn, Male-driven evolution among Eoaves? A test of the replicative
219 division hypothesis in a heterogametic female (ZW) system. *J. Mol. Evol.* **49**, 750–759
220 (1999).
- 221 12. H. Ellegren, A.-K. Fridolfsson, Sex-specific mutation rates in salmonid fish. *J. Mol. Evol.* **56**,
222 458–463 (2003).
- 223 13. D. A. Filatov, D. Charlesworth, Substitution rates in the X- and Y-linked genes of the plants,
224 *Silene latifolia* and *S. dioica*. *Mol. Biol. Evol.* **19**, 898–907 (2002).
- 225 14. C.-A. Whittle, M. O. Johnston, Male-driven evolution of mitochondrial and chloroplastial
226 DNA sequences in plants. *Mol. Biol. Evol.* **19**, 938–949 (2002).

- 227 15. T. Grimm, *et al.*, On the origin of deletions and point mutations in Duchenne muscular
228 dystrophy: most deletions arise in oogenesis and most point mutations result from events in
229 spermatogenesis. *J. Med. Genet.* **31**, 183–186 (1994).
- 230 16. C. Lázaro, *et al.*, Sex differences in mutational rate and mutational mechanism in the NF1
231 gene in neurofibromatosis type 1 patients. *Hum. Genet.* **98**, 696–699 (1996).
- 232 17. G. Achaz, S. Gangloff, B. Arcangioli, The quiescent X, the replicative Y and the Autosomes
233 <https://doi.org/10.1101/351288>.
- 234 18. C. R. Primmer, N. Saino, A. P. Møller, H. Ellegren, Unraveling the Processes of
235 Microsatellite Evolution Through Analysis of Germ Line Mutations in Barn Swallows
236 *Hirundo rustica*. *Mol. Biol. Evol.* **15**, 1047–1047 (1998).
- 237 19. J. Brohede, C. R. Primmer, A. Møller, H. Ellegren, Heterogeneity in the rate and pattern of
238 germline mutation at individual microsatellite loci. *Nucleic Acids Res.* **30**, 1997–2003
239 (2002).
- 240 20. K. D. Makova, S. Yang, F. Chiaromonte, Insertions and deletions are male biased too: a
241 whole-genome analysis in rodents. *Genome Res.* **14**, 567–573 (2004).
- 242 21. J. B. S. Haldane, The rate of spontaneous mutation of a human gene. *J. Genet.* **31**, 317
243 (1935).
- 244 22. Z. Gao, *et al.*, Overlooked roles of DNA damage and maternal age in generating human
245 germline mutations. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 9491–9500 (2019).
- 246 23. S. Gangloff, *et al.*, Quiescence unveils a novel mutational force in fission yeast. *Elife* **6**
247 (2017).
- 248 24. J. B. Drost, W. R. Lee, The developmental basis for germline mosaicism in mouse and
249 *Drosophila melanogaster*. *Genetica* **102-103**, 421–443 (1998).
- 250 25. V. L. Bauer, C. F. Aquadro, Rates of DNA sequence evolution are not sex-biased in
251 *Drosophila melanogaster* and *D. simulans*. *Mol. Biol. Evol.* **14**, 1252–1257 (1997).
- 252 26. D. J. Begun, P. Whitley, Reduced X-linked nucleotide polymorphism in *Drosophila*
253 *simulans*. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5960–5965 (2000).
- 254 27. A. J. Betancourt, D. C. Presgraves, W. J. Swanson, A test for faster X evolution in
255 *Drosophila*. *Mol. Biol. Evol.* **19**, 1816–1819 (2002).
- 256 28. B. A. Counterman, D. Ortíz-Barrientos, M. A. F. Noor, Using comparative genomic data to
257 test for fast-X evolution. *Evolution* **58**, 656–660 (2004).
- 258 29. K. Thornton, D. Bachtrog, P. Andolfatto, X chromosomes and autosomes evolve at similar
259 rates in *Drosophila*: no evidence for faster-X protein evolution. *Genome Res.* **16**, 498–504
260 (2006).
- 261 30. D. J. Begun, *et al.*, Population genomics: whole-genome analysis of polymorphism and
262 divergence in *Drosophila simulans*. *PLoS Biol.* **5**, e310 (2007).
- 263 31. T. T. Hu, M. B. Eisen, K. R. Thornton, P. Andolfatto, A second-generation assembly of the

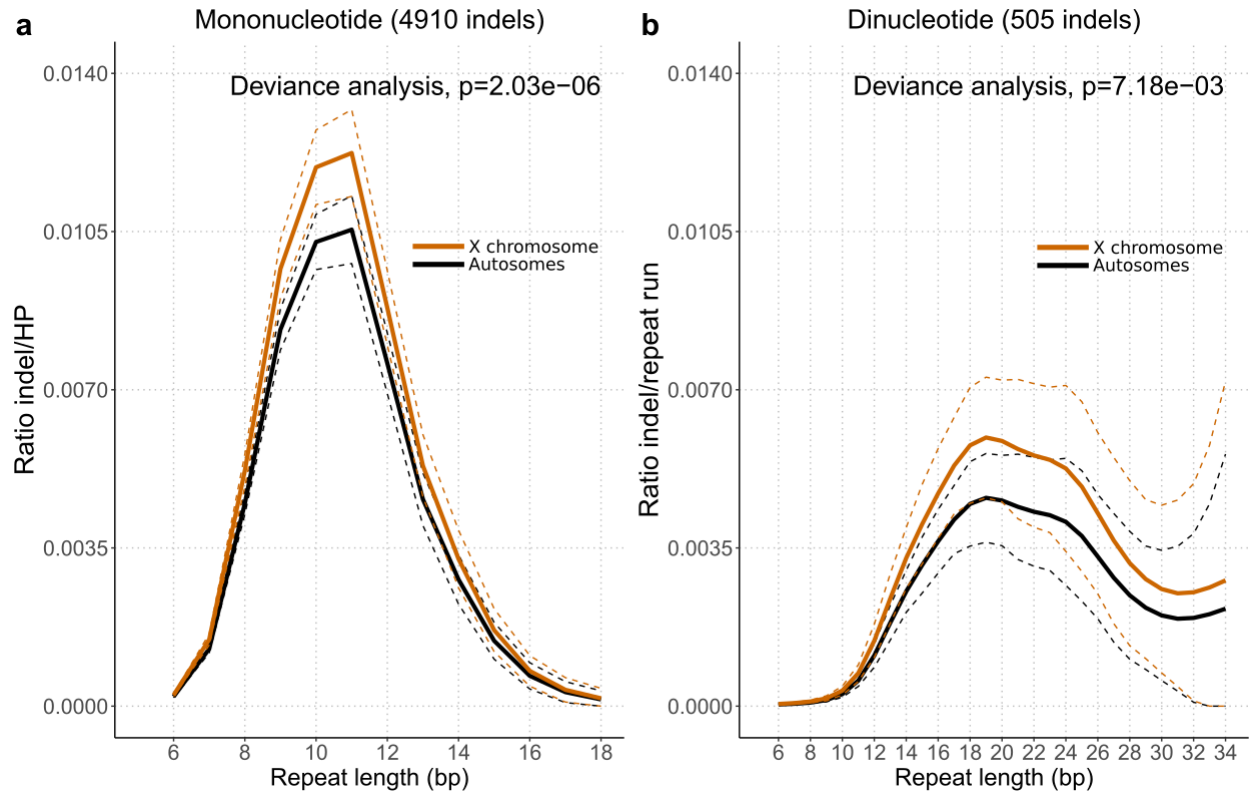
- 264 *Drosophila simulans* genome provides new insights into patterns of lineage-specific
265 divergence. *Genome Res.* **23**, 89–98 (2013).
- 266 32. D. Garrigan, S. B. Kingan, A. J. Geneva, J. P. Vedanayagam, D. C. Presgraves, Genome
267 diversity and divergence in *Drosophila mauritiana*: multiple signatures of faster X evolution.
268 *Genome Biol. Evol.* **6**, 2444–2458 (2014).
- 269 33. B. Charlesworth, J. L. Campos, B. C. Jackson, Faster-X evolution: Theory and evidence
270 from *Drosophila*. *Mol. Ecol.* **27**, 3753–3771 (2018).
- 271 34. D. Bachtrog, Evidence for male-driven evolution in *Drosophila*. *Mol. Biol. Evol.* **25**, 617–619
272 (2008).
- 273 35. P. D. Keightley, *et al.*, Analysis of the genome sequences of three *Drosophila melanogaster*
274 spontaneous mutation accumulation lines. *Genome Res.* **19**, 1195–1201 (2009).
- 275 36. L. Törmä, C. Burny, V. Nolte, K.-A. Senti, C. Schlötterer, Transcription-coupled repair in
276 *Drosophila melanogaster* is independent of the mismatch repair pathway. *bioRxiv*,
277 2020.04.07.029033 (2020).
- 278 37. C. Flores, W. Engels, Microsatellite instability in *Drosophila* spellchecker1 (MutS homolog)
279 mutants. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2964–2969 (1999).
- 280 38. G. Levinson, G. A. Gutman, Slipped-strand mispairing: a major mechanism for DNA
281 sequence evolution. *Mol. Biol. Evol.* **4**, 203–221 (1987).
- 282 39. C. Schlötterer, Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365–371
283 (2000).
- 284 40. H. Ellegren, Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* **5**,
285 435–445 (2004).
- 286 41. B. Brinkmann, M. Klintschar, F. Neuhuber, J. Hühne, B. Rolf, Mutation rate in human
287 microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum.*
288 *Genet.* **62**, 1408–1415 (1998).
- 289 42. B. Harr, C. Schlötterer, Long microsatellite alleles in *Drosophila melanogaster* have a
290 downward mutation bias and short persistence times, which cause their genome-wide
291 underrepresentation. *Genetics* **155**, 1213–1220 (2000).
- 292 43. B. Meier, *et al.*, Mutational signatures of DNA mismatch repair deficiency in *C. elegans* and
293 human cancers. *Genome Res.* **28**, 666–675 (2018).
- 294 44. D. Bachtrog, S. Weiss, B. Zangerl, G. Brem, C. Schlötterer, Distribution of dinucleotide
295 microsatellites in the *Drosophila melanogaster* genome. *Mol. Biol. Evol.* **16**, 602–610
296 (1999).
- 297 45. S. N. Wood, Fast stable restricted maximum likelihood and marginal likelihood estimation of
298 semiparametric generalized linear models. *J. R. Stat. Soc. Series B Stat. Methodol.* **73**, 3–
299 36 (2011).
- 300 46. T. F. C. Mackay, *et al.*, The *Drosophila melanogaster* Genetic Reference Panel. *Nature*
301 **482**, 173–178 (2012).

- 302 47. P. Modrich, Mechanisms in eukaryotic mismatch repair. *J. Biol. Chem.* **281**, 30305–30309
303 (2006).
- 304 48. P. D. Sniegowski, P. J. Gerrish, T. Johnson, A. Shaver, The evolution of mutation rates:
305 separating causes from consequences. *Bioessays* **22**, 1057–1066 (2000).
- 306 49. H. Ellegren, Characteristics, causes and evolutionary consequences of male-biased
307 mutation. *Proc. Biol. Sci.* **274**, 1–10 (2007).
- 308 50. S. Kruglyak, R. T. Durrett, M. D. Schug, C. F. Aquadro, Equilibrium distributions of
309 microsatellite repeat length resulting from a balance between slippage events and point
310 mutations. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10774–10778 (1998).
- 311 51. R Core Team, R: A Language and Environment for Statistical Computing (2018).
- 312 52. A. McKenna, *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing
313 next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- 314 53. H. Long, M. G. Behringer, E. Williams, R. Te, M. Lynch, Similar Mutation Rates but Highly
315 Diverse Mutation Spectra in Ascomycete and Basidiomycete Yeasts. *Genome Biol. Evol.* **8**,
316 3815–3821 (2016).
- 317 54. The Bioconductor Dev Team, *BSgenome.Dmelanogaster.UCSC.dm6: Full genome*
318 *sequences for Drosophila melanogaster (UCSC version dm6)* (2014).
- 319 55. T. Thiel, W. Michalek, R. K. Varshney, A. Graner, Exploiting EST databases for the
320 development and characterization of gene-derived SSR-markers in barley (*Hordeum*
321 *vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422 (2003).
- 322 56. S. N. Wood, N. Pya, B. Säfken, Smoothing Parameter and Model Selection for General
323 Smooth Models. *J. Am. Stat. Assoc.* **111**, 1548–1563 (2016).
- 324 57. O. Miettinen, M. Nurminen, Comparative analysis of two rates. *Stat. Med.* **4**, 213–226
325 (1985).



326

327 Figure 1. Mutation rate differences between the autosomes (black) and the X chromosome
328 (orange). a) SNV mutation rates across all sites and b) separated for A/T and G/C pairs. c) Indel
329 mutation rates across all sites. The bars represent 95% confidence intervals using exact Poisson
330 tests, the dots represent the actual estimates.



331

332 Figure 2. Occurrence of indels in mononucleotide (a) and dinucleotide (b) repeats on the
333 autosomes (black) and the X chromosome (orange). The x-axis represents the repeat length. The
334 y-axis represents the ratio of indels normalized for the prevalence of repeat type in the genome,
335 as described in (43). Binomial generalized additive models with cubic spline have been fitted. The
336 predicted values are in plain lines, 95% confidence intervals are represented in dotted lines. The
337 reported p-values are obtained from a deviance analysis contrasting a model with and without the
338 chromosomal status (autosomes or X) covariate.

339 Table 1. Mutation rate estimates.

Mutation type	Context	Number of mutations	Mutation rate (95% Poisson CI)
SNVs	Autosomes	6161	9.938×10^{-7} (9.691×10^{-7} - 1.019×10^{-6})
SNVs	X chromosome	1184	8.714×10^{-7} (8.225×10^{-7} - 9.225×10^{-7})
SNVs	C/G Autosomes	3063	1.156×10^{-6} (1.116×10^{-6} - 1.198×10^{-6})
SNVs	C/G X chromosome	590	1.005×10^{-6} (9.255×10^{-7} - 1.089×10^{-6})
SNVs	A/T Autosomes	3098	8.727×10^{-7} (8.423×10^{-7} - 9.040×10^{-7})
SNVs	A/T X chromosome	594	7.699×10^{-7} (7.092×10^{-7} - 8.343×10^{-7})
Indel	Autosomes	3968	6.401×10^{-7} (6.203×10^{-7} - 6.603×10^{-7})
Indel	X chromosome	1704	1.254×10^{-6} (1.195×10^{-6} - 1.315×10^{-6})

340