Transposable elements in individual genotypes of *Drosophila simulans*

Sarah Signor[1]

[1]Department of Biological Sciences, North Dakota State University

sarah.signor@ndsu.edu

Key words: transposable elements, *Drosophila simulans*, inbred lines, copy number, genetic variation

Abstract

Transposable elements are mobile DNA sequences that are able to copy themselves within a host's genome. Within insects they often make up a substantial proportion of the genome. While they are the subject of intense research, often times when copy number is estimated it is estimated only at the population level, or in a limited number of individuals within a population. I was interested in whether or not transposable elements accumulate at different rates in different genetic backgrounds – the variance between individual genotypes in transposable element accumulation. Using two populations of *Drosophila simulans* from California and Africa I estimated transposable element copy number in individual genotypes. Some active transposable elements seem to be a property of the species, while others of the populations. I find that in addition to population level differences in transposable element load certain genotypes accumulate transposable elements at a much higher rate than others. Most likely active transposable elements are fairly rare, and were inherited only by specific genotypes that were used to create the inbred lines. Whether or not this reflects dynamics in natural populations, where transposable elements may accumulate in specific genotypes and maintain themselves in the population rather than being active at low levels population wide, is an open question.

Introduction

Transposable elements are mobile DNA sequences that are able to move around the genome, potentially increasing their copy number in the process. They are abundant and contribute substantially to genetic variation in populations, influencing genome evolution (Biémont *et al.*, 1999; Bennetzen, 2000; Feschotte, 2008). The number and location of transposable element insertions can vary substantially between species, populations, and individuals (Vieira *et al.*, 1999; Vieira & Biémont, 2004; C Vieira, 2008; Kofler & Schlötterer, 2015; Kofler *et al.*, 2015b; Jakšić *et al.*, 2017; Kofler *et al.*, 2018). Introgression of transposable elements between species is also common, and has been reported frequently (Kofler *et al.*, 2015a; Hill & Betancourt, 2018). In addition, they make up a substantial proportion of the genome in many invertebrates and plants (Bennetzen, 2000; Tenaillon *et al.*, 2010; Kofler *et al.*, 2015b). Transposable elements are divided into three categories: long terminal repeat retrotransposons, short interspersed nuclear element retrotransposons, and transposons. The latter moves through DNA excision and repair, while the former two transpose via an RNA intermediate.

The most well studied transposable element population in insects is that of *Drosophila melanogaster*. It has been reported that in *D. melanogaster* total copy number of transposable elements may be similar between individuals, but the location of insertions can vary considerably (Barrón *et al.*, 2014). Overall copy number is thought to be maintained by transposition-selection balance, though new transposable elements may initially have high transposition rates as the host machinery evolves new defenses (Yang & Nuzhdin, 2003; Pasyukova, 2004; Slotkin & Martienssen, 2007; Johnson, 2010; Lee & Langley, 2012; Romero-Soriano & Garcia Guerreiro, 2016). Transposable elements also accumulate in heterochromatic regions, potentially because these regions are thought to be ‘transcriptionally silent’, and thus the deleterious effects of their

insertion are mitigated (Dimitri, 2003). However, they can also contribute to variation in quantitative traits, differences in fitness, and changes in gene expression (Mackay 1984; Mackay 1989, Shrimpton et al. 1990, Mackay et al. 1992, Long et al. 2000) . Some transposable elements have been associated with increases in fitness due to changes in gene regulation (Aminetzach *et al.* 2005; Mateo *et al.*, 2014)).

The recent explosion of population level genomic data has greatly increased our ability to estimate transposable element load in populations. Much of this work on transposable element copy number has occurred in *Drosophila*, where the existence of multiple sequenced inbred panels lend themselves to estimating copy number and insertion site frequency . However, differences in the number of transposable element insertions are thought to accumulate in inbred laboratory lines because active copies of transposable elements are rare and will be inherited only in some lines (Nuzhdin *et al.*, 1997). This is potentially mirroring what occurs in natural populations, where transposable element site heterogeneity may potentially be due to frequent transpositions in rare flies rather than low levels of transposition population-wide (Nuzhdin, 2000). Many studies of transposable element insertions compare populations which have been in the laboratory for different amounts of time, wild flies and laboratory flies, and they estimate population-level insertion rate using Pool-seq. While Pool-seq is an effective tool for estimating population-level frequency, understanding the dynamics of transposable elements in inbred lines requires the ability to estimate the variance between genotypes in transposable element copy number, as well as the repeatability of transposable element load between populations. Here we use previously sequenced lines of *D. simulans* from two populations to directly estimate transposable element copy number in individual genotypes.

One population was collected in Africa in 2002 and 2006 and I do not know how long they were held in the lab prior to being sequenced, however six of the stocks were sequenced from originally collected material (wild). A second Californian population was held in the lab for two years prior to sequencing while they were being inbred. I will look at variance in transposable element copy number between inbred genotypes, differences between the wild and inbred lines, and differences between the populations in the mean and variance of transposable element copy number.

Methods

Fly lines

The 22 African *D. simulans* isofemale lines were collected by William Ballard in 2002 from Madagascar and Peter Andolfatto in 2006 from Kenya (Jackson *et al.*, 2017). They were inbred in the lab for nine generations, during which time six were lost and DNA was collected from the initial stocks. The raw reads are 90 bp paired end illumina sequencing and they were downloaded from SRA PRJEB7673 (Jackson *et al.*, 2017). The first read from each pair was used for mapping. The California lines were collected from the Zuma Organic Orchard in Los Angeles, CA on two consecutive weekends of February 2012, from a pile of rotting strawberries (Signor *et al.*, 2017; Signor & Nuzhdin, 2018; 2019). Reads were single end 100 bp and this project has been deposited at the SRA under accession SRP075682.

Mapping and copy number estimation

We mapped the reads using bwa mem version 0.7.15 to the *D. simulans* 2.02 assembly and the 179 consensus transposable element sequences from EMBL, downloaded from Flybase.org (Li 2015). Bam files were sorted and index with samtools v.1.9 and optical duplicates were removed

using picard MarkDuplicates (http://picard.sourceforge.net) (Li *et al.*, 2009; McKenna *et al.*, 2010). Reads with a mapping quality of below 15 were removed (this removes reads which map equally well to more than one location). Using read coverage to determine copy number has been compared to other methods and is neither permissive nor conservative (Srivastav & Kelleher, 2017). We therefore estimated transposable element copy number per family by comparing average coverage of the transposable sequence to average coverage of chromosome arm 2L in R, and computed population averages and individual copy number (Hill *et al.*, 2015). P-values of comparisons between means and variances were corrected for multiple testing using Bonferroni correction.

SNPs and summary statistics

We called SNPs within the transposable elements and the genomes using GATK Haplotypecaller (McKenna *et al.*, 2010). SNPs were filtered for a minimum depth of four. SNPs were not filtered for missing calls given that not at all individuals will share insertions. Tajima's *D* was estimated in windows of 1 kb using VCFtools, and prior to estimation indels and SNPs with more than two alleles were removed (Danecek *et al.*, 2011). Tajima's D was not intended to estimate selection on the transposable elements, but rather to understand if some transposable element insertions were more recent than others, indicated by the frequency and number of SNPs.

Results and Discussion

Population-level variation

Of the 128 elements examined in the population, 85 have different mean numbers of insertions between the two populations (t-test, Bonferroni corrected p = .05/128). Of those, only 17 are higher in the African populations, suggesting that overall the CA population has more

137 transposable element insertion sites. Indeed, overall Californian *D. simulans* have an average of
138 1,797 insertions per genotype, while African *D. simulans* have 1,496. The five elements with the
139 largest difference in copy number in Californian *D. simulans* compared to African are the *INE-1*,
140 *Tc1*, *transib2*, *1360*, and *Cr1a*. These are present on average in 37 more copies in Californian *D.*
141 *simulans*. The elements more common in Africa include *baggins*, *X-element*, *R1A1-element*,
142 *diver2*, *Rt1c*, *Rt1b*, *jockey*, *ninja*, *Osvaldo*, *gypsy*, *Bari1*, *gypsy8*, *Tirant*, *mariner*, *BS3*, *G5A*, and
143 *aurora-element*. *G5A*, *gypsy8*, *X-element*, *baggins*, *R1A1-element*, *copia*, *G7*, *invader6*, *diver*,
144 *hobo*, *invader3*, *Rt1b*, *Rt1c*, *Bari1*, and *Tirant* also have significantly different (and higher)
145 variance compared to Californian *D. simulans* (F-test, Bonferroni corrected p = 0.05/128). 20
146 elements have different and larger variance in the Californian *D. simulans* compared to African
147 (F-test, Bonferroni corrected p = 0.05/128). This includes *Stalker4*, *Tc3*, *frogger*, *F-element*, *412*,
148 *accord*, *G4*, *rover*, *17.6*, *HMS-Beagle*, *gypsy6*. The most abundant transposable elements in the
149 California population were the highly abundant *INE-1*, *Cr1a*, *G6*, *1360*, *transib2*, *Tc1*, *baggins*,
150 and *roo*. In the African populations this is similar, *INE-1*, *Cr1a*, *G6*, *baggins*, *R1A1-element*,
151 *diver2*, *X-element*, and *HB*, followed by *1316* and *roo*.
152 *Stalker4, Stalker*, *Bari2*, *Tc3*, *G7*, and *Tart-C* were never present as more than a fraction
153 of an element in any individual, and in general coverage was limited to small regions within
154 these elements. This suggests that if these elements are present they are old and degraded. *G3*
155 and *hopper2* are estimated as being present in ~1 copy per individual in both populations,
156 however closer inspection reveals that all reads are mapping to the start and end of the elements.
157 This suggests that the only copy or copies present have internal deletions. For the *G-element* all
158 but a small fraction of reads map to one 140 bp sequence, but they map abundantly to that
159 particular sequence, such that the copy number is estimated as close to 6 per individual. Again,

160 reads which map equally well to more than one location were filtered out, thus this does not
161 represent non-specific mapping to repetitive elements. These patterns are not seen in the other
162 transposable elements, where coverage is even across the length of the element (Supplementary
163 Figures 1,2,4,5 as examples)

164 The *D. melanogaster pogo* and *Helitron* elements were not present in these populations,
165 which has been previously noted, suggesting that these transposable elements are not present in
166 *D. simulans* (Kofler *et al.*, 2015b). A full-length version of *Quasimodo* (two copies) and *gypsy6*
167 (one copy) were present in one individual. In other individuals reads map to very specific small
168 regions of the transposable elements, suggesting old and degraded copies. *Stalker3* is also
169 present in one genotype as a full-length copy. There is no mapping in the African populations
170 and very little in Californian populations, suggesting that old or degraded copies are not present
171 in the rest of the genotypes sampled.

172 Tajima's *D* of transposable elements is not intended as a measure of selection, given that
173 the reads are coming from different places in the genome. However, a large negative Tajima's *D*
174 might indicate recent activity of a single copy of the transposable element, as the copies have
175 more low frequency SNPs. The most negative Tajima's D in California belonged to *G6*, *TAHRE*,
176 *Juan*, *gypsy7*, *flea*, *Bari1* and *frogger* (less than -1.5). In African *D. simulans* this includes
177 *Tabor*, *Transpac*, *flea*, *Juan*, *Bari1*, *G6*, and *accord.* Thus in both populations *G6*, *flea*, *Bari1*,
178 and *Juan* likely have recent activity (Supplementary Figure 1 & 2). This is consistent with other
179 work on *Juan*, which suggests it is actively transposing in the species (Kofler *et al.*, 2015b). Note
180 that African *D. simulans* have a largely neutral genomic Tajima's *D* when split between the two
181 populations or considered together, while Californian *D. simulans* is elevated genome-wide
182 ((Signor *et al.*, 2017), Supplementary Figure 3). An elevated Tajima's *D* could indicate a number

183 of things, such as older invasions or multiple, diverged active elements in the populations. The
184 majority of differences in mean and variance in the population are due to transposable elements
185 with a positive Tajima's *D*, suggesting that these differences are due to different rates of gain and
186 loss of abundant transposable elements over time or low level spread of multiple diverged
187 copies.

188 *The p-element*

189 The *p-element* recently invaded *D. simulans* from *D. melanogaster* as described in (Kofler *et al.*,
190 2015a), however Pool-seq cannot tie *p-element* insertions to specific individuals, only determine
191 the average number of insertions. What was reported previously was 0.4 insertions in Florida
192 populations, and 29 in South Africa (Kofler *et al.*, 2015a). What we see in the California
193 population is an average of 2 insertions, however that is because the majority of individuals do
194 not have any insertions (137 individuals have less than .3 estimated copies, Figure 1). The
195 remaining individuals have between .5 and 39 copies. We recovered the single reported SNP in
196 the *p-element*, a G to A transition at position 2040. It is interesting that it is not invading
197 genotypes in the population at the same rate, but rather reaching high copy number in some
198 genotypes and not others (Nuzhdin 2000). However, given the observed 'outliers' (see next
199 section), one genotype for a given transposable element with high copy number, this may be
200 reflective of how some transposable elements invade populations. It is also possible that *p-*
201 *elements* are just proliferating in strains that were infected prior to collection (Nuzhdin &
202 Mackay 1997). This was observed previously in lab strains of *D. melanogaster*, though
203 contamination and introgression may also have played a role (Rahman *et al.* 2015).

204 *Transposable elements in individual genomes*

205 Some transposable elements have considerably higher copy number in particular genotypes
206 compared to the population average. For example, in one genotype *Dsim\ninja* is present in 29
207 copies, compared to the population mean of three. *Dsim\ninja* has 10 fixed differences and 27
208 polymorphisms in this strain from the California population, and the population average is 7.5
209 fixed differences and 264 polymorphisms (Supplementary Figure 4 & 5). This suggests that
210 *Dsim\ninja* was recently active in this genome. This is true of several transposable elements
211 which have outliers in the population. *Stalker2* has an outlier with 17 fixed SNPs and 8
212 polymorphisms, compared to a population average of 14 fixed SNPs and 43 polymorphisms. The
213 *gypsy10* element has a single outlier individual with 203 fixed differences compared to *D.*
214 *melanogaster*, and 77 polymorphisms. The population mean is 158 fixed differences and 222
215 polymorphisms. Other transposable elements with large outliers in the California population
216 include *opus*, *blood*, *GATE*, *diver*, *Tabor*, *INE-1*, *diver2*, *idefix*, *1731*, *412*, and *297*.
217 Sampling of the African populations was much more limited thus less genotype-specific
218 variation is sampled, and indeed only two transposable elements had large outliers, both in the
219 same individual from Madagascar: *copia*, and *diver*. This individual had 11 fixed differences in
220 *copia*, and 20 polymorphisms, while the population average is 10 fixed differences and 78
221 polymorphisms (and 20 copies compared to 4-11 for the rest of the population). For *diver* this
222 individual had 20 fixed differences and 84 polymorphisms, compared to a population average of
223 12 fixed differences and 215 polymorphisms (and 30 copies compared to 4-10 for the rest of the
224 population). The outlier individual is one that was inbred in the lab. In general being inbred in
225 the lab is not affecting overall transposable element copy number however, as comparing
226 between lines that were sequenced directly upon collection and those that there inbred, there is
227 no significant difference between the mean number of transposable elements for any

228 transposable element family. All of this is to suggest that transposable element load varies
229 significantly between genotypes within a population, though which transposable elements are
230 able to 'wake up' and multiply may be shared between genotypes in different populations
231 (*diver*). That is to say that some recent transposable element activity is shared across the species
232 (*G6*), populations (*p-element*) and some is specific to genotypes within a population. Those that
233 'wake up' in individual lines appears to be due to sampling of individuals that are permissive or
234 contain active transposable elements, rather than an overall increase in transposable element
235 activity in inbred lines.

236 *Comparison to other studies*

237 Previously *HMS Beagle*, *blood*, *flea*, *opus*, *Stalker*, and *F-element* were described as being
238 present in 0-1 copies in most populations from a large survey, with some populations having a
239 significant number of insertions (Biémont *et al.*, 1999). In our populations all of these elements
240 with the exception of *Stalker* were present in on average more than six insertions. Of these only
241 *flea* has more low frequency polymorphisms suggesting that it has spread within the genome
242 recently (CA, Tajima's D -1.7). (Vieira *et al.*, 1999) detected *17.6* at one insertion site in *D.*
243 *simulans*, and here it was present on average in 1 (CA) to 0.5 (AF) copies in these populations,
244 which is more frequent than recent papers such as (Kofler *et al.*, 2015b) which detected four
245 copies in the population of ~800 which was sampled. (Vieira *et al.*, 1998) estimated 0-30 copies
246 of *412* and 20-100 copies of *roo* per genome including heterochromatin. We observed 5-42
247 copies of *412* in both populations, with a single outlier at 73. For *roo* we observed 20-75 copies
248 in both populations. Thus our estimates are consistent with previous estimates for these elements
249 that include heterochromatin. It has also been suggested that *412* is recently active (Biémont *et*
250 *al.*, 1999).

251 *Tirant* has previously been reported as having higher copy number in African *D.*
252 *simulans*, potentially due to a recent mobilization of the element (Fablet *et al.*, 2006). We find
253 that pattern here, including a higher variance in the African populations where copy number
254 ranges from 2 to 6.68, compared to 2-3.8 in California (Fablet *et al.*, 2006). Viera et al. 1999
255 sampled diverse *D. simulans* populations and found ~1 *Tirant* element per genome, though this
256 was limited to euchromatic regions and older *Tirant* insertions may potentially have accumulated
257 in heterochromatic regions. It has previously been suggested than *Dmau\mariner* activity is
258 higher in populations that have spread out of Africa, though here the *Dmau\mariner* element has
259 a higher copy number in Africa than in the Californian *D. simulans*, from 0-5 with an average of
260 2.33, compared to 0-3 with an average of 1.22 (Picot *et al.* 2008). For *mariner2* we find copy
261 number to be relatively stable in both populations (9.6 in California, 7.3 in Africa), while in
262 Kofler et al. 2015 *mariner2* is only reported in six copies in the entire population of ~800,
263 though again this is limited to euchromatic regions.
264 The *G6* element has a large difference from previously reported values, with an average
265 of 66 insertions in Californian *D. simulans* and 69 in African. However only 37 insertions where
266 reported total for a previously estimated population of ~800 *D. simulans* isofemale lines (Kofler
267 *et al.*, 2015b). Only 18% of reads in Kofler et al. (2015) map to euchromatin, but this is not a
268 large enough fraction to explain the difference. In addition, in the populations reported here the
269 *G6* element has primarily low frequency polymorphisms (Tajima's D = -2.4), suggesting that this
270 is a recent expansion of copy number. The population examined in Kofler et al. (2015) is from
271 South African in 2013, a later collection date than either population here, though it is possible
272 that the *G6* activation is geographically limited to more northern populations. Overall our
273 estimates are higher than the work of Kofler et al. 2015, which only estimates more than one

insertion per line for four transposable elements (*1360*, *hobo*, *roo* and *Tc-2*). This is conservative compared to other estimates of euchromatic insertions, for example Vieira et al. 1999 use *in situ* hybridization to estimate copy number and found more than one insertion on average for nineteen elements (of 36 measured, *412*, *blood*, *burdock*, *copia*, *coral*, *flea*, *gypsy*, *HMS beagle*, *mdg3*, *opus*, *roo*, *Tirant*, *dor*, *F-element*, *Helena*, *I-element*, *jockey*, *Bari1*, and *hobo*). There is other evidence this work may be conservative, for example (Kaminker *et al.*, 2002) estimated 146 *roo* insertions in the euchromatin of the *D. melanogaster* reference genome, while (Kofler *et al.*, 2015b) estimates 1966 total for a population of 554 (approximately 3.6 per individual). This may also be due to inherent variability between populations.

Conclusions

Transposable element load is an attribute of species, populations, and individual genotypes. Elements which were not active in most of the population wake up in particular genotypes and spread rapidly, other elements reach high abundance across populations but were not noted in previous studies as being particularly prominent (*G6*). In inbred laboratory lines active transposable element copies may be inherited by some lines and not others, and active transposable elements can accumulate over time (Nuzhdin *et al.*, 1997). This can cause differences over time in the number of insertions within a line and large differences between lines in transposable element copy number (Nuzhdin *et al.*, 1997). This may also be reflective of natural patterns in which transposable elements proliferate in particular genotypes rather than at low levels in the population as a whole (Nuzhdin, 2000). Now that the Californian *D. simulans* lines have been held in the laboratory for eight years it would be interesting to see how the genotype-specific patterns of transposable element insertions have continued to change. Overall,

looking at variance between individuals is an important part of understanding the ways in which transposable elements maintain themselves in populations.

Acknowledgements: I would like to thank C. and S. Emery for helpful commentary on the manuscript. I am also thankful to J. Butler and T. Robert for help in the laboratory.

The author declares that she has no conflicting interests.

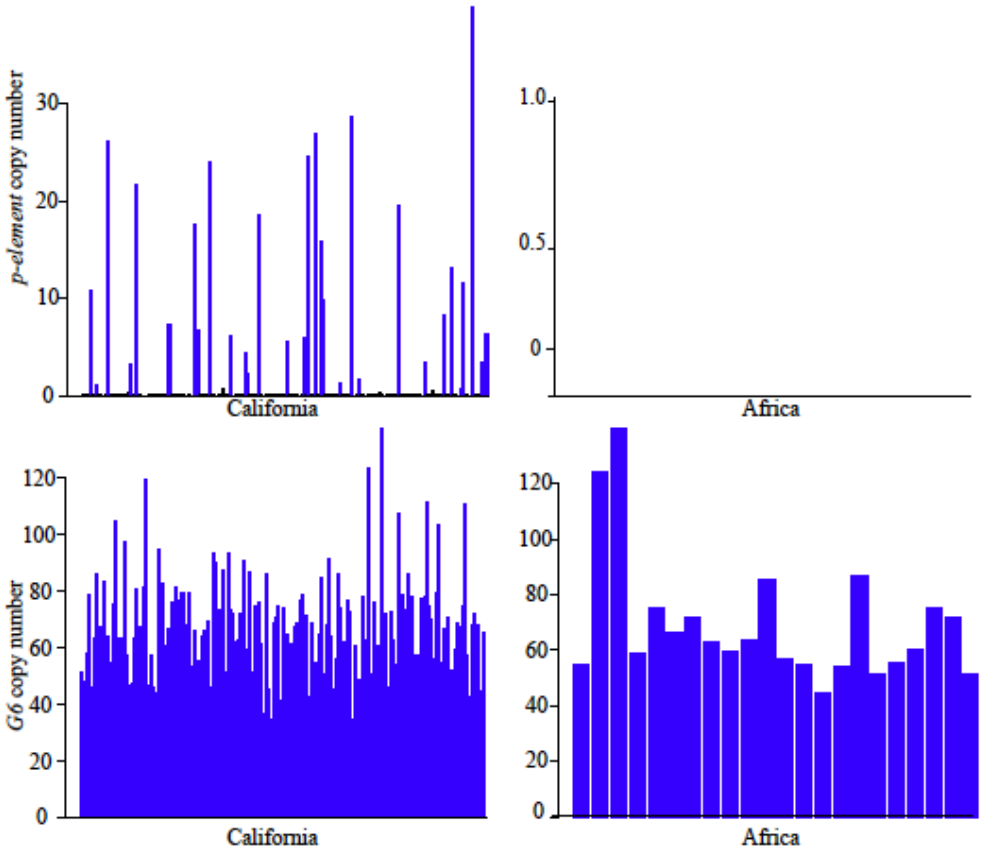

Figure 1: Estimated copy number for the *p-element* and *G6* in Californian and African *D. simulans*. Each bar represents an individual from the population. As expected, the *p-element* was

not found in the African population sampled in the early 2000's, but by 2012 when the Californian *D. simulans* was sampled it had invaded. *G6* has a high copy number in both populations of *D. simulans*, which wasn't recorded in previous studies on African *D. simulans*.

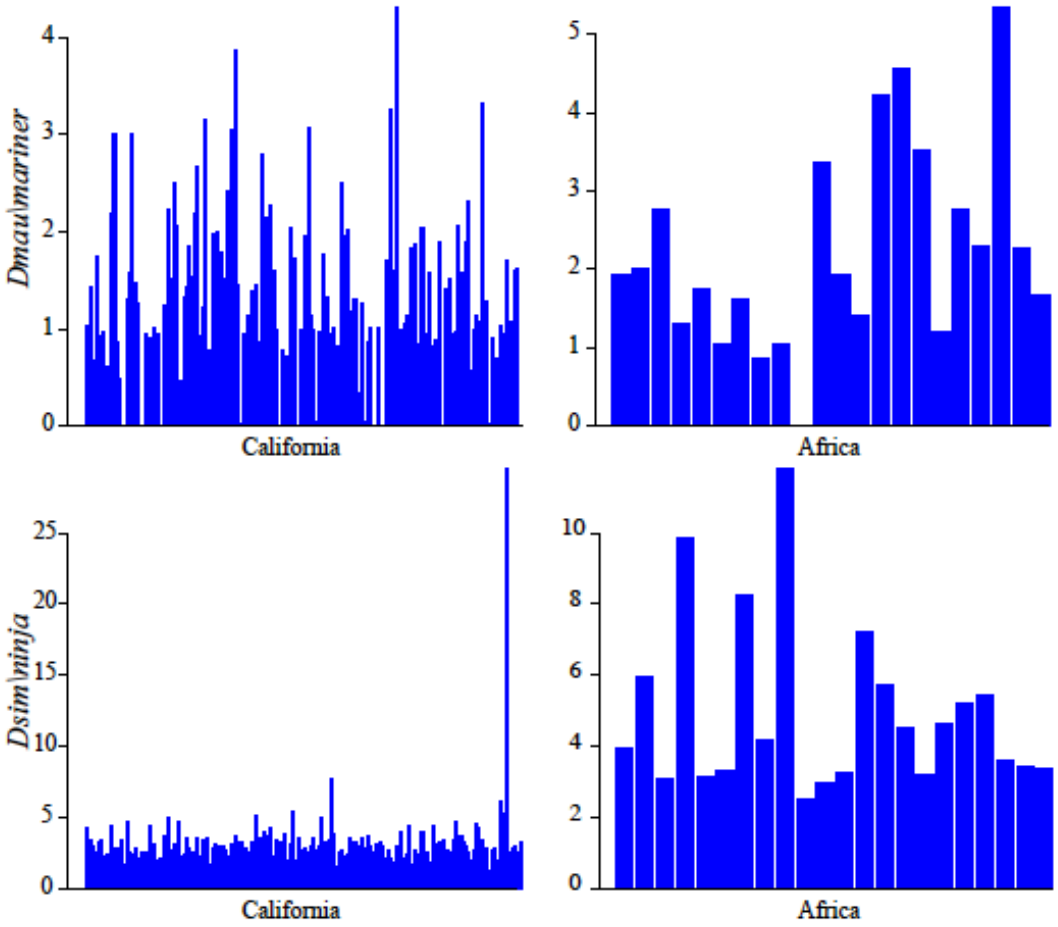

Figure 2: Estimated copy number for the two non-melanogaster transposable elements included here, Dsim\ninja and Dmau\mariner. Variance between individual genomes is large for transposable elements, for example in California Dsim\ninja varies from 1-29, and in Africa from 2-12. The individual with 29 copies of Dsim\ninja has 10 fixed SNPs within the element and 27 polymorphisms compared to other population members (for example, 239 polymorphisms and 5 fixed SNPs), suggesting that it has been recently active within that genome.

Element	CA Average	Variance	AF Average	Variance	Element	CA Average	Variance	AF Average	Variance	Element	CA Average	Variance	AF Average	Variance
297	18.44	7.87	12.11	8.31	*GATE*	26.98	12.73	28.32	13.07	*NOF*	1.04	0.21	0.97	0.32
412	27.78	31.66	14.77	5.39	*gtwin*	4.85	0.73	4.78	0.87	*opus*	7.64	5.03	6.85	9.54
1360	64.03	37.76	35.07	20.84	*gypsy*	4.97	1.06	5.92	2.02	*Osvaldo*	5.44	0.85	6.96	1.38
1731	6.69	6.86	7.10	3.65	*gypsy10*	8.37	5.13	6.34	3.94	*P-element*	2.36	44.92	-	-
17.6.	1.10	0.13	0.50	0.00	*gypsy11*	3.73	0.29	2.42	0.08	*pogo*	-	-	-	-
3S18	11.89	5.66	12.59	4.38	*gypsy12*	3.97	0.75	4.51	0.86	*Porto1*	10.02	0.65	7.49	0.39
accord	6.05	10.35	2.54	1.17	*gypsy2*	0.95	0.02	0.81	0.01	*Q-element*	0.95	0.03	0.72	0.01
accord2	4.84	0.81	3.23	0.41	*gypsy3*	3.45	0.95	3.08	0.73	*Quasimodo*	0.19	0.03	0.08	0.00
aurora-element	0.70	0.02	0.89	0.02	*gypsy4*	2.07	0.04	1.98	0.08	*R1-2*	1.13	0.05	1.21	0.11
baggins	52.65	15.51	62.04	42.68	*gypsy5*	3.68	1.12	2.26	0.52	*R1A1-element*	34.07	72.24	49.39	301.81
Bari1	4.92	4.58	18.05	30.07	*gypsy6*	0.39	0.01	0.31	0.00	*R2-element*	39.52	65.11	30.05	74.21
Bari2	0.12	0.00	0.11	0.00	*gypsy7*	0.86	0.01	0.64	0.00	*roo*	50.89	66.75	33.38	43.96
blood	12.12	11.30	7.98	10.57	*gypsy8*	4.77	0.69	6.38	1.95	*rooA*	19.29	1.30	18.61	1.19
BS	2.93	0.27	2.94	0.17	*gypsy9*	3.14	0.24	2.46	0.23	*rover*	1.24	0.15	0.58	0.03
BS3	6.41	0.63	6.95	0.51	*HB*	43.04	20.11	40.52	32.24	*Rt1a*	3.61	0.26	3.91	0.38
BS4	3.31	0.13	2.93	0.07	*Helena*	29.20	17.08	22.24	10.19	*Rt1b*	17.43	6.41	24.71	8.67
Burdock	12.73	4.35	9.27	3.04	*Helitron*	-	-	-	-	*Rt1c*	20.60	14.16	27.78	41.37
Circe	6.56	1.79	5.88	2.80	*HeT-A*	6.52	4.70	1.50	1.59	*S-element*	1.41	0.06	0.83	0.05
copia	9.22	4.00	6.52	11.85	*HMS-Beagle*	13.44	2.59	10.49	2.17	*S2*	0.73	0.01	0.57	0.00
Cr1a	125.69	88.15	97.02	39.36	*HMS-Beagle2*	0.82	0.01	0.69	0.00	*springer*	2.10	0.10	2.11	0.09
diver	6.75	8.19	8.14	26.95	*hobo*	32.87	12.41	31.79	32.23	*Stalker*	0.28	0.00	0.19	0.00
diver2	35.58	21.62	44.05	32.27	*hopper*	23.58	16.70	15.88	11.24	*Stalker2*	2.86	3.49	2.31	1.14
Dm88	8.43	0.45	8.34	0.34	*hopper2*	0.77	0.06	0.66	0.05	*Stalker3*	0.08	0.03	-	-
Dmau\mariner	1.22	0.72	2.33	1.54	*I-element*	37.33	35.49	21.37	13.78	*Stalker4*	0.09	0.00	0.06	0.00
Doc	25.20	24.77	23.59	37.24	*Idefix*	3.70	1.91	2.64	0.57	*Tabor*	9.95	44.94	5.81	3.25
Doc2-element	30.43	6.53	28.56	8.46	*INE-1*	193.45	296.57	116.27	104.78	*TAHRE*	3.09	1.10	1.01	1.35
Doc3-element	20.57	1.17	19.21	1.48	*invader1*	4.62	0.11	4.55	0.18	*TART-A*	1.05	0.24	0.54	0.09
Doc4-element	3.68	0.25	2.99	0.07	*invader2*	6.09	1.50	4.29	0.92	*TART-B*	0.94	0.20	0.36	0.15
Dsim\ninja	3.14	4.99	4.93	5.85	*invader3*	6.71	1.93	6.47	5.80	*TART-C*	0.02	0.00	0.01	0.00
F-element	28.41	11.23	19.78	1.48	*invader4*	3.24	0.08	3.07	0.06	*Tc1*	54.48	36.25	31.21	10.81
FB	22.44	20.74	13.75	14.15	*invader5*	0.66	0.01	0.56	0.00	*Tc1-2*	12.45	2.47	8.46	1.85
flea	13.71	10.09	10.85	13.78	*invader6*	5.85	1.55	6.20	3.67	*Tc3*	0.15	0.00	0.06	0.00
frogger	1.05	0.06	0.89	0.01	*Ivk*	14.98	2.26	12.87	2.72	*Tirant*	2.77	0.06	3.58	2.17
Fw2	6.50	0.20	5.64	0.14	*jockey*	8.73	3.10	11.61	4.33	*Tom1*	26.03	6.41	17.00	7.94
Fw3	3.01	0.07	2.89	0.16	*jockey2*	5.55	0.15	4.68	0.04	*transib1*	19.12	8.84	12.41	4.99
G-element	4.22	0.93	4.08	1.26	*Juan*	21.46	27.13	18.68	13.75	*transib2*	55.12	44.06	27.48	39.43
G2	25.99	15.12	23.93	16.78	*looper1*	2.24	0.11	1.56	0.07	*transib3*	2.81	0.07	2.07	0.07
G3	0.91	0.01	0.76	0.01	*mariner2*	9.61	0.54	7.34	0.28	*transib4*	4.14	0.17	2.80	0.16
G4	4.10	0.23	3.52	0.06	*Max-element*	29.78	17.50	32.43	15.97	*Transpac*	8.21	4.81	5.35	4.09
G5	15.05	2.32	15.99	2.77	*McClintock*	13.49	8.00	6.86	2.81	*X-element*	34.10	8.98	41.49	23.77
G5A	10.93	0.66	11.67	2.70	*mdg1*	6.43	1.00	4.95	0.85	*Xanthias*	9.43	3.14	8.90	6.90
G6	66.62	309.64	69.36	532.88	*mdg3*	7.02	4.73	5.94	3.01	*ZAM*	2.92	0.12	2.01	0.23
G7	0.18	0.00	0.20	0.01	*micropia*	8.01	4.05	6.53	2.71					

Table 1: Summary statistics for copy number of transposable elements from the Californian and African *D. simulans*.

Aminetzach YT, Macpherson JM, Petrov DA. 2005. Pesticide resistance via transposition-

mediated adaptive gene truncation in *Drosophila*. *Science* **309**:764–7.

Barrón, M.G., Fiston-Lavier, A.-S., Petrov, D.A. & González, J. 2014. Population genomics of transposable elements in *Drosophila*. *Annu. Rev. Genet.* **48**: 561–581.

Bennetzen, J.L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**: 251–269.

Biémont, C., Vieira, C., Borie, N. & Lepetit, D. 1999. Transposable elements and genome evolution: the case of *Drosophila simulans*. *Genetica* **107**: 113–120.

C Vieira, C.B. 2008. Selection against transposable elements in *D. simulans* and *D. melanogaster*. *Genet. Res.* **68**: 9-115

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Dimitri, P. 2003. Colonization of heterochromatic genes by transposable elements in *Drosophila*. *Mol. Biol. Evol.* **20**: 503–512.

Fablet, M., McDonald, J.F., Biémont, C. & Vieira, C. 2006. Ongoing loss of the *tirant* transposable element in natural populations of *Drosophila simulans*. *Gene* **375**: 54–62.

Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**: 397–405.

Hill, T. & Betancourt, A.J. 2018. Extensive exchange of transposable elements in the *Drosophila pseudoobscura* group. 1–14. Mobile DNA.

Hill, T., Schlötterer, C. & Betancourt, A. 2015. Hybrid dysgenesis in *Drosophila simulans* due to a rapid global invasion of the P- element. *PLoS* **12**: e1005920.

Jackson, B.C., Campos, J.L., Haddrill, P.R., Charlesworth, B. & Zeng, K. 2017. Variation in the intensity of selection on codon bias over time causes contrasting patterns of base composition evolution in *Drosophila*. *Genome Biol. Evol.* evw291–22.

Jakšić, A.M., Kofler, R. & Schlötterer, C. 2017. Regulation of transposable elements: interplay between TE-encoded regulatory sequences and host-specific trans-acting factors in *Drosophila melanogaster*. *Mol. Ecol*. **26**: 5149-5159.

Johnson, N.A. 2010. Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends Genet.* **26**: 317–325.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., *et al.* 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biology* **3**: research0084.1.

Kofler, R. & Schlötterer, C. 2015. Low levels of transposable element activity in *Drosophila mauritiana*: causes and consequences. *bioRxiv*.

364 Kofler, R., Hill, T., Nolte, V., Betancourt, A.J. & Schlötterer, C. 2015a. The recent invasion of
365 natural *Drosophila simulans* populations by the *P-element*. *Proc. Nat. Acad. Sci. USA* **112**:
366 6659–6663.

367 Kofler, R., Nolte, V. & Schlötterer, C. 2015b. Tempo and mode of transposable element activity
368 in *Drosophila*. *PLoS Genet.* **11**: e1005406.

369 Kofler, R., Senti, K.-A., Nolte, V., Tobler, R. & Schlötterer, C. 2018. Molecular dissection of a
370 natural transposable element invasion. *Genome Res.* **28**: 824–835.

371 Lee, Y.C.G. & Langley, C.H. 2012. Long-term and short-term evolutionary impacts of
372 transposable elements on *Drosophila*. *Genetics* **192**: 1411–1432.

373 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* 2009. The Sequence
374 Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

375 Li, H. 2015. Aligning sequence reads, clone sequences and assemble contigs with BWA-MEM
376 *arXiv* 1–3.

377 Long, A.D., Lyman, R.F., Morgan, A.H., Langley, C.H., Mackay, T.F. 2000. Both naturally
378 occurring insertions of transposable elements and intermediate frequency polymorphisms at
379 the *achaete-scute* complex are associated with variation in bristle number in *Drosophila*
380 *melanogaster*. *Genetics* **154**:1255–69.

381 Mackay, T.F.C. 1984. Jumping genes meet abdominal bristles: hybrid dysgenesis-induced
382 quantitative variation in *Drosophila melanogaster*. *Genet. Res.* **44**: 231-237.
383
384 Mackay, T.F.C. 1989. Transposable elements and fitness in *Drosophila melanogaster*. *Genome*
385 **31**: 284-295.
386
387 Mackay, T.F., Lyman, R.F., Jackson, M.S., 1992. Effects of P element insertions on quantitative
388 traits in *Drosophila melanogaster*. *Genetics* **130**: 315-332.
389
390 Picot ,S., Wallau, G.L., Loreto, E.L.S., Heredia, F.O., Hua-Van A., Capy P. 2008. The *mariner*
391 transposable element in natural populations of *Drosophila simulans*. *Heredity* **101**: 53–9.
392
393 Rahman R., Chirn, G.-W., Kanodia, A., Sytnikova, Y.A., Brembs, B., Bergman, C.M., *et al.*
394 2015. Unique transposon landscapes are pervasive across *Drosophila melanogaster*. *Nucleic*
395 *Acids Res.* **43**:10655–72.
396
397 Mateo, L., Ullastres, A. & González, J. 2014. A Transposable element insertion confers
398 xenobiotic resistance in *Drosophila*. *PLoS Genet.* **10**: e1004560.

399 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* 2010.
400 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
401 sequencing data. *Genome Res.* **20**: 1297–1303.

402 Nuzhdin, S.V. 2000. Sure facts, speculations, and open questions about the evolution of
403 transposable element copy number. In: *Transposable Elements and Genome Evolution* (J. F.
404 McDonald, ed), pp. 129–137. Springer Netherlands, Dordrecht.

405 Nuzhdin, S.V., Pasyukova, E.G. & Mackay, T.F. 1997. Accumulation of transposable elements
406 in laboratory lines of *Drosophila melanogaster*. *Genetica* **100**: 167–175.

407 Pasyukova, E.G. 2004. Accumulation of transposable elements in the genome of *Drosophila*
408 *melanogaster* is associated with a decrease in fitness. *Journal of Heredity* **95**: 284–290.

409 Romero-Soriano, V. & Garcia Guerreiro, M.P. 2016. Expression of the retrotransposon *Helena*
410 reveals a complex pattern of TE deregulation in *Drosophila* hybrids. *PLoS ONE* **11**:
411 e0147903.

412 Signor, S. & Nuzhdin, S. 2018. Dynamic changes in gene expression and alternative splicing
413 mediate the response to acute alcohol exposure in *Drosophila melanogaster*. *Heredity* **121:**
414 342-360.

415 Signor, S.A. & Nuzhdin, S.V. 2019. Evolution of phenotypic plasticity in response to ethanol
416 between sister species with different ecological histories (*Drosophila melanogaster and D.*
417 *simulans*). *bioRxiv*.

418 Signor, S.A., New, F.N. & Nuzhdin, S. 2017. A large panel of *Drosophila simulans* reveals an
419 abundance of common variants. *Genome Biol. Evol.* **10**: 189–206.

420 Shrimpton, A.E., Mackay, T.F., Brown, A.J. 1990. Transposable element-induced response to
421 artificial selection in *Drosophila melanogaster*: molecular analysis of selection lines.
422 *Genetics* **125**:803–11.

423 Slotkin, R.K. & Martienssen, R. 2007. Transposable elements and the epigenetic regulation of
424 the genome. *Nat. Rev. Genet.* **8**: 272–285.

425 Srivastav, S.P. & Kelleher, E.S. 2017. Paternal induction of hybrid dysgenesis in *Drosophila*
426 *melanogaster* is weakly correlated with both *P-element* and *hobo* element dosage. *G3* **7**:
427 1487–1497.

428 Tenaillon, M.I., Hollister, J.D. & Gaut, B.S. 2010. A triptych of the evolution of plant
429 transposable elements. *Trends Plant Sci.* **15**: 471–478.

430 Vieira, C. & Biémont, C. 2004. Transposable element dynamics in two sibling species:
431 *Drosophila melanogaster* and *Drosophila simulans*. *Genetica* **120**: 115–123.

432 Vieira, C., Aubry, P., Lepetit, D. & Biémont, C. 1998. A temperature cline in copy number for
433 *412* but not *roo*/B104 retrotransposons in populations of *Drosophila simulans*. *Proc. R. Soc.*
434 *B* **265**: 1161–1165.

435 Vieira, C., Lepetit, D., Dumont, S. & Biémont, C. 1999. Wake up of transposable elements
436 following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol.* **16**: 1251–1255.

437 Yang, H.-P. & Nuzhdin, S.V. 2003. Fitness costs of *Doc* expression are insufficient to stabilize
438 its copy number in *Drosophila melanogaster. Mol. Biol. Evol.* **20**: 800–804.

439