

Mobile DNA

Phenotypic and genomic analysis of P elements in natural populations of *Drosophila melanogaster*

--Manuscript Draft--

Manuscript Number:								
Full Title:	Phenotypic and genomic analysis of P elements in natural populations of <i>Drosophila melanogaster</i>							
Article Type:	Research							
Funding Information:	<table border="1"> <tr> <td>Wellcome Trust (096602/B/11/Z)</td> <td>Dr Michael G. Nelson</td> </tr> <tr> <td>Human Frontier Science Program Young Investigator (RGY0093/2012)</td> <td>Dr Casey Bergman</td> </tr> </table>	Wellcome Trust (096602/B/11/Z)	Dr Michael G. Nelson	Human Frontier Science Program Young Investigator (RGY0093/2012)	Dr Casey Bergman			
Wellcome Trust (096602/B/11/Z)	Dr Michael G. Nelson							
Human Frontier Science Program Young Investigator (RGY0093/2012)	Dr Casey Bergman							
Abstract:	<p>The <i>Drosophila melanogaster</i> P transposable element provides one of the best cases of horizontal transfer of a mobile DNA sequence in eukaryotes. Invasion of natural populations by the P element has led to a syndrome of phenotypes known as "P-M hybrid dysgenesis" that emerges when strains differing in their P element composition mate and produce offspring. Despite extensive research on many aspects of P element biology, questions remain about the stability and genomic basis of variation in P-M dysgenesis phenotypes. Here we report the P-M status for a number of populations sampled recently from Ukraine that appear to be undergoing a shift in their P element composition. Gondal dysgenesis assays reveal that Ukrainian populations of <i>D. melanogaster</i> are currently dominated by the P' cytotype, a cytotype that was previously thought to be rare in nature, suggesting that a new active form of the P element has recently spread in this region. We also compared gondal dysgenesis phenotypes and genomic P element predictions for isofemale strains obtained from three worldwide populations of <i>D. melanogaster</i> in order to guide further work on the molecular basis of differences in cytotype status across populations. We find that the number of euchromatic P elements per strain can vary significantly across populations but that total P element numbers are not strongly correlated with the degree of gondal dysgenesis. Our work shows that rapid changes in cytotype status can occur in natural populations of <i>D. melanogaster</i>, and informs future efforts to decode the genomic basis of geographic and temporal differences in P element induced phenotypes</p>							
Corresponding Author:	Iryna Kozeretska, Ph.D. D.Sc. Taras Shevchenko National University of Kyiv Kyiv, Ukraine UKRAINE							
Corresponding Author Secondary Information:								
Corresponding Author's Institution:	Taras Shevchenko National University of Kyiv							
Corresponding Author's Secondary Institution:								
First Author:	Iryna Kozeretska, Ph.D. D.Sc.							
First Author Secondary Information:								
Order of Authors:	<table border="1"> <tr> <td>Iryna Kozeretska, Ph.D. D.Sc.</td> </tr> <tr> <td>Vladyslav Bondarenko</td> </tr> <tr> <td>Vadym Shulga</td> </tr> <tr> <td>Svitlana Serga</td> </tr> <tr> <td>Andrii Rozhok</td> </tr> <tr> <td>Oleksandra Protsenko</td> </tr> <tr> <td>Michael G. Nelson</td> </tr> </table>	Iryna Kozeretska, Ph.D. D.Sc.	Vladyslav Bondarenko	Vadym Shulga	Svitlana Serga	Andrii Rozhok	Oleksandra Protsenko	Michael G. Nelson
Iryna Kozeretska, Ph.D. D.Sc.								
Vladyslav Bondarenko								
Vadym Shulga								
Svitlana Serga								
Andrii Rozhok								
Oleksandra Protsenko								
Michael G. Nelson								

	Casey Bergman
Order of Authors Secondary Information:	
Opposed Reviewers:	

[Click here to view linked References](#)

bioRxiv preprint doi: <https://doi.org/10.1101/047910>; this version posted April 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Phenotypic and genomic analysis of *P* elements in natural populations of *Drosophila melanogaster*

**Kozeretska I.A.*¹, Bondarenko V.¹, Shulga V.I.¹, Serga S.V.¹, Rozhok A.²
I., Protsenko A.V.¹, Nelson M.G.³, Bergman C.M.³**

1 – Taras Shevchenko National University of Kyiv, 01601, 64 Volodymyrska str, Kyiv, Ukraine

2 – University of Colorado, School of Medicine, Aurora, CO, USA

3 – Faculty of Life Sciences, University of Manchester, Manchester M21 0RG, United Kingdom

* Address for Correspondence:

iryna.kozeretska@gmail.com

Keywords: *P* element, *Drosophila melanogaster*, Uman', transposable elements, population genomics

Abstract

24
25
26 The *Drosophila melanogaster* *P* transposable element provides one of the best cases of horizontal
27 transfer of a mobile DNA sequence in eukaryotes. Invasion of natural populations by the *P* element
28 has led to a syndrome of phenotypes known as “P-M hybrid dysgenesis” that emerges when strains
29 differing in their *P* element composition mate and produce offspring. Despite extensive research on
30 many aspects of *P* element biology, questions remain about the stability and genomic basis of
31 variation in P-M dysgenesis phenotypes. Here we report the P-M status for a number of populations
32 sampled recently from Ukraine that appear to be undergoing a shift in their *P* element composition.
33 Gondal dysgenesis assays reveal that Ukrainian populations of *D. melanogaster* are currently
34 dominated by the P’ cytotype, a cytotype that was previously thought to be rare in nature,
35 suggesting that a new active form of the *P* element has recently spread in this region. We also
36 compared gondal dysgenesis phenotypes and genomic *P* element predictions for isofemale strains
37 obtained from three worldwide populations of *D. melanogaster* in order to guide further work on
38 the molecular basis of differences in cytotype status across populations. We find that the number of
39 euchromatic *P* elements per strain can vary significantly across populations but that total *P* element
40 numbers are not strongly correlated with the degree of gondal dysgenesis. Our work shows that
41 rapid changes in cytotype status can occur in natural populations of *D. melanogaster*, and informs
42 future efforts to decode the genomic basis of geographic and temporal differences in *P* element
43 induced phenotypes.

Introduction

44
45
1
246 A substantial portion of eukaryotic genomes is represented by transposable elements (TEs). These
3
447 TE families include those that colonized genomes long ago during the evolution of the host species
5
448 and groups, but also those that have appeared in their host genomes recently. The genome of
7
449 *Drosophila melanogaster* is thought to have been colonized by the *P* element family of transposons
9
50 at least 70 years ago as a result of a horizontal transmission event from *D. willistoni* [1, 2], a species
11
51 that inhabits South America, the Caribbean and southern parts of North America. Laboratory strains
12
52 of *D. melanogaster* established from wild populations before the 1950s did not contain *P* element,
14
53 while by the late 1970s this TE family was found in all populations worldwide [1]. The specific
16
54 molecular and evolutionary mechanisms that promoted the global spread of the *P* element over such
18
55 a short period are not still clearly understood.

20
56
22
57 The presence of *P* elements induces a number of phenotypes in *Drosophila* that can be
23
58 characterized by the so-called “P-M hybrid dysgenesis” assay [3]. Among the most prominent *P*
25
59 element induced phenotypes is gonadal dysgenesis (GD), which is the key marker determining P-M
27
60 status in particular strains of flies [1, 4–10]. In the P-M system, fly strains can be classified by their
29
61 phenotypes as follows: P-strains have both *P* element transposition inducing and repressing
31
62 abilities, P'-strains only have the inducing ability, Q-strains only have the repressing ability, and M-
32
63 strains have neither [3–4, 11]. An M-strain carrying some *P* sequences in the genome is called M'
34
64 to distinguish it from true M-strains, which are completely devoid of *P* elements [12]. Hybrid
36
65 dysgenesis phenotypes were previously thought to be controlled by repressor proteins of various
38
66 kinds that are encoded by *P* elements of different sizes [13]. More recent work has shown that these
40
67 phenotypes mostly arise due to RNA interference mechanisms mediated by piRNAs produced by
42
68 telomeric *P* elements, such as TP5 and TP6, and the effects are amplified by RNAs produced by
43
69 other *P* elements [14].
45
70

47
71 Natural populations of *D. melanogaster* are currently thought to have a relatively stable
49
72 distributions of *P* elements and P-M status. For example, Australian populations demonstrate a
51
73 north-south cline of the frequency of various *P* element copies [6], which underwent only minor
52
74 changes in the frequencies of truncated and full-size copies of the *P* element a decade later [13]. In
53
75 North American populations, P and Q cytotypes prevailed in the 1970s-1980s [1, 5]. Two decades
54
76 later, no significant changes were detected, except for an increase in the proportion of truncated *P*
55
77 element copies [15]. A similar phenotypic stability has also been observed in populations from
56
78 Eurasia, Australia, Japan and Taiwan [7-8, 16]. Cytotype status has also remained stable for over 30
58
79
60
80

79 to 40 years in some Japanese populations [9, 17–18]. Some authors propose that these protracted
80 periods of stability follow the initial transition from M cytotype to P, Q and M' cytotypes [9].

81
82 Studies of *P* element insertions at the molecular level provide the best opportunity to characterize
83 the dynamics of this TE in natural populations. Such analyses have been hampered by the lack of
84 unbiased and practical methods to characterize the frequencies of *P* element insertions at individual
85 sites across large numbers of strains. In the past, estimates of TE insertion frequencies were
86 obtained by *in situ* hybridization to polytene chromosomes [19–23]. However, the resolution of *in*
87 *situ* hybridization is limited, with highly truncated insertions being missed entirely and nearby
88 insertions not being distinguished from one another. More recent studies have used targeted PCR to
89 survey populations for TE insertions identified in the reference genome sequence [24–28]. These
90 surveys are necessarily biased towards insertions with high population frequencies [24], and are not
91 applicable in the case of the *P* element since the ISO-1 strain used for the reference genome
92 sequence is an M strain lacking any *P* elements [28].

93
94 Application of whole-genome next generation sequencing (NGS) technologies has improved many
95 areas of population genetics research. To date, hundreds of re-sequenced genomes of *D.*
96 *melanogaster* exist and can be freely used for population and genomic analyses [29–33]. Moreover,
97 a number of computational algorithms have been designed for *de novo* TE insertion discovery,
98 annotation, and population analysis in *Drosophila* [34–40]. Comparison of different methods for
99 detecting TEs in *Drosophila* NGS data has shown that they identify different subsets of TE
100 insertions [41]. Determining which TE detection method is best for specific biological applications
101 remains an area of active research.

102
103 Here we report the recent P-M status of a number of populations from Ukraine sampled in 2012–
104 2013. Previous studies of flies collected in Ukraine from 1936 through the 1980's revealed only M
105 and M' cytotypes [1, 42–43]. More recent collections in Ukraine from 2008 and 2009 revealed only
106 M' cytotypes [44], consistent with the stability in cytotype status observed in other regions around
107 the world. Here we present evidence that the P' cytotype is now prevalent in several locations in
108 Ukraine, suggesting a relatively rapid recent shift in the cytotype status of these populations. To
109 better understand the molecular basis for differences in cytotype status among populations, we also
110 compared GD phenotypes and genomic *P* element predictions for isofemale strains obtained from
111 three worldwide populations of *D. melanogaster*. This analysis revealed that the number of
112 euchromatic *P* elements per strain can vary significantly across populations but is not correlated
113 with the degree of GD. Finally, we show that ability to predict *P* elements in pooled population

115 differences in the number of *P* elements across populations using pool-seq must control for read
116 depth per strain. Our work shows that rapid changes in cytotype status can occur in natural
117 populations of *D. melanogaster*, and informs future efforts to decode the genomic basis of
118 differences in *P* element induced phenotypes over time and space.

7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

119

Materials and Methods

120

2

121

4

122

6

123

8

124

10

125

11

126

13

127

15

128

17

129

18

130

20

131

22

132

24

133

26

134

27

135

29

136

31

137

33

138

35

139

36

140

38

141

40

142

42

143

44

144

46

145

48

146

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

Sample collection. Flies were collected at fruit processing facilities and in apple orchards in late August to early September in 2012 and 2013 at seven locations in Ukraine: Uman' (48°45'45.26"N – 30°14'38.97"E), Varva (50°29'33.30"N – 32°42'50.93"E), Inkerman (44°61'77.64"N – 33°61'44.14"E), Chornobyl (51°16'13.73"N – 30°13'19.63"E), Kyiv (50°21'9.06"N – 30°28'57.70"E), Poliske (51°14'9.49"N – 29°23'51.60"E), Chornobyl Nuclear Power Plant (NPP) (51°22'30.01"N – 30°8'21.33"E). Single gravid females were used to establish isofemale strains. For each population, isofemale strains collected from one year were used for analysis.

Gonadal dysgenesis phenotyping. The cytotype of a particular isofemale strain was determined by the GD assay, for which two types of crosses were used: cross A (M-strain (Canton S) females x tested males) and A* (tested females x P-strain (Harwich) males) [4, 42]. Three A and three A* crosses were routinely made for each isofemale strain and 50 F1 females of each cross were dissected for gonadal status inspection. Both unilateral and bilateral ovary reduction was counted. The GD score for each isofemale strain was calculated using the formula $\%GD = \frac{1}{2}(\%GD_1) + \%GD_2$, where $\%GD_1$ stands for the percent of individuals with unilateral reduction of the ovary, and $\%GD_2$ means the percent of individuals with bilateral gonadal reduction. P-M characteristics were defined as in [5]. The result of GD tests for each population was shown as a two-dimensional graph, plotting A %GD versus A* %GD, following [45].

In addition to collecting GD phenotypes for populations from Ukraine, we also re-analyzed GD assay data from [46]. These cytotype data are for isofemale strains from three geographic regions: North America (Athens, Georgia, USA), Europe (Montpellier, France), and Africa (Accra, Ghana), described in [47]. Definitions of A and A* crosses in [46] are inverted relative to those proposed by [4], and were standardized prior to re-analysis here.

Genome-wide identification of P element insertions. Whole genome shotgun sequences described in [31] from the three populations analyzed by [46] [North America (Athens, Georgia, USA), Europe (Montpellier, France), and Africa (Accra, Ghana)] were downloaded from the European Nucleotide Archive (ERP009059). These data were collected using a uniform library preparation and sequencing strategy (thus mitigating many possible technical artifacts) and include data for both individual isofemale strains and pools of isofemale strains [31]. Two pool-seq samples were analyzed for each population [North America (15 and 30 strains), Europe (20 and 39 strains), and Africa (15 and 32 strains)].

154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265

P element insertions were identified by TEMP (revision d2500b904e2020d6a1075347b398525ede5feae1; [38]) and RetroSeq (revision 700d4f76a3b996686652866f2b81fetc6f0241e0; [48]) using the McClintock pipeline (revision 3ef173049360d99aaf7d13233f9d663691c73935; [49]). McClintock was run across the major chromosome arms (chr2L, chr2R, chr3L, chr3R, chr4, chrY, and chrX) of the dm6 version of the Release 6 reference genome [50] using the following options: -C -m "retroseq temp" -i -p 12 -b. Reference TE insertions needed for TEMP were generated automatically by McClintock using RepeatMasker (version open-4.0.6). The *D. melanogaster* TE library used by McClintock to predict reference and non-reference TE insertions is a slightly modified version of the Berkeley *Drosophila* Genome Project TE data set v9.4.1 described in [51, 52].

In addition to providing "raw" output for each component method in the standardized zero-based BED6 format, McClintock generates "filtered" output tailored for each method. TEMP output was filtered to (i) eliminate predictions where the start or end coordinates had negative values, (ii) retain predictions where there is sequence evidence supporting both ends of an insertion, and (iii) retain predictions that have a ratio of reads supporting the insertion to non-supporting reads of >10%. Likewise, RetroSeq predictions were filtered to (i) eliminate predictions where two different TE families shared the same coordinates and (ii) retain predictions assigned a call status of greater than or equal to six as defined by [53].

Graphical and statistical analyses were performed in the R programming environment (version 3.2.1).

Results

Cyotype status of natural populations from Ukraine. To determine cyotypes in natural populations of *D. melanogaster*, we carried out GD assays for 241 isofemale strains from 7 locations in Ukraine (Table 1, Fig. 1, File S1). We identified two main cyotypes in these populations, namely M [12] and P' [11]; strains with Q or P cyotype were not detected. Previous work suggests M-strains detected here are actually likely to be M' [44]. Interestingly, the majority of lines were of P' (86%; 207/241), a cyotype which has not been previously reported in this region and is thought to be rare in nature [1, 42–44]. Most populations demonstrated a higher proportion of P'-strains relative to M'-strains, except for the population from Inkerman, which was equally represented with P'-strains and M'-strains.

Differences in GD phenotypes for isofemale strains among Ukrainian populations were observed in both inducing ability (One-way ANOVA; $F=44.5$, 6 d.f., $P<2e-16$) and susceptibility (One-way ANOVA; $F=160.46$, 6 d.f., $P<2e-16$) (Fig. S1). The average inducing ability of a strain varied from 11.0% (Inkerman) to 51.1% (Poliske) in A crosses of the GD test. However, in most populations (Uman', Kyiv, Chornobyl, Varva), the inducing ability values were much closer to each other (21.4–26.4%). The greatest range in inducing ability within a population was observed from Chornobyl, in which the %GD varied from 4–65%.

The average susceptibility to *P* transposition varied from 24.4% (Inkerman) to 84.5% (Kyiv) in A* crosses of the GD test. In most populations (Uman', Kyiv, Chornobyl NPP, Poliske), the susceptibility values were very high (>73.7%), with Chornobyl and Varva having a more intermediate average susceptibility. We did not detect strains with susceptibility below 10%. For A* crosses, the greatest range in susceptibility was observed in the population from Chornobyl.

Overall, most populations from Ukraine demonstrated a moderate ability to induce *P* transposition, with moderate to high susceptibilities to *P* transposition. One exception to this rule is for the population from Inkerman, which demonstrated low inducing abilities and low susceptibility. In general, all the populations showed less variation in inducing ability (A crosses) relative to susceptibility (A* crosses). In addition, most strains from each population clustered together reasonably well in phenotypic space, with the exception of the Chornobyl population, which exhibited substantial variation for both GD phenotypes.

Comparison of cyotype status and P element insertions in individual strains from North

212 *America, Europe and Africa*
213
214 Answering questions about how cytotype status varies geographically and temporally requires a
215 better understanding of how genomic *P* element insertion profiles relate to *P* element induced
216 phenotypes. As a preliminary step towards this goal, we identified *P* element insertions in publicly
217 available genomic data [31] for a panel of 43 isofemale strains from three global regions with
218 previously-published GD phenotypes [46]. In contrast to the Ukrainian populations reported here,
219 isofemale strains from these populations were mainly P-strains, M-strains and Q-strains (Fig. 2,
220 Table 2, File S2). According to our genomic analysis below, strains that are defined phenotypically
221 M are actually M' strains. Additionally, strains from these populations do not differ significantly in
222 the degree of their inducing ability (One-way ANOVA; $F=0.06$, 2 d.f., $P=0.94$) or susceptibility
223 (One-way ANOVA; $F=1.66$, 2 d.f., $P=0.2$) (Fig. S2A–B).

224
225 We predicted *P* element insertions in the genomes of these strains using two independent
226 bioinformatics methods – TEMP and RetroSeq – to ensure that our conclusions are not dependent
227 on the idiosyncrasies of a single TE detection software package (File S2, File S3). We also
228 investigated the effects of the standard filtering of TEMP and RetroSeq output performed by
229 McClintock, a meta-pipeline that runs and parses multiple TE insertion detection methods. Across
230 all three populations, the numbers of *P* elements predicted per strain were well correlated across
231 strains, regardless of the method of analysis and filtering ($r \geq 0.712$) (Fig. S3). The correlation in the
232 number of *P* elements predicted per strain was highest for the filtered TEMP and filtered RetroSeq
233 datasets ($r=0.945$), suggesting that the post-processing steps performed by McClintock improve the
234 consistency of TE predictions made by these methods on isofemale strains. For filtered datasets, the
235 number of *P* elements predicted per strain was 52–134 for TEMP and 52–162 for RetroSeq.

236
237 We tested whether the number of *P* elements per strain is associated with GD phenotypes and found
238 neither cross A nor A* to be significantly linearly correlated with the filtered number of predictions
239 made by TEMP or RetroSeq ($p > 0.11$; Fig. 3). Similar results were obtained for the raw output of
240 these methods as well (Fig. S4). However, we did find evidence that strains from the North
241 American populations carry fewer euchromatic *P* element insertions in their genomes relative to the
242 European and African populations, regardless of the TE detection method and filtering (One-way
243 ANOVA; $F > 9.26$; 2 d.f., $P < 5e-4$) (Fig. 3; Fig. S2C–F). These results suggest that there is no simple
244 relationship between the number of euchromatic *P* elements and GD phenotypes at the level of
245 individual strains, and that systematic differences in the abundance of *P* elements per strain may not
246 lead to differences in the frequency of P-M phenotypes at the population level.

Population difference in *P* element insertion numbers can be observed in pool-seq samples.

Pooled-strain sequencing (pool-seq) is a cost-effective strategy to sample genomic variation across large numbers of strains and populations [54]. To address whether the differences among populations we observed in the number of *P* elements predicted in isofemale strain data are also seen in pool-seq data, we predicted *P* element insertions in pool-seq samples from the same populations (Table 3, File S3). Two pool-seq samples are available for each population that differ in the number of individuals/isofemale strains used: North America (n=15 and n=30), Europe (n=20 and n=39), and Africa (n=15 and n=32). The smaller pools from each population include one individual from the same isofemale strains analyzed above; the larger pools contain one individual from the same strains as the smaller pools, plus individuals from additional isofemale strains from the same population that were not sequenced as isofemale strains. Thus, the smaller pool samples are a nested subset of the larger pool samples, and pool-seq samples from the same population are not fully independent from one another. However, larger and smaller strain pool samples contain similar numbers of reads (~44 million read pairs per sample).

The numbers of *P* element insertions identified by TEMP and RetroSeq in these pool-seq samples are given in the Table 3. For the raw output, TEMP predicted more insertions in the larger strain pools relative to the smaller strain pools, as expected for a method designed to capture TE insertions that are polymorphic within a sample [38]. However, the McClintock-filtered TEMP output included many fewer insertions per strain pool than in the raw output, as well as fewer insertions in the larger strain pools relative to the smaller strain pools. These effects are likely because of the requirement for default McClintock filtering of TEMP predictions to have a ratio of reads supporting the insertion to non-supporting reads of >10%. McClintock filtering reduced the total number of RetroSeq predictions only by less than 2-fold, with fewer insertions being predicted for the larger strain pools in both raw and filtered RetroSeq output. Regardless of method or filtering, when the number of strains in a pool-seq sample was used as a scaling factor, fewer insertions were predicted per strain in the larger strain pools relative to smaller strain pools for all of the populations studied. Because smaller pools contain a subset of the same strains that are present in larger pools, these results suggest a dilution effect whereby, at a fixed sequencing coverage, *P* element insertions that are predicted in smaller pools cannot be detected in larger pools, even though they are in fact present in the sample.

281 In spite of this dilution effect, African pool-seq samples tended to have more insertions per strain
282 than North American samples as seen in the isofemale strain datasets. This result is most clearly
283 demonstrated for the comparison between North American and African samples with 15 strains
284 where African sample has more predicted insertions regardless of TE detection method and
285 filtering. These results suggest that, despite dilution effects, pool-seq samples can capture general
286 population trends in total *P* element insertion numbers seen in isofemale strain sequencing.

Discussion

We analyzed the P-M status of strains from seven natural populations of *D. melanogaster* from Ukraine. Strains from one of these populations (Uman') have a long history of study (starting from 1966) [1]. Another population from Magarach (Crimea) was represented with isofemale strains collected from the wild in 1936 [1, 42–43]. Rozhok et al (2009) most recently characterized the P-M status of nine populations from Ukraine and revealed only the M' cytotype, which is consistent with data from Europe [7]. Thus, flies collected in Ukraine in from the 1930s to 2009 appeared to have harbored only M or M' cytotypes. The results presented here indicate that Ukrainian populations of *D. melanogaster* are currently dominated by the P' cytotype, a cytotype that was previously thought to be rare in nature but has recently been found in the southern USA and Turkey [10, 15]. The sudden prevalence of the P' cytotype in these populations suggests that a new active form of the *P* element has recently spread in Ukraine.

Several studies before 2000 describe P-M status as phenotypically stable in populations from Japan, Eurasia, Africa, Oceania, and North America [1, 4–10]. The apparent rapid changes in the P-M status in Ukraine do not follow this general pattern, however the mechanisms underlying these changes remain unclear. It has been suggested that high activity of TEs may result from the stresses that accompany adaptation to new climates [55]. Such a response to stress has been suggested for populations with M' cytotype [15] which has previously been found in Ukrainian populations [44].

We also performed an exploratory analysis of genomic data from three worldwide populations of *D. melanogaster* with published GD phenotypes [31, 46–47]. The main aim of this analysis was to inform future computational and sampling strategies to decode the genomic basis of variation in *P* element induced phenotypes such as those observed in Ukraine. However, our results allowed us to draw several preliminary conclusions about the detection of *P* element insertions in *D. melanogaster* population genomic data and the genomic basis of GD phenotypes.

For samples derived from isofemale strains, we find that two different TE detection methods (TEMP and RetroSeq) generate well-correlated numbers of *P* element predictions per strain, but that filtering by McClintock improves the overall correlation between these methods (mainly by reducing the number of presumably false positive TEMP predictions). Analysis of pool-seq samples revealed larger differences between TE detection methods and a larger effect of McClintock filtering. Regardless of method or filtering, we found that there is a diminishing return on the number of *P* element insertions detected per individual/strain in pool-seq samples for a given

322 element across pool-seq samples unless the number of strains and/or the read depth is carefully
323 controlled. Diminishing returns per strain also implies there are allele frequency threshold effects
324 for calling *P* element insertions in NGS samples, which may also impact attempts to integrate
325 results from pool-seq data with those based on isofemale strains.
326

327
328 Regardless of prediction method, we found no simple linear correlation between the strength of GD
329 phenotypes and the number of euchromatic *P* element insertions across isofemale strains (Fig. 3).
330 Strains with very different number of euchromatic *P* elements appear to be able to generate very
331 similar phenotypic outcomes. These results may imply that GD phenotypes are determined by
332 specific full-length *P* element insertions that are active rather than overall numbers that include
333 many inactive defective copies, or that GD phenotypes are controlled by heterochromatic *P* element
334 insertions not detected here. Alternatively, the lack of correlation between the number of *P* element
335 insertions and GD phenotypes may imply that the genomic sequence data is not of sufficient depth
336 in these samples, or that the bioinformatic methods are not sensitive enough to detect all *P* element
337 insertions in an isofemale strain and introduce noise that obscures a true relationship.
338

339 We did however observe differences in the number of predicted *P* element insertions at the
340 population level (Fig. 3), even though we found no strong differences in the levels of GD
341 phenotypes across these populations. Specifically, the North American population had the fewest
342 predicted *P* element insertions, regardless of the TE detection method or filtering (Fig. 3, Fig. S2).
343 Evidence for fewer *P* element insertions per strain in the North American population could also be
344 detected in pool-seq samples, especially when the number of strains is controlled for (Table 3).
345 Thus it is possible to detect clear differences in euchromatic *P* element insertion profiles among
346 populations using either isofemale strains or pool-seq, however interpreting how these profiles
347 relate to P-M phenotypes at the strain or population level remains an open challenge.
348

349 In conclusion, based on the increase in a novel P' cytotope, we suggest that at the present time there
350 is a new *P* element type that is active in Ukrainian populations of *D. melanogaster*. Further analysis
351 of genomic data from these populations is needed to understand the dynamics of this invasion and
352 connect variation in GD phenotypes to *P* element insertions at the molecular level.
353

354 Acknowledgements

355 This work was supported by Wellcome Trust PhD Studentship (096602/B/11/Z) to MGN and a
356 Human Frontier Science Program Young Investigator grant RGY0093/2012 to CMB.
357

Authors declare no competing interests

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

358

References

1. Anxolabéhère D, Kidwell MG, Periquet G. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. *Mol Biol Evol.* 1988; 5(3): 252–69.
2. Daniels SB, Preston KR, Strausbaugh LD, Kidwell MG, Chovnick A. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. *Genetics.* 1990; 124(2): 339–55.
3. Kidwell MG, Kidwell JF, Sved JA. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics.* 1977; 86(4): 813–33.
4. Engels WR, Preston CR. Components of hybrid dysgenesis in a wild population of *Drosophila melanogaster*. *Genetics.* 1980; 95(1): 111–28.
5. Kidwell MG. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc Nat Acad Sci USA* 1983; 80(6): 1655-9.
6. Boussy IA, Healy MJ, Oakshott JG, Kidwell MG. Molecular analysis of the P–M gonadal dysgenesis cline in eastern Australian *Drosophila melanogaster*. *Genetics.* 1988; 119(4): 889–902.
7. Bonnivard E, Higuët D. Stability of European natural populations of *Drosophila melanogaster* with regard to the P–M systems: a buffer zone made up of Q populations. *J Evol Biol.* 1999; 12(4): 633–47.
8. Itoh M, Sasai N, Inoue Y, Watada M. *P* elements and P–M characteristics in natural populations of *Drosophila melanogaster* in the southernmost islands of Japan and in Taiwan. *Heredity.* 2001; 86(2): 206–12.
9. Itoh M, Fukui T, Kitamura M, Uenoyama T, Watada M, Yamaguchi M. Phenotypic stability of the P–M system in wild populations of *Drosophila melanogaster*. *Genes Genet Syst.* 2004; 79(1): 9–18.
10. Onder BS, Bozcuk AN. P - M phenotypes and their correlation with longitude in natural populations of *Drosophila melanogaster* from Turkey. *Russian Journal of Genetics.* 2012; 48(12): 1170–76.
11. Quesneville H, Anxolabéhère D. Dynamics of Transposable Elements in Metapopulations: A Model of *P* Element Invasion in *Drosophila*. *Theor Popul Biol.* 1998; 54(2): 175-93.
12. Bingham PM, Kidwell MG, Rubin GM. The molecular basis of P–M hybrid dysgenesis: the role of the *P* element, a P-strain-specific transposon family. *Cell.* 1982; 29(3): 995–1004.

- 390 13. Ogura K, Woodruff RC, Itoh M, Boussy IA. Long-term patterns of genomic *P* element
391 content and P-M characteristics of *Drosophila melanogaster* in eastern Australia. *Genes Genet Syst.*
392 2007; 82(6): 479-87.
- 393 14. Thopr MW, Chapman EJ, Simmons MJ. Cytotype regulation by telomeric *P* elements in
394 *Drosophila melanogaster*: variation in regulatory strength and maternal effects. *Genet Res Camb.*
395 2009; 91(5): 327–36.
- 396 15. Itoh M, Takeuchi N, Yamaguchi M, Yamamoto M, Boussy IA. Prevalence of full-size *P*
397 and *KP* elements in North American populations of *Drosophila melanogaster*. *Genetica.* 2007;
398 131(1): 21–8.
- 399 16. Boussy IA, Itoh M, Rand D, Woodruff RC. Origin and decay of the *P* element-associated
400 latitudinal cline in Australian *Drosophila melanogaster*. *Genetica.* 1998; 104(1): 45–70.
- 401 17. Gamo S, Sakajo M, Ikeda K, Inoue YH, Sakoyama Y, Nakashima-Tanaka E. Temporal
402 distribution of *P* elements in *Drosophila melanogaster* strains from natural populations in Japan.
403 *Jpn J Genet.* 1990; 65(5): 277–85.
- 404 18. Matsuura E, Takada S, Kato H, Niizeki S, Chigusa S. Hybrid dysgenesis in natural
405 populations of *Drosophila melanogaster* in Japan. III. The P–M system in and around Japan.
406 *Genetica.* 1993; 90(1): 9–16.
- 407 19. Eanes WF, Wesley C, Charlesworth B. Accumulation of *P* elements in minority inversions
408 in natural populations of *Drosophila melanogaster*. *Genet Res.* 1992; 59(1): 1–9.
- 409 20. Biémont C, Lemeunier F, Garcia Guerreiro MP, Brookfield JF, Gautier C, Aulard
410 S, Pasyukova EG. Population dynamics of the *copia*, *mdg1*, *mdg3*, *gypsy*, and *P* transposable
411 elements in a natural population of *Drosophila melanogaster*. *Genet Res.* 1994; 63(3): 197–212.
- 412 21. Charlesworth B, Lapid A, Canada D. The distribution of transposable elements within and
413 between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and
414 distribution. *Genet Res.* 1992; 60(2): 103–14.
- 415 22. Charlesworth B, Jarne P, Assimacopoulos S. The distribution of transposable elements
416 within and between chromosomes in a population of *Drosophila melanogaster*. III. Element
417 abundances in heterochromatin. *Genet Res.* 1994; 64(3): 183–97.
- 418 23. Maside X, Bartolome C, Assimacopoulos S, Charlesworth B. Rates of movement and
419 distribution of transposable elements in *Drosophila melanogaster*: in situ hybridization vs Southern
420 blotting data. *Genet Res.* 2001;78(2):121–36.
- 421 24. Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. Size matters: non-LTR
422 retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol.* 2003; 20(6):
423 880-92.

- 424 25. Lipatov M, Lenkov K, Petrov DA, Bergman CM. Paucity of chimeric gene-transposable
425 element transcripts in the *Drosophila melanogaster* genome. BMC Biol. 2005; 3–24.
- 426 26. Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, Gonzalez J. Population genomics of
427 transposable elements in *Drosophila melanogaster*. Mol Biol Evol. 2010; 28(5): 1633–44.
- 428 27. Blumenstiel J P, Chen X, He M, Bergman, CM. An age-of-allele test of neutrality for
429 transposable element insertions. Genetics. 2014; 196(2): 523-38.
- 430 28. Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Celniker SE.
431 Genome Biol. 2002; 3(12): research0084.1–84.2.
- 432 29. Mackay TF, Richards S, Stone EA., Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y,
433 Magwire MM, Cridland JM et al. The *Drosophila melanogaster* genetic reference panel. Nature.
434 2012; 482(7485): 173–8.
- 435 30. Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM., Crepeau MW, Duchon
436 P, Emerson JJ, Saelao P, Begun DJ et al. Population Genomics of sub-saharan *Drosophila*
437 *melanogaster*: African diversity and non-African admixture. PLoS Genet. 2012; 8(12), e1003080.
- 438 31. Bergman CM, Haddrill PR. Strain-specific and pooled genome sequences for populations of
439 *Drosophila melanogaster* from three continents. F1000Res. 2015; 4–31.
- 440 32. Grenier JK, Arguello JR, Moreira MC, Gottipati S, Mohammed J, Hackett SR, Boughton R,
441 Greenberg AJ, Clark AG. Global diversity lines - a five-continent reference panel of sequenced
442 *Drosophila melanogaster* strains. G3 (Bethesda). 2015; 5(4): 593–603.
- 443 33. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley
444 CH, Pool JE. The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila*
445 *melanogaster* genomes, including 197 from a single ancestral range population. Genetics. 2015;
446 199(4): 1229–41.
- 447 34. Fiston-Lavier AS, Carrigan M, Petrov, D. A, Gonzalez J. T-lex: a program for fast and
448 accurate assessment of transposable element presence using next-generation sequencing data.
449 Nucleic Acids Res. 2011; 39(6): e36—e36.
- 450 35. Linheiro RS, Bergman CM. Whole genome resequencing reveals natural target site
451 preferences of transposable elements in *Drosophila melanogaster*. PLoS One. 2012; 7(2): e30008—
452 e30008.
- 453 36. Kofler R, Betancourt AJ, Schlotterer C. Sequencing of pooled DNA samples (Pool-Seq)
454 uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. PLoS
455 Genet. 2012; 8(1): e1002487.
- 456 37. Cridland JM, Macdonald SJ, Long AD, Thornton KR. Abundance and distribution of
457 transposable elements in two *Drosophila* QTL mapping resources. Mol Biol Evol. 2013; 30(10):
458 2311–27.

- 459 38. Zhuang J, Wang J, Theikapp W, Weng Z. TEMP: a computational method for analyzing
460 transposable element polymorphism in populations. *Nucleic Acids Res.* 2014; 42(11): 6826–38.
- 461 39. Nakagome M, Solovieva E, Takahashi A, Yasue H, Hirochika H, Miyao A. Transposon
462 Insertion Finder (TIF): a novel program for detection of de novo transpositions of transposable
463 elements. *BMC Bioinforma.* 2014; 15:71.
- 464 40. Fiston-Lavier AS, Barrón MG, Petrov DA, González J. T-lex2: genotyping, frequency
465 estimation and re-annotation of transposable elements using single or pooled next-generation
466 sequencing data. *Nucleic Acids Res.* 2015; 43(4): e22.
- 467 41. Rahman RG, Chirn A, Khanodia YA, Sytnikova B, Brembs CM, Bergman & N.C. Lau.
468 Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. *Nucleic*
469 *Acids Res.* 2015; 43(22): 10655–72.
- 470 42. Anxolabéhère D, Nouaud D, Periquet G, Tchen P. *P* element distribution in Eurasian
471 populations of *Drosophila melanogaster*: A genetic and molecular analysis. *Proc Natl Acad Sci*
472 *USA.* 1985; 82(16): 5418–22.
- 473 43. Ronsseray S, Lehmann M, Anxolabéhère D. Copy number and distribution of *P* and *I*
474 mobile element in *Drosophila melanogaster* population. *Chromosoma.* 1989; 98(3): 207–14.
- 475 44. Rozhok AI, Bilousov O, Protsenko OV, Zhuk O, Kozeretska IA. Occurrence of *P* element
476 in natural populations of *Drosophila melanogaster* in Ukraine. *Drosophila Information Service.*
477 2009; 92–73.
- 478 45. Yamamoto A, Hihara E, Watanabe TK. Hybrid dysgenesis in *Drosophila melanogaster*:
479 predominance of Q factor in Japanese populations and its change in the laboratory. *Genetica.* 1984;
480 63(1): 71–77.
- 481 46. Ignatenko OM, Zakharenko LP, Dorogova NV, Fedorova SA. *P* elements and the
482 determinants of hybrid dysgenesis have different dynamics of propagation in *Drosophila*
483 *melanogaster* populations. *Genetica.* 2015; 43(6): 751–9.
- 484 47. Verspoor RL, Haddrill PR. Genetic diversity, population structure and Wolbachia infection
485 status in a worldwide sample of *Drosophila melanogaster* and *D. simulans* populations. *PLoS One.*
486 2011; 6(10): e26318.
- 487 48. Keane TM, Wong K, Adams DJ. RetroSeq: transposable element discovery from next-
488 generation sequencing data. *Bioinformatics.* 2013; 29(3): 389–90.
- 489 49. McClintock. <https://github.com/bergmanlab/mcclintock>.
- 490 50. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Celniker SE. The Release
491 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.* 2015; 25(3): 445–58.
- 492 51. Sackton TB, Kulathinal RJ, Bergman CM, Quinlan AR, Dopman EB, Carneiro M, Marth
493 GT, Hartl DL, Clark AG. Population genomic inferences from sparse high-throughput sequencing
494 of two populations of *Drosophila melanogaster*. *Genome Biol Evol.* 2009; 18(1): 449–65.

495
496
1
497
3
498
4
499
6
500
8
501
9
10
502
11
503
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

52. Berkeley *Drosophila* Genome Project. The data set v9.1.1 transposon sequences. https://github.com/cbergman/transposons/blob/master/misc/D_mel_transposon_sequence_set.fa.

53. Nellåker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, Flint J, Adams DJ, Frankel WN, Ponting CP. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* 2012 15; 13(6): R45.

54. Schlötterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nat Rev Genet.* 2014; 15(11): 749–63.

55. Vieira C, Biémont C. Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. *Genetica.* 2004; 120(1): 115–23.

504

Tables

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

Table 1. Gondal dysgenesis levels and P-M status for isofemale strains of *D. melanogaster* obtained from natural populations in Ukraine. Previous work suggests M-strains reported here are likely to be M' [44].

Population	Year of collection	Cross A (%GD±SD)	Cross A* (%GD±SD)	# strains	M	Q	P'	P
Uman'	2012	21.4±6.9	83.1±8.5	38	3	0	35	0
Kyiv	2012	24.9±6.4	84.4±7.5	49	2	0	47	0
Chornobyl	2012	24.7±15.0	52.5±16.3	40	5	0	35	0
Inkerman	2013	11.0±4.9	24.4±9.2	23	11	0	12	0
Poliske	2013	51.1±8.6	73.7±13.2	18	0	0	18	0
Chornobyl NPP	2013	13.6±6.6	75.2±8.2	34	12	0	22	0
Varva	2013	26.4±8.8	35.8±11.0	39	1	0	38	0
Total				241	34	0	207	0

511 **Table 2. Gonadal dysgenesis levels and P-M status for isofemale strains of *D. melanogaster***
512 obtained from natural populations in North America, Europe and Africa. Phenotypically M-strains
513 are in fact M'-strains based on analysis of genomic data (File S2, File S3).

Population	Year of collection	Cross A (%GD±SD)	Cross A* (%GD±SD)	# strains	M	Q	P'	P
N. America	2009	15.1±13.4	6.4±9.1	14	3	4	0	7
Europe	2010	6.1±12.2	6.8±14.0	17	2	12	1	2
Africa	2010	16.3±22.8	6.0±5.9	12	1	8	1	2
Total				43	6	24	2	11

516 **Table 3. Numbers of P element insertions identified by TEMP and RetroSeq in pool-seq samples**
517 from three worldwide populations of *D. melanogaster*. Columns labeled raw and filtered represent
518 output generated by each method before or after standard filtering by McClintock, respectively (see
519 Materials and Methods for details).

Population	# strains	TEMP (raw)	TEMP (filtered)	RetroSeq (raw)	RetroSeq (filtered)
N. America	15	684	53	312	219
N. America	30	1,101	27	259	159
Europe	20	1,003	85	372	245
Europe	39	1,395	46	278	171
Africa	15	958	110	505	348
Africa	32	1,681	28	329	193

Figure Legends

Figure 1. Results of GD assays for isofemale strains from natural populations from Ukraine. The graphs present %GD for the populations for cross A (vertical axis) and cross A* (horizontal axis). Each dot represents an isofemale strain. The P-M status for various sectors of GD phenotypic space defined by A and A* crosses are according to [44] and shown in panel A.

Figure 2. Results of GD tests for isofemale strains from natural populations from North America, Europe and Africa. The graphs present %GD for the populations for cross A (vertical axis) and cross A* (horizontal axis) using data from [45]. A and A* cross definitions in [45] are inverted relative to those proposed by [4] and were standardized prior to analysis here. Each dot represents an isofemale strain. Definition of P-M status by A and A* crosses according to [44] is the same as in Fig. 1A.

Figure 3. Relationship between %GD in A and A* crosses and numbers of euchromatic *P* element insertions identified by TEMP or RetroSeq for isofemale strains from natural populations from North America, Europe and Africa. %GD data are from [45] and are the same standardized values as in Figure 2. Numbers of *P* elements predicted by TEMP or RetroSeq shown here are after standard filtering by McClintock (see Materials and Methods for details). Results for unfiltered raw output of TEMP or RetroSeq are shown in Fig. S4. Each triangle represents an isofemale strain.

Supplemental Files

File S1. Tab separated value formatted file with %GD data from A and A* crosses, P-M cytotype status, and population for isofemale strains from Ukraine.

File S2. Tab separated value formatted file with %GD data from A and A* crosses, P-M cytotype status, population, and numbers of predicted *P* elements in raw and filtered output from TEMP and RetroSeq, respectively, for isofemale strains from three global regions. GD data are taken from [46] and were standardized to definitions proposed by [4] prior to re-analysis here.

File S3. Zip archive of browser extensible data (BED) files of predicted *P* element locations in genome sequences from 50 isofemale strains and 6 pool-seq samples from three global regions. Each sample has four BED files corresponding to raw (*raw.bed) and filtered (*nonredundant.bed) output from TEMP and RetroSeq, respectively.

Figure S1. Distributions of %GD in A and A* crosses for isofemale strains within and between populations from Ukraine. Distributions are shown as boxplots with black lines representing median values, boxes representing the interquartile range (IQR), whiskers representing the limits of values for strains that lie within 1.5 x IQR of the upper or lower quartiles, and circles representing strains that lie outside 1.5 x IQR of the upper or lower quartiles.

Figure S2. Distributions of %GD in A and A* crosses and numbers of predicted *P* element insertions for isofemale strains within and between populations from three worldwide regions. Distributions are shown as boxplots with black lines representing median values, boxes representing the interquartile range (IQR), whiskers representing the limits of values for strains that lie within 1.5 x IQR of the upper or lower quartiles, and circles representing strains that lie outside 1.5 x IQR of the upper or lower quartiles. Numbers of *P* elements predicted by TEMP or RetroSeq shown are before (raw) and after (filtered) standard filtering by McClintock.

Figure S3. Correlation among methods in the numbers of predicted *P* element insertions for a worldwide sample of isofemale strains from North America, Europe and Africa. Numbers of *P* elements predicted by TEMP or RetroSeq shown are before (raw) and after (filtered) filtering by McClintock. Each circle represents an isofemale strain.

576 **Figure S4. Relationship between %GD in A and A* crosses and numbers of euchromatic P element**
577 insertions identified by TEMP or RetroSeq for isofemale strains from natural populations from
578 North America, Europe and Africa. %GD data are from [45] and are the same standardized values
579 as in Figure 2. Numbers of P elements predicted by TEMP or RetroSeq shown are raw output prior
580 to standard filtering by McClintock. Results for McClintock-filtered output of TEMP and RetroSeq
581 are shown in Fig. 3. Each circle represents an isofemale strain.

5
1
3
5
7
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 1

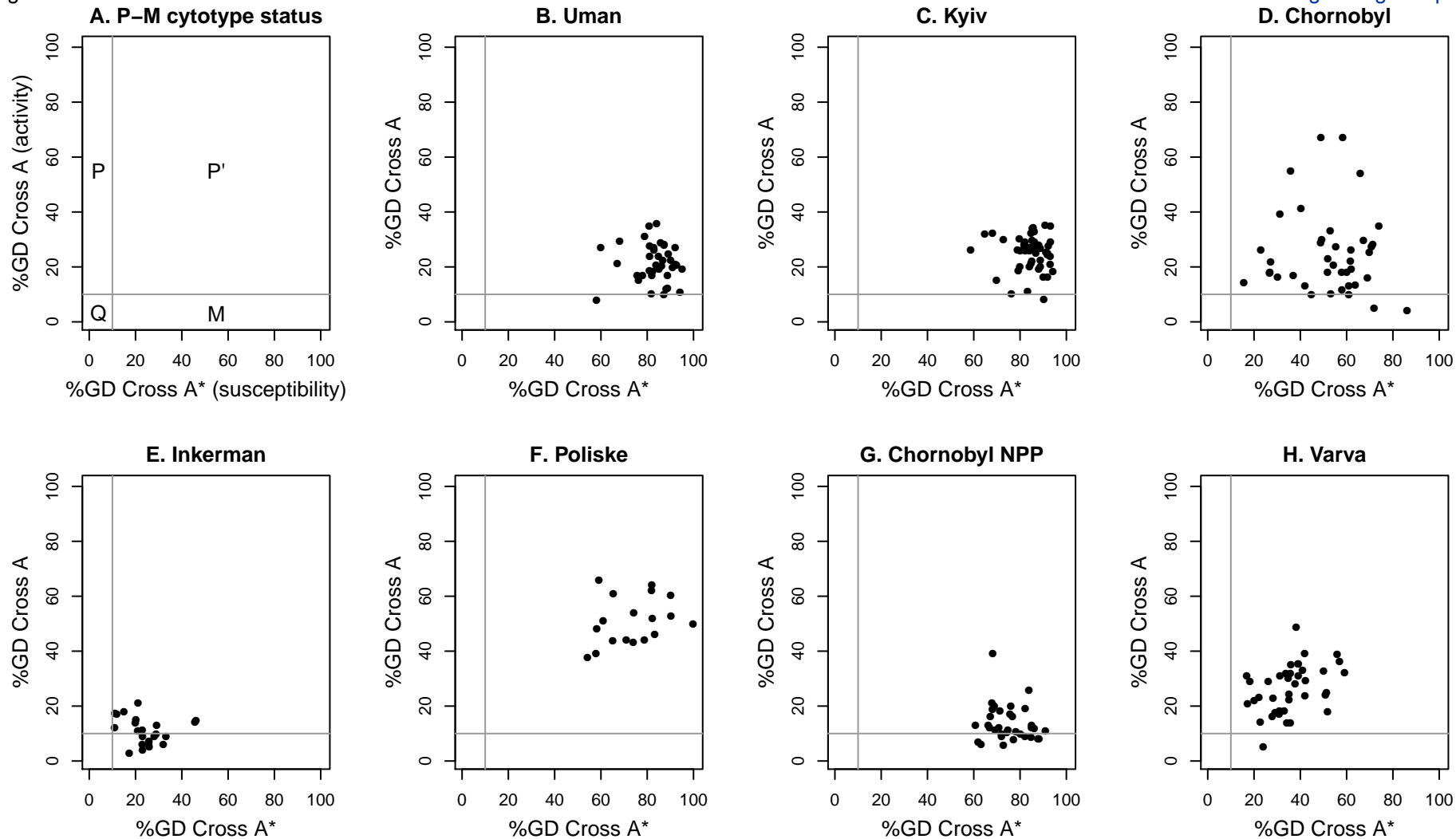
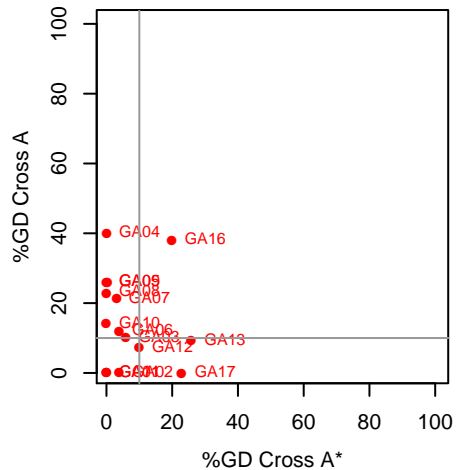
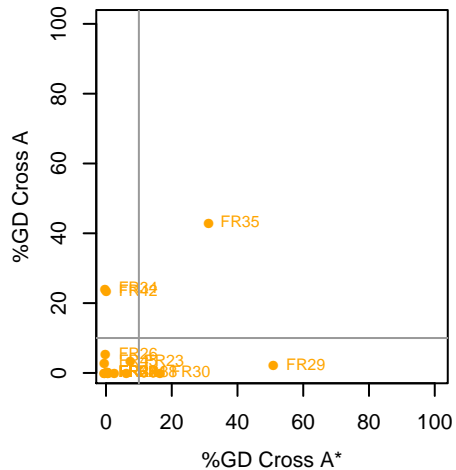


Figure 2

A. North America (Athens, Georgia, USA)



B. Europe (Montpellier, France)



C. Africa (Accra, Ghana)

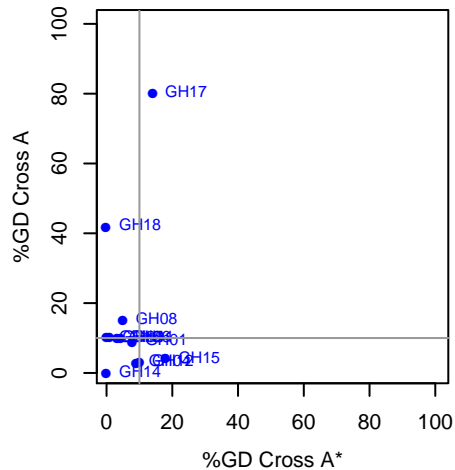
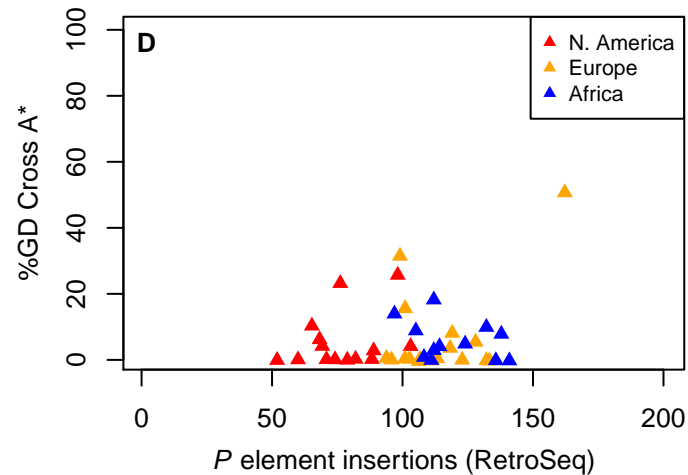
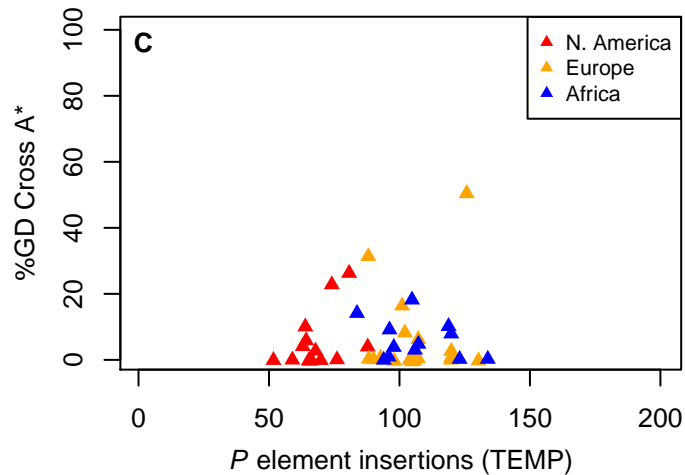
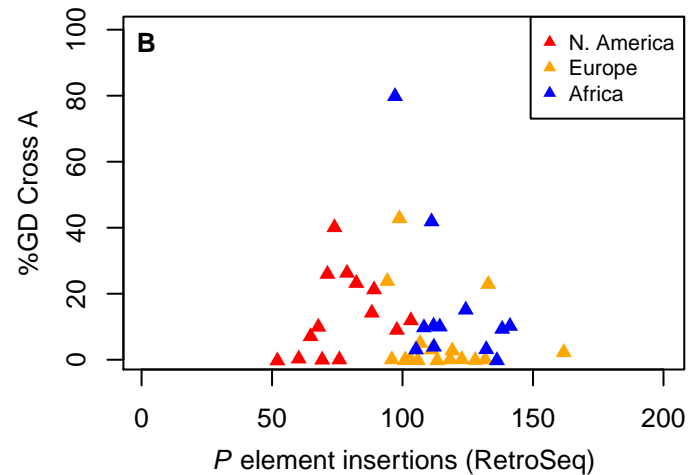
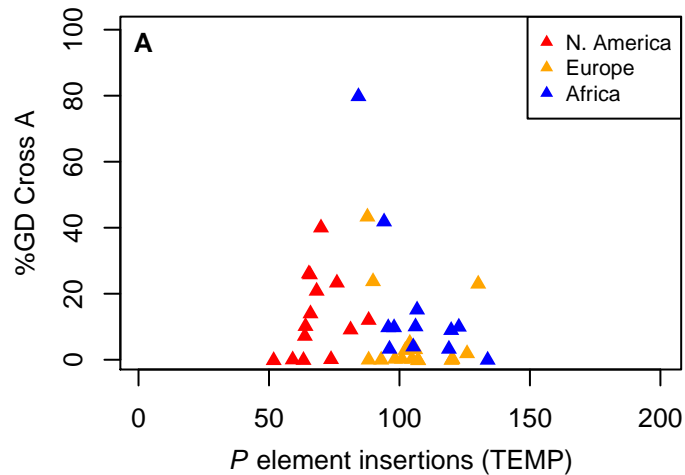
[Click here to download Figure Figure2.pdf](#)

Figure 3

[Click here to download Figure Figure3.pdf](#)



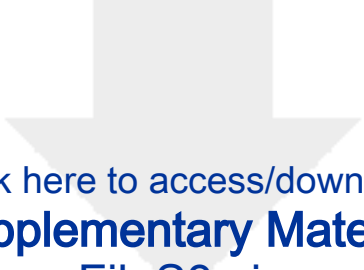
Click here to access/download
Supplementary Material
FileS1.tsv





Click here to access/download
Supplementary Material
FileS2.tsv





Click here to access/download
Supplementary Material
FileS3.zip

