

# Hybrid dysgenesis in *Drosophila simulans* due to a rapid global invasion of the *P*-element

Tom Hill<sup>1</sup>, Christian Schlötterer<sup>1</sup>, Andrea J. Betancourt<sup>1\*</sup>

<b>Supplementary Materials and Methods</b>	<b>2</b>
1. Fly strains	2
2. Hybrid dysgenesis assays	2
3. Sterility assays	3
4. Analysis of RNAseq data	3
5. Analysis of whole genome sequence data	3
6. PCR	4
7. RT-PCR	4
8. Supplementary References	5
<b>Supplementary Figures</b>	<b>6</b>
<b>Figure S1.</b> Hybrid dysgenic crosses and controls	6
<b>Figure S2.</b> Map of cytotypes	8
<b>Figure S3.</b> Sterility assay results	9
<b>Figure S4.</b> TE coverage comparison between DI & DS type lines	10
<b>Figure S5.</b> Association between dysgenesis and P-element copy number	11
<b>Figure S6.</b> RT-PCR results	12
<b>Supplementary Tables</b>	<b>13</b>
<b>Table S1.</b> Strains used	13
<b>Table S2.</b> Association between TE presence and cytotype	16
<b>Table S3.</b> Primer sequences	19

## **Text S1. Supplementary Materials and Methods**

### **1. Fly strains.**

We used *D. simulans* isofemale lines collected across multiple locations over 15 years (see Supplementary Table 1 for details). The species identity of all lines was confirmed either by visual inspection of male genitalia or, when this was not possible, by amplification of PCR products that yield species-specific product lengths (using primer sequences for Argonaute2 (FBgn0087035), kindly provided by Darren Obbard: Mel\_Sim\_F 5'-CCCTAAACCGCAAGGATGGAG-3', D\_sim\_303\_R 5'-GTCCACTTGTGCGCCACATT-3', D\_mel\_394\_R 5'-CCTTGCCTTGGGATATTATTAGGTT-3'). Flies were reared on molasses-yeast-agar *Drosophila* medium. To obtain sufficient offspring from crosses that normally show high levels of cytoplasmic incompatibility, we cured lines of *Wolbachia* by raising larvae on food supplemented with 0.05mg/ml tetracycline-hydrochloride for 2 generations before performing crosses. After curing, flies were test for *Wolbachia* using PCR, with primer sequences wsp81F (5'-TGGTCCAATAAGTGATGAAGAAAC-3') and wsp691R (5'-AAAAATTAAACGCTACTCCA-3'), and a cycling program of 94°C for 5 minutes followed by 30 cycles of 94°C 30 seconds, 55°C for 1 minute and 72°C for 2 minutes, followed by 72°C for 10 minutes (1). If known, the infection status of lines prior to curing is given in Supplementary Table 1. Crosses performed using the same lines, but with and without cured flies had similar results (shown in Figure S1).

### **2. Hybrid dysgenesis assays.**

To assay for hybrid dysgenesis, we performed reciprocal crosses between lines, using 5 virgin males and females from each line. Flies were left to lay eggs for a total of 9 days at 29°C, transferred to a fresh vial every 3 days, and offspring reared at 29°C (28°C is the lowest temperature at which dysgenesis is observed in *D. melanogaster* (2)). We dissected 30 F1 females from each cross, and assessed them for the presence or absence of two normal ovaries. We scored the number of females lacking one or both ovaries as dysgenic, following (3). We used Fisher's exact tests to test for a difference between reciprocal crosses in the

number of dysgenic female offspring, allowing us to roughly categorize the strains as DI (dysgenesis inducing, having an excess of dysgenic offspring when crossed to susceptible females), DS (dysgenesis susceptible, having an excess of dysgenic offspring when crossed to inducing males), or DR (dysgenesis resistant, no excess in either direction). Some heterogeneity occurred within strains; and some strains could not be unambiguously classified (4 borderline DI/DR types, and 2 borderline DR/DS types, and 3 that tested as DI and DS ). These strains were not used in the analysis; in all, 177 strains were unambiguously cytotyped.

### **3. Sterility Assays.**

To assay for sterility, reciprocal crosses were performed as described previously at 29°C, newly emerged offspring from these crosses were collected as virgins as separated based on the date of collection. Following this, crosses between individual females and control virgin males (raised at 25°C) were established at 25°C in the following subsets: Dysgenic females aged 3-9 days, non-dysgenic females aged 3-9 days, Dysgenic Females aged 10-16 days, non-dysgenic females aged 10-16 days. Flies were left to mate and lay eggs for a total of 7 days and offspring from these crosses were collected every 2 days following this. We scored crosses by the number of offspring produced. We used a Wilcoxon Rank Sum Test to assess for a significant difference in the number of offspring for each direction ( $p < 0.05$ ).

### **4. Analysis of RNAseq data.**

Paired end Illumina reads (accession number PRJEB7936 (4)) were mapped to a *D. simulans* reference genome produced from a Madagascar strain (5) and 179 *Drosophila* transposable element (TE) sequences (<http://flybase.org/>; (6,7)) using GSNAP (parameters: -n =1, -N = 1)(8).

### **5. Analysis of whole genome sequence data.**

We used paired-end Illumina reads (SRA PRJEB7936; (4)) collected from 13 barcoded individual F1 offspring of the M252 reference strain females crossed to males from Florida (2010) isofemale lines, and the paired end reads from the M252 reference strain (5). We mapped these data to the *D. simulans* M252 reference genome using BWA-SW (default parameters) (9–11) and following the mapping protocol described in Kofler et al (2012). The average coverage per sequenced F1 cross was 36.36. We counted the number of insertions per

individual sequenced in two ways: First, we used PoPoolationTE [11] to identify and count TE insertions from different families, with the requirement that each insertion be supported by at least 3 reads. Insertions were considered to be potentially specific to the Florida lines if they occurred at count frequencies of 15-85% (as we expect insertions to be heterozygous) and were not present in the M252 sequence. Second, due to the fragmented nature of the *D. simulans* genome sequence, some insertion sites may be difficult to locate. We therefore also estimated TE copy number per family *via* average coverage of the TE sequence compared to average coverage of chromosome arm 2L.

## **6. PCR.**

Primers for TEs were designed based on the EMBL sequences (Supplementary Table 6, <http://flybase.org/>; (6,7)). The following PCR program was used: 94°C for 5 minutes followed by 30 cycles of 94°C 30 seconds, 55°C for 1 minute and 72°C for 1-3 minutes (depending on the expected length of the PCR product, 1 minute per kilobase); 72°C for 10 minutes. PCR products were run on a 1% agarose gel to test for the presence or absence of a TE. To survey for the P-element, primers were designed for each exon of P-element, shown in the following table. The presence of full P-elements was assessed using the forward primer for exon 0 and the reverse primer for exon 3 and using the TIR primer. Known P-types from *D. melanogaster* lines were used as positive controls in all PCRs. When possible, PCR results were confirmed with independent DNA extractions (though this was not possible for wild collected flies from Portugal and African samples from Michel Veuille, Supplementary Table S1). DNA quality was checked with a positive control PCR for an essential gene.

Sanger sequencing of a subset of PCR products was used to confirm that the amplicons correspond to P-element sequence; the sequence of additional sequencing primers is available upon request.

## **RT-PCR.**

RNA was isolated from Florida, Georgia, Harvard, Maryland and Madagascar lines using Peqlab TriFast RNA isolation protocol, and extractions checked on a 1% agarose gel. Following this, RT-PCR was performed using QIAGEN OneStep RT-PCR Kit. To assess the presence or absence of full P-element RNA, the following PCR program was used: 45°C for 30 minutes for reverse transcription.

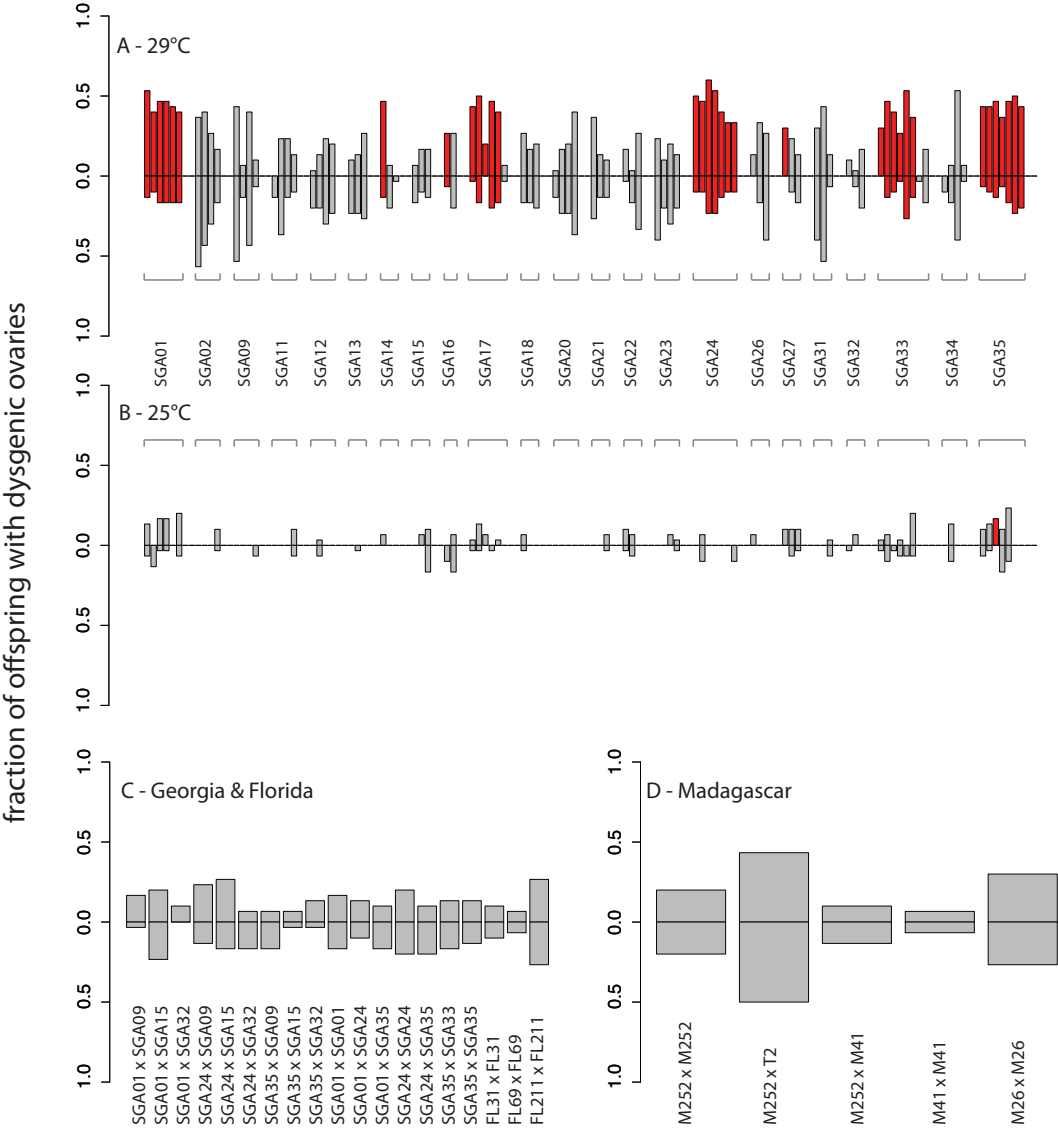
95C for 15 minutes followed by 30 cycles of 94C for 15 seconds, 55C for 1 minute and 68C for 6 minutes; 68C for 10 minutes. We used the forward primer for exon 0 and the reverse primer for exon 3 to test for the presence of full expressed P-elements.

## 8. Supplementary References

1. Braig HR, Zhou W, Dobson SL, Neill SLO. Cloning and Characterization of a Gene Encoding the Major Surface Protein of the Bacterial Endosymbiont *Wolbachia pipientis*. J Bacteriol. 1998;180(9):2373–8.
2. Kidwell JF, Kidwell MG, Sved JA. Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility, and male recombination. Genetics. 1977;86:813–33.
3. Engels WR, Preston CR. Hybrid Dysgenesis in *Drosophila melanogaster*: The Biology of Female and Male Sterility. Genetics. 1979;92(2257):161–74.
4. Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. The recent invasion of natural *Drosophila simulans* populations by the P-element. Proc Natl Acad Sci. 2015;112(21):6659–63.
5. Palmieri N, Nolte V, Chen J, Schlötterer C. Genome assembly and annotation of a *Drosophila simulans* strain from Madagascar. Mol Ecol Resour. 2014;
6. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLoS Comput Biol. 2005;1(2):0166–75.
7. Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, et al. The Transposable Elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol. 2002;3:RESEARCH0084.
8. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinforma Oxf Engl. 2010 Apr 1;26(7):873–81.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009;25(14):1754–60.
10. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009;25(16):2078–9.
11. Kofler R, Betancourt AJ, Schlötterer C. Sequencing of pooled DNA Samples ( Pool-Seq ) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. PloS Genet. 2012;8(1):1–16.

Supplementary Figures

Figure S1



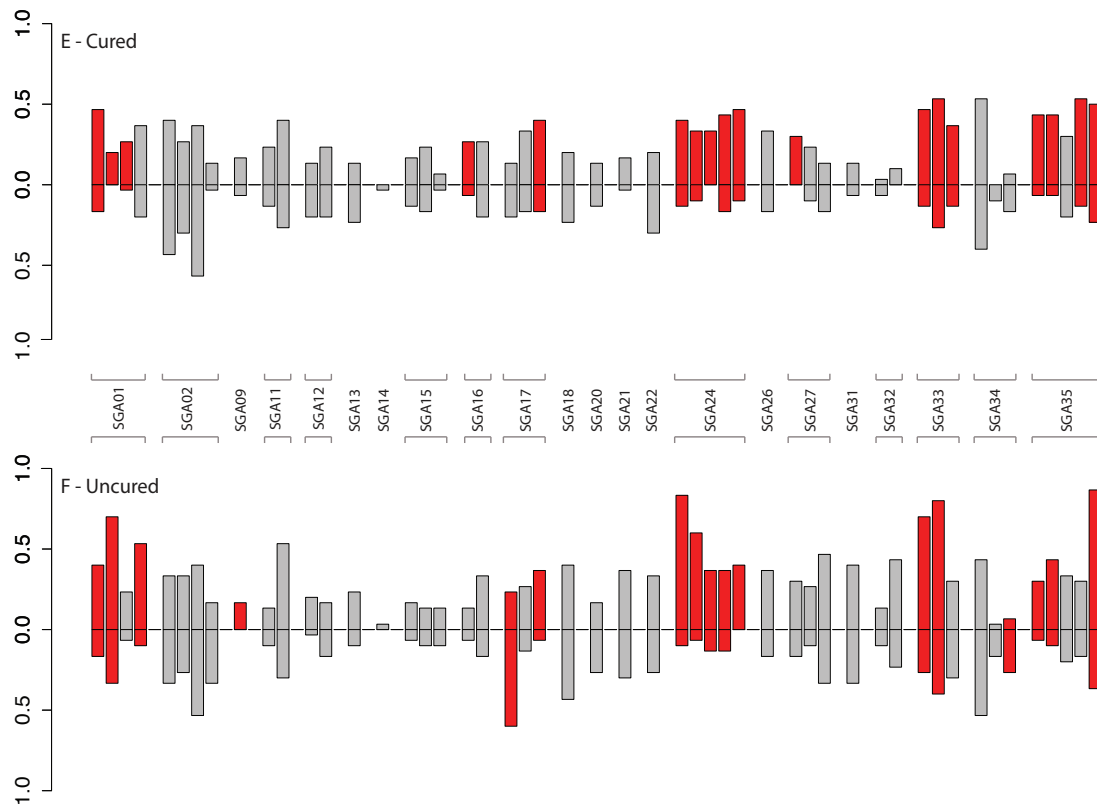


Figure S1. Hybrid dysgenic crosses and controls. Bar plots show the fraction of dysgenic offspring for both directions of each set of reciprocal crosses. Bars are coloured red when a significant difference between each reciprocal cross is found (Fisher's Exact Test,  $p < 0.05$ ). **A.** Initial crosses between Georgia (2009) and Madagascar (2004) flies at 29° C, with the fraction of dysgenic F1 females shown when the Georgia strain is paternal (positive direction) or maternal (negative direction). **B.** A subset of the crosses in panel A were repeated at 25° C, a non-dysgenic temperature in *D. melanogaster*; repeated crosses are shown in the same position as in panel A. **C.** Fraction of dysgenic offspring from crosses within and between Georgia and Florida lines at 29° C. The first strain named in the cross is the paternal strain in the positive direction and the maternal strain in the negative direction. **D.** Fraction of dysgenic offspring from crosses within and between Madagascar lines at 29° C. The first strain named in the cross is the paternal strain in the positive direction and the maternal strain in the negative direction. **E.** Fraction of dysgenic offspring from crosses between Georgia and Madagascar lines at 29C after curing of wolbachia with tetracycline hydrachloride for 2 generations. **F.** Fraction of dysgenic offspring from crosses between Georgia and Madagascar lines at 29C before curing.



**Figure S2**

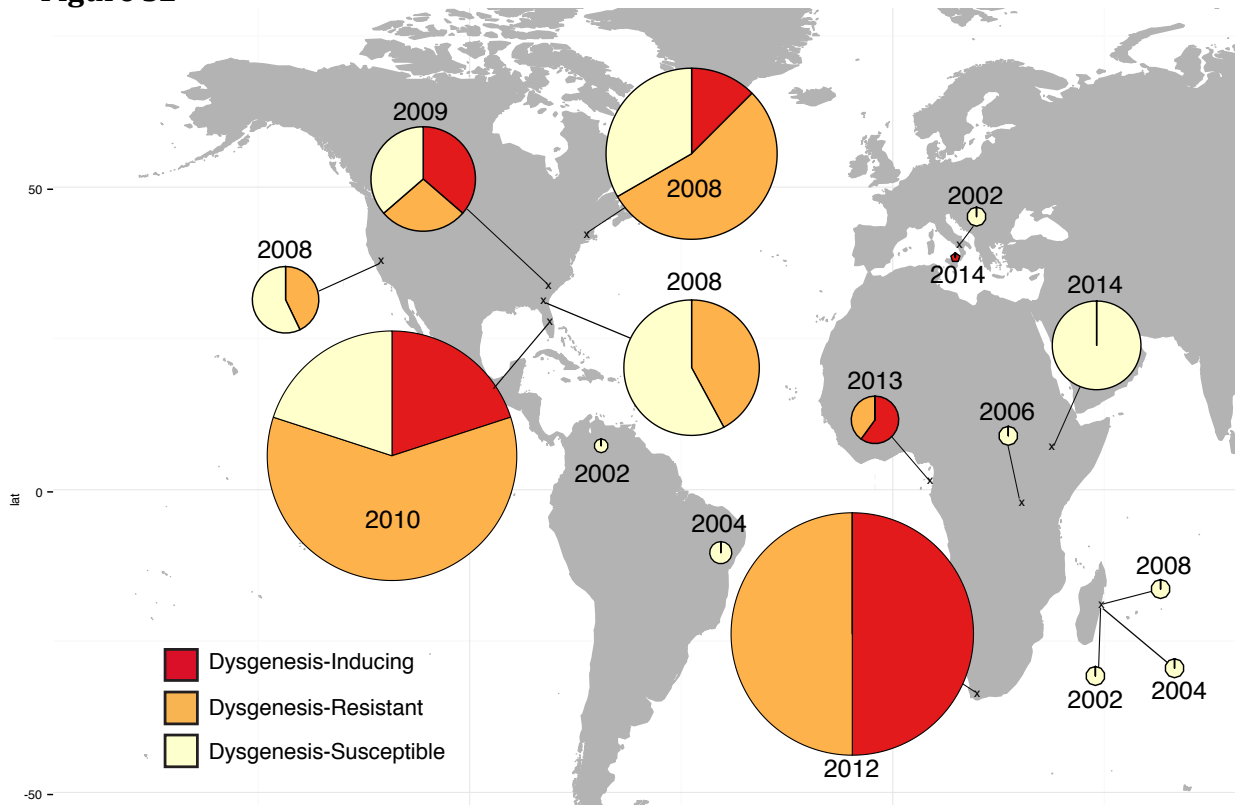


Figure S2. Map shows the approximate location where strains were collected; pie charts show the proportion of each population which were dysgenesis-inducing (red), dysgenesis-resistant (orange) and dysgenesis-susceptible (yellow). This summarizes the data given in Table 2. The area of the pie chart is proportional to the number of strains sampled.

**Figure S3**

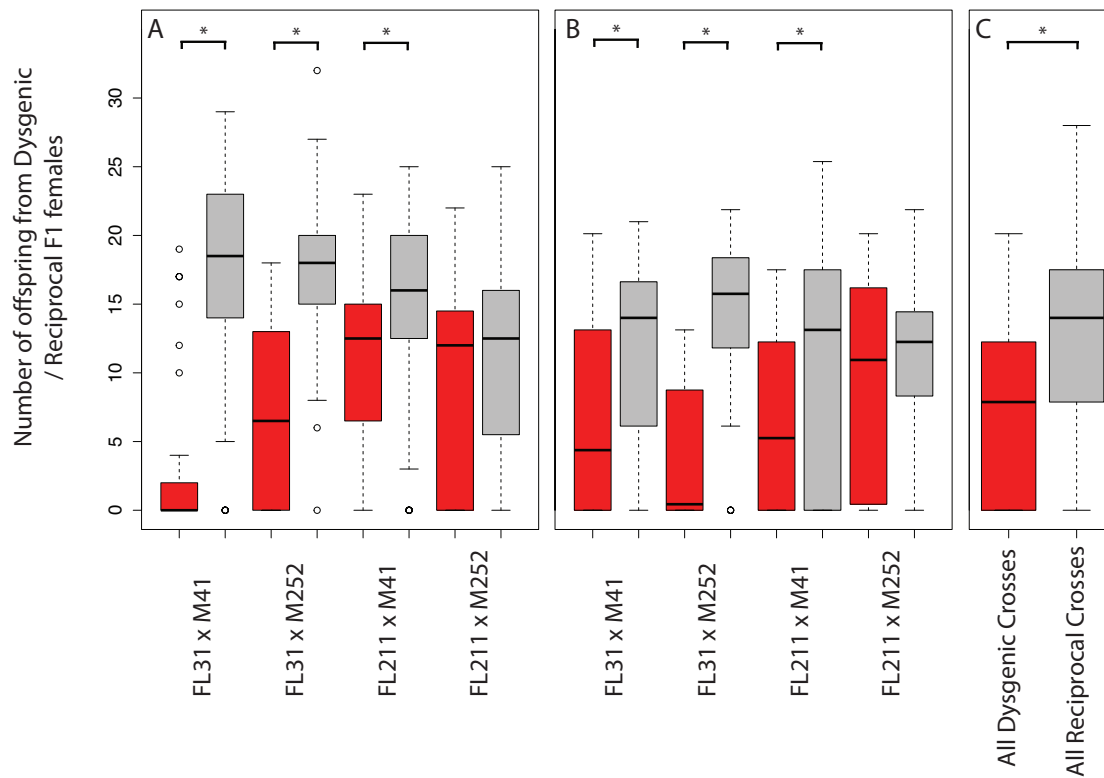
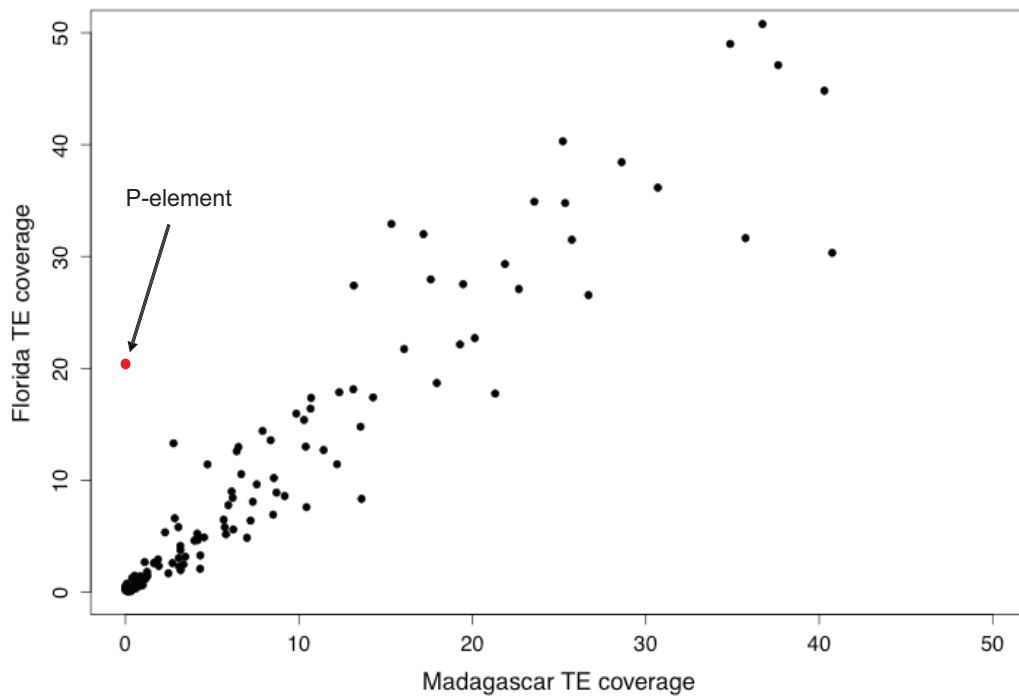


Figure S3. Plots showing the number of offspring produced by the female offspring from both the dysgenic (red) and reciprocal (grey) cross. Female parents were dissected after crossing to confirm the presence of gonadal dysgenesis (mean of 50.6% of female parents were dysgenic, compared to a mean of 8.6% dysgenic female parents in the reciprocal cross). **A.** Females were mated after 3-9 days of aging. **B.** Females were mated after 10-16 days of aging. **C.** Combined results for all crosses and ages. Crosses that show a significant difference in the number of offspring for each direction are marked with a star (Wilcoxon Rank Sum Test  $p < 0.05$ ).

**Figure S4**



**Figure S4.** Average coverage for TEs in the sequenced Florida (2010) x Madagascar (2004) isofemale lines vs. coverage for TEs in M252 (2004). This Florida sequence data (SRA:PRJEB7936) and Madagascar data (SRA:SRX504933) was mapped to the *D. simulans* reference genome alongside Flybase TE sequences to identify differences in coverage for these TEs. Only one TE was found in Florida but not in Madagascar, P-element.

**Figure S5**

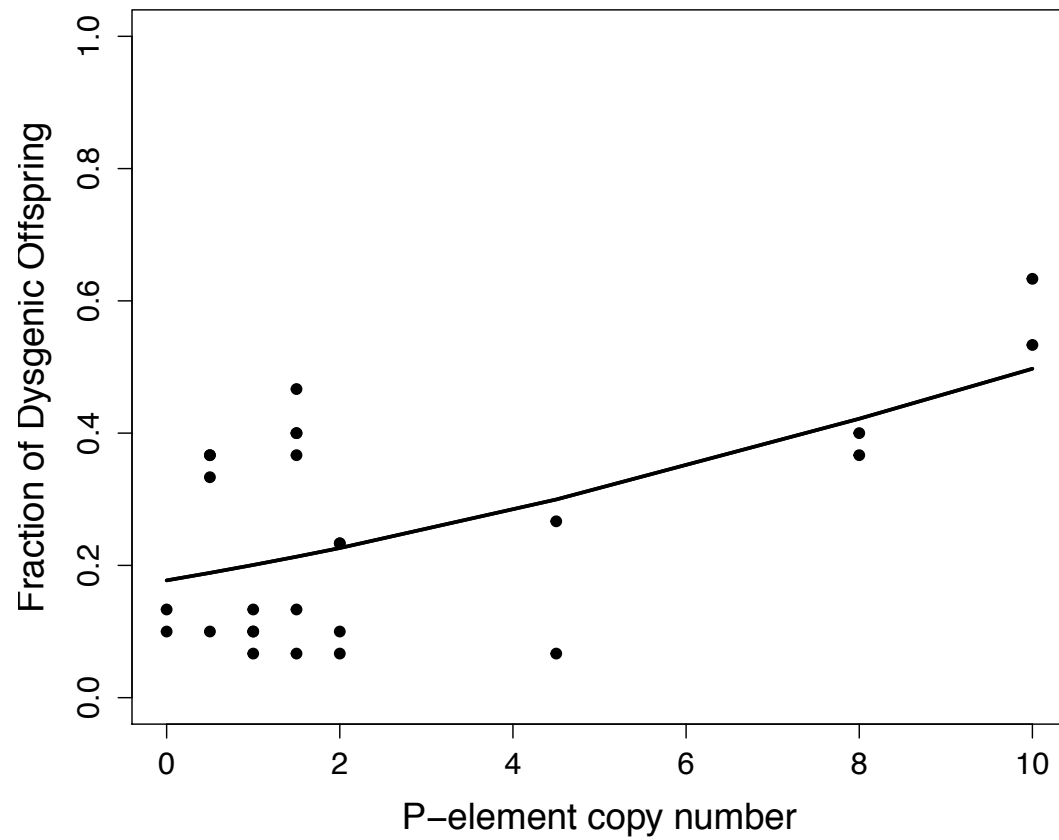


Figure S5. Estimated copy number of P-elements (estimated via dividing the average coverage of P-element in a sample by the average coverage of chromosome 2R) vs. the fraction of dysgenic offspring observed in a dysgenic cross. Line shows the fit of a binomial generalised linear model (Supplementary Table 2; z value = 5.87,  $p = 4.25 \times 10^{-9}$ ).

**Figure S6**

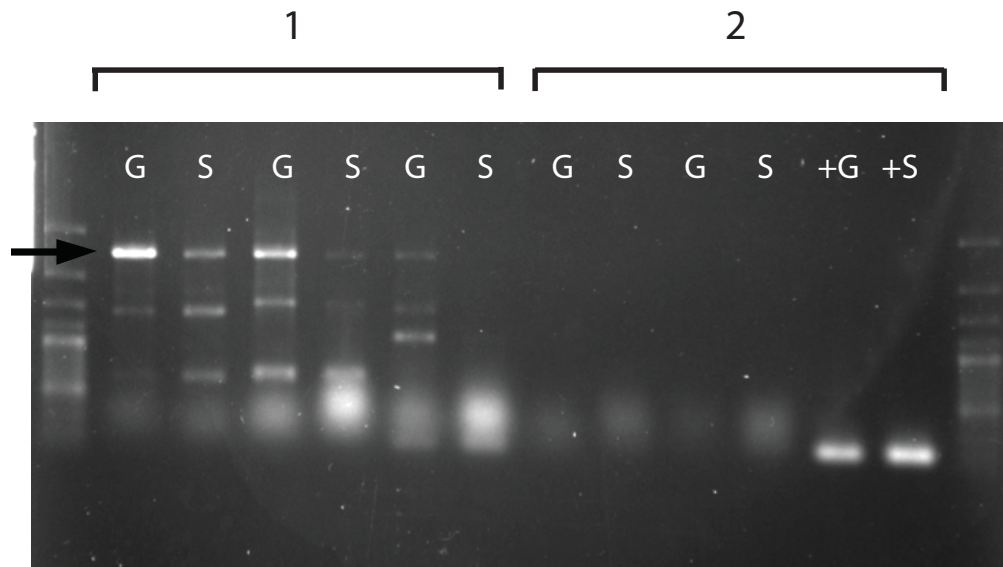


Figure S6. RT-PCR product from DI and DS lines. Group 1 consists of female flies from three DI lines roughly dissected into soma (S) and ovaries (G). The top band, at 2.6kb, corresponds to the expected size of the full-length P-element transcript after splicing. Group 2 consists of female flies from two DS lines with a positive control (*D. melanogaster* Harwich strain).

**Supplementary Table 1.** Fly strains used in this study, including time and date of collection, and collector. The fraction of strains with P-element are shown for each strain.

population	n	species	Locality	year	collected by	Lines	proportion of strains with P-element exon				Number of strains with P-element			
							exon 0	exon 1	exon 2	exon 3	full	partial	none	total
SGA	22	<i>D. simulans</i>	Athens, Georgia, USA	2009	P.Haddrill	SGA1-2, 9, 11-18, 20-22, 24, 26-27, 31-35	0.68	0.5	0.45	1	7	15	0	22
ANDA	7	<i>D. simulans</i>	Andasibe, Madagascar	2008	J.R. David	A1, 3-6, 9, 11	0.56	0.84	0.14	0.28	0	6	1	7
M	5	<i>D. simulans</i>	unknown, Madagascar	2004	B. Ballard	M4, 26, 41, 252, 258	0	0	0	0	0	0	5	5
Mad	1	<i>D. simulans</i>	unknown, Madagascar	2009	E. Helmich	M5	1	0	0	0	0	1	0	1
TANA	7	<i>D. simulans</i>	Antananarivo, Madagascar	2008	J.R. David	T1-2, 6-10	0.42	0.42	0.14	0.71	0	5	2	7
FL	36	<i>D. simulans</i>	Tampa, Florida, USA	2010	R. Tobler	FL3, 6, 11, 15, 26-27, 30-31, 35, 69, 82, 96-98, 101, 108, 113, 116, 131, 136, 141, 150, 155, 164-165, 168, 174, 183-184, 187, 189, 198, 200, 203, 208, 211	0.81	0.56	0.56	0.81	17	12	7	36
SA	34	<i>D. simulans</i>	Kanonkop, South Africa	2013	H. van Schalkwyk	SA1104-1105, 1109, 1112, 1116, 1118, 1120, 1123-1125, 1128-1129, 1134, 1140, 1144, 1146, 1148, 1157, 1161-1163, 1165, 1175-1176, 1183-1184, 1187-1189, 1191-1192, 1195, 1197, 1198	1	1	1	1	34	0	0	34
HAR	14	<i>D. simulans</i>	Harvard, Massachusetts, USA	2008	A. Paaby	HAR1-3, 5-8, 10, 13-14, 16-17, 20, 21	0.28	0	0	0.14	0	4	10	14
DAV	19	<i>D. simulans</i>	Davis, California, USA	2009	M. Turelli	DAV1-8, 11, 13, 15-16, 18-20, 23, 26, 28	0.05	0.05	0	0	0	1	18	19
MED	17	<i>D. simulans</i>	Media, Pennsylvania, USA	2008	A. Paaby	MED1, 3-14, 16-17, 19, 20	0.12	0.059	0	0	0	2	15	17
CHU	13	<i>D. simulans</i>	Churchville, Maryland, USA	2008	A. Paaby	CHU2, 4-6, 10-12, 15, 17-19, 21, 22	0.69	0.31	0.23	0.69	5	4	4	13
HIN	19	<i>D. simulans</i>	South Carolina, USA	2008	A. Paaby	HIN1-18, 20	0.053	0.053	0.053	0.053	0	1	18	19
IRF	13	<i>D. simulans</i>	Jasper, Florida, USA	2008	A. Paaby	IRF1, 3-4, 6-8, 10-14, 17, 18	0.077	0.077	0.077	0	0	2	11	13
LPS	10	<i>D. simulans</i>	Morven, Georgia, USA	2008	A. Paaby	LPS1-6, 9, 12-14	0	0	0	0	0	0	10	10
PD	12	<i>D. simulans</i>	Padova, Italy	2002	E. Tauber	PD7-8, 18-19, 29, 38, 42-43, 52-53, 55, 67	0	0	0	0	0	0	12	12
BGT	2	<i>D. simulans</i>	Guaymaral, close to Boqota, Columbia	2002	P. Orozco-terWengel	BGT4, 6	0	0	0	0	0	0	2	2
TU	1	<i>D. simulans</i>	Djerba, Tunisia	2001	M. Puchinger	TU46	0	0	0	0	0	0	1	1
ZOM	2	<i>D. simulans</i>	Zomba, Malawi	2001	C. Niessinger	ZOM3, 4	0	0	0	0	0	0	2	2
KIB	2	<i>D. simulans</i>	Kibale, Uganda	2001	M. Imhof	KIB11, 73	0	0	0	0	0	0	2	2
ABU	9	<i>D. simulans</i>	Ubatuba, Brazil	2004	J. F. Garcia	ABU4, 7, 16, 18, 23-26, 28	0	0	0	0	0	0	9	9
Mandra	6	<i>D. simulans</i>	Mandraka Park, Madagascar	2008	J.R. David	MAND2-5, 7, 9	0.67	0.67	0	0.5	0	4	2	6

population	n	species	Locality	year	collected by	Lines	proportion of strains with P-element exon				Number of strains with P-element			
							exon 0	exon 1	exon 2	exon 3	full	partial	none	total
YA	3	<i>D. simulans</i>	Nairobi, Kenya	2013	D.Matute	YA9, 19, 29	0.33	0.33	0	0.33	0	3	0	3
ST	7	<i>D. simulans</i>	Sao Tome, Sao Tome and Principe	2013	D.Matute	ST1,ST14,ST16,ST5,ST6,ST8	1	1	1	1	7	0	0	7
BS	5	<i>D. simulans</i>	Antananarivo, Madagascar	2013	D.Matute	BS0, 4, 8	1	1	1	1	5	0	0	5
BIO	15	<i>D. simulans</i>	Bioko Norte, Equatorial Guinea	2013	D.Matute	BIOKO B11,BIOKO C1.3,BIOKO C1.5,BIOKO C4,BIOKO F5,BIOKO FH7,BIOKO L9,BIOKO LOZ,BIOKO LP88,BIOKO MC,BIOKO R2,BIOKO RIM,BIOKO SJ	1	1	1	1	15	0	0	15
SREN	3	<i>D. simulans</i>	Sorrento, Italy	2014	M.Pegoraro	SREN17, 52, 61	1	0.33	0.33	0.66	1	2	0	3
MD	16	<i>D. simulans</i>	Antananarivo, Madagascar	2002	B. Ballard	MD6, 13, 15, 42, 63, 71, 73, 75, 105-106, 199, 221, 223, 233, 242, 251	0	0	0	0	0	0	16	16
NS	12	<i>D. simulans</i>	Nairobi, Kenya	2006	B. Ballard	NS5, 31-33, 39-40, 50, 71, 78-79, 113, 137	0	0	0	0	0	0	12	12
PORT	60	<i>D. simulans</i>	North Portugal, Portugal	2006	E.Sucena	DNA from 60 wild-caught flies	0.13	0.18	0.16	0.18	3	10	47	60
POR	36	<i>D. simulans</i>	North Portugal, Portugal	2013	E.Sucena	DNA from 36 wild-caught flies	1	1	1	1	36	0	0	36
IT	38	<i>D. simulans</i>	Bologna, Italy	2014	M.F.Shou	IT1-38	1	1	1	1	38	0	0	38
CRO	50	<i>D. simulans</i>	Zabqreb, Croatia	2014	A.M. Jaksic	CRO1-50	1	1	1	1	50	0	0	50
ED	1	<i>D. simulans</i>	Dodola, Ethiopia	2008	J.Pool	ED7	1	0	0	1	0	1	0	1
KN	1	<i>D. simulans</i>	Nyahururu, Kenya	2009	J.Pool	KN86	1	0	0	1	0	1	0	1
UK	1	<i>D. simulans</i>	Kisoro, Uganda	2012	R.Corbett	UK6	1	0	0	1	0	1	0	1
UB	1	<i>D. simulans</i>	Bundibugyo, Uganda	2012	R.Corbett	UB5	1	0	0	1	0	1	0	1
TZ	1	<i>D. simulans</i>	Uyole, Tanzania	2009	L.Nsemwa	TZ5	0	0	0	0	0	0	1	1
KR	1	<i>D. simulans</i>	Mariqat, Kenya	2009	J.Pool	KR56	0	0	0	0	0	0	1	1
EZ	1	<i>D. simulans</i>	Ziway, Ethiopia	2008	J.Pool	EZ108	0	0	0	0	0	0	1	1
SP	1	<i>D. simulans</i>	Phalaborwa, South Africa	2010	J.Pool	SP1	1	1	1	1	1	0	0	1
KT	1	<i>D. simulans</i>	Thika, Kenya	2009	J.Pool	KT24	1	0	0	1	0	1	0	1
BA	4	<i>D. simulans</i>	Puglia, Italy	2014	O.Rota-Stabelli	Bari1-4	1	1	1	1	4	0	0	4
RIV	2	<i>D. simulans</i>	Riverside,California, USA	1988	M.Turelli	Riv84, 88	0	0	0	0	0	0	2	2
YOL	2	<i>D. simulans</i>	Yolo,California, USA	2010	M.Turelli	yolo0, 10	0	0.5	0	0	0	1	1	2
IRV	1	<i>D. simulans</i>	Irvine,California, USA	2014	M.Turelli	Irv14	1	1	1	1	2	0	0	2
Mex1	1	<i>D. simulans</i>	Oaxaca, Mexico	2002	S.Castrezana	14021-0251.180	0	0	0	0	0	0	1	1
Mex2	1	<i>D. simulans</i>	Sonora, Mexico	2009	T.Markow	14021-0251.270	0	0	0	0	0	0	1	1
Mex3	1	<i>D. simulans</i>	Guanajuato, Mexico	2014	T.Markow	14021-0251.312	1	1	1	1	1	0	0	1
Mex4	1	<i>D. simulans</i>	Cusco, Peru	2009	D.Lindsley	14021-0251.279	0	1	0	0	0	1	0	1
CAM1	10	<i>D. simulans</i>	Nyasoso, Cameroon	2004	M. Vieulle	CAM1.1-1.10	0	0	0	0	0	0	10	10
CAM2	25	<i>D. simulans</i>	Nyasoso, Cameroon	2007	M. Vieulle	CAM2.1-2.25	0	0	0	0	0	0	25	25
CAM3	18	<i>D. simulans</i>	Nyasoso, Cameroon	2010	M. Vieulle	CAM3.1-3.18	0	0	0	0	0	0	18	18
TANZ	41	<i>D. simulans</i>	Liqnees, Tanzania	1996	M. Vieulle	TANZ1-41	0	0	0	0	0	0	41	41
ZIMB	6	<i>D. simulans</i>	Nyanqa, Zimbabwe	1997	M. Vieulle	ZIMB1-6	0	0	0	0	0	0	6	6
AA	11	<i>D. simulans</i>	Addis Ababa, Ethiopia	2014	W.Miller	AA1-14	0	0	0	0	0	0	12	12

population	n	species	Locality	year	collected by	Lines	proportion of strains with P-element exon				Number of strains with P-element			
							exon 0	exon 1	exon 2	exon 3	full	partial	none	total
LD	10	<i>D. sechellia</i>	La Digue, The Seychelles	2014	D.Matute	LD8, 10, 12-13, 16-20, 118	0	0	0	0	0	0	10	10
DNF	11	<i>D. sechellia</i>	Denis, The Seychelles	2014	D.Matute	DNF3-4, 10, 12, 20, 29, 47, 51, 66, 118	0	0	0	0	0	0	11	11
ANRO	9	<i>D. sechellia</i>	Mahé, The Seychelles	2014	D.Matute	ANRO1, 5-7, 16, 26, 102, 110, 219	0	0	0	0	0	0	9	9



**Supplementary Table 2.** Association between cytotypes of *D. simulans* strains and infection status for individual TE families. Note that individual exons for the P-element could be amplified from all 51 DS strains; the results shown here are for amplification of the full P-element.

FlyBase Accession	TE Identity	NCBI Accession	No. of Reads Mapping to TE	DI strain type		DS strain type		Dysgenic Cross		Non-Dysgenic Cross	
				present	absent	present	absent	Z value	p-value	Z value	p-value
FBgn0000004	17.6	X01472	2	51	0	41	0	-0.39149	0.730	-1.18624	0.80632
FBgn0000005	297	X03431	130	51	0	41	0	-2.01491	0.893	-1.42197	0.81195
FBgn0000007	1731	X07656	1430	51	0	41	0	-1.67161	0.921	0.24463	0.58643
FBgn0000155	roo	AY180917	2046	51	0	41	0	-3.30206	0.267	-1.66303	0.47847
FBgn0000199	blood	AY180916	690	51	0	41	0	-0.99903	0.944	-1.5414	0.70209
FBgn0000349	copia	X02599	162	51	0	41	0	-1.87497	0.678	-1.14494	0.56334
FBgn0000481	Doc	X17551	668	51	0	41	0	-1.06952	0.451	-0.33114	0.96105
FBgn0000638	FB	V00246	62	51	0	41	0	-2.74924	0.379	-0.9269	0.64241
FBgn0000652	F-element	AC005198	66	51	0	41	0	-1.7608	0.844	-1.18336	0.90021
FBgn0001100	G-element	X06950	2	51	0	41	0	0.46267	0.149	-0.55518	0.75958
FBgn0001167	gypsy	M12927	82	51	0	41	0	-3.00189	0.415	-1.14864	0.70764
FBgn0001181	HB	X01748	36	51	0	41	0	-2.37633	0.759	-1.31492	0.85416
FBgn0001207	HMS-Beagle	AF365402	222	51	0	41	0	-2.23781	0.801	-0.66547	0.62644
FBgn0001210	hobo	M69216	1070	51	0	41	0	-2.22755	0.820	-0.75427	0.97104
FBgn0001249	I-element	M14954	234	51	0	41	0	-2.48112	0.916	-1.05276	0.94185
FBgn0001283	jockey	M22874	56	51	0	41	0	-1.97423	0.692	-0.72319	0.75113
FBgn0002651	Dmau\mariner	M14653	356	51	0	41	0	-2.32102	0.728	-1.04207	0.9305
FBgn0002697	mdg1	X59545	108	51	0	41	0	-0.5451	0.342	0.23008	0.61598
FBgn0002698	mdg3	X95908	116	51	0	41	0	-2.46209	0.363	-1.48184	0.71795
FBgn0002745	micropia	X14037	40	51	0	41	0	-2.45071	0.870	-0.97708	0.44451
FBgn0003007	opus	AY180918	1226	51	0	41	0	-1.86519	0.891	-1.8927	0.784
<b>FBgn0003055</b>	<b>P-element</b>	<b>X06779</b>	<b>1220</b>	<b>45</b>	<b>6</b>	<b>0</b>	<b>41</b>	<b>5.87411</b>	<b>4.25E-09</b>	<b>0.372</b>	<b>0.17028</b>
FBgn0003519	Stalker	AF420242	18	51	0	41	0	-2.45438	0.598	-1.59377	0.5864
FBgn0003908	R1A1-element	X51968	208	51	0	41	0	-0.36968	0.617	-1.06801	0.82507
FBgn0003909	R2-element	X51967	120	51	0	41	0	-0.81102	0.714	-1.18112	0.94652
FBgn0004082	Tirant	X93507	0	51	0	41	0	-1.40231	0.719	-1.46964	0.82176
FBgn0004141	HeT-A	U06920	4	51	0	41	0	-0.28359	0.041	-0.38739	0.84494
FBgn0004904	TART-A	AY561850	8684	51	0	41	0	0.03328	0.417	-0.84346	0.8967
FBgn0004904	TART-B	U14101	2996	51	0	41	0	-1.09765	0.868	-0.85201	0.96626
FBgn0005384	3S18	U23420	124	51	0	41	0	-1.50212	0.746	-1.02926	0.72679
FBgn0005673	1360	AC005453	1873	51	0	41	0	-1.70071	0.937	-0.6999	0.81417
FBgn0005773	Bari1	X67681	786	51	0	41	0	-2.4699	0.271	-1.25788	0.68563
FBgn0010103	aurora-element	AB022762	68	51	0	41	0	-1.12437	0.838	-0.48058	0.80053
FBgn0010302	Burdock	U89994	808	51	0	41	0	-3.13347	0.679	-0.80122	0.95492
FBgn0013017	Dtei\I-element	M28878	2	51	0	41	0	-1.21917	0.874	-0.8261	0.95352

FlyBase Accession	TE Identity	NCBI Accession	No. of Reads Mapping to TE	DI strain type		DS strain type		Dysgenic Cross		Non-Dysgenic Cross	
				present	absent	present	absent	Z value	p-value	Z value	p-value
FBgn0000004	17.6	X01472	2	51	0	41	0	-0.39149	0.730	-1.18624	0.80632
FBgn0000005	297	X03431	130	51	0	41	0	-2.01491	0.893	-1.42197	0.81195
FBgn0000007	1731	X07656	1430	51	0	41	0	-1.67161	0.921	0.24463	0.58643
FBgn0000155	roo	AY180917	2046	51	0	41	0	-3.30206	0.267	-1.66303	0.47847
FBgn0000199	blood	AY180916	690	51	0	41	0	-0.99903	0.944	-1.5414	0.70209
FBgn0000349	copia	X02599	162	51	0	41	0	-1.87497	0.678	-1.14494	0.56334
FBgn0000481	Doc	X17551	668	51	0	41	0	-1.06952	0.451	-0.33114	0.96105
FBgn0000638	FB	V00246	62	51	0	41	0	-2.74924	0.379	-0.9269	0.64241
FBgn0000652	F-element	AC005198	66	51	0	41	0	-1.7608	0.844	-1.18336	0.90021
FBgn0001100	G-element	X06950	2	51	0	41	0	0.46267	0.149	-0.55518	0.75958
FBgn0001167	gypsy	M12927	82	51	0	41	0	-3.00189	0.415	-1.14864	0.70764
FBgn0001181	HB	X01748	36	51	0	41	0	-2.37633	0.759	-1.31492	0.85416
FBgn0001207	HMS-Beagle	AF365402	222	51	0	41	0	-2.23781	0.801	-0.66547	0.62644
FBgn0001210	hobo	M69216	1070	51	0	41	0	-2.22755	0.820	-0.75427	0.97104
FBgn0001249	I-element	M14954	234	51	0	41	0	-2.48112	0.916	-1.05276	0.94185
FBgn0001283	jockey	M22874	56	51	0	41	0	-1.97423	0.692	-0.72319	0.75113
FBgn0002651	Dmau\mariner	M14653	356	51	0	41	0	-2.32102	0.728	-1.04207	0.9305
FBgn0002697	mdg1	X59545	108	51	0	41	0	-0.5451	0.342	0.23008	0.61598
FBgn0002698	mdg3	X95908	116	51	0	41	0	-2.46209	0.363	-1.48184	0.71795
FBgn0013099	Dvir\Tv1	AF056940	1	51	0	41	0	-1.33233	0.710	-0.89187	0.89622
FBgn0014947	flea	Z27119	390	51	0	41	0	-0.69113	0.887	-1.4806	0.4231
FBgn0014967	hopper	X80025	114	51	0	41	0	-1.47803	0.883	-1.39311	0.77382
FBgn0015168	Dsim\ninja	D83207	1246	51	0	41	0	-1.91035	0.448	-1.5188	0.80386
FBgn0015945	GATE	AJ010298	788	51	0	41	0	-1.28988	0.965	-1.13806	0.87296
FBgn0023131	ZAM	AJ000387	58	51	0	41	0	-2.17033	0.990	-1.61192	0.6531
FBgn0026065	Idefix	AJ009736	2	51	0	41	0	-2.47527	0.820	-1.28902	0.85659
FBgn0026416	INE-1	U66884	8	51	0	41	0	-2.5185	0.696	-0.39372	0.64254
FBgn0026443	Dyak\TART	AF468026	152	51	0	41	0	-0.86517	0.519	-0.80413	0.63823
FBgn0040267	Transpac	AF222049	204	51	0	41	0	-0.39943	0.056	-1.01755	0.95726
FBgn0041728	Rt1a	AJ278684	16	51	0	41	0	-0.42303	0.567	-0.97235	0.75333
FBgn0042231	X-element	AF237761	536	51	0	41	0	-0.88247	0.625	-0.92549	0.97344
FBgn0042682	Rt1b	AF281636	642	51	0	41	0	-0.36436	0.552	-0.88279	0.98929
FBgn0044997	Dfun\Isfun-1	AJ309320	1	51	0	41	0	-1.85151	0.930	-1.23279	0.92776
FBgn0046110	Juan	AY180919	1111	51	0	41	0	-1.57674	0.384	-0.57506	0.99869
FBgn0063432	gypsy5	AE003485	69	51	0	41	0	-2.76696	0.940	-1.7717	0.95716
FBgn0063434	gypsy3	AC007477	301	51	0	41	0	-1.35685	0.671	-0.12006	0.97611
FBgn0063435	gypsy2	AL035631	100	51	0	41	0	-1.34233	0.451	-0.97145	0.8725
<b>FBgn0063450</b>	<b>Tom1</b>	<b>Z24451</b>	<b>118</b>	<b>51</b>	<b>0</b>	<b>31</b>	<b>10</b>	<b>-2.7071</b>	<b>0.733</b>	<b>-0.9084</b>	<b>0.70185</b>
FBgn0063755	Osvaldo	AY089271	172	51	0	41	0	-0.21893	0.673	-0.7358	0.85443
FBgn0063782	accord2	AF541947	464	51	0	41	0	-1.6155	0.725	-0.7682	0.91337

FlyBase Accession	TE Identity	NCBI Accession	No. of Reads Mapping to TE	DI strain type		DS strain type		Dysgenic Cross		Non-Dysgenic Cross	
				present	absent	present	absent	Z value	p-value	Z value	p-value
FBgn0000004	17.6	X01472	2	51	0	41	0	-0.39149	0.730	-1.18624	0.80632
FBgn0000005	297	X03431	130	51	0	41	0	-2.01491	0.893	-1.42197	0.81195
FBgn0000007	1731	X07656	1430	51	0	41	0	-1.67161	0.921	0.24463	0.58643
FBgn0000155	roo	AY180917	2046	51	0	41	0	-3.30206	0.267	-1.66303	0.47847
FBgn0000199	blood	AY180916	690	51	0	41	0	-0.99903	0.944	-1.5414	0.70209
FBgn0000349	copia	X02599	162	51	0	41	0	-1.87497	0.678	-1.14494	0.56334
FBgn0000481	Doc	X17551	668	51	0	41	0	-1.06952	0.451	-0.33114	0.96105
FBgn0000638	FB	V00246	62	51	0	41	0	-2.74924	0.379	-0.9269	0.64241
FBgn0000652	F-element	AC005198	66	51	0	41	0	-1.7608	0.844	-1.18336	0.90021
FBgn0001100	G-element	X06950	2	51	0	41	0	0.46267	0.149	-0.55518	0.75958
FBgn0001167	gypsy	M12927	82	51	0	41	0	-3.00189	0.415	-1.14864	0.70764
FBgn0001181	HB	X01748	36	51	0	41	0	-2.37633	0.759	-1.31492	0.85416
FBgn0001207	HMS-Beagle	AF365402	222	51	0	41	0	-2.23781	0.801	-0.66547	0.62644
FBgn0001210	hobo	M69216	1070	51	0	41	0	-2.22755	0.820	-0.75427	0.97104
FBgn0001249	I-element	M14954	234	51	0	41	0	-2.48112	0.916	-1.05276	0.94185
FBgn0001283	jockey	M22874	56	51	0	41	0	-1.97423	0.692	-0.72319	0.75113
FBgn0002651	Dmau\mariner	M14653	356	51	0	41	0	-2.32102	0.728	-1.04207	0.9305
FBgn0002697	mdg1	X59545	108	51	0	41	0	-0.5451	0.342	0.23008	0.61598
FBgn0002698	mdg3	X95908	116	51	0	41	0	-2.46209	0.363	-1.48184	0.71795
FBgn0063897	Stalker4	AF541949	16	51	0	41	0	-2.32908	0.674	-1.58999	0.68398
FBgn0063917	McClintock	AF541948	76	51	0	41	0	-2.45278	0.833	-1.02479	0.88787
FBgn0063919	Max-element	AJ487856	8042	51	0	41	0	-1.18875	0.784	-0.90187	0.90862
FBgn0066148	Dvir\TART	AY219709	52	51	0	41	0	-1.36247	0.013	0.33571	0.00176
FBgn0067380	invader6	NT_033778	788	51	0	41	0	-2.25478	0.564	-1.81172	0.69662
FBgn0067381	hopper2	AF541950	2	51	0	41	0	-2.77127	0.099	-1.20656	0.18121
FBgn0067382	gypsy9	AE002591	6	51	0	41	0	-1.24682	0.409	-1.1465	0.56632
FBgn0067383	gypsy8	AE003788	4	51	0	41	0	0.18771	0.475	-0.55607	0.99382
FBgn0067384	gypsy7	AE003788	4	51	0	41	0	-0.70493	0.560	-0.9531	0.86706
FBgn0067385	gypsy12	AE003789	216	51	0	41	0	-0.4693	0.144	-0.97919	0.65992
FBgnnnnnnnn	Dpse\mini-me	AC131959	1510	51	0	41	0	-1.30225	0.683	-0.7871	0.81431

**Supplementary Table 3.** Primers used for identifying the presence or absence of  
Primers used for identifying the presence or absence of P-element

Start position and direction	Sequence
Exon0 – 94 forward	GGTTGTGTGCGGACGAATTT
Exon0 – 378 reverse	CTGGTTCAGGCTCTATCACTTT
Exon1 – 615 forward	TCTACGCAAAATCTTCACGGAC
Exon1 – 1144 reverse	CTGATATACCGAGCTCTGTCCA
Exon2 – 1241 forward	TCCTGCAGATGACCATTAAAGA
Exon2 – 1900 reverse	TTAAACTGCAGTGGAGTGGGAT
Exon3 – 2181 forward	GGACAACTCTGAAAGCTGGC
Exon3 – 2545 reverse	CGTTTCGCGCTGCTAATATTAA
TIR – 3-31 forward & reverse	TGATGAAATAACATAAGGTGGTCCCGTCG