Size Matters: Non-LTR Retrotransposable Elements and Ectopic Recombination in Drosophila

Dmitri A. Petrov, Yael T. Aminetzach, Jerel C. Davis, Douda Bensasson, and Aaron E. Hirsh

Department of Biological Sciences, Stanford University

The *Drosophila melanogaster* genome contains approximately 100 distinct families of transposable elements (TEs). In the euchromatic part of the genome, each family is present in a small number of copies (5–150 copies), with individual copies of TEs often present at very low frequencies in populations. This pattern is likely to reflect a balance between the inflow of TEs by transposition and the removal of TEs by natural selection. The nature of natural selection acting against TEs remains controversial. We provide evidence that selection against chromosome abnormalities caused by ectopic recombination limits the spread of some TEs. We also demonstrate for the first time that some TE families in the Drosophila euchromatin appear to be only marginally affected by purifying selection and contain many copies at high population frequencies. We argue that TEs in these families attain high population frequencies and even reach fixation as a result of low family-wide transposition rates leading to low TE copy numbers and consequently reduced strength of selection acting on individual TE copies. Fixation of TEs in these families should provide an upward pressure on the size of intergenic sequences counterbalancing rapid DNA loss through small deletions. Copy-number–dependent selection on TE families caused by ectopic recombination may also promote diversity among TEs in the Drosophila genome.

Introduction

Transposable elements (TEs) are ubiquitous and extremely active agents of genome variability and evolution (Craig et al. 2002). They contribute the bulk of DNA in most eukaryotic genomes (Kidwell 2002), including our own, and are responsible for a sizable proportion of visible mutations (Finnegan 1992). It is clear that a thorough understanding of the regulation of the activity and abundance of TEs is essential to our comprehension of genome function and evolution.

The molecular nature, population dynamics, and evolution of TEs in the Drosophila genome have been the subject of intense investigation for the last 20 years (Finnegan and Fawcett 1986; Berg and Howe 1989; Charlesworth and Langley 1989; Charlesworth, Sniegowski, and Stephan 1994; Nuzhdin 1999; Bartolome, Maside, and Charlesworth 2002; Craig et al. 2002; Kaminker et al. 2002). These studies have provided evidence that TEs in the Drosophila genome fall into a large assortment (~100) of diverse families, with each family present in a limited copy number in the euchromatic portion of the genome (<150 per genome), and most copies present at very low population frequencies (<5%).

Spread of TEs in the Drosophila genome is likely to be limited both by regulation of the transposition rate and by natural selection against individual TE copies as TEs become more numerous. Regulation of transposition by either TE-driven or host-driven mechanisms undoubtedly takes place (Laski, Rio, and Rubin 1986; Kidwell 1989; Lozovskaya, Hartl, and Petrov 1995; Petrov et al. 1995; Lohe and Hartl 1996; Nuzhdin et al. 1998; Ketting et al. 1999; Aravin et al. 2001; Robert et al. 2001) and is very important in determining the population dynamics of TEs. However, by itself, the regulation of transposition rate is insufficient to explain all features of TE distributions. In particular, the low population frequencies of individual copies of TEs in Drosophila euchromatin must be due to natural selection against individual TE copies (Charlesworth and Charlesworth 1983; Kaplan and Brookfield 1983; Langley, Brookfield, and Kaplan 1983).

Several distinct but not mutually exclusive hypotheses about the nature of selection against individual TE copies have been proposed (for review see Nuzhdin 1999). First, individual TE copies may be deleterious because they disrupt genes, by affecting either their coding capacity or their regulation (''gene-disruption model'') (Finnegan 1992; McDonald et al. 1997). Second, translation of TE-encoded proteins may be costly, and these proteins may generate deleterious effects by nicking chromosomes and disrupting cellular processes (''deleterious TE-product expression model'') (Nuzhdin 1999). Finally, a high copy number of TEs could be harmful because ectopic recombination among numerous dispersed and heterozygous TEs generates strongly deleterious chromosome rearrangements (''ectopic recombination model'') (Montgomery, Charlesworth, and Langley 1987).

Many previous studies focused on specifically testing the ectopic recombination model by looking at low versus high recombination areas of the Drosophila genome. These studies have generally (but not always) found a higher abundance of TEs in areas of low recombination (Charlesworth, Lapid, and Canada 1992*a*, 1992*b*; Hoogland and Biemont 1996). Because it is believed that areas of low recombination also experience reduced rates of ectopic recombination (Langley et al. 1988; Montgomery et al. 1991; Goldman and Lichten 1996, 2000), these results can be taken to support the ectopic recombination model. However, in addition to presumably having a lower rate of ectopic recombination, low-recombination areas also have lower densities of genes (Adams Celniker, and Holt 2000), are likely to permit lower levels of gene expression (Becker 1995; Henikoff 1995; Lu, Ma, and Eissenberg 1998; Birchler, Bhadra, and Bhadra 2000), and experience less efficient selection because of the Hill-Robertson effect (Hill and Robertson 1966; although see Charlesworth and Charlesworth 1983 for an investigation of the

Key words: transposable elements, ectopic recombination, copy number maintenance.

E-mail: dpetrov@stanford.edu.

Mol. Biol. Evol. 20(6):880–892. 2003
DOI: 10.1093/molbev/msg102

Hill-Robertson effect on TEs in Drosophila). Thus all three explicit selection hypotheses predict higher copy numbers and higher population frequencies of TEs in genomic regions of low recombination. The fact that this prediction is borne out empirically does not discriminate well among the three current hypotheses.

In this study we attempt to obviate these difficulties by focusing exclusively on one class of TEs (non-long terminal repeat [LTR] retroelements) in the high-recombination areas of the *D. melanogaster* genome. Non-LTR elements are abundant in the Drosophila genome (Berezikov, Bucheton, and Busseau 2000), and are attractive as a model system for a number of reasons. Because they evolve primarily, or possibly even exclusively, through vertical transmission (Malik, Burke, and Eickbush 1999) and cannot excise precisely from the genome, we can provisionally ignore horizontal transfer or excision in considering their population dynamics. In addition, they commonly generate 5′-truncated DOA (dead-on-arrival) elements as a natural outcome of transposition (Luan et al. 1993). These DOA elements are not transcribed and do not encode functional proteins. Thus they cannot generate potentially deleterious transcripts and proteins, allowing us to discount selection against deleterious expression of TE-encoded proteins as a force acting against individual DOA copies.

Moreover, the variable size of DOA elements at the time of transposition may allow us to discriminate between the ectopic recombination and gene-disruption hypotheses. Specifically, we reasoned that selection against the deleterious effects of ectopic recombination should affect longer elements more strongly than shorter ones, as they represent longer targets for homologous pairing (Dray and Gloor 1997). In a sense, the variation in length among newly transposed non-LTR elements allows us to study variation of the recombination rate among individual TEs, rather than among whole genomic areas. This in turn allows us to escape the confounding correlations of background recombination rate, gene density, and chromatin states in the interpretation of the results. It also allows us to simplify the analysis further by concentrating only on the high recombination areas of the *D. melanogaster* genome, thus reducing the probability of selective sweeps and background selection (Smith and Haigh 1974; Berry, Ajioka, and Kreitman 1991; Charlesworth, Morgan, and Charlesworth 1993; Hudson and Kaplan 1995) playing a significant role in determining the population dynamics of the studied TEs.

Our analysis provides new evidence that selection against ectopic recombination, rather than against costly expression of TE proteins or gene disruption by individual TEs, limits the spread of at least some non-LTR elements in the Drosophila genome. We also demonstrate that some non-LTR families appear to be under very weak purifying selection, in that they include many insertions that reach high population frequencies and even fixation in the *D. melanogaster* euchromatin. Combined with the evidence for the importance of ectopic recombination, the observation of TEs at high frequencies suggests that transposition rates vary significantly among TE families, and possibly over longer time scales for the same TE family. We discuss a hypothesis whereby transposition rate for a particular TE family can decline sharply for a period of time, leading to reduced copy numbers and ectopic recombination rates among remaining TE copies. During these periods selection acting on the remaining TE copies may be sufficiently weak to allow fixation of multiple TEs in the Drosophila euchromatin by genetic drift.

Materials and Methods

Identifying TE Insertions in the Drosophila Genome

Consensus sequences of each of the full-length transposable elements were used as *blastn* queries on the 13th Nov 2000 release of the *D. melanogaster* genome at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?DATABASE=drosoph), with the E (expect) set at 0.0001 (expected number of false positives per search) and otherwise default settings. We used the following query sequences: M22874 (*Jockey*); *X*17551 (*Doc*); *X*77571 pos 651.5776 (*BS*); AF237761.1 (*X*). These query sequences were taken from Flybase (http://flybase.bio.indiana.edu/transposons/lk/melanogaster-transposon.html). They are thought to represent the consensus sequences of full-length, active copies of TEs in these families. Hits less than 2.5 kb apart in the same contig were treated as parts of the same transposable element. All insertions were further verified through visual inspection of alignments with their respective consensus sequences. The divergences of the individual copies from the query sequences are listed in table 1. Using a perl script (available by request) we identified the distance to the nearest gene (including predicted genes) for every element by using the estimated start and end positions of the element reported in the Blast output and the annotations for the contig in which it occurred (table 1).

Determining the Length of TEs

The length of non-LTR elements can change after insertion through secondary insertions and deletions. In our analysis we use two different measures of TEs. One measure is the current length of each TE. The other is our best guess of the original length of the TE prior to any secondary deletions and insertions. To estimate the original length of the elements we used both the alignment with the consensus sequence of the element and the presence of target site duplications generated at the time of transposition. The presence of such sites indicates strongly that an element has not been secondarily truncated at either end. In cases where we did not find target site duplications we assumed that the length of the consensus sequence from the 5′ edge of the alignment with a particular TE to the 3′ end of the consensus sequence represents the length of this TE at the time of transposition.

Drosophila Strains

We used the sequenced strain (y^1; cn^1; bw^1) as a positive control in our polymerase chain reactions (PCRs). The population sampling was done in 10 American and 8 Tunisian strains. The American strains are isofemale

Table 1
The Population and Genomic Data for the Four Studied Non-LTR Families

Family	TE Name	Size at Insertion (bp)	Size (bp)	Map Position	To Gene (bp)	Primer+	Primer−	Pooled Freq.	Wi Freq.	Tunisia Freq.	Divergence Relative to Consensus Copy
BS	2987*BS*	108	108	7C-7D	NO GENES	TTTATGCTTCGCAAGGTTAG	CTGAACTGAACATTCTTGACG	100	100	100	0.88%
BS	3443*BS*	334	334	7E1-E2	20631	TTTGTTTGATTCCGTTTGG	TCAATTGACATGCAATGTTTC	50	56	43	0.65%
BS	3457*BS*	131	131	58D8-E3	977	GCAATTTCCAGAGGCGTAGCA	CGACGATTTTAATAAATCATA	73	71	75	0.00%
BS	3494*BS*	35	35	12C4-C7	3385	TCAACATCTGGCTGAGAAAC	GGCACGCATAAATAAAAACA	100	100	100	5.56%
BS	3504*BS*	83	83	15D4-15E1	0	GTCTCCTTAACGGTTGGACT	TTTTGGACCAGAACAATCAC	100	100	100	2.38%
BS	3560*BS*-1	540	540	65E1-E5	751	GGGCAAGGGCAACTACTGGAC	GCTGATTTGCATAAAACGCTA	6	11	0	0.75%
BS	3560*BS*-2	126	126	65D6-E6	675	GCATTTAAACTGAACGCAACC	TAGCCATAAATAATGACGAAC	59	56	63	0.00%
BS	3618*BS*	112	112	28A5-28A6	709	CTTGTTTTTGCTCTCCCTCT	TTAACTGGCGCTTAAGAATG	53	56	50	0.00%
BS	3717*BS*	356	356	90A1-90A2	9053	CGACTGTCAATCAGAGACGAA	CTGCATATATAAATCATACGC	13	0	25	0.58%
BS	3730*BS*	800	800	92E7-92E8	199	GAGGAGGTCAACAACATATTC	GCGTCATTTACTGGTTCTACT	38	38	38	0.55%
BS	3748*BS*	164	164	96A3-96A5	2186	GGTTCCGCTCCACTGTCATGT	TAAATAAATGAAACCGTCGAA	8	13	0	0.95%
Doc	3435*Doc*	4750	4750	5B2-5B3	625	TTTGAACGACAACTGCTTGG	GCACTGATTGCTACGGATAAGTT	0	0	0	0.28%
Doc	3437*Doc*	4744	4746	5E6-6A2	20178	TTTATAAAATTGGCAAGGAAT	GGGCTGTAATTGCAGTCCACA	0	0	0	0.33%
Doc	3477*Doc*	429	429	63C2-63C3	0	TATTCGGAACTATCTAACATA	TGTTTTCTTGCCACTTCTTTA	0	0	0	0.25%
Doc	3496*Doc*	4839	4839	12F4-12F5	135	CGACTGAATTTCAATTTGATC	TTTGGTAACCTCTGCTCAACA	0	0	0	0.08%
Doc	3533*Doc*	3859	2937	70F1-70F4	280	ATGGCATCTTGAATAGTGGTT	GAATCCACAATGATGACCTTC	0	0	0	0.52%
Doc	3544*Doc*	4829	4829	68C13-68C15	0	CTAAAGTGCCTGTCTCACAGC	TTGAGATTTCCAACAAACAGT	0	0	0	0.28%
Doc	3548*Doc*	2315	2275	67D10-67D11	451	GCATTTAAATTGGATGTGTTG	ATTTTTCATACGGTCGATCGG	0	0	0	0.44%
Doc	3551*Doc*	1631	1631	67B4-67B7	1921	TTGGGAAAGGGGTGCTAATCT	GCTAATTAGCGAAATGATGTT	0	0	0	0.25%
Doc	3552*Doc*	2774	2774	67B1-67B2	0	CGGAGAAAAACAAGGATGTAT	ATTTCCGTTCTAATTGAGTTC	0	0	0	0.36%
Doc	3561*Doc*-1	4833	4833	65C3-65D3	18076	CGGGCGACGGAGGAGGACTTT	GCGGGGCCACCACTACACACG	0	0	0	0.19%
Doc	3561*Doc*-2	1963	1963	65C5-65D2	6860	TATTTCAGGGGGTAAGGATTG	TAATAATTTAAAGCGGCGTAA	0	0	0	0.00%
Doc	3564*Doc*	2317	2317	65A2-65A2	346	GGGAATTTCGGAAACGTAGAT	GGAGGACCCTACACTGTTAGA	6	11	0	0.10%
Doc	3609*Doc*	2325	2325	25D5-25D6	508	ATATTGCTTTACAATTATCTG	CCACGCGAATTATGTCTGTTT	0	0	0	0.75%
Doc	3749*Doc*-1	4736	4738	96A20-96B10	0	GCGCCTTCGTAGGAGATTTAG	AATTGTTTTTCGATACATAAC	0	0	0	0.26%
Doc	3749*Doc*-2	2328	2328	96B7-96B10	392	TGTTATGCCTGTACCATTATG	TCCGAAATTCAATTAGGTTGC	0	0	0	0.21%
Doc	*3751Doc*	4513	4513	96C8-96C9	0	CGGCTGCCATTACACTT	GACAACTACGCCACTAATATG	76	100	50	0.53%
Jockey	3432J-1	5020	5020	4C10-4C11	6766	TCATGACAAAGCGAAAGTCGG	ACCCTATCTCAAGTCTAGAAG	0	0	0	0.56%
Jockey	3432J-2	368	366	4C9-4C12	3803	GGAATTTGTTGAACGGGATGC	TTCTGCGGTTGAAATCTAAGC	0	0	0	0.30%
Jockey	3434J	1192	1167	4F2-4F4	1503	CCAAATAAAATGTTTCGAAGA	TGGCTGTTGTTGTTCGAGCAC	0	0	0	0.43%
Jockey	3435J	152	42	5A11-5A13	258	TGGTTTGAGCTCCGCTTAATG	TTGAGCGCAGAAGGAATCCTA	100	100	100	21.43%
Jockey	3441J	376	376	7B5-7B8	6268	CTACTAGTTCCACGTGATCTT	ACACCCTTTTACCTAAGCATA	6	10	0	0.00%
Jockey	3445J	372	372	8B8-8C1	500	TGTTTCTCCTTGCCGTTTCCT	AAACGCGCCGCATGTA	0	0	0	0.00%
Jockey	3447J	1452	1445	8D10-8E7	515	GCCAGAAACTCGACCTCCT	CCATCATTGCTCCGATAGATA	0	0	0	0.28%
Jockey	3457J-1	3498	3485	58E3-58E7	20820	CATGGAATATCCGAATTGGTA	GGAAAGCTCCAAGACACGAAC	0	0	0	0.32%
Jockey	3457J-2	1723	1713	58D4-58D6	948	TTGAGTACGAAAAATTAGTTA	ATGTACGGTCAATGAAGTTAG	0	0	0	0.53%
Jockey	3457J-3	373	370	58D4-58D6	11613	CTTCTCGCACTTTATGGACCC	TATTCAAAAGCCCGATAA	0	0	0	0.00%
Jockey	3458J	353	345	59A1-59A2	1198	TGCTGCAACTCAACTATGTCT	ATTTTATAGGGCGCATATGTG	0	0	0	0.00%
Jockey	3478J	382	370	63D2-63D3	3797	ATGGCAAGTGCTGCAGTCGTA	AAATGTTGCTTTTAACAAGTG	0	0	0	0.00%
Jockey	3492J-1	1080	1080	11E13-11F5	9783	CTTTGCCTGCGATGATTAACT	ATTTTTATGGCCTACTCTCA	0	0	0	0.29%
Jockey	3492J-2	616	616	11E13-11F5	11047	CTTTGCCTGCGATGATTAACT	ATTTTTATGGCCTACTCTCA	0	0	0	0.52%
Jockey	3503J	353	339	15B1-15B4	0	CATCGGCTACGTGATTGTG	GAGCTTGCGTTGCTGGAATAG	6	0	13	0.00%
Jockey	3545J	3148	3148	68B3-68C1	8186	TCGCTATTTTTGAAGACTTAC	CCAATTTGATGTGCATCTTA	0	0	0	0.48%
Jockey	3551J-1	5010	5010	67B8-67B9	0	GAAACTTTTTATCCGGGCTAC	AGCGCCAAATTTACCACGTC	0	0	0	0.39%
Jockey	3551J-2	284	284	67B8-67B9	0	GGGAATATGTGCAAGGG	TTTTATGTTTCGATGGATACG	0	0	0	0.00%
Jockey	3552J	609	568	67A1-67A3	4096	TTTACTGAACTATGCCTACAG	AAGAATTTTCTCCGTTAGATA	0	0	0	1.27%
Jockey	3564J	5090	2583	65A1-65A2	5186	GGCCTGCAGTTGAGTCCT	CCTACTACTACTTTGCGTCC	6	11	0	0.70%

Table 1
Continued

Family	TE Name	Size at Insertion (bp)	Size (bp)	Map Position	To Gene (bp)	Primer+	Primer−	Pooled Freq.	Wi Freq.	Tunisia Freq.	Divergence Relative to Consensus Copy
Jockey	3565J	262	262	64E1-64E2	913	TTTCTTGCCAAATGACGACAG	TTAATGTGTTCGCTACCTTGA	8	0	14	0.00%
Jockey	3575J	5007	5007	25B5-25B7	2591	GAAAAATACTCGCACGGTC	TGATAAATTGAATGAACACGC	0	0	0	0.52%
Jockey	3585J	248	248	22A4-22B1	4153	CAAGGATCAAGGACGTTCAAG	GGCGGTGTGCTAAGTC	0	0	0	0.00%
Jockey	3619J	5000	109	28D5-28D10	37375	TGGCTAAAAGCACGAGGGATT	CTTTGGGTCGATGTAACTGTT	6	0	13	1.05%
Jockey	3624J	2894	2894	30B3-30B5	715	CATTTTTCACCAGGAACGTTA	AGGAAATCGCATGGTTGAC	0	0	0	0.35%
Jockey	3746J	365	365	95D4-95D7	9465	ATTTATGTGGTTCGCTTACGC	TGTATTGCTGCCCCGTCCTGT	0	0	0	0.30%
Jockey	3757J	319	319	97C1-97C2	1879	CTTGTCTGCCATTTGCCGTCT	ACATTTCAATTTGCGCTTACC	0	0	0	0.64%
Jockey	3764J-1	5000	2413	98C2-98C3	1131	TAAATCAAGCCGGAACAC	GCCCAGCTAAGCGATCCTAAT	7	14	0	0.58%
Jockey	3764J-2	370	370	98C2-98C3	1131	CTAAGGAATCTGCGTTACTGA	CCGACCGATTTGGTAGAATGC	0	0	0	0.30%
Jockey	3766J	5000	5000	98E3-98E4	5926	GCGCAGGGTCTCCGAACACTT	AGCCCAAGATATAATCCAATC	0	0	0	0.50%
Jockey	3769J-1	1201	1201	99A7-99B1	1045	GGGCAACACAATGAGACACTT	TCCCACAACCCACTGACATAC	12	22	0	0.25%
Jockey	3769J-2	247	247	99A7-99B1	87957	TGCGTTGTACCCTACCGT	AAACTAACAGGGCATATAACT	0	0	0	0.00%
Jockey	3794J	3232	3232	56E4-56E5	3678	CCCAACAATAACAATCGATGT	TCTTTGAGCATCCAATTTGAG	0	0	0	0.19%
Jockey	3801J	286	286	55B1-55B2	459	ACGGCTCCAATCGATGAGTCA	GAAAGGCCGGAGACTAGGTCA	0	0	0	0.00%
Jockey	3803J	341	341	54C10-54C12	0	AACATAGAGAAAATCGGTTAT	GTTCAATTTCGACACGGCTAA	6	0	13	0.88%
Jockey	3804J	5016	5016	54A1-54A3	6816	AAAAGTTTGTGGCACTAAAT	AAAAATAGTGGACGAAGATTC	0	0	0	0.53%
Jockey	3812J	254	254	51D8-51E1	11020	TATCCAGCAATCCCATAGTGA	GCCTTAGTTGCCTACAGTTGT	12	22	0	0.00%
X	3488*X*	1252	1252	11A11-11A12	0	AAAGAAAACGCCGAGCAAGGT	GATGACTGCCAGCAAGTTGAC	63	67	57	0.34%
X	3499*X*	1510	326	13D2-13D4	413	TGTGTGCCATAAAGTAATCAG	AAAACGAAACGAAAATAGGTA	100	100	100	24.21%
X	3505*X*	409	216	15E7-15F1	11234	GCAACAAAGGTAAACTTATAT	CGCTCCCCGGATGGTACATCT	100	100	100	7.41%
X	3535*X*	4747	4747	70D2-70D5	12611	TCCCCTCTTGTGACTCGTTGC	TTGTCGGTGGCCCTGTAGGAG	0	0	0	0.19%
X	3546*X*	1657	1657	68A3-68A4	50	GCATTATGAAAATTACGTTCT	TACAGGATCAGCTGCTACCAG	13	13	13	0.19%

strains (Wi1, Wi3, Wi15, Wi18, Wi35, Wi41, Wi45, Wi68, Wi69, Wi83) that were collected at the Wolfskill Orchard, Davis, CA, by Sergey Nuzhdin, and that have been further subjected to over 30 generations of brother-sister matings (S. Nuzhdin, personal communication). The African strains are isofemale lines (T1, T3, T12, T13, T17, T18, T27, T28) collected in Tunisia, Africa, by Charles Aquadro. The high level of isogenicity in all of these strains was further confirmed by the fact that 6 TE insertions that were highly polymorphic across the tested strains (frequency ranging from 42% to 75%), did not show a single case of presence/absence polymorphism within any of the strains. Overall there were 53 cases of the presence and 41 cases of the absence of one of these TEs.

Population Assays

The population frequencies of TEs were assayed by amplifying individual TEs using primer pairs in the flanking regions. The primers were designed with the aid of the Oligo 6 software package (Molecular Biology Insights, Inc., 1988). The primer sequences are provided in table 1.

DNA Sequencing

We verified the PCR identification of the fixed TEs by sequencing them in a number of strains (see *Results*). In each case PCR reactions were enzymatically cleaned with Exonuclease I and Shrimp Alkaline Phosphatase (1 unit of Shrimp Alkaline, 5 units of Exonuclease I, 1.2 μl of 10-fold reaction buffer-0.2 M Tris, and 0.1 M $MgCl_2$ added to a 10 μl PCR reaction; mixture was incubated at 37°C for 45 min, and enzymes were inactivated at 70°C for 15 min), and were cycle-sequenced in quarter-reactions according to the ABI 377 sequencing protocol with Big Dye (4 μl of the cleaned PCR products, 2 μl of the Big Dye, 5 μl of the sequencing buffer [160 mM Tris Ph 9.0, 10 mM $MgCl_2$], 0.17 μl primer previously diluted to 20 μM, 5 μl ddH_2O) under standard cycling conditions (96°C for 1 min, 24 cycles of: 96°C for 10 s, 1.0°C/s to 50.0°C, 50°C for 5 s, 1.0°C/s to 60.0°C, 60.0°C for 4 min, 1.0°C/s to 96.0°C). These reactions were precipitated using ethanol and $MgSO_4$ as described in the ABI sequencing manual, and the sequences were visualized on an ABI 377 automated sequencer. The primers used for amplifying the TEs (table 1) were also used for sequencing. The representative sequences have been deposited to GenBank under the accession numbers AY226801 through AY226814.

Estimating Intensity of Natural Selection

We make use of a diffusion approximation and the resulting sojourn time density function (Ewens 1979, Eqs. 4.22–4.26 & 5.47; Nagylaki 1974) to estimate the probability that an element is at a particular frequency in the population. We assume an infinite number of insertion sites, as in Kaplan and Brookfield (1983). (For an approach that also makes use of the diffusion approximation, but applies slightly different simplifying assumptions, see Charlesworth and Charlesworth [1983]). We assume that the fitness of individuals who are homozygous for the element is $1+s,\ s>-1$; heterozygotes have fitness $1+hs,\ hs>-1$; and homozygotes without the element have fitness 1. Let y represent the vector $\{x, N, s, h\}$, where $x, 0 \le x \le 1$, is the frequency of an element in the population of N diploid individuals. Under these assumptions, the drift and diffusion terms of the diffusion approximation of the standard Wright-Fisher model are $m[y] = 2Ns(1-x)x(h+x-2hx)$ and $v[x] = x(1-\mathrm{x})$. Let $\bar{\tau}[y]\Delta x$ be the expected amount of time that an element that is initially present as a single copy spends on the frequency interval $I{:}(x, x+\Delta x)$ before it is absorbed at $x=0$ or $x=1$. Under the standard assumptions of the diffusion approximation,

$$\bar{\tau}[y] = \frac{2}{v[x]\psi[y]g[0,1]}(g[1/2N,1]g[0,x]\theta[p-x] + g[0,1/2N]g[x,1]\theta[x-p])$$

where

$$\theta[z] = \begin{cases} 1, & z>0 \\ 1/2, & z=0; \\ 0, & z<0 \end{cases} \quad \psi[y] = e^{-2\int m[y]/v[x]dx}; \quad \text{and} \quad g[\mathrm{a},\mathrm{b}] = \int_a^b \psi[y]dx.$$

Assuming that the number of elements in the population is large, and that the population is at transposition–selection equilibrium, the frequency spectrum of elements is given approximately by $F[y] = \bar{\tau}[y]/\int_0^1 \bar{\tau}[y]dx$. Because we measured the frequencies of elements that were initially identified in the sequenced *D. melanogaster* genome, the appropriate distribution is not $F[y]$, but rather the distribution of element frequency conditional on the element being present in the first individual sampled (the sequenced genome). Taking $F[y]$ as the prior and x as the probability that an element at frequency x in the population is present in the sequenced genome, we obtain the posterior probability density function $F'[y] = x\bar{\tau}[y]/\int_0^1 x\bar{\tau}[y]dx$. At a given site, j, at which the sequenced genome bears an element, the probability that i of k sequences sampled from the population also bear the element is then $P_j[N,s,h] = \binom{k}{i}\int_0^1 F'[y]x^i(1-x)^{k-i}dx$. Assuming that unlinked sites are independent, the likelihood of the frequency data across n separate sites is simply $L[N,s,h] = \Pi_{j=1}^{n} P_j$. The maximum likelihood estimate of s (s*) or h (h*) was found by maximizing L, given the frequency data and fixed values of the other parameters (N and s or h). The 95% confidence intervals around s* were calculated by numerically solving for s such that $2\ln[L[N,s^*,h]/L[N,s,h]] = a$, where a was chosen such that $\int_0^a \chi^2_{[1]} = 0.025$ or $\int_0^a \chi^2_{[1]} = 0.975$. We estimated selection coefficients (s*) assuming semi-dominance ($h = 1/2$) and complete dominance ($h = 1$). The two models yielded similar estimates of s*, and qualitative conclusions were identical. We also investigated a model allowing for underdominance, which is likely to be a more appropriate depiction of the operation of selection against ectopic recombination events. In this model, we assumed that individuals homozygous for the element experience only a small decrease in fitness ($Ns = -1$), whereas heterozygotes may experience a larger fitness decrement,

Table 2
Numbers of Mapped TEs in the High and Low Recombination Areas of *D. melanogaster* Euchromatin

	Doc	*Jockey*	*BS*	*X*
High recombination areas[a]	16	37	11	5
Low recombination areas[b]	35	24	29	16
Total	51	60	40	21

[a] High recombination areas of *D. melanogaster* euchromatin defined as in (Charlesworth 1996) (*X*: 3C3-15F3; 2L: 22A1-31A1; 2R: 50F9-59F8; 3L: 62A12-71A1; 3R: 89F4-99F1).

[b] Low recombination areas of *D. melanogaster* euchromatin are defined as all mapped location other than (*X*: 3C3-15F3; 2L: 22A1-31A1; 2R: 50F9-59F8; 3L: 62A12-71A1; 3R: 89F4-99F1).

reflecting the increased probability of ectopic recombination in heterozygotes. Thus, *s* was set to a small negative value while the heterozygous effect, *h**, was estimated. Again, qualitative conclusions about the strength of selection against each family of TE were unchanged. Estimates reported in *Results* were calculated assuming semi-dominance and an effective population size of 10^6 (Kreitman 1983). Qualitative conclusions were unchanged under different assumptions about dominance and with $N_e = 10^5$ (Schug et al. 1998).

Detecting Adaptive Events Within a Maximum Likelihood Framework

To detect putative adaptive insertions of TEs, we used a likelihood ratio test for heterogeneity in the selection coefficients of transposable elements that belong to a single family. This test works by comparing two nested models of transposable element selective effects: Model 1 (M_1) assumes that all TEs are subject to the same strength of selection, and Model 2 (M_2) allows each TE to possess one of two different selection coefficients. Under M_1, only one free parameter, s^*, the selection coefficient of all TEs of a given family, is estimated from the frequency data. Under M_2, three free parameters are estimated from the data: s_1 and s_2 are two distinct selection coefficients, and p is the proportion of elements with selection coefficient s_1. The likelihood of the data given M_1 is $L[N,s^*,h]$, the function described in the previous section. The likelihood of the data given M_2 is calculated similarly, but with

$$P_j[N,s,h] = p \times \binom{k}{i} \int_0^1 F'[y_1]x^i(1-x)^{k-i}\,dx + (1-p) \times \binom{k}{i} \int_0^1 F'[y_2]x^i(1-x)^{k-i}\,dx$$

in which $y_1 = \{x,N,s,h\}$. Likelihood tests for heterogeneous selection coefficients were performed for the same assumptions about N and h as described above.

Results

Identifying Individual Non-LTR Elements in the Drosophila Euchromatin

We examined the distribution and population dynamics of four different non-LTR elements present in Drosophila euchromatin. All the elements belong to the *Jockey*-clade (Malik, Burke, and Eickbush 1999), and their active copies are approximately equal in length: *Jockey* (Mizrokhi et al. 1985) – 5.1 kb, *Doc* (Bender, Spierer, and Hogness 1983) – 4.8 kb, *BS* (Udomkit et al. 1995) – 5.2 kb, and *X* element (Tudor et al. 2001) – 4.8 kb. These elements do not transpose into specific sites and consequently are dispersed across the genome. In this study, we focused on the high recombination (HR) areas in the *D. melanogaster* euchromatin (*X*: 3C3–15F3; 2L: 22A1–31A1; 2R: 50F9–59F8; 3L: 62A12–71A1; 3R: 89F4–99F1) (Charlesworth 1996). This simplifies the analysis by reducing the probability of selective sweeps and background selection (Smith and Haigh 1974; Berry, Ajioka, and Kreitman 1991; Charlesworth, Morgan, and Charlesworth 1993; Hudson and Kaplan 1995).

We identified all unambiguous copies of these four elements in the HR euchromatin of the sequenced *D. melanogaster* genome using Blast (table 1 and table 2), designed PCR primers to the flanking regions of each copy (table 1), and used them to assess the frequency of individual elements in 18 natural strains of *D. melanogaster* collected in North America (California, USA) and Africa (Tunisia) (see *Materials and Methods*). We used 10 different, isofemale, highly inbred strains from North America and 8 different isofemale strains from Tunisia. Because PCR failed in ~11% of the cases, the number of tested strains was reduced from 18 to an average of 16 per TE insertion. The rate of PCR failure did not correlate significantly with TE length (Kendall's $\tau = -0.017$, $P = 0.83$) or vary significantly among TE families (G-test, 3 df; $P = 0.96$). In the course of the experiments we further verified the isogenicity of the strains by finding that 6 highly polymorphic TE insertions (varying from 42% to 75% in frequency across the strains) did not show a single case of a polymorphism within the strains.

A Putatively Adaptive Recent Insertion of a *Doc* Element

One copy of *Doc* was present in all tested American strains (9 out of 9) and was highly polymorphic (4 out of 8) in the Tunisian strains. Additional sampling demonstrated that the frequency of this element varied sharply across the worldwide sample of Drosophila populations, with generally higher frequencies in the American than in the African populations (data not shown). This element is the only one out of 16 sampled *Doc* elements that is present at an appreciable frequency (greater than 10%) in any population. We used a likelihood approach (see *Materials and Methods*) to demonstrate that this copy of *Doc* is an outlier. If we pool the data across all populations we can test the model that all *Doc* elements are subject to the same strength of purifying selection. The model that assumes that all *Doc* copies have the same selective coefficients generates the MLE of selection at $N_e s = -8.6$ with the $\ln L = -22.5$. The model that allows for two coefficients improves the likelihood ($\ln L = -10.2$) and estimates that 93.8% of TEs in the sample (~15 out of 16) have $N_e s = -136$ and 6.2% (~1 out of 16) have $N_e s = 3$. Thus we have clear evidence of heterogeneity of selective coefficients in the sample (LRTstatistic = 24.6, using χ^2 distribution with 2 df, $P \ll 0.0001$). The frequent copy is also clearly the outlier: only the removal of the frequent copy of the *Doc* element from the analysis removes apparent heterogeneity ($P \approx 1$).

Furthermore, this *Doc* copy apparently truncates a conserved phosphotransferase-encoding gene (CG10618), suggesting that this insertion is likely to have a selective effect. The unusually high frequency of this *Doc* copy, the sharp variability of its frequency in different populations of *D. melanogaster*, and the reasonable expectation of the presence of a selective effect are all signs of the putative adaptive effect generated by the insertion of this *Doc* element (or of its tight linkage to an adaptive mutation in a neighboring locus). We are investigating these possibilities.

Note that the other TE families (*Jockey*, *BS*, ad *X*) showed no sign ($P \approx 1$) of heterogeneity of selection coefficients and that all other TEs were present at indistinguishable frequencies in the US and Tunisian populations, either for TEs within each family (results not shown) or for all TEs pooled together (tests of H_0: Tunisian and American frequencies are drawn from the same distribution for all TEs; Wilcoxon signed-ranks test, $P = 0.48$; *t*-test, 67 df, $P = 0.36$; *t*-test after the angular ($\arcsin\sqrt{\text{p}}$) transformation of the data, 67 df, $P = 0.39$). Testing individual TE copies also revealed no instances of disparate frequencies in American and Tunisian populations. All TEs fixed in one population were fixed in the other, and the TEs present in intermediate frequencies in one population were also present in intermediate frequency in the other population (G-test, 1 df; *P* values range from 1 to 0.77). Transposable elements that were not found in one population were either not found in the other population or present in very low frequencies (maximum frequency was 2 out of 8 found for 3717BS in the Tunisia population; this frequency is not significantly different from the 0 out of 8 frequency found for the same element in the American population; Fisher's exact test, $P = 0.48$). In the remainder of the analysis we will exclude the frequent *Doc* element and will discuss the data pooled across the Tunisian and American populations for the rest of the elements.

Frequency Spectra Vary Sharply Across TE Families

The frequency distribution for each family is shown in figure 1. Different element families clearly exhibit distinct frequency distributions (Kruskal-Wallis rank test, $P < 0.001$ both for all TEs and for the subset of polymorphic TEs only). There appear to be two distinct kinds of frequency distributions. One is exemplified by *Jockey* and *Doc*, in which all polymorphic copies are present at low frequencies. In contrast, *BS* and *X* elements are found at all frequencies: low, intermediate, and high. Whereas we cannot distinguish the frequency distribution of *Jockey* from that of *Doc* (Mann-Whitney rank test, $P = 0.12$; *t*-test, 49 df, $P = 0.13$) or *BS* versus *X* families (Mann-Whitney rank test, $P = 0.54$; *t*-test, 9 df, $P = 0.46$), the combined distribution of *Jockey* and *Doc* elements is sharply different from the combined distribution of *BS* and *X* elements (Mann-Whitney rank test, $P < 0.0001$).

Interestingly, there appears to be a rough negative correlation between the copy number of an element and the population frequency of polymorphic copies (table 2). *Doc* and *Jockey* have many copies in the euchromatin, but each copy is present at a very low population frequency. The reverse is true for *X* and *BS* elements.

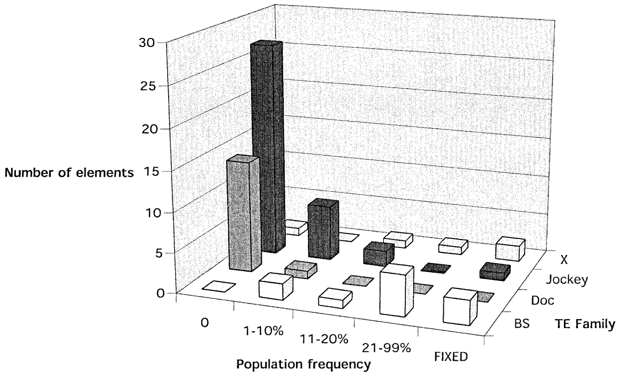

FIG. 1.—Histogram of population frequencies. Elements at frequency 0 were found only in the sequenced *D. melanogaster* genome. The putatively adaptive insertion of a *Doc* element is excluded.

Estimating the Intensity of Selection

To understand the nature of population forces acting on these elements, we conducted maximum likelihood analysis of the strength of natural selection consistent with the observed frequency distributions of polymorphic elements. We assumed transposition–selection balance and used a diffusion approximation to obtain the expected frequency spectrum of elements as a function of the strength of selection. We adjusted this distribution to account for the fact that, by studying only copies that were initially found in the sequenced genome, we effectively presampled elements in proportion to their population frequencies (see *Materials and Methods*).

Using this probability distribution, we find evidence for strong ($N_e s \ll -1$) purifying selection acting on *Doc* (95% confidence interval: $-4300 < N_e s < -23$) and *Jockey* elements ($-63 < N_e s < -14$). Such strong purifying selection is entirely consistent with previous studies of TEs in Drosophila euchromatin (Charlesworth and Langley 1989). Surprisingly, the frequency distributions of *BS* ($-2.7 < N_e s < 2.9$) and *X* elements show no signs of purifying natural selection ($-7.0 < N_e s < 3.0$). Whereas we can easily reject neutrality ($N_e s = 0$) for *Doc* and *Jockey* (log-likelihood ratio of the maximum likelihood versus likelihood value found by setting $s = 0$, $P \ll 0.001$ in both cases), we cannot do so for *BS* (log-likelihood test, $P = 0.5$) or *X* (log-likelihood test, $P = 0.15$).

Fixation of TEs in the Drosophila Euchromatin

We determined that several of the elements were present in all of the tested strains (2 *X*, 3 *BS*, and 1 *Jockey*). We verified this observation by diagnostically sequencing each of these six elements in several strains (minimum 2 and maximum 17 strains). In each case we confirmed the presence of the identified TE copy in all tested strains with the sequence of the junctions >99% similar to that found in the sequenced Drosophila genome (data not shown).

The comparison of the sequences of these TEs (taken from the *D. melanogaster* genome sequence database) with the full-length consensus sequences of these elements (table 1) is consistent with fixation of these elements in *D. melanogaster*. These 6 elements are significantly more divergent from their respective consensus sequences than the elements determined to be polymorphic (Mann-Whitney, $P < 0.0001$). Based on the level of divergence and using the rate of neutral evolution of 2% per Myr (Moriyama and Powell 1997), we can estimate that the three *BS* elements fixed fairly recently (~0.4 to 3 MYA). The two fixed *X* elements and one *Jockey* element appear much more ancient (3.5 to 10 Myr), with multiple point substitutions riddling their sequence (*Jockey* element – 21% divergence from the consensus *Jockey* sequence; *X* elements – 7% to 24% divergence from the consensus *X* sequence).

Discussion

There are several competing hypotheses about the nature of selective forces acting to restrict copy number of TEs in the Drosophila chromatin and to generate patterns of low population frequency of individual TE copies. The three hypotheses are (1) selection against deleterious effects of TE insertions on neighboring genes (''gene-disruption model'') (Finnegan 1992; McDonald et al. 1997), (2) selection against deleterious effects of TE-generated products (''deleterious TE-product expression model'') (Nuzhdin 1999), and (3) selection against deleterious products of ectopic recombination among dispersed homologous TEs (''ectopic recombination model'') (Montgomery, Charlesworth, and Langley 1987).

Which model or models can account for the patterns that we see in these four non-LTR families? First of all, the TE-product expression model is not applicable in this case. Although it may in fact be very important for other types of TEs, it cannot explain strong selection acting against the *Doc* and *Jockey* elements because most of them are 5′ truncated, promoter-less, and thus likely untranscribed and untranslated copies (12 out of 15 *Doc* and 32 out of 37 *Jockey* elements). Excluding the full-length elements does not affect any of the results (data not shown).

Rejecting the ''Gene-Disruption'' Model

In contrast, the model of selection against gene disruption should apply to these non-LTR families to the same extent as it would to any other dispersed TE family. Non-LTR families in general, and *Jockey* and *Doc* elements in particular, are known to induce visible mutations (Driver et al. 1989; O'Hare et al. 1991; White and Jacobson 1996), and so we know that they can disrupt genes. Can the ''gene-disruption'' model explain our results?

The data for the *Jockey* and *Doc* families are consistent with this model, assuming that there are no or extremely few truly neutral sites in the Drosophila euchromatin into which these elements can insert (Charlesworth 1991). This follows from the fact that all *Jockey* and *Doc* copies are rare (except a single element apparently affected by positive selection), and thus they must all affect neighboring genes in a subtle but decidedly ($N_e s \ll -1$) deleterious manner. However, the observation that some non-LTR families such as *BS* and *X* have many frequent elements and overall are distributed in a neutral or (nearly neutral) manner undermines this interpretation. Indeed, under the gene-disruption model we have to infer that there are plenty of truly neutral sites into which *BS* and *X* elements can insert and also that these elements avoid all of the deleterious sites into which *Jockey* and *Doc* elements insert with inevitability.

To explain these contrasting patterns across different non-LTR families under the gene-disruption model, we need to find a reason for why individual *Doc* and *Jockey* elements are significantly more deleterious than *BS* and *X* elements. On the face of it, this deleterious effect doesn't appear very likely—the four families are very similar in sequence organization as they all belong to the same *Jockey*-family of non-LTR elements (Malik, Burke, and Eickbush 1999). They are also similar in the mode of transposition and have similar lengths of functional elements. Nevertheless, it is possible that *Doc* and *Jockey* have a different and particularly deleterious insertion site preference, for instance exclusively inside or very near genes, while *BS* and *X* elements have a different, non-deleterious insertion site preference, for instance always far away from genes. We fail, however, to find any evidence that this is the case. *Jockey* and *Doc* copies in our sample are not on average closer to genes than *BS* and *X* elements (*t*-test, $P = 0.29$, 1-tailed test, 65 df; Mann-Whitney test, $P = 0.66$) (fig. 2). Similarly there is no correlation between the population frequency of individual TEs and their distance to genes (Kendall's $\tau = -0.1$, $P = 0.22$).

It is also possible, however, that the copies of *Jockey* and *Doc* elements are more deleterious even when they land at the same distance from genes as *BS* and *X* elements. *Jockey* and *Doc* elements could have specific regulatory sequences interfering with gene regulation even at long distances. For example, it is possible that they possess yet undiscovered active sequences similar to the *Su*(*Hw*) binding sites, such as that present in *gypsy* elements (Harrison et al. 1989; Smith and Corces 1992), which can disrupt promoter–enhancer interactions of genes (Geyer, Spana, and Corces 1986; Harrison et al. 1989; Scott, Taubman, and Geyer 1999). Because such sequences are generally present at the untranslated 5′ end of TEs, longer TE copies would carry such sequences more often. Thus they might be very deleterious, even at very long distances, by disrupting long-range promoter–enhancer interactions. Longer TE copies might also be more deleterious because of their bulk. Compatible with this interpretation, *Jockey* and *Doc* elements are on average much longer than *BS* and *X* elements (fig. 3; comparison of the polymorphic elements, Mann-Whitney test, $P = 0.01$; comparison of all copies, Mann-Whitney test, $P = 0.001$).

However, there is a way we can test this scenario. In our data, there are many short *Jockey* elements that are as short as the *BS* elements. The comparison of their population frequencies reveals that short *Jockey* (<800 bp) elements are substantially less frequent than the *BS* elements (also all <800 bp) (Mann-Whitney test, $P < 0.0001$). Note that within the elements shorter than 800

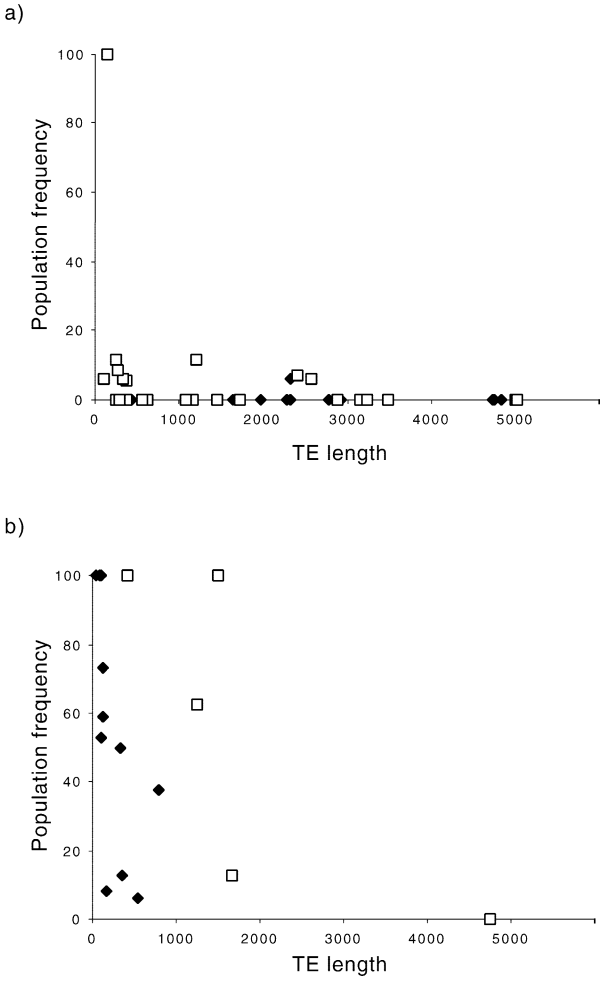

FIG. 2.—The relationship between element length and population frequency. a) *BS* (black squares) and × (open squares); b) *Doc* (black squares) and *Jockey* (open squares). The length of the fixed elements (100% frequency) is our best estimate of its length at the time of transposition. The length of a polymorphic element is its current length. The putatively adaptive insertion of a *Doc* element is excluded.

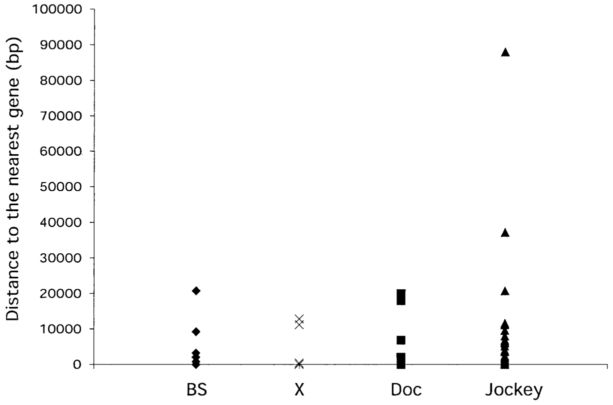

FIG. 3.—Distance to the nearest gene for each TE copy. The putatively adaptive insertion of a *Doc* element is excluded.

bp, *Jockey* and *BS* elements have indistinguishable length distributions (*t*-test, 25 df, $P = 0.58$; Mann-Whitney test, $P = 0.22$). This suggests that short *Jockeys* are still more strongly deleterious than *BS* elements, despite their similarly short size.

To explain this based on the gene-disruption model, short *Jockey* elements containing a 3′ portion of the reverse transcriptase coding sequence, the poly-A signal, and the poly-A tail must be inevitably deleterious even at large distances from genes, whereas the short *BS* and *X* element containing very similar sequences must generally be entirely neutral. Although we cannot formally eliminate this possibility, we consider it highly unlikely. We thus provisionally reject the gene-disruption model for these four TE families.

The Ectopic-Recombination Model Is Consistent with the Data

The final described model is the ''ectopic recombination model,'' whereby selection acts against deleterious effects of recombination among dispersed homologous TE copies. Can this model explain our data? Because TE copies of any particular family can recombine only with other copies from the same family, selection acting on any one family of TEs is entirely independent of selection on any other family. At the same time, copies within a given family should be subject to correlated levels of selection. In this way this model predicts that variation in the strength of selection can vary systematically among different TE families. Moreover, in selection–transposition balance, the families that transpose less frequently should equilibrate under lower copy numbers and should be subject to lower strength of selection. The frequency of ectopic recombination should be a monotonically increasing function of the copy number (Montgomery, Charlesworth, and Langley 1987) and the length of polymorphic elements (Dray and Gloor 1997). Thus selection strength could easily vary across families, with the families containing fewer and/or shorter polymorphic copies predicted to be under weaker selection.

Our results appear consistent with all of these predictions. The strength of selection varies strongly across the TE families, with *Jockey* and *Doc* families being under stronger selection than *BS* and *X* elements. With only four data points, we do not have the power to test whether *BS* and *X* have significantly fewer copy numbers than *Jockey* and *Doc*, but there appears to be a tendency in this direction (table 2). We do have enough power, however, to test whether polymorphic *Jockey* and *Doc* elements are on average much longer than polymorphic *BS* and *X* elements, and, as mentioned, indeed they are (fig. 3). The average and median lengths of polymorphic *BS* and *X* elements (929 ± 411 bp and 356 bp, respectively) are much shorter than the average and median lengths of polymorphic *Doc* and *Jockey* elements (2,000 ± 249 bp and 1,538 bp, respectively). These differences are significant (Mann-Whitney test, $P = 0.009$).

Selection Discriminates Among TEs Based on Length Within Families

The ectopic recombination model explains well why selection acts family by family and why selection is stronger in the families with more numerous and longer copies. It makes additional predictions, however. In particular the length of TEs should matter not only among families but also within them. The longer TEs should be more deleterious than shorter TEs within families and should be present at lower population frequencies on average.

To test this prediction we assessed whether the length of TEs within a given family correlates negatively with the population frequency. Because there was only a single *Doc* element in a non-zero frequency within the sampled strains (table 1), we excluded *Doc* elements from this analysis. In all other cases we did find negative correlation between the length of TEs within a family and the population frequency (*X*, Kendall's $\tau = -0.7$, 1-tailed $P = 0.035$; *BS*, Kendall's $\tau = -0.67$, 1-tailed $P = 0.002$; *Jockey*, Kendall's $\tau = -0.28$, 1-tailed $P = 0.008$). Note that we are using the current length of polymorphic TEs and our best estimate of the length at insertion for fixed TEs. This is because we are attempting to understand the parameters of the population process of frequency change prior to fixation, and most secondary deletions are likely to have happened after fixation. This analysis is conservative for our purposes.

If we limit the analysis to the polymorphic elements only, we still detect negative correlation between the current length and the population frequency for the *Jockey* elements (*Jockey*, Kendall's $\tau = -0.22$, 1-tailed $P = 0.03$) and for the *BS* and *X* elements, albeit at a marginally significant level, (*X*, Kendall's $\tau = -1$, 1-tailed $P = 0.06$; *BS*, Kendall's $\tau = -0.43$, 1-tailed $P = 0.07$). As expected, the fixed elements were significantly shorter at the time of integration than the polymorphic elements are at present time within *Jockey* (Wilcoxon rank test, 1-tailed $P = 0.045$) and *BS* families (Wilcoxon rank test, 1-tailed $P = 0.007$). The fixed *X* elements are on average shorter than the polymorphic ones, but this difference is not significant (Wilcoxon rank test, 1-tailed $P = 0.13$). However the small number of *X* elements (5) makes meaningful comparisons difficult.

Interestingly, if instead of the current length of the polymorphic TEs, we look at the inferred length of polymorphic TEs at the time of their insertion, we no longer find a negative correlation between the length of *Jockey* elements and their population frequency (Kendall's $\tau = -0.063$, $P = 0.58$). Closer inspection shows that out of 9 polymorphic *Jockey* elements at non-zero frequency in the studied strains, three were full-length elements at the time of insertion that subsequently suffered large secondary deletions (2,507 bp, 2,587 bp, and 4,891 bp). No *Jockey* element at zero frequency suffered such large deletions. The population frequency of these three *Jockey* elements is substantially higher than the rest of the *Jockey* elements (Mann-Whitney test, $P = 0.005$)

Two explanations for this pattern are possible. One is that the *Jockey* elements present at non-zero frequencies are somewhat older on average and therefore had more time to suffer deletions than the elements present at zero frequencies. However, the analysis of point substitutions does not lend much support to this proposition. *Jockey* elements found in zero frequency and those found in non-zero frequency are not significantly different in their divergence from the consensus sequence measured in the number of nucleotide differences per nucleotide (G-test, 1 df, $P = 0.3$). The other possible explanation is that full-length elements cannot reach non-zero frequencies observable in our sample (>5%), unless they become substantially shorter through secondary deletions. In this way secondary deletions may lower the strength of purifying selection acting on the long elements. Further supporting this interpretation is the finding that the rate of large (>400 bp) deletions relative to the rate of nucleotide substitution is substantially higher among non-zero frequency *Jockey* elements than among the zero frequency ones (Fisher's exact test, 1 df, $P = 0.018$).

Ectopic Recombination as an Example of Selection Based on Homologous Interactions

The overall result of this study is to suggest that, at least in these four families of TEs, selection does not operate on the individual effects of TEs on the neighboring genes, but rather operates at the level of families of homologous TEs. Moreover, selection gets stronger with the increase of the copy number and length of individual TE copies. All of these features are consistent with selection acting against the products of ectopic recombination. However, they are also consistent with selection based on other homology-dependent interactions. The presence of homologous DNA and RNA sequences in the cell leads to a multitude of profound phenotypic effects affecting chromatin state and levels of gene expression (Henikoff and Dreesen 1989; Fanti et al. 1998; Ketting et al. 1999; Pal-Bhadra, Bhadra, and Birchler 1999, 2002; Sass and Henikoff 1999; Wu and Morris 1999; Aravin et al. 2001). Thus it is entirely possible that selection against many dispersed, long, homologous TE copies is mediated not (or not exclusively) by ectopic recombination, but through some other homology-dependent, epigenetic effect. Our analysis does not distinguish among these possibilities.

Transposition-Selection Balance with Variable Transposition Rates Between Families or Within Families over Time

The variation in the strength of selection among TE families in our study most likely reflects variation in the family-specific rate of transposition. Indeed, in transposition–selection balance the copy number equilibrates at a level where the rate of TE elimination by selection matches the rate of transposition. The stronger level of selection acting against TEs within a family thus implies a higher rate of transposition of TEs within that family. Thus the most straightforward way to interpret our data is to postulate a higher rate of transposition for *Jockey* and *Doc* elements than for *BS* and *X* elements.

It is possible that *Jockey* and *Doc* are just more active TEs than *BS* and *X*. However, it is also possible that transposition rate within a family varies through time and we have simply caught *Jockey* and *Doc* during their active phase, whereas we have caught *BS* and *X* during their slow phase. Under this hypothetical scenario, *BS* and *X* used to transpose at high rates (similar to those currently observed for *Jockey* and *Doc*), were present in high copy numbers, and were under strong selection based on ectopic recombination (or some other homology-dependent selective mechanism). We know that transposition rate can itself evolve, and the presence of polymorphic modifiers of transposition in Drosophila populations has been documented (Nuzhdin et al. 1998). It is possible that fixation of repressors of transposition for *BS* and *X* families led to a sharp reduction of transposition rates, leading to a reduction of the copy number of the *BS* and *X* elements as a result of drift and purifying selection. Eventually, the copy number became sufficiently low, and the strength of selection sufficiently weak, to allow the shorter elements to drift to higher frequencies. In this scenario the rate of transposition must have remained low for a long enough time (on the order of $4N_e$ generations) that many short elements were able to reach high population frequencies, and some even had enough time to reach fixation.

Some features of our data hint that this second model may be a reasonable possibility. For instance, if *BS* elements have always been transposing at low rates with many copies reaching fixation, we should find many fixed copies of different ages. In fact, we do find three fixed copies, but all of them are fairly young (inserted <3 MYA based on the level of divergence from the consensus sequence). One explanation is that the older copies have all been lost or have become unrecognizable through frequent deletions (Petrov, Lozovskaya, and Hartl 1996; Pritchard and Schaeffer 1997; Petrov et al. 1998; Petrov and Hartl 1998; Ramos-Onsins and Aguade 1998; Robin et al. 2000; Blumenstiel, Hartl, and Lozovsky 2002; Petrov 2002). It is also possible, however, that *BS* elements were transposing fast and thus were not fixing through drift prior to 3 MYA. We also found a single fixed *Jockey* element that was fixed approximately 10 MYA (based on its 21.4% divergence from the consensus sequence). If the rate of transposition and the strength of selection against *Jockey* elements were both as high 10 MYA as they are today, we need to postulate that this *Jockey* element was swept to fixation by positive selection based on its local effect. Without positive selection, probability of fixation of new mutations with $N_e s$ of -26 (as estimated for *Jockey*) is astronomically small ($\sim 7 \times 10^{-49}$). Its small size (152 bp) is simply a coincidence under the scenario of positive selection, but it would be naturally predicted under the scenario in which it drifted to fixation during a period of low transposition rates of *Jockey* elements.

These considerations are clearly insufficient to distinguish between the model where transposition rate varies mostly among TE families and stays relatively constant within a family, and the model where transposition rate varies sharply for a particular family through time, with long periods ($\geq 4\ N_e$ generations) of high and low transposition rates. However, future studies could resolve this question. In particular studies of the same families in multiple Drosophila species can establish whether the same families (such as *Jockey* and *Doc*) are always present in high copy numbers and low population frequencies and whether the reverse is true for other families (such as *BS* and *X*). Studies of age distribution of fixed TEs for more TE families may also shed light on this issue.

Selection at the Level of Ectopic Recombination and Genome Evolution

If the neutral attainment of intermediate population frequency and even fixation is a consistent feature of some TE families or a periodic feature of many TE families, why haven't we seen more fixed and high-frequency TEs before? The probable explanation is that mostly short elements reach fixation. In view of the recent demonstration that a high rate of DNA loss through small (<400 bp) deletions affects most or all sequences in the Drosophila genome (Petrov, Lozovskaya, and Hartl 1995; Pritchard and Schaeffer 1997; Petrov et al. 1998; Petrov and Hartl 1998; Ramos-Onsins and Aguade 1998; Robin et al. 2000; Blumenstiel, Hartl, and Lozovsky 2002; Petrov 2002), it seems likely that fixed elements have a relatively short persistence time.

However, even the high-frequency polymorphic or recently fixed short TEs that have not yet been deleted, may have been overlooked in the past. Most surveys of TE frequencies have been conducted either by conducting in situ hybridization with polytene chromosomes or by surveying particular genomic regions. In situ hybridization is quite inefficient when dealing with very short regions of homology, whereas population surveys of particular genomic regions bias the analysis in favor of high-copy polymorphic TEs, as these are more likely to be captured segregating in a predetermined region. As the families containing high-frequency polymorphic copies are likely to be present in low copy numbers, they are going to be underrepresented in population samples based on predefined chromosomal regions.

Despite the relative paucity of detectable fixed TEs, fixation of TEs may be of great importance in Drosophila genome size evolution. The observation of TE fixation, and the possibility that it occurs from time to time for a large number TE families, may provide the counterbalancing force to persistent DNA loss through frequent small deletions (Petrov, Lozovskaya, and Hartl 1996; Pritchard and Schaeffer 1997; Petrov et al. 1998; Petrov and Hartl 1998; Ramos-Onsins and Aguade 1998; Robin et al. 2000; Blumenstiel, Hartl, and Lozovsky 2002; Petrov 2002).

There also may be implications of these findings for the evolution of the transposable elements themselves. If the principal deleterious effect of TEs is due to ectopic recombination, multiple TEs should be able to coexist without burdening the host, provided they do not recombine with each other. The risk of ectopic recombination should therefore impose a strong selective pressure for rapid sequence divergence of TEs. This may be one reason for the evolution of such a large number of TE families in the Drosophila lineage (Charlesworth and Langley 1989).

Acknowledgments

We thank Sergey Nuzhdin, Charles Aquadro, and Casey Bergman for the gifts of the *D. melanogaster* strains and helpful discussions. We are also grateful to Marc Feldman, Virginia Walbot, Krista Ingram, Emile Zuckerkandl, Mark Siegal, and Justin Fay for extremely helpful comments on the manuscript. The comments by the editor and the three anonymous referees improved the manuscript substantially. The work was supported by a Stanford University Office of Technology Licensing award to D.A.P; A.E.H. was supported by National Institutes of Health grants GM 28016 and GM 28428 to Marcus W. Feldman; and J.C.D. was supported in part by a National Science Foundation graduate fellowship.

Literature Cited

Adams, M. D., S. E. Celniker, R. A. Holt, et al. (194 co-authors). 2000. The genome sequence of *Drosophila melanogaster*. Science **287**:2185–2195.

Aravin, A. A., N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, and V. A. Gvozdev. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. Curr. Biol. **11**:1017–1027.

Bartolome, C., X. Maside, and B. Charlesworth. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol. Biol. Evol. **19**: 926–937.

Becker, P. B. 1995. Drosophila chromatin and transcription. Semin. Cell Biol. **6**:185–190.

Bender, W., P. Spierer, and D. S. Hogness. 1983. Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the Bithorax complex in *Drosophila melanogaster*. J. Mol. Biol. **168**:17–33.

Berezikov, E., A. Bucheton, and I. Busseau. 2000. A search for reverse transcriptase-coding sequences reveals new non-LTR retrotransposons in the genome of *Drosophila melanogaster*. Genome Biol. **1**:research 0012.1–0012.15.

Berg, D. E., and M. M. Howe. 1989. Mobile DNA. American Society for Microbiology, Washington, D.C.

Berry, A. J., J. W. Ajioka, and M. Kreitman. 1991. Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics **129**:1111–1117.

Birchler, J. A., M. P. Bhadra, and U. Bhadra. 2000. Making noise about silence: repression of repeated genes in animals. Curr. Opin. Genet. Dev. **10**:211–216.

Blumenstiel, J. P., D. L. Hartl, and E. R. Lozovsky. 2002. Patterns of insertion and deletion in contrasting chromatin domains. Mol. Biol. Evol **19**:2211–2225.

Charlesworth, B. 1991. Transposable elements in natural populations with a mixture of selected and neutral insertion sites. Genet. Res. **57**:127–134.

———. 1996. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. **68**:131–149.

Charlesworth, B., and D. Charlesworth. 1983. The population dynamics of transposable elements. Genet. Res. **42**:1–27.

Charlesworth, B., and C. H. Langley. 1989. The population genetics of Drosophila transposable elements. Annu. Rev. Genet. **23**:251–287.

Charlesworth, B., A. Lapid, and D. Canada. 1992*a*. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. Genet. Res. **60**:103–114.

———. 1992*b*. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. II. Inferences on the nature of selection against elements. Genet. Res. **60**:115–130.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics **134**:1289–1303.

Charlesworth, B., P. Sniegowski, and W. Stephan. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371**:215–220.

Craig, N. L., R. Craigie, M. Gellert, and A. M. Lambowitz. 2002. Mobile DNA II. American Society of Microbiology, Washington D.C., p. 1204.

Dray, T., and G. B. Gloor. 1997. Homology requirements for targeting heterologous sequences during P-induced gap repair in *Drosophila melanogaster*. Genetics **147**:689–699.

Driver, A., S. F. Lacey, T. E. Cullingford, A. Mitchelson, and K. O'Hare. 1989. Structural analysis of *Doc* transposable elements associated with mutations at the white and suppressor of forked loci of *Drosophila melanogaster*. Mol. Gen. Genet. **220**:49–52.

Ewens, W. H. 1979. Mathematical population genetics. Springer-Verlag, New York.

Fanti, L., D. R. Dorer, M. Berloco, S. Henikoff, and S. Pimpinelli. 1998. Heterochromatin protein 1 binds transgene arrays. Chromosoma **107**:286–292.

Finnegan, D. J. 1992. Transposable elements. Curr. Opin. Genet. Dev. **2**:861–867.

Finnegan, D. J., and D. H. Fawcett. 1986. Transposable elements in *Drosophila melanogaster*. Pp. 1–62 *in* N. MacLean, ed. Oxford Surveys of Eukaryotic Genes. Oxford University Press, Oxford.

Geyer, P. K., C. Spana, and V. G. Corces. 1986. On the molecular mechanism of gypsy-induced mutations at the yellow locus of *Drosophila melanogaster*. EMBO J. **5**:2657–2662.

Goldman, A. S., and M. Lichten. 1996. The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. Genetics **144**:43–55.

Goldman, A. S., and M. Lichten. 2000. Restriction of ectopic recombination by interhomolog interactions during *Saccharomyces cerevisiae* meiosis. Proc. Natl. Acad. Sci. USA **97**: 9537–9542.

Harrison, D. A., P. K. Geyer, C. Spana, and V. G. Corces. 1989. The gypsy retrotransposon of *Drosophila melanogaster*: mechanisms of mutagenesis and interaction with the suppressor of hairy-wing locus. Dev. Genet. **10**:239–248.

Henikoff, S. 1995. Gene silencing in Drosophila. Curr. Top. Microbiol. Immunol. **197**:193–208.

Henikoff, S., and T. D. Dreesen. 1989. Trans-inactivation of the Drosophila brown gene: evidence for transcriptional repression and somatic pairing dependence. Proc. Natl. Acad. Sci. USA **86**:6704–6708.

Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet. Res. **8**:269–294.

Hoogland, C., and C. Biemont. 1996. Chromosomal distribution of transposable elements in *Drosophila melanogaster*: test of the ectopic recombination model for maintenance of insertion site number. Genetics **144**:197–204.

Hudson, R. R., and N. L. Kaplan. 1995. Deleterious background selection with recombination. Genetics **141**:1605–1617.

Kaminker, J. S., C. M. Bergman, B. Kronmiller et al. (12 co-authors). 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol. **3**:research0084-4.

Kaplan, N. L., and J. F. Brookfield. 1983. Transposable element in mendelian populations. III. Statistical results. Genetics **104**:485–495.

Ketting, R. F., T. H. Haverkamp, H. G. van Luenen, and R. H. Plasterk. 1999. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell **99**:133–141.

Kidwell, M. G. 1989. Regulatory aspects of the expression of P-M hybrid dysgenesis in Drosophila. Pp. 183–194 *in* M. E. Lambert, J. F. McDonald, I. B. Weinstein, eds. Transposable elements as mutagenic agents. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

———. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica **115**:49–63.

Kreitman, M. 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304**:412–417.

Langley, C. H., J. F. Brookfield, and N. L. Kaplan. 1983. Transposable elements in mendelian populations. I. A theory. Genetics **104**:457–471.

Langley, C. H., E. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth. 1988. On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. **52**:223–235.

Laski, F. A., D. C. Rio, and G. M. Rubin. 1986. Tissue specificity of Drosophila *P* element transposition is regulated at the level of mRNA splicing. Cell **44**:7–19.

Lohe, A. R., and D. L. Hartl. 1996. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol. Biol. Evol. **13**:549–555.

Lozovskaya, E. R., D. L. Hartl, and D. A. Petrov. 1995. Genomic regulation of transposable elements in Drosophila. Curr. Opin. Genet. Dev. **5**:768–773.

Lu, B. Y., J. Ma, and J. C. Eissenberg. 1998. Developmental regulation of heterochromatin-mediated gene silencing in Drosophila. Development **125**:2223–2234.

Luan, D. D., M. H. Korman, J. L. Jacubczak, and T. H. Eickbush. 1993. Reverse transcriptase of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell **72**:595–605.

Malik, H. S., W. D. Burke, and T. H. Eickbush. 1999. The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. **16**:793–805.

McDonald, J. F., L. V. Matyunina, S. Wilson, I. K. Jordan, N. J. Bowen, and W. J. Miller. 1997. LTR retrotransposons and the evolution of eukaryotic enhancers. Genetica **100**: 3–13.

Mizrokhi, L. I., L. A. Obolenkova, A. F. Priimagi, Y. V. Ilyin, T. I. Gerasimova, and G. P. Georgiev. 1985. The nature of unstable insertion mutations and reversions in the locus cut of *Drosophila melanogaster*: molecular mechanism of transposition. EMBO J. **4**:3781–3787.

Montgomery, E., B. Charlesworth, and C. H. Langley. 1987. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. Genet. Res. **49**:31–41.

Montgomery, E. A., S. M. Huang, C. H. Langley, and B. H. Judd. 1991. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. Genetics **129**:1085–1098.

Moriyama, E. N., and J. R. Powell. 1997. Synonymous substitution rates in Drosophila: mitochondrial versus nuclear genes. J. Mol. Evol. **45**:378–391.

Nagylaki, T. 1974. The moments of stochastic integrals and the distribution of sojourn times. Proc. Natl. Acad. Sci. USA **71**:746–749.

Nuzhdin, S. V. 1999. Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica **107**:129–137.

Nuzhdin, S. V., E. G. Pasyukova, E. A. Morozova, and A. J. Flavell. 1998. Quantitative genetic analysis of *copia* retrotransposon activity in inbred *Drosophila melanogaster* lines. Genetics **150**:755–766.

O'Hare, K., M. R. Alley, T. E. Cullingford, A. Driver, and M. J. Sanderson. 1991. DNA sequence of the *Doc* retroposon in the white-one mutant of *Drosophila melanogaster* and of secondary insertions in the phenotypically altered derivatives white-honey and white-eosin. Mol. Gen. Genet. **225**: 17–24.

Pal-Bhadra, M., U. Bhadra, and J. A. Birchler. 1999. Cosuppression of nonhomologous transgenes in Drosophila involves mutually related endogenous sequences. Cell **99**:35–46.

———. 2002. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol. Cell **9**:315–327.

Petrov, D. A. 2002. DNA loss and evolution of genome size in Drosophila. Genetica **115**:81–91.

Petrov, D. A., Y.-C. Chao, E. C. Stephenson, and D. L. Hartl. 1998. Pseudogene evolution in Drosophila suggests a high rate of DNA loss. Mol. Biol. Evol. **15**:1562–1567.

Petrov, D. A., and D. L. Hartl. 1998. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol. Biol. Evol. **15**:293–302.

Petrov, D. A., E. R. Lozovskaya, and D. L. Hartl. 1996. High intrinsic rate of DNA loss in Drosophila [see comments]. Nature **384**:346–349.

Petrov D. A., J. L. Schutzman, D. L. Hartl, and E. R. Lozovskaya. 1995. Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. Proc. Natl. Acad. Sci. USA **92**:8050–8054.

Pritchard, J. K., and S. W. Schaeffer. 1997. Polymorphism and divergence at a Drosophila pseudogene locus. Genetics **147**: 199–208.

Ramos-Onsins, S., and M. Aguade. 1998. Molecular evolution of the Cecropin multigene family in Drosophila. functional genes vs. pseudogenes. Genetics **150**:157–171.

Robert, V., N. Prud'homme, A. Kim, A. Bucheton, and A. Pelisson. 2001. Characterization of the flamenco region of the *Drosophila melanogaster* genome. Genetics **158**:701–713.

Robin, G. C., R. J. Russell, D. J. Cutler, and J. G. Oakeshott. 2000. The evolution of an α-esterase pseudogene inactivated in the *Drosophila melanogaster* lineage. Mol. Biol. Evol. **17**:563–575.

Sass, G. L., and S. Henikoff. 1999. Pairing-dependent mislocalization of a Drosophila brown gene reporter to a heterochromatic environment. Genetics **152**:595–604.

Schug, M. D., C. M. Hutter, K. A. Wetterstrand, M. S. Gaudette, T. F. Mackay, and C. F. Aquadro. 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Mol. Biol. Evol. **15**:1751–1760.

Scott, K. C., A. D. Taubman, and P. K. Geyer. 1999. Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength. Genetics **153**:787–798.

Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. **23**:23–35.

Smith, P. A., and V. G. Corces. 1992. The suppressor of Hairy-wing binding region is required for gypsy mutagenesis. Mol. Gen. Genet. **233**:65–70.

Tudor, M., A. J. Davis, M. Feldman, M. Grammatikaki, and K. O'Hare. 2001. The *X* element, a novel LINE transposable element from *Drosophila melanogaster*. Mol. Genet. Genomics **265**:489–496.

Udomkit, A., S. Forbes, G. Dalgleish, and D. J. Finnegan. 1995. *BS*: a novel LINE-like element in *Drosophila melanogaster*. Nucleic Acids Res. **23**:1354–1358.

White, L. D., and J. W. Jacobson. 1996. Insertion of the retroposable element, *Jockey*, near the Adh gene of *Drosophila melanogaster* is associated with altered gene expression. Genet. Res. **68**:203–209.

Wu, C. T., and J. R. Morris. 1999. Transvection and other homology effects. Curr. Opin. Genet. Dev. **9**:237–246.

David Rand, Associate Editor

Accepted January 31, 2003