Research

The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective

Joshua S Kaminker*†, Casey M Bergman†‡, Brent Kronmiller‡§, Joseph Carlson‡, Robert Svirskas¶, Sandeep Patel‡, Erwin Frise‡, David A Wheeler¥, Suzanna E Lewis*, Gerald M Rubin*‡#, Michael Ashburner** and Susan E Celniker‡

Addresses: *Department of Molecular and Cellular Biology, University of California, Berkeley, CA 94720, USA. ‡*Drosophila* Genome Project, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. ¶Amersham Biosciences, 2100 East Elliot Rd, Tempe, AZ 85284, USA. #Howard Hughes Medical Institute, ¥Human Genome Sequencing Center and Department of Molecular and Cell Biology, Baylor College of Medicine, Houston, TX 77030, USA. **Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK. §Current address: Department of Bioinformatics and Computational Biology, Iowa State University, Ames, IA 50011, USA. †These authors contributed equally to this work.

Correspondence: Michael Ashburner. E-mail: ma11@gen.cam.ac.uk

Published: 23 December 2002

Genome ***Biology*** 2002, **3(12)**:research0084.1–0084.20

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/12/research/0084

(Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 7 October 2002
Revised: 11 November 2002
Accepted: 25 November 2002

Abstract

Background: Transposable elements are found in the genomes of nearly all eukaryotes. The recent completion of the Release 3 euchromatic genomic sequence of *Drosophila melanogaster* by the Berkeley *Drosophila* Genome Project has provided precise sequence for the repetitive elements in the *Drosophila* euchromatin. We have used this genomic sequence to describe the euchromatic transposable elements in the sequenced strain of this species.

Results: We identified 85 known and eight novel families of transposable element varying in copy number from one to 146. A total of 1,572 full and partial transposable elements were identified, comprising 3.86% of the sequence. More than two-thirds of the transposable elements are partial. The density of transposable elements increases an average of 4.7 times in the centromere-proximal regions of each of the major chromosome arms. We found that transposable elements are preferentially found outside genes; only 436 of 1,572 transposable elements are contained within the 61.4 Mb of sequence that is annotated as being transcribed. A large proportion of transposable elements is found nested within other elements of the same or different classes. Lastly, an analysis of structural variation from different families reveals distinct patterns of deletion for elements belonging to different classes.

Conclusions: This analysis represents an initial characterization of the transposable elements in the Release 3 euchromatic genomic sequence of *D. melanogaster* for which comparison to the transposable elements of other organisms can begin to be made. These data have been made available on the Berkeley *Drosophila* Genome Project website for future analyses.

Background

Transposable element sequences are abundant yet poorly understood components of almost all eukaryotic genomes [1]. As a result, many biologists have an interest in the description of transposable elements in completely sequenced eukaryotic genomes. The evolutionary biologist wants to understand the origin of transposable elements, how they are lost and gained by a species and the role they play in the processes of genome evolution; the population geneticist wants to know the factors that determine the frequency and distribution of elements within and between populations; the developmental geneticist wants to know what roles these elements may play in either normal developmental processes or in the response of the organism to external conditions; finally, the molecular geneticist wants to know the mechanisms that regulate the transposition cycle of these elements and how they interact with the cellular machinery of the host. It is for all of these reasons and more that a description of the transposable elements in the recently completed Release 3 genomic sequence of *D. melanogaster* is desirable.

Our understanding of transposable elements owes much to research on *Drosophila*. Over 75 years ago, Milislav Demerec discovered highly mutable alleles of two genes in *D. virilis*, *miniature* and *magenta* ([2-4], reviewed in [5]). Both genes were mutable in soma and germline and, for the *miniature-3α* alleles, dominant enhancers of mutability were also isolated by Demerec. In retrospect, it seems clear that the mutability of these alleles was the result of transposition of mobile elements. The dominant enhancers may have been particularly active elements or mutations in host genes that affect transposability (see below). There matters stood until McClintock's analysis of the *Ac* and *Ds* factors in maize, which led to the discovery of transposition [6] and the discovery of insertion elements in the *gal* operon of *Escherichia coli* (see [7]).

Green [8] synthesized the available evidence to make a strong case for insertion as a mechanism of mutagenesis in *Drosophila*. Concurrently, Hogness' group had begun a molecular characterization of two elements in *D. melanogaster*, *412* and *copia* [9,10] and provided evidence that they were transposable [11-13]. Glover [14] unknowingly characterized the first eukaryotic transposable element at the molecular level, the insertion sequences of 28S rRNA genes. The discovery of male recombination [15], and two systems of hybrid dysgenesis in *D. melanogaster* (see [16,17]) bridged the gap between genetic and molecular analyses. The discovery of the transposable elements that cause hybrid dysgenesis, the *P* element [18] and the *I* element [19], led to the first genomic analyses of transposable elements in a eukaryote.

The publication of the Release 1 genomic sequence in March 2000 [20] and the Release 2 genomic sequence in October 2000 encouraged several studies on the genomic distribution and abundance of transposable elements in *D. melanogaster* [21-25]. Unfortunately, neither release was suitable for rigorous analysis of its transposable elements. In the whole-genome shotgun assembly process, repetitive sequences (including transposable elements) were masked by the SCREENER algorithm and remained as gaps between unitigs [26]. During the repeat-resolution phase of the whole-genome assembly, an attempt was made to fill these gaps. However, comparisons of small regions sequenced by the clone-by-clone approach versus the whole-genome shotgun method show that this process did not produce accurate sequences for transposable elements [26,27]. These results demonstrate that rigorous analyses of the transposable elements, or any other repetitive sequence, requires a sequence of higher quality, now publicly available as Release 3 [28]. For the first time, the nature, number and location of the transposable elements can reliably be analyzed in the euchromatin of *D. melanogaster*.

Results and discussion

Identification of known and novel transposable elements

Eukaryotic transposable elements are divided into those that transpose via an RNA intermediate, the retrotransposons (class I elements), and those that transpose by DNA excision and repair, the transposons (class II elements [1]). Within the retrotransposons, the major division is between those that possess long terminal repeats (LTR elements) and those that do not (LINE and SINE elements [29]). Among the transposons, the majority transpose via a DNA intermediate, encode their own transposase and are flanked by relatively short terminally inverted repeat structures (TIR elements). *Foldback* (*FB*) elements, which are characterized by their property of reannealing after denaturation with zero-order kinetics, are quite distinct from prototypical class I or II elements, and have been included in our analyses [30]. Other classes of repetitive elements, such as *DINE-1* [31-33], which are structurally distinct from all other classes, have not been included in this study.

We used a criterion of greater than 90% identity over more than 50 base-pairs (bp) of sequence to assign individual elements to families (see Materials and methods for details; a classification is shown in the additional data available with the online version of this paper (see Additional data files)). Subsequently, in order to ensure proper inclusion of elements in appropriate families, we generated multiple alignments for all families of transposable element represented by multiple copies. This allowed us to identify and remove spurious hits to highly repetitive regions of the genome, and it also enabled us to distinguish sequences of closely related families that share extensive regions of similarity.

A summary by class of the total number of complete and partial transposable elements in the Release 3 *Drosophila*

euchromatic sequence is presented in Table 1, and detailed results for individual families of transposable element are listed in Table 2. Including those described here, there are 96 known families of transposable elements in *D. melanogaster*: 49 LTR families, 27 LINE-like families, 19 TIR families and the FB family. We have identified 1,572 full or partial elements from 93 of these 96 families (Table 1). In total, 3.86% (4.5 Mb) of the Release 3 sequence is composed of transposable elements. Previous analysis of both the euchromatic and heterochromatic sequences has suggested that 9% of the *Drosophila* genome is composed of repetitive elements [34]. One reason for this difference may be that the proportion of transposable element sequences in heterochromatic regions is higher than the genomic average [22,35].

As shown in Table 1 and Figure 1, the different classes vary in their contribution to the *Drosophila* euchromatin both in amount of sequence and number of elements. LTR elements make up the largest proportion of the euchromatin (2.65%), more sequence than the sum of all other classes of element (LINE-like elements 0.87%, TIR elements 0.31%, and *FB* elements 0.04%). LTR elements are also the most numerous class of transposable element in the euchromatic sequences (682) followed by LINE-like (486), TIR (372), and *FB* (32) elements. The largest family representing each of the three major classes is *roo* (146 copies; LTR), *jockey* (69 copies; LINE-like), and *1360* (105 copies; TIR) (Table 2). The average size of all transposable elements in our study is 2.9 kilobases (kb), smaller than the 5.6 kb average length of middle repetitive DNA, estimated from reassociation kinetics [36].

Three of the 96 families are not described in this paper because they have not been found in the euchromatic portion of this sequence; these are the *P* element, *R2* and *ZAM*. It is not surprising that we did not find any *P* elements, as the sequenced strain was selected to be free of them. While we did not find *R2* and *ZAM* elements in the euchromatin, both of these elements were identified in unmapped scaffolds that derive from the heterochromatin [37]. The *R2* element has previously been found only within the 28S rDNA locus and in heterochromatin [38]. Strains of *D. melanogaster* are known to exist in which *ZAM* elements occur in low copy number in heterochromatic sequences [39]. The absence of the telomere-associated *HeT-A* and *TART* from the euchromatic portions of all chromosomes except chromosome 4 is not unexpected; the tandem arrays of these two elements are flanked by Taq microsatellite sequences [40,41] which are difficult to assemble and are under-represented in the current version of this sequence.

We discovered eight new families of transposable element within the Release 3 sequences. Two are members of the TIR class: *Bari2* (EMBL: AF541951) and *hopper2* (EMBL: AF541950). Six are members of the LTR class: *frogger* (EMBL: AF492763), *rover* (EMBL: AF492764), *cruiser* (a.k.a. *Quasimodo*) (EMBL: AF364550), *McClintock* (EMBL:AF541948), *qbert* (EMBL: AF541947), and *Stalker4* (EMBL: AF541949). We identified *Bari2* (four copies) by querying the *D. melanogaster* genome using a *Bari1*-like element isolated from *D. erecta* (EMBL: Y13853). The *Bari2* element shares 52% amino acid identity with the *Bari1* element; overall these elements share less than 50% nucleotide identity throughout their sequence. The *hopper2* (five copies) and *Stalker4* (two copies) families were identified by an analysis of the multiple alignment of the *hopper* and *Stalker* families, respectively. These alignments indicate distinct subfamilies on the basis of both nucleotide divergence and structural rearrangements over large regions of their alignment. The *hopper2* and *hopper* elements share 70% amino-acid identity throughout their predicted open reading frames (ORFs). However, outside their ORFs, these elements are quite divergent and do not share significant nucleotide identity (< 30%). The *Stalker4* element shares nearly 100% identity over the length of the predicted protein-coding domains, but these elements share only 50% nucleotide identity over the remainder of their sequence. A similar amount of conservation is seen between the previously defined [42] *Stalker* and *Stalker2* sequences. The *frogger* element (one partial copy) was identified on the basis of its LTRs and a predicted protein-coding ORF that is 73% similar at the amino-acid level to that of the *Dm88* family. The *rover* family (six copies) was identified in a BLAST search for repetitive elements in the genome; it is most closely related to the *17.6* element (71% amino acid identity). The *cruiser* family (14 copies) was identified during the finishing project by virtue of its LTRs, and is most closely related in sequence to the *Idefix* family, sharing 60% amino acid identity. The *qbert* family was identified by searching for regions of the genome that share similarity with protein-coding ORFs represented in our transposable element dataset. The *qbert* family (one copy) is most highly related to the *accord* family and shares regions of similarity that are 66% identical at the amino-acid level. The *McClintock* family (two copies), identified by its presence in a repeat region near the centromere of chromosome 4, is most closely related to the *17.6* family. *McClintock* shares 86% amino-acid identity with the protein-coding ORFs of *17.6*; elsewhere these elements are quite divergent, sharing less than 50% identity over the first 5,000 bp of the *McClintock* element.

We have also discovered several other sequences with high sequence similarity to the protein-coding regions of transposable elements, but they are not associated with repeats (see also [24]). These elements cannot easily be classified into particular families. Although we have not included them in this analysis, they have been included in the Release 3 annotation of the genome [43]. While some may be examples of functional host genes derived from transposable elements, such as are known in humans (see, for example [44]) and ciliates (see, for example [45,46]), others may reflect remnants of elements that have become functionally constrained in the host genome (see, for example [47]).

Table 1

An overview of the numbers of transposable elements in the euchromatic genome of *D. melanogaster*

Class	Arm	Total transposable element sequence (in bp)	% of arm	Total number of transposable elements	Number full length	% Full length	Number of transposable elements per Mb in genome	Number of transposable elements per Mb in proximal 2 Mb
All families	X	828,370	3.80	276	83	30.43	12.67	50
	2L	878,471	3.95	305	100	32.79	13.73	58.5
	2R	870,914	4.29	313	84	26.84	15.42	89
	3L	938,947	4.02	288	100	34.72	12.33	66.5
	3R	866,971	3.11	288	102	35.76	10.33	24.5
	4	127,874	10.33	102	9	8.82	82.40	-
	Total	4,511,547		1,572	478			
	Average		3.86			30.53	13.46	57.70
LTR	X	628,924	2.89	134	54	41.04	6.15	19.00
	2L	603,536	2.72	127	67	52.76	5.72	20.00
	2R	573,034	2.82	140	54	38.57	6.90	30.50
	3L	618,441	2.65	117	58	49.57	5.01	24.00
	3R	621,272	2.23	154	67	44.16	5.52	7.50
	4	44,121	3.56	10	4	40.00	8.08	-
	Total	3,089,328		682	304			
	Average		2.65			44.87	5.84	20.20
LINE-like	X	136,348	0.63	71	18	25.35	3.26	14.50
	2L	185,499	0.83	98	20	20.41	4.41	18.50
	2R	225,984	1.11	109	18	16.51	5.37	37.50
	3L	251,077	1.08	106	27	25.47	4.54	24.00
	3R	176,355	0.63	70	19	27.14	2.51	4.50
	4	37,399	3.02	32	1	3.12	25.85	-
	Total	1,012,662		486	103			
	Average		0.87			21.19	4.16	19.80
TIR	X	45,324	0.21	59	7	11.86	2.71	14.50
	2L	82,761	0.37	76	11	14.47	3.42	18.00
	2R	69,291	0.34	62	11	17.74	3.05	20.50
	3L	52,743	0.23	57	12	21.05	2.44	16.50
	3R	63,359	0.23	60	14	23.33	2.15	12.50
	4	44,195	3.57	58	3	5.17	46.85	-
	Total	357,673		372	58			
	Average		0.31			15.59	3.19	16.40
FB	X	17,774	0.08	12	4	33.33	0.55	2.00
	2L	6,675	0.03	4	2	50.00	0.18	2.00
	2R	2,605	0.01	2	1	50.00	0.1	0.50
	3L	16,686	0.07	8	3	37.50	0.34	2.00
	3R	5,985	0.02	4	2	50.00	0.14	0.00
	4	2,159	0.17	2	1	50.00	1.62	-
	Total	51,884		32	13			
	Average		0.04			40.62	0.27	1.30

For each class, the total numbers of each family of element, together with the numbers (and percentage of elements) that are full length is given for each chromosome arm. Column 3 gives the total base pairs contained within transposable elements, column 4 the percentage of each chromosome arm composed of transposable element sequences, column 8 the number of elements per Mb, and column 9 the numbers of elements within the most proximal 2 Mb of each of the five major chromosome arms. Differences in density and amount of transposable element sequence were tested by binning major chromosomal arms into 50-kb windows and testing significance by Mann-Whitney *U* tests. The only significant difference ($p < 0.05$) observed, either including or excluding the proximal 2 Mb, was an increase in density and amount of sequence on the X chromosome relative to 3R.

Table 2

The transposable elements of *D. melanogaster*

Class	Family	Canonical length	X	2L	2R	3L	3R	4	Total number	Number full length	Number partial	Number in proximal 2 Mb	Average pairwise distance
LTR	*17.6*	7439	2	0	3	5	2	0	12	7	5	4	0.006
	1731	4648	1	0	0	0	1	0	2	1	1	1	0.000
	297	6995	22	12	6	7	10	0	57	18	39	12	0.032
	3S18	6126	4	0	1	1	0	0	6	4	2	3	0.075
	412	7566	8	0	7	11	5	0	31	24	7	6	0.024
	accord	7404	0	0	1	0	0	0	1	0	1	0	-
	aurora	4263	0	0	2	1	0	0	3	1	2	3	0.074
	blastopia	5034	5	2	7	1	2	0	17	13	4	4	0.016
	blood	7410	1	11	2	3	5	0	22	22	0	6	0.001
	Burdock	6411	2	4	4	0	3	0	13	7	6	4	0.002
	Circe	6356	0	0	2	0	0	0	2	0	2	2	0.057
	copia	5143	4	13	4	5	4	0	30	26	4	3	0.002
	diver	6112	1	1	3	1	3	0	9	9	0	1	0.002
	diver2	4917	0	4	3	2	0	0	9	0	9	9	0.032
	Dm88	4558	0	0	2	0	30	0	32	0	32	2	0.015
	frogger	2483	0	1	0	0	0	0	1	1	0	1	-
	GATE	8507	1	0	16	0	0	3	20	0	20	17	0.077
	gtwin	7411	2	2	0	2	0	0	6	2	4	3	0.038
	gypsy	7469	0	1	1	0	0	0	2	1	1	1	0.000
	gypsy2	6841	1	0	0	2	0	0	3	1	2	2	0.067
	gypsy3	6973	0	0	2	0	0	0	2	1	1	2	0.038
	gypsy4	6852	0	1	0	0	1	0	2	1	1	2	0.041
	gypsy5	7369	1	0	0	1	0	0	2	1	1	0	0.005
	gypsy6	7826	0	1	0	0	0	0	1	0	1	1	-
	HMS-Beagle	7062	4	5	1	0	3	0	13	9	4	3	0.054
	Idefix	7411	1	2	0	3	1	0	7	2	5	5	0.022
	invader1	4032	0	0	4	3	18	1	26	1	25	7	0.023
	invader2	5124	1	4	3	2	0	0	10	3	7	10	0.053
	invader3	5484	2	5	2	2	5	0	16	3	13	5	0.044
	invader4	3105	0	4	2	1	1	1	9	2	7	2	0.068
	invader5	4038	1	4	0	1	0	0	6	0	6	6	0.068
	McClintock	6450	0	0	0	1	0	1	2	2	0	1	0.002
	mdg1	7480	5	2	9	6	3	0	25	13	12	2	0.012
	mdg3	5519	3	5	2	2	4	0	16	8	8	8	0.009
	micropia	5457	1	0	0	0	4	0	5	2	3	1	0.010
	opus	7521	3	6	6	6	3	0	24	16	8	5	0.003
	qbert	7650	0	0	1	0	0	0	1	1	0	1	-
	Quasimodo	7387	2	7	0	4	1	0	14	5	9	5	0.016
	roo	9092	35	22	31	31	27	0	146	58	88	22	0.012
	rooA	7621	1	0	0	1	1	2	5	0	5	3	0.045
	rover	7318	3	0	1	0	2	0	6	3	3	3	0.035
	springer	7546	2	1	4	1	3	0	11	5	6	5	0.061
	Stalker	7256	3	1	0	5	3	0	12	3	9	7	0.014
	Stalker2	8119	4	0	5	2	1	1	13	4	9	6	0.015
	Stalker4	7379	1	0	0	0	1	0	2	2	0	0	0.001
	Tabor	7345	1	2	0	0	0	0	3	2	1	0	0.001
	Tirant	8526	4	3	3	4	5	1	20	15	5	5	0.001
	Transpac	5249	2	1	0	0	2	0	5	5	0	0	0.000
	ZAM	8435	0	0	0	0	0	0	0	0	0	0	-

Table 2 *(continued from the previous page)*

The transposable elements of *D. melanogaster*

Class	Family	Canonical length	X	2L	2R	3L	3R	4	Total number	Number full length	Number partial	Number in proximal 2 Mb	Average pairwise distance
LINE-like	*baggins*	5453	1	2	10	0	0	1	14	0	14	12	0.076
	BS	5142	2	6	6	7	8	0	29	6	23	6	0.028
	Cr1a	4470	1	5	17	21	2	10	56	1	55	42	NC
	Doc	4725	5	16	5	19	10	0	55	30	25	7	0.006
	Doc2	4789	0	0	1	0	0	0	1	0	1	1	-
	Doc3	4740	0	1	6	1	0	1	9	0	9	8	0.065
	F	4708	2	7	10	10	11	2	42	16	26	11	0.019
	G	4346	1	2	0	0	0	0	3	0	3	1	0.059
	G2	3102	1	9	1	2	1	0	14	2	12	2	0.036
	G3	4605	1	1	2	0	0	0	4	0	4	3	0.029
	G4	3856	0	5	0	2	3	1	11	0	11	6	0.100
	G5	4856	0	5	0	1	2	1	9	0	9	6	0.063
	G6	2042	1	1	0	1	0	0	3	1	2	3	0.006
	Helena	1317	2	1	1	1	2	0	7	0	7	5	0.097
	HeT-A	6083	0	0	0	0	0	3	3	0	3	0	0.018
	I	5371	7	5	6	3	5	2	28	8	20	9	0.037
	Ivk	5402	3	3	0	1	0	0	7	2	5	4	0.070
	jockey	5020	16	9	15	13	14	2	69	12	57	4	0.003
	jockey2	3428	4	2	0	0	1	3	10	0	10	5	0.112
	Juan	4236	4	1	1	1	2	0	9	6	3	3	0.001
	R1	5356	2	1	2	2	0	3	10	2	8	7	0.049
	R2	3607	0	0	0	0	0	0	0	0	0	0	-
	Rt1a	5108	0	0	6	5	2	0	13	5	8	6	0.053
	Rt1b	5183	4	6	14	7	5	1	37	5	32	22	0.095
	Rt1c	5443	11	2	2	1	0	1	17	1	16	14	0.177
	TART	10654	0	0	0	0	0	1	1	0	1	0	-
	X	4740	3	8	4	8	2	0	25	6	19	12	0.049
TIR	*1360*	1177	10	18	16	20	11	30	105	10	95	48	0.087
	Bari1	1728	0	1	1	0	2	1	5	5	0	0	0.002
	Bari2	1064	0	2	0	0	1	1	4	1	3	3	0.043
	HB	1653	5	7	6	3	7	4	32	5	27	21	0.114
	H	2959	5	11	2	1	5	0	24	1	23	3	0.069
	hopper	1435	5	2	3	2	3	0	15	11	4	6	0.048
	hopper2	1680	0	3	0	2	0	0	5	1	4	0	0.042
	looper1	1881	1	0	0	0	2	0	3	0	3	2	0.048
	mariner2	912	5	5	2	2	0	3	17	4	13	13	0.144
	NOF	4347	2	0	2	2	1	0	7	0	7	4	0.013
	P	2907	0	0	0	0	0	0	0	0	0	0	-
	pogo	2121	9	12	5	7	11	0	44	5	39	5	0.020
	S	1736	9	4	11	11	12	4	51	14	37	27	0.079
	S2	1735	1	3	3	3	2	1	13	0	13	8	0.268
	Tc1	1666	2	1	4	1	2	11	21	1	20	7	0.085
	transib1	2167	0	0	2	0	0	0	2	0	2	2	0.000
	transib2	2844	3	5	2	2	0	0	12	0	12	6	0.082
	transib3	2883	0	1	1	1	1	3	7	0	7	4	0.088
	transib4	2656	2	1	2	0	0	0	5	1	4	4	0.152
FB	*FB*	1492	12	4	2	8	4	2	32	13	19	13	0.066

The canonical length of each element (in bp) is shown in column 3, the total numbers of each family on each chromosome arm in columns 4-9, the grand totals for each family in column 10, and the numbers that are full length, partial and in the most proximal 2 Mb of the major chromosome arms, in columns 11-13. Partial elements are defined as those whose length is less than 97% of the canonical element. The average pairwise distance within each family is shown in column 14. The *Cr1a* family could not be reliably aligned and therefore average pairwise distance was not computed (NC).

Chromosomal distribution of elements

The percentage of each chromosome arm composed of transposable elements varies between 3.11% and 4.29%, except for chromosome 4, which is over 10% transposable elements (Table 1). The average transposable element density is 10-15 per million bases (Mb) for the major chromosome arms, and over 82 per Mb for chromosome 4. These densities are greater than the estimate of 5 per Mb derived from lower-resolution cytological methods [48], presumably because unclustered elements and partial elements may give weak *in situ* hybridization signals. In contrast to previous findings and theoretical expectations [22,49], we found no evidence for a reduction in density of transposable elements on the X chromosome relative to the major autosome arms (Table 1).

The densities of LINE-like elements and TIR elements on chromosome 4 are from five to ten times higher than their densities on the major chromosome arms, 25.85 per Mb and 46.85 per Mb, respectively, compared to 2.51-5.37 per Mb and 2.15-3.42 per Mb, respectively (Table 1, see also [22]). By contrast, the density of LTR elements on chromosome 4 is only slightly higher (8.08 per Mb) than on the five major chromosome arms (5.01-6.90 per Mb). Moreover, the percentage of chromosome 4 that is composed of LTR elements is only slightly higher than that of the major chromosome arms (3.56% versus 2.23-2.89%). Thus, the difference in density of transposable elements on chromosome 4 is predominantly due to an order-of-magnitude increase in the number of LINE-like and TIR elements.

Transposable element density is also known to vary along the major chromosome arms [20-22]. As shown in Figure 2 and discussed in Table 1, the density of transposable elements increases in the proximal euchromatin, here defined as the proximal 2 Mb of the assembly of each of the five major chromosome arms constituting about 10% of the euchromatic sequence analyzed. On the major chromosome arms, 36.7% (577/1,572) of the elements are located in proximal euchromatin, consistent with previous observations that the density of transposable elements is higher in heterochromatic regions of the genome [50-54]. These proximal sequences represent the transition between euchromatin and heterochromatin. Of 14 families located exclusively within the proximal 2 Mb, 12 are low copy number (defined here as less than 8 copies). Elements belonging to low copy number families show some tendency to be located in these regions; 78 of 142 elements that belong to low copy number families are located in the proximal 2 Mb of the chromosome arms.

Finally, although the densities of transposable elements in the proximal euchromatin and chromosome 4 are both elevated with respect to the euchromatic average (58 and 82 elements per Mb, respectively), the composition of the elements in these regions is quite different. The increase in transposable element density in the proximal regions of the chromosome arms is due to increased numbers of elements belonging to all structural classes (Figure 2), while the increase in elements on chromosome 4 is due almost exclusively to LINE-like and TIR elements.

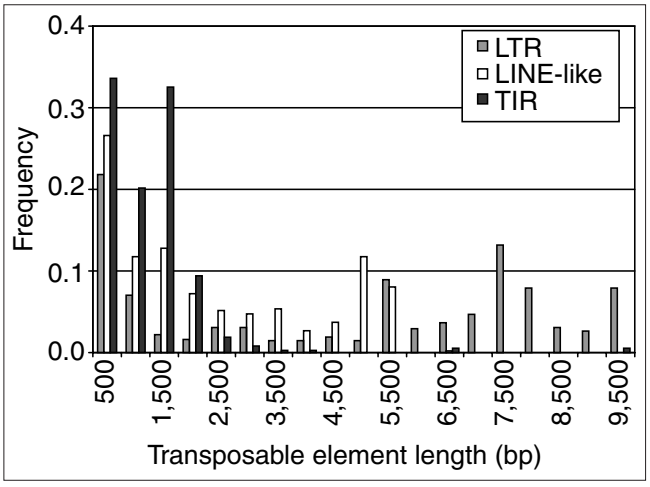

Figure 1
Frequency distribution of transposable element lengths in the *Drosophila* genome. Plotted are the lengths (in bp) of individual elements by functional class: LTR (gray), LINE-like (white), and TIR (black). Pairwise tests among all three classes (LTR versus LINE-like, LTR versus TIR, and LINE-like versus TIR) reveal that the distribution of individual element lengths differ significantly between functional classes (Mann-Whitney *U* test, $p < 1 \times 10^{-6}$).

Analysis of structural variation

Transposable elements can be autonomous or defective with respect to transposition. Defective elements often exhibit deletions in ORFs or terminal repeats which are necessary for transposition. Assuming that canonical elements represent full-length active copies, we defined any element less than 97% of the length of the canonical member of their family as partial. On the basis of this criterion, more than two-thirds (1,094/1,572) of the elements in the Release 3 sequence are partial (Table 1). The proportion of partial elements is reasonably uniform among major chromosome arms (64-73%); 463 of 1,001 (46.3%) partial elements on the major chromosome arms lie within the proximal 2 Mb. In contrast, 91% of the transposable elements on chromosome 4 are partial. As LINE-like and TIR elements make up 88% of the elements on chromosome 4, these data indicate differences in proportions of partial elements between classes. In fact, 79% of LINE-like elements and 84% of TIR elements are partial, whereas only 55% of LTR elements are partial (Table 1).

Twenty-five of the 93 families (26.8%) represented in Release 3 are composed entirely of partial elements (8 LTR, 11 LINE-like, 6 TIR). An additional 17 families have only one full-length element. Fifteen of the 25 partial-only families are low copy number (less than 8 copies; 5 LTR, 6 LINE-like, 4 TIR) and 10 are high copy number (3 LTR, 5 LINE-like, 2 TIR). The majority of elements (133/196) in these 25 low

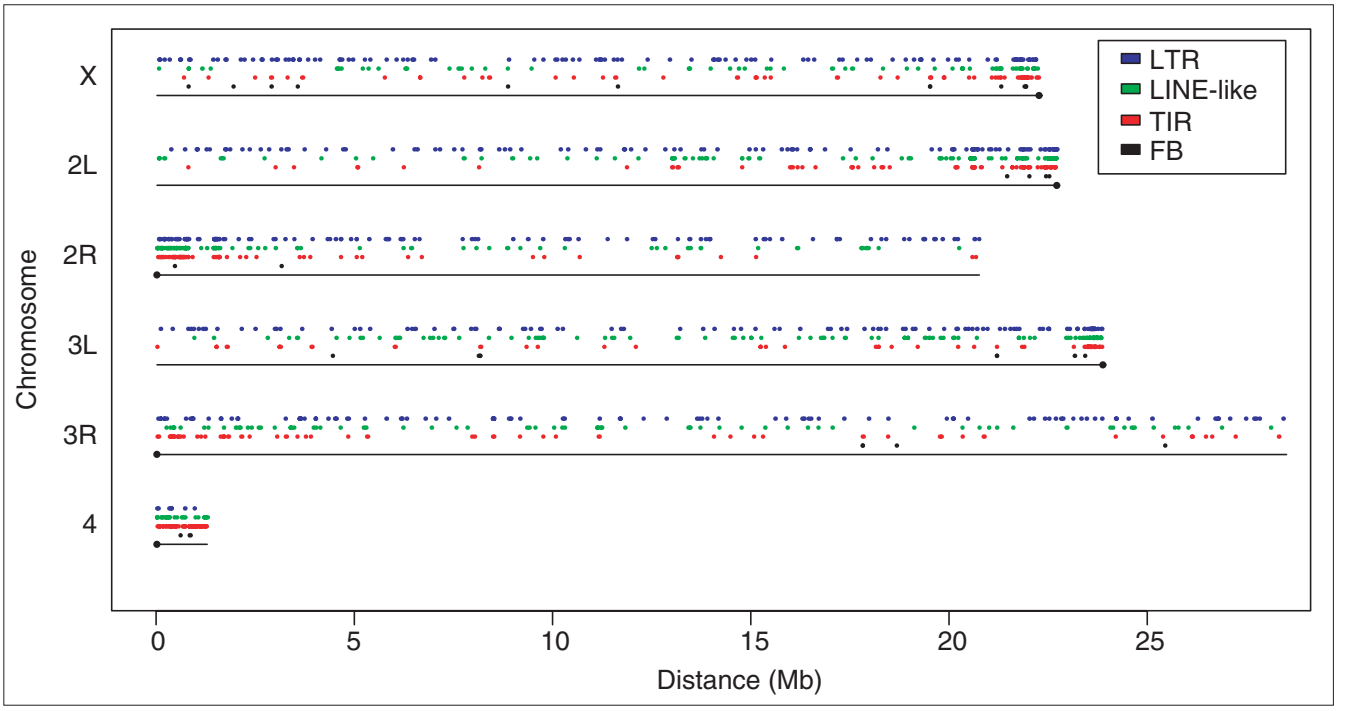

Figure 2
Distribution of transposable elements along chromosome arms. For each chromosome arm, the centromeres are indicated by circles. Each colored tick marks the start coordinate of an element belonging to one of the four classes of element (see key). Note the large number of LTR elements (blue) relative to the other classes on the major chromosome arms, and the higher number of LINE-like (green) and TIR (red) elements relative to the number of LTR elements seen for chromosome 4. While there is a relatively even distribution of transposable elements throughout the majority of each arm, there is a significant increase in the density of all classes of element in the proximal euchromatin (see also Table 1).

copy number families are found in the proximal 2 Mb or on chromosome 4; all elements for 9 of these 25 families are found exclusively in these regions of the genome.

Analysis of the distribution of transposable element lengths scaled relative to the length of their canonical sequence shows that all three classes have bimodal distributions of scaled element lengths, but differ significantly from one another (Figure 3). The bimodal shape of these distributions presumably reflects the boundary states of the dynamic process of deletion, excision and transposition. Only a very small number of LINE-like (6) and TIR (16) elements exceed their canonical length, consistent with the fact that deletions occur more frequently than insertions in *D. melanogaster* [55]. A higher number of LTR elements (30) exceed the length of their canonical sequence, but on average these elements are less than 2% longer than their canonical length.

We characterized the distribution of structural variation for a representative element from each of the three major classes by determining the proportion of sequences represented in multiple alignments for a given nucleotide site (Figure 4). The resulting plot for the LINE-like *jockey* family approximates a negative exponential distribution starting from the 3′ end (Figure 4a). LINE-like elements become deleted preferentially at their 5′ ends, as a consequence of their mechanism of transposition [56]. The TIR element *pogo* shows a very different pattern; internal deletions predominate, leaving the inverted repeat termini intact [57] (Figure 4b). By analogy with patterns of deletion in *P* elements [58], these deleted elements will be non-autonomous with respect to transposition and presumably arise when double-stranded gap repair is interrupted [59,60]. By contrast, for the representative LTR element *roo*, there is a relatively uniform pattern of structural variation across the element, with the exception of two apparent deletion hotspots, at coordinates approximately 1 kb and approximately 8 kb, both of which occur in regions that are expected to be coding (Figure 4c).

One class of defective LTR elements, solo LTR sequences, has been known for some time in *Drosophila* [61] and other species [62-64]. These presumably arise by exchange between the two LTRs flanking an element, with the loss of the reciprocal product, a small circular molecule. In *Saccharomyces cerevisiae*, 85% of all LTR element insertions are solo LTRs [65]. We screened for solo LTRs of each family of element, using a criterion of 80% identity to the canonical LTR sequence of each

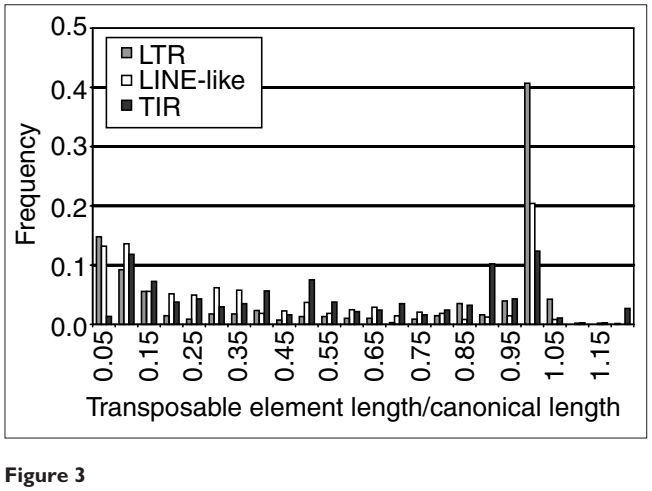

Figure 3
Frequency distribution of transposable element lengths scaled relative to their canonical lengths. Plotted are the scaled lengths of individual elements by functional class: LTR (gray), LINE-like (white), and TIR (black). Mann-Whitney *U* tests among all three classes (LTR versus LINE-like, LTR versus TIR, and LINE-like versus TIR) reveal that the distribution of scaled element lengths differ significantly between functional classes (Mann-Whitney *U* test, $p < 1 \times 10^{-4}$).

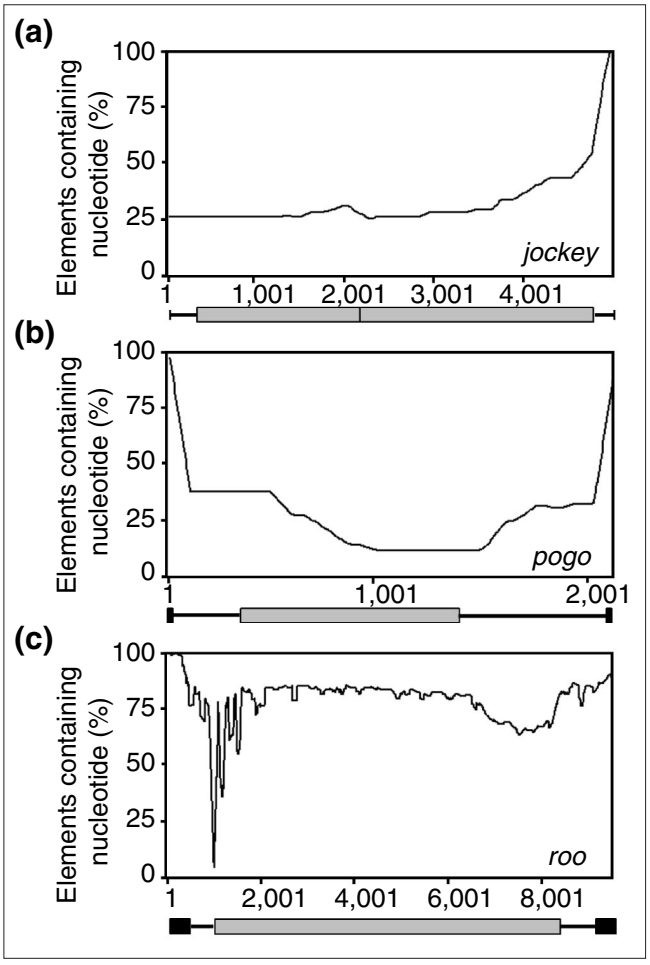

Figure 4
Structural variation within three common transposable elements: **(a)** *jockey*; **(b)** *pogo*; **(c)** *roo*. Multiple alignments generated from each respective family were used to approximate genomic variation. Each position along the length of the multiple alignment (*x*-axis) was measured for the presence of a nucleotide. The percentage of elements within an alignment that contained the nucleotide was determined and is indicated along the *y*-axis. A schematic drawing representing a *jockey*, *pogo*, or *roo* element is shown directly below each panel; coding regions are indicated by light-gray boxes and repeats (*pogo* and *roo*) are represented by black boxes.

family. Only 58 solo LTRs were identified, of which 14 are *roo* LTR elements.

Analysis of sequence variation within families

Point mutations in coding regions of the *gypsy* family of retrotransposons correlate with both transposition frequency and copy number [66]. We identified only one full-length *gypsy* class element (FBti0019898). Sequence comparison of this *gypsy*'s ORF2 with that of the 'active' strain ORF2 shows these two ORFs to be identical and suggests that the single full-length *gypsy* element is 'active' in the sequenced strain. Other families of element have also been found to be polymorphic with respect to their coding potential in *Drosophila*. Kalmykova *et al*. [67] found that most *1731* elements have the +1 frameshift between their *gag* and *pol* gene regions typical of LTR elements; some do not, however, and instead express a Gag-Pol fusion protein. The single full-length *1731* element in Release 3 (FBti0020325) is of the latter type.

Sequence variation within families of element was estimated by analyzing the average pairwise distance within each family after multiple alignment (Table 2). These data show that intra-family variation ranges from complete identity to 26.8% average pairwise distance (see *S2* family), but with only seven families having greater than 10% average pairwise distance. Analysis of the distribution of intra-family average pairwise distances by functional class shows that LTR families have lower levels of average pairwise distance relative to LINE-like or TIR families (Figure 5). The average sequence divergence for the LTR class of elements is only 2.6%; for the LINE-like and TIR classes it is 5.4% and 7.7%, respectively. These estimates of intra-family variation are remarkably similar to those of Wensink [68] who, by studying the kinetics of DNA reassociation of cloned middle repetitive sequences, showed that families of middle repetitive sequence exhibited on the order of 3-7% sequence divergence. These data illustrate differences in the amount of within-family variation seen between classes of element, which could be due to a variety of distinct mechanisms including differences in genealogical history or selective constraints.

Nesting and clustering of transposable elements

The nesting of transposable elements is common in plant genomes [69-72]. For our analysis, transposable elements that have inserted within another element are termed nests;

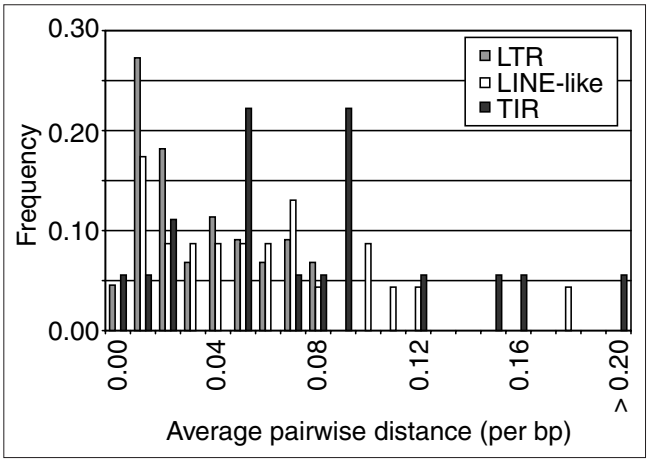

Figure 5
Frequency distribution of within-family average pairwise distances. Plotted are average pairwise distances (per bp) for individual transposable element families by functional class. Mann-Whitney *U* tests reveal that intra-family average pairwise distances differ significantly between LTR families and LINE-like families ($p < 0.005$), and between LTR families and TIR families ($p < 0.0005$), but not between LINE-like and TIR families ($p < 0.311$).

groups of transposable elements located within 10 kb of each other are defined as clusters. We found 62 nests or clusters of transposable elements containing 328 full or partial elements. This indicates that about 21% of transposable elements in this study are either inserted into another element or positioned adjacent to another element. The number of nested or clustered elements per arm ranges from 1.4 to 3.6 per Mb. The density of such elements is much higher in the proximal regions of the euchromatic arms; of the 62 nests or clusters, 25 are within the proximal 2 Mb regions of the major chromosome arms [73]. Eighty-nine percent of the elements belonging to nests or clusters are partial, in contrast to 69% of all elements. LTR elements are nested or clustered more often (29.3%) than either LINE-like elements (12.0%) or TIR elements (15.8%). This is presumably due to the larger proportion of LTR elements present in the *Drosophila* euchromatin.

Foldback (*FB*) elements often contain non-*FB* sequences [30,74]. Both *NOF* and *HB* elements have been found flanked by *FB* arms [30,75]. We identified two *HB* elements immediately adjacent to *FB* elements and four examples of *NOF* elements inserted into *FB* elements.

Patterns of element nesting can be very complex, as has been observed in other species [72], and may involve elements of the same class or elements of different classes. The insertion of a transposable element may trigger a runaway process, since it will provide a target into which other elements may insert without deleterious consequences [76]. In a sample of 31 simple nests each involving only two or three elements, we observed all nine possible combinations of nesting among the LTR, LINE-like and TIR classes. The largest euchromatic complex of elements is on chromosome arm 3R (coordinate approximately 8.3 Mb); it is a complex of 30 fragments of *Dm88*, 18 fragments of *invader1* and three fragments of *micropia* elements, occupying 32.4 kb. Many of these fragments are identical; for example, of the 18 *invader1* fragments, nine represent bases 1-424 of the canonical *invader1* LTR sequence, three represent bases 143-424, two bases 80-424 and two bases 1-108. Losada *et al*. [77] have suggested that some novel transposable elements have evolved by nesting, in particular, that the *Circe* element arose as a consequence of the insertion of the *Loa*-like element of *D. silvestris* into the *Ulysses*-like element of *D. virilis*.

Several complex nests involve many different families of element. The nest near the base of 2L (coordinate approximately 20.1 Mb), for example, involves 11 different families of all three major classes of element. Large clusters containing only one family of element are also found. For example, there is a complex of seven *GATE* elements at coordinate approximately 14.2 Mb on chromosome arm 2R and a complex of six *mdg1* elements at approximately 5.7 Mb on the same chromosome arm.

Some transposable elements are present as large tandem arrays. For example, the *Tc1*-like *Bari1* is organized as a tandem array in the heterochromatin at the base of chromosome arm 2R [78]. Tandem LTR element pairs have also been found in the *D. melanogaster* genome (for example, FBti0019752 and FBti0019753); here, two *roo* elements share an internal LTR. A number of different mechanisms have been suggested to result in tandem *Ty1* and *Ty5* elements in *S. cerevisiae* [65,79]; all involve recombination between either linear cDNAs or circular DNA generated by LTR transposition and a chromosomal element. The mechanism(s) by which tandem elements arise in *Drosophila* is not known.

Insertion-site preferences of natural transposable elements

Transposable elements insert at a staggered cut in chromosomal DNA; after repair, this results in a duplication of the target sequence. For the *R1* and *R2* LINE-like elements, there is high insertion-site specificity for sites within the 28S rDNA gene [38]. For some LTR retrotransposons, a preference for AT-rich sequences has been known for some time [80-82]. We estimated the physical characteristics of 500 bp of DNA flanking the insertion sites of three high-copy number elements, *roo* (LTR), *jockey* (LINE-like) and *pogo* (TIR). In our analysis, we included only elements for which the duplicated target sequence could be unambiguously identified. Within the individual element families, we found no recognizable motif in the nucleotide sequence of the repeats flanking insertions (data not shown). However, analysis of different physical characteristics of these

sequences revealed distinct characteristics for each of the three families of element (Figure 6). Our data suggest that *roo* and *pogo* prefer to insert in sequences of either higher than average (*roo*) or lower than average (*pogo*) denaturation temperatures; this may reflect functional differences in the insertion mechanism of these elements. There is no obvious bias in the sequences into which *jockey* elements insert. Analysis of other high-copy families of element revealed distinct insertion site characteristics, suggesting that the characteristics shown in Figure 6 are not shared within classes of element (not shown).

LTR retrotransposons use a tRNA primer for first-strand synthesis during transposition. In *S. cerevisiae*, 90% of LTR retrotransposons are within 750 bp of tRNA genes, and there are an average of 1.2 insertions per tRNA gene [65]. Our data suggest no relationship between the location of tRNAs and transposable elements in *D. melanogaster*. Of 313 elements on chromosome arm 2R, only five are within 10 kb of a tRNA gene, or tRNA gene cluster. However, Saigo [83] has described an association of a tRNA pseudogene and the 3′ end of a *copia* element, possibly resulting from an aberrant reverse transcription; an initiating tRNA:Met pseudogene has also been described as being associated with repetitive sequences [84].

It has been known for many years that the *P* element shows a marked preference to insert immediately 5′ to genes or within 5′ exons [85]. This preference presumably reflects the chromatin environment at the time of *P*-element transposition. We analyzed the position of transposable elements with respect to the closest known or predicted gene from the Release 3 reannotation [86]. There are 551 elements located 5′ to transcribed regions, 585 elements located 3′ to transcribed regions, and 436 elements within transcribed regions. These ratios are consistent for all classes of element, suggesting that there is no insertion site bias with respect to genes. We also find no bias of insertion with respect to the transcribed strand.

The proportion of transposable elements is higher in intergenic regions than in transcribed regions. Only 27.7% (436/1,572) transposable elements map within regions that are annotated as transcribed, although over 50% of the major chromosome arms are predicted to be transcribed. This result suggests that a large proportion of transposable element insertions in transcribed regions have deleterious effects and are not incorporated into the genome of *D. melanogaster*. As with total numbers of transposable elements on each chromosome arm, we see no reduction in the number of transposable elements inserted within transcribed regions on the X chromosome. Of the 436 transposable elements inserted within genes, 79 are on the X chromosome, which is within the range seen on the other major chromosome (70-88). This is consistent with the percent of coding/non-coding sequence on the X chromosome (51.9%) relative to that of the other chromosome arms (53.8%).

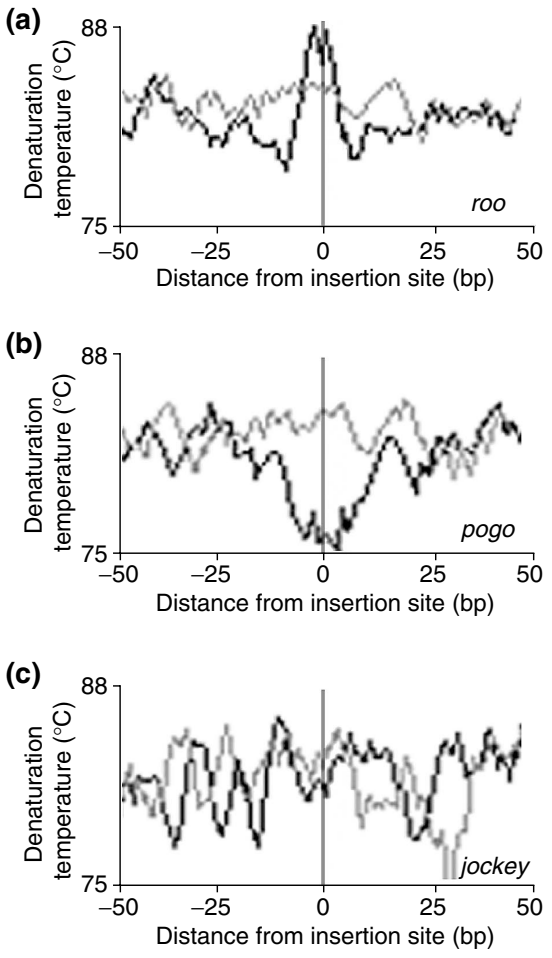

Figure 6
Structural characteristics of DNA insertion sites. Sites for **(a)** *roo*; **(b)** *pogo*; **(c)** *jockey*. Genomic sequence flanking the insertion site of each element was extracted from our dataset. Those elements for which duplicated target sequences could not be identified were discarded. The remaining sequences from each family were centered on the repeat (vertical gray line) and the average denaturation temperature across all sequences was determined using a 3-bp window size. In each panel, the light horizontal gray line represents the average denaturation temperature of random genomic sequence and the horizontal black line represents the average denaturation temperature of the experimental set of sequences. The *x*-axis represents the distance (in bp) from the insertion site and the *y*-axis represents the temperature (°C). The sequences flanking the *roo* (a) and *pogo* (b) elements have opposite characteristics; the *roo* sequences have a higher than average denaturation temperature whereas the *pogo* sequences have a lower than average denaturation temperature. The average denaturation temperature of the sequence flanking the *jockey* elements does not differ from that of the random sequence.

All 436 transposable elements that map within transcribed regions are predicted to be within introns (see also [22]). However, during the reannotation of the genome [86],

coding exons were not annotated in sequences with homology to transposable elements. Thus it is possible that a small number of transposons within transcribed regions actually are inserted into a coding exon. It is worth noting that, of the four mutations known to be carried by the sequenced strain one (bw^1), and possibly a second (sp^1), are mutated by the insertion of *412* elements.

In a recent study of five protein-coding genes located in the proximal regions of chromosome arms, Dimitri *et al.* [35] found that introns contain 50% transposable element sequence; this contrasts with euchromatic introns, which contain only 0.11% transposable element sequence.

Transposable elements in completely sequenced genomes

We can make a preliminary comparison of the transposable elements of *D. melanogaster* with those of the other fully sequenced eukaryote genomes: *S. cerevisiae* [87], *Schizosaccharomyces pombe* [64], *Caenorhabditis elegans* [88] and *Arabidopsis thaliana* [89].

In *S. cerevisiae*, all transposable elements are of the LTR class [62]; five different families are known and these comprise 3.1% of the entire genome in *S. cerevisiae*. The majority of these are solo LTRs (85%) [65]. This is in contrast to the few solo LTRs found in the Release 3 sequences of *D. melanogaster*. Transposable elements are quite rare in the sequenced strain of *S. pombe*; only 11 intact, and three defective, *Tf2* LTR elements are known [64,90].

In *C. elegans*, all three major classes of transposable element are found. There are 19 families of LTR retrotransposons, with at most three full-length members [25,63]. There are three families of LINE-like retrotransposons, *Rte-1*, *Sam* and *Frodo*, with about 30 elements overall; 11 of *Sam*, three of *Frodo* and 10-15 of *Rte-1* [91,92]. There are seven families of TIR (*Tc*) elements, with copy numbers between 61 and 294 [93], nine families of *mariner*-like element, with from one to 66 copies ([45], quoted in [46]), and five families of short DNA elements [94], with copy numbers from 81 to 1,204 [93]. Given the occurrence of retrotransposons in both *C. elegans* and *D. melanogaster*, it is interesting that the genomes of both species are characterized by very few retrotransposed pseudogenes [95-97]. These may be generated at a low rate, or may be deleted quickly as suggested by studies of the lineages of *Helena* elements in *D. virilis* [98] and *D. melanogaster* [55].

Transposable elements are far more abundant in the genome of *A. thaliana* than in the euchromatic genomes of *C. elegans* or *D. melanogaster*. In *Arabidopsis*, over 5,500 transposable elements exist, representing 10% of the 'euchromatic' sequence [89]. The pericentromeric heterochromatin and the heterochromatic knob on chromosome 4 of *Arabidopsis* have a very high density of transposable elements and other repeats [89,99-101]. As in *Drosophila*, certain families of elements appear only in these heterochromatic regions, for example the *Arabidopsis* retrotransposon *Athila*. In contrast to *Drosophila*, class I and class II elements in *Arabidopsis* show very different chromosomal distributions, the former in the centromeric regions and the latter flanking these regions.

One class of element that is absent or so far unrecognized in the genome of *D. melanogaster* are the MITEs, miniature inverted repeat elements, characterized as short (under 500 bp) elements with inverted repeat termini and without a transposase necessary for autonomous transposition. Elements similar to MITEs have been described in *D. subobscura* and its relatives [102]. Whether or not MITEs are indeed a separate class of element, or simply represent internally deleted (and hence non-autonomous) TIR elements is unclear [103]. In *C. elegans*, these elements are abundant, with 5,000 elements in four sequence families; they show a non-random chromosomal distribution [104]. MITEs are characteristic of plant genomes; in *A. thaliana*, Surzycki and Belknap [105] have identified three families with a copy number of about 90. In maize there are an estimated 6,000 copies of the *mPIF* family of MITE elements alone, and there is evidence for autonomous family members [106].

An additional class of element not identified in the *Drosophila* genome are the SINEs (short interspersed elements) [29,107]. This class of element is often closely associated with LINE elements and it has been proposed that the transposition of SINEs utilizes proteins encoded by LINE elements [107]. The *DINE-1* family of repetitive element, isolated in *Drosophila* [32], shares weak similarity to the SINEs, but this family lacks important structural features typical of other members of this class.

Comparison of sequence and cytological data

In this paper we have described the transposable elements of the euchromatin of *D. melanogaster*, as represented by the Release 3 sequences. Because Release 3 represents only a single sequence from the y^1; cn^1 bw^1 sp^1 isogenic strain first constructed in J. Kennison's laboratory in the early 1990s [108], it is important to determine whether the composition of transposable elements in this strain is typical of the species as a whole.

It is well established that *Drosophila* strains vary in the number and location of transposable elements; these differences are often taken as *de facto* evidence of transposability [12,13]. Large differences in the abundance of transposable elements have been observed between laboratory strains for families such as *gypsy*, *Bari1*, *ZAM* and *Idefix* [109-112]. Such variation in transposable element copy number may be associated with a mutation either of the element itself, or of host genes that would normally regulate copy number

(see, for example, the role of the *flamenco* gene in regulating *gypsy* activity [113]; see [114,115] for a review of host-element interactions). Transposable elements also differ between laboratory strains and natural populations of *D. melanogaster* (Table 3, see [116] for review). Perhaps the most salient examples are those elements that cause hybrid dysgenesis - the *P* element, the *I* element and the *H* element (a.k.a. *hobo*) - which are either wholly absent from, or defective in, most laboratory strains, but abundant today in natural populations [58,117,118]. Only one of the *H* elements appears to be full length, and comparison of its coding sequence to that of the canonical suggests that this element is active in the sequenced strain. Further, eight of the *I* elements identified in the sequenced strain appear to be of similar length and sequence to the active canonical element. The ORFs of these elements are very similar to those of the canonical, suggesting that they too might be active.

There has been extensive sampling of laboratory and natural strains of *D. melanogaster* for euchromatic transposable elements by the method of *in situ* hybridization [48,116]. These samples provide estimates of transposable element abundance that are relevant to compare with our data, as both types of studies sample euchromatic sequences. As shown in Table 3, the sequenced $y^1; cn^1\ bw^1\ sp^1$ isogenic strain is a typical *D. melanogaster* strain, at least with respect to the numbers of euchromatic elements. Overall, the Spearman rank order correlation coefficient between the number of elements of each family in Release 3 and the average mid-point of the ranges seen in other strains is 0.86 ($p < 10^{-6}$). The correlation between these two types of data is imperfect as closely located elements (within 100 kb) of the same family and grossly deleted elements are not resolved by the method of *in situ* hybridization. Moreover, the copy number of any individual element may be very different in different strains (see above), and certain elements (for example, *copia*, *Doc*, *roo*) may dramatically increase in copy number in particular laboratory strains (see [119,120]). Nevertheless, this strong correlation suggests that results based on analysis of the sequenced strain may be representative of the species as a whole.

As previously noted, 25 of the 93 families of transposable element represented in the Release 3 euchromatic sequence have only partial elements. Full-length copies of these families may be discovered in other strains of *D. melanogaster*, in the heterochromatin, or in closely related species. Indeed, full-length copies of the *aurora* family are present in *D. simulans* (*Dsim/ninja* [121,122]), of the *mariner* clade in *D. mauritiana* [123], and of the *Helena* family in *D. virilis* [124]. Our understanding of the evolutionary dynamics of the transposable elements will also be immeasurably improved by comparative studies between *D. melanogaster* and other Diptera, such as *Anopheles gambiae* and *D. pseudoobscura*.

Materials and methods

The sequence and other datasets

The sequence releases

Release 1 of the 'complete' euchromatic sequence of the genome of *D. melanogaster* was made available in March 2000 [20,26]. As explained in Background, this sequence, by and large derived from an assembly of a 12.8x whole-genome shotgun sequence, is not suitable for the analysis of repeated sequences. Nor was the subsequent release made available from the Berkeley *Drosophila* Genome Project's (BDGP) website, Release 2 (October 2000), which filled 330 sequence gaps (but left some 1,300) and had improved the order and orientation data for scaffolds [125]. Release 3 is the first high-quality complete sequence of this genome; it is of Phase 3 quality [126]; that is, not only is the sequence quality itself high (an estimated sequence error rate of less than 1 in 30,000 bp), but there are very few gaps.

There are two caveats with respect to the assembly used in this analysis (see [28] for details). The first is that regions containing arrays of similar elements may suffer from assembly artifacts. Resolving these will require the use of a constrained assembly program. The second is that some parts of the distal X chromosome and of chromosome arm 3L are unfinished in Release 3 [28]. The transposable elements of these unfinished sequences were included in analyses of the abundance and distribution of elements, but were excluded from all analyses that required the alignment of element sequences. Seventy-three elements are included within Tables 1 and 2, but not included in alignments. These elements will be clearly indicated in our datasets (see below).

In Release 1 there were 3.8 Mb of sequence that could not be mapped to any chromosome arm [20]. These were assembled into unmapped scaffolds that included sequences from gaps in the euchromatic arms and sequences from the centric 'heterochromatin', including the entire Y chromosome [127]. The highly repetitive nature of these sequences makes them difficult to assemble, either from a whole genome shotgun or from a cloned-based sequencing strategy. The 'heterochromatin' also includes sequences that cannot be readily cloned, for instance the satellite DNA sequences and, perhaps, others. The distinct genetic and cytological nature of the pericentromeric regions of the *Drosophila* chromosomes, both metaphase and polytene interphase, has been known for many years [128,129] and its structure and properties clearly result from the nature of its sequences [130].

We defined euchromatin as all sequence that has been assembled into a chromosome arm scaffold, and heterochromatin as the rest (unmapped scaffolds, see [37]). We recognize, of course, that the transition from euchromatin to heterochromatin is not abrupt; indeed we show that the characteristics of the most basal regions of the chromosome arms differ in their sequence organization from the regions distal to them.

Table 3

A comparison of the numbers of euchromatic transposable elements in the Release 3 sequence with those estimated from natural populations and laboratory stocks by *in situ* hybridization

Class	Family	Number in Release 3	Range	Midpoint	Midpoint average	Source	Reference
LTR	*17.6*	12	8–17	12.5	12.5	7-10 isogenic lines, Beltsville, MD	[51]
	1731	2	8–15	11.5	11.5	7-10 isogenic lines, Beltsville, MD	[51]
	297	57	18–35	26.5	26.6	7-10 isogenic lines, Beltsville, MD	[51]
			20–30	25		4 lab stocks	[13]
			23	23		20 isogenic lines, Raleigh, NC	[49]
			32	32		182 inbred lines	[145]
	3S18	6	25	25	18.5	8 lab stocks	[146]
			12	12		6 lab stocks	[19]
	412	31	18–38	28	25.8	7-10 isogenic lines, Beltsville, MD	[51]
			26–32	29		4 lab stocks	[13]
			21	21		20 isogenic lines, Raleigh, NC	[49]
			31	31		7 natural populations	[147]
			20	20		182 inbred lines	[145]
	blood	22	14	14	14.0	182 inbred lines	[145]
	copia	30	20–43	31.5	24.3	4 lab stocks	[13]
			12–25	18.5		18 inbred strains, Azerbaidjan	[148]
			23	23		7 natural populations	[147]
	gypsy	2	1	1	3.0	7 natural populations	[147]
			5	5		182 inbred lines	[145]
	HMS-Beagle	13	11	11	11.0		Hey and Eanes (unpublished), [116]
	Idefix	7	4–20	12	12.0	6 lab stocks	[110]
	mdg1	25	11–23	17	20.3	7-10 isogenic lines, Beltsville, MD	[51]
			15–27	21		20 lab stocks	[149]
			14–22	18		17 inbred stocks, Azerbaidjan	[150]
			25	25		18 inbred strains, Azerbaidjan	[148]
	mdg3	16	5–18	11.5	11.5	20 lab stocks	[149]
	opus	24	12–23	17.5	15.0	7-10 isogenic lines, Beltsville, MD	[51]
			10–15	12.5		2 lab stocks	[151]
	roo	147	55–75	65	63.0	7-10 isogenic lines, Beltsville, MD	[51]
			61	61		20 isogenic lines, Raleigh, NC	[49]
	Stalker	12	2–6	4	4.0	8 lab stocks	[146]
	Tirant	20	3–13	8	9.5	10 wild-type stocks	[152]
			6–16	11		3 lab stocks	[153]
	ZAM	0	0–15	7.5	7.8	4 lab stocks	[111]
			1–15	8		3 lab stocks	[39]
LINE-like	*Doc*	55	20–30	25	25.0	2 lab stocks	[154]
	F	42	25–30	27.5	34.3	1 lab stock	[155]
			41	41			Hey and Eanes (unpublished), [116]
	I	28	13–21	17	12.3	18 inbred strains, Azerbaidjan	[148]
			0–15	7.5		6 lab stocks	[19]
	jockey	69	32–40	36	29.0	7-10 isogenic lines, Beltsville, MD	[51]
			22	22		182 inbred lines	[145]
TIR	*1360*	105	19–39	29	29	4 lab strains	[156]
	Bari1	5	4	4	9.8	4 lab stocks	[157]
			2–29	15.5		46 lab stocks and natural populations	[109]
	H	24	8–60	34	27.0	17 inbred stocks, Azerbaidjan	[150]
			25–27	26		natural populations, Greece	[158]
			21	21		182 inbred lines	[145]
	NOF	7	0–2	1	1.0	8 lab strains	[159]
	S	51	24–91	57.5	57.5	10 lab stocks and natural populations	[160]
FB	*FB*	32	20–30	25	22.5	8 lab strains	[159]
			20	20		182 inbred lines	[145]

These data are illustrative of the published literature, not an exhaustive survey. The estimated range of copy number per family, range midpoint for each source, and midpoint averages across all sources are shown in columns 4, 5 and 6, respectively.

The dataset we used for unmapped scaffolds is from a new assembly provided by Celera Genomics ([28] and E. Myers and G. Sutton, personal communication). This assembly, WGS3, is the result of improvements to the Celera assembly algorithms [26]. WGS3 assembles 115.5 Mb into 14 mapped scaffolds, and leaves 22.2 Mb in 2,761 scaffolds, each less than 1 Mb in length, assigned to unmapped scaffolds. One reason for the increase in size of unmapped scaffolds between the first and third whole-genome shotgun assemblies is the inclusion in WGS3 of some 809,000 extra sequence reads not included in the two previous assemblies [127]. The 22.2 Mb of sequence in unmapped scaffolds includes sequences that properly belong to the euchromatin as well as heterochromatic sequences. For this reason we simply used this dataset as a subject sequence set for searching, by BLAST, for elements that we had been unable to discover in the euchromatic sequence. A description of the transposable elements of the heterochromatin, and of the telomeres, of *D. melanogaster* is the subject of a publication from the *Drosophila* Heterochromatin Genome Project [37].

Reference datasets

A reference dataset of 'canonical' sequences of transposable elements was built by M. Ashburner, P. Benos and G. Liao during the early stages of developing methods for *Drosophila* genome annotation. It was first used during the annotation of the 2.9 Mb '*Adh* region' [131] and has been maintained subsequently, and made public, by the Cambridge and Berkeley groups [132]. As new sequences were published by others these were added to this file. Most of these were 'real' sequences, although some from Repbase [23,42] were consensus sequences. In addition, we made a determined effort to discover, from the evolving Release 3 sequence, 'complete' sequences of the many elements known only from small sequence fragments, for example of their LTR regions, as well as all new elements identified in this study. The following elements were available too late to be included in our analyses, but will be included in further updates of the data: *Tc3-like* [133], *ninja* (EMBL: AF520587) and the elements 'DREF', 'BG.DS00797' and 'CG.13775' of Robertson [46]. The sequences described by Robertson [46] may be examples of functional host genes derived from transposable elements, such as are known in humans (for example [44]) and ciliates (for example [45], and H. Robertson, personal communication).

Nomenclature

Many transposable elements of *D. melanogaster* have been described and named independently by several research groups. In this paper we use the names adopted by FlyBase, which attempts to reflect priority of publication (or sequence release). There are, in addition to those described here, many elements in FlyBase that have never been associated with a sequence or a restriction map. In the absence of further evidence, nothing more can be said about these, which will be marked as being of 'uncertain status' in their FlyBase records.

Available datasets

The following datasets are freely available for download [134] and are maintained by FlyBase. When using these resources please note, and publish, the Release numbers associated with the files.

File 1. A file containing a single ('canonical') sequence of each family of element; this is a frozen dataset of the sequences used to search the genome for transposable elements.

File 2. A file of annotated 'canonical' sequences, one for each identified family of transposable element. This file is, in effect, an update version of file 1. These sequences were chosen as the longest discovered in the genome with (where relevant and where possible) intact ORFs. There are a few families for which no intact element could be found. We have then attempted to construct an intact element from the available data. Such artifices are noted in the records. These data will be updated when new information becomes available, and will be further annotated by FlyBase. Each release will be archived.

File 3. A file, in FASTA format, of each individual element that has been discovered. The following data are to be found on the header line of each record:

>family_name,FBgn_id,FBti_id,chromosome_arm:Release 3_coordinates

FBgn_id is the FlyBase record for the family, FBti_id is the unique identifier of each occurrence of an element and the coordinates are from the Release 3 data. In addition to the sequence of each element, each record includes 500 bp of 5′ and 3′ flanking sequence. These data will be regularly updated, in step with each new Release of the assembled sequence. Each release will be archived.

File 4. The alignments of elements within a family used for the current analysis. This file is in MASE format [135] with each element identified by its FBti number. This is a frozen dataset that will not be updated by the BDGP.

File 5. The nested transposable elements and element complexes are available as an independent dataset. Included within each sequence is 500 bp of flanking sequence on each side of the element complex. Each nest or complex has a unique FBti identifier number in FlyBase; in addition each component of a nest or complex has its own FBti identifier number. In the FASTA header line for each sequence in this file the data included are:

>FBti_of_nest_or_complex,FBti_of_component,chromosome _arm:coordinates

Comparison with other datasets

To support our claim that the Release 1 sequence is an inadequate substrate for rigorous analysis we have compared

the sequences of transposable elements in that release with those of Release 3. We determined the identity of elements in the two releases by a comparison of the 500 bp on their 5′ flanks. Our results suggest that many, if not most, of the sequences from Release 1 are artifacts of that assembly. Of the 1,572 elements characterized in Release 3 only 381 (24%) were correctly determined in Release 1. Of the 1,191 (76%) sequences that were not correctly sequences in Release 1, 483 contained Ns, 45 were completely absent, and 663 contained an average of 34 incorrectly identified nucleotides per element. The complete data are available from [134].

Analytical methods

Identification of known transposable elements

WU-BLASTN 2.0 [136] was used to search all chromosome arms for regions of similarity to each element in the Release 3 dataset. The parameters for the BLAST search were M = 3, N = 3, Q = 3, R = 3, X = 3 and S = 3. BLAST searches were done on a 32-node dual PIII Linux-based compute farm supplied by Linux Network. Distribution of BLAST jobs to the cluster was managed by the Portable Batch System (PBS [137]). Individual BLAST jobs were submitted via pbsrsh, an rsh-like program (E.F., unpublished work). In addition, PBS was optimized and modified for the BDGP to handle a large number of queued jobs (E.F., unpublished work).

BLAST reports were generated by searching a single chromosome arm with each individual element. The results were then parsed to generate a list of the coordinates of all high-scoring pairs (HSPs) that were at least 50 bp long and whose query and subject sequences had a pairwise identity of at least 90%. All HSPs on this list that were within 10 kb of each other and summed to greater than 100 bp were pooled into a 'span'. Each span was bounded by two coordinates - a start coordinate that corresponds to the lowest coordinate of any HSP in a particular span, and an end coordinate that corresponds to the highest coordinate of any HSP in the same span. A master list was then generated that contained all spans for all elements on a particular arm. Any spans (for the same or different elements) that had overlapping coordinates were examined further by an analysis of the sequences of the HSPs. While this identified a small number of spurious spans that did not correspond to real elements, the majority of these instances correspond to the nested elements discussed below. Start and end coordinates for all spans belonging to each element were used to extract genomic sequences for multiple sequence alignment (see below). In some rare instances where it was not possible to differentiate the element to which the HSP belonged, overlapping coordinates were recorded. Spurious sequences that did not align with other family members were removed from both the list of spans and the multiple alignments. Other attempts to define transposable element families on the basis of sequence identity have used a 90% cutoff with reference to the protein sequence of the reverse transcriptase motif of LTR-elements [25,138]. For LINE-like transposons, Berezikov *et al.* [24] used a 70% nucleic acid sequence identity criterion over 200 bp.

Identification of new transposable elements through genome-genome comparison

The first approach to discovering new transposable elements was by an all-by-all BLAST using chromosome arms 2L, 2R, 3R, 4 and the proximal half of the X. The chromosome arms were divided into 20-kb segments, each segment overlapping the previous by 10 kb. We used the NCBI-BLAST 2.0 to compare each 20-kb section against the others. Hits with greater than 95% identity and 1,000 bp long were parsed and used as query sequences in a BLAST against the canonical element sequence dataset. Redundant results were removed. The coordinates of the repeats were parsed and known repeats were tagged. New repeats were reviewed in CONSED [139] for the presence of ORFs and repeat structure.

Identification of new transposable elements through isolation of LTR sequences

A second approach was taken to identify single-copy elements containing LTRs. Each chromosome arm was divided into 1,000-bp pieces with neighboring pieces overlapping each other by 500 bp. WU-BLASTN 2.0 was used to search each chromosome arm for all regions of similarity to each 1000-bp piece (parameters: M = 3, N = 3, Q = 3, R = 3, X = 3 and S = 3). The BLAST report from such a search was parsed to generate a list of all HSPs that were at least 100 bp long and whose query and subject sequences had a pairwise identity of at least 95%. Then, all HSPs on this list greater than 500 bp apart and less than 15 kb apart were pooled into a span. As above, each span was bounded by a start coordinate which corresponds to the lowest coordinate of any HSP in a particular pool and an end coordinate which corresponds to the highest coordinate of any HSP in the same pool. Each set of coordinates was compared to the list of coordinates of transposable elements identified in the screen for known elements and these were eliminated from this list. Then, the coordinates of the remaining spans were used to extract genomic sequence from the finished chromosome arms. Each piece of genomic sequence was then compared to the coding sequence of the known transposable elements using WU-TBLASTX 2.0 (with default parameters). Any span that produced a hit with a $E < 10^{-8}$ was analyzed by searching through the non-redundant protein database at the NCBI using NCBI-BLASTX [126].

Alignment and calculation of evolutionary distances

Preliminary multiple alignments of elements within families were made using the default settings of DIALIGN v2-1 [140]. The resulting multiple alignments were visualized in the SEAVIEW alignment editor [135]. Subsequent realignment was done using the CLUSTALW (1.7.4) [141] implementation internal to SEAVIEW with manual refinement. Multiple alignments were used to calculate average pairwise distance within families using Kimura's 2-parameter substitution model

(transition:transversion ratio = 2:1) [142] as implemented in the DNADIST program of the PHYLIP package [143].

Physical characteristics of element insertion sites

To analyze the physical properties of the insertion sites of transposable elements we used the programs developed by Liao *et al.* [144]. The flanking sequences of elements with canonical ends were aligned, centered on a single copy of the element's target site sequence (that duplicated on insertion). The sequences were then analyzed for A-philicity, propeller twist, duplex stability and denaturation temperature, as described in [144]. As a baseline we used a randomly generated 500-bp sequence set of the same base composition as the overall genome of *D. melanogaster* (G. Liao, personal communication). These analyses were performed with 49 *roo* element sequences, 12 *jockey* sequences and 28 *pogo* sequences. Additional analyses were carried out using elements from the following families: *copia*, *blood*, *412* and *Doc*.

Additional data files

A table showing the classification of transposable elements in the genus *Drosophila* is available with the online version of this paper.

Acknowledgements

This work was supported by NIH grant HG00750 to G.M.R., by NIH grant HG00739 to FlyBase (W.M. Gelbart) and by programme grant G822559 from the Medical Research Council to M.A., D. Gubb and S. Russell. C.M.B is supported by NIH training grant T32 HL07279 to E. Rubin. Research was conducted at the Lawrence Berkeley National Laboratory under Department of Energy contract DE-AC0376SF00098, University of California. M.A. thanks Jean Wiborg for her help in the logistics of his visits to Berkeley. We thank Patrizio Dimitri, Bernardo Carvalho and Gary Karpen for allowing us to quote from their unpublished work, and Bob Levis, Mike Young, Stu Tsubota, Francois Payre, and Andy Flavell, for information on individual elements. We also thank Christian Biemont, Bernardo Carvalho, Patrizio Dimitri, Dan Hartl, Sergey Nuzhdin, Dmitri Petrov, Hugh Robertson, and Alfredo Ruiz, for their comments on the manuscript of this paper. We also thank the anonymous reviewers for their helpful suggestions and comments.

References

1. Craig NL, Craigie R, Gellert M, Lambowitz AM (Eds): *Mobile DNA II*. Washington, DC: ASM Press; 2002.
2. Green MM: **Genetic instability in *Drosophila melanogaster*: Mutable miniature (m^{mu}).** *Mutat Res* 1975, **29:**77-84.
3. Demerec M: **Miniature-α - a second frequently mutating character in *Drosophila virilis*.** *Proc Natl Acad Sci USA* 1926, **12:**687-690.
4. Demerec M: **Magenta-α - a third frequently mutating character in *Drosophila virilis*.** *Proc Natl Acad Sci USA* 1927, **13:**249-253.
5. Demerec M: **Unstable genes.** *Bot Rev* 1935, **1:**233-248.
6. McClintock B: **The origin and behavior of mutable loci in maize.** *Proc Natl Acad Sci USA* 1950, **36:**344-355.
7. Starlinger P: **Mutations caused by the integration of IS*1* and IS2 into the *gal* operon**. In *DNA Insertion Elements*, *Plasmids*, *and Episomes*, Edited by Bukhari AI, Shapiro JA, Adhya SL. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1977:25-30.
8. Green MM: **A case for DNA insertion mutants in *Drosophila melanogaster***. In *DNA Insertion Elements*, *Plasmids*, *and Episomes*, Edited by Bukhari AI, Shapiro JA, and Adhya SL. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1977:437-445.
9. Rubin GM, Finnegan DJ, Hogness DS: **The chromosomal arrangement of coding sequences in a family of repeated genes.** *Prog Nucleic Acid Res Mol Biol* 1976, **19:**221-226.
10. Finnegan DJ, Rubin GM, Young MW, Hogness DS: **Repeated gene families in *Drosophila melanogaster*.** *Cold Spring Harb Symp Quant Biol* 1978, **42:**1053-1063.
11. Ilyin YV, Tchurikov NA, Ananiev EV, Ryskov AP, Yenikolopov GN, Limborska SA, Maleeva NE, Gvozdev VA, Georgiev GP: **Studies on the DNA fragments of mammals and *Drosophila* containing structural genes and adjacent sequences.** *Cold Spring Harb Symp Quant Biol* 1978, **42:**959-969.
12. Young MW: **Middle repetitive DNA: a fluid component of the *Drosophila* genome.** *Proc Natl Acad Sci USA* 1979, **76:**6274-6278.
13. Strobel E, Dunsmuir P, Rubin GM: **Polymorphisms in the chromosomal locations of elements of the *412*, *copia* and *297* dispersed repeated gene families in *Drosophila*.** *Cell* 1979, **17:**429-439.
14. Glover DM: **Cloned segment of *Drosophila melanogaster* rDNA containing new types of sequence insertion.** *Proc Natl Acad Sci USA* 1977, **74:**4932-4936.
15. Hiraizumi Y: **Spontaneous recombination in *Drosophila melanogaster* males.** *Proc Natl Acad Sci USA* 1971, **68:**268-270.
16. Kidwell MG: **Hybrid dysgenesis in *Drosophila melanogaster*: The relationship between the P-M and I-R interaction systems.** *Genet Res* 1979, **33:**205-217.
17. Engels WR, Preston CR: **Hybrid dysgenesis in *Drosophila melanogaster*: the biology of female and male sterility.** *Genetics* 1979, **92:**161-174.
18. Bingham PM, Kidwell MG, Rubin GM: **The molecular basis of P-M hybrid dysgenesis: the role of the *P* element, a P-strain-specific transposon family.** *Cell* 1982, **29:**995-1004.
19. Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ: **The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor.** *Cell* 1984, **38:**153-163.
20. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, *et al.*: **The genome sequence of *Drosophila melanogaster*.** *Science* 2000, **287:**2185-2195.
21. Rizzon C, Marais G, Gouy M, Biemont C: **Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome.** *Genome Res* 2002, **12:**400-407.
22. Bartolome C, Maside X, Charlesworth B: **On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*.** *Mol Biol Evol* 2002, **19:**926-937.
23. Jurka J: **Repbase update: a database and an electronic journal of repetitive elements.** *Trends Genet* 2000, **16:**418-420.
24. Berezikov E, Bucheton A, Busseau I: **A search for reverse transcriptase-coding sequences reveals new non-LTR retrotransposons in the genome of *Drosophila melanogaster*.** *Genome Biol* 2000, **1:**research0012.1-0012.15.
25. Bowen NJ, McDonald JF: ***Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside.** *Genome Res* 2001, **11:**1527-1540.
26. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, *et al.*: **A whole-genome assembly of *Drosophila*.** *Science* 2000, **287:**2196-2204.
27. Benos PV, Gatt MK, Murphy L, Harris D, Barrell B, Ferraz C, Vidal S, Brun C, Demaille J, Cadieu E, *et al.*: **From first base: the sequence of the tip of the *X* chromosome of *Drosophila melanogaster*, a comparison of two sequencing strategies.** *Genome Res* 2001, **11:**710-730.
28. Celniker SE, Wheeler DA, Kronmiller B, Carlson JW, Halpern A, Patel S, Adams M, Champe M, Dugan SP, Frise E, *et al.*: **Finishing a whole genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence.** *Genome Biol* 2002, **3:**research0079.1-0079.14.
29. Deininger PL: **SINEs: Short interspersed repeated DNA elements in higher eukaryotes**. In *Mobile DNA*, Edited by Berg DE, Howe MM. Washington, DC: American Society of Microbiology; 1989:619-637.
30. Truett MA, Jones RS, Potter SS: **Unusual structure of the *FB* family of transposable elements in *Drosophila*.** *Cell* 1981, **24:**753-763.
31. Wilder J, Hollocher H: **Mobile elements and the genesis of microsatellites in dipterans.** *Mol Biol Evol* 2001, **18:**384-392.

32. Locke J, Howard LT, Aippersbach N, Podemski L, Hodgetts RB: **The characterization of *DINE-1*, a short, interspersed repetitive element present on chromosome and in the centric heterochromatin of *Drosophila melanogaster*.** *Chromosoma* 1999, **108:**356-366.
33. Locke J, Podemski L, Roy K, Pilgrim D, Hodgetts R: **Analysis of two cosmid clones from chromosome 4 of *Drosophila melanogaster* reveals two new genes amid an unusual arrangement of repeated sequences.** *Genome Res* 1999, **9:**137-149.
34. Spradling AC, Rubin GM: ***Drosophila* genome organization: conserved and dynamic aspects.** *Annu Rev Genet* 1981, **15:**219-264.
35. Dimitri P, Junakovic N, Arca B: **Colonization of heterochromatic genes by transposable elements in *Drosophila melanogaster*.** *Mol Biol Evol* 2002, in press.
36. Manning JE, Schmid CW, Davidson N: **Interspersion of repetitive and nonrepetitive DNA sequences in the *Drosophila melanogaster* genome.** *Cell* 1975, **4:**141-155.
37. Hoskins RA, Smith CD, Carlson J, Carvalho AB, Halpern A, Kaminker JS, Kennedy C, Mungall CJ, Sullivan BA, Sutton GG, *et al.*: **Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly.** *Genome Biol* 2002, **3:**research0085.1-0085.16.
38. Jakubczak JL, Burke WD, Eickbush TH: **Retrotransposable elements *R1* and *R2* interrupt the rRNA genes of most insects.** *Proc Natl Acad Sci USA* 1991, **88:**3295-3299.
39. Baldrich E, Dimitri P, Desset S, Leblanc P, Codipietro D, Vaury C: **Genomic distribution of the retrovirus-like element *ZAM* in *Drosophila*.** *Genetica* 1997, **100:**131-140.
40. Beissmann H, Walter MF, Mason JM: **Telomeres in *Drosophila* and other insects**. In *Telomeres and Telomerases: Cancer and Biology*, Edited by Krupp G, Parwaresch R. Georgetown, TX: Landes Biosciences; 2002.
41. Pardue ML, DeBaryshe PG: ***Drosophila* telomeres: two transposable elements with important roles in chromosomes.** *Genetica* 1999, **107:**189-196.
42. **Repbase Update** [http://www.girinst.org/Repbase_Update.html]
43. **Berkeley *Drosophila* Genome Project: GadFly genome annotation database of *Drosophila*** [http://www.fruitfly.org/annot]
44. Robertson HM, Zumpano KL: **Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome.** *Gene* 1997, **205:**203-217.
45. Witherspoon DJ, Doak TG, Williams KR, Seegmiller A, Seger J, Herrick G: **Selection on the protein-coding genes of the TBE1 family of transposable elements in the ciliates *Oxytricha fallax* and *O. trifallax*.** *Mol Biol Evol* 1997, **14:**696-706.
46. Robertson HM: **Evolution of DNA transposons in eukaryotes**. In *Mobile DNA II*, Edited by Craig NL, Craigie R, Gellert M, Lambowitz AM. Washington, DC: ASM Press; 2002.
47. Maside X, Bartolome C, Charlesworth B: **S-element insertions are associated with the evolution of the *Hsp70* genes in *Drosophila melanogaster*.** *Curr Biol* 2002, **12:**1686-1691.
48. Charlesworth B, Langley CH: **The population genetics of *Drosophila* transposable elements.** *Annu Rev Genet* 1989, **23:**251-287.
49. Montgomery E, Charlesworth B, Langley CH: **A test for the role of natural selection in the stabilisation of transposable element copy number in a population of *Drosophila melanogaster*.** *Genet Res* 1987, **49:**31-41.
50. Carmena M, Gonzalez C: **Transposable elements map in a conserved pattern of distribution extending from beta-heterochromatin to centromeres in *Drosophila melanogaster*.** *Chromosoma* 1995, **103:**676-684.
51. Charlesworth B, Jarne P, Assimacopoulos S: **The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. III. Element abundances in heterochromatin.** *Genet Res* 1994, **64:**183-197.
52. Junakovic N, Terrinoni A, Di Franco C, Vieira C, Loevenbruck C: **Accumulation of transposable elements in the heterochromatin and on the Y chromosome of *Drosophila simulans* and *Drosophila melanogaster*.** *J Mol Evol* 1998, **46:**661-668.
53. Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi S, Marchetti E, Caizzi R, Caggese C, Gatti M: **Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin.** *Proc Natl Acad Sci USA* 1995, **92:**3804-3808.
54. Dimitri P: **Constitutive heterochromatin and transposable elements in *Drosophila melanogaster*.** *Genetica* 1997, **100:**85-93.
55. Petrov DA, Hartl DL: **High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups.** *Mol Biol Evol* 1998, **15:**293-302.
56. Finnegan DJ: **Transposable elements: how non-LTR retrotransposons do it.** *Curr Biol* 1997, **7:**R245-R248.
57. Tudor M, Lobocka M, Goodell M, Pettitt J, O'Hare K: **The *pogo* transposable element family of *Drosophila melanogaster*.** *Mol Gen Genet* 1992, **232:**126-134.
58. Engels WR: **P elements in *Drosophila melanogaster***. In *Mobile DNA*, Edited by Berg DE, Howe MM. Washington, DC: American Society of Microbiology; 1989.
59. Engels WR, Johnson-Schlitz DM, Eggleston WB, Sved J: **High-frequency P element loss in *Drosophila* is homolog dependent.** *Cell* 1990, **62:**515-525.
60. Hsia AP, Schnable PS: **DNA sequence analyses support the role of interrupted gap repair in the origin of internal deletions of the maize transposon, MuDR.** *Genetics* 1996, **142:**603-618.
61. Carbonare BD, Gehring WJ: **Excision of *copia* element in a revertant of the white-apricot mutation of *Drosophila melanogaster* leaves behind one long-terminal repeat.** *Mol Gen Genet* 1985, **199:**1-6.
62. Boeke JD: **Transposable elements in *Saccharomyces cerevisiae***. In *Mobile DNA*, Edited by Berg DE, Howe MM. Washington, DC: American Society of Microbiology; 1989.
63. Ganko EW, Fielman KT, McDonald JF: **Evolutionary history of *Cer* elements and their impact on the *C. elegans* genome.** *Genome Res* 2001, **11:**2066-2074.
64. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, *et al.*: **The genome sequence of *Schizosaccharomyces pombe*.** *Nature* 2002, **415:**871-880.
65. Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF: **Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence.** *Genome Res* 1998, **8:**464-478.
66. Lyubomirskaya NV, Smirnova JB, Razorenova OV, Karpova NN, Surkov SA, Avedisov SN, Kim AI, Ilyin YV: **Two variants of the *Drosophila melanogaster* retrotransposon *gypsy* (*mdg4*): structural and functional differences, and distribution in fly stocks.** *Mol Genet Genomics* 2001, **265:**367-374.
67. Kalmykova A, Maisonhaute C, Gvozdev V: **Retrotransposon *17.31* in *Drosophila melanogaster* changes retrovirus-like expression strategy in host genome.** *Genetica* 1999, **107:**73-77.
68. Wensink PC: **Sequence homology within families of *Drosophila melanogaster* middle repetitive DNA.** *Cold Spring Harb Symp Quant Biol* 1978, **42:**1033-1039.
69. Fu H, Park W, Yan X, Zheng Z, Shen B, Dooner HK: **The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome.** *Proc Natl Acad Sci USA* 2001, **98:**8903-8908.
70. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL: **The paleontology of intergene retrotransposons of maize.** *Nat Genet* 1998, **20:**43-45.
71. SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, *et al.*: **Nested retrotransposons in the intergenic regions of the maize genome.** *Science* 1996, **274:**765-768.
72. Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z: **Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum.** *Proc Natl Acad Sci USA* 1999, **96:**7409-7414.
73. O'Hare K, Chadwick BP, Constantinou A, Davis AJ, Mitchelson A, Tudor M: **A 5.9-kb tandem repeat at the euchromatin-heterochromatin boundary of the X chromosome of *Drosophila melanogaster*.** *Mol Genet Genomics* 2002, **267:**647-655.
74. Caceres M, Puig M, Ruiz A: **Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions.** *Genome Res* 2001, **11:**1353-1364.
75. Harden N, Ashburner M: **Characterization of the *FB-NOF* transposable element of *Drosophila melanogaster*.** *Genetics* 1990, **126:**387-400.
76. Walbot V, Petrov DA: **Gene galaxies in the maize genome.** *Proc Natl Acad Sci USA* 2001, **98:**8163-8164.
77. Losada A, Abad JP, Agudo M, Villasante A: **The analysis of *Circe*, an LTR retrotransposon of *Drosophila melanogaster*, suggests that an insertion of non-LTR retrotransposons into**

LTR elements can create chimeric retroelements. *Mol Biol Evol* 1999, **16:**1341-1346.
78. Caizzi R, Caggese C, Pimpinelli S: ***Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization.** *Genetics* 1993, **133:**335-345.
79. Ke N, Voytas DF: **High frequency cDNA recombination of the *Saccharomyces* retrotransposon *Ty5*: The LTR mediates formation of tandem elements.** *Genetics* 1997, **147:**545-556.
80. Inouye S, Yuki S, Saigo K: **Sequence-specific insertion of the *Drosophila* transposable genetic element *17.6*.** *Nature* 1984, **310:**332-333.
81. Freund R, Meselson M: **Long terminal repeat nucleotide sequence and specific insertion of the *gypsy* transposon.** *Proc Natl Acad Sci USA* 1984, **81:**4462-4464.
82. Tanda S, Shrimpton AE, Chueh LL, Itayama H, Matsubayashi H, Saigo K, Tobari YN, Langley CH: **Retrovirus-like features and site specific insertions of a transposable element, tom, in *Drosophila ananassae*.** *Mol Gen Genet* 1988, **214:**405-411.
83. Saigo K: **A *copia* primer pseudogene possibly generated by an aberrant reverse transcription of a *copia*-related element in *Drosophila*.** *Nucleic Acids Res* 1986, **14:**7815.
84. Sharp S, DeFranco D, Silberklang M, Hosbach HA, Schmidt T, Kubli E, Gergen JP, Wensink PC, Soll D: **The initiator tRNA genes of *Drosophila melanogaster*: evidence for a tRNA pseudogene.** *Nucleic Acids Res* 1981, **9:**5867-5882.
85. Spradling AC, Stern DM, Kiss I, Roote J, Laverty T, Rubin GM: **Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project.** *Proc Natl Acad Sci USA* 1995, **92:**10824-10830.
86. Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell K, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnik SE, *et al.*: **Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review.** *Genome Biol* 2002, **3:**research0083.1-0083.22.
87. Goffeau A, Aert R, Agostini-Carbone L, Ahmed A, Aigle M, Alberghina L, Albermann K, Albers M, Aldea M, Alexandraki D, *et al.*: **The yeast genome directory.** *Nature* 1997, **387 (Suppl):**1-105.
88. The *C. elegans* Sequencing Consortium: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282:**2012-2018.
89. The *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408:**796-815.
90. Weaver DC, Shpakovski GV, Caputo E, Levin HL, Boeke JD: **Sequence analysis of closely related retrotransposon families from fission yeast.** *Gene* 1993, **131:**135-139.
91. Marin I, Plata-Rengifo P, Labrador M, Fontdevila A: **Evolutionary relationships among the members of an ancient class of non-LTR retrotransposons found in the nematode *Caenorhabditis elegans*.** *Mol Biol Evol* 1998, **15:**1390-1402.
92. Youngman S, van Luenen HG, Plasterk RH: ***Rte-1*, a retrotransposon-like element in *Caenorhabditis elegans*.** *FEBS Lett* 1996, **380:**1-7.
93. Duret L, Marais G, Biemont C: **Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*.** *Genetics* 2000, **156:**1661-1669.
94. Devine SE, Chissoe SL, Eby Y, Wilson RK, Boeke JD: **A transposon-based strategy for sequencing repetitive DNA in eukaryotic genomes.** *Genome Res* 1997, **7:**551-563.
95. Jeffs P, Ashburner M: **Processed pseudogenes in *Drosophila*.** *Proc R Soc Lond B Biol Sci* 1991, **244:**151-159.
96. Harrison PM, Echols N, Gerstein MB: **Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome.** *Nucleic Acids Res* 2001, **29:**818-830.
97. Wang W, Brunet FG, Nevo E, Long M: **Origin of *sphinx*, a young chimeric RNA gene in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2002, **99:**4448-4453.
98. Petrov DA, Lozovskaya ER, Hartl DL: **High intrinsic rate of DNA loss in *Drosophila*.** *Nature* 1996, **384:**346-349.
99. Kapitonov VV, Jurka J: **Molecular paleontology of transposable elements from *Arabidopsis thaliana*.** *Genetica* 1999, **107:**27-37.
100. Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, *et al.*: **Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.** *Nature* 1999, **402:**761-768.
101. Mayer K, Schuller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Dusterhoft A, Stiekema W, Entian KD, Terryn N, *et al.*: **Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.** *Nature* 1999, **402:**769-777.
102. Miller WJ, Nagel A, Bachmann J, Bachmann L: **Evolutionary dynamics of the *SGM* transposon family in the *Drosophila obscura* species group.** *Mol Biol Evol* 2000, **17:**1597-1609.
103. Kapitonov VV, Jurka J: **POGON1, a 'bona fide' family of nonautonomous DNA transposons.** *Repbase Repts* 2002, **2:**7.
104. Surzycki SA, Belknap WR: **Repetitive-DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes.** *Proc Natl Acad Sci USA* 2000, **97:**245-249.
105. Surzycki SA, Belknap WR: **Characterization of repetitive DNA elements in *Arabidopsis*.** *J Mol Evol* 1999, **48:**684-691.
106. Zhang X, Feschotte C, Zhang Q, Jiang N, Eggleston WB, Wessler SR: **P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases.** *Proc Natl Acad Sci USA* 2001, **98:**12572-12577.
107. Weiner AM: **SINEs and LINEs: the art of biting the hand that feeds you.** *Curr Opin Cell Biol* 2002, **14:**343-350.
108. Brizuela BJ, Elfring L, Ballard J, Tamkun JW, Kennison JA: **Genetic analysis of the *brahma* gene of *Drosophila melanogaster* and polytene chromosome subdivisions 72AB.** *Genetics* 1994, **137:**803-813.
109. Caggese C, Pimpinelli S, Barsanti P, Caizzi R: **The distribution of the transposable element *Bari-1* in the *Drosophila melanogaster* and *Drosophila simulans* genomes.** *Genetica* 1995, **96:**269-283.
110. Desset S, Conte C, Dimitri P, Calco V, Dastugue B, Vaury C: **Mobilization of two retroelements, *ZAM* and *Idefix*, in a novel unstable line of *Drosophila melanogaster*.** *Mol Biol Evol* 1999, **16:**54-66.
111. Leblanc P, Desset S, Dastugue B, Vaury C: **Invertebrate retroviruses: *ZAM* a new candidate in *D. melanogaster*.** *EMBO J* 1997, **16:**7521-7531.
112. Kim AI, Belyaeva ES, Aslanian MM: **Autonomous transposition of *gypsy* mobile elements and genetic instability in *Drosophila melanogaster*.** *Mol Gen Genet* 1990, **224:**303-308.
113. Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A: ***Flamenco*, a gene controlling the *gypsy* retrovirus of *Drosophila melanogaster*.** *Genetics* 1995, **139:**697-711.
114. Labrador M, Corces VG: **Interactions between transposable elements and the host genome**. In *Mobile DNA II*, Edited by Craig NL, Craigie R, Gellert M, Lambowitz AM. Washington, DC: ASM Press; 2002.
115. Labrador M, Corces VG: **Transposable element-host interactions: regulation of insertion and excision.** *Annu Rev Genet* 1997, **31:**381-404.
116. Biemont C, Cizeron G: **Distribution of transposable elements in *Drosophila* species.** *Genetica* 1999, **105:**43-62.
117. Streck RD, MacGaffey JE, Beckendorf SK: **The structure of *hobo* transposable elements and their insertion sites.** *EMBO J* 1986, **5:**3615-3623.
118. Crozatier M, Vaury C, Busseau I, Pelisson A, Bucheton A: **Structure and genomic organization of *I* elements involved in I-R hybrid dysgenesis in *Drosophila melanogaster*.** *Nucleic Acids Res* 1988, **16:**9199-9213.
119. Pasyukova EG, Nuzhdin SV: ***Doc* and *copia* instability in an isogenic *Drosophila melanogaster* stock.** *Mol Gen Genet* 1993, **240:**302-306.
120. Nuzhdin SV, Mackay TF: **The genomic rate of transposable element movement in *Drosophila melanogaster*.** *Mol Biol Evol* 1995, **12:**180-181.
121. Shevelyov YY: ***Aurora*, a non-mobile retrotransposon in *Drosophila melanogaster* heterochromatin.** *Mol Gen Genet* 1993, **239:**205-208.
122. Kanamori Y, Hayashi H, Yamamoto MT: **Molecular identification of the active ninja retrotransposon and the inactive aurora element in *Drosophila simulans* and *D. melanogaster*.** *Genes Genet Syst* 1998, **73:**385-396.
123. Medhora MM, MacPeek AH, Hartl DL: **Excision of the *Drosophila* transposable element *mariner*: identification and characterization of the *Mos* factor.** *EMBO J* 1988, **7:**2185-2189.
124. Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER: **Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*.** *Proc Natl Acad Sci USA* 1995, **92:**8050-8054.

125. **Berkeley *Drosophila* Genome Project: Release 2 Notes** [http://www.fruitfly.org/annot/release2.html]
126. **National Center for Biotechnology Information** [http://ncbi.nlm.nih.gov]
127. Carvalho AB, Vibranovski MD, Carlson J, Celniker S, Hoskins R, Rubin GM, Sutton GG, Adams MD, Myers EW, Clark AG: **Y chromosome and other heterochromatic sequences of the *Drosophila melanogaster* genome: how far can we go?** *Genetica* 2002, in press.
128. Heitz E: **Uber α- und β-Heterochromatin sowie Konstanz und Bau der chromomeren bei *Drosophila*.** *Biol Zentbl* 1934, **54:**588-609.
129. Painter TS, Muller HJ: **Parallel cytology and genetics of induced translocations and deletions in *Drosophila*.** *J Hered* 1929, **20:**287-298.
130. Miklos GL, Yamamoto MT, Davies J, Pirrotta V: **Microcloning reveals a high frequency of repetitive sequences characteristic of chromosome 4 and the β-heterochromatin of *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 1988, **85:**2051-2055.
131. Ashburner M, Misra S, Roote J, Lewis SE, Blazej R, Davis T, Doyle C, Galle R, George R, Harris N, *et al.*: **An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: the *Adh* region.** *Genetics* 1999, **153:**179-219.
132. **Berkeley *Drosophila* Genome Project: Natural transposon elements in Release 3 sequence** [http://www.fruitfly.org/p_disrupt/TE.html]
133. Tu Z, Shao H: **Intra- and inter-specific diversity of *Tc3*-like transposons in nematodes and insects and implications for their evolution and transposition.** *Gene* 2002, **282:**133-142.
134. **Berkeley *Drosophila* Genome Project** [http://www.fruitfly.org]
135. Galtier N, Gouy M, Gautier C: **SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny.** *Comput Appl Biosci* 1996, **12:**543-548.
136. **WU BLAST 2.0** [http://blast.wustl.edu/]
137. **Portable Batch System** [http://www.openpbs.org]
138. Bowen NJ, McDonald JF: **Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements.** *Genome Res* 1999, **9:**924-935.
139. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8:**195-202.
140. Morgenstern B: **DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment.** *Bioinformatics* 1999, **15:**211-218.
141. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
142. Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.** *J Mol Evol* 1980, **16:**111-120.
143. Felsenstein J: **Phylip.** 1993, Department of Genetics, University of Washington: Seattle, WA.
144. Liao GC, Rehm EJ, Rubin GM: **Insertion site preferences of the *P* transposable element in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2000, **97:**3347-3351.
145. Dominguez A, Albornoz J: **Rates of movement of transposable elements in *Drosophila melanogaster*.** *Molec Gen Genet* 1996, **251:**130-138.
146. Georgiev PG, Kiselev SL, Simonova OB, Gerasimova TI: **A novel transposition system in *Drosophila melanogaster* depending on the *Stalker* mobile genetic element.** *EMBO J* 1990, **9:**2037-2044.
147. Vieira C, Biemont C: **Selection against transposable elements in *D. simulans* and *D. melanogaster*.** *Genet Res* 1996, **68:**9-15.
148. Biemont C, Gautier C: **Localisation polymorphism of *mdg-1*, *copia*, *1* and *P* mobile elements in genomes of *Drosophila melanogaster*, from data of inbred lines.** *Heredity* 1988, **1988:**335-346.
149. Belyaeva ES, Ananiev EV, Gvozdev V: **Distribution of mobile dispersed genes (*mdg-1* and *mdg-3*) in the chromosomes of *Drosophila melanogaster*.** *Chromosoma* 1984, **90:**16-19.
150. Biemont C, Gautier C, Heizmann A: **Independent regulation of mobile element copy number in *Drosophila melanogaster* inbred lines.** *Chromosoma* 1988, **96:**219-294.
151. Whalen JH, Grigliatti TA: **Molecular characterization of a retrotransposon in *Drosophila melanogaster*, *nomad*, and its relationship to other retrovirus-like mobile elements.** *Mol Gen Genet* 1998, **260:**401-409.
152. Molto MD, Paricio N, Lopez-Preciado MA, Semeshin VF, Martinez-Sebastian MJ: ***Tirant*: a new retrotransposon-like element in *Drosophila melanogaster*.** *J Mol Evol* 1996, **42:**369-375.
153. Viggiano L, Caggese C, Barsanti P, Caizzi R: **Cloning and characterization of a copy of *Tirant* transposable element in *Drosophila melanogaster*.** *Gene* 1997, **197:**29-35.
154. Vaury C, Chaboissier MC, Drake ME, Lajoinie O, Dastugue B, Pelisson A: **The *Doc* transposable element in *Drosophila melanogaster* and *Drosophila simulans*: genomic distribution and transcription.** *Genetica* 1994, **93:**117-124.
155. Di Nocera PP, Digan ME, Dawid IB: **A family of oligo-adenylate-terminated transposable sequences in *Drosophila melanogaster*.** *J Mol Biol* 1983, **168:**715-727.
156. Kholodilov NG, Bolshakov VN, Blinov VM, Solovyov VV, Zhimulev IF: **Intercalary heterochromatin in *Drosophila*. III. Homology between DNA sequences from the Y chromosome, bases of polytene chromosome limbs, and chromosome *4* of *D. melanogaster*.** *Chromosoma* 1988, **97:**247-253.
157. Berghella L, Dimitri P: **The heterochromatic rolled gene of *Drosophila melanogaster* is extensively polytenized and transcriptionally active in the salivary gland chromocenter.** *Genetics* 1996, **144:**117-125.
158. Zabalou S, Alahiotis SN, Yannopoulos G: **A three-season comparative analysis of the chromosomal distribution of *P* and *hobo* mobile elements in a natural population of *Drosophila melanogaster*.** *Hereditas* 1994, **120:**127-140.
159. Harden N: **Characterization of the transposable element *FB-NOF* in *Drosophila melanogaster*.** In *Department of Genetics*. Cambridge: University of Cambridge; 1989: 178.
160. Merriman PJ, Grimes CD, Ambroziak J, Hackett DA, Skinner P, Simmons MJ: ***S elements*: a family of *Tc1*-like transposons in the genome of *Drosophila melanogaster*.** *Genetics* 1995, **141:**1425-1438.