Available online at www.sciencedirect.com
ScienceDirect
Gene 385 (2006) 96–102
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Retrogene movement within- and between-chromosomes in the evolution of *Drosophila* genomes ☆

Hongzheng Dai [a], Toshio F. Yoshimatsu [c], Manyuan Long [a,b,c,*]

[a] *Committee on Genetics, The University of Chicago, 1101 E 57th Street, Chicago, IL 60637, USA*
[b] *Department of Ecology and Evolution, The University of Chicago, 1101 E 57th Street, Chicago, IL 60637, USA*
[c] *Biological Sciences Collegiate Division, The University of Chicago, 924 E 57th Street, Chicago, IL 60637, USA*

Received 3 February 2006; received in revised form 24 April 2006; accepted 24 April 2006
Available online 23 September 2006

Abstract

Recent genomic analyses in *Drosophila* and mammals of inter-chromosomal retroposition have revealed that during evolution the retroposed genes that show male-biased expression tend to leave the X chromosome and opt for autosomal positions. Such a phenomenon may be a process of general, genomic and evolutionary relevance. It contributed to the unexpected overrepresentation of male-biased genes on the autosomes recently observed in microarray expression experiments. In this paper, we report our genomic analysis of within-chromosomal retroposition in *Drosophila melanogaster*, and compare it with the previously identified pattern of the between-chromosomal retroposition. We find that a surfeit of autosomal retroposed genes originated from parental genes located on the same chromosome, in contrast to the X chromosome in which only few genes retroposed in cis. Such an autosomal proximity effect implicates a role of the mutation process for retroposition in determining chromosomal locations of autosome-derived retroposed genes. Furthermore, this phenomenon supports the hypothesis that natural selection favors the retroposition of genes out of the X chromosome. Analyses of a large expression database for *D. melanogaster* genes revealed that the vast majority of the X-derived autosomal retroposed genes had evolved testis expression functions, consistent with other previous genomic analyses.

1. Introduction

What evolutionary forces determine the position of a gene in a genome? Our previous genomic analyses of retroposition between chromosomes have indicated that distributions of retrogenes in the genomes of *Drosophila* (Betrán et al., 2002) and mammals (Emerson et al., 2004) are not random with respect to the X chromosome and autosomes. One favorable criterion to study retroposed gene system is that the duplicate genes can be readily defined as a retroposition process in these organisms. Most parental genes contain introns while the newly evolved paralogous daughter genes lack introns due to their origin from spliced, non-primary gene transcripts (Long et al., 2003; Khil et al., 2005). Such feature allows us to characterize the direction of gene movement, where we can observe the migration pattern during the genome evolution *en masse*. Further, a comparison of the genomic distribution of functional retroposed genes and nonfunctional retropseudogenes may facilitate the detection of the role of natural selection rather than a mechanistic bias in the determination of gene positions (Betrán et al., 2004).

Our previous analysis revealed unexpected patterns of dominantly unidirectional gene movement from the *Drosophila* X chromosome to autosomes (Betrán et al., 2002). These results, however, were based on a relatively small sample size due to limited expression data available. Therefore, it appeared necessary to re-examine the previous conclusions using the upgraded datasets. In this study, we will investigate three issues using genomic data, especially the expression dataset, available in GenBank. First, given the fact that a large number of between-chromosomal retroposition events have been identified previously, we ask whether there is a comparable level of retroposition within chromosomes (i.e. the parental gene and the retroposed new genes are within the same chromosome). Second, if there exists a large number of within-chromosomal

☆ Presented at the 6th Anton Dohrn Workshop entitled "Evolutionary Genomics", Ischia, Italy, 30 October–3 November 2005.

* Corresponding author. Department of Ecology and Evolution, The University of Chicago, 1101 E 57th Street, Chicago, IL 60637, USA. Tel.: +1 773 702 0557; fax: +1 773 702 9740.
E-mail address: mlong@midway.uchicago.edu (M. Long).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.04.033

retroposition events, it would be important to know whether there are any patterns for such movement. Third, our previous analysis revealed that a vast majority of the X-derived autosomal retrogenes had evolved testis expression. This suggests that testes function is the target of evolutionary forces responsible for influencing the genomic locations of retrogenes. Hence, we will ask whether there is any expression pattern associated with the retrogenes evolved from the within-chromosomal retroposition.

Although the main interest of our previous analysis (Betrán et al., 2002) was inter-chromosomal retroposition, the preliminary and incomplete search of intra-chromosomal retroposition evinced few retroposed genes in the X chromosome. However, lack of within-chromosomal analysis of chromosomes 2 and 3 made the initial analysis inconclusive for the X chromosome. One cannot determine whether the under-represented within-X chromosomal retroposition is a consequence of natural selection, or is based on a biased function of a low within-X chromosomal retroposition rate. Such a bias could for example be a consequence of differential chromatin accessibility of the X chromosome relative to the autosomes. In this report, we investigated the patterns of intra-chromosomal retroposition, and the correlation between locations of retroposed genes and recombination rates. We found that the pattern of within-chromosomal retroposition was in support of previous conclusions based on the between-chromosomal retroposition.

2. Materials and methods

2.1. Identification of retroposed genes

Methods similar to those of Emerson et al. (2004) were employed to identify retroposed genes. The whole *Drosophila melanogaster* genome sequence data (Ensembl36 BDGP4) was obtained from Ensembl website (www.ensembl.org). The cDNA sequences of the *Drosophila* genome were aligned against all unspliced gene sequence of itself in an all-by-all comparison using BLAST (tblastx). In the initial screening, we rejected all alignments that aligned over less than 50% of both genes, and with amino acid identity of less than 50%. Subsequent screening on the output data was done to purge any occurrence of redundancy and to sort out the aligned pairs into different categories (i.e. a combination of a number of exons in the parental gene and the direction of gene movements by retroposition event). We then identified the base pair positions of the alignments in the sequence pairs using bl2seq. The purpose of the second alignment was three fold: (a) to identify the precise base pair positions of the alignments to obtain the conserved nucleotide sequences; (b) to reject alignments shorter than 150 bp or longer than 5000 bp; and (c) to determine whether the alignment occurs within the coding region. Because one of the hallmarks of the retroposed gene is that it is intron-less whereas the parental gene conserves introns, it was crucial to ascertain that all homolog sequences occur in the coding regions. To verify the splice signal sequences that define exon–intron structure, we manually inspected the aligned gene sequences against its reference sequence obtained from FlyBase (www.flybase.org). We also disregarded single exon genes that hit other single exon gene, as such pairs are not clearly retroposition events. Because chromosome 4 contains only 83 genes and contributes few retroposed genes, we neglected this chromosome in our analyses.

2.2. K_A and K_S estimation and K_A/K_S ratio test

To discern genes that are likely to be functional, we used a commonly accepted method where the likelihood ratio test is employed to determine whether K_A/K_S between the parental and retrogene pairs are smaller than 0.5 ($P<0.05$) (Emerson et al., 2004; Betrán et al., 2002). The rates between the number of non-synonymous substitution per site per time period (K_A) and the number of synonymous substitution per site per time period (K_S) for each gene pair contrast the two types of substitution events. A ratio significantly smaller than one is considered to indicate functional constraint in general. Depending on the selective constraint on the parental gene, however, K_A/K_S ratio smaller than unity but higher than 0.5 would be considered as expected ratio for divergence between functionless new retrogene duplicate and a functional parental gene (Li, 1997). The codeml program PAML 3.14b (Yang, 1997; http://abacus.gene.ucl.ac.uk/software/paml.html) was used to calculate the ratio. We performed two runs, one with one fixed to 0.5 and another estimating omega. The log likelihood value of the fixed omega model (l_0) was compared to the free model (l_i), and tested the statistical significance by comparing twice the log likelihood difference, $2\Delta l=2(l_i-l_0)$, to a χ^2 distribution with one degree of freedom (Yang, 1998). The K_A/K_S ratio was considered significantly smaller than 0.5 if the free model was significantly more likely than the fixed omega model.

2.3. Expected number of retropositions

We used the expectation formula developed by Betrán et al. (2002) to measure the expected number of retropositions. The expected frequency (P_{KL}), by which $P_{X\rightarrow A}$, $P_{A\rightarrow X}$, and $P_{A\rightarrow A}$ indicate the direction of the retroposition from parental gene to the new gene ($A\rightarrow A$ is bi-directional, i.e. includes $A_2\rightarrow A_3$ and $A_3\rightarrow A_2$), can be calculated using the equation

$$P_{KL}=\frac{\sum N_i L_j f_{ij}}{\sum\sum N_i L_j f_{ij}}$$

N_i and L_j are the proportions of gene number at the source chromosome i and the euchromatic size of the targeted chromosome, respectively, and f_{ij} is the frequency of occurrence of retroposition to a given chromosome in the population. According to genome data (Adams et al., 2000) and the existence of males and females in the population, i, j=X, 2 and 3, N_i=0.17, 0.38, 0.45; L_j=0.19, 0.36, 0.44 (chromosome 4 ignored for its minuscule size); and f_{ij}=0.75 for j=X and 1 for j=2 or 3; reflecting the relative population size of the X chromosome and autosomes. When i=j, the expectation within chromosomes is calculated.

2.4. Statistical analyses of the direction of distribution by retroposition

The filtered and sorted data of retroposed genes indicated the direction of movements (i.e. A→X, X→A, or A→A, by which A→A indicates bi-directionality and includes both self→self and self→other chromosomes). To test the pattern of gene movements, we compared the observations to the expectations using the formula

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where i is the direction of movement (i.e. A→X, X→A, A→A, or X→X), E_i is the expected number of retroposition, and O_i is the number of genes observed to deviate as indicated by i. Because previous work demonstrated movement of retroposition out of the X chromosome (Emerson et al., 2004; Betrán et al., 2002), we first analyzed the movement of genes to and out of the X chromosome (A→X and X→A) as well as retroposition between chromosomes 2 and 3 (A→A). We assume that the statistic X^2 follows a χ^2 distribution. Because our original interest was to investigate the overall pattern of movement of retroposition, we also analyzed the gene movement between and within chromosomes (i.e. self→self and self→other) using χ^2 test as denoted above, with one degree of freedom. In addition, we used Monte Carlo simulation to determine the probability of X^2 values. Some 1,000,000 simulated movements were tested, by which the proportion P of simulated χ^2 statistic that exceeded the observed χ^2 statistic was calculated. For example, of the 1,000,000 iterations, 21,593 cases were recognized to exceed the 46 self→self-retroposition movements experimentally observed, leading to P=21,593/1,000,000=0.021593. We found the two methods gave highly similar probabilities and the Monte Carlo simulation gave slightly lower probabilities for several tests we conducted in this study. We thus reported the more conserved χ^2 probabilities.

We note here that the number of pseudogenes in *Drosophila* genome is tremendously reduced compared to mammals (Petrov et al., 1996), not granting significant material for the comparison between functional retrogenes and the distribution of retro-pseudogenes, as we did in mammals previously (Emerson et al., 2004).

2.5. Expression patterns of retroposed genes

The UniGene system of NCBI (www.ncbi.nlm.nih.gov/entrez/unigene) was used to search for tissue expression information for identified retroposed genes and their parental genes. Our previous result from a small sample (22 retroposed genes and 22 parental genes) revealed that most of the X-derived autosomal genes (10 out of 11 X-derived autosomal genes) evolved testis expression patterns (Betrán et al., 2002). We took advantage of the expanded expression database (www.ncbi.nlm.nih.gov/unigene, January 2006) and generated larger retroposed gene sets to retest this pattern. This database collected all published expression data from ESTs, microarray tests, cDNA libraries from a number of tissues of *D. melanogaster*. We discarded the uninformative MI (mixed transcripts), although it shows that the gene is expressed.

3. Results

3.1. Between-chromosomal retroposition and gene movement out of the X chromosome

In a previous paper (Betrán et al., 2002), 24 retroposed genes with their parental genes were identified using criteria of 70% amino acid sequence identity. In this study, we decrease the identity criteria to 50% of amino acid sequence to test the distribution of gene movement with a larger sample size and longer evolution history. We identified 81 between- and 47 within-chromosomal retroposition events (128 events in total). Among the 81 between-chromosomal events, we observed 43 events deriving from the X chromosome to autosome, while only 10 events occurred originating from autosomes to the X chromosome (Table 1, Fig. 1A). Considering the gene number and the euchromatin length of each chromosome, following a similar random insertion model for expectation in a previous analysis (Betrán et al., 2002), we calculated the expected numbers of retroposition events for three directions: (i) X→A; (ii) A→X; (iii) A→A. Then, the statistical significance was tested using χ^2 test and Monte Carlo simulation. A highly significant biased distribution of retroposition events was observed (χ^2=39.12, df=2, $P=2\times10^{-8}$), which is consistent with previous results obtained using small sample size (Betrán et al., 2002) where retroposed genes demonstrate a tendency to move out of the X chromosome. A similar pattern has also been reported in mammalian genomes (Emerson et al., 2004).

3.2. Retroposition within chromosomes and a deficiency of X-linked retrogenes

Previous analysis indicated a scarceness of X-linked retroposed genes that originated from X-linked parental genes (Betrán et al., 2002). However, it is unclear whether this deficiency is a result of constraints such as chromatin compaction or other epigenetic effects that led to a lower retroposition rate within chromosomes, or a possible consequence of natural selection against insertion of retroposed genes into the X chromosome in general. The latter could have become manifested by inter-chromosomal retropositions, which revealed movement out of the X chromosome under some form of selection. We detected retroposition events within chromosomes and within chromosomal arms (Fig. 1B). The large

Table 1
Biased distribution of retrogenes between chromosomes

Direction	Observed numbers	Expected numbers	Excess (%)
A→A	28	46.6	−39.9
A→X	10	15.9	−37.1
X→A	43	18.5	132.4

χ^2=39.13, df=2, $P=2\times10^{-8}$.

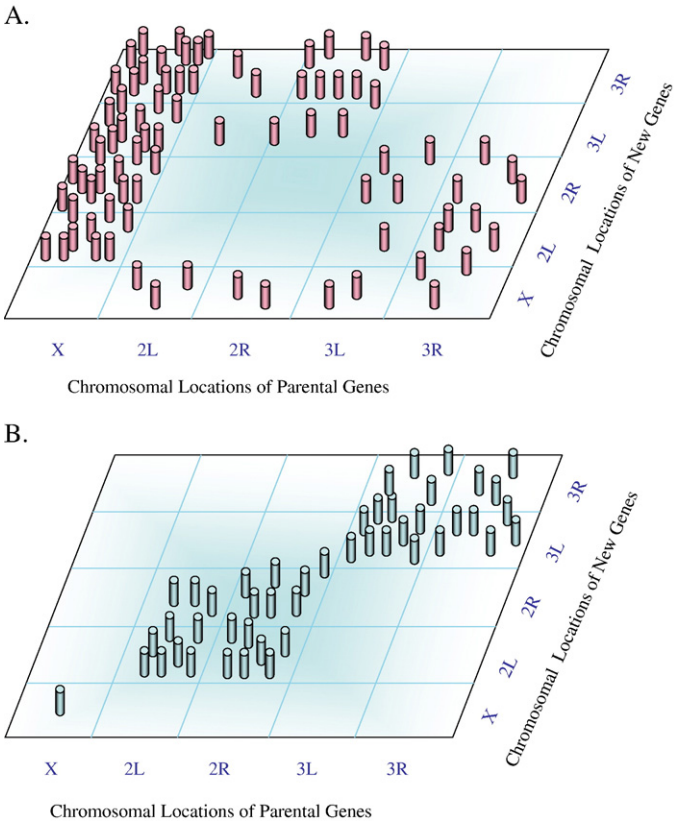

Fig. 1. Chromosomal locations of retroposed genes and their parental genes. In the two dimensions five arms of the three chromosomes (X, 2 and 3) are labeled to indicate the positions of retroposed genes (new genes) and their parental genes. A. The between-chromosomal analysis. B. The within-chromosomal analysis.

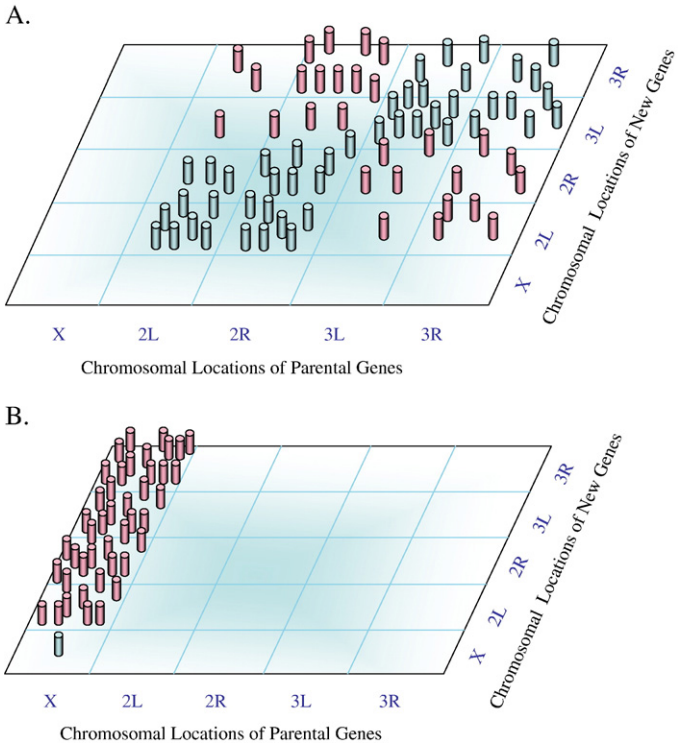

Fig. 2. A. Distribution of the within- and between-chromosomal retropositions in autosomes 2 and 3. B. Distribution of the X-derived retroposed genes.

sample obtained provides an opportunity to compare the distribution between- and within-chromosomes.

Using the criteria in this investigation (Sections 2.1 and 2.2), we identified only one retroposed gene on the X chromosome that was generated by a X-linked parental gene (Fig. 1B). This is consistent with the conclusion of a previous analysis (Betrán et al., 2002). Considering that the scarcity of X-linked retroposed genes is likely related to selection acting on the X chromosome, we separately analyzed the retroposition events that took place within chromosomes 2 and 3.

We identified altogether 46 retroposed genes that were generated by the parental genes in the same autosomes 2 and 3 (Fig. 1B, Table 2). 26 out of 46 pairs are within the same arm, while the remaining 20 pairs were between two arms in the same chromosome. Although statistically insignificant, there seems to exist a trend for retroposition to maintain association with the originating chromatin as shown by a higher frequency of retroposed events within the same arm rather than between arms movement. Moreover, this trend makes one more prediction — that within-chromosome retroposition may take place more frequently than between-chromosomes. In this search, we identified 46 pairs that are within the same chromosomes, while only 28 pairs are between chromosomes 2 and 3 (Fig. 2A, Table 3). Compared to the expected numbers (Table 3), the within-chromosomal retroposition frequency shows a 25% excess, whereas between-chromosomal retroposition frequency falls 24.7% short. However, the X chromosome differs entirely from this autosomal trend. The totaled proportion of retroposition frequency within- and between-autosomes 2 and 3 is 62% (46/74) and 37.8% (28/74) respectively. This gives a prediction that among 44 X-derived retroposed genes (43+1, Fig. 2B), 27 should have been derived from X→X retroposition and 17 X→A retroposition. These predicted numbers are significantly different from the observed 1 X→X event versus 43 X→A events ($\chi^2=64.84$, $df=1$; $P<10^{-6}$). Thus, the scarceness of the X-derived and X-linked retroposed genes is inconsistent with the hypothesis of lower within-chromosomal retroposition rate, supporting the previous

Table 2
Biased distribution of retrogenes within autosomes: comparison of chromosome arms

	Observed numbers	Expected numbers	Excess (%)
Within-arm	26	22.4	16.1
Different arms	20	23.6	−15.3

$\chi^2=2.52$, $df=1$, $P=0.1124$.

Table 3
Biased distribution of retrogenes within autosomes: comparison of whole chromosomes

	Observed numbers	Expected numbers	Excess (%)
Within chromosome	46	36.8	25.0
Different chromosomes	28	37.2	−24.7

$\chi^2=4.56$, $df=1$, $P=0.032$.

conclusion that natural selection is likely the evolutionary force acting against retroposed genes on the X chromosome.

3.3. Vast majority of X-derived autosomal retrogenes are expressed in adult male testes

Based on the above study of both inter- and intra-chromosomal retroposition rates, we rediscovered a consistent pattern that a high excess of X-derived retroposed genes migrated to autosomes. The X chromosome has not only a deficiency of retrogenes deriving from the autosome-linked parental genes, but also a dearth of retrogenes originating from the same X chromosome. From the UniGene dataset, we identified 29 retroposed genes with expression data that originated from the parental genes in different chromosomes (Table 4). Among them, we found that 72.4% (21 out of 29) of X-derived autosomal genes were expressed in adult male testes, suggesting that the vast majority of X-derived autosomal retrogenes have evolved testis specific expression patterns. In particular, 24% of these X-derived autosomal retrogenes evolved testis expression that was not detected in parental genes. Furthermore, we also observed that the new genes seem to be expressed in fewer tissues or organs (2.0 tissues per retroposed gene) in comparison with the parental genes (3.6 tissues per retroposed gene), suggesting that new retroposed genes have more specific functions involved in male reproduction. The proportions of the retroposed genes involving

Table 4
Identified X→A retroposition events and expression patterns

Parental genes				Retroposed genes				Identity
Gene ID	Exon	Chr	Expression	Gene ID	Exon	Chr	Expression	
CG2998	2	X	H, E, SG, T, HE, FB	CG15527	1	3R	UN	74.58
CG8636	2	X	H, HE, E,	CG10881	1	3R	H	69.92
CG9172	3	X	T, H, HE, E	CG2014	1	3R	UN	89.6
CG3560	3	X	H, HE, B,	CG17856	1	3R	UN	89.16
CG32672	3	X	T, H, E, SG, HE, FB	CG12334	1	3R	T, FB	82.76
CG3422	3	X	E	CG17268	1	3R	T	77.59
CG3774	5	X	E, H	CG14511	1	3R	UN	56.85
CG2915	5	X	T, E, H, SG, FB	CG4408	1	3R	E, FB, H	64.92
CG14206	6	X	T, E, H, SG, FB, HE	CG12275	1	3R	E	79.59
CG2076	6	X	T, E, H, HE	CG1287	1	3R	T, H	70.92
CG9032	7	X	E, H, HE, FB	CG31477	1	3R	UN	76.92
CG32684	48	X	T, E, H, B, HE	CG31202	1	3R	H	51.71
CG2621	110	X	T, E, H, HE	CG31003	1	3R	T, H	71.76
CG15645	3	X	T, FB	CG13732	1	3L	T, H, FB, E	74.59
CG32581	3	X	H, E	CG32847	1	3L	T, E, FB	72.31
CG1633	4	X	E, H, HE	CG6888	1	3L	T	62.32
CG8918	4	X	H	CG32238	1	3L	UN	61.18
CG2025	4	X	H, E	CG10588	1	3L	UN	56.83
CG5703	4	X	T, E, H	CG6485	1	3L	T	55.61
CG12410	5	X	HE, E,	CG11582	1	3L	UN	77.97
CG12530	5	X	HE, E, H, FB, SG	CG8556	1	3L	HE, E, SG	70.95
CG1404	5	X	T, HE, H, E, FB	CG7815	1	3L	T, H, FB	58.69
CG12359	6	X	T, H, E	CG32110	1	3L	T, H, E	56.77
CG14214	2	X	H, HE, E, FB	CG8860	1	2R	T	98.55
CG14816	2	X	T, E, H	CG15874	1	2R	UN	61
CG4575	2	X	UN	CG7786	1	2R	UN	51.72
CG8893	3	X	T, H, HE, E, FB, SG	CG12055	1	2R	T, H, E	95.86
CG9091	3	X	T, H, HE, E, FB, SG	CG9873	1	2R	UN	77.01
CG17437	3	X	H, HE, E	CG10931	1	2R	UN	58.9
CG8310	4	X	UN	CG8186	1	2R	T, H, HE, E, SG	88.29
CG1696	4	X	T, H, E	CG8584	1	2R	T	58.27
CG12157	9	X	H, E	CG8330	1	2R	T, E, FB	75.96
CG2033	10	X	T, H, HE, E, FB, SG	CG12324	1	2R	E	96.1
CG2096	11	X	H, HE, E	CG10930	1	2R	T	65.34
CG9360	2	X	H, HE, E, SG	CG9150	1	2L	HE	58.33
CG5254	2	X	E, SG	CG9582	1	2L	T	58.16
CG1740	4	X	HE, H, E, SG	CG10174	1	2L	UN	85.61
CG8931	4	X	T, H, E, FB	CG5755	1	2L	T, H, FB	69.86
CG2713	5	X	T, HE, H, E	CG6691	1	2L	T	64.63
CG12101	6	X	T, H, HE, E	CG2830	1	2L	T, H,	66.36
CG14222	6	X	E	CG31730	1	2L	UN	55.88
CG4199	16	X	T, E, HE, H	CG10700	1	2L	E	59.39
CG2694	17	X	T, E, H	CG11322	1	2L	T, H	66.52

Note: E: Embryo; H: Head; T: Testis; FB: Fat body; UN: Unknown; HE: Hemocyte; SG: Salivary gland; SO: Sensory organ; Chr: chromosome; Identity: protein sequence identity.

between- or within autosomes that are expressed in adult male testis are lower (58.3% between-chromosomal retroposed genes and 62% within-chromosomal retroposed genes).

4. Discussion

Our present observations are based on a significantly increased dataset of *D. melanogaster*. The results add strong support to our previous conclusion made from a relatively small sample size, by which excess of retroposed genes diverged from the X-linked parental genes. These retroposed genes favored autosomal locations, and mostly evolved testis expression (Betrán et al., 2002). One may wonder what underlying mechanisms would be responsible for the excess of retroposed genes moving to autosomes and why they evolved testis expression. Further, what is the consequence of such a process for the chromosomal distribution of male-specific genes?

There are a number of hypothetical mechanisms to account for the biased distribution (see Betrán et al., 2002, 2004; Emerson et al., 2004). In general, mutation based hypotheses, e.g. different expression levels between X-linked and autosomal genes, are not able to explain the observed excess of retrogenes that move to autosomes. However, several forms of selection-based hypotheses are potentially able to account for the observations. These include the hypothetical germline X-inactivation in *Drosophila* (Lifschytz and Lindsley, 1972), the Wu–Xu model of the sexual antagonism hypothesis (Wu and Xu, 2003), and the dominant advantageous mutation hypothesis (Charlesworth et al., 1987). However, none of these hypotheses can account for all phenomena of male expression genes. For example, male germline X-inactivation cannot explain the observed excess of male genes in accessory glands in autosomes (Swanson et al., 2001), although this hypothesis may explain the excess of autosomal retroposed genes that are expressed in adult male testis. Male germline X-inactivation in *Drosophila* was inferred from genetic analysis and cytological observations (reviewed by Lifschytz and Lindsley, 1972, 1974). Cytologically, precocious condensation of the sex chromosomes in *D. melanogaster* was mentioned (Lifschytz and Lindsley, 1972) and genetical analyses (Lindsley, 1965) showed that 75% of translocations between the X chromosome and autosomes are male sterile, which was thought to be a consequence of the interference of the hypothetical X-inactivation process. However, these are not direct evidence, which has yet to be found.

However, mammalian meiotic sex chromosome inactivation (MSCI) has been directly studied in detail (Richler et al., 1992; Ayoub et al., 1997; Huynh and Lee, 2005). MSCI has been known to play a role in determination of the male-biased genes in mammals (Reinke, 2004; Betrán et al., 2002; Emerson et al., 2004). Nevertheless, genomic analyses and experiments showed MSCI also cannot explain the autosomal distribution of all male genes. For example, in a recent study Wang et al. (2005) observed some degree of postmeiotic reactivation, during which mammalian male genes were also observed to favor autosomal positions (Emerson et al., 2004). Therefore, it is likely that there are multiple genetic principles responsible for excess autosomal distribution of male genes.

Compared to the uncertainty in identifying the underlying mechanisms, the consequence of the observed phenomena is more obvious. In this study, the dataset included not only those relatively young retroposed genes but also evolutionarily more ancient genes that suggest the same biased distribution. Both datasets, i.e. young retro-paralogs as well as evolutionary older retro-paralogs, sustain a process in which retrogenes with parent genes on the X chromosome left the X to become preferentially accommodated on autosomes. This fact indicates that the positive selection for the genes escaping out of the X chromosome is ancient and continued over a long evolutionary period. 38 of the 43 X→A gene pairs have a $K_S > 1.0$ and many of such retroposed genes have orthologs detectable in the genome sequences of *D. virilis* and/or *D. pseudoobscura*, suggesting their ages older than 46 mys or 60–65 mys (Powell, 1997). For example, retroposed genes CG12334, CG8556, CG8186, and CG9150 have orthologs in both *D. pseudoobscura* and *D. virilis*, indicating ages older than 60–65 mys. The existence of these ancient retroposed genes indicates an extended evolutionary time of gene movement, suggesting a prevailing mechanism to enrich genes of testis-biased expression on autosomes during the evolution of the *Drosophila* genome. This mechanism does not support a general notion that the X chromosome is a hot bed for male genes (Bainbridge, 2003). Conclusions similar to ours were reached in *C. elegans* as well (Reinke, 2002). This enriching process may have contributed to the recently observed high proportion of male genes on autosomes in *Drosophila* (Ranz et al., 2003; Parisi et al., 2003). A similar enriching process may also be a factor for the location of mammalian spermatogenesis-related genes that are relatively under-represented on the X chromosome (Khil et al., 2004, 2005; Emerson et al., 2004; Marques et al., 2005).

Acknowledgements

We would like to thank all members in M. L.'s laboratory for valuable discussion and Esther Betrán of the University of Texas at Arlington for critical reading of the manuscript. This work was supported by a National Science Foundation CAREER award and National Institutes of Health R1 grant in USA to M.L.

References

Adams, M.D., et al., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Ayoub, N., Richler, C., Wahrman, J., 1997. Xist RNA is associated with the transcriptionally inactive XY body in mammalian male meiosis. Chromosoma 106, 1–10.

Betrán, E., Thornton, K., Long, M., 2002. Retroposed new genes out of *Drosophila*. Genome Res. 12, 1854–1859.

Betrán, E., Emerson, J.J., Kaessmann, H., Long, M., 2004. Sex chromosomes and male functions: where do new genes go? Cell Cycle 3, 873–875.

Bainbridge, D., 2003. The X in Sex: how the X Chromosome Controls our Lives. Harvard University Press, Boston.

Charlesworth, B., Coyne, J.A., Barton, N.H., 1987. The relative rate of evolution of sex chromosomes and autosomes. Am. Nat. 130, 113–146.

Emerson, J.J., Kaessmann, H., Betrán, E., Long, M., 2004. Extensive gene traffic on the mammalian X chromosome. Science 303, 537–540.

Huynh, K.D., Lee, J.T., 2005. X-chromosome inactivation: a hypothesis linking ontogeny and phylogeny. Nat. Rev. Genet. 6, 410–418.

Khil, P.P., Smirnova, N.A., Romanienko, P.J., Camerini-Otero, R.D., 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. Nat. Genet. 36, 642–646.

Khil, P.P., Oliver, B., Camerini-Otero, R.D., 2005. X for intersection: retrotransposition both on and off the X chromosome is more frequent. Trends Genet. 21, 3–7.

Li, W.-H., 1997. Molecular Evolution. Sinauer Associates Sunderland, MA.

Lifschytz, E., Lindsley, D.L., 1972. The role of X-chromosome inactivation during spermatogenesis (*Drosophila*-allocycly-chromosome evolution-male sterility dosage compensation). Proc. Natl. Acad. Sci. U. S. A. 69, 182–186.

Lifschytz, E., Lindsley, D.L., 1974. Sex chromosome activation during spermatogenesis. Genetics 78, 23–331.

Lindsley, D.L., 1965. Chromosomal function at the supragenic level. In: Valencia, J.I., Grell, R.F. (Eds.), Genes and Chromosomes—Structure and Function. Nat. Cancer Inst. Monogra, vol. 18, pp. 275–290.

Long, M., Betran, E., Thornton, K., Wang, W., 2003. The origin of new genes: glimpses from the young and old. Nat. Rev., Genet. 4, 865–875.

Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A., Kaessmann, H., 2005. Emergence of young human genes after a burst of retroposition in primates. PLoS Biol. 3, e357.

Parisi, M., et al., 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. Science 299, 697–700.

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

Powell, J.R., 1997. Progress and Prospects in Evolutionary Biology: the *Drosophila* Model. Oxford University Press, Oxford.

Ranz, J.M., Castillo-Davis, C.I., Meiklejohn, C.D., Hartl, D.L., 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science 300, 1742–1745.

Reinke, V., 2002. Functional exploration of the *C. elegans* genome using DNA microarrays. Nat. Genet. 32, 541–546 (Suppl).

Reinke, V., 2004. Sex and the genome. Nat. Genet. 36, 548–549.

Richler, C., Soreq, H., Wahrman, J., 1992. X inactivation in mammalian testis is correlated with inactive X-specific transcription. Nat. Genet. 2, 192–195.

Swanson, W.J., Clark, A.G., Waldrip-Dail, H.M., Wolfner, H.M., Aquadro, M.F., 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc. Natl. Acad. Sci. U. S. A. 98, 7375–7379.

Wang, P.J., Page, D.C., McCarrey, J.R., 2005. Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. Hum. Mol. Genet. 14, 2911–2918.

Wu, C.-I., Xu, E.Y., 2003. Sexual antagonism and X inactivation — the SAXI hypothesis. Trends Genet. 19, 243–247.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comp. Appl. BioSci. 13, 555–556.

Yang, Z., 1998. Likelihood ratio test for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15, 568–573.