Letter to the Editor

Genome Size and Intron Size in *Drosophila*

Etsuko N. Moriyama, Dmitri A. Petrov,† and Daniel L. Hartl‡*

*Department of Ecology and Evolutionary Biology, Yale University; †Harvard University Society of Fellows; and
‡Department of Organismic and Evolutionary Biology, Harvard University

Petrov, Lozovskaya, and Hartl (1996) demonstrated that unconstrained regions of the non-long terminal repeat retrotransposable element *Helena* lose DNA at an unusually high rate in species of the *Drosophila virilis* species group. More recent data also indicate a high rate of DNA loss of *Helena* in species of the *Drosophila melanogaster* species subgroup (Petrov and Hartl 1998). Based on these observations, the authors suggested that the paucity of pseudogenes in *Drosophila* is the product of rampant deletion of DNA in regions not subjected to selective constraints, and they further extrapolated that different deletion rates may contribute to the divergence in genome size among taxa.

Their assumption is that such a high rate of deletion is not confined to *Helena* elements alone. The sizes of any unconstrained regions, such as introns and other noncoding regions, would also be decreased to the extent that selection allows, and consequently the genome size would be reduced. Supporting evidence was obtained from vertebrate genes (Hughes and Hughes 1995; Ogata, Fujibuchi, and Kanehisa 1996). In accordance with the difference in genome size, human genes have significantly longer introns than do avian or rodent homologs.

The genome size of *D. virilis* is 0.34–0.38 pg per haploid genome, while those of *D. melanogaster* and *Drosophila pseudoobscura* are 0.18–0.21 pg (Powell 1997). Even taking into account the different proportions of the genome devoted to pericentromeric heterochromatin, the genome of *D. virilis* is considerably larger than that of *D. melanogaster* and *D. pseudoobscura* (Hartl and Lozovskaya 1995). The difference in the size of the euchromatic genome between *D. virilis* and *D. melanogaster* is about 36% (150 Mb vs. 110 Mb). If differences in genome size of such magnitude are due to different rates of accumulation of small deletions and insertions throughout the euchromatic genome, then we can predict that *D. virilis* genes should have longer introns than those of *D. melanogaster* and *D. pseudoobscura.*

We compared the lengths of 115 complete introns collected from 42 homologous genes between *D. melanogaster* and *D. virilis,* and 60 introns from 22 homologous genes between *D. melanogaster* and *D. pseudoobscura* (11 genes are common in the three species). Figure 1 shows the comparisons of intron lengths among these species. *Drosophila virilis* genes tend to have much longer introns than do those of *D. melanogaster* (fig. 1*a*), while *D. pseudoobscura* genes do not show such a clear difference (fig. 1*b*). The difference in intron length between *D. melanogaster* and *D. virilis* is significant at the 1% level ($P = 0.002$, randomization test for paired comparisons, table 1). The mean (and median) intron lengths for the two species are 283 (79) and 394 (82) bp, respectively. The difference in the means is 39%, which is in surprisingly good agreement with the size difference of the two euchromatic genomes mentioned above (36%). No significant difference in intron length was observed between *D. virilis* and *D. pseudoobscura*, probably due to the small sample size (11 genes, 25 introns).

Mount et al. (1992) classified *Drosophila* introns by their lengths: ‘‘short introns’’ (80 bp or shorter) and ‘‘long introns’’ (longer than 80 bp). The splicing mechanisms of short introns are considered to be different from those of long introns (Mount et al. 1992; Mount 1993). When we examined the lengths of these two groups of introns separately, different patterns of intron length variation were found (table 1). Between *D. melanogaster* and *D. virilis,* long introns are significantly longer in *D. virilis* (the mean difference is 196 bp; $P = 0.011$), whereas short introns are significantly longer in *D. melanogaster* (although the mean difference is only 2 bp; $P = 0.032$). On the other hand, *D. pseudoobscura* has significantly longer short introns than *D. melanogaster* ($P = 0.006$). Although the sample size for the comparison with *D. pseudoobscura* is small, the nonparametric Wilcoxon signed-ranks test also showed a significant difference in short-intron length ($P = 0.011$). Most of the changes in intron length are observed within the same length group. When the introns compared are short in one species and long in the other (the category designated ‘‘between’’ in table 1), in 14 cases out of 23, *D. virilis* had the long intron.

In addition to the size changes, there were nine cases of intron loss (or gain), eight of them from the comparison with *D. virilis.* Because the mechanism of intron loss/gain may differ from that of indels within introns, these nine cases were not included in table 1. All of the losses/gains were found for short introns, with only one exception for *trithorax* (a 203-bp intron is found in the *D. virilis* gene but not in the *D. melanogaster* gene). For six cases out of nine, *D. melanogaster* genes have no introns at the corresponding sites. Curiously, the other three cases, in which *D. melanogaster* has the corresponding introns whereas *D. virilis* does not, were all found in the same gene, *brown.*

These results imply that the mechanisms leading to the difference in long-intron size may have contributed

Key words: *Drosophila melanogaster, Drosophila pseudoobscura, Drosophila virilis,* intron length, genome size.

Address for correspondence and reprints: Etsuko N. Moriyama, Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, Connecticut 06520-8106. E-mail: moriyama@peaplant.biology.yale.edu.

Mol. Biol. Evol. 15(6):770–773. 1998

to the difference in overall genome size between *D. virilis* and *D. melanogaster.* Short introns in *D. melanogaster* are predominantly shorter than those of *D. pseudoobscura,* whereas *D. melanogaster* has slightly longer short introns than even *D. virilis.* Such small-scale changes in short introns do not seem to be correlated with any major difference in genome size. The mechanisms leading to the origin and fixation of indels appear to be different among the *Drosophila* lineages and/or between long and short introns.

The average size of deletions in *Helena* elements is approximately 25 bp, although approximately half of the deletions are in the size range 1–10 bp (Petrov, Lozovskaya, and Hartl 1996; Petrov and Hartl 1998). In contrast, the mean length difference in short introns is very short (2 or 4 bp depending on the species compared; table 1), which makes it seem likely that fixed differences in short-intron length are strongly skewed by selective constraints. (The number and size distribution of deletions in long introns cannot be ascertained, because the introns are too divergent in sequence to be aligned.)

Akashi (1996) found that protein size of *D. melanogaster* is larger than that of its sibling species, *D. simulans,* whereas introns (both short and long) do not show differences in length between the two species. Natural selection seems to be responsible for the difference in protein length. Relaxed selective constraints in the *D. melanogaster* genome compared with *D. simulans* appear to allow disadvantageous longer proteins to persist (Akashi 1996). We compared the lengths of 54 genes between *D. melanogaster* and *D. virilis.* The coding sequences of 20 genes were longer in *D. melanogaster,* and 26 were longer in *D. virilis* (randomization test for paired comparisons; $P = 0.09$). On the other hand, of the 30 genes compared between *D. melanogaster* and *D. pseudoobscura,* only 6 had longer coding sequences in *D. melanogaster,* and 16 were longer in *D. pseu-*

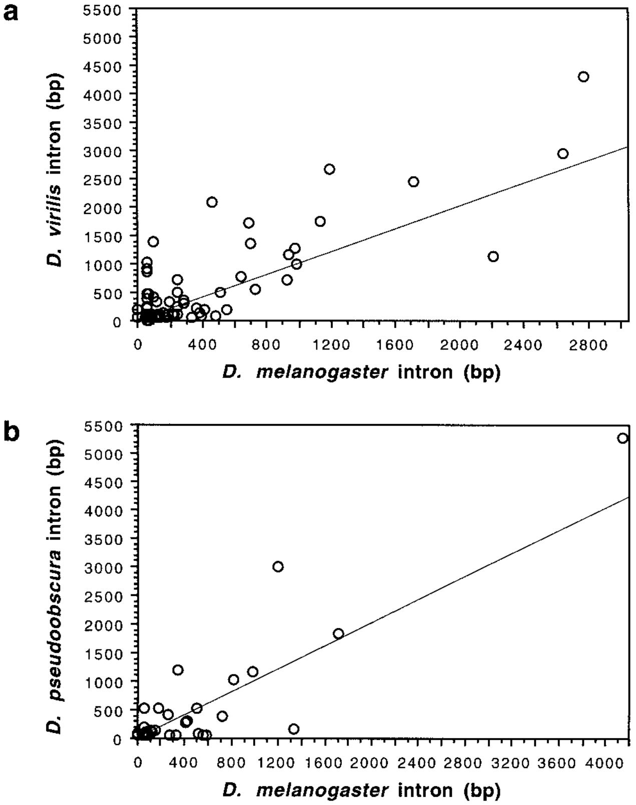

FIG. 1.—Comparisons of intron lengths between *Drosophila melanogaster* and *Drosophila virilis* (*a,* 115 introns) and between *D. melanogaster* and *Drosophila pseudoobscura* (*b,* 60 introns). The diagonal line is for reference only, to show where the intron lengths are identical between the two species. Only nonoverlapped complete introns and those located within coding regions were used in the analyses. GenBank accession numbers are available on request. Species abbreviations: Dm (*D. melanogaster*), Dp (*D. pseudoobscura*), and Dv (*D. virilis*).

Table 1
Comparisons of Intron Lengths Between *Drosophila melanogaster, Drosophila pseudoobscura,* and *Drosophila virilis*

COMPARISON[a]	NUMBER OF CASES[b]		MEAN DIFF.[c]	
(sp. 1 vs. sp. 2)	sp. 1 > sp. 2	sp. 1 < sp. 2	(sp. 1 − sp. 2)	*P* VALUE[d]
Dm vs. Dv [total; 107].	50	51	−111.0	0.002
Dm vs. Dv [short; 42]	22	15	2.4	0.032
Dm vs. Dv [long; 42].	19	22	−196.0	0.011
Dm vs. Dv [between; 23]	9	14	−162.7	0.020
Dm vs. Dp [total; 58].	21	36	−31.9	0.263
Dm vs. Dp [short; 28]	6	21	−4.0	0.006
Dm vs. Dp [long; 17].	7	10	−156.6	0.174
Dm vs. Dp [between; 13]	8	5	70.9	0.166

[a] Lengths of homologous introns were compared between species 1 (sp. 1) and 2 (sp. 2). ''Total'' data set includes all of homologous introns for the two species. The ''short'' data set includes only introns 80 bp or shorter for both species, and the ''long'' data set includes those introns longer than 80 bp for both species. When two species have introns in different length categories, the comparisons are presented as ''between.'' The number of introns compared is shown in brackets. Species abbreviations are Dm (*D. melanogaster*), Dp (*D. pseudoobscura*), and Dv (*D. virilis*).

[b] Number of cases in which species 1 has longer introns than species 2 (sp. 1 > sp. 2) or vice versa (sp. 1 < sp. 2).

[c] The mean difference of the intron lengths (bp) between species 1 and species 2.

[d] Randomization test for the paired comparisons. In each test, species names (sp. 1 and sp. 2) were assigned at random to each pair of introns, and then the distribution of the mean differences was obtained from 5,000 random replications. Virtually the same results were obtained by the parametric paired *t*-test.

doobscura (randomization test for paired comparisons; $P = 0.03$). This significant difference is consistent with that observed in short-intron length between these two species. Can both results, longer short introns and longer coding sequences in *D. pseudoobscura* genes, be explained by possible weaker intensity of natural selection in this species? This possibility seems to be contradicted by other evidence. For example, the average level of DNA polymorphism, in either coding or noncoding regions, is more than twofold higher in *D. pseudoobscura* than in *D. melanogaster*, and there are many more replacement polymorphisms in *D. melanogaster* (Moriyama and Powell 1996). These observations suggest relative inefficiency of natural selection in *D. melanogaster*, most likely due to a reduction in effective population size.

Charlesworth (1996) has argued that diverse genome sizes may result not only from differences in deletion rates, but also from differences in selective constraints. Natural selection may operate at the level of the enzymes involved in DNA synthesis and repair that govern the creation and size distribution of deletions and insertions. Natural selection may also operate at the level of the individual DNA sequences themselves, provided that a small deletion produces a large enough selection coefficient to overcome the effects of random genetic drift. While selection at some level is undoubtedly an important factor in the evolution of genome size, there is as yet no evidence to suggest that particular deletions are selectively advantageous in and of themselves because they result in a smaller genome. If this were the case, one would expect to observe a correlation between the age of a sequence and the aggregate size of the deletions within it, because a shorter sequence should persist longer if a shorter sequence is selectively advantageous. The data from *Helena* in both the *D. virilis* species group and the *D. melanogaster* species subgroup show no such correlation. There is a correlation between the age of a *Helena* sequence (as assessed by the number of nucleotide substitutions in it) and the number of deletions, but there is not a correlation between age and aggregate deletion size (Petrov and Hartl 1998).

The length difference in long introns appears to be consistent with the difference in genome size between *D. melanogaster* and *D. virilis.* It is not clear whether *D. melanogaster* has accumulated more deletions or *D. virilis* has accumulated more insertions. More insertions have been observed in *Helena* elements in *D. melanogaster* than in *D. virilis*, although the difference is not statistically significant (Petrov and Hartl 1998). In both species, the rate and size of small deletions far exceed those of small insertions. But if there are so many deletions relative to insertions, and the deletions are, on average, larger, then why is each *Drosophila* genome not stripped to its minimal size? The answer would seem to be that the deletions must be counteracted by a relatively small number of large insertions. (Large insertions would not have been detected in the *Helena* data, given the manner in which the sequences were ascertained.) Furthermore, large insertions are less likely to be severely detrimental than are large deletions, because any deletion, if large enough, will probably eliminate important DNA sequences. This sets an effective upper limit to the size of a deletion that can become fixed, but not to the size of an insertion. The implication for genome evolution in *Drosophila* is that a large number of relatively small deletions may be offset by a small number of relatively large insertions.

One obvious potential source of relatively large insertions is the movement of transposable elements. To look for known transposable elements or their remnants in our intron data, BLAST homology searches were conducted on 50 intron sequences longer than 500 bp (including 20 *D. melanogaster* and 21 *D. virilis* sequences) against the nonredundant nucleotide database (http://www.ncbi.nlm.nih.gov/BLAST/). We found only one instance: a 2,078-bp intron of *D. virilis sevenless* has homology in a short region with an IS*Y3* insertion sequence (L13721; Steinemann and Steinemann 1993). The low number of matches does not necessarily imply that transposable elements do not constitute part of the long introns. Possible rapid degeneration of such sequences through nucleotide substitution and the accumulation of small indels may have precluded their detection through BLAST searches.

In any case, the larger sizes of long introns in *D. virilis* than in *D. melanogaster* suggest that the mechanisms governing the increase or decrease in size of DNA sequences operate more or less uniformly throughout the euchromatin and affect single-copy DNA in long introns as well as repetitive sequences like *Helena.*

Acknowledgments

We thank J. R. Powell for his comments on the manuscript. This work was supported in part by NIH grants GM33741 and HG01250 to D.L.H.

LITERATURE CITED

AKASHI, H. 1996. Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster.* Genetics **144**:1297–1307.

CHARLESWORTH, B. 1996. The changing sizes of genes. Nature **384**:315–316.

HARTL, D. L., and E. R. LOZOVSKAYA. 1995. The *Drosophila* genome map: a practical guide. R. G. Landes, Austin, Tex.

HUGHES, A. L., and M. K. HUGHES. 1995. Small genomes for better flyers. Nature **377**:391.

MORIYAMA, E. N., and J. R. POWELL. 1996. Intraspecific nuclear DNA variation in *Drosophila.* Mol. Biol. Evol. **13**: 261–277.

MOUNT, S. M. 1993. Messenger RNA splicing signals in *Drosophila* genes. Pp. 333–358 *in* G. MARONI, ed. An atlas of *Drosophila* genes: sequences and molecular features. Oxford University Press, New York.

MOUNT, S. M., C. BURKS, G. HERTZ, G. D. STORMO, O. WHITE, and C. FIELDS. 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Nucleic Acids Res. **20**:4255–4262.

OGATA, H., W. FUJIBUCHI, and M. KANEHISA. 1996. The size differences among mammalian introns are due to the accumulation of small deletions. FEBS Lett. **390**:99–103.

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

PETROV, D. A., E. R. LOZOVSKAYA, and D. L. HARTL. 1996. High intrinsic rate of DNA loss in *Drosophila*. Nature **384**: 346–349.

POWELL, J. R. 1997. Progress and prospects in evolutionary biology: the *Drosophila* model. Oxford University Press, New York.

STEINEMANN, M., and S. STEINEMANN. 1993. A duplication including the Y allele of *Lcp2* and the *TRIM* retrotransposon at the *Lcp* locus on the degenerating *neo-Y* chromosome of *Drosophila miranda*: molecular structure and mechanisms by which it may have arisen. Genetics **134**:497–505.

CHARLES F. AQUADRO, reviewing editor

Accepted March 4, 1998