Letter to the Editor

Protein Variation in *Drosophila simulans*, and Comparison of Genes from Centromeric Versus Noncentromeric Regions of Chromosome 3

David J. Begun

Section of Evolution and Ecology, University of California–Davis

Several studies of *Drosophila melanogaster* have reported patterns of DNA variation in genes located near centromeres or near the telomere of the X chromosome (Aguadé, Miyashita, and Langley 1989; Langley et al. 1993, 2000; Wayne and Kreitman 1996). There is good evidence of reduced crossing over for these regions in *D. melanogaster*. However, there are relatively few population data for such regions in *Drosophila simulans* (Begun and Aquadro 1991; Martin-Campos et al. 1992; Hilton, Kliman, and Hey 1994; Wayne and Kreitman 1996). Here I report data from population samples of four genes located near the centromere of chromosome 3 in *D. simulans*: *Hem-protein, CKII-α, Gelsolin,* and *Amalgam*. The physical locations of these genes in *D. melanogaster* are 79E2, 80A, 82A1–3, and 84A5, respectively. Because there are no detectable chromosomal rearrangements between species in these regions (Ashburner 1989), I assume that the locations of these genes are the same in *D. simulans*. The sequences used were from a set of inbred lines derived from flies collected in the Wolfskill Orchard in Winters, Calif. (Begun and Whitley 2000). I also report sequence data from *Drosophila yakuba* for *Hem-protein, CKII-α,* and *Gelsolin*. For some analyses, I used previously published data from sequences of 40 *D. simulans* genes distributed across the X chromosome and chromosome arm 3R (Begun and Whitley 2000). Sequences were analyzed with the DnaSP program (Rozas and Rozas 1999). Silent mutations were classified as preferred or unpreferred (Sharp and Lloyd 1993) by using the outgroup method as described by Akashi (1996). New sequences reported here can be found in GenBank under accession numbers AY052156—AY052190.

The four genes located near the centromere of chromosome 3 of *D. simulans* (table 1) show reduced heterozygosity (θ) at silent sites compared with other genes on chromosome 3. The mean silent θ for these four genes (0.009) is significantly lower (Mann-Whitney, $P = 0.015$) than the mean for 19 genes located more distally (0.035). Despite the clear difference in the levels of silent polymorphism for centromeric versus more distal genes, the mean silent divergence of four centromeric genes (0.097) is not significantly different from the mean of other genes (0.108) on chromosome 3. Mean replacement θ values for noncentromeric ($\theta = 0.0013$) and centromeric genes ($\theta = 0.0009$) are not significantly different. Although the ratio of replacement θ to replacement divergence is smaller for the centromeric genes (0.15) than for the noncentromeric genes (0.44), the difference between regions is not significant.

In contrast to results from analysis of silent variation on the X chromosome and chromosome 3 in *D. simulans* (Begun and Whitley 2000), there is no good evidence for reduced amino acid polymorphism on the X chromosome. The mean replacement θ for X-linked genes ($n = 21$, $\theta = 0.0013$) is exactly the same as the estimate for noncentromeric chromosome 3 genes ($n = 19$, $\theta = 0.0013$). Similarly, replacement divergence values for the X chromosome (0.014) and chromosome 3 (0.011) are not significantly different (Mann-Whitney, $P = 0.75$). The ratio of replacement θ to replacement divergence is smaller for the X chromosome (0.21) than for chromosome 3 (0.44), but again the difference is not significant (Mann-Whitney, $P = 0.24$).

The different statistical conclusions on the effect of chromosomal location for silent versus replacement variation in *D. simulans* might be a result of reduced statistical power associated with low levels of amino acid variation (compared with those of silent variation) or might reflect a genuine difference between the dynamics of silent and replacement variation. More extensive sampling of replacement variation across the genome would be required to resolve this issue.

The two sampled genes located closest to the centromere are *CKII-α,* about 26 bands to the left of the centromere on 3L, and *Gelsolin,* about 8 bands to the right of the centromere on 3R. Both have severely reduced levels of silent heterozygosity (table 1) relative to the average for the chromosome (0.035; Begun and Whitley 2000). Silent θ for *Hem-protein* is relatively high ($\theta = 0.015$), although this gene is located only ~42 bands distal to the centromere on 3L. The same is true for *Amalgam* (about 97 bands to the right of the centromere on 3R), which is about as polymorphic as *Hem-protein*. These data suggest that severe reductions in silent heterozygosity are restricted to only a very small euchromatic region (perhaps 1,000 kb) on each side of the centromere of chromosome 3 in *D. simulans*. Few data exist on patterns of recombination near *D. melanogaster* centromeres, and even less information is available for *D. simulans*. However, the small amount of *D. simulans* genetic data (Sturtevant 1929; True, Mercer, and Laurie 1996) indicates a minimal centromere effect relative to *D. melanogaster*. Given the established relationship between recombination and polymorphism, the genetic and population data from *D. simulans* are consistent with a dramatic reduction in recombination over a small physical region near the centromere. Such regions may be ideal for detailed study of the effects of variable recombination on sequence evolution.

Key words: *Drosophila simulans*, protein variation, DNA polymorphism, natural selection, population.

Address for correspondence and reprints: David J. Begun, Section of Evolution and Ecology, University of California, Davis, California 95616. E-mail: djbegun@ucdavis.edu.

Mol. Biol. Evol. 19(2):201–203. 2002

Table 1
Silent and Replacement Heterozygosity (θ) and Divergence at Four Genes Located Near the Centromere of Chromosome 3 in *Drosophila simulans*

Gene	Cyt.[a]	Silent θ[b]	Replacement θ[b]	Silent divergence	Replacement divergence
Hem-protein	79E2	0.015	0.0005	0.097	0.0026
CKII-α	80A	0.000	0.0000	0.072	0.0000
Gelsolin	82A1–3	0.003	0.0020	0.088	0.0210
Amalgam	84A5	0.016	0.0011	0.105	0.0067

NOTE.—Silent divergence estimates between *D. simulans* and *D. melanogaster* are Jukes-Cantor corrected. The numbers of silent sites surveyed for *Hem-protein*, *CKII-α*, *Gelsolin*, and *Amalgam* were 260, 167, 232, and 235, respectively. Eight chromosomes were sampled for each gene.

[a] The polytene band location of the homologous gene in *D. melanogaster*.

[b] Watterson (1975).

Table 2
Replacement Polymorphism (θ) and Divergence (*D*) in *Drosophila simulans* Genes

Gene	Cyt.[a]	No. of Sites	θ[b]	*D*[c]	θ/*D*
Chromosome 3R					
Gld (*n* = 11)	84C	727	0.0006	0.0059	0.107
Rh3 (*n* = 5)	85F	850	0.0000	0.0024	0.000
mir (*n* = 6)	86A	924	0.0000	0.0045	0.000
nos (*n* = 7)	86D	545	0.0038	0.0321	0.117
eld (*n* = 7)	87B	717	0.0023	0.0084	0.272
Hsc70 (*n* = 7)	88E	980	0.0009	0.0005	1.745
CP190 (*n* = 7) ...	88E	970	0.0030	0.0168	0.176
ry (*n* = 8)	90A	1,029	0.0040	0.0044	0.915
hyd (*n* = 8)	92E	1,131	0.0003	0.0064	0.053
Rel (*n* = 7)	93D	1,871	0.0015	0.0502	0.031
pit (*n* = 7)	93F	849	0.0005	0.0025	0.190
AP50 (*n* = 8)	94B	977	0.0000	0.0000	—
T-cpl (*n* = 8)	94B	864	0.0000	0.0064	0.000
fzo (*n* = 8)	94E	1,052	0.0026	0.0358	0.072
AATS (*n* = 7)	95CD	1,014	0.0020	0.0116	0.174
tld (*n* = 7)	96A	732	0.0022	0.0045	0.497
Osbp (*n* = 8)	96B	814	0.0005	0.0083	0.057
boss (*n* = 5)	96F	1,167	0.0008	0.0029	0.282
Tpi (*n* = 9)	99E	561	0.0007	0.0002	3.300
Chromosome I (X)					
runt (*n* = 8)	19E	1,140	0.0004	0.0004	0.091
G6pd (*n* = 8)	18E	1,191	0.0000	0.0202	0.000
bnb (*n* = 8)	17E	715	0.0043	0.0140	0.308
r (*n* = 6)	15A	909	0.0000	0.0327	0.000
mei-218 (*n* = 8) ..	15D	853	0.0015	0.0886	0.017
sog (*n* = 8)	13D	920	0.0000	0.0011	0.000
g (*n* = 7)	12B	751	0.0015	0.0027	0.545
Yp3 (*n* = 8)	12BC	830	0.0012	0.0143	0.087
v (*n* = 8)	10A	868	0.0000	0.0012	0.000
Yp2 (*n* = 6)	9A	798	0.0000	0.0063	0.000
otu (*n* = 6)	7F	818	0.0043	0.0508	0.084
sn (*n* = 8)	7D	929	0.0000	0.0011	0.000
dec-1 (*n* = 7)	7C	1,065	0.0041	0.0244	0.168
ct (*n* = 6)	7B	848	0.0041	0.0049	0.837
sqh (*n* = 7)	5D	365	0.0000	0.0000	—
X (*n* = 7)	5D	1,087	0.0005	0.0030	0.156
ovo (*n* = 8)	4E	1,021	0.0005	0.0060	0.084
mei-9 (*n* = 6)	4B	901	0.0014	0.0140	0.100
per (*n* = 6)	3B	1,279	0.0027	0.0018	1.502
z (*n* = 6)	3A	606	0.0000	0.0033	0.000
Pgd (*n* = 7)	2D	693	0.0016	0.0058	0.273

[a] See Begun and Whitley (2000) for details on cytological locations of genes.

[b] Estimates of θ for X-linked genes have been multiplied by 4/3 to make them directly comparable to estimates from autosomal loci.

[c] *D* is the average pairwise divergence (uncorrected) between the *D. simulans* population sample and a single *D. melanogaster* allele.

Regions of normal recombination in *D. simulans* exhibit a ratio of unpreferred to preferred fixations of ~2 (Takano 1998; Begun 2001). For three genes located near the centromere of chromosome 3 (table 3), the ratio of unpreferred to preferred fixations (1:8) is significantly different (*G*-test, $P = 0.01$) from the ratio observed for more distally located genes on chromosome 3 (37:18). There should be no effect of recombination rate on the proportion of preferred versus unpreferred fixations for genes evolving at mutation-selection-drift equilibrium—we expect roughly equal numbers of fixations in the two presumptive fitness classes regardless of the recombinational environment (Akashi 1996). A possible explanation for the observed heterogeneity is that the recombination rate in the centromere region of *D. simulans* has recently increased. Genes with a long history of reduced recombination are expected to have a higher proportion of unpreferred to preferred codons at equilibrium than are genes located in regions of more extensive crossing over. This is because low rates of crossing over decrease the efficacy of purifying selection against very slightly deleterious (e.g., unpreferred) mutations. If the recombination rate in a region of historically low recombination recently increased, then we would expect to observe a transient increase in the fixation rate for very slightly beneficial mutations as such regions ''recover'' from a history of ineffectual purifying selection. Further investigation of silent fixations in centromeric regions of *D. simulans* and further genetic analysis of species in the melanogaster subgroup will be required to evaluate this hypothesis.

Six of 44 genes analyzed here (*Rel*, *runt*, *G6pd*, *r*, *mei-218*, and *otu*) deviate significantly from the neutral model in heterogeneity tests of silent and replacement polymorphism from *D. simulans* and fixed differences from *D. melanogaster*. Five of the six are X-linked, with the only exception being *Rel* (although it should be noted that significance was only marginal for *runt* and *mei-218*). Deviations in all six genes are in the direction of ''too many'' amino acid differences between species. Moreover, five of the six (with the exception being *runt*) exhibit high rates of protein evolution compared with other genes in these two species (table 2), consistent with the notion that these genes have experienced recurrent directional selection of amino acid mutations.

Acknowledgments

I thank P. Whitley for lab work and C. Langley, W. Stephan, and anonymous reviewers for comments. This

Table 3
Numbers of Unpreferred (U) and Preferred (P) Variants in 3 Genes Located Near the Centromere (*Hem-protein, CKII-α,* and *Gelsolin*) Versus 13 More Distally Located Genes on Chromosome 3 in *Drosophila simulans*

	Polymorphic		Fixed	
	U	P	U	P
Centromere . . .	4	5	1	8
Distal	136	25	37	18

work was supported by the NIH and the Sloan Foundation.

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.

Aguadé, M., N. Miyashita, and C. H. Langley. 1989. Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122**: 607–615.

Ashburner, M. 1989. Drosophila: a laboratory handbook. Cold Spring Harbor Press, New York.

Begun, D. J. 2001. The frequency distribution of nucleotide variation in *Drosophila simulans*. Mol. Biol. Evol. **18**: 1343–1352.

Begun, D. J., and C. F. Aquadro. 1991. Molecular population genetics of the distal region of the X chromosome in Drosophila: evidence for genetic hitchhiking of the *yellow-achaete* region. Genetics **129**:1147–1158.

Begun, D. J., and P. Whitley. 2000. Reduced X-linked nucleotide variation in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **97**:5960–5965.

Hilton, H., R. M. Kliman, and J. Hey. 1994. Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution **48**:1900–1913.

Langley, C. H., B. P. Lazzaro, W. Phillips, E. Heikkinen, and J. M. Braverman. 2000. Linkage disequilibrium and the site frequency spectra in the *su(s)* and *su(w^a)* regions of the *Drosophila melanogaster* X chromosome. Genetics **156**: 1837–1852.

Langley, C. H., J. MacDonald, N. Miyashita, and M. Aguadé. 1993. Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **90**:1800–1803.

Martin-Campos, J. M., J. M. Comeron, N. Miyashita, and M. Aguadé. 1992. Intraspecific and interspecific variation at the y-ac-sc region of Drosophila simulans and Drosophila melanogaster. Genetics **130**:805–816.

Rozas, J., and R. Rozas. 1999. DnaSP 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15**:174–175.

Sharp, P. M., and A. T. Lloyd. 1993. Codon usage. Pp. 378–397 *in* G. Maroni, ed. An atlas of Drosophila genes: sequences and molecular features. Oxford University Press, Oxford, England.

Sturtevant, A. H. 1929. Contributions to the genetics of *Drosophila simulans*. Carnegie Inst. Wash. Publ. **399**:1–62.

Takano, T. S. 1998. Rate variation of DNA sequence evolution in the Drosophila lineages. Genetics **149**:959–970.

True, J. R., J. M. Mercer, and C. C. Laurie. 1996. Differences in crossover frequency and distribution among three sibling species of Drosophila. Genetics **142**:507–523.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**:256–276.

Wayne, M. L., and M. Kreitman. 1996. Reduced variation at concertina, a heterochromatic locus in Drosophila. Genet. Res. **69**:101–108.

Wolfgang Stephan, reviewing editor

Accepted August 20, 2001