Nucleotide Variation at the *runt* Locus in *Drosophila melanogaster* and *Drosophila simulans*

Joanne A. Labate,[1] *Christiane H. Biermann, and Walter F. Eanes*

Department of Ecology and Evolution, State University of New York at Stony Brook

Intra- and interspecific nucleotide variation for the major developmental gene *runt* in *Drosophila* was studied in *D. melanogaster* and *D. simulans.* The 1.5- kb protein-coding region and the 0.4-kb intron of the *runt* gene were sequenced for 11 alleles in each species. The *D. melanogaster* alleles originated from east Africa. Estimated parameters of intraspecific variation in *D. melanogaster* (exons: $\theta = 0.018$, $\pi = 0.018$; intron: $\theta = 0.014$, $\pi = 0.014$) and *D. simulans* (exons: $\theta = 0.007$, $\pi = 0.005$; intron: $\theta = 0.008$, $\pi = 0.005$) were below average for other X-linked genes, while divergence between species (exons: $D = 0.094$; intron: $D = 0.069$) fell within the normal range for both silent and replacement changes. This estimate for *runt,* along with published values for three other genes in regions of normal recombination, show east African *D. melanogaster* to be roughly twice as polymorphic as *D. simulans.* The majority of nucleotide variation, silent and replacement, in both species was found to be selectively neutral using various statistical tests (HKA, McDonald-Kreitman, Tajima, and Fu and Li tests). Monte Carlo simulations of the coalescent process significantly rejected a Wright-Fisher model with respect to an amino acid polymorphism and the distribution of polymorphic sites among the *D. simulans* lines. This indicated an old lineage and may reflect ancestral population substructuring in *D. simulans.*

Introduction

Studies of nucleotide polymorphism in a variety of genes have revealed generalities about the nature of natural selection and historical population structure in the sibling species *Drosophila melanogaster* and *Drosophila simulans.* In *D. melanogaster,* an important observation has been the association of reduced silent site heterozygosity with regions of low recombination (Aguadé, Miyashita, and Langley 1989; Berry, Ajioka, and Kreitman 1991; Begun and Aquadro 1992). Another observation has been higher levels of restriction site polymorphism associated with collections of *D. melanogaster* from east Africa in comparison with cosmopolitan populations (Begun and Aquadro 1993). In *D. simulans,* the number of loci is limited and the sample sizes are smaller, but the main observations have been higher heterozygosity levels and patterns of silent site divergence that argue for historically larger population sizes than those of *D. melanogaster* (Akashi 1996; Moriyama and Powell 1996). Some of the *D. simulans* genes also possess curious genealogical features. One early observation was the incidence of shared lineages with its island endemics (Coyne and Kreitman 1986; Hey and Kliman 1993; Kliman and Hey 1993). There are other unusual features. For example, not only is mtDNA sequence diversity in *D. simulans* low, but there is also a highly derived lineage that accounts for almost all of the polymorphic sites (Ballard and Kreitman 1994). Rand, Dorfsman, and Kann (1994) discovered two distinct lineages for the *ND5* gene within a sample of *D. simulans* mtDNA. Likewise, the *Tpi* locus in *D. simulans* shows an old lineage defined by a significantly large number of derived mutations (Hasson et al. 1998). These results, along with the observations on the *vermilion* locus by Begun and Aquadro (1995*b*), suggest that *D. simulans* may possess greater population subdivision than originally envisioned.

In this study, we examine sequence variation in the *runt* gene in *D. melanogaster* and *D. simulans.* We had several original motivations for examining this gene. The earlier study of *G6pd* (or *Zw*) showed relatively high levels of polymorphism in *D. melanogaster,* but notably low levels of polymorphism in *D. simulans* (Eanes et al. 1996). The *G6pd* locus is located near the proximal end of the X chromosome, at band 18E12–13, and the *su(f)* locus at polytene band 20E shows drastically reduced levels of polymorphism in both cosmopolitan (Langley et al. 1993) and east African populations (Begun and Aquadro 1995*a*). The *runt* gene falls roughly halfway between these two genes on the polytene cytological map at 19E1–2, and therefore has relevance to the question of where heterozygosity levels may decrease relative to the centromeric region. Another motivation was the underrepresentation of complex genes involved in early development in the evolutionary studies of genes in this species pair. One such early developmental gene, *transformer* (*tra*), has greatly reduced levels of silent polymorphism (Walthour and Schaeffer 1994), yet also exhibits extremely elevated levels of interspecific divergence in amino acid replacements. In contrast, a recent study of the *dpp* gene, which is also involved in metamorphosis, showed relatively normal levels of silent polymorphism while appearing very highly constrained at the amino acid level (Richter et al. 1997). The *runt* gene is involved in several major early developmental processes, including segmentation (Nüsslein-Volhard and Wieschaus 1980), sex determination (Duffy and Gergen 1991), and neurogenesis (Duffy, Kania, and Gergen 1991). It is an excellent candidate for an early developmental gene.

Materials and Methods

The *D. melanogaster runt* alleles originated from seven isofemale lines collected from Zimbabwe (desig-

[1] Present address: USDA/ARS and Department of Agronomy, Iowa State University.

Key words: polymorphism, population genetics, *runt,* molecular evolution, *Drosophila melanogaster, Drosophila simulans.*

Address for correspondence and reprints: USDA/ARS and Department of Agronomy, Iowa State University, Ames, Iowa 50011. E-mail: jlabate@iastate.edu.

Mol. Biol. Evol. 16(6):724–731. 1999

nated Z) and four isofemale lines from Okavango Delta, Botswana, collected in 1985 (OK). The *runt* gene was made homozygous by crossing a single male from an isofemale line to virgin FM6/N^{264-84} females and extracting the X chromosome to homozygosity (Eanes et al. 1989). Five *D. simulans runt* alleles originated from Davis Peach Farm, Mount Sinai, N.Y., in 1990 (DPF), three from Montpelier, France, in 1991 (MT), and three from Vera Cruz, Mexico, in 1990 (VC). These were extracted using an attached-X stock provided by J. A. Coyne (University of Chicago).

DNA from isochromosomal OK lines was extracted, and the locus *runt* was PCR-amplified and sequenced using standard protocols (Hasson et al. 1998). Sequencing primers were spaced approximately every 300 bp; all polymorphisms were confirmed on both strands. Occasionally, small gaps of less than 10 nt (less than 1%–2% of the total sequence) were not legible for one strand. One possible PCR-originated error was observed, a deletion of about 20 nt within an exon in one line. Because this was only seen once, it was considered to be an artifact.

All sequences have been deposited in GenBank under accession numbers AF077059–AF077080. The original *D. melanogaster runt* sequence was reported by Kania et al. (1990). Within a 12-alanine homopolymeric stretch (their nucleotide positions 304–339), we found a 13th codon for alanine in all *D. melanogaster* and *D. simulans* lines. Our numbering system also differs from theirs in that we designated the translation initiation site as position +1 and included 414 nt for the intron between exons 1 and 2.

Four standard tests of neutrality of the observed nucleotide variation were used, including the fit of the silent polymorphism frequency spectrum to the neutral expectation (Tajima 1989; Fu and Li 1993) and contrasts of interlocus polymorphism and divergence under the neutral theory (Hudson, Kreitman, and Aguadé 1987; McDonald and Kreitman 1991). For the HKA tests, the *runt* locus was contrasted with the *ver* and *G6pd* loci, because east African sequence data are available for these loci (Eanes et al. 1996; Begun and Aquadro 1995*b*), both are X-linked, and both are located in regions of relatively normal recombination. For *D. simulans*, the same loci were used, because they constitute the largest data sets for that species and are also X-linked (Eanes et al. 1996; Begun and Aquadro 1995*b*). In the HKA tests, the divergence between the random pair of alleles designated Z11 and DPF2 was used.

To evaluate the potential impact of amino acid polymorphisms on distortion of the genealogical relationships in our samples of alleles, we carried out Monte Carlo simulations of the Wright-Fisher process to investigate the significance of specific features of the sequence data with respect to observed amino acid mutations (Hudson 1990, 1993). The specific features, defined a priori, were (1) the number of silent polymorphisms associated with subsets of alleles bearing particular derived mutations and (2) the number of ''fixed'' mutations identified between those same subsets. Singleton amino acid mutations were examined with respect to the latter feature. For example, given *m* observed polymorphisms in a total sample of *n* genes, what is the probability that a subset of *i* alleles (or a singleton allele) will have *j* or more sites uniquely fixed? A large number of replicate trees were generated, each with *m* polymorphic sites, by simulating a coalescent process using an adaptation of the algorithm proposed by Hudson (1990). The proportion of simulated trees with fixed differences equal to or greater than the observed number was determined. A sample of 10,000 replicate trees was generated in each simulation (Eanes et al. 1996).

Results and Discussion

The entire 1.53-kb coding region and the 0.41-kb intron (totaling 1,947 nt including insertions/deletions) were sequenced on both strands for 11 *D. melanogaster* and 11 *D. simulans* isochromosomal lines. Nucleotide position 1 (table 1) is the first site of the methionine start codon. Exon 1, the intron, and exon 2 span positions 1–709, 710–1123, and 1124–1947, respectively. Twenty-three polymorphic sites were found in the *D. melanogaster* exons; three of these were amino acid polymorphisms (table 1 and fig. 1). The intron contained 17 nucleotide and three insertion/deletion polymorphisms. The 11 *D. simulans* sequences contained 9 polymorphisms in the exons, including a single amino acid polymorphism, and 13 polymorphic sites in the introns, including 3 indels. Fixed site differences between *D. melanogaster* and *D. simulans* included 16 synonymous differences, 5 nonsynonymous differences, and no indels in the exons, and 16 nucleotide differences and 8 small (1–5 nucleotide) indels in the intron (table 1 and fig. 1).

The estimate of interspecific divergence based on effective numbers of silent sites and using a random pair of sequences (Z11 and DPF2) was 0.094 in the *runt* exons and 0.069 in the intron. These are typical values. Estimates of θ (Watterson 1975) and π (Nei and Li 1979) are given in table 2 and were based on effective numbers of silent sites (Kreitman 1983) of 385 for the exons and 409 and 404 for *D. melanogaster* and *D. simulans* introns, respectively. *Drosophila melanogaster* was more variable than *D. simulans*; however, we used as our sample of *D. melanogaster* isofemale lines from Okavango Delta, Botswana, and Zimbabwe. Begun and Aquadro (1993) have shown these populations to be two to three times as polymorphic as most cosmopolitan populations from temperate climates, and we were interested in focusing on a population that was more likely to be in ''genetic equilibrium'' than cosmopolitan populations. The original inference of higher polymorphism came from restriction site variation; however, only limited data exist at the direct sequence level of individual genes for east African lines. For the five genes now reported, the *runt, G6pd, ver,* and *Tpi* loci show levels of polymorphism in east African *D. melanogaster* that are actually higher than those in *D. simulans.* The only exception is the *Gld* locus, but this is in a region of low recombination (Hamblin and Aquadro 1997). Between loci, the observed levels of silent polymorphism at *runt*

Table 1
Polymorphic and Diverged Sites Identified in 11 *runt* Alleles Each of *Drosophila simulans* and *Drosophila melanogaster*

D. simulans polymorphism

Nucleotide Position

	57	84	133	579	873	938	943	1011	1024	1028	1040	1041	1056	1077	1078	1086	1098	1152	1251	1671	1674	1836
MEL Z11	T	G	A	G	T	A	T	C	C	T	C	C	A	C	G	C	C	T	G	C	C	C
VC910	.	G	G	.	.	.	C	.	.	.	3	4	.	.	.	.	T	C	.	.	T	.
DPF4	.	T	T	.	C	T	A	T	A	5	A	.	T	T	.	.	.	.	.	A	.	.
MT6	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	.
DPF1	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	T
DPF2	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	T
DPF3	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	T
DPF87	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	T
MT5	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	T
MT12	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	T
VC902	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	T
VC903S	C	T	T	T	.	T	A	T	.	5	A	.	.	T	A	T	.	C	A	.	.	T

D. melanogaster polymorphism

Line	87	130	372	381	630	767	792	793	818	855	866	870	871	902	903	905	925	928	964	994	1002	1003	1024	1041	1104	1284	1341	1357	1422	1455	1479	1560	1587	1608	1611	1662	1704	1710	1746	1785	1818	1819	1869
SIM DPF2	T	G	C	C	C	C	G	C	G	C	A	C	C	1	A	T	A	T	C	C	C	T	C	C	T	C	C	C	G	C	C	C	G	C	C	G	C	C	G	C	G	G	C
Z16	C	A	.	.	.	G	.	.	.	.	.	.	.	T	G	2	C	.	.	.	.	.	.	.	C	.	.	.	T	.	T	T	A	.	.	.	.	.	A	A	T	T	.
OK13	C	A	.	.	.	G	.	.	.	.	.	.	.	T	G	2	C	.	.	.	.	.	.	.	C	.	.	.	T	.	T	T	A	.	.	.	.	.	A	A	T	T	.
OK5	.	A	.	.	.	G	C	G	.	.	.	.	.	T	G	2	C	.	.	.	.	.	.	.	C	.	.	.	T	.	T	T	A	.	.	.	.	.	A	A	T	T	.
Z27	C	A	.	.	.	.	.	.	.	.	.	.	.	T	G	2	C	.	.	.	.	.	.	.	C	.	T	.	T	.	T	T	A	.	.	.	T	T	A	.	T	T	.
Z11	C	A	.	.	A	.	.	.	.	.	C	.	.	T	.	2	T	.	.	.	.	A	.	.	A	.	.	.	.	G	.	T	A	.	.	.	T	T	A	.	.	.	.
Z42	C	.	.	.	A	.	.	.	.	.	C	.	.	T	.	2	T	.	.	.	.	.	.	.	A	.	.	.	.	G	.	T	A	.	.	.	T	.	A	A	.	.	.
Z55	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	2	C	A	T	.	T	.	.	.	C	.	.	.	.	.	.	.	A	.	.	.	.	.	A	A	.	.	.
Z74	.	.	.	.	.	.	.	.	.	.	.	T	T	T	G	2	C	A	T	.	T	.	.	.	C	T	.	.	.	.	.	.	.	.	A	.	.	.	A	A	.	.	.
Z64	.	.	.	A	.	.	.	.	A	.	.	.	.	1	.	.	T	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	T
OK17	.	.	.	.	.	.	.	.	A	A	.	T	.	1	.	.	C	.	.	.	.	A	A	4	C	.	.	.	.	.	T	T	.	.	.	A	.	.	A	A	.	.	.
OK94	.	.	T	.	.	.	.	.	A	.	.	T	.	1	.	.	T	.	.	.	.	A	A	.	C	.	T	T	.	.	.	.	.	A	.	.	.	.	.	A	.	.	T

Fixed diverged sites

	128	129	133	342	501	582	624	744	750	773	811	880	882	925	929	942	943	950	953	958	959	960	989	1015	1033	1038	1040	1047	1055	1068	1104	1245	1266	1278	1380	1398	1399	1416	1452	1545	1549	1569	1707	1851	1860
														C																	C														
MEL	C	C	A	T	A	A	T	A	6	A	A	G	T	T	T	T	T	T	8	G	C	T	C	G	I	T	C	11	A	G	A	G	A	C	A	G	T	G	T	T	T	C	T	T	A
SIM	T	T	T	C	G	C	C	G	I	7	C	A	A	A	G	A	A	C	A	C	G	9	T	A	10	A	3	C	12	T	T	A	C	G	G	T	G	A	G	A	C	A	C	A	C
			G														C										A																		

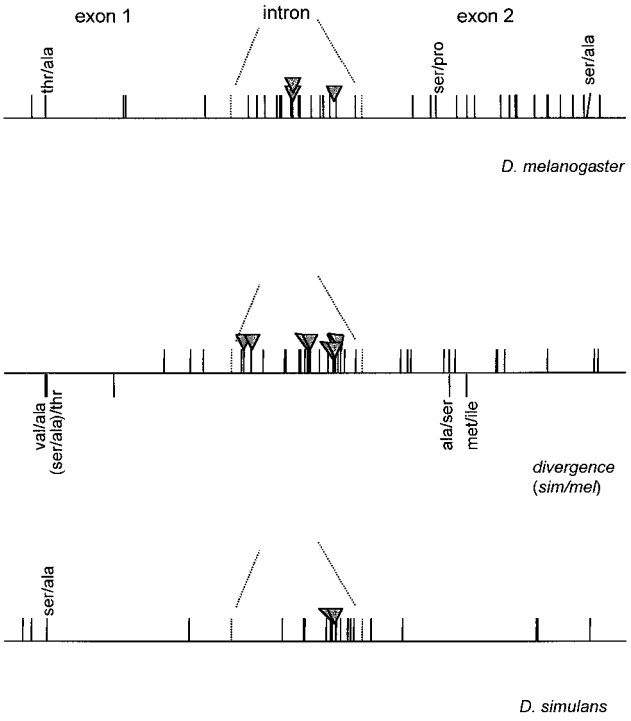

FIG. 1.—The distribution of polymorphic and diverged sites across the *runt* coding region for 11 *D. melanogaster* and 11 *D. simulans* alleles. All amino acid polymorphic sites are labeled; triangles indicate insertion/deletion polymorphisms.

were substantially lower than levels observed for east African lines sampled for the *G6pd* ($\theta = 0.027$; Eanes et al. 1996), *ver* ($\theta = 0.042$; Begun and Aquadro 1995*b*), and *Tpi* ($\theta = 0.038$; Hasson et al. 1998), but not *Gld* ($\theta = 0.001$; Hamblin and Aquadro 1997). Likewise, estimates of θ and π for *D. simulans* were notably lower than those for genes in regions with normal levels of recombination (Moriyama and Powell 1996). These results suggest that heterozygosity may be reduced in this region of the chromosome, but it must be recognized that all HKA tests of interlocus contrasts were not significant (table 3).

There is no evidence for interspecific selection on amino acid replacements. All amino acid substitutions in *D. melanogaster* and *D. simulans runt* loci, both within and between species, differed by two or fewer physicochemical properties and are considered to be conservative changes (Taylor 1986). The ratio of amino acid to silent polymorphisms within *D. simulans* and *D. melanogaster* was typical of X-linked loci (Begun 1996). The replacement : synonymous mutation ratios pooled across species were 4:29 for polymorphisms and 5:16 for fixed differences. Neither this contrast nor the individual partitionings of the data by species were statistically significant by the McDonald-Kreitman test (table 3). Considering the results for *runt* and *dpp* (Richter et al. 1997), the extremely high amino acid substitution rate at *tra* (O'Neil and Belote 1992) is not typical of early developmental genes.

A sliding-window analysis was used to explore contrasting patterns of intra- and interspecific variation

Table 2
Estimates of θ[a] and π for *Drosophila melanogaster* and *Drosophila simulans runt* Loci Based on Synonymous Polymorphism and Effective Numbers of Silent Sites

		EXON[b]		INTRON	
	n	$\hat{\theta}$	$\hat{\pi}$	$\hat{\theta}$	$\hat{\pi}$
D. melanogaster	11	0.018	0.018	0.014	0.014
D. simulans.	11	0.007	0.005	0.008	0.005

[a] Because of X-linkage, $\theta = 3N_e\mu$.
[b] Estimates are based on 385 silent-site equivalents.

across a gene (Hudson and Kaplan 1988; Kreitman and Hudson 1991). Because expected polymorphism and divergence within a region share the mutation rate (μ) in the neutral parameter θ, the plot of nucleotide position versus intraspecific polymorphism across a locus should coincide in shape with the corresponding plot of interspecific divergence. Nucleotide position versus mean variation within species (Watterson 1975) and between species (Li and Graur 1991, eq. 3.19) was plotted across the *runt* coding region (fig. 2). The positions of the intraspecific amino acid polymorphisms are shown. The two measures appear concordant across the *D. melanogaster* sequence. There was no apparent excess (or deficiency) of heterozygosity in the regions containing amino acid polymorphisms within either species, as might be indicative of selection. There was a stretch of relatively reduced intraspecific variation in *D. simulans*, approximately between positions 1350 and 1600.

In *D. melanogaster*, the putative functional regions showed a relatively reduced rate of evolution at silent sites when compared with the remainder of the exons. Polymorphic and diverged sites were scattered throughout the coding region, but the cluster of functional domains in the 5′ end of the gene (codons 1–315; Kania et al. 1990) appeared to be slightly more conserved. Estimates of θ for synonymous sites in codons 1–315 and in codons 316–510 were 0.009 and 0.032, respectively, and estimates of divergence for a random pair of alleles between species for these regions were 0.062 and 0.117, respectively. This conservation of synonymous sites could be caused by selection for translational accuracy within functional domains (Akashi 1994).

The Monte Carlo simulations (Hudson 1990, 1993) were used to address two types of specific null hypotheses that are associated with the distribution of silent-site variation within and between sets of alleles defined by specific amino acid polymorphism (Hudson et al. 1994; Eanes et al. 1996). The distribution of synonymous polymorphisms within and between sets can reflect historical selection acting on the amino acid substitution that has resulted in rapid increases in allele frequency or, conversely, the persistence of specific amino acid mutations and the subsequent accumulation of differences between sets. In *D. melanogaster*, we have three amino acid variants at nucleotide positions 130 (Ala/Thr), 1357 (Pro/Ser), and 1819 (Ala/Ser) to address. There were 37 polymorphisms in the entire set

Table 3
Summary Statistics for Tajima, Fu and Li, McDonald-Kreitman, and HKA Tests

	n	Tajima's *D*	Fu and Li's *D*	McDonald-Kreitman (*G*)	HKA
Drosophila melanogaster.....	11	+0.002	+0.239	0.807	*runt* vs. *G6pdh*, 1.57 *runt* vs. *ver*, 0.017
Drosophila simulans.....	11	−1.427	−1.541	0.624	*runt* vs. *G6pdh*, 0.148 *runt* vs. *ver*, 1.52

of $n = 11$ sequences. In the first case, eight silent polymorphisms were segregating in the four *runt* sequences bearing the Ser change at position 1819, with two ''fixed'' differences between sets. In the case of the Thr polymorphism at nucleotide 133, there were 18 polymorphisms associated with 5 lines bearing the Thr change (they overlapped with the Ser change; see table 1), and there were zero fixed differences associated with this set. Finally, there were two fixed differences associated with the single line (OK94) bearing the Pro/Ser change at nucleotide 1357. In all cases, the observed distributions of polymorphic sites and fixed differences were commonly observed in simulations of the coalescence process. Thus, there is little evidence for historical selection acting on these particular amino acid polymorphisms. These simulations are predicated

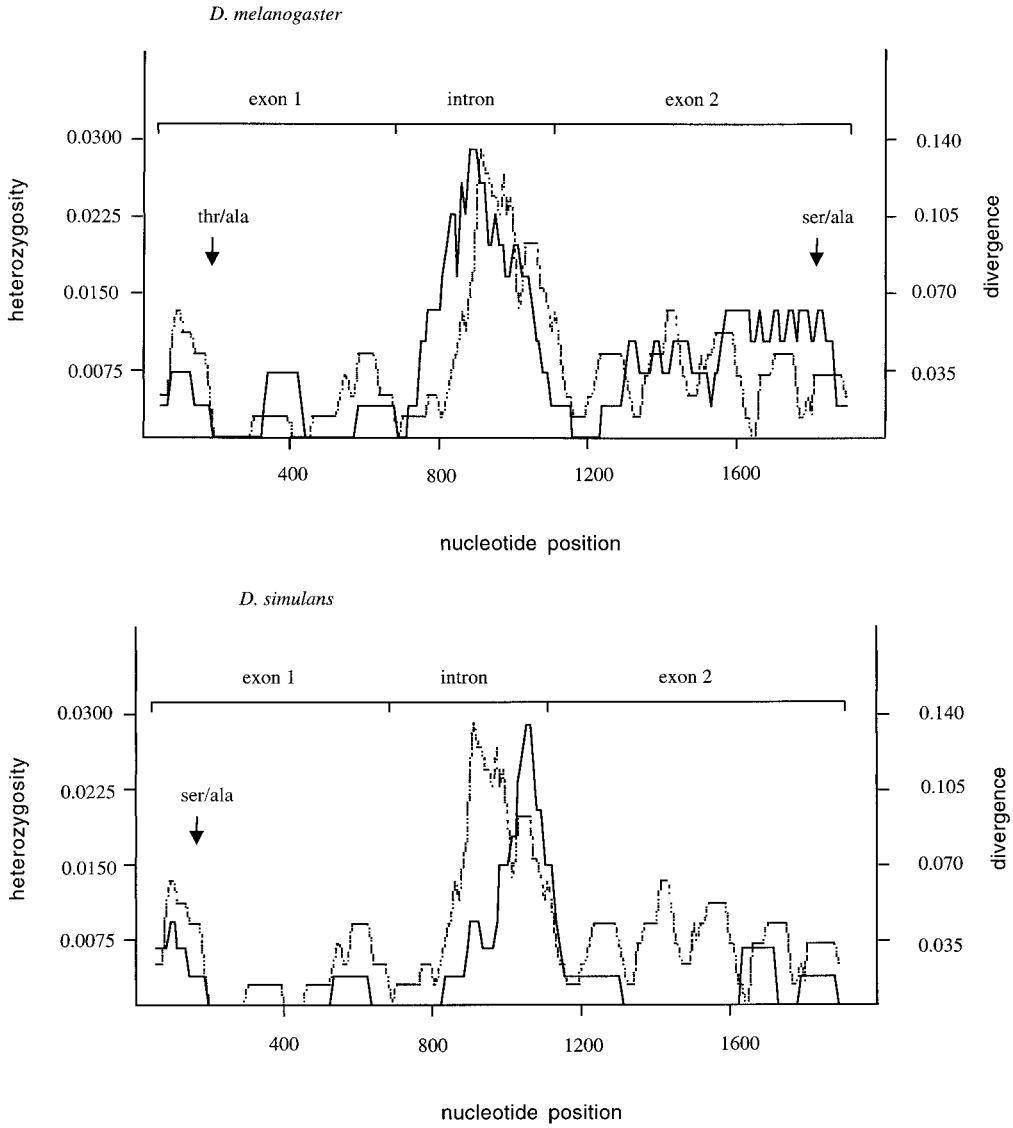

FIG. 2.—Sliding window of heterozygosity within *D. melanogaster* or *D. simulans* (solid line) and divergence between *D. melanogaster* and *D. simulans* (dotted line) *runt.* The window's size was 100 nt, and it was moved along the *x*-axis in 10-nt increments. The *y*-axis values were based on absolute numbers of sites within the window rather than effective numbers.

on the assumption of no recombination among lineages, and while the presence of recombination does not invalidate the simulations, it does reduce the power to detect historical selection and its imprint on the diversity of linked sequences.

The collection of 11 sequences from *D. simulans* showed a single amino acid polymorphism (Ser/Ala at nucleotide 133) associated with line VC910. Because *D. melanogaster* possessed a Thr at this position, it was impossible to establish the polarity of this change. In the total sample, there were 22 silent changes, but 11 were contributed as singletons by this line. In our simulations, the occurrence of a single lineage possessing an amino acid mutation and with this number of changes (or greater) was rarely observed ($P = 0.0013$). This suggested that, with respect to the sample of *D. simulans runt* sequences and this amino acid polymorphism, the population was not at Wright-Fisher equilibrium. This indicates the retention of an old lineage in association with the amino acid polymorphism, and the possibility of a hitchhiking event whereby the most derived lineages have increased in frequency sufficiently recently that recombination and mutation have yet to obscure the event.

The HKA (Hudson, Kreitman, and Aguadé 1987), Tajima (1989), McDonald-Kreitman (1991), and Fu and Li (1993) tests applied to the *runt* locus failed to reject the null hypothesis that the majority of intraspecific and interspecific nucleotide variation within the locus is selectively neutral (table 3). In contrast, the sample of *D. simulans runt* alleles gave a highly significant result with respect to fixed differences in the Monte Carlo simulation. Both Tajima's and Fu and Li's tests gave negative *D* statistics consistent in direction with the excess of singleton mutations. The apparent contradiction between rejection of the Wright-Fisher neutral population model using Monte Carlo simulation versus the other tests (Tajima's and Fu and Li's) is due to both the lack of statistical power in the latter tests and the specific and conditional nature of the hypotheses being tested by simulation.

The significant lack of variation in the set of *D. simulans runt* sequences, along with the unique derived lineage, has several biological explanations. One possibility is the catching of an incipient ''selective sweep'' event (Kaplan, Hudson, and Langley 1989). The most derived sequences may have recently experienced rapid directional selection acting either directly on the amino acid site or on a closely linked site. Alternatively, this may be an old balanced polymorphism that has recently experienced a population bottleneck. Background selection would not appear to be a viable explanation. Like selective sweeps, background selection reduces steady-state levels of polymorphism; however, it does not produce genealogies or allele frequency spectra that deviate from neutrality (Braverman et al. 1995; Charlesworth, Charlesworth, and Morgan 1995). It would be interesting to examine this polymorphism across a larger geographic scale. It should be noted that hitchhiking and background selection are both associated with regions of low recombination. Despite its proximity to polytene region 20, genetic exchange is usually assumed to be nearly normal in this region (see Kliman and Hey 1993). The inferred minimum number of recombination events (Hudson and Kaplan 1985) in the *D. melanogaster* sample was relatively large (11), as was the estimated recombination parameter ($3Nc = 30$; Hudson 1987). In the *D. simulans* sample, there was no evidence of any recombination event.

Finally, the deep genealogical structure for *runt* in *D. simulans* could also indicate past population structuring. Unfortunately, the data for most gene sequences in *D. simulans* are sparse, averaging four to six sequences per gene (Moriyama and Powell 1996). The probability of recovering the deepest lineage (coalescing at the oldest node) is $(n - 1)/(n + 1)$ (Kliman and Hey 1993), so it is likely that the deepest lineage has been missed in some other data sets. As with *runt, Tpi* (Hasson et al. 1998), *ver* (Begun and Aquadro 1995*b*), and the *ND5* (Rand, Dorfsman, and Kann 1994) and *cytochrome b* (Ballard and Kreitman 1994) genes of mtDNA, show deep lineages in *D. simulans* and are generally characterized by asymmetry in the number of derived silent mutations. This is not seen in all loci, not even in some with substantial sample sizes (see Eanes et al. 1996; Hamblin and Aquadro 1996). Nevertheless, it is expected that in a subdivided population, all genes will not diverge equally, nor after admixture will all lineages be retained for all loci. The imbalance in mutation accumulation seen in some genes is possibly a remnant of ancestral subdivision with population-size-specific accumulation of differences (Akashi 1996). Recent examination of the phylogenic relationships of the island endemic species *Drosophila mauritiana* and *Drosophila sechellia,* which still share some polymorphic lineages with *D. simulans,* have suggested sequential island hopping because of the number of shared derived mutations between the island endemics (Caccone et al. 1996). However, an ancient subdivision from which both island species are independently founded may also provide an explanation. These results for the *runt* locus add to the growing list of loci suggesting that *D. simulans* is not a species in genetic equilibrium.

Acknowledgments

This work was supported by National Institute of General Medical Sciences grant GM-45247 to W.F.E. This is contribution number 1029 from the Graduate Program in Ecology and Evolution, State University of New York at Stony Brook.

LITERATURE CITED

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.

AKASHI, H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics **136**:927–935.

———. 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.

BALLARD, W. O. J., and M. KREITMAN. 1994. Unraveling selection in the mitochondrial genome of *Drosophila*. Genetics **138**:757–772.

BEGUN, D. J. 1996. Population genetics of silent and replacement variation in *Drosophila simulans* and *D. melanogaster*: X/autosome differences? Mol. Biol. Evol. **13**:1405–1407.

BEGUN, D. J., and C. F. AQUADRO. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356**:519–520.

———. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365**:548–550.

———. 1995*a*. Evolution at the tip and base of the X chromosome in an African population of *Drosophila melanogaster*. Mol. Biol. Evol. **12**:382–390.

———. 1995*b*. Molecular variation at the *vermilion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140**:1019–1032.

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

BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY, and W. STEPHAN. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140**:783–796.

CACCONE, A., E. N. MORIYAMA, J. M. GLEASON, L. NIGRO, and J. R. POWELL. 1996. A molecular phylogeny for the *Drosophila melanogaster* subgroup and the problem of polymorphism data. Mol. Biol. Evol. **13**:1224–1232.

CHARLESWORTH, D., B. CHARLESWORTH, and M. T. MORGAN. 1995. The pattern of neutral molecular variation under the background selection model. Genetics **141**:1619–1632.

COYNE, J. A., and M. KREITMAN. 1986. Evolutionary genetics of two sibling species, *Drosophila simulans* and *D. sechellia*. Evolution **40**:673–691.

DUFFY, J. B., and J. P. GERGEN. 1991. The *Drosophila* segmentation gene *runt* acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene *Sex-lethal*. Genes Dev. **5**:2176–2187.

DUFFY, J. B., M. A. KANIA, and J. P. GERGEN. 1991. Expression and function of the *Drosophila* gene *runt* in early stages of neural development. Development **113**:1223–1230.

EANES, W. F., J. W. AJIOKA, J. HEY, and C. WESLEY. 1989. Restriction-map variation associated with the *G6PD* polymorphism in natural populations of *Drosophila melanogaster*. Mol. Biol. Evol. **6**:384–397.

EANES, W. F., M. KIRCHNER, J. YOON, C. H. BIERMANN, I.-N. WANG, M. A. MCCARTNEY, and B. C. VERRELLI. 1996. Historical selection, amino acid polymorphism and lineage-specific divergence at the *G6pd* locus in *Drosophila melanogaster* and *D. simulans*. Genetics **144**:1027–1041.

FU, Y.-X., and W.-H. LI. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693–709.

HAMBLIN, M. T., and C. F. AQUADRO. 1996. High nucleotide sequence variation in a region of low recombination in *Drosophila simulans* is consistent with the background selection model. Mol. Biol. Evol. **13**:1133–1140.

———. 1997. Contrasting patterns of nucleotide sequence variation at the *glucose dehydrogenase* (*Gld*) locus in different populations of *Drosophila melanogaster*. Genetics **145**:1053–1062.

HASSON, E., I.-N. WANG, L.-W. ZENG, M. KREITMAN, and W. F. EANES. 1998. Nucleotide variation in the *triosephosphate isomerase* (*Tpi*) locus of *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **15**:756–769.

HEY, J., and R. M. KLIMAN. 1993. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10**:804–822.

HUDSON, R. R. 1987. Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50**:245–250.

———. 1990. Gene genealogies and the coalescent process. Pp. 1–44 *in* D. FUTUYMA and J. ANTONOVICS, eds. Oxford surveys in evolutionary biology. Oxford University Press, New York.

———. 1993. The how and why of generating gene genealogies. Pp. 23–36 *in* N. TAKAHATA and A. G. CLARK, eds. Mechanisms of molecular evolution. Sinauer, Sunderland, Mass.

HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIATOWSKI, and F. J. AYALA. 1994. Evidence for positive selection in the *superoxide dismutase* (*Sod*) region of *Drosophila melanogaster*. Genetics **136**:1329–1340.

HUDSON, R. R., and N. L. KAPLAN. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111**:147–164.

———. 1988. The coalescent process in models with selection and recombination. Genetics **120**:831–840.

HUDSON, R. R., M. KREITMAN, and M. AGUADÉ. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

KANIA, M. A., A. S. BONNER, J. B. DUFFY, and J. P. GERGEN. 1990. The *Drosophila* segmentation gene *runt* encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. Genes Dev. **4**:1701–1713.

KAPLAN, N. L., R. R. HUDSON, and C. H. LANGLEY. 1989. The ''hitchhiking effect'' revisited. Genetics **123**:887–899.

KLIMAN, R. M., and J. HEY. 1993. DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics **133**:375–387.

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

KREITMAN, M., and R. R. HUDSON. 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127**:565–582.

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.

LI, W.-H., and D. GRAUR. 1991. Fundamentals of molecular evolution. Sinauer, Sunderland, Mass.

MCDONALD, J. H., and M. KREITMAN. 1991. Adaptive protein evolution at the *ADH* locus in *Drosophila*. Nature **351**:652–654.

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

NEI, M., and W.-H. LI. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76**:5269–5273.

NÜSSLEIN-VOLHARD, C., and E. WIESCHAUS. 1980. Mutations affecting segment number and polarity in *Drosophila*. Nature **287**:795–801.

O'NEIL, M. T., and J. M. BELOTE. 1992. Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. Genetics **131**:113–128.

RAND, D. M., M. DORFSMAN, and L. M. KANN. 1994. Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. Genetics **138**:741–756.

RICHTER, B., M. LONG, R. C. LEWONTIN, and E. NITASAKA. 1997. Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in *Drosophila.* Genetics **145**:311–323.

TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

TAYLOR, W. R. 1986. The classification of amino acid conservation. J. Theor. Biol. **119**:205–218.

WALTHOUR, C. S., and S. W. SCHAEFFER. 1994. Molecular population genetics of sex determination genes—the *transformer* gene of *Drosophila melanogaster*. Genetics **136**:1367–1372.

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

DAVID M. RAND, reviewing editor

Accepted February 15, 1999