Copyright © 1994 by the Genetics Society of America

Neutral and Non-Neutral Evolution of Drosophila Mitochondrial DNA

David M. Rand, Michele Dorfsman and Lisa M. Kann

Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912

Manuscript received March 11, 1994
Accepted for publication July 21, 1994

ABSTRACT

To test hypotheses of neutral evolution of mitochondrial DNA (mtDNA), nucleotide sequences were determined for 1515 base pairs of the NADH dehydrogenase subunit 5 (ND5) gene in the mitochondrial DNA of 29 lines of *Drosophila melanogaster* and 9 lines of its sibling species *Drosophila simulans*. In contrast to the patterns for nuclear genes, where *D. melanogaster* generally exhibits much less nucleotide polymorphism, the number of segregating sites was slightly higher in a global sample of nine ND5 sequences in *D. melanogaster* ($s = 8$) than in the nine lines of *D. simulans* ($s = 6$). When compared to variation at nuclear loci, the mtDNA variation in *D. melanogaster* does not depart from neutral expectations. The ND5 sequences in *D. simulans*, however, show fewer than half the number of variable sites expected under neutrality when compared to sequences from the *period* locus. While this reduction in variation is not significant at the 5% level, HKA tests with published restriction data for mtDNA in *D. simulans* do show a significant reduction of variation suggesting a selective sweep of variation in the mtDNA in this species. Tests of neutral evolution based on the ratios of synonymous and replacement polymorphism and divergence are generally consistent with neutral expectations, although a significant excess of amino acid polymorphism within both species is localized in one region of the protein. The rate of mtDNA evolution has been faster in *D. melanogaster* than in *D. simulans* and the population structure of mtDNA is distinct in these species. The data reveal how different rates of mtDNA evolution between species and different histories of neutral and adaptive evolution within species can compromise historical inferences in population and evolutionary biology.

MITOCHONDRIAL DNA is widely used as a genetic marker in population and evolutionary biology. It is generally assumed that the variation in mtDNA sequences is neutral with respect to fitness and that the patterns of nucleotide variation can be used to infer the evolutionary histories of populations and closely related species. In recent years several reports have described non-neutral behavior of mtDNA haplotypes in experimental populations (MACRAE and ANDERSON 1988; FOS *et al.* 1990; NIGRO and PROUT 1990). As general tests for the neutrality of mtDNA, these results have been inconclusive either because the experimental design has been challenged (SINGH and HALE 1990) or because nuclear-cytoplasmic interactions have been implicated as the basis for the presumed fitness differences of mtDNAs (FOS *et al.* 1990; NIGRO and PROUT 1990). While direct experimentation will continue to play an important role in addressing the *mechanisms* of mitochondrial and nuclear coevolution, its ability to distinguish between neutral *vs.* adaptive molecular evolution in the *history* of a species remains uncertain (GILLESPIE 1991).

Nucleotide sequence data offer an alternative and potentially more powerful means of testing hypotheses of neutrality. Here one seeks to infer, from the patterns of nucleotide polymorphism and divergence, whether the observed variation is consistent with specific predictions of the neutral theory of molecular evolution (KIMURA 1983). One such prediction is that the levels of nucleotide polymorphism within a species will be correlated with levels of sequence divergence between species (*e.g.*, HUDSON *et al.* 1987). An additional prediction of the neutral theory is that the ratio of synonymous: replacement substitutions will be the same within and between species (MCDONALD and KREITMAN 1991). In Drosophila a rapidly growing number of nuclear loci has been subjected to these (and other) tests of neutrality (*e.g.*, KREITMAN and HUDSON 1991; BEGUN and AQUADRO 1991, 1993, 1994; BERRY *et al.* 1991; MARTIN-CAMPOS *et al.* 1992; SCHAEFER and MILLER 1992; LANGLEY *et al.* 1993). Given the widespread use of mtDNA in population biology, the application of these methods to mtDNA sequences seems long overdue.

Because mtDNA does not recombine, it provides an interesting context in which to detect departures from neutral expectations. If an advantageous mutation occurred in an mtDNA variant that lead to its fixation in a population, all other polymorphisms on that molecule would be fixed as a single linkage group (the hitchhiking effect; MAYNARD-SMITH and HAIGH 1974). Evidence for such a "selective sweep" was recently reported for a region of the fourth chromosome in Drosophila, a part of the genome that generally does not recombine (BERRY *et al.* 1991). Similarly, other regions of the Drosophila genome that experience low rates of recombination have been shown to exhibit low levels of nucleotide polymorphism (*e.g.*, AGUADÉ *et al.* 1989; BEGUN and AQUADRO

1991, 1993; MARTIN-CAMPOS *et al.* 1992; STEPHAN and MITCHELL 1992). While these patterns may be the result of a recent selective sweep of one or more advantageous mutations, background selection against deleterious mutations may also play a role in the reduction of linked neutral polymorphism in regions of low recombination (CHARLESWORTH *et al.* 1993).

As a first step in testing some of the neutral predictions regarding mtDNA, nucleotide sequences were determined for 1515 base pairs of the NADH dehydrogenase subunit 5 (ND5) gene in nine wild lines each of *Drosophila melanogaster* and *Drosophila simulans* from diverse localities around the world and 10 lines each of *D. melanogaster* from Arvin, California and Zimbabwe, Africa. The ND5 gene was chosen because it is the longest protein coding gene in mtDNA and is one of the most divergent genes between *D. melanogaster* and *Drosophila yakuba* (GARESSE 1988). Under neutrality, one would predict that this gene would be more variable within species which would provide the best opportunity in a coding region to detect the loss of variation from a potential selective sweep.

The data provide no evidence for departures from neutral mtDNA evolution in the samples of *D. melanogaster*. In contrast, the levels of variation in the ND5 gene of *D. simulans* are lower than expected. Although the current sequence data fail to reject neutrality at the 5% level, an analysis of existing restriction enzyme data (HALE and SINGH 1991b; SOLIGNAC *et al.* 1986) does reveal significantly less variation in *D. simulans* than expected under neutrality. Tests of the ratio of synonymous and replacement substitutions within and between species (MCDONALD and KREITMAN 1991) fail to reject neutral mtDNA evolution except in a very localized region of the ND5 gene. However, different rates of mtDNA evolution in the two species are evident in the ND5 sequences. In general, the patterns of nucleotide variation support earlier studies suggesting that the evolutionary histories of mtDNA have been very distinct in *D. simulans* and *D. melanogaster* (SOLIGNAC *et al.* 1986; BABA-AÏSSA *et al.* 1988; SATTA and TAKAHATA 1990; HALE and SINGH 1991a,b). The data illustrate the problems of inferring population histories from mtDNA variation when rates of molecular evolution differ between species and different histories of adaptive and neutral evolution have shaped molecular variation within species.

MATERIALS AND METHODS

Fly strains: Nine strains of both *D. melanogaster* and *D. simulans* were chosen from diverse localities around the world. Six lines of *D. melanogaster* were obtained from RAMA SINGH and LARRY HALE and were used in their earlier studies (HALE and SINGH 1991a,b) of mtDNA restriction site variation (ARG, La Plata, Argentina; BRO, Brownsville, Texas; CAF, Brazzaville, Congo; FRA, Villeurbanne, France; IND, Varanasi, India; JAP, Jume, Japan). Three additional lines were obtained from ANDREW CLARK (DVA, Death Valley, California; DEN, Egå, Denmark; PEN, Rothrock State Forest, Pennsylvania).

The nine strains of *D. simulans* were obtained from the National Drosophila Species Resource Center (Bowling Green State University, Bowling Green, Ohio; name of sequence, collecting locality for line and stock center number are, respectively: SIM.0HAI, Kenscoff, Haiti, 14021-1251.0; SIM.1GUY, Georgetown, Guyana, 14021-1251.1; SIM.2COL, Leticia, Colombia, 14021-1251.2; SIM.4AUS, Australia, 14021-1251.4; SIM.5PER, Lima, Peru, 14021-1251.5; SIM.6CAL, Nueva, California, 14021-1251.6; SIM.7AFR, Raratango, Cook Island, Africa, 14021-1251.7; SIM.8 MEX, Tamazunchale, Mexico, 14021-1251.8; SIM.9NGU, Gorak, New Guinea, 14021-1251.9). The collecting localities for the *D. simulans* lines are all from regions where the cosmopolitan, siII haplotype is found (BABA-AÏSSA *et al.* 1988; HALE and SINGH 1991b), and the siII haplotype was confirmed by Southern blot analyses (data not shown).

Two random samples of 10 lines each of *D. melanogaster* from Arvin, California (Arv2, Arv3, Arv8, Arv11, Arv12, Arv13, Arv15, Arv19, Arv21, Arv23) and Zimbabwe, Africa (Zim3, Zim5, Zim6, Zim7, Zim8, Zim10, Zim11, Zim18, Zim22, Zim24) were obtained from C. F. AQUADRO and D. J. BEGUN.

The "diverse" samples of both *D. melanogaster* and *D. simulans* were chosen for comparison to the widely cited data of KREITMAN (1983) and KREITMAN and HUDSON (1991), where sequences from the *Adh* region were presented for individual lines from diverse localities around the world. While such samples violate some of the assumptions of both the HKA test and the TAJIMA's test (see below), the use of these samples will provide a comparison to a number of other published HKA tests that have used the KREITMAN and HUDSON data as a reference locus (*e.g.*, BERRY *et al.* 1991; BEGUN and AQUADRO 1991, 1994). The "random" samples from Arvin, California and Zimbabwe were chosen for direct comparison to recently published data from nuclear loci for these same lines (BEGUN and AQUADRO 1993, 1994).

DNA amplification and sequencing: DNA template was prepared from a single female from each of these strains by placing the fly in a microcentrifuge tube, macerating the tissue with a sterile pipette tip holding 50 μl of 10 mM Tris (pH 8.2), 1 mM EDTA and 25 mM NaCl and proteinase K added to a concentration of 200 μg/ml (GLOOR and ENGELS 1991). The resulting homogenate was incubated at 37° for 20–30 min and the proteins were denatured by heat treatment at 95° for 2–3 min.

Sequencing templates were prepared by polymerase chain reaction (PCR) amplification of two overlapping fragments covering the 1500-bp region sequenced. A 1.39-kb fragment was amplified with primers 880R: 5′CCAAAAAGAGGCATATCACT3′ and 2230L: 5′AGCTATAGCTGCTCCTACAC3′ [number = 3′ nucleotide of primer based on the published sequence of GARESSE (1988) and the letter indicates the direction of elongation with respect to GARESSE (1988), Figure 2]. An overlapping, 1.77 kb fragment was amplified using primers 1280R 5′GACCTCCAAAATATTCTGAT3′ and 3017L 5′TAGAAGAGGTAAAATTCGAG3′. Nine additional primers spaced approximately 300 bp apart on each strand were used as internal sequencing primers on these templates. DNA was amplified in 50-μl reactions containing 1 μl of fly homogenate (see above), 5 μl of Promega Mg-free 10X buffer, 6 μl of 25 mM $MgCl_2$, 50 pmol of each primer and 2 units of Promega *Taq* polymerase. Double-stranded products were precipitated with ammonium acetate and ethanol at room temperature (KREITMAN and LANDWEBER 1989). The PCR products were used in three sequencing reactions.

The sequences were determined in both directions using dideoxy methods. The double-stranded templates were heated to 100° for 3 min and transferred to a dry ice-ethanol bath.

These "snap-cooled" templates were melted in the presence of labeling mixture with 50 pmol of sequencing primer and transferred to the chain termination reactions (U.S. Biochemical Corp. with Sequenase, version 2.0). Sequencing gels were run "short" with 1.0 M sodium acetate in the lower buffer chamber (SHEEN and SNEED 1988) resolving sequence from about 50 bp through 350 bp from the primer. "Long" gel runs without sodium acetate generally resolved sequence from 200 bp through 500 bp from the same primer.

Sequence comparisons and analyses: Sequences were aligned by eye. Heterozygosity per nucleotide site, π, was estimated from the average number of nucleotide differences between all pairs of sequences within a species (NEI 1987, Equation 10.6). A different measure of nucleotide heterozygosity, θ, was also estimated from the number of polymorphic or "segregating" sites (WATTERSON 1975; NEI 1987, Equation 10.3). Both π and θ are estimates of the neutral parameter, which for autosomal regions is $4N_e\mu$, where N_e is the effective population size and μ is the neutral mutation rate. Since mtDNA is effectively haploid and is transmitted through the female germline, θ and π are estimates of $2N_{ef}\mu$ where N_{ef} is the effective population size of females. Assuming equal numbers of males and females and no paternal leakage of mtDNA, estimates of θ and π for mtDNA should be multiplied by four for comparison to a typical autosomal region (or multiplied by three for comparison to *X*-linked genes). It should be noted that as estimates of nucleotide heterozygosity, θ and π are only valid when the samples are drawn from a randomly mating population at equilibrium for mutation and genetic drift. Since some of the sequences described here are a diverse sample from localities around the world, this assumption may well be violated (*e.g.*, Begun and Aquadro 1993).

The sequences were tested for departures from the neutral expectations using the HKA test (HUDSON *et al.* 1987), TAJIMA's (1989) test and the MCDONALD and KREITMAN (1991) test. The HKA test examines whether the levels of nucleotide polymorphism at different loci are consistent with the levels of divergence between species at these loci. This test was modified to account for the different effective population size of mitochondrial genes and other regions of the genome for which the appropriate data exist (polymorphism in *D. melanogaster* and *D. simulans* and divergence between these species; see, *e.g.* BEGUN and AQUADRO 1991; FORD *et al.* 1994). Modifications were also made to account for different sample sizes of alleles sequenced (or restriction mapped) in mtDNA and the reference loci (see, *e.g.*, BERRY *et al.* 1991), as well as for different "effective" number of nucleotides surveyed by restriction enzymes or direct sequencing (HUDSON *et al.* 1987; BEGUN and AQUADRO 1991; BERRY *et al.* 1991; FORD *et al.* 1994).

Phylogenetic analyses: Phylogenetic trees of the sequences from diverse samples of *D. melanogaster* and *D. simulans* and the homologous sequence from *D. yakuba* (CLARY and WOLSTENHOLME 1985) were reconstructed using both parsimony (PAUP 3.0s, SWOFFORD 1991) and distance methods (MEGA 1.01, KUMAR *et al.* 1993). The *D. yakuba* sequence was assigned as an outgroup in the parsimony analyses with a constraint matrix of character state transformations defined by the frequencies of nucleotide changes in the 19 sequences (State Changes option of MacClade 3.01, MADDISON and MADDISON 1992). Bootstrap replications were performed using a branch and bound algorithm. In the distance methods, pairwise distances between all sequences (n = 9 *D. melanogaster*, n = 9 *D. simulans* and the single *D. yakuba*) were determined using TAMURA's (1992) distance which adjusts for biased base composition.

RESULTS

The nucleotide positions that are variable within either *D. melanogaster* or *D. simulans* are presented in Table 1 (diverse samples of *D. melanogaster* and *D. simulans*). Table 1 also shows the nucleotide present at the homologous position in *D. yakuba* (CLARY and WOLSTENHOLME 1985). Figure 1 presents a comparison of the 1515 bp in *D. melanogaster* and *D. simulans*. The published sequence (GARESSE 1988) was not used in estimating nucleotide polymorphism or divergence. This sequence is derived from two different clones, one from a Canton-S stock (up to position 1858) and the other from a wild French stock (position 1858 through the rRNA genes; GARESSE 1988). Moreover, the published sequence has a six bp insertion relative to all other strains (plus *D. simulans* and *D. yakuba*) that adds an isoleucine and a glycine to the open reading frame (boxed nucleotides at position 1859 in Figure 2 of GARESSE 1988). It should be noted that the *Eco*RI cloning site defining the junction between the Canton-S clone and the clone of the wild French stock's mtDNA lies in this 6-bp insertion raising the possibility that this insertion is a cloning artifact. The absence of these 6 bp from the 38 new sequences reported here seems to support this possibility.

Patterns of sequence divergence: The numbers and proportions of nucleotide differences between *D. melanogaster*, *D. simulans* and *D. yakuba* are presented in Table 2. Sequence divergence was determined from all pairwise differences between the nine diverse sequences of *D. melanogaster* and *D. simulans* plus the published sequence of *D. yakuba*. Net sequence divergence between *D. melanogaster* and *D. yakuba* (10.6%) is slightly more than twice of that between *D. melanogaster* and *D. simulans* is (4.8%). The divergence between *D. simulans* and *D. yakuba* (9.3%) is less than twice of that between *D. melanogaster* and *D. simulans* is (4.8%). This suggests a faster rate of substitution along the lineage leading to *D. melanogaster* (see phylogenetic analysis below). To test this, the method of TAJIMA (1993) was applied. Using *D. yakuba* as a known outgroup and a comparison of the ND5 sequences between *D. melanogaster* and *D. simulans* [see the data in Figure 1, Table 1 and Figure 2 of GARESSE (1988)], there are 43 sites where *D. melanogaster* has a fixed, derived nucleotide [m_1 = 43; see TAJIMA (1993) expression 2a] and 24 sites where *D. simulans* has a fixed, derived nucleotide (m_2 = 24; m_2 = 27 if one considers only sim.1, sim.2, sim.6 and sim.9). This excess of derived nucleotides along the *D. melanogaster* lineage is significant [expression 4 in TAJIMA (1993); X^2 = 5.388, d.f. = 1, $P < 0.025$). However, considering m_2 = 27 with the polymorphisms in sim.1, sim.2, sim.6 and sim.9, the difference is not quite significant at the 5% level: X^2 = 3.657, d.f. = 1, $P < 0.06$.

Using a simple relative rate approach ($D_{rel} = D_{mel\text{-}yak} - D_{sim\text{-}yak}$), the D_{rel} for total sites = 10.6% − 9.3% = 1.3%, D_{rel} for synonymous sites = 7.45% and D_{rel} for replacement (nonsynonymous) sites = 0.25%. Thus, the faster rate of evolution in *D. melanogaster* appears to be greatest at synonymous sites. If synonymous substitutions are strictly neutral in the ND5 gene, this suggests a higher mutation rate in the *D. melanogaster* lineage. Faster

TABLE 1

Polymorphic nucleotide sites in *D. melanogaster* and *D. simulans*

Position (Garesse):	1358	1470	1502	1505	1523	1531	1715	1775	1913	2003	2030	2132	2161	2314	2603	2621	2625	2752	2840
Position (Figure 1):	1479	1367	1335	1332	1314	1306	1122	1062	930	840	813	711	682	529	240	222	218	91	3
Type of change:	S	R	R	S	S	R	S	S	S	S	S	I	R	R	S	S	R	R	S
Consensus:	A	A	A	A	A	G	A	C	A	A	T		T	T	A	A	T	A	A
Line																			
ARG	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
BRO	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
CAF	·	·	·	·	·	·	G	·	G	·	·	·	·	·	G	G	·	·	·
DEN	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
DVA	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
FRA	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
IND	·	·	·	·	·	·	·	·	·	·	C	·	C	C	G	·	·	·	·
PEN	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV11	·	·	·	·	·	A	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV12	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV13	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV15	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV19	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV21	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ARV23	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ZIM3	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
ZIM5	·	·	·	·	·	·	G	·	·	G	·	·	·	·	G	·	·	·	·
ZIM6	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
ZIM7	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
ZIM8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	C	·	·
ZIM10	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
ZIM11	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
ZIM18	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
ZIM22	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·
ZIM24	·	·	·	·	·	·	·	T	·	·	·	·	·	·	G	·	·	·	·
SIM.0HAI	·	·	·	·	·	·	·	·	·	·	·	TAA	·	·	·	·	·	C	G
SIM.1GUY	G	·	T	G	G	·	·	·	·	·	·	·	·	·	·	·	·	C	G
SIM.2COL	G	·	T	G	G	·	·	·	·	·	·	·	·	·	·	·	·	·	G
SIM.4AUS	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	G
SIM.5PER	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	G
SIM.6CAL	G	·	T	G	G	·	·	·	·	·	·	·	·	·	·	·	·	C	G
SIM.7AFR	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	G
SIM.8MEX	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·
SIM.9NGU	G	·	T	G	G	·	·	·	·	·	·	·	·	·	·	·	·	C	G
YAK	·	·	C	·	·	·	G	·	·	·	·	·	·	·	·	·	C	C	·

Note: The two column headings listed as "Position" indicate the base position in GARESSE (1988) (increasing to the right) and the base position indicated in Figure 1 (decreasing to the right). The nucleotides shown as "Consensus" are the consensus nucleotide for North American *D. melanogaster* and are shown on the bottom line of Figure 1. The type of change indicates a synonymous (S), a replacement (R) or an insertion (I). A dot (·) indicates the same nucleotide. The nucleotides shown for YAK are from CLARY and WOLSTENHOLME (1985).

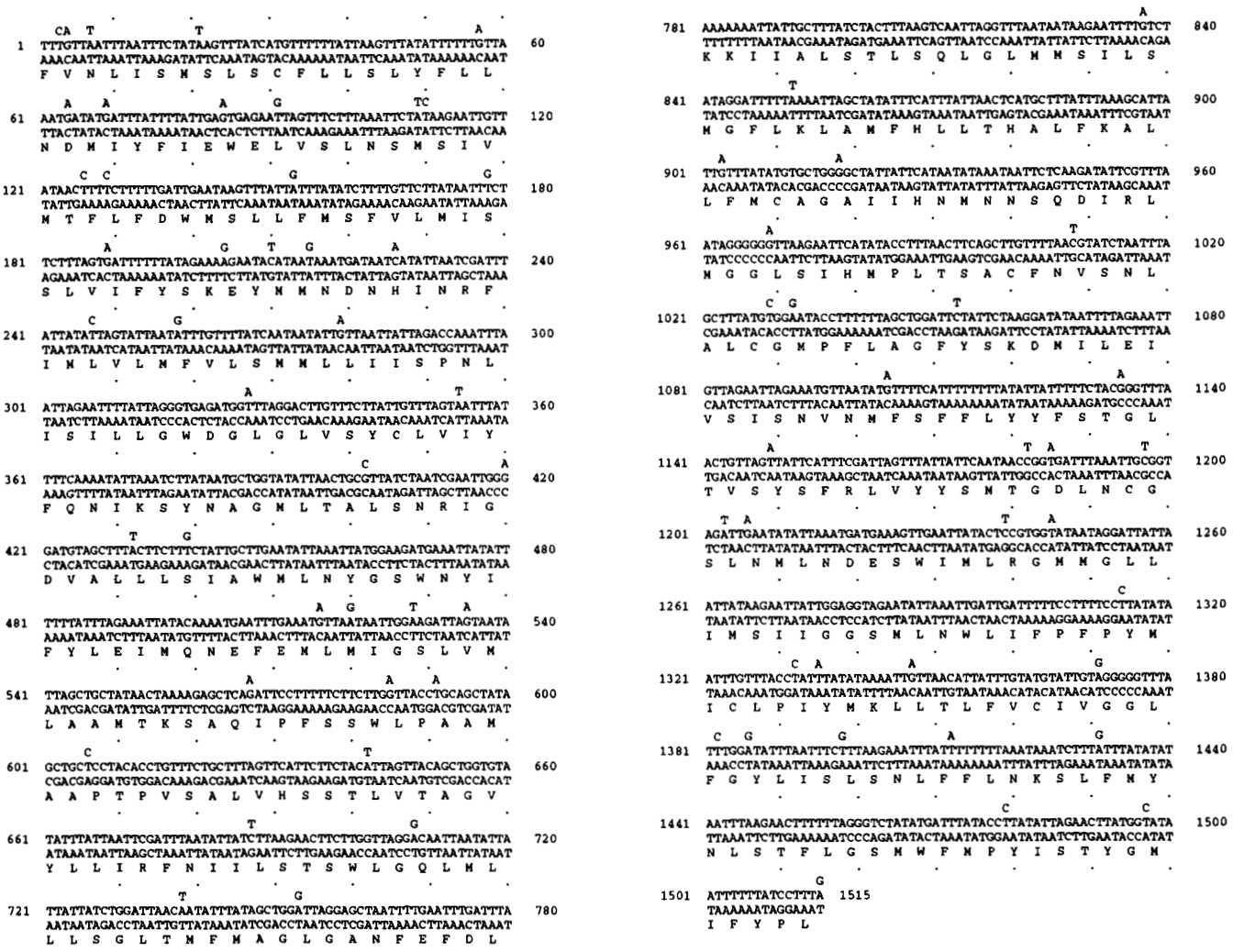

FIGURE 1.—Nucleotide and amino acid sequence of ND5 gene from *D. melanogaster* and *D. simulans*. The *D. melanogaster* sequence is presented in the opposite direction from that in GARESSE (1988) so that the bottom strand presented here is the strand in Figure 2 of GARESSE. Position 1 corresponds to position 2842 and position 1515 corresponds to position 1322 in GARESSE. Note that the duplication of six bp reported at position 1859 in GARESSE was not found in the sequences reported here. Differences between *D. melanogaster* and *D. simulans* are indicated by a single nucleotide above the sequence. The single letter amino acid code is shown below the sequence and is based on the top strand in the figure.

rates of evolution in *D. melanogaster* have been reported for some mitochondrial genes [or portions of genes: ND2, COI (SHARP and LI 1989), COI, COII, ATPase6 (KANEKO *et al.* 1993)].

The ratio of synonymous to replacement substitutions is slightly higher in the *mel-sim* comparison than in the intraspecific or *mel-yak* comparisons. This pattern is also true for the ND2 gene [ratios of synonymous to replacement substitutions for *mel* vs. *sim*, *mel* vs. *yak* and *sim* vs. *yak* are 17.9, 4.9 and 5.8, respectively; see Tables 2 and 3 of KANEKO *et al.* (1993)]. However, the COI and ATPase6 genes show very different ratios for these species comparisons (32.5, 36.1 and 73.2 for COI and 7.7, 13.3 and 17.6 for ATPase 6).

The patterns of transitional and transversional changes in the sequences indicate that transitions outnumber transversions at close evolutionary distances (more than 2:1 within species and between *D. melanogaster* and *D. simulans*). The bias toward transitions drops off to 1:1 in the comparisons to *D. yakuba*. This decrease in the transition:transversion bias has been noted in previous sequence comparisons of mtDNA in Drosophila (DESALLE *et al.* 1987a; SATTA *et al.* 1987).

The patterns of synonymous and replacement substitution between *D. melanogaster* and *D. simulans* vary considerably across the ND5 gene. Figure 2 presents the patterns of divergence in a sliding window of 100 bp that is moved across the data set (*i.e.*, along Figure 1 from right to left) one variable site at a time. The greatest amount of synonymous divergence is in a region with no amino acid divergence (the left half of Figure 2 which corresponds to the carboxy-terminal half of the ND5 protein; note that ND5 is transcribed from the minor

TABLE 2

Nucleotide differences between species

	No. of differences[a,b]			Corrected proportion of differences[a,c]		
	mel	*sim*	*yak*	*mel*	*sim*	*yak*
Total sites						
mel	2.4	72.1	152	0.0016	0.0492	0.1076
sim	69.6	2.7	134	0.0475	0.0018	0.0944
yak	150	132		0.1060	0.0926	
Synonymous sites[d]						
mel	1.7	54.4	105	0.0053	0.1860	0.4059
sim	52.6	1.9	90	0.1805	0.0057	0.3322
yak	103	88.1		0.4007	0.3265	
Replacement sites[d]						
mel	0.7	17.7	47.2	0.0006	0.0151	0.0410
sim	16.9	0.8	44.3	0.0145	0.0006	0.0386
yak	46.5	43.5		0.0404	0.0379	
Ratio (synonymous/replacement)						
mel	2.43	3.07	2.22	8.83	12.32	9.90
sim	3.11	2.38	2.03	12.45	9.50	8.61
yak	2.22	2.03		9.92	8.61	
Transitions						
mel	3	50	74	0.0016	0.0343	0.0551
sim		3	66	0.0329	0.0013	0.0483
yak				0.0535	0.0470	
Transversions						
mel	0	23	78	0.0000	0.0156	0.0543
sim		1.5	68	0.0154	0.0005	0.0475
yak				0.0543	0.0470	
Ratio (transitions/transversions)						
mel		2.17	0.95		2.1940	1.0147
sim		2.00	0.97	2.1408	2.4503	1.0161
yak				0.9853	0.9999	

[a] Within each of the nine comparisons for a given type of data, values presented are the average of all pairwise differences (or proportions) between sequences. The upper right of each 3 × 3 matrix shows the gross differences and the lower left shows the values correcting for intraspecific variation [the values on the diagonal; NEI (1987, p. 276)].

[b] Number of differences are uncorrected, observed values.

[c] Corrected proportions were calculated using the JUKES CANTOR (1969) method for total, synonymous and replacement sites and the TAMURA (1992) method for proportions of transitions and transversions.

[d] The values for corrected proportions of synonymous and replacement sites are the divergence per synonymous and replacement site, respectively.

coding strand of Drosophila mtDNA which is from right to left in Figure 2 of GARESSE (1988) and from left to right in Figure 1 of this report). In the right (amino-terminal) half of the comparison, there are some windows of 100 bp where there are as many replacement as synonymous differences between *D. melanogaster* and *D. simulans*. If the 1515 bases are divided into two regions of 758 nucleotides, the numbers of nonsynonymous and synonymous substitutions, respectively, are three and 29 in the left half, and 13 and 27 in the right half. This difference is significant (Chi-square with continuity correction = 4.24, d.f. = 1, $P < 0.039$).

Nucleotide polymorphism in *D. melanogaster*: Data on the number of segregating sites and nucleotide heterozygosity are presented in Table 3. The diverse *D. melanogaster* sequences show eight variable nucleotide positions ($S = 8$) over the 1515 bp surveyed. Three of these eight polymorphisms are replacements and two of the three replacement polymorphisms are in the sequence from India. All of the polymorphisms are transitions, five A:G and three C:T.

The estimate of nucleotide heterozygosity based on the number of segregating sites ($\theta = 0.0019$) is slightly, but not significantly, larger than that based on the number of pairwise differences between sequences ($\pi = 0.0016$). This is to be expected in a diverse sample with one or a few sequences drawn from different geographic regions since such a sample is likely to have an excess of unique or low frequency polymorphisms. This is reflected in the negative value of TAJIMA's *D* statistic. The two halves of the data set (N-terminal and C-terminal; see sequence divergence above) show patterns very similar to the entire 1515-bp region. HALE and SINGH (1991a) report an average pairwise difference (π) of 0.0053 estimated from 4-, 4.5- and 6-cutter restriction analyses of entire mtDNAs in 144 lines of *D. melanogaster*. This larger estimate of π presumably reflects the greater number of collecting localities and flies surveyed by HALE and SINGH (1991a).

The random samples of *D. melanogaster* from Arvin, California, and Zimbabwe, Africa, show, as expected, considerably less variation than the diverse sample (see

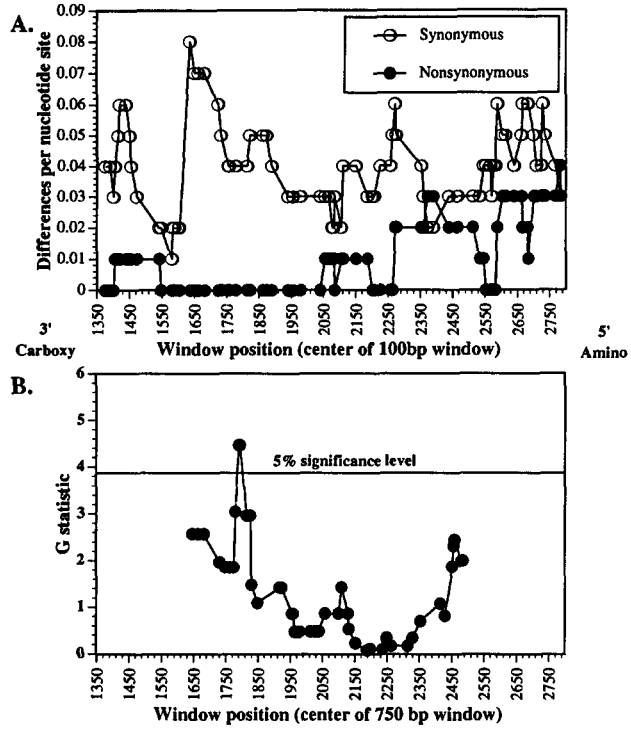

FIGURE 2.—Patterns of synonymous and nonsynonymous (replacement) substitution between *D. melanogaster* and *D. simulans* across the ND5 gene. (A) A window spanning 100 bp was moved across the data set (from right to left of Figure 1) and the number of silent and replacement differences between the two species per nucleotide site was tabulated. The left part of panel A shows low levels of amino acid replacement substitution but the highest level of synonymous substitution. Toward the right half of panel A the levels of synonymous and replacement substitutions are more similar. These differences between the two halves of the data are significant (see RESULTS). (B) A window of 750 bp was moved across the data set (across Table 1 from left to right) and the *G*-statistic was calculated for a MCDONALD-KREITMAN test comparing the numbers of synonymous and replacement polymorphisms within species to the number of synonymous and replacement substitutions between species. The data are generally consistent with a neutral model, but in one region of the gene there is a window of 750 bp that show an excess of amino acid replacement polymorphism. This excess lies in the region where the level of amino acid substitution between species is lowest.

Table 3). There is a single nonsynonymous polymorphism in one line of the Arvin sample while the other nine lines from Arvin, and the three additional U.S. lines in the diverse sample, are identical. The 10 sequences from Zimbabwe, Africa, show four segregating sites, one of which is nonsynonymous. Thus, the U.S. sample has lower levels of mtDNA polymorphism ($\pi = 0.0001$) than the Zimbabwe sample ($\pi = 0.0006$), a pattern that was observed by HALE and SINGH (1991a) for restriction analyses of mtDNAs from North American vs. Euro-African samples. Similarly, data from several nuclear loci indicate that nucleotide polymorphism in the Arvin sample is lower than in the Zimbabwe sample (BEGUN and AQUADRO 1993, 1994).

Population structure in *D. melanogaster*: There is clear differentiation between the Arvin, California, and Zimbabwe samples indicated by two fixed or nearly fixed positions (see Tables 1 and 4). At the two positions where these samples are differentiated (1715 and 2603) the diverse sample of *D. melanogaster* also shows evidence for differentiation between the United States and Europe/Africa (see Table 1). Measures of population subdivision indicate that the differentiation between the United States and Africa for mtDNA is substantial and significant (see Table 4). While a comparable statistic for differentiation between North America and Africa was not presented by HALE and SINGH (1991a), it is clear from their data that this distinction is strong. Restriction haplotype #7 was found in 47 of 52 (90%) lines from North America, but was found only once in 36 lines (3%) from the Euro-African region (Table 4 of HALE and SINGH 1991a).

The differentiation of mtDNA between the United States and Africa is higher than that for nuclear gene regions (BEGUN and AQUADRO 1993, 1994). This is to be expected given the lower effective population size of mtDNA. Assuming equal numbers of males and females, the effective population sizes of mitochondrial and *X*-linked genes should differ by a factor of three. Notably, the F_{ST} values reported by BEGUN and AQUADRO (1993) for *X*-linked genes that are experiencing "normal" levels of recombination are quite close to one-third the F_{ST} for mtDNA (see Table 4). These samples are directly comparable since the mitochondrial ND5 sequences are derived from some of the same female lines of *D. melanogaster* that BEGUN and AQUADRO used to estimate F_{ST} for these *X*-linked genes. The ratios of the F_{ST} values would be expected to equal the inverse of the ratios of the effective population sizes for two different genetic markers only if the combined term Nm (effective population size of marker times the migration rate) in the expression $F_{ST} = 1/(Nm + 1)$ is large ($Nm \sim 5$–10). Given the large population size ($\sim 10^6$ AQUADRO *et al.* 1988) and presumed high migration rate (SINGH and RHOMBERG 1987; DAVID and CAPY 1988; HALE and SINGH 1991a,b) in *D. melanogaster*, an $Nm > 5$ may well hold.

In contrast, genes at the tip of the *X* chromosome which experience low levels of recombination show much higher F_{ST} values between the Arvin and Zimbabwe samples (BEGUN and AQUADRO 1993). This additional level of population differentiation at more distal portions of the *X* chromosome is presumably due to the effects of independent hitchhiking events of selectively advantageous or deleterious mutations in these regions of low recombination [BEGUN and AQUADRO (1993); see also STEPHAN and MITCHELL (1992)]. To the extent that the differentiation between North America and Zimbabwe in *D. melanogaster* at *X*-chromosome loci experiencing "normal" levels of recombination (BEGUN and

TABLE 3

Estimates of nucleotide variation at the ND5 locus

Species	Sample (n)	No. Sites	No. variable sites	θ^a	π^b	TAJIMA's D^c
D. melanogaster	Diverse (9)	1515	8	0.0019 ± 0.0007	0.0016 ± 0.0013	-0.855
D. melanogaster	Arvin, CA (10)	1515	1	0.0002 ± 0.0001	0.0001 ± 0.0003	-1.112
D. melanogaster	Zimbabwe (10)	1515	4	0.0009 ± 0.0003	0.0006 ± 0.0007	-1.245
D. melanogaster	N-term, diverse (9)	758	4	0.0019 ± 0.0007	0.0016 ± 0.0015	-0.689
D. melanogaster	C-term, diverse (9)	758	4	0.0019 ± 0.0007	0.0016 ± 0.0015	-0.843
D. simulans	Diverse (9)	1515	6	0.0015 ± 0.0005	0.0018 ± 0.0010	0.907
D. simulans	N-term, diverse (9)	758	4	0.0010 ± 0.0004	0.0006 ± 0.0009	-1.363
D. simulans	C-term, diverse (9)	758	2	0.0019 ± 0.0007	0.0029 ± 0.0015	2.073*

[a] Estimate is from WATERSON (1975) and error assumes no recombination [see NEI (1987, Equations 10.2 and 10.3)].
[b] Estimate is from NEI (1987, expression 10.6) and error is the square root of expression 10.9 (NEI 1987). Both estimates are of 2 $N\mu$ where N is the effective number of females and μ is the neutral mutation rate.
[c] From TAJIMA (1989) expression 38; * indicates a departure from neutral expectations at the 5% level.

TABLE 4

Estimates of population subdivision in *D. melanogaster*

Samples compared	F_{ST}[a]	X^{2}[b]	d.f.	$P<$
mtDNA-ND5 sequences				
Arvin, California *vs.* Zimbabwe[c]	0.74	17.40	5	0.005
United States *vs.* Euro-Africa[d]	0.73	16.38	4	0.005
X-linked loci				
Pgd[e]	0.25			
G-6pd[e]	0.30			
vermilion[e]	0.32			
white[e]	0.28			
Mean for *X*-linked loci	0.29			
mtDNA:*X* ratio	2.57			

[a] F_{ST} estimated using the method of HUDSON *et al.* (1992).
[b] Test of the null hypothesis that the genetic distance between populations $Dij = 0$. See NEI (1987, p. 227).
[c] Ten sequences from each population; data from Table 2.
[d] Four American lines (ARG, BRO, DEA, PEN) *vs.* three Euro-African lines (CAF, FRA, DEN); data from Table 1.
[e] Data are for United States *vs.* Zimbabwe samples analyzed using four-cutter analysis (BEGUN and AQUADRO 1993).

AQUADRO 1993) reflects neutral levels of population differentiation, the above comparisons of F_{ST} values for nuclear and mitochondrial markers suggest essentially neutral population subdivision of mtDNA in *D. melanogaster*. However, additional samples and better estimates of Nm for each marker are needed before firm conclusions of this sort can be made.

Nucleotide polymorphism in *D. simulans*: The diverse *D. simulans* sequences show six polymorphic nucleotide sites, and one three base pair insertion (TAA at position 2132) that adds a leucine to an (A + T)-rich stretch coding for leucines and a methionine. Only base-pair substitutions were considered in estimating nucleotide heterozygosity. Of the six single base polymorphisms, two are nonsynonymous transversions (A:T at position 1502 and C:A at position 2752) and four are synonymous A:G transitions. Four of the polymorphisms lie in the C-terminal half of the sequence and are at high frequency in the sample (4/9), defining two lineages within *D. simulans*. Three of these four polymorphisms are closely spaced in the sequence (positions 1502, 1505 and 1523). None of the polymorphisms are shared between *D. melanogaster* and *D. simulans*. At position 2840 the polymorphism within *D. simulans* is either the retention of an ancestral state (A in both *D. melanogaster* and *D. yakuba*) or a back mutation to A. The latter explanation is more parsimonious given the many fixed differences between *D. melanogaster* and *D. simulans*.

The two estimates of nucleotide heterozygosity show slightly different patterns in the two species, θ being lower, and π being higher in *D. simulans* than in *D. melanogaster* (Table 3). This difference reflects the greater number of high frequency polymorphisms in *D. simulans*, resulting in a positive value of TAJIMA's *D* statistic. While it is apparent that the levels of nucleotide variation in *D. simulans* are not significantly lower (see standard errors in Table 3), this is distinct from the patterns of variation in nuclear genes where *D. simulans* generally shows more nucleotide variation then *D. melanogaster* (AQUADRO *et al.* 1988; BEGUN and AQUADRO 1991; HEY and KLIMAN 1993). It should be noted, however, that *D. simulans* has been reported to show lower levels of nucleotide heterozygosity than *D. melanogaster* in regions of the genome experiencing low rates of recombination (BEGUN and AQUADRO 1991). In both species, hitchhiking of selectively advantageous or deleterious mutations has been invoked to account for the reductions of nucleotide heterozygosity relative to loci experiencing "normal" rates of recombination (BEGUN and AQUADRO 1991; BERRY *et al.* 1991).

Population structure in *D. simulans*: In contrast to *D. melanogaster*, where fixed differences in mtDNA are evident between continents, the polymorphisms in *D. simulans* show little geographic structure Notably, the two distinct types of mtDNA evident within the *D. simulans* sample are found in both the New and Old World (see *Phylogenetic analysis* below). This suggests that the differentiation of these two mtDNAs occurred prior to the expansion of *D. simulans* to a cosmopolitan distribution. The measure of F_{ST} proposed by HUDSON *et al.* (1992) actually gives a negative value (−0.286) for the ND5 sequences in *D. simulans* when the six New World lines are compared to the three lines from Africa, New

TABLE 5

HKA statistics for sequence and restriction-site polymorphism and divergence at nuclear and mitochondrial genes

Nuclear Locus (n, bp surveyed)	mtDNA (n, bp surveyed)	S_1/ES_1	S_2/ES_2	D_1/ED_1	D_2/ED_2	T	X^2	$P <$
Polymorphism in *D. melanogaster*								
Adh 5', worldwide (11, 1243)[a]	ND5, worldwide, silent sites (9, 285)	30/26.0	5/4.0	77/81.0	54/50.0	8.15	2.73	0.1
Adh 5', worldwide (11, 1243)[a]	ND5, worldwide, all sites (9, 1515)	30/25.6	8/5.4	77/81.4	72/67.6	8.30	2.21	0.3
Adh 5', U.S. (5, 1243)[a]	ND, Arvin, Calif., silent sites (10, 285)	18/14.6	0/3.4	77/80.4	54/50.6	10.43	2.47	0.2
Adh 5', U.S. (5, 1243)[a]	ND5, Arvin, Calif., all sites (10, 1515)	18/14.5	1/4.5	77/80.5	72/68.5	10.55	1.86	0.4
Adh 5', U.S. (5, 1243)[a]	ND5, U.S., silent sites (13, 285)	18/14.4	0/3.6	77/80.6	54/50.4	10.65	2.77	0.1
Adh 5', U.S. (5, 1243)[a]	ND5, U.S., all sites (13, 1515)	18/14.2	1/4.7	77/80.8	72/68.2	10.83	2.17	0.2
Pgd, Arvin, Calif., (34, S:1248, D:4203)[b]	ND5, Arvin, Calif., all sites (10, 1515)	8/7.0	1/2.0	196/197.0	72/71.0	33.09	0.45	0.6
Pgd, Arvin, Zimbabwe, (48, S:1248, D:4203)[b]	ND5, Arvin, Zimbabwe, all sites (10, 1515)	9/10.2	4/2.8	196/194.8	72/73.2	23.65	0.39	0.8
period, worldwide, all sites (6, S:1868, D:1838)[c]	ND5, worldwide, all sites (9, 1515)	29/25.8	8/11.3	65/68.3	72/68.8	5.15	0.51	0.6
period, U.S., all sites (4, S:1868, D:1838)[c]	ND5, U.S., all sites (13, 1515)	13/8.9	1/5.1	65/69.1	72/67.9	13.40	2.60	0.2
Polymorphism in *D. simulans*								
period, worldwide, all sites (6, S:1871, D:1838)[c]	ND5, worldwide, all sites (9, 1515)	54/43.6	6/16.4	65/75.4	72/61.6	3.03	3.07	0.09
period, worldwide, all sites (6, S:1871, D:1838)[c]	RE, worldwide (69, S:>416; D:230)[d,e]	54/48.7	4/9.3	65/70.3	14/8.7	2.36	4.76	0.04

All HKA tests involve comparisons of the polymorphism and divergence at a nuclear locus with polymorphism and divergence in mtDNA. All tests are within and between *D. melanogaster* and *D. simulans*. HKA tests were modified to account for the differences in effective population sizes between the nuclear locus and mtDNA. For *Adh*-mtDNA comparisons, a scaling factor of four was used; for comparisons using *period* or *Pgd* and mtDNA a scaling factor of three was used, reflecting the nuclear-to-mtDNA ratios of the expected values of $S/\Sigma_{i=1}^{n-1}(1/i)$. The first two columns show the locus, a description of the geographic nature of the sample and the sites used (n = number of alleles sampled, bp surveyed = number of bases sequenced or restriction mapped). All tests considered "effective" number of nucleotides [see HUDSON *et al.* (1987) and MATERIALS AND METHODS]. When the numbers of effective sites differed for estimates of polymorphism and divergence, these numbers are listed next to S: or D: respectively, in the first or second column. ND5 refers to the mtDNA sequences reported here; *RE* refers to restriction enzyme data from the cited reference. S_1 is the observed, and ES_1 is the expected, number of segregating sites at the nuclear locus. S_2/ES_2 are the comparable values for mtDNA. D_1 is the observed, and ED_1 is the expected, number of differences between *D. melanogaster* and *D. simulans* at the nuclear locus. D_2/ED_2 are the comparable values for mtDNA. Test statistics shown are distributed approximately as a chi-square with one degree of freedom (5% significance at 3.84). T = expected divergence time in units of $2N$ generations.

[a] Data from KREITMAN and HUDSON (1991).

[b] Data from BEGUN and AQUADRO (1994).

[c] Data from KLIMAN and HEY (1993).

[d] Polymorphism data from HALE and SINGH (1991b); 104 restriction sites were detected, some of which were six-cutters; the number listed here for sites surveyed is assuming all were four cutters and is thus a conservative estimate.

[e] Divergence data from SOLIGNAC *et al.* (1986).

Guinea and Australia. This stems from a higher average number of pairwise differences between lines within the New World than between lines from the two "populations." This is of little significance because the sample of *D. simulans* is not appropriate for calculating F_{ST}.

HALE and SINGH (1991b) also found two frequent restriction haplotypes in *D. simulans*, plus two additional haplotypes at lower frequency. While they reported a relatively high G_{ST} value (0.55) among 69 lines from around the world, they found no endemic haplotypes and the two common haplotypes were often present in the same populations. Thus the sequence data presented here are consistent with earlier restriction studies (HALE and SINGH 1991a,b) indicating lower levels of mtDNA variation, and less population structure of mtDNA in *D. simulans* than in *D. melanogaster*.

Tests of neutrality: The method of HUDSON *et al.* (1987) was used to test for departures from neutrality. None of the samples of mtDNA sequence from *D. melanogaster* reported here deviate significantly from neutral expectations when compared to different regions of the nuclear genome (see Table 5). The HKA test of the ND5 sequences in *D. simulans* compared to the six sequences from the *period* locus reported by KLIMAN and HEY (1993) is notable in that the mtDNA sequences show fewer than half the expected number of segregating sites. Although the HKA statistic is not significant, a *G*-test using the same data but adjusting the polymorphisms in mtDNA for the difference in effective population size ($3 \times S_2 = 18$; see Table 5) is highly significant ($G = 15.16$, $P < 0.0001$; note that such a test does not adjust for the possibility that polymorphisms are not independent of one another). Using the published data on restriction site variation in *D. simulans* (HALE and SINGH 1991b) and restriction site differences between *D. simulans* and *D. melanogaster* (SOLIGNAC *et al.* 1986), mtDNA variation in *D. simulans* does show a significant HKA test statistic in comparison to the *period* sequence reported by KLIMAN and HEY (1993) (see Table 5).

It is important to note that the assumption of a randomly mating population is violated in the HKA tests using the "diverse" samples of *D. melanogaster* (and the *D. simulans* sample). Previous work has shown that population subdivision of *D. melanogaster* exists among different geographic regions of the globe (BEGUN and AQUADRO, 1993, 1994; HALE and SINGH 1991a). This may also be true for *D. simulans* (HALE and SINGH 1991b; but possibly to a lesser extent than in *D. melanogaster* as indicated above). Hence, the HKA results from these samples should be interpreted with caution since these diverse, overdispersed samples will typically show more polymorphism that random samples from single localities (*cf.* Table 3).

The MCDONALD and KREITMAN (1991) test examines whether the ratio of silent to replacement substitutions is the same within and between species. Applied to the current data, this approach fails to reject the neutral model. For 29 *D. melanogaster* sequences *vs.* nine *D. simulans* sequences there are 15 fixed replacement (FR), 54 fixed synonymous (FS), seven polymorphic replacement (PR) and 11 polymorphic synonymous (PS) sites ($G = 2.08$, $P < 0.5$). When the *D. simulans* clade containing sim.1, sim.2, sim.6 and sim.9 (see Table 1 and Figure 3) is used as the interspecific comparison, the numbers suggest an excess of polymorphic replacements (FR = 16, FS = 58, PR = 6, PS = 7), although the test is not significant ($G = 3.18$, $P < 0.1$).

There is, however, considerable spatial variation in the outcome of this test (Figure 2). If one considers a sliding window of 750 bp that is moved across the data set, the carboxy-terminal half of the data (left half of Table 1; 29 *D. melanogaster*, 9 *D. simulans*) does indicate a region where the ratios of synonymous and replacement polymorphism and divergence are not consistent with neutral expectations (see Figure 2B). This deviation is significant (FR = 2, FS = 25, PR = 4, PS = 7; $G = 4.47$, $P < 0.05$). As is evident from this figure, the significance of this one peak hangs on a single variable nucleotide site since the significance drops off rapidly as the window moves along. By comparing the top and bottom halves of Figure 2, it is evident that the lowest level of replacement substitution between species overlaps the peak in the *G* statistic of the MCDONALD-KREITMAN test. Hence the significant *G* statistic stems from the fact that the level of replacement polymorphism does not drop in this window of reduced amino acid divergence. A pattern of excess replacement polymorphism has been observed in the mitochondrial ATPase 6 gene in *D. melanogaster* (KANEKO *et al.* 1993) and at the mitochondrial ND3 gene in mice (NACHMAN *et al.* 1994). Two plausible explanations for these observations are that amino acid polymorphisms are slightly deleterious or that selective constraints on protein sequences have been relaxed within species (NACHMAN *et al.* 1994).

The Tajima test examines whether the average number of pairwise nucleotide differences between haplotypes in a sample (π) is larger or smaller than that expected (θ) from the observed number of segregating sites. Under the assumptions of a random mating population and an infinite sites equilibrium model of neutral DNA evolution, the difference between π and θ is expected to be zero (*i.e.*, TAJIMA's $D = 0$). A positive value of TAJIMA's *D* indicates possible balancing selection or population subdivision. A negative value of TAJIMA's *D* suggests recent directional selection, a population bottleneck or background selection of slightly deleterious alleles (TAJIMA 1989). With the exception of the carboxy-terminal half of the *D. simulans* sequences, none of the samples deviate significantly from neutral expectations (see Table 3). TAJIMA's *D* is negative in the *D. melanogaster* samples while it is positive in *D. simulans*. As noted above this reflects the presence of polymorphisms in the *D. melanogaster* samples that are observed in a single line (six out of eight segregating sites

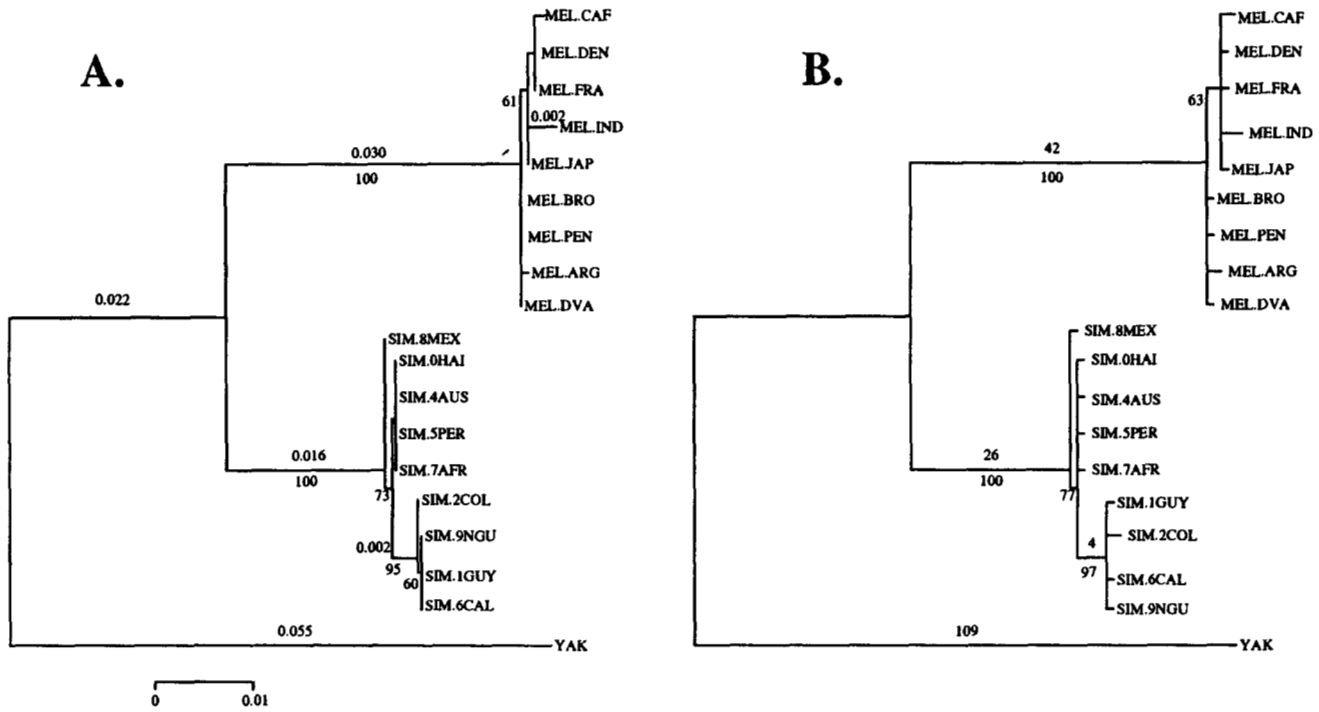

FIGURE 3.—Phylogenetic analysis of ND5 sequences. (A) A neighbor joining tree of the nine sequences each from the diverse samples of *D. melanogaster* and *D. simulans*, plus the homologous published sequence from *D. yakuba* (CLARY and WOLSTENHOLME 1985). The tree is based on TAMURA (1992) distances of all variable sites. Branch lengths are shown above the branches, bootstrap percentages from 100 bootstrap replications are shown below the branches. (B) A 50% majority rule consensus tree based on a 100 bootstrap replications of a branch an bound search in PAUP (SWOFFORD 1991). Numbers above the branches are the number of unambiguous character state changes along the branch and the numbers below the branches are bootstrap percentages. The placement of Arvin, California, and Zimbabwe sequences within *D. melanogaster* would be the same as for MEL.BRO and MEL.CAF, respectively. Tree length = 193, consistency index = 0.97 (0.95 excluding uninformative characters).

have a frequency of 1/9 in the diverse *D. melanogaster* sample while only two of the six polymorphisms in *D. simulans* are present at a frequency of 1/9).

With the exception of the sample from Arvin, California, the values of TAJIMA's *D* should be treated with caution since the diverse sequences from *D. melanogaster* and *D. simulans* are worldwide samples. As suggested by data presented above (Tables 1 and 4) and elsewhere (BEGUN and AQUADRO 1993, 1994), TAJIMA's (1989) assumption of a random mating population may well be violated for *D. melanogaster*. There is some evidence that mating may not be random within the Zimbabwe population due to sexual isolation between apparent strains or "incipient species" within *D. melanogaster* (C.-I. WU, personal communication). Since the data from Zimbabwe do not depart from a neutral model, it seems unlikely that nonrandom mating is clouding this result. However, this possibility should be considered in sequence surveys using samples from this locality. While it is not clear that the assumption of panmixia is violated for the diverse sample of *D. simulans*, the significant TAJIMA's *D* in the carboxy half of the data set should also be treated with caution since this reflects an arbitrary partitioning of the data set.

Phylogenetic analysis: Some of the issues presented in the preceding sections are evident in the structure of phylogenetic trees obtained from either parsimony or distance methods. Figure 3A presents a neighbor-joining (NJ) tree derived from TAMURA (1992) distances which adjusts for the strong bias in A + T content (MEGA) (see KUMAR *et al.* 1993). Figure 3B presents the 50% majority rule consensus tree obtained from a bootstrap analysis using a branch and bound search (PAUP version 3.0s; SWOFFORD 1991). The faster rate of mtDNA evolution in *D. melanogaster* is evident in the evolutionary distances (numbers above branches in Figure 3A) and in the number of substitutions unique to specific lineages (numbers above branches in Figure 3B). Bootstrap analysis with both branch and bound searches and the NJ approach indicate a New World-Old World split in the *D. melanogaster* sequences (numbers below branches). These analyses show strong support for the phylogenetic split within *D. simulans* but it is evident from the geographic origin of the lines of *D. simulans* present in either clade that this split does not correlate with a New World-Old World split.

DISCUSSION

General considerations: The motivation for this study was to test the neutrality of mtDNA by applying several statistical tests to estimates of nucleotide polymorphism and divergence. The data presented here for the ND5 gene from both worldwide and local, random samples of

D. melanogaster provide evidence that mtDNA evolution is consistent with neutral models in this species. In a comparable worldwide sample of *D. simulans*, the ND5 sequences provide no statistical evidence for deviation from neutral molecular evolution, although the levels of mtDNA variation are less than half that of a neutral expectation based on comparisons to nucleotide variation at the *period* locus. Using previously published data from restriction surveys (HALE and SINGH 1991b; SOLIGNAC *et al.* 1986), there is significantly less mtDNA variation in *D. simulans* when compared to variation at the *period* locus in this species (see Table 5).

The overall patterns of synonymous and replacement polymorphism and divergence are also consistent with neutral expectations (*cf.* MCDONALD and KREITMAN 1991), but a neutral model can be rejected and accepted in different regions of the ND5 gene. The current analyses of mtDNA variation and divergence thus provide evidence for both neutral and non-neutral evolution in Drosophila and indicate that the evolutionary histories of mtDNA have been very different in *D. melanogaster* and *D. simulans* (*cf.* HALE and SINGH 1991a,b). It is notable that the sequence data reveal different rates of mtDNA evolution for this species pair and very different patterns of population and phylogenetic structure within species. Evolutionary inferences from mtDNA (or other loci) commonly use interspecific comparisons to date evolutionary events within species. The reliability of this endeavor depends critically on knowledge about differential rates of evolution and distinct genealogical relationships of haplotypes within both species under comparison. A faster rate of mtDNA evolution in *D. melanogaster* and a selective sweep of mtDNA variation in *D. simulans* will lead to incorrect inferences regarding the "mitochondrial eve" in the latter species (see below).

Contrasting histories of *D. melanogaster* and *D. simulans*: The generally higher levels of nucleotide variation at nuclear loci in *D. simulans* than in *D. melanogaster* has been taken as evidence for a greater historical effective population size in *D. simulans* (AQUADRO *et al.* 1988). The nucleotide data are distinct from data on allozymes (SINGH and RHOMBERG 1987) which indicate *less* variation in *D. simulans* than in *D. melanogaster*. If allozymes are more visible to selection, the differences between the nuclear DNA and allozyme data would be consistent with the notion that natural selection is more effective in larger populations. However, this assumes that the majority of the polymorphisms at allozyme loci are not being maintained by balancing selection, in which case a larger effective population size would result in *more* variation at allozyme loci in *D. simulans* (as measured by heterozygosity; see AQUADRO *et al.* 1988).

In light of the patterns for nuclear loci, the reduced level of mtDNA variation in *D. simulans* further indicates a population history distinct from that of *D. melanogaster*. There are both neutral and non-neutral explanations for the apparently inconsistent pattern of mtDNA variation in *D. simulans*. HALE and SINGH (1991b) proposed a scenario where low levels of migration among populations in *D. simulans* could affect nuclear and mtDNA variation differently. At the appropriate level, limited migration could allow nuclear variation to be maintained, but given the lower effective population size of mtDNA, its effective migration rate could be sufficiently low to allow drift within local populations to reduce mtDNA variation. The effective population sizes of autosomal and mitochondrial genes should differ by a factor of four. Given the evidence for paternal leakage in *D. simulans* (KONDO *et al.* 1990) and a higher variance in mating success in males than in females (BATEMAN 1948) the differences in effective population sizes is probably less than a factor of four (or less than three for *X*-linked loci). Acknowledging that evidence for panmixia of nuclear genes and subdivision of mtDNA has been reported in *Drosophila mercatorum* (DESALLE *et al.* 1987b), the conditions for the appropriate balance of migration and drift seem a bit restrictive to be the sole explanation for such a clear reduction of mtDNA variation below neutral expectations in *D. simulans*.

An alternative explanation for the reduced mtDNA variation in *D. simulans* is one or more selective sweeps of an advantageous mutation. While such an hypothesis is appealing for non-recombining mtDNA, the presence of two distinct lineages of mtDNA within the sample of *D. simulans* (see Figure 3) requires a more complex explanation than a simple selective sweep (note that the lines of *D. simulans* studied here belong to the cosmopolitan siII mitochondrial haplotype; see MATERIALS AND METHODS and SOLIGNAC *et al.* 1986). Since the two mtDNA lineages are found in widely separated geographic locations, they presumably existed prior to the expansion of *D. simulans* out of Africa (HALE and SINGH 1991b; DAVID and CAPY 1988).

A selective sweep may have occurred before the divergence of the two lineages, and the low level of variation reflects that which accumulated between subpopulations during a period of geographic fragmentation of the ancestral African population. Alternatively, during a period of geographic separation of the ancestral African population of *D. simulans*, purifying selection may have eliminated all but one of the existing mtDNA haplotypes in each of the two subpopulations. A third possibility is that two independent mutations drove selective sweeps of distinct mtDNA haplotypes in each of two geographically isolated subpopulations. A further, not mutually exclusive, possibility is that the patterns of mtDNA variation in *D. simulans* have been altered by the presence of a Wolbachia endosymbiont (HOFFMANN *et al.* 1986; TURELLI and HOFFMANN 1991). Since this microorganism is involved in cytoplasmic incompatibility, it seems likely that it could contribute to a departure from strictly neutral evolution of mtDNA. Given that associations between cytoplasmic incompatibility types and mtDNA haplotypes have been reported in *D. simulans* (HALE

and HOFFMANN 1990; MONTCHAMP-MOREAU *et al.* 1991), *Wolbachia* endosymbionts may have played a role in the divergence of mtDNA haplotypes as well as contributing to a pattern of non-neutral mtDNA evolution (but see ROUSSET *et al.* 1992). While it is misguided to invoke an explanation that is exclusively neutral or adaptive in nature, since the pattern of mtDNA variation in *D. simulans* has the footprint of a mixture of evolutionary forces, it does appear that selection has been one of these forces.

Sequential periods of neutral and non-neutral mtDNA evolution: A traditional view of mtDNA evolution is one where new mutations define a bifurcating and hierarchically nested tree of haplotypes. Under neutrality the amount of variation within a species will reflect an equilibrium of mutation and genetic drift. Because mtDNA is a single linkage group it seems likely that periodic selective sweeps have occurred in the evolutionary history of mtDNAs in many species, analogous to periodic selection of new mutations in bacterial chemostats (ATWOOD *et al.* 1951). Are these types of events frequent enough that sequence analyses will be able to detect them in mtDNA?

Our ability to detect selective sweeps by sampling from a natural population will be influenced, among other factors, by the amount of time taken for the variant to reach fixation (*i.e.*, fitness differential between the "new" and "old" haplotypes) and the amount of time elapsed since fixation was reached. This is a result of additional mutations occurring on the genetic background of the new haplotype both during the polymorphic phase and after the fixation event (assuming an adaptive mutation). While it may take many generations for a population to return to equilibrium after a selective sweep, the accumulation of new polymorphisms will slowly dilute our ability to detect such events using statistical tests based on nucleotide variation. Moreover, the relatively small size of the mitochondrial genome (about 16–18 kb containing 13 protein coding genes, 22 transfer RNAs and 2 ribosomal RNAs) presents a small "target" on which adaptive or deleterious (see CHARLESWORTH 1993) mutations might occur. As a result, a small number of mutations that might drive a selective sweep will "hit" mtDNA per unit time, further reducing the chances that the footprint of reduced variation will be detectable. That virtually all regions of the nuclear genome of Drosophila experiencing low rates of recombination appear to exhibit reduced variation (*e.g.*, AGUADÉ *et al.* 1989; MARTIN-CAMPOS *et al.* 1992; STEPHAN and MITCHELL 1992; BERRY *et al.* 1991; BEGUN and AQUADRO 1991, 1992, 1993) may be attributed to the presence of many genes in these regions which are effectively large targets for non-neutral mutations by virtue of their linkage relationship (see Figure 6 in AQUADRO and BEGUN 1993).

Even if selective sweeps in mtDNA are rare and hard to detect statistically, one prediction of recurrent adaptive fixations is that mtDNA should exhibit fewer segregating sites than that expected under neutrality. As indicated in Table 5, this prediction appears to be upheld qualitatively as most of the HKA tests from both *D. melanogaster* and *D. simulans* reveal that the observed mtDNA variation is less than that expected under neutrality.

If the amount of observed mtDNA variation in *D. simulans* reflects the accumulation of new polymorphisms after a selective sweep, a rough estimate can be obtained of the number of generations, t, since the elimination of variation from a selective sweep. Following the coalescent approach taken by BERRY *et al.* (1991), the total time on a genealogy is $T_{tot} = nt$ where n is the sample size. The expected number of polymorphisms is $S = T_{tot}\mu k$, where μ is the mutation rate per nucleotide site per generation and k is the number of effectively silent sites. In *D. simulans*, $S = 4$ silent sites, $k = 285$, $n = 9$ and μ is taken as 1×10^{-8} substitutions per site per year (SHARP and LI 1989 report a range of $8–16 \times 10^{-9}$ for several genes in mtDNA). Assuming 10 generations per year, t is estimated to be 1.5×10^6 generations ago. Alternatively one can use a molecular clock approach and apply the rate of evolution at synonymous sites between *D. melanogaster* and *D. simulans* to the amount of synonymous divergence between the two lineages of mtDNA within *D. simulans*. With 18.6% synonymous divergence between species (Table 2) in approximately 3 million years (see HEY and KLIMAN 1993), the three synonymous substitutions between the two mtDNA lineages within *D. simulans* (0.78% diverged; see Figure 3) suggests a divergence time of 1.3×10^5 years. A value of 1.6×10^5 is obtained using total sites. Considering 10 generations per year, these estimates agree well with the coalescent approach. These calculations *assume* neutrality. The central point however, is that these are times back to the common ancestor of DNA sequences which may be decoupled from the evolutionary history of organisms when the evolution of DNA departs from neutrality (see Figure 4).

Gene trees and organismal histories: Two general types of deviations from neutral evolution (Figure 4A) within species can lead to errors in dating intraspecific events. Under balancing selection (or some other form of diversifying selection), one will observe greater amount of nucleotide variation than expected. When the interspecific calibration is applied to this excess variation, the inferred common ancestor of the sequences is older than the actual organismal common ancestor (see Figure 4B). In contrast, following a selective sweep, the common ancestor of alleles is moved closer to the recent, and a naive interspecific calibration leads to a error in the assignment of an organismal common ancestor (*e.g.*, a mitochondrial *vs.* and organismal "eve"; see Figure 4C). When non-neutral evolution within species is coupled with accelerated (or reduced) rates of evolution in the reference sequence of the sister

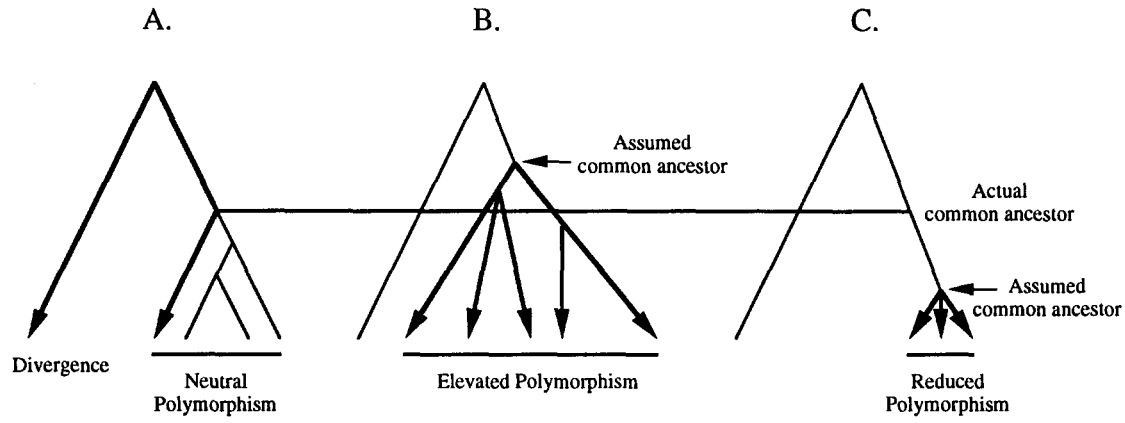

FIGURE 4.—Non-neutral evolution of DNA can compromise historical inference in population and evolutionary biology. (A) Neutral DNA evolution where the application of a rate of evolution determined from an interspecific comparison (thick lines with arrows) to the genealogy of haplotypes within a species provide an accurate estimate of both the common ancestor of DNA sequences and organisms (horizontal line). (B) With balancing or other diversifying selection (or relaxed selective constraint) operating within species, the application of the interspecific rate of DNA evolution to intraspecific variation results in an assumed common ancestor older than the organismal common ancestor. (C) Following a selective sweep of variation the common ancestor of the haplotypes is rendered more recent than the organismal common ancestor. Further misleading inferences about intraspecific events can be made if the rates of evolution between species differ.

taxon, these problems of dating evolutionary events may be exacerbated. The likelihood of making completely erroneous inferences will depend on the degree to which branch lengths have been altered and the confidence limits on evolutionary rates [which can be quite wide; see Figures 1 and 2 in HILLIS and MORITZ (1990)].

In light of the historical biogeography of *D. melanogaster* and *D. simulans*, the genealogies of mtDNAs in these species seem to suggest that the expansion of *D. simulans* out of Africa occurred more recently than that of *D. melanogaster* (HALE and SINGH 1991b). Recognizing that mtDNA evolution has been faster in *D. melanogaster* than in *D. simulans* and that a selective sweep may have reduced mtDNA variation in the latter species, the actual dispersal of the flies may have been closer in time. Similar types of reassessments could apply to other organisms, and notably, to the human mitochondrial eve hypothesis (CANN *et al.* 1987) if non-neutral events have pushed the coalescent time of human mtDNAs closer to the present. Given the growing bits of evidence for departures from neutral evolution of mtDNA (WHITTAM *et al.* 1986; EXCOFFIER 1990; NACHMAN *et al.* 1994; BALLARD and KREITMAN 1994), evolutionary inferences based on patterns of mtDNA variation should consider selection as a force that needs to be addressed directly rather than simply ignored.

We would like to thank CHIP AQUADRO and DAVID BEGUN for providing fly stocks and for helpful comments on many aspects of this study. MICHAEL NACHMAN provided a very helpful review of the manuscript. This work was supported by National Science Foundation grant DEB-9120293 to D.M.R.

Note added in proof: We recently identified an additional fixed synonymous difference between *D. melanogaster* and *D. simulans* that is not reported in Figure 1. Position 318 of our Figure 1 is an "A" in *D. simulans*. This corresponds to position 2525 in Figure 2 of GARESSE (1988). Although this increases the value of D_2 in HKA tests and the values reported in Table 2, it does not change any of the conclusions in this report.

LITERATURE CITED

AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989 Reduced variation in the *yellow-achete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615.

AQUADRO, C. F., and D. J. BEGUN, 1993 Evidence for and implications of genetic hitchhiking in the *Drosophila* genome, pp. 159–178 in *Mechanisms of Molecular Evolution*, edited by N. TAKAHATA and A. G. CLARK. Sinauer Associates, Sunderland, Mass.

AQUADRO, C. F., K. M. LADO and W. A. NOON, 1988 The *rosy* region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. Genetics **119:** 1165–1190.

ATWOOD, K. C., L. K. SCHNEIDER and F. J. RYAN, 1951 Selective mechanisms in bacteria. Cold Spring Harbor Symp. Quant. Biol. **16:** 345–354.

BABA-AÏSSA, F., M. SOLIGNAC, N. DENNEBOUY and J. R. DAVID, 1988 Mitochondrial DNA variability in *Drosophila simulans*: quasi absence of polymorphism within each of the three cytoplasmic races. Heredity **61:** 419–426.

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

BATEMAN, A. J., 1948 Intra-sexual selection in *Drosophila*. Heredity **2:** 349–368.

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

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.

BEGUN, D. J., and C. F. AQUADRO, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **356:** 519–520.

BEGUN, D. J., and C. F. AQUADRO, 1994 Evolutionary inferences from variation at the 6-phosphogluconate dehydrogenase locus in

natural populations of Drosophila: selection and geographic differentiation. Genetics **136:** 155–171.

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.

CANN, R. L., M. STONEKING and A. C. WILSON, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

CLARY, D. O., and D. R. WOLSTENHOLME, 1985 The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. **22:** 252–271.

DAVÍD, J., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

DESALLE, R., T. FREEDMAN, E. M. PRAGER and A. C. WILSON, 1987a Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. **26:**157–164.

DESALLE, R., A. TEMPLETON, I. MORI, S. PLETSCHER and J. S. JOHNSON, 1987b Temporal and spatial heterogeneity of mtDNA polymorphisms in natural populations of *Drosophila mercatorum*. Genetics **116:** 215–223.

EXCOFFIER, L., 1990 Evolution of human mitochondrial DNA: Evidence for departure from a pure neutral model of populations at equilibrium. J. Mol. Evol. **30:** 125–139.

FORD, M. J., C. K. YOON and C. F. AQUADRO, 1994 Molecular evolution of the *period* gene in *Drosophila athabasca* Mol. Biol. Evol. **11:** 169–182.

FOS. M. A. DOMINGUEZ, A. LATORRE and A. MOYA, 1990 Mitochondrial DNA evolution in experimental populations of *Drosophila subobscura*. Proc. Nat. Acad. Sci. USA **87:** 4198–4201.

GARESSE, R., 1988 *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations. Genetics **118:** 649–663.

GILLESPIE, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, Oxford.

GLOOR, G., and W. R. ENGELS, 1991 Single fly DNA preps for PCR. Drosophila Inf. Newsl. **1**. (Available from WRENGELS@ WISCMACC. BITNET.)

HALE, L. R., and A. A. HOFFMANN, 1990 Miotchondrial DNA polymorphism and cytoplasmic incompatibility in natural populations of *Drosophila simulans*. Evolution **44:** 1383–1387.

HALE, L. R., and R. S. SINGH, 1991a A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history *vs.* selection in the genetic structure of geographic populations. Genetics **129:** 103–117.

HALE, L. R., and R. S. SINGH, 1991b Contrasting patterns of genetic structure and evolutionary history as revealed by mitochondrial DNA and nuclear gene-enzyme variation between *Drosophila melanogaster* and *Drosophila simulans*. J. Genet. **70:** 79–90.

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.

HILLIS, D. M, and C. MORITZ, 1990 An overview of aplications of molecular systematics, pp. 502–515 in *Molecular Systematics*, edited by D. M. HILLIS and C. MORITZ. Sinauer Associates, Sunderland, Mass.

HOFFMANN, A. A., M. TURELLI and G. M. SIMMONS, 1986 Unidirectional incompatibility between populations of *Drosophila simulans*. Evolution **40:** 692–701.

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

HUDSON, R. R, M. SLATKIN and W. P. MADDISON, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

KANEKO, M., Y. SATTA, E. T. MATSUURA and S. CHIGUSA, 1993 Evolution of the mitochondrial ATPase 6 gene in *Drosophila*: unusually high level of polymorphism in *D. melanogaster*. Genet. Res. **61:** 195–204.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

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.

KONDO, R., Y. SATTA, E. T. MATSUURA, H. ISHIWA, N. TAKAHATA *et al.*, 1990 Incomplete maternal transmission of mitochondrial DNA in Drosophila. Genetics **126:** 657–663.

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.

KREITMAN, M., and L. LANDWEBER, 1989 A strategy for producing single stranded DNA in the polymerase chain reaction: a direct method for genomic sequencing. Gene Anal. Tech. **6:** 84–88.

KUMAR, S., K. TAMURA and M. NEI, 1993 *MEGA: Molecular Evolutionary Genetic Analysis*. Institute of Molecular Evolutionary Genetics, The Pennsylvania State University.

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

MACRAE, A. F., and W. W. ANDERSON, 1988 Evidence for non-neutrality of mitochondrial DNA haplotypes in *Drosophila pseudoobscura*. Genetics **120:** 485–494.

MADDISON, W. P, and D. R. MADDISON, 1992 *MacClade: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, Mass.

MARTIN-CAMPOS, J. M., J. M. CAMERON, 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.

MAYNARD-SMITH, J., and J. HAIGH, 1974 The hitchhiking effect of a favorable gene. Genet. Res. **23:** 23–35.

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

MONTCHAMP-MOREAU, C., J.-F. FERVEUR and M. JACQUES, 1991 Geographic distribution and inheritance of three cytoplasmic incompatibility types in *Drosophila simulans* Genetics **129:** 399–407.

NACHMAN, M. W., S. N. BOYER and C. F. AQUADRO, 1994 Non-neutral evolution at the mitochondrial ND3 gene in mice. Proc. Natl. Acad. Sci. USA. (in press).

NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NIGRO, L., and T. PROUT, 1990 Is there selection on RFLP differences in mitochondrial DNA? Genetics **125:** 551–555.

ROUSSET, F. D. VAUTRIN and M. SOLIGNAC, 1992 Molecular identification of *Wolbachia*, the agent of cytoplasmic incompatibility in *Drosophila simulans*, and variability in relation with host mitochondrial types. Proc. R. Soc. Lond. B **247:** 163–168.

SATTA, Y., and N. TAKAHATA, 1990 Evolution of *Drosophila* mitochondrial DNA and the history of the *melanogaster* subgroup. Proc. Natl. Acad. Sci. USA **87:** 9558–9562.

SATTA, Y., H. ISHIWA and S. I. CHIGUSA, 1987 Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. Mol. Biol. Evol. **4:** 638–650.

SCHAEFFER, S. W., and E. L. MILLER, 1992 Molecular population genetics of an electrophoretically monomorphic protein in the Alcohol dehydrogenase region of *Drosophila pseudoobscura* Genetics **132:** 163–178.

SHARP, P., and W.-H. LI, 1989 On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Evol. **28:** 398–402.

SHEEN, J., and B. SNEED, 1988 Electrolyte gradient gels for DNA sequencing. BioTechniques **6:** 942–944.

SINGH, R. S., and L. R. HALE, 1990 Are mitochondrial DNA variants selectively non-neutral. Genetics **124:** 995–997.

SINGH, R. S., and L. R. RHOMBERG, 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and geographic differentiation. Genetics **117:** 255–271.

SOLIGNAC, M., M. MONNEROT and J.-C. MOUNOLOU, 1986 Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Evol. **23:** 31–40.

STEPHAN, W. and S. J. MITCHELL, 1992 Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. Genetics **132:** 1039–1045.

SWOFFORD, D. L., 1991 PAUP: Phylogenetic analysis using parsimony. Illinois Natural History Survey.

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

TAJIMA, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135:** 599–607.

TAMURA, K., 1992 Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol. Biol. Evol. **9:** 678–687.

TURELLI, M., and A. A. HOFFMANN, 1991 Rapid spread of an inherited incompatibility factor in California *Drosophila*. Nature **353:** 440–442.

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

WHITTAM, T. S., A. G. CLARK, M. STONEKING, R. L. CANN and A. WILSON, 1986 Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. Proc. Natl. Acad. Sci. USA **83:** 9611–9615.

Communicating editor: A. G. CLARK