J Mol Evol (2000) 51:64–75
DOI: 10.1007/s002390010067

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 2000

Comparative Genomics of Mitochondrial DNA in *Drosophila simulans*

J. William O. Ballard

The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL 60605-2496, USA

Received: 16 September 1999 / Accepted: 14 March 2000

Abstract. The current study compares the nucleotide variation among 22 complete mitochondrial genomes of the three distinct *Drosophila simulans* haplotypes with intron 1 of the *alcohol dehydrogenase-related* locus. This is the first study to investigate the sequence variation of multiple complete mitochondrial genomes within distinct mitochondrial haplotypes of a single species. Patterns of variation suggest distinct forces are influencing the evolution of mitochondrial DNA (mtDNA) and autosomal DNA in *D. simulans.* First, there is little variation within each mtDNA haplotype but strong differentiation among them. In contrast, there is no support for differentiation of the mitochondrial haplotypes at the autosomal locus. Second, there is a significant deficiency of mitochondrial variation in each haplotype relative to the autosomal locus. Third, the ratio of nonsynonymous to synonymous substitutions is not equal in all branches of the well-resolved phylogeny. There is an excess of nonsynonymous substitutions relative to synonymous substitutions within each *D. simulans* haplotype. This result is similar to that previously observed within the mtDNA of distinct species. A single evolutionary force may be causally linked to the observed patterns of mtDNA variation—a rickettsia-like microorganism, *Wolbachia pipientis,* which is known to directly influence mitochondrial evolution but have a less direct influence on autosomal loci.

Key words: Comparative genomics — Genus *Drosophila* — Mitochondrial DNA — Neutral evolution

Introduction

The purpose of this study is to examine nucleotide variation in the complete mitochondrial genome and in intron 1 of the *alcohol dehydrogenase–related* (*Adhr*) locus of *Drosophila simulans.* Twenty-two isofemale lines of the three *D. simulans* haplotypes (*si*I, -II, and -III) and two lines of *D. melanogaster* were included. Investigating the variation in complete mitochondrial genomes enables differentiation of evolutionary forces acting at individual loci and whole genomes. Including multiple individuals of each haplotype and species permits investigation of polymorphism and divergence. Furthermore, by comparing the variation in mtDNA to that in a nuclear gene, a more refined understanding of the suite of evolutionary forces acting on the mitochondrial genome is possible.

This is the first study to investigate the sequence variation of multiple complete mitochondrial genomes within distinct mitochondrial haplotypes of a single species. Considering complete mitochondrial genomes is important because it is not known if genome-wide processes can be inferred from a small region of nonrecombining mitochondrial DNA (mtDNA), particularly in some tests involving polymorphism and divergence (Rand and Kann 1998). Nachman (1998) and Rand and Kann (1998) screened GenBank for data sets that included multiple individuals from single species and sufficient intraspecific variation so that statistical analyses could be performed. In only one case was a comparison of all genes possible (Nachman et al. 1996). Ballard (2000a) conducted a maximum likelihood sliding-window analysis of nine *Drosophila* mtDNA genomes and detected four disjunct regions that had a significantly different nucleotide substitution process from that derived from the complete sequence. Two of these regions had significantly elevated rates of evolution in distinct lineages of *D. simulans.* Ballard (2000a) compared mitochondrial genomes of *D. yakuba,* three isofemale lines of *D. simulans,* two of *D. melanogaster* and *D. mauriti-*

ana, and one of *D. sechellia.* In that study, the three isofemale lines of *D. simulans* and two of *D. mauritiana* represented single representatives of distinct haplotypes.

The three *D. simulans* haplotypes were identified by Solignac and Monnerot (1986), who divided 13 isofemale lines of *D. simulans* into three mtDNA cleavage morphs based on 12 restriction enzymes. Baba-Aïssa et al. (1988) extended the survey to include 217 isofemale lines. The *si*I type is known to occur in New Caledonia, Hawaii, Tahiti, and the Seychelles Islands. The *si*II type has a worldwide distribution, but has not been collected on the Pacific Islands of New Caledonia, Hawaii, and Tahiti. The *si*III haplotype has been collected from Madagascar and Reunion Island, where it is sympatric with the *si*II type. In the coding region of the mtDNA, Baba-Aïssa et al. (1988) noted that the three mitochondrial types differed by 10 to 15 restriction sites among the 40 identified. Within a given type, however, variability was absent or was restricted to a single site. There is also evidence for reduced levels of sequence variation in *si*II at both the NADH dehydrogenase subunit 5 (ND5) (Rand et al. 1994) and cytochrome *b* loci (Ballard and Kreitman, 1994).

It is possible that independent advantageous mutations caused the reduction in mtDNA diversity in each haplotype. A reduction in polymorphism may result from genetic hitchhiking of neutral mutations in response to selection on another part of the genome (Maynard Smith and Haigh 1974; Kaplan et al. 1989). Thus, if *Drosophila* mtDNA does not recombine (Ballard 2000a), the selective fixation of any mutation in the mtDNA will lead to the concomitant fixation of all variants in that genome. However, selection need not even be acting on the mitochondrial genome itself: any maternally inherited factor could potentially influence haplotypic diversity. One such factor is the rickettsia-like microorganism *Wolbachia* (Turelli and Hoffmann 1995). At least five strains of *Wolbachia* infect *D. simulans* (Hoffmann et al. 1986, 1996; Hoffmann and Turelli 1988; O'Neill and Karr 1990; Montchamp-Moreau et al. 1991; Rousset et al. 1992; James and Ballard 2000).

The influence of *Wolbachia* on mitochondrial variation has been studied most extensively in California (USA). There, two types of *si*II lines have been identified that display cytoplasmic incompatibility: the R-line (originally collected at Riverside, CA), which is infected with *Wolbachia,* and the uninfected W-line (originally collected at Watsonville, CA) (Hoffmann et al. 1986; Hoffmann and Turelli 1988; O'Neill and Karr 1990). In *D. simulans,* cytoplasmic incompatibility may be expressed when a male harboring a strain of *Wolbachia* mates with a female that does not carry that same strain or is uninfected. Sperm enters the egg normally, but defects in fertilization cause a reduction in egg hatchability. Because of this cytoplasmic incompatibility, an R-type *Wolbachia* strain (*w*Ri) and its linked mtDNA variant will spread rapidly through a population (Hoffmann et al. 1986; Turelli and Hoffmann 1995; Turelli et al. 1992). The parasite is occasionally lost from infected flies so that once a population becomes infected the barrier to gene flow is not complete (Hoffmann and Turelli 1988; Hoffmann et al. 1990). Consistent with this observation, Ballard et al. (1996) observed significantly less variability in mtDNA compared to that of a nuclear gene, based on the HKA test (Hudson et al. 1987), in both a *Wolbachia*-infected and the pooled sample of *Wolbachia*-uninfected and -infected flies.

The purpose of this study is to test specific hypotheses concerning the evolution of mtDNA in *D. simulans.* With a nonrecombining genome all genes share a single genealogy and all should be equally affected by genetic hitchhiking of neutral mutations in response to selection on any part of the genome. To amplify the data needed to test for non-neutrality, whole genome polymorphism data was obtained. To perform additional statistical tests, including HKA and HBA tests (Hudson et al. 1987, 1992), I also conducted a study of variation at intron 1 of the *Adhr* locus. This autosomal region was sequenced because it has considerable variation, and the pattern is consistent with a neutrally evolving locus (Sumner 1991). The mitochondrial data show that there is little variation within each of the three mitochondrial haplotypes but considerable differentiation among them. In contrast, there is no differentiation among the three mitochondrial haplotypes at the autosomal locus. A single evolutionary force may be linked to the observed patterns. Alternately, a variety of independent events may have generated the observed patterns of polymorphism and divergence.

Materials and Methods

Lines and Loci Employed. The complete mitochondrial genomes, excluding the A + T rich region, from 24 *Drosophila* isofemale lines were sequenced. The A + T rich region is difficult to align (Lewis et al. 1994; Inohira et al. 1997) and so is not included. *D. mauritiana* and *D. sechellia* are excluded because the phylogenetic relationships of these taxa at the *Adhr* locus are unclear (Ballard 2000b).

Two isofemale lines of *D. melanogaster* are used—Oregon R and Zimbabwe 53 (Table 1). Isofemale lines taken from North America and Zimbabwe have clearly differentiated mtDNA (Rand et al. 1994) and many nuclear variants are not shared between the two localities (Begun and Aquadro 1993). The sequence of *D. melanogaster* collated by Lewis et al. (1995) is not employed in this study because it is a chimera of different isofemale lines.

Six isofemale lines of *si*I, seven of *si*II, and nine of *si*III were included to maximize geographic and mitochondrial diversity. Though the samples within each mitochondrial haplotype were chosen at random, the combined data set is not strictly a random sample of *D. simulans.*

The first intron of *Adhr* was chosen because the pattern of variability at this locus is compatible with a neutral equilibrium model in *D. simulans* (Sumner 1991). *Adhr* is downstream of *Adh* and is related to that gene by an ancient tandem duplication. *Adhr* consists of three nonoverlapping open reading frames separated by two introns that are conserved between *D. pseudoobscura* and *D. simulans* (Schaeffer and Aquadro 1987).

Table 1. List of isofemale lines employed in this study

Species	Haplotype	Line	Location
D. melanogaster	—	Oregon R	Roseburg, OR, USA[a]
	—	Zimbabwe 53	Sengwa, Africa[b]
D. simulans	*si*I	HW00	Honolulu, Hawaii[c]
	*si*I	HW09	Honolulu, Hawaii[c]
	*si*I	TT00	Moorea, Tahiti[d]
	*si*I	TT01	Papeete, Tahiti[c]
	*si*I	NC37	Noumea, New Caledonia[d]
	*si*I	NC48	Noumea, New Caledonia[d]
	*si*II	DSR	Riverside, CA, USA[f]
	*si*II	DSW	Watsonville, CA, USA[f]
	*si*II	SC00	Mahe, Seychelles[d]
	*si*II	RU35	St. Pierre, Reunion Is.[e]
	*si*II	MD106	Ansirabe, Madagascar[e]
	*si*II	MD225	Joffreville, Madagascar[e]
	*si*II	C167	Nanyuki, Kenya[g]
	*si*III	RU00	Bois des Nefles, Reunion Is.[h]
	*si*III	RU259	St. Dennis, Reunion Is.[h]
	*si*III	RU01	St. Dennis, Reunion Is.[h]
	*si*III	RU07	Salazie, Reunion Is.[e]
	*si*III	MDW86	Antananarivo, Madagascar[h]
	*si*III	MD111	Ansirabe, Madagascar[e]
	*si*III	MD112	Ansirabe, Madagascar[e]
	*si*III	MD119	Ansirabe, Madagascar[e]
	*si*III	MD221	Joffreville, Madagascar[e]

[a] D.E. Lancefield
[b] R.R. Ramey and L. Brown
[c] K. Kaneshero
[d] M. Solignac
[e] J.W.O. Ballard
[f] A. Hoffmann
[g] A.V. Olembo
[h] R. Russell

DNA Preparation, Polymerase Chain Reaction (PCR), and Cycle Sequencing. DNA from all lines was extracted using the PureGene® Kit (Gentra) following the Isolation from Fixed Tissue protocol. In all cases, the DNA was extracted from individuals less than 14 days of age. In *D. melanogaster*, deletions in the mtDNA do not increase with age (embryos to 60-day-old adults; Schwarze et al. 1998).

The 15,034-bp mitochondrial genome was PCR amplified in 14 overlapping fragments. To minimize the possibility of contamination, each genome was completed before the next was commenced. Negative controls confirmed that there was no contamination. To sequence the mitochondrial molecule, 84 to 96 cycle sequencing reactions were employed. Both strands were sequenced using Taq-Dye Deoxy Terminator Cycle sequencing (Ballard 2000a). Sequences were imported into the Sequencher® software program and the chromatograms investigated. Six inconsistencies between the sequences derived from independent PCR products were resolved by reamplifying and resequencing the region of interest. These were considered PCR errors. In this study, there was no evidence of heteroplasmy or large-scale deletions (Solignac et al. 1983; Kondo et al. 1990). The mitochondrial sequence data were aligned against the published *D. yakuba* sequence (Clary and Wolstenholme 1985).

Pooling genes transcribed from opposite strands has the potential to mask any transcriptional or translational biases. In an attempt to minimize this bias, each triplet transcribed from the minor strand was reverse complemented. This procedure retains the order of amino acids but reverses the relative positions of the first and third codons. The tRNAs and rRNAs transcribed from the minor strand were then complemented. Synonymous (does not cause an amino acid change in a protein-coding region), nonsynonymous (causes an amino acid change), and silent (noncoding in this study) substitutions were then defined in DnaSP 3.14 (Rozas and Rozas 1997).

The 501-bp region of intron 1 of *Adhr* was amplified (Ballard et al. 1996), cloned and a single copy sequenced (Ballard 2000a). Two to four cycle sequencing reactions were employed to sequence this region. The *Adhr* data were aligned against the previously published sequence of Cohn et al. (1984) and Ballard et al. (1996) using Sequencer®. No PCR errors were detected.

Phylogenetic Analyses. In an attempt to represent the number of insertion and deletion (indel) events accurately, each presumed event was parsimoniously scored by inserting a "1" into the matrix at appropriate sites. Gaps were then treated as missing data.

The genealogical relationship of the mitochondrial genome and the *Adhr* locus was analysed by parsimony using PAUP* (Swofford 1998). Bootstrapping (Efron 1982; Felsenstein 1985) was used to test monophyly. For this study, 1,000 pseudosamples were generated to estimate the bootstrap proportions.

This is the first study, to my knowledge, that specifically includes autosomal data from representatives of each distinct mitochondrial haplotype. Previous analyses of *D. simulans* autosomal loci have suggested that this species is broadly distributed with little population subdivision. The mitochondrial data suggest there is extensive population subdivision. It may be predicted that the *Adhr* and mtDNA genealogies will not be correlated if they are independently segregating in a randomly mating population or evolving under distinct evolutionary processes. As an alternative, consider the possibility that the distinct mtDNA haplotypes occur in populations that have distinct evolutionary trajectories and are not randomly mating. In this latter case it may be expected that the autosomal and mitochondrial data will be correlated. To investigate whether the autosomal and mtDNA data are significantly associated, the incongruence length difference (ILD) test (Farris et al. 1995) was employed to test the null hypothesis that the mitochondrial and autosomal loci are evolving under homogeneous biological processes. This random-partitioning test is an extension of a measure originally reported by Mickevich and Farris (1981) and is based on the null hypothesis of congruence.

Sequence Investigations. Nucleotide diversity (π) and the neutral parameter (θ) based on the number of segregating sites were calculated

for *D. melanogaster* and *D. simulans.* Nucleotide diversity (within each species or haplotype) and divergence (between species) were calculated by the sliding window method. In this method, a window (1,000 bp in length) was moved along the sequences in steps of 50 bp. Nucleotide diversity and divergence were calculated in each window, and the value assigned to the nucleotide at the midpoint of the window.

The HBK test (Hudson et al. 1992) was employed to investigate whether the mitochondrial haplotypes are genetically subdivided at the autosomal locus. The test is a nonparametric, permutation-based statistical test that uses molecular variation in samples of DNA sequences from two or more geographic localities. In this case, "localities" are not defined geographically but by their mitochondrial haplotype. The null hypothesis for this test is that there is random association of each mitochondrial haplotype with each nuclear allele. The alternative is that each mitochondrial haplotype is associated with specific nuclear alleles. Hudson et al. (1992) examined the power of the test to detect subdivision under a selectively neutral Wright-Fisher Island model with Monte Carlo simulations. With small population sizes in one or more localities the rank statistic Z* and the statistic based on χ^2 with the *p* values estimated by permutation (1,000 in this study) (Roff and Bentzen 1989) have the greatest power and are employed.

Tests of the Neutral Model of Molecular Evolution. The HKA test (Hudson et al. 1987) was used to compare the levels of silent and synonymous variation in mitochondrial and autosomal loci. The HKA test is a conservative test of an equilibrium neutral model's prediction that polymorphism within species and divergence between species will be positively correlated. The level of mtDNA polymorphism in each species/haplotype was compared to the corresponding level in *Adhr*. Divergence was estimated by calculating the mean sequence divergence of each *D. simulans* gene with that of *D. melanogaster*. The HKA test is based on the assumption that each sample is taken from a single randomly mating population. For each test, the effective population size of the mitochondrial genes was considered to be one-quarter that of the autosomal gene. The first test considers the polymorphism in the 22 lines of *D. simulans* with the divergence from the two lines of *D. melanogaster*. These results should be treated with caution because the total sample is clearly not taken from a panmictic population. The three subsequent tests consider the polymorphism within each haplotype (*si*I, -II, and -III, respectively) with the divergence from *D. melanogaster*. The *si*I haplotype is fixed in the Pacific Islands of New Caledonia, Tahiti, and Hawaii (James and Ballard 2000), and this may be considered a random sample of these islands. The *si*II haplotype has a worldwide distribution and is probably fixed in all regions but the Pacific Islands, Madagascar, Reunion Island, and the Seychelles (Baba-Aïssa et al. 1988; James and Ballard 2000). The *si*III haplotype occurs in Madagascar and Reunion Island, where it accounts for about a third (58/189) of *D. simulans* (James and Ballard 2000). Consequently, the HKA test was modified to consider the proportion of *si*III individuals in the population. It is assumed that there is no barrier to autosomal gene flow between the *si*II and *si*III haplotypes in Madagascar and Reunion Island (perhaps a simplifying assumption).

To investigate protein evolution, the MK test (McDonald and Kreitman 1991) was used to compare the number of nonsynonymous to synonymous substitutions in the mitochondrial genome. A post hoc *t* test was then conducted. The post hoc *t* indicated what each cell in the table contributes to the Chi-square statistic. They are calculated to follow a standard normal distribution and absolute values greater than 1.96 indicate that the cell in question is significantly different ($p < 0.05$) than expected by chance. The MK test of neutrality is based on the prediction that the ratio of nonsynonymous to synonymous differences should be equal in all branches of the phylogenetic tree. These analyses do not include a multiple hit correction and therefore may underestimate the number of fixed synonymous substitutions in some lineages. To summarize the substitution pattern within and among species, the index of neutrality (NI) is presented (Rand and Kann 1996). This index reflects the extent to which the levels of amino acid variation within species depart from the strictly neutral model. Strict neutrality has an index of 1.0. Values greater than 1.0 typically indicate an excess of amino acid variation within species, and values less than 1.0 indicate an excess of nonsynonymous substitutions among species, relative to the number of synonymous substitutions.

In this study, I employ Tajima's (1989) D, Fu and Li's (1993) D*, and Fu's (1997) Fs to test whether the observed mutation patterns in the mitochondrial and autosomal data are consistent with a neutral model of molecular evolution. Simmonsen et al. (1995) found that Tajima's D is generally more powerful against the alternative hypothesis of a selective sweep than Fu and Li's D* or F*. However, Fu (1997) found that Fu and Li's (1993) tests are more powerful than Tajima's test in the case of background selection. Tajima's D tests whether there is a significant difference in the estimates of θ derived from the number of segregating sites and the average heterozygosity. Fu and Li's (1993) D* test statistic is based on the differences between the number of singletons and the total number of mutations (Fu and Li 1993). Fu's (1997) Fs is related to Strobeck's statistic S (Strobeck 1987; Fu 1997) and tends to be positive when there is an excess of old mutations. It is the most powerful of the three tests at detecting hitchhiking when there is no recombination (Fu 1997). These tests assume that each sample is taken from a single randomly mating population. This assumption is violated in this study, as approximately equal numbers of each haplotype were included. The consequences of violating this assumption are considered in the Results and the Discussion.

Results

Phylogenetic Analyses

The alignment of the 24 mitochondrial genomes (GenBank #AF200828-9, AF200833-54) is straightforward for the majority of the sequence. However, 76 bp of 15,034 bp are deleted from all analyses because it is not possible to unequivocally determine the indels that included positions 5,535–5,584 and 6,022–6,047. The region between 5,535–5,584 forms the majority of an intervening sequence between COIII and the glycine tRNA. The region between positions 6,022–6,047 is located in an intervening sequence between ND3 and the alanine tRNA. Sequence data from intron 1 of *Adhr* were obtained for all 24 isofemale lines (GenBank #AF201423-4, AF201428-49), and all 501 sites were readily aligned.

The mitochondrial genome and the autosomal locus were analyzed independently (Bull et al. 1993). There is no evidence for recombination in the mitochondrial genome of *Drosophila* (Ballard 2000a), while there is extensive recombination in intron 1 of *Adhr* (Ballard et al. 1996). As may be expected for independently segregating chromosomes, the ILD test (Farris et al. 1995) shows the regions are not evolving under a single homogeneous process. The sum of lengths for the original partition was 1,141 steps. This is significantly longer than the sum of lengths of 999 randomized partitions of the same size ($p = 0.001$).

There are three distinct mitochondrial haplotypes in *D. simulans* (Fig. 1A and Table 2). Within each haplotype, there is very little variation (Table 3). Within *si*I, there are 16 segregating sites, within *si*II 12, and within

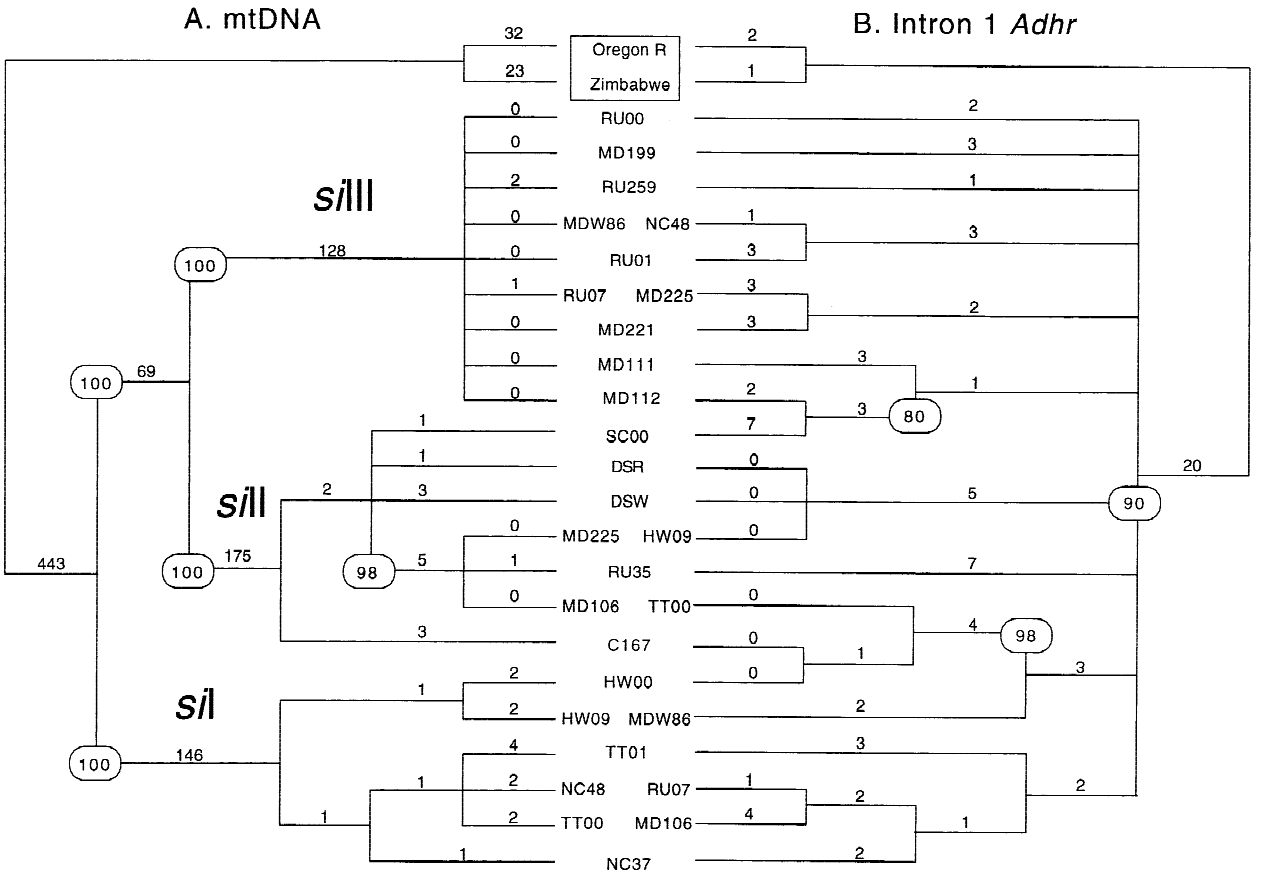

Fig. 1. Strict consensus trees generated from the mitochondrial and autosomal data. Taxa are placed in the center aisle. The two taxa in the box are lines of *D. melanogaster* and are the designated outgroup. If a single taxon is present in a given row it applies to both the mitochondrial and autosomal data sets. Where two taxa occur in a single row the left taxon refers to the mitochondrial data set and the right taxon to the autosomal data set. Both data sets were bootstrapped 1,000 times and the resulting proportions (> 70%) are shown in circles. The number of substitutions is shown above each branch. **A** Mitochondrial data. A total of the 14,958 bp are included: 14,023 characters are constant, 51 variable characters are parsimony uninformative, while 884 are parsimony informative. Five equally parsimonious trees of length 1,050 steps were found. These trees have a consistency index of 0.92. **B** Intron 1 of the *Adhr* locus. A total of 501 characters are included: 433 characters are constant, 26 variable characters are parsimony uninformative, and 42 characters are parsimony informative. Twenty-six equally parsimonious trees of length 91 steps were found. These trees have a consistency index of 0.77.

Table 2. Variation among the three distinct *D. simulans* mtDNA haplotypes

Haplotypes compared	Number of fixed differences	Average number of differences	K*	K at synonymous sites
*si*I and *si*II	384	392.07	2.62E-2	0.13
*si*I and *si*III	339	345.00	2.30E-2	0.11
*si*II and *si*III	300	305.19	2.04E-2	0.10

* Nucleotide divergence

*si*III just 3. No nodes within the *si*I or *si*III haplotypes are supported by more than a 70% bootstrap (Fig. 1A). Within the *si*II haplotype there is strong support for monophyly of the three lines from Madagascar and Reunion (MD106, MD225, and RU35) relative to the isofemale lines from the Seychelles (SC00), California (DSR and DSW), and Kenya (C167). On this lineage, three of the five substitutions (Fig. 1A) are fixed, while two are homoplasious. Here, the ancestral *si*II haplotype is defined as *si*IIA and the derived haplotype as *si*IIB.

The nearly allopatric geographic substructure of the mitochondrial haplotypes is obviously not reiterated at *Adhr*. Data from intron 1 of the *Adhr* locus do not show the same population substructure (Fig. 1B). Notably, there is strong support for monophyly of TT00, HW00, and C167 (*si*I, *si*I, and *si*II, respectively), of DSR, DSW, and HW09 (*si*II, *si*II, and *si*I, respectively), and of MD112 and SC00 (*si*III and *si*II, respectively).

Sequence Investigations

MtDNA polymorphism data and estimates of nucleotide divergence are summarized in Tables 3 and 4 and Figs. 2, 3, and 4. The 53 segregating sites in the two *D. melanogaster* lines are evenly distributed over the mitochondrial genome (Fig. 2). In contrast to this even distribution, there is an obvious decline in both *D. simulans*

Table 3. Variation within the mtDNA of *D. melanogaster* and *D. simulans*

	All data	*D. melanogaster*	*D. simulans*	*si*I	*si*II	*si*III
n	24	2	22	6	7	9
Seg. sites	922	53	532	16[a]	12[a]	3[a]
Variants	16	2	14	6	5	3
π^b	—	3.56E-3	1.58E-2	3.9E-4	3.2E-4	4.0E-5
θ^c	—	3.56E-3	9.97E-3	4.7E-4	3.3E-4	7.0E-5

[a] These mutations are not shared
[b] Nucleotide diversity (per site)
[c] Neutral parameter (per site)

Table 4. Protein coding variation in *D. simulans* mitochondrial DNA

Locus	Region[a]	Sites	π^b total	π synon.	π nonsyn.	K^c total
ND2	246–1,265	1,020	1.82E-2	6.94E-2	5.03E-3	4.03E-2
COI	1,480–3,012	1,533	2.06E-2	8.49E-2	9.20E-4	4.90E-2
COII	3,092–3,772	681	1.74E-2	6.45E-2	4.57E-3	4.55E-2
ATP8	3,920–4,072	153	2.34E-2	7.68E-2	1.04E-2	4.81E-2
ATP6	4,075–4,743	669	1.82E-2	6.93E-2	2.85E-3	4.48E-2
COIII	4,749–5,531	783	1.55E-2	6.77E-2	1.50E-4	4.99E-2
ND3	5,671–6,018	348	2.33E-2	1.04E-1	3.30E-3	5.48E-2
ND5	6,500–8,215	1,713	1.97E-2	7.62E-2	4.06E-3	5.49E-2
ND4	8,302–9,636	1,335	1.67E-2	6.26E-2	3.85E-3	4.21E-2
ND4L	9,642–9,926	285	1.46E-2	6.33E-2	2.23E-3	3.48E-2
ND6	10,070–10,588	519	2.16E-2	7.55E-2	7.40E-3	3.48E-2
Cyt*b*	10,598–11,728	1,131	1.87E-2	8.00E-2	1.25E-3	4.77E-2
ND1	11,825–12,757	933	2.00E-2	7.94E-2	3.10E-3	4.23E-2
Coding	1–15,034	11,103	1.88E-2	7.46E-2	3.20E-3	4.63E-2

[a] Does not include initiation or termination codons
[b] Nucleotide diversity
[c] Nucleotide divergence

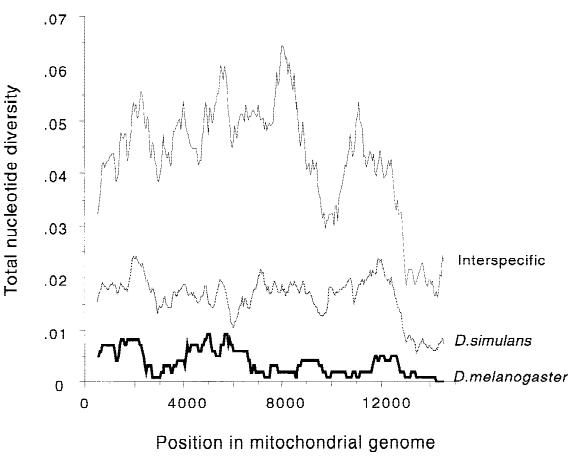

Fig. 2. Sliding window plot of total nucleotide diversity of *D. simulans* and *D. melanogaster* and divergence between them. The window is 1,000 bp in length and moves in 50-bp steps. The median position is plotted on the X axis.

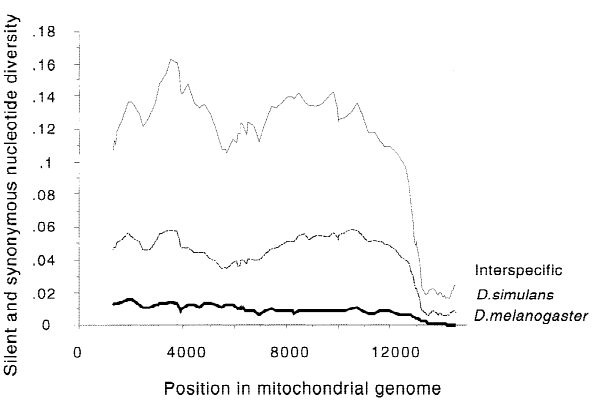

Fig. 3. Sliding window plot of silent and synonymous nucleotide diversity for *D. simulans* and *D. melanogaster* and divergence between them. The window is 1,000 bp in length and moves in 50-bp steps. The median position is plotted on the X axis.

nucleotide diversity and interspecific divergence in the large and small subunit rRNAs (lrRNA = 12,836–14,166 and srRNA = 14,240–15,034). There is also a distinct drop in interspecific divergence around position 9,600 (Fig. 2) that is not observed in the sliding-window plot of the silent and synonymous divergence (Fig. 3). This drop in interspecific divergence occurs in the ND4L and ND6 loci (Table 4). The estimates of total nucleotide diversity for *D. melanogaster* are similar to those obtained for restriction site data (π_r = 2.0E-3 restriction site π) by Hale and Singh (1987). In contrast, the esti-

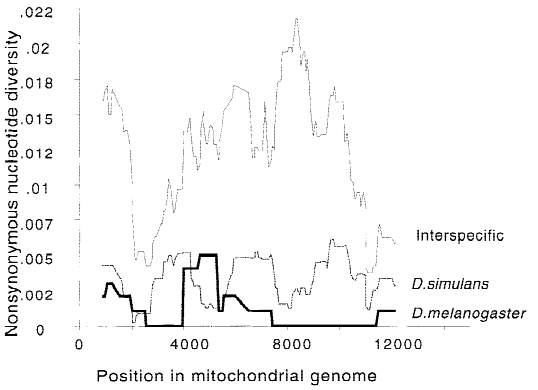

Fig. 4. Sliding window plot of nonsynonymous nucleotide diversity of *D. simulans* and *D. melanogaster* and divergence between them. The window is 1,000 bp in length and moves in 50-bp steps. The median position is plotted on the X axis.

mates for *D. simulans* nucleotide diversity are generally an order of magnitude lower than those calculated from restriction site data by Baba-Aïssa et al. (1988) (*D. simulans* $\pi_r = 1.87$, *si*I $\pi_r = 3.3E\text{-}2$, *si*II $\pi_r = 6.0E\text{-}3$, and *si*III $\pi_r = 0.00$). In this study, there are just 16, 12, and 3 segregating sites in *si*I, *si*II, and *si*III, respectively (Table 3).

Two standardized estimators of nucleotide polymorphism, nucleotide diversity and the neutral parameter, indicate a lower level of nucleotide polymorphism for *D. melanogaster* than *D. simulans* (Table 5). This result is consistent with previous studies (Aquadro 1992). Concordant with the known geographical distributions of the *D. simulans* haplotypes the levels of π are in the order *si*II > *si*I > *si*III.

The HBK test was employed to test the null hypothesis that there is a random association of each mitochondrial haplotype with each autosomal allele. In this study, Z* and χ^2 based on permutation are employed because they have the greatest power when the sample size is small (Hudson et al. 1992). In all cases the results are not statistically significant and the null hypothesis cannot be rejected ($n_1 = 6$, $n_2 = 7$, $n_3 = 9$; Z* probability = 0.29 and χ^2 probability based on permutation = 0.85). There is some evidence of population subdivision at the autosomal locus between *si*II and *si*III based on Z* ($p = 0.06$), but not on χ^2 based on permutation ($p = 0.48$). There is no evidence of population subdivision between *si*I and *si*II or between *si*I and *si*III employing either Z* ($p = 0.90$ and 0.18, respectively) or χ^2 ($p = 1.00$ for both).

Tests of Neutrality

Assuming that intron 1 of *Adhr* is evolving as a neutral marker (and this may not be the case in all populations), the HKA tests show that mtDNA diversity in each haplotype is suppressed below its neutral equilibrium value (Table 6). This suggests that genetic hitchhiking must have recently swept each *D. simulans* haplotype of variation. Indeed there are only nine, seven, and two silent and synonymous polymorphisms in the complete mitochondrial genomes of six lines of *si*I, seven of *si*II, and nine of *si*III, respectively. In contrast, there is no significant difference between the levels of polymorphism in the mitochondrial genome and intron 1 of *Adhr* in *D. simulans* relative to their divergence from *D. melanogaster*.

Statistical tests of protein evolution suggest that the ratio of nonsynonymous to synonymous substitutions is not equal in all mitochondrial lineages ($\chi^2_8 = 38.59$, $p < 0.001$). Post hoc *t* tests were conducted and mapped onto the phylogeny (Fig. 5). These data show a mosaic of substitution patterns. However, one consistent pattern emerges. Within *D. melanogaster* and each *D. simulans* haplotype there is an excess of nonsynonymous substitutions relative to synonymous changes. This is significant within the *si*I and *si*II haplotypes (6/8 and 5/3, respectively). To further investigate protein evolution in each gene, the index of neutrality (NI) is presented within and among species (Fig. 6). These data suggest that cytochrome *b* exhibits the highest NI and ND6 the lowest. However, following Bonferroni correction the NI of no individual gene is significantly different from that expected by chance.

Within each mitochondrial haplotype Tajima's (1989) D, Fu and Li's (1993) D*, and Fu's (1997) Fs tend to be negative (Table 7). In the nine lines defined by having the *si*III mitochondrial haplotype, Fs is significantly negative for intron 1 of *Adhr*. When the three haplotypes are combined these statistical tests of polymorphism are significantly negative for the autosomal locus and significantly positive for the mtDNA data. These data suggest that distinct evolutionary forces are acting on the mtDNA and autosomal DNA (aDNA) of *D. simulans* (Table 7). However, this latter result should be treated with caution because the 22 lines of *D. simulans* included in this study were not chosen at random. Rather, they were selected to represent approximately equal numbers of each haplotype. The equal frequency of highly differentiated haplotypes will cause an artificial increase in polymorphic sites with variants segregating at rather intermediate frequencies. This may cause these statistics to be artificially positive.

Discussion

Ballard (2000a) applied a comparative genomics approach to differentiate processes that are acting at a higher level from those acting on a single gene. The current study extends this research program to investigate the nucleotide variation within and among mitochondrial and autosomal DNA of *D. simulans* lineages. Overall at least three lines of evidence suggest that significantly different forces are influencing the evolution

Table 5. Variation at intron 1 of the alcohol dehydrogenase-repeated (*Adhr*) locus of *D. melanogaster* and *D. simulans*

	All data	*D. melanogaster*	*D. simulans*	*si*I	*si*II	*si*III
n	24	2	22	6	7	9
Seg. sites	51	2	39	17	27	19
Variants	22	2	18	5	6	9
π^a	—	4.18E-3	1.54E-2	1.48E-2	2.03E-2	1.16E-2
θ^b	—	4.18E-3	2.39E-2	1.51E-2	2.40E-2	1.52E-2

[a] Nucleotide diversity (per site)
[b] Neutral parameter (per site)

Table 6. HKA test comparing the intraspecific silent and synonymous polymorphism in *D. simulans* to the interspecific divergence from *D. melanogaster*[a]

	D. simulans[b]		*si*I[b]		*si*II[b]		*si*III[bc]	
	mtDNA	*Adhr*	mtDNA	*Adhr*	mtDNA	*Adhr*	mtDNA	*Adhr*
Intraspecific polymorphism data								
No. segregating sites (obs)	466	39	9	17	7	27	2	9
No. segregating sites (exp)	463.80	40.20	20.49	5.53	25.87	8.13	6.69	2.04
Total number of sites[d]	6,229.13	459	6,258.09	477	6,224.78	477	6,233.19	461
Sample size	22	22	6	6	7	7	9	9
Interspecific divergence data								
Mean no. differences (obs)	504.23	21.23	493.71	21.67	524.82	21.85	495.22	20.44
Mean no. differences (exp)	506.43	19.03	482.22	33.14	505.95	40.72	490.42	26.15
Total number of sites[d]	6,229.13	443	6,231.09	460	6,224.78	460	6,233.19	445
χ square		0.055		14.01***		23.93***		5.23*

[a] Asterisks show the p values ($p < 0.05$*, $p < 0.01$**, $p < 0.001$***)
[b] The effective population size of the mitochondrial loci was considered to be 1/4 that of the autosomal loci
[c] The effective population size of the mitochondrial loci was corrected for its frequency (0.33) in the population
[d] Excludes sites with alignment gaps

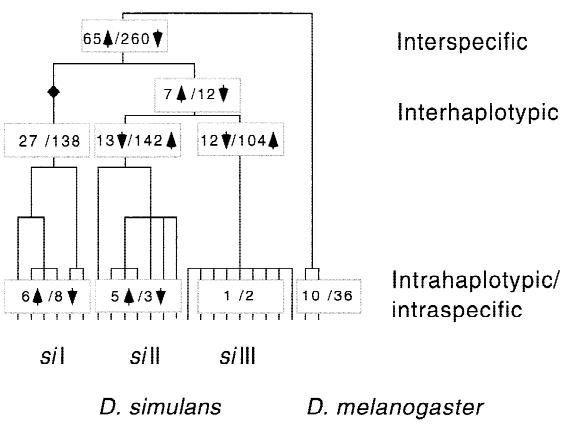

Fig. 5. Nonsynonymous/synonymous substitutions mapped onto the phylogenetic tree (Fig. 1). Arrows denote a significant departure from that expected by chance ($p < 0.05$ determined by post hoc t test). A ♦ has been placed on the interhaplotypic lineage to *si*I to indicate the phylogenetic position of the lineage to *D. sechellia* (Ballard 2000a).

of *D. simulans* mtDNA and aDNA. First, the mtDNA haplotypes are geographically subdivided, whereas the genealogy of *Adhr* clearly indicate extensive gene flow between the geographic regions. Second, there is a significant deficiency of mitochondrial variation in each *D. simulans* haplotype relative to the autosomal variation. Third, the ratio of synonymous to nonsynonymous substitutions is not equal in all branches of the well resolved phylogeny. The obvious question that must be addressed, then, is what evolutionary mechanism can influence the mtDNA so dramatically but yet have no detectable affect on *Adhr*. One potential candidate that is known to directly influence mitochondrial evolution but have a less direct influence on autosomal loci is *Wolbachia.* At least five distinct strains of *Wolbachia* are known to infect *D. simulans* (James and Ballard 2000).

Studies of phenotypic variation and of allozymes in *D. simulans* suggested that there is little population subdivision within the species (Hyytia et al. 1985; Choudhary and Singh 1987). In contrast, DNA data suggest there is population subdivision of *D. simulans* both within Africa and between Africa and the New World. Begun and Aquadro (1995) showed significant differentiation ($F_{ST} = 0.25$) at the vermilion locus at synonymous sites between a North American and a Central African population. Furthermore, they showed that most of the sequence variants were not shared. Hamblin and Veuille (1999) surveyed *vermilion* and *G6pd* and showed that there is significant population subdivision in four African populations. Irvin et al. (1998) surveyed for variation at 34 microsatellite loci and observed less vari-

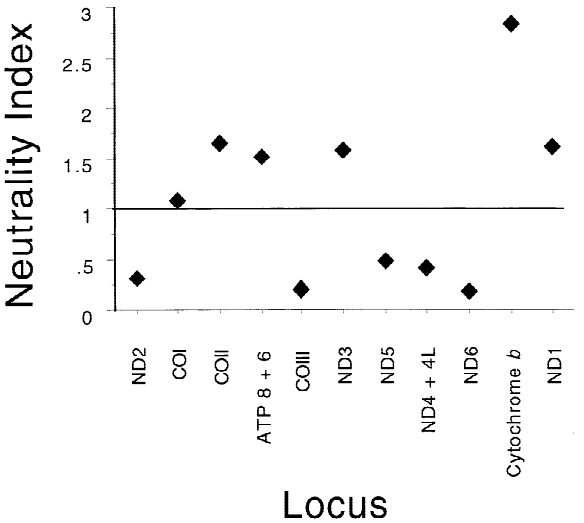

Fig. 6. Neutrality indices determined from MK tests comparing *D. melanogaster* with *D. simulans*. Strict neutrality has an index of 1.0. Values greater than 1.0 typically indicate an excess of amino acid variation within species, whereas values less than 1.0 indicate an excess of nonsynonymous changes, relative to the number of synonymous substitutions.

ability in North and South American *D. simulans* populations than for a Zimbabwe population. Unfortunately, none of these studies included representatives of each major mitochondrial type. In this study, there is no evidence of a congruent signal in the mtDNA data and the autosomal data from *Adhr*. Notably, bootstrap pseudosampling of the nuclear data support monophyly of three clades [(MD112, SC00), (DSR, DSW, HW09), (TT00, C167, HW00)], each of which contains lines with distinct mitochondrial haplotypes (Fig. 1). These data show that there is no evidence for genetic differentiation at *Adhr*.

HKA tests demonstrate that there is a relative reduction in mitochondrial variability in each haplotype compared to the autosomal locus. A single evolutionary force may be linked to this pattern. *Wolbachia* is one possible candidate for inducing a selective sweep. Alternatively, it is possible that distinct evolutionary forces are acting on the mtDNA and aDNA of the three haplotypes. Turelli (1994) suggested that it is unlikely that *Wolbachia* could directly reduce gene flow between populations infected with different strains of the parasite. Rather, any reduction in gene flow may occur following allopatric host–parasite coevolution. In contrast, Wade and Chang (1995) suggest that, under some circumstances, a *Wolbachia*-induced selective sweep may cause hitchhiking of autosomal DNA.

The MK test shows that, in the mtDNA, there is a significant excess of nonsynonymous substitutions relative to synonymous changes within the *si*I and *si*II haplotypes (6/8 and 5/3, respectively; Fig. 5). The *si*III haplotype has a small number of segregating sites but exhibits the same trend. This may be caused by the accumulation of slightly deleterious nonsynonymous substitutions within each haplotype, which rarely become fixed among haplotypes. These intrahaplotypic results are similar to those commonly observed in the mtDNA of closely related species at the cytochrome *b* gene (Nachman 1998; Rand and Kann 1998). Within the interhaplotypic *si*I lineage there is no significant difference in the number of synonymous and nonsynonymous changes (27/138). However, if *D. sechellia* is included (Ballard 2000a) this lineage can be divided into two distinct segments (Fig. 5). One of these segments is ancestral to *si*I and *D. sechellia*, and the more terminal segment includes only *si*I. Ballard (2000a) observed a relative excess of nonsynonymous changes in the ancestral segment of the branch. There is a deficiency of nonsynonymous substitutions relative to synonymous changes in the interhaplotypic branch to *si*II (13/142; Fig. 5). Ballard (2000a) observed that the COI locus has an elevated rate of synonymous substitutions in this branch. There is also a deficiency of nonsynonymous changes relative to synonymous substitutions in the interhaplotypic branch to *si*III (12/104; Fig. 5). Ballard (2000a) noted that there was an excess of synonymous S codons (changes from A/T $\rightarrow$ C/G) and a deficiency of silent changes in this lineage. *D. simulans si*III has been collected on Madagascar and Reunion Island, and it is likely to have a small historical population size. In this study, there is an excess of nonsynonymous substitutions in the interspecific lineage between *D. simulans* and *D. melanogaster*. If this lineage is divided into two segments by the addition of *D. yakuba* (Ballard 2000a) the relative excess of nonsynonymous substitutions occurs in the lineage from the common ancestor to *D. melanogaster*. Data from nuclear loci suggest that the ancestral effective population size of *D. melanogaster* was smaller than that of *D. simulans* (Aquadro et al. 1988; Akashi 1996). If true, the relative excess of amino acid substitutions on the lineage to *D. melanogaster* may be the result of a relaxation of selection against deleterious mutations.

Finally, polymorphism statistical tests (i.e., Tajima's D, Fu and Li's D*, and Fu's Fs) tend to be negative within *D. simulans* haplotypes (Table 7), suggesting recent directional selection, background selection, or a population expansion. These results are consistent with the phylogenetic hypothesis and the HKA test results that there has been genetic hitchhiking within each haplotype. When the complete mtDNA data set is considered, these tests are significantly positive (Table 7). Positive values may indicate possible balancing selection, population subdivision, excess of old mutations, a deficiency of young mutations, or a sampling strategy that has artificially increased the number of variants at intermediate frequency. The total sample does not correspond to a random population sample. In fact, the sampling strategy was to include approximately equal subsamples of the

Table 7. Polymorphism tests of the neutral theory of molecular evolution[a]

	D. simulans		*si*I		*si*II		*si*III	
	mtDNA	*Adhr*	mtDNA	*Adhr*	mtDNA	*Adhr*	mtDNA	*Adhr*
Tajima's D	2.40*	−1.38	−1.06	−0.31	−0.15	−0.88	−1.51	−1.16
Fu and Li's D*	1.64*	−1.41	−1.11	−0.25	−0.35	−0.82	−1.68	−1.21
Fu's Fs	16.53**	−7.86***	−1.87	0.24	0.2	−0.09	−0.38	−4.66**

[a] Asterisks show the p values ($p < 0.05$*, $p < 0.01$**, $p < 0.001$***)

three highly differentiated haplotypes. This strategy will cause an artificial increase of polymorphic sites with variants segregating at intermediate frequencies. This alone may have caused the positive sign of the latter statistic.

A primary finding of this study is that significantly different evolutionary forces are influencing the evolution of *D. simulans* mtDNA and aDNA. One possible mechanism driving this is the host–parasite interaction between the *D. simulans* haplotypes and the strains of *Wolbachia.* The *si*I haplotype has been collected in New Caledonia, Tahiti, Hawaii, and the Seychelles Islands (Baba-Aïssa et al. 1988; James and Ballard 2000). The *Wolbachia* strain *w*Ha infects more than 99% of all *si*I haplotype flies assayed (Turelli and Hoffmann 1995; James and Ballard 2000). Most of these flies from New Caledonia and the Seychelles Islands are also infected with the *w*No strain of *Wolbachia.* These double infections appear to result from horizontal transfer (Montchamp-Moreau et al. 1991; Rousset et al. 1992; James and Ballard 2000). The *si*II type has a worldwide distribution, but has not been collected on the Pacific Islands of New Caledonia, Hawaii, and Tahiti. Phylogenetic analysis of the *si*II haplotype suggests there are two subtypes *si*IIA and *si*IIB (James and Ballard 2000; Ballard unpublished data). The *Wolbachia* strain *w*Ri infects the *si*IIA haplotype, while the *w*Au strain infects the *si*IIB haplotype (Hoffmann et al. 1986, 1996, Hoffmann and Turelli 1988; O'Neill and Karr 1990; James and Ballard 2000). The *si*III haplotype occurs in Madagascar and Reunion Island. About 30% of these flies are infected with the *w*Ma strain of *Wolbachia* (Rousset et al. 1992; James and Ballard 2000).

The *w*Ha and *w*Ri strains of *Wolbachia* cause cytoplasmic incompatibility and are bidirectionally incompatible with each other (O'Neill and Karr 1990). Infected females can produce normal numbers of progeny if they mate with an uninfected male or a male infected with the same strain of bacteria, but uninfected females produce fewer progeny when they mate with infected males (Hoffmann et al. 1986; Hoffmann and Turelli 1988; O'Neill and Karr 1990; James and Ballard 2000). Once a threshold infection frequency is reached, infected females are at a selective advantage over uninfected individuals, and the infection may be expected to sweep through uninfected populations (Caspari and Watson 1959). In these cases, the mtDNA will be carried passively as the microorganism sweeps through a population causing a reduction in mtDNA diversity. In this study, HKA tests indicate a significant deficiency of mitochondrial variation in the *si*I and the *si*II mitochondrial haplotypes compared to intron 1 of the *Adhr* locus. The *Wolbachia*-induced bidirectional incompatibility combined with the high levels of infection in the field (close to 100% for *si*I and 40–70% for *si*IIA; Turelli and Hoffmann 1995; James and Ballard 2000) provide a compelling mechanism for the maintenance of allopatry of the *si*I and *si*IIA haplotypes. An infected *si*IIA *w*Ri individual would produce low numbers of offspring if it mated with a *si*I *w*Ha fly. Similarly, an *si*IIA female that is not infected with *Wolbachia* (*w*-) is incompatible with a *si*I *w*Ha male. An *si*IIA *w*-male would produce normal numbers of offspring if mated with a *si*I *w*Ha–infected female. In this case there would be gene flow, but the offspring would be *si*I *w*Ha. This hypothesis predicts that *si*IIA *w*Ri and *si*I *w*Ha flies are unlikely to occur in a stable equilibrium except when both combinations are equally fit. Baba-Aïssa et al. (1988) reported that *si*I and *si*II flies were collected from the Seychelles. However, they do not report on the frequencies of each haplotype or of their *Wolbachia* infection status.

Neither the *w*Au nor the *w*Ma strain of *Wolbachia* induces high levels of cytoplasmic incompatibility and the *si*IIB and *si*III haplotypes can be found at intermediate frequency in Madagascar (James and Ballard 2000). In this study, an HKA test exhibited a significant deficiency of mitochondrial variation in the *si*III haplotype compared to intron 1 of the *Adhr* locus. Ballard (unpublished) has also found evidence for reduced mitochondrial diversity of *si*IIB in Madagascar. If *w*Ma and *w*Au directly effect mitochondrial diversity, it is likely that the strains either caused incompatibility in the recent past and/or they significantly increase the fitness of infected individuals (Turelli 1994). Hoffmann et al. (1996) reported that females infected with the *w*Au strain produced more eggs than lines cured of infection with tetracycline, however, this difference was not significant. An alternative explanation is that populations with the *si*IIB and the *si*III mtDNA are small or recently derived.

This study investigates the nucleotide variation in

complete mitochondrial genomes and intron 1 of *Adhr* to distinguish between evolutionary forces acting at these loci. Overall at least three lines of evidence suggest that significantly different forces are influencing the evolution of *D. simulans* mtDNA and aDNA. It is possible that a variety of independent events within each haplotype generated the observed patterns of polymorphism and divergence. Alternately, a single factor, such as maternally inherited *Wolbachia*, may have caused the significantly different substitution patterns that are observed in the mtDNA and aDNA.

Acknowledgments. I thank Kirrie Ballard and Andrew Berry for help collecting flies; ICIPE for assistance in obtaining collection permits; Michael Ashburner, Tim Karr, Robin Russell, Michel Solignac, and Michael Turelli for providing isofemale lines; Jeremy Kirchman and Matt Dean for help with sequencing; and Richard Hudson for advice on the statistical tests. Adam Eyre-Walker, Martin Kreitman, and two anonymous reviewers made significant and constructive comments on the manuscript. All sequencing was done in the Pritzker Laboratory of Molecular Systematics and Evolution, Field Museum. Funds were provided by National Science Foundation (NSF) Grant No. DEB-9702824 and the Field Museum Marshall Field Fund.

References

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

Aquadro CF (1992) Why is the genome variable? Insights from *Drosophila.* Trends Genet 8:355–362

Aquadro CF, Lado KM, Noon WA (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

Baba-Aïssa F, Solignac M, Dennebouy N, David JR (1988) Mitochondrial DNA variability in *Drosophila simulans:* quasi absence of polymorphism within each of the three cytoplasmic races. Heredity 61:419–426

Ballard JWO (2000a) Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. J Mol Evol 51:48–63

Ballard JWO (2000b) When one is not enough: introgression in *Drosophila.* Mol Biol Evol 17:1126–1130

Ballard JWO, Kreitman M (1994) Unraveling selection in the mitochondrial genome of *Drosophila.* Genetics 138:757–772

Ballard JWO, Hatzidakis J, Karr TL, Kreitman M (1996) Reduced variation in *Drosophila simulans* mitochondrial DNA. Genetics 144:1519–1528

Begun D, Aquadro CF (1993) African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365:548–550

Begun DJ, Aquadro CF (1995) Molecular variation at the *vermillion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans.* Genetics 149:1019–1032

Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ (1993) Partitioning and combining data in phylogenetic analysis. Syst Biol 42:384–397

Caspari E, Watson GS (1959) On the evolutionary importance of cytoplasmic incompatibility in mosquitoes. Evolution 13:568–570

Choudhary M, Singh RS (1987) A comprehensive study of genetic variation in *Drosophila melanogaster*. III. Variations in genetic structure and their causes between *Drosophila melanogaster* and its sibling species *Drosophila simulans.* Genetics 117:697–710

Clary DO, Wolstenholme DR (1985) The mitochondrial molecule of *Drosophila yakuba:* nucleotide sequence, gene organization and genetic code. J Mol Evol 22:252–271

Cohn VH, Thompson MA, Moore GP (1984) Nucleotide sequence comparison of the *Adh* gene in three drosophilids. J Mol Evol 20:31–37

Efron B (1982) The jackknife, the bootstrap, and other resampling plans. Conf Board Math Sci Soc Ind Appl Math 38:1–92

Farris JS, Källersjö M, Kluge A, Bult C (1995) Testing significance of incongruence. Cladistics 10:315–319

Felsenstein J (1985) Confidence limits on phylogenies: an approach using bootstrap. Evolution 39:783–791

Fu YX (1997) Statistical test of neutrality of mutations against population growth hitchhiking and background selection. Genetics 147: 915–925

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Hale LR, Singh RS (1987) Mitochondrial DNA and genetic structure in populations of *Drosophila melanogaster*. Mol Biol Evol 4:622–637

Hamblin MT, Veuille M (1999) Population structure among African and derived populations of *Drosophila simulans:* evidence for ancient subdivision and recent admixture. Genetics 153:305–317

Hoffmann AA, Turelli M (1988) Unidirectional incompatibility in *Drosophila simulans:* inheritance, geographic variation and fitness effects. Genetics 119:435–444

Hoffmann AA, Turelli M, Simmons GM (1986) Unidirectional incompatibility between populations of *Drosophila simulans.* Evolution 40:692–701

Hoffmann AA, Turelli M, Harshman LG (1990) Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans.* Genetics 126:933–948

Hoffmann AA, Clancy D, Duncan J (1996) Naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. Heredity 76:1–8

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. Mol Biol Evol 9:138–151

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Hyytia P, Capy P, David JR, Singh RS (1985) Enzymatic and quantitative variation in European and African populations of *Drosophila simulans.* Heredity 54:209–217

Inohira K, Hara T, Matsuura ET (1997) Nucleotide sequence divergence in the A+T rich region of mitochondrial DNA in *Drosophila mauritiana.* Mol Biol Evol 14:814–822

Irvin SD, Wetterstrand, Hutter CM, Aquadro CF (1998) Genetic variation and differentiation at microsatellite loci in *Drosophila simulans:* evidence for founder effects in new world populations. Genetics 150:777–790

James AC, Ballard JWO (2000). Expression of cytoplasmic incompatibility in *Drosophila simulans* and its impact on infection frequencies and distribution of *Wolbachia pipientis.* Evolution (in press)

Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking effect" revisited. Genetics 123:887–899

Kondo R, Satta Y, Matsuura ET, Ishiwa H, Takahato N, Chigusa SI (1990) Incomplete maternal transmission of mitochondrial DNA in *Drosophila.* Genetics 126:657–663

Lewis DL, Farr CL, Farquhar AL, Kaguni LS (1994) Sequence organization, and evolution of the A+T region of *Drosophila* mitochondrial DNA. Mol Biol Evol 11:523–538

Lewis DL, Farr CL, Kaguni LS (1995) *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequences and evolutionary comparisons. Insect Mol Biol 4:263–278

Maynard Smith J, Haigh J (1974) The hitchhiking effect of a favorable gene. Genet Res 23:23–25

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila.* Nature 351:652–654

Mickevich MF, Farris JS (1981) The implications of congruence in *Menidia.* Syst Zool 30:351–370

Montchamp-Moreau C, Ferveur J-F, Jacques M (1991) Geographic distribution of three cytoplasmic incompatibility types in *Drosophila simulans.* Genetics 129:399–407

Nachman MW (1998) Deleterious mutations in animal mitochondrial DNA. Genetica 103:61–69

Nachman MW, Brown WM, Stoneking M, Aquadro CF (1996) Nonneutral mitochondrial variation in humans and chimpanzees. Genetics 142:953–963

O'Neill SL, Karr TL (1990) Bidirectional incompatibility between conspecific populations of *Drosophila simulans.* Nature 384:178–180

Rand DM, Kann LM (1996) Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila,* mice and humans. Mol Biol Evol 13:735–748

Rand DM, Kann LM (1998) Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica 102–103:393–407

Rand DM, Dorfsman M, Kann LM (1994) Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. Genetics 138:741–756

Roff DA, Bentzen P (1989) The statistical analysis of mitochondrial DNA polymorphisms: chi^2 and the problem of small samples. Mol Biol Evol 6:539–545

Rousset F, Vautrin D, Solignac M (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

Rozas J, Rozas R (1997) DnaSP version 3.0: a novel software package for extensive molecular population genetics analysis. Comp Applic Biosci 13:307–311

Schaeffer SW, Aquadro CF (1987) Nucleotide sequence of the *Adh* gene region of *Drosophila pseudoobscura:* evolutionary change and evidence for an ancient gene duplication. Genetics 117:61–73

Schwarze SR, Weindruch R, Aiken JM (1998) Decreased mitochondrial RNA levels without accumulation of mitochondrial deletions in aging *Drosophila melanogaster.* Mutat Res 382:99–107

Simmonsen KL, Churchill G, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141: 413–429

Solignac M, Monnerot M (1986) Race formation and introgression within *Drosophila simulans, D. mauritiana* and *D. sechellia* inferred from mitochondrial DNA analysis. Evolution 40:531–539

Solignac M, Monnerot M, Monnerot J-C (1983) Mitochondrial heteroplasmy in *Drosophila mauritiana.* Proc Nat Acad Sci 80:6942–6946

Strobeck C (1987) Average number of nucleotide difference in a sample from a single subpopulation: a test for population subdivision. Genetics 117:149–153

Sumner CJ (1991) Nucleotide polymorphism in the alcohol dehydrogenase duplicate locus of *Drosophila simulans:* implications for the neutral theory. Undergraduate thesis, Princeton University, Princeton, NJ

Swofford DL (1998) PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4b3. Sinauer Associates, Sunderland, MA

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 105:437–460

Turelli M (1994) Evolution of incompatibility-inducing microbes and their hosts. Evolution 48:1500–1513

Turelli M, Hoffman AA, McKechnie SW (1992) Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. Genetics 132:713–723

Turelli M, Hoffmann AA (1995) Cytoplasmic incompatibility in *Drosophila simulans:* dynamics and parameter estimates from natural populations. Genetics 140:1319–1338

Wade J, Chang NW (1995) Increased male fertility in *Tribolium confusum* beetles after infection with the intracellular parasite *Wolbachia.* Nature 373:72–74