J Mol Evol (2000) 51:48–63
DOI: 10.1007/s002390010066

JOURNAL OF MOLECULAR EVOLUTION

Comparative Genomics of Mitochondrial DNA in Members of the *Drosophila melanogaster* Subgroup

J. William O. Ballard

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

Received: 30 July 1999 / Accepted: 16 March 2000

Abstract. In this study, a comparative genomics approach is employed to investigate the forces that shape evolutionary change in the mitochondrial DNA (mtDNA) of members of the *Drosophila melanogaster* subgroup. This approach facilitates differentiation of the patterns of variation resulting from processes acting at a higher level from those acting on a single gene. The mitochondrial genomes of three isofemale lines of *D. simulans* (*si*I, -II, and -III), two of *D. melanogaster* (Oregon R and a line from Zimbabwe), and *D. mauritiana* (*ma*I and -II), and one of *D. sechellia* were sequenced and compared with that derived from *D. yakuba.* Data presented here indicate that at least three broad mechanisms shape the evolutionary dynamics of mtDNA in these taxa. The first set of mechanisms is intrinsic to the molecule. Dominant processes may be interpreted as selection for an increased rate of replication of the mtDNA molecule, biases in DNA repair, and differences in the pattern of nucleotide substitution among strands. In the genes encoded on the major strand (62% of the coding DNA) changes to or from C predominate, whereas on the minor changes to or from G predominate. The second set of mechanisms affects distinct lineages. There are evolutionary rate differences among lineages, possibly owing to population demographic changes or changes in mutational biases. This is supported by the heterogeneity found in synonymous, nonsynonymous, and silent substitutions. The third set of mechanisms differentially affects distinct genes. A maximum-likelihood sliding-window analysis detected four disjunct regions that have a significantly different nucleotide substitution process from that derived from the complete sequence. These data show the potential for comparative genomics to tease apart subtle forces that shape the evolution of DNA.

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

Introduction

This study was motivated by the observation that the evolutionary rates of synonymous changes at two localized regions of mitochondrial DNA (mtDNA) are significantly different in distinct mitochondrial lineages of *Drosophila simulans* (Ballard 1998). This paper attempts to understand this unexpected heterogeneity in mutational features and rates among the mitochondrial genomes. It also aims to determine how frequently this heterogeneity in the pattern of substitutions is observed in distinct lineages of mtDNA from closely related species within the *D. melanogaster* subgroup.

To assign changes to the appropriate lineage a representative from each major mitochondrial haplotype within the *D. melanogaster* subgroup was included. Within *D. simulans* there are three major mitochondrial types *si*I, -II, and -III (Solignac et al. 1986; Satta and Takahata 1990; Montchamp-Moreau et al. 1991). A representative of each haplotype was included. A single representative of *D. sechellia* and both major mitochondrial lineages of *D. mauritiana* (*ma*I and -II) and *D. melanogaster* were included (Solignac and Monnerot 1986; Solignac et al. 1986; Rand et al. 1994). This study does not include multiple representatives of each lineage. However, variation within each lineage is very low (Solignac

et al. 1986; Solignac and Monnerot 1986; Baba-Aïssa et al. 1988; Ballard and Kreitman 1994; Rand et al. 1994; Ballard 2000).

To ascribe changes to the correct region of mtDNA, the complete mitochondrial genome was sequenced. Sequencing complete genomes permits the detection of patterns that operate at a higher level than the individual gene, but that nevertheless affect the gene (Blattner et al. 1997; Alm et al. 1999). The vast majority of studies focus on a single gene, and, because these studies cannot investigate processes above the level of the gene, they are vulnerable to misinterpreting patterns as belonging to individual gene processes. Targeting specific loci can determine whether they are under selection (Kreitman 1983; McDonald and Kreitman 1991; Ballard and Kreitman 1995), while contrasting multiple loci can determine whether they are under similar selective pressures (Hudson et al. 1987). This is the first study that attempts to apportion patterns of evolutionary change to the appropriate level.

The evolution of the mtDNA molecule in *Drosophila* is both dynamic and complex. Ballard (1998) observed that rates of synonymous changes in two regions of mtDNA are significantly different in distinct mitochondrial lineages of *D. simulans.* One 2,531-bp region includes the majority of cytochrome oxidase I (COI), NADH dehydrogenase subunit 2 (ND2), and three tRNAs (Satta and Takahata 1990). The second region, 1,137 bp in length, includes the complete cytochrome *b* gene (Ballard and Kreitman 1994). Within the 2,531-bp region there are 19 changes on the *si*I lineage, 45 on the *si*II lineage, and 22 on the *si*III lineage. A relative rate test (Tajima 1993) using *D. melanogaster* as the outgroup shows a highly significant increase in the evolutionary rate of the *si*II lineage compared to both *si*I and *si*III. In contrast, in the 1,137-bp region there are 26 substitutions on the *si*I lineage and 0 and 5 on the *si*II and *si*III lineages, respectively.

There are at least three broad mechanisms that may induce the observed heterogeneity in the pattern of synonymous changes at the two localized regions of mtDNA. First, there may be rapid shifts in mutational tendencies that are gene-specific (or possibly mitochondrial region–specific). Second, there may be selection, either weak or strong, for synonymous changes that also shifts rapidly due to changes in the environment or in the demography of the population. Third, heterogeneity between regions may arise from any of the above mechanisms plus recombination. The observed localized changes in evolutionary rate cannot be explained by selection at linked sites with genetic hitchhiking (Maynard Smith and Haigh 1974; Kaplan et al. 1989) or by background selection (Charlesworth et al. 1993). Nor can they be explained by interactions with a maternally inherited factor (Ballard and Kreitman 1994; Ballard et al. 1996; Ballard 1998, 2000).

Biases toward A/T-ending codons has been observed in the mtDNA of *Drosophila* (Clary and Wolstenholm 1985). This suggests that there may be a bias in the mutation rate toward A and T. This may be caused by factors that impinge on rates of DNA damage (Martin 1995) and the availability of each nucleotide in the cellular medium of the mitochondrion (Xia 1996). Mutational biases have also been shown to occur on each strand of DNA (Garesse 1988; Rand and Kann 1998). Garesse (1988) noted that T $\rightarrow$ C substitutions predominate on the major strand of *D. melanogaster* (62% of the coding DNA), while A $\rightarrow$ G substitutions predominate in the genes encoded on the minor strand. Rand and Kann (1998) further noted that in *D. melanogaster* and *D. pseudoobscura* there was a clear bias for T $\rightarrow$ C fixations in the cytochrome *b* gene (major strand) and A $\rightarrow$ G fixations in the ND5 gene (minor strand). In this study, genes transcribed from the minor strand are complemented.

The pattern of substitutions in mtDNA lineages may be affected by changes in the environment or demography of the population. In mtDNA, it has been suggested that slightly deleterious nonsynonymous or synonymous substitutions accumulate within species (Nachman et al. 1994; Nachman 1998; Ballard and Kreitman 1994; Rand et al. 1994; Rand and Kann 1998). Ohta (1972, 1973) predicted that very slightly deleterious mutants are governed by random drift in small populations because they behave as if selectively neutral, but in large populations, these same mutants are effectively selected against. Small populations may also accumulate deleterious mitochondrial mutations at an increased rate because there is a greater probability of each individual failing to pass on the best mitochondrial haplotype (Bergstrom and Prichard 1998). The exact population sizes of each major mitochondrial haplotype within the *D. melanogaster* subgroup are not known. However, some predictions can be made about relative abundances. For example, *D. sechellia* and *D. mauritiana* are island endemics with small population sizes (Tsacas and David 1974; Lachaise et al. 1988; Hey 1994). Concordant with the known geographic distributions of the *D. simulans* haplotypes the levels of nucleotide diversity are in the order *si*II > *si*I > *si*III (Ballard 2000). The *si*II haplotype has a worldwide distribution but has not been collected in the same Pacific Islands as *si*I. The *si*I haplotype has only been collected on Indian/Pacific Islands and may have been subjected to a series of population size reductions. The *si*III haplotype has been collected from Madagascar and Reunion, where it is sympatric with the *si*II haplotype.

This study also tests for recombination among the distinct lineages. A number of recent studies have suggested that mtDNA may recombine in humans and some fungi (Saville et al. 1998; Eyre-Walker et al. 1999; Hagelberg et al. 1999; Awadalla et al. 1999). Several methods

for detecting recombination have been proposed (Milkman and Crawford 1983; Hudson and Kaplan 1985; Stephens 1985; Dykhuizen and Green 1991; Maynard Smith 1992). In this study, the homoplasy test (Maynard Smith and Smith 1998) and the sliding-window maximum likelihood method (Grassly and Holmes 1997) are employed. The homoplasy test is a modification of the "cladistic" method of Hudson and Kaplan (1985) and aims to determine whether any observed homoplasies have occurred by mutation or there has been recombination. This test is appropriate when the sequences differ by about 1–5% of nucleotides. The sliding-window maximum likelihood method aims to detect regions that differ significantly from the model that is derived from the complete data set. The latter method has the advantage that it can detect regions with different evolutionary histories (recombination) as well as regions with significantly different rates (selection). In this study, there is no evidence for recombination between the mitochondrial genomes included. The absence of recombination simplifies the analysis because there is one genealogy. This enhances the utility of the *Drosophila* mitochondrial genome as a model system for investigating genome-wide effects.

Materials and Methods

Lines Employed. The complete mitochondrial genomes from eight isofemale lines were sequenced and the data compared with that derived from *D. yakuba* (Clary and Wolstenholme 1985). *D. yakuba* is a canonical outgroup of the *melanogaster* subgroup. The A+T rich region is difficult to align (Lewis et al. 1994; Inohira et al. 1997) and is not included in this analysis.

Two isofemale lines of *D. melanogaster* are included—Oregon R and Zimbabwe 53. The Oregon R line was collected in Roseburg, OR, around 1925 by D.E. Lancefield. Lewis et al. (1994) published the A+T rich region of this line. The *D. melanogaster* Zimbabwe 53 line, referred to here as *D. melanogaster* Zimbabwe, was collected in 1990 at the Sengwa Wildlife Preserve in Africa. 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 *D. melanogaster* sequence of Lewis et al. (1995) is not included because it is a chimera of different isofemale lines.

Three isofemale lines of *D. simulans* are included, one from each distinct mitochondrial haplotype. Distinct strains of *Wolbachia* are known to infect each mitochondrial haplotype. In some cases *Wolbachia* causes bidirectional incompatibility between lines infected with distinct strains (Montchamp-Moreau et al. 1991; Rousset et al. 1992). This incompatibility has the potential to reduce gene flow between mitochondrial lineages and is one explanation for the persistence of the three distinct haplotypes (James and Ballard 2000). The *si*I haplotype is known to occur only on Tahiti, New Caledonia, Hawaii, and the Seychelles (Baba-Aïssa et al. 1988). The author collected *D. simulans si*I TT01 in Papeete on Tahiti Nui in 1998. The *si*II haplotype has a worldwide distribution but has not been collected on any Pacific Islands (Baba-Aïssa et al. 1988). *D. simulans si*II DSR was collected in 1984 from Riverside, CA (DSR), by A. Hoffman (LaTrobe University, Melbourne). The *si*III haplotype has been collected from Madagascar and Reunion Island, where it is sympatric with the *si*II haplotype (Solignac and Monnerot 1986). *D. simulans si*III MD199 was collected by the author in Joffreville, Madagascar, in 1998. These three lines are referred to here as *D. simulans si*I, -II, and -III, respectively.

Two isofemale lines of *D. mauritiana* are included, one from each distinct haplotype (*ma*I and -II) (Solignac and Monnerot 1986). O. Kitagawa collected *D. mauritiana ma*I BG1 in Mauritius in 1981 and *D. mauritiana ma*II G52 in Mauritius in 1985. These two lines are referred to here as *D. mauritiana ma*I and -II, respectively.

One isofemale line of *D. sechellia* is included. The line was the first collected line of the species. It was described by Isacas and Bächli (1981).

DNA Preparation, Polymerase Chain Reaction (PCR), and Cycle Sequencing. DNA from all lines was extracted using the PureGene® Kit (Gentra) following the DNA isolation from fixed tissue protocol. In all cases, the DNA was extracted from individuals 7–14 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). Fourteen overlapping fragments were PCR amplified for 32 cycles. Each cycle consisted of 15-s denaturation at 94°C, 5-s annealing between 50–54°C, and 90-s extension at 72°C. Each amplicon was electrophoresed on a 1% agarose gel to verify size and the remainder of the reaction or reactions were cleaned and concentrated with Microcon 100s (Amicon).

Both strands were sequenced using Taq-Dye Deoxy Terminator Cycle sequencing (Applied Biosystems) employing 100–200 ng of the amplicon (500–2,000 bp) and 25 ng of primer. Eighty-four to 96 independent cycle sequencing reactions were employed to sequence both strands of each mitochondrial molecule. For each cycle sequencing reaction, 120 ng of template and 25 ng of primer were added to 4.25 μl of Taq-Dye Deoxy Terminator premix (Applied Biosystems) prior to cycle sequencing for 25 cycles. The cycling profile was 15-s denaturation at 95°C, 5-s annealing at 50°C, and 4-min extension at 65°C. Cycle sequencing reactions were NH_4OAc precipitated and dried. The dried samples were resuspended in 4 μl of deionized formamide and 50 m*M* EDTA (5:1) and 2 μl electrophoresed on an Applied Biosystems 377 DNA sequencer.

Sequences were imported into the Sequencher® software program and the chromatograms investigated. Each sequence was completed before the next isofemale line was commenced. There was no evidence of heteroplasmy or large-scale deletions that have been reported in some lines of *D. mauritiana* (Solignac et al. 1983; Kondo et al. 1990). Three inconsistencies between the sequences derived from independent PCR products were resolved by reamplifying and resequencing the region of interest. These were considered PCR errors.

Genome Investigations. The sequence data were aligned against the published *D. yakuba* sequence (Clary and Wolstenholme 1985) using Sequencher. The tRNAs were defined by tRNAscan-SE v1.11 (Lowe and Eddy 1997) and by comparison with the *D. yakuba* sequence.

Pooling genes transcribed from both strands has the potential to mask any strand-based biases. In an attempt to minimize this bias, each triplet transcribed from the negative strand loci was reverse complemented. This procedure retains the order of amino acids but reverses the relative positions of the first and third codons. The data file was then validated by comparing the amino acid sequences with the original data file. The tRNAs and rRNAs transcribed from the minor strand were then complemented. The result is that all loci can be analyzed in a single sequence with minimal alteration of nucleotide position. For consistency, all references to specific positions and base changes refer to this modified sequence. The aligned sequences were imported into DnaSP 3.14 (Rozas and Rozas 1997) where the protein coding regions, tRNAs, rRNAs, and intervening sequences were defined.

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 alleles was analyzed by parsimony and maximum likelihood using PAUP* (Swofford 1998). For all likelihood analyses, the HKY model (Hasegawa et al. 1985) enforcing a molecular clock with two substitution types was employed. The proportion of constant sites

and the transition/transversion (Ti/Tv) ratio were estimated from the data.

Bootstrapping (Efron 1982; Felsenstein 1985) was used to test monophyly. For this study, 1,000 pseudosamples were generated to estimate the bootstrap proportions. The completely resolved tree derived from the complete data set was well supported by both parsimony and likelihood analyses and is referred to as the "true" mitochondrial genealogy. It is not assumed that this is the species tree.

To investigate whether there was any recombination between the six sequences in the *D. simulans* triad, I employed the homoplasy test (Maynard Smith and Smith 1998) and the sliding-window maximum likelihood method (Grassly and Holmes 1997). The homoplasy test determines if there is a statistically significant excess of homoplasies in the phylogenetic tree derived from the data set compared to an estimate of the number of homoplasies expected by repeated mutation in the absence of recombination. An excess of homoplasies is considered a signature of recombination. In this study a 3,493-bp data set was constructed. This data set contained third codon positions, omitting initiation and termination codons and codons where there was amino acid variation. Homoplasy ratios were calculated with "datain.bas" and then "exph.bas" as reported in Maynard Smith and Smith (1998). This is a number with expectation 0 for a clonal population, and 1.0 for free recombination. The effective number of sites was considered to be in the range 2,095–3,493.

Genome-Wide Biases. There is a marked A+T bias in *Drosophila* mtDNA. Various hypotheses have been proposed to explain differences in nucleotide composition. There may be a mutation bias influenced by factors that impinge on rates of DNA damage (Martin 1995). This bias may be caused by the susceptibility of dGTP and dCTP to oxidative damage and the preferential incorporation of dATP opposite abasic sites by polymerase. It may also be caused by hydrolytic deamination, which can convert cytosine to uracil. There may also be selection for a high A/T content. A/T-rich genomes may replicate more quickly than G/C-rich genomes and, if all else is equal, have a selective advantage in a heteroplasmic population. In a natural population of *D. simulans*, heteroplasmy has been observed with rates of about 6% (Kondo et al. 1990). The free energy of base stacking depends on the particular combination of adjacent bases, so the overall free energy differs for each doublet combination of base pairs. Lewin (1997) reported that the overall free energy of A•U doublet pairs has ΔG values between −0.9 and −1.0, doublets containing one A•U and one G•C vary between −1.7 and −2.3 and doublets containing only G•C vary from −2.0 to −3.4 kcal/mol. The underlying mutation and selection biases may be exposed in partitions with low functional constraint. In the mitochondrial genome it may be hypothesized that the intervening spacer regions are one such partition. In this study, the proportion of each nucleotide in the intervening spacer regions is compared with that in each codon position, in the tRNAs, and in the rRNAs.

There may also be a codon bias in mtDNA. Xia (1996) argued that, within a codon family, a codon ending with a nucleotide that is most abundant in the cellular medium should be used most frequently in the mRNA. In mitochondria, this is likely to be adenine because the concentration of cellular ATP is much higher in mitochondria than the other three nucleotides (Bridger and Henderson 1983). Xia's (1996) transcriptional hypothesis of codon usage predicts relative A-richness and T-deficiency in the coding strand. To investigate these alternate hypotheses the mean number of A- and T-ending codons at fourfold degenerate synonymous codons was compared.

Mutational biases may occur on each strand of DNA (Garesse 1988; Rand and Kann 1998). To investigate strand-specific mutation biases, the synonymous substitutions at third positions are mapped onto the well-supported genealogy using PAUP* (Swofford 1998) and MacClade (Maddison and Maddison 1992).

Substitutions in Each Lineage. If a sufficiently large proportion of mutations in mtDNA are nearly neutral, then mtDNA lineages may be affected by changes in the environment or demography of the population. Ballard (2000) showed that the pattern of substitutions within each of the distinct *D. simulans* haplotypes is not consistent with neutrality, and is similar to that previously reported between distinct species (Nachman 1998; Rand and Kann 1998). To investigate the pattern of substitutions between genes within lineages and between lineages, three contingency analyses were conducted. These tests are modifications of the MK test (McDonald and Kreitman 1991), and they are appropriate when there is no recombination.

The first analysis compared the number of 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 in each lineage. All substitutions were parsimoniously mapped onto the well-resolved phylogeny to investigate the number of changes in each lineage. There was only one nucleotide difference between *si*III and *ma*I, so these terminal lineages were not considered. The null hypothesis is that the proportion of synonymous, nonsynonymous, and silent changes in each lineage should be the same if this ratio is maintained under a strictly neutral model of molecular evolution. As an alternative, small-population-size lineages may accumulate deleterious mitochondrial mutations at an increased rate (Ballard and Kreitman 1994; Nachman et al. 1994; Rand et al. 1994; Bergstrom and Prichard 1998). Silent substitutions are expected to be very close to neutral. However, some are likely to be deleterious. Specifically, silent substitutions that affect tRNA or rRNA secondary structure might be deleterious. In humans, for example, the tRNA Leucine^{UUR} gene is an etiologic hot spot (Moraes et al. 1993). These analyses do not include a multiple hit correction and therefore may underestimate the number of fixed synonymous substitutions in some lineages.

In the second analysis, the directions of synonymous changes in each lineage were compared (Ballard and Kreitman 1994; Akashi 1995). For mitochondrial genes, codon usage cannot be causally linked to tRNA abundances because there is typically only one tRNA for each amino acid. Consequently, it cannot be stated that a codon is "preferred" (Ikemura 1981, 1982). In *Drosophila* mtDNA, there is an A/T bias in synonymous positions. This result is consistent with a bias in the substitutions to A/T compared to C/G. However, it is unlikely that this is the only evolutionary force acting on synonymous sites. Ohta (1972, 1973) predicted that mutants that are mildly deleterious in large populations might be neutral in small populations. To investigate codon usage, all synonymous changes were parsimoniously mapped onto each ingroup lineage using the "true" genealogy. Sites on the branch to *D. yakuba* were not included because they were employed to polarize changes. Synonymous changes were categorized as mutating to W-codons (C/G $\rightarrow$ A/T), to S-codons (A/T $\rightarrow$ C/G), or to WS-codons (A $\leftrightarrow$ T and C $\leftrightarrow$ G). The IUPAC code defines a W as either an A or T and an S as either a C or G. The neutral prediction is that the proportion of the three kinds of synonymous changes should be the same among lineages. As an alternative, consider a model of weak selection against S-codons (Ballard and Kreitman 1994). The prediction is that if synonymous S-codons are slightly deleterious they will drift to fixation more rapidly in smaller populations. Conversely, it may be expected that W-codon mutations will accumulate more quickly in larger populations. WS-codon mutants, under this scenario, are expected to be very close to neutral.

The third analysis contrasts the direction of synonymous changes (W-, S-, and WS-codons) with the number of nonsynonymous and silent substitutions. This analysis facilitates direct comparison of all substitution types. As previously stated, the null hypothesis is that rate differences among mtDNA lineages are expected to apply equally to all loci. However, if S-codons are slightly deleterious, nonsynonymous substitutions and S-codon changes are expected to drift to fixation in small effective populations. W-codon changes are likely to accumulate in larger populations, whereas most silent changes and WS-codon mutants are predicted to be very close to neutral. Post hoc *t* tests were then performed and these results overlaid onto the "true" genealogy on a lineage by lineage basis. These *t* values are a form of standardized residual that indicates what each cell in the table contributes to the

Chi-square statistic. Since they are calculated to follow a standard normal distribution, absolute values greater than 1.96 (for a 0.05 probability level) indicate that the cell in question is significantly different from that expected by chance.

Sliding-Window Analyses. Variability between the nine mitochondrial genomes was analyzed using the sliding-window approach. These analyses started at position 1 and terminated at position 15,034. The 76 unalignable bases and the A+T region were excluded. The A+T region is 2–5 kb in length and extends from position 15,035 to position 1. Given the large size of this region, it was not considered biologically informative to continue the sliding-window past position 15,034.

Small windows are subject to large stochastic variation, while large windows may smooth over small regions of inconsistency. As such, sliding-window analyses are a compromise. Employing SWAN (ver. 1) (Proutski and Holmes 1998), data were plotted as moving averages shifting 40 bp with a window of 1,000 bp. The variability in each column of the alignment was estimated as an entropy function of the nucleotide variation using the following equation:

$$\text{Variability} = -\sum_{i=\text{A,C,G,T}} \frac{n_i}{N} \ln \frac{n_i}{N}$$

where n_i = the numbers of each nucleotide (A, C, G, or T) in a column of the alignment and N = the total number of sequences analyzed.

Variation between the genomes was then investigated with the sliding-window maximum likelihood method, implemented in PLATO (ver 2.11). This method aims to detect regions that conflict with a single phylogenetic topology and nucleotide substitution process derived from the entire sequence. Such deviation along sequences, called spatial phylogenetic variation by Grassly and Holmes (1997), may reflect recombination or varying selective forces along the sequence. This approach calculates the likelihoods for each site independently. It then generates a measure of the average likelihood of a given window with respect to the rest of the sequence. Maximum values of this method are associated with regions showing low likelihoods given the maximum likelihood model derived from the complete sequence. This measure is calculated for window sizes from 10 nucleotides up to half the sequence length and in all possible positions. Regions fewer than 10 bases are excluded from the analysis because they are deemed uninformative and subject to spurious patterns. Grassly and Holmes (1997) showed that the null distribution of this measure for each region size was approximately normal, permitting Z values to be calculated. A potential problem with this analysis is inflated type 1 error. To correct for multiple comparisons, Grassly and Holmes (1997) employed the Bonferroni correction, dividing α by the number of regions tested.

The tree topology, branch lengths, and Ti/Tv ratio were estimated from the data (as described above) and enforced in a HKY (Hasegawa et al. 1985) sliding-window analysis (Grassly and Holmes 1997). Preliminary PLATO analyses were conducted from positions 1–6,000, 4,500–10,500, and 9,000–15,034. Subsequent tests considered positions 1–7,500 and 7,500–15,034 because the preliminary analyses did not suggest that the region around position 7,500 was anomalous. The large memory requirements of this method precluded a single test of the complete genome.

The sliding-window maximum likelihood analyses identified four regions conflicting with a single phylogenetic topology and nucleotide substitution process along the entire sequence. The Kishino and Hasegawa (1989) test was then employed to investigate whether each of these regions had a significantly different topology from the "true" genealogy. However, the results from these tests should be interpreted with caution as, strictly, this test is only valid when two trees being compared are selected on an a priori basis. A region of DNA with a significantly different topology would support the hypothesis that the mitochondrial genome had recombined. Again, no support was found for recombination. One alternative is that there have been one or more changes in evolutionary rate in each anomalous region.

Two additional sets of analyses were performed to investigate the possibility of evolutionary rate variation within the mtDNA genome. The first set of analyses compared the number of parsimony inferred changes in each lineage in each of the four anomalous regions to that determined from the complete data set. Post hoc tests were then employed to test for evolutionary rate variation in specific lineages. In the second set of tests, two additional sliding-window analyses were conducted. These analyses were performed as previously described, but the rRNAs were excluded because there was no evidence for anomalous rates of evolution in either locus. Exclusion of the rRNAs permitted each test to be completed in a single analysis. The first sliding-window analysis considered synonymous substitutions (where synonymous and nonsynonymous changes both occurred at a single site, the nonsynonymous substitutions were replaced with a "?"). The second considered nonsynonymous substitutions (in this case, synonymous substitutions were replaced with a "?"). These analyses do not distinguish inter- and intraspecific/haplotypic variability in terminal branches. Therefore, when necessary, the region of interest was sequenced from additional isofemale lines to gain an estimate of intraspecific/haplotypic variability. These mtDNA genomes were completed and are included in Ballard (2000).

Results

Genomic Investigations

The alignment is straightforward for the majority of the sequences (GenBank #AF200828-32, AF200834, AF200841, and AF200852). However, it is not possible to determine unequivocally the indel events that include positions 5,535–5,584 and 6,022–6,047. As a consequence, 76 bp are deleted from all analyses. 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.

Protein Coding: In *ma*II the putative initiation codon for COI is TTAA. This is the first report of TTAA being used as an initiation codon. ATAA is the COI initiation codon in *D. melanogaster*, *D. sechellia*, and *si*I, while GTAA is the initiation codon at this locus in *ma*I, *si*II, and −III. Clary and Wolstenholme (1983) and de Bruijn (1983) reported that the COI initiation codon in *D. yakuba* and *D. melanogaster* was ATAA. Satta et al. (1987) reported the initiation codon for *D. simulans* from Nairobi, Kenya, was GTAA.

The total number of triplet codons in *D. simulans*, *D. sechellia*, and *D. mauritiana* is 3,723 (Table 1). It is one less in *D. melanogaster* (Table 1). The single amino acid deletion in both *D. melanogaster* lines is observed in the ND5 locus, four amino acids 3′ of the termination codon. In all other sequences, it is a phenylalanine (TTT). Clary and Wolstenholme (1985) reported that the number of triplet codons is 12 more than observed in *D. simulans*, *D. sechellia*, and *D. mauritiana*. In all taxa sequenced in this study, there is either a TAA or TAG termination codon 12 codons 3′ of where Clary and Wolstenholme (1985) proposed that the ND1 locus terminates in *D. yakuba*. Clary and Wolstenholme (1985) reported a leu-

Table 1. The number of occurrences of each codon in the mitochondrial genomes of *D. yakuba*, *D. melanogaster* Oregon R, *D. melanogaster* Zimbabwe (top row from left to right), *D. sechelia*, *D. mauritiana ma*II, *D. mauritiana ma*I (middle row from left to right), *D. simulans si*I, *D. simulans si*II, and *D. simulans si*III (bottom row from left to right). The total number of triplet codons in *D. yakuba*, *D. sechellia*, *D. mauritiana*, and *D. simulans* is 3,723 and in *D. melanogaster* 3,722

Phe	UUU	311	317	319	**Ser**	UCU	121	129	129	**Tyr**	UAU	141	145	142	**Cys**	UGU	40	39	39
		310	317	311			131	133	130			145	149	157			38	37	39
		311	312	311			130	129	130			150	146	157			38	37	39
	UUC	17	15	15		UCC	4	6	6		UAC	28	20	21		UGC	2	2	2
		18	14	18			7	6	5			21	18	12			2	3	1
		17	18	18			6	6	5			18	22	12			2	3	1
Leu	UUA	541	542	540		UCA	102	96	96	**TER**	UAA	6	6	6	**Trp**	UGA	96	97	97
		541	539	538			92	90	96			6	6	6			94	94	94
		549	543	538			97	97	96			7	7	6			97	95	94
	UUG	25	24	24		UCG	3	5	4		UAG	0	1	1		UGG	6	4	4
		20	23	21			3	4	2			1	1	1			7	4	7
		16	22	21			0	2	2			0	0	1			4	6	7
	CUU	36	30	33	**Pro**	CCU	79	78	79	**His**	CAU	65	70	72	**Arg**	CGU	8	14	13
		30	31	32			72	74	76			69	66	66			13	13	13
		31	29	32			74	75	76			67	63	66			13	12	13
	CUC	2	2	1		CCC	3	6	6		CAC	12	9	7		CGC	0	0	0
		0	1	0			8	7	5			10	11	11			0	0	0
		0	1	0			6	7	5			10	14	11			0	0	0
	CUA	19	0	20		CCA	45	41	42	**Gln**	CAA	70	65	64		CGA	45	44	44
		28	23	25			49	47	46			67	66	68			44	42	42
		23	20	25			49	46	46			68	67	68			44	44	42
	CUG	2	0	0		CCG	3	4	3		CAG	0	5	6		CGG	6	1	2
		1	1	0			2	3	4			4	5	3			2	4	4
		0	0	0			2	3	4			3	4	3			2	3	4
Ile	AUU	344	353	354	**Thr**	ACU	97	95	95	**Asn**	AAU	192	192	193	**Ser**	AGU	33	30	30
		341	340	340			102	99	103			187	196	189			28	31	30
		347	348	340			98	98	103			190	192	189			27	27	30
	AUC	15	15	15		ACC	3	5	5		AAC	13	10	10		AGC	1	0	0
		24	21	22			5	5	3			16	10	14			2	1	1
		18	20	22			6	6	3			15	12	14			1	2	1
Met	AUA	194	217	216		ACA	85	77	77	**Lys**	AAA	75	81	81		AGA	73	74	74
		200	203	205			81	78	78			79	77	77			75	75	76
		207	209	205			80	78	78			76	78	77			76	77	76
	AUG	18	14	14		ACG	2	2	2		AAG	8	6	5		AGG	0	0	0
		21	19	19			2	2	0			6	8	8			0	0	0
		17	13	19			1	0	0			9	7	8			0	0	0
Val	GUU	90	82	82	**Ala**	GCU	125	129	129	**Asp**	GAU	54	58	59	**Gly**	GGU	67	62	61
		78	81	83			136	138	135			58	58	56			59	70	65
		78	77	83			140	130	135			56	54	56			60	60	65
	GUC	3	1	1		GCC	9	8	8		GAC	10	8	7		GGC	2	1	2
		8	1	1			8	6	9			8	7	9			0	1	0
		6	3	1			4	13	9			9	12	9			1	1	0
	GUA	92	92	92		GCA	37	27	27	**Glu**	GAA	81	76	76		GGA	129	138	137
		91	96	99			23	28	26			76	73	74			140	129	131
		92	97	99			25	26	26			75	75	74			135	135	131
	GUG	8	8	8		GCG	2	4	4		GAG	1	2	2		GGG	22	18	19
		7	9	7			5	2	2			2	5	4			20	19	24
		10	11	7			3	2	2			3	3	4			24	24	24

cine (TTG) at this site. I confirmed this sequence in the same line of *D. yakuba*, so it is unlikely to be a sequencing artifact. Either the ND1 locus is 12 bases longer in *D. yakuba* or ND1, like COII, ND2, ND4, and ND5, is incompletely terminated and must acquire a complete termination codon by polyadenylation. In this study, it is assumed that ND1 is incompletely terminated. For reference, Table 1 includes a modified codon table for *D. yakuba*.

To investigate the substitution rates in each protein coding locus Ks and Ka (Nei and Gojobori 1986) were calculated and plotted as the average of all pairs. Ks is defined as the number of synonymous substitutions per synonymous site for any pair of sequences. Ka is defined as the number of nonsynonymous substitutions per nonsynonymous site for any pair of sequences. Figure 1 shows that Ks to Ka ratio is highest for COI, COII, COIII, ND3, cytochrome *b*, ND4L, and ND1. The Ks to

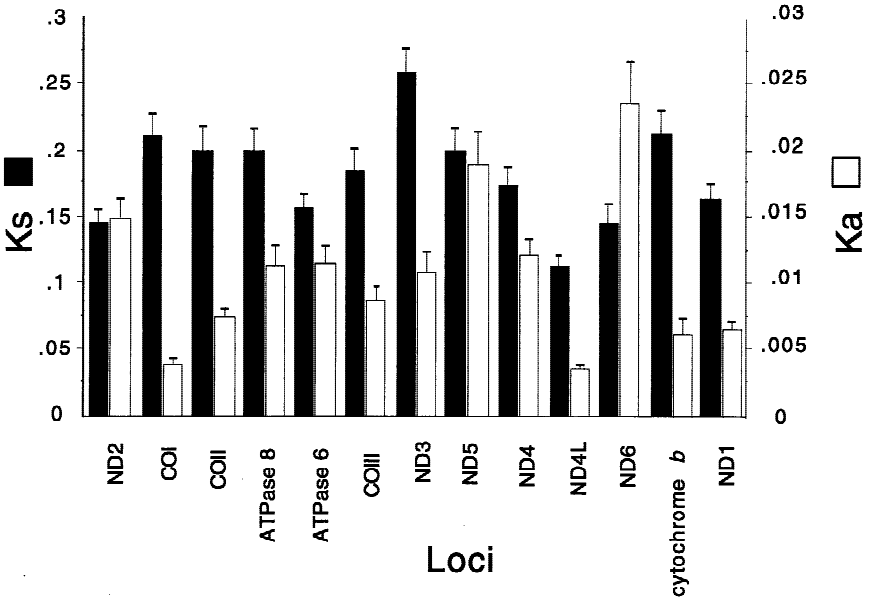

Fig. 1. Graph of average Ks and Ka between all pairs for each protein coding locus. Note that the Y axes are not on the same scale.

Ka ratio is lowest for ND2, ND5, and ND6 and intermediate for ATPase 8, ATPase 6, and ND4.

tRNAs: There are 1,459 bp in the 22 tRNAs. The phenylalanine tRNA is the most variable, with 11 variable positions (Fig. 2). Parsimony analysis inferred that there are seven variable sites in the lineage to *D. yakuba*, one of which is a compensatory mutation in the amino acid arm. In contrast, there is no variation in the glutamine tRNA. All other tRNAs have between one and nine variable positions.

Large Subunit rRNA (lrRNA) and Small Subunit rRNA (srRNA): There are 1,331 bp in the lrRNA and 795 bp in the srRNA. The single nucleotide difference between the mtDNA genomes of *si*III and *ma*I occurs in the lrRNA. There is a T insertion at position 13,252 in *ma*I, *si*I, and -II relative to the other taxa including *si*III. Solignac and Monnerot (1986) previously showed that there were no restriction site differences between isofemale lines of *si*III and *ma*I. Similarly, Satta and Takahata (1990) found no differences between lines of *si*III and *ma*I in a 2,531-bp region that included the COI locus. Ballard (2000) suggested that this might be the result of introgression. *D. mauritiana* (males) and *D. simulans* (females) can be crossed in the laboratory, resulting in sterile male and fertile female F1 hybrids (Robertson 1983). If the fertile F1 female is then backcrossed to males of *D. mauritiana* the mitochondrial genome will remain the *si*III type, but the nuclear DNA will be a mosaic of two genomes.

Phylogenetic Analyses

Figure 3 can be considered the "true" genealogy. Maximum parsimony analysis of the 14,958-bp data set with indel events scored generates a single well-supported topology. Maximum likelihood analysis generates the same topology. The homoplasy test (Maynard Smith and Smith 1998) does not suggest that there is any recombination among members of the *D. simulans* triad when the effective number of sites is in the range of 0.6–1 S (homoplasy ratio's = 0–0.25). S is the number of third position sites, excluding initiation and termination codons and codons at which there is amino acid variation.

These data do not support monophyly of the haplotypes of *D. simulans* or *D. mauritiana* relative to *D. sechellia* (Fig. 3). *D. simulans si*I and *D. sechellia* form a monophyletic assemblage, while *si*III and *ma*I are sister taxa. As described above, it is possible that there has been introgression of *si*III mtDNA into *D. mauritiana. D. simulans si*III has been collected on Reunion Island and Madagascar (Solignac and Monnerot 1986), and *D. mauritiana* has only been collected in Mauritius (Tsacas and David 1974). It is possible that *si*III occurs on Mauritius but has not been collected on the island. Moreover, the introgressed mtDNA (*ma*I) may have a selective advantage over the *ma*II mtDNA. Solignac and Monnerot (1986) reported that 88% of *D. mauritiana* isofemale lines they collected were the introgressed haplotype *ma*I. Niki et al. (1989) reported that, when *D. mauritiana ma*I mitochondria were microinjected into *D. melanogaster*, there was a preferential fixation of foreign mtDNA in the absence of selective conditions.

Genome-Wide Biases

In the mitochondrial genome, it is hypothesized that the intervening spacer regions are under low functional con-

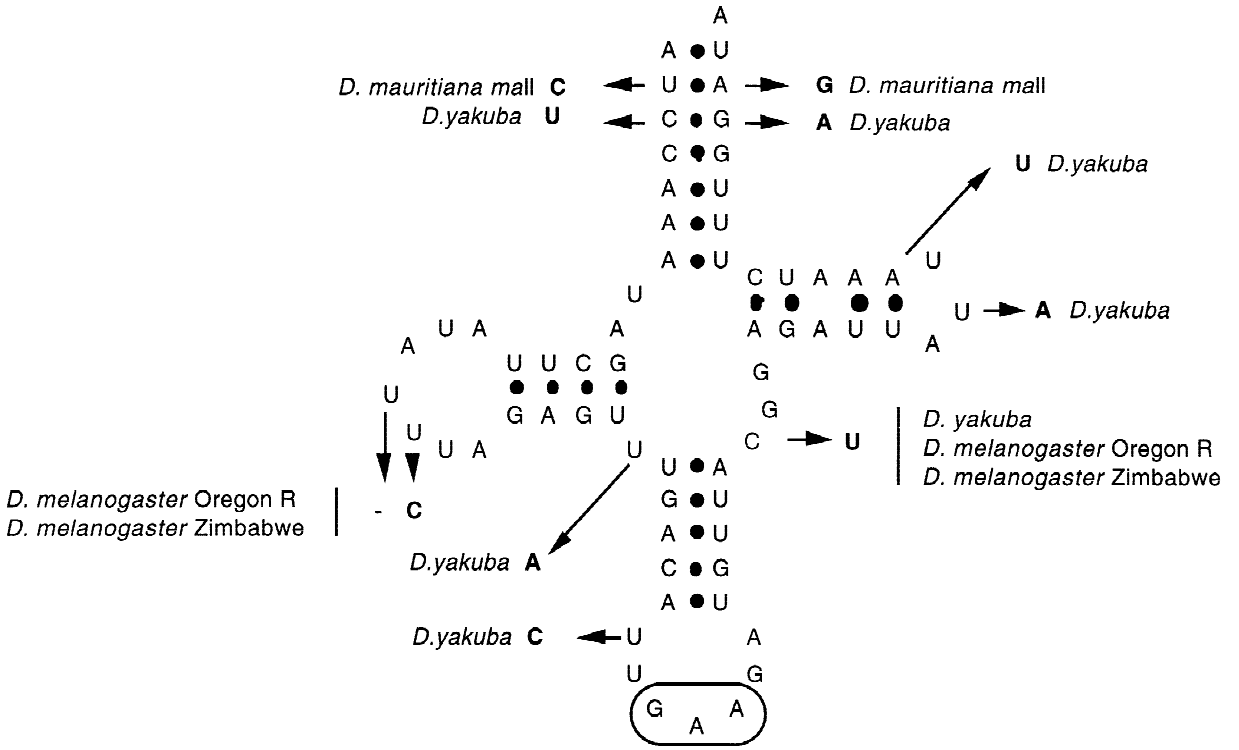

Fig. 2. The phenylalanine tRNA of *D. simulans si*I. Changes in the nine sequences included in this study are mapped onto the secondary structure.

straint and may reveal the underlying mutation biases. To investigate the bias between different partitions, the proportion of each nucleotide in the intervening spacer regions was compared with those in each codon position, in the tRNAs, and in the rRNAs (Table 2). Overall, the mean proportion of A/T nucleotides in the spacer regions (0.90) is significantly less extreme than third codon positions (0.94) as determined by a paired sign test ($p < 0.01$). One explanation for this result is that the intervening spacers are subject to selection. An alternate explanation is that there is selection for increased A/T-ending codons.

Xia (1996) proposed that maximizing transcriptional efficiency would cause an increase in A-ending codons. A prediction of this hypothesis is relative A-richness and T-deficiency in the coding strand. To test this prediction the mean number of A- and T-ending codons at the eight fourfold degenerate synonymous codons was compared. *D. mauritiana ma*I was excluded because it is likely the result of an introgression event. The data do not show a significant transcriptional bias in *Drosophila* mtDNA (Xia 1996). There is a mean of 623.5 T-ending and 594.75 A-ending fourfold degenerate codons. Of the eight fourfold degenerate codons, four are A-biased. This difference is not significant by a paired *t* test (mean difference is −11.14, $t_7 = -0.85$, $p = 0.40$). There is also no discernable trend if the anticodon is considered or if the analysis is repeated for each strand of DNA.

To investigate strand-specific mutational biases on the major and minor strands, the unambiguous third codon position synonymous substitutions where mapped onto the well-supported genealogy using MacClade. A contingency table analysis was then performed on the number of changes, to and from each nucleotide, on each strand (Fig. 4). Overall there is a significant difference in the direction of substitutions ($\chi^2_7 = 156.45$, $p < 0.001$ if C → A, and G → T substitutions are excluded because of the low number of changes). A post hoc *t* test was then performed on the contingency table. On the major strand there are relatively more changes from A or T → C, and from C → T than observed on the minor strand (t = 2.72, 7.07, and 5.78 respectively, $p < 0.05$). On the minor strand there are relatively more changes from A or T → G, and from G → A than the major strand (t = 8.62, 5.04, and 2.17 respectively, $p < 0.05$). These data support Garesse (1988) and Rand and Kann (1998).

Substitutions in Each Lineage

If mutations in mtDNA are nearly neutral, then the pattern of substitutions may be affected by changes in the environment or demography of the population. Contingency table analyses were employed to investigate the homogeneity of the substitution processes in each lineage. The first analysis compared the number of synonymous, nonsynonymous, and silent substitutions in each lineage (Table 3). In the second analysis, the direction of synonymous changes in each lineage was compared ($\chi^2_{18} = 92.27$, $p < 0.001$). The third analysis compared the numbers of W-codons, S-codons, WS-codons, and nonsynonymous and silent mutations in each lineage ($\chi^2_{36} = 130.68$, $p < 0.01$). Post hoc *t* values from the latter con-

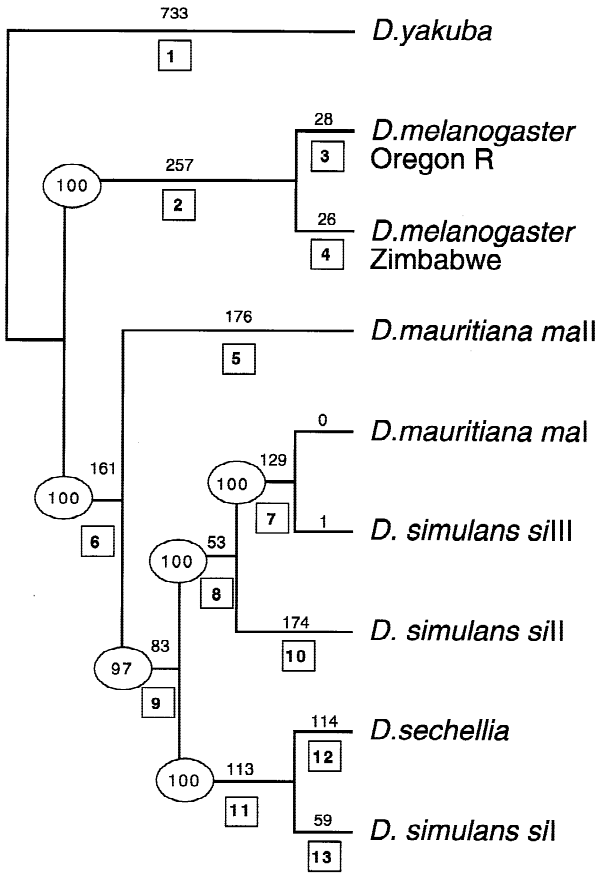

Fig. 3. Phylogenetic relationships of the *D. melanogaster* subgroup determined by 14,958 bp of mtDNA. Of these 13,309 are constant, 861 variable parsimony uninformative, and 788 parsimony informative. Parsimony analysis shows that there is a single most parsimonious tree of length 2,107 steps (Consistency Index = 0.85). The numbers in circles denote the bootstrap percentage from 1,000 pseudoreplicates. The numbers above lines refer to the branch length determined by parsimony. The boxed numbers distinguish each lineage. Maximum likelihood analysis (see text for details) generates the same topology (-Ln = 29468.66, Ti/Tv = 3.96, proportion of invariable sites estimated by likelihood = 0.84).

tingency table were then plotted onto the "true" genealogy (Fig. 5).

The pattern of substitutions has an underlying theme, but it is not upheld in a consistent manner. Generally speaking, there are relative excesses of nonsynonymous changes and S-codon mutants in populations that are likely to be small and a relative excess of W-codon mutants in populations that are likely to be large. Lineages 2 (to the common ancestor of *D. melanogaster*) and 10 (to *si*II) are not concordant with predictions. In both of these cases, S-codon mutants and nonsynonymous changes have significantly opposite trends (Fig. 5). It is not clear why this pattern is observed, however, the nearly neutral model of evolution does not account for all the variation in these mitochondrial data in all lineages. One possible explanation is that S-codon mutants or amino acid changes are not deleterious in all lineages. An alternate explanation is that there was a change in the mutational bias in these lineages.

Table 2. Proportion of each nucleotide in seven partitions in the mitochondrial genomes of nine species of *Drosophila.* In each case the standard error is less than 0.01

	Proportion of each nucleotide			
Partition	A	C	G	T
First codon	0.31	0.1	0.2	0.39
Second codon	0.45	0.03	0.02	0.49
Third codon	0.45	0.03	0.03	0.49
tRNAs	0.38	0.1	0.13	0.39
lrRNA	0.41	0.06	0.11	0.41
srRNA	0.41	0.08	0.13	0.39
Intervening	0.48	0.06	0.05	0.42

The *si*I haplotype has been collected from the Seychelles, Tahiti, New Caledonia, and Hawaii (Solignac and Monnerot 1986; Lachaise et al. 1988). As such, this population probably has been subject to a series of bottlenecks. Despite this expectation, there is no evidence for rate acceleration in the terminal branch to *si*I. There is, however, an excess of nonsynonymous changes in lineage 11 ($t = 1.98$, $p < 0.05$) relative to the alternatives.

D. mauritiana and *D. sechellia* are island endemics with small population sizes (Tsacas and David 1974; Lachaise et al. 1988; Hey 1994). In both lineage 5 to *ma*II and lineage 12 to *D. sechellia* (Fig. 5), there is a significant excess of S-codon mutants ($t = 2.73$ and 2.08, respectively, $p < 0.05$). In lineage 5 to *ma*II, there is a deficiency of W-codon mutants ($t = 3.88$, $p < 0.05$) relative to WS-codon mutants and nonsynonymous and silent changes (Fig. 5). In lineage 12 to *D. sechellia*, there is a deficiency of WS-codon mutants ($t = 2.95$, $p < 0.05$) relative to the alternatives (Fig. 5).

*D. simulans si*III has been collected on Madagascar and Reunion Island. James and Ballard (2000) collected 189 isofemale lines of *D. simulans* from Madagascar and observed that 58 were *si*III and 131 were *si*II. Consistent with the observation that *si*III has a small historical population size, there is an excess of S-codon mutants in lineage 7 ($t = 1.98$, $p < 0.05$) and deficiency of silent changes ($t = 2.08$, $p < 0.05$) relative to the alternatives (Fig. 5).

It is difficult to interpret the pattern of substitutions associated with the radiation of the simulants triad. However, it is likely that *D. simulans* expanded out of islands in the Indian Ocean (Lachaise et al. 1988). It now occurs worldwide. In lineage 6 to the simulans triad (*D. mauritiana, D. simulans,* and *D. sechellia*), there is a deficiency of S-codon mutants ($t = 4.62$, $p < 0.05$) and excess of WS-codon mutants ($t = 4.47$, $p < 0.05$) relative to the alternatives (Fig. 5).

Sliding-Window Analyses

A sliding-window approach was taken to detect regions that conflict with a single nucleotide substitution process

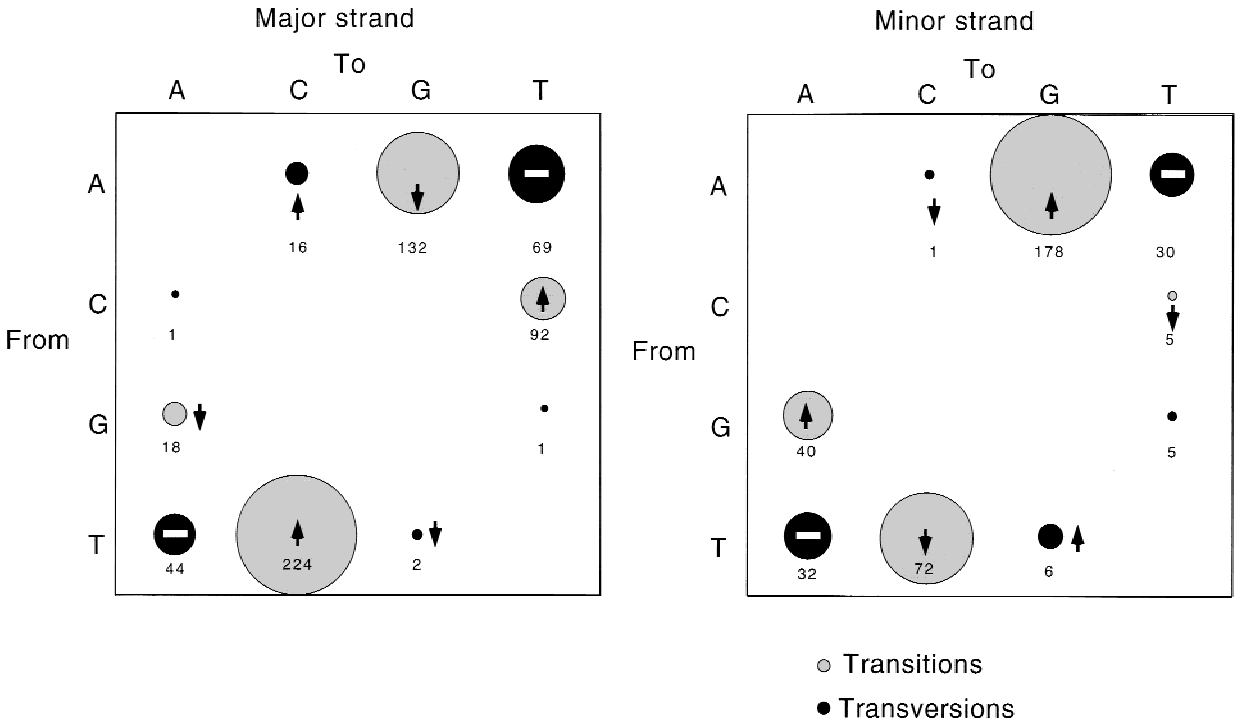

Fig. 4. Third codon position synonymous nucleotide substitutions in the major and minor strands mapped onto the well-resolved genealogy. Each number indicates the total number of changes between states. The circles represent the total proportion of each substitution type. A contingency table analysis was performed on the total number of changes to and from each nucleotide on each strand (see text for details). A post hoc *t* test was then conducted (C → A, and G → T substitutions are excluded because of the low number of changes). Significant post hoc *t* values (t > |1.96|) are represented on the figure with an arrow: ⬆ indicates a relative excess of this substitution type, ⬇ indicates a relative deficiency of this substitution type, and − indicates no significant difference between the number of changes on the major and minor coding strands.

derived from the entire sequence. The highest variability among sequences in the sliding-window plot is observed around positions 1,480 and 6,640 (Fig. 6A). The COI locus occurs between positions 1,476–3,015, and the ND5 locus occurs between positions 6,499–8,216. The lowest variability is observed around position 9,200 and positions 12,770–15,034. Position 9,200 is within the ND4 locus (8,301–9,639). The region from 12,770–15,034 includes the leucine^{CUN} and valine tRNAs, and both rRNAs.

As expected, the variability between all nine sequences is highly correlated with the variability within each of four clades. However, there are four distinct areas of incongruence (Figs. 6B and 6C). The divergence between *si*II and -III is relatively high between positions 1,000–2,500 ($r = 0.24$) and 10,000–12,500 ($r = 0.34$) (Fig. 6B). The divergence between *D. sechellia* and *si*I is relatively low between positions 6,000–7,000 ($r = 0.57$) and relatively high between 7,500–9,500 ($r = 0.39$) (Fig. 6C).

Maximum likelihood sliding-window analyses also suggest that four regions are inconsistent with the substitution process derived from the complete sequence. Region 1 (positions 1,536–2,439) occurs in the COI locus ($Z = 6.901$, $p < 0.05$). Region 2 (6,451–6,847) spans the phenylalanine tRNA (Fig. 2) and the 5′ region of the ND5 locus ($Z = 5.07$, $p < 0.05$). Region 3 (7,547–9,256) encompasses the histidine tRNA and part of the ND5 and ND4 loci ($Z = 9.15$, $p < 0.05$). Region 4 (10,270–12,495) spans cytochrome *b*, the serine^{UCN} tRNA, and the majority of ND6 and ND1 ($Z = 7.954$, $p < 0.05$).

In Region 1, there is an excess of synonymous substitutions in the interspecific lineage to *si*II. It occurs within the 2,531-bp region sequenced by Satta and Takahata (1990) and shown by Ballard (1998) to have a high rate of substitution in the *si*II lineage. Contingency table analysis shows that there is a significant excess of substitutions to *si*II between positions 1,536–2,439 ($t = 2.03$, $p < 0.05$). A sliding-window analysis of synonymous sites detects an anomalous region in the COI locus between positions 1,536–2,442 ($Z = 10.55$, $p < 0.05$). The COI locus of *si*II lines from Kenya, from the Seychelles, and a second line collected from western USA are identical with the line employed in this study (GenBank #AF200839, AF200844, AF200840, respectively).

Contingency table analysis shows that there is a long branch to *D. yakuba* in Region 2 ($t = 1.76$, $p < 0.05$). There are large numbers of changes in the phenylalanine tRNA ($Z = 5.4$, $p < 0.05$) (Fig. 2) and an anomalous region of nonsynonymous changes between positions

Table 3. Investigation of the number of substitutions inferred by parsimony in each branch

Lineage (Fig. 3)	Substitution				
	Synonymous				
		(W-, S-, and WS-codons)	Nonsynonymous	Silent	Total
1	498	*	116	119	733
2	162	(14, 111, 37)	50	44	256[a]
3[b]	20	(3, 16, 1)	5	3	28
4[b]	16	(1, 14, 1)	5	5	26
5	131	(5, 107, 19)	18	27	176
6	116	(27, 55, 34)	19	26	161
7	107	(21, 76, 10)	12	10	129
8	38	(4, 24, 10)	8	7	53
9	62	(15, 41, 6)	7	14	83
10	141	(30, 104, 7)	13	20	174
11	79	(11, 54, 14)	21	13	113
12	88	(16, 69, 3)	15	11	114
13	46	(9, 34, 3)	4	9	59
Total	1,504	(1,006)	293	308	2,105[c]
	Synonymous *cf.* Nonsynonymous		Nonsynonymous *cf.* Silent	Silent *cf.* Synonymous	Overall
	$G^2_{12} = 32.93$		$G^2_{12} = 12.07$	$G^2_{12} = 18.95$	$G^2_{24} = 47.35$
	$p = 0.001$		$p = 0.44$	$p = 0.09$	$p = 0.003$

* W-, S-, and WS-codons are not included for the *D. yakuba* lineage because they were used to polarize changes. As a consequence, the total number of W-, S-, and WS-codons is 498 less than the total number of synonymous mutations.

[a] This is one step shorter than shown on Fig. 3 because the amino acid deletion to *D. melanogaster* in ND5 is not included.

[b] Not included in the statistical analysis of synonymous substitutions because of the low number of substitutions.

[c] Note that this is two steps shorter than the most parsimonious tree. The silent changes on the lineage to *si*III and the amino acid deletion to *D. melanogaster* in ND5 are not included.

6,752–6,873 ($Z = 6.72$, $p < 0.05$). The anomaly might occur in an internal lineage or in the intraspecific branch to the isofemale line of *D. yakuba* included in this study.

Region 3 is likely to be the result of multiple changes in evolutionary rates among lineages. There appears to be a high relative divergence between *D. sechellia* and *si*I in this region (Fig. 6C). There are more changes than expected in lineage 7 to *si*III/*ma*I and fewer than expected in lineage 13 to *si*I. However, post hoc analysis does not detect a significant rate change in either lineage ($t = 1.65$ and -1.5, respectively, $p > 0.05$). Sliding-window analysis of nonsynonymous sites suggests there is an anomalous region between positions 7,976–7,989 ($Z = 8.21$, $p < 0.05$). Rand et al. (1994) previously noted that there was heterogeneity in the number of nonsynonymous changes in this region between *D. melanogaster* and *si*II.

In Region 4 there is a rate increase in lineage 11 to *D. sechellia* and *si*I ($t = 2.03$, $p < 0.05$). Region 4 spans the region shown by Ballard (1998) to have a high substitution rate in the lineage to siI. However, Ballard (1998) did not include *D. sechellia.* Inclusion of *D. sechellia* in this study permitted the lineage to *si*I to be divided into two branches (lineages 11 and 13 in Fig. 3). Region 4 is a large region that spans cytochrome *b*, the serine^{UCN} tRNA, and the majority of ND6 and ND1. Sliding-window analysis of the nonsynonymous sites suggests there is an anomalous region in the ND6 gene between positions 10,318–10,388 ($Z = 12.29$, $p < 0.05$). In lineage 11 there is a significant excess of substitutions in ND6 ($t = 2.68$, $p < 0.05$), an excess in the cytochrome *b* gene ($t = 1.73$, $p < 0.1$), and a deficiency in the ND1 gene ($t = -0.43$, $p > 0.05$).

The topology for each of the four regions does not differ significantly from that generated for the genome, as determined by Kishino and Hasegawa (1989) tests. This result supports the tenet that there is no recombination between these lineages. Employing parsimony, Region 1 has the same topology as that generated from the complete data set. Regions 2, 3, and 4 do not have significantly different topologies (Region 2 $t = 1.41$, Region 3 $t = 0.38$, Region 4 $t = 1.25$, $p > 0.05$). Employing likelihood, there is no significant difference in the topology of the "true" genealogy and that derived from each region (Region 1 $t = 0.77$, Region 2 $t = 0.61$, Region 3 $t = 0.0$, Region 4 $t = 0.8$, $p > 0.05$). The anomaly of these four regions is not caused by a change in Ti/Tv ratio (Region 1 Ti/Tv = 3.90; Region 2 Ti/Tv = 2.74; Region 3 Ti/Tv = 4.68; Region 4 Ti/Tv = 4.82; complete sequence Ti/Tv = 3.96).

Discussion

Sequence evolution, in principle, can be explained by

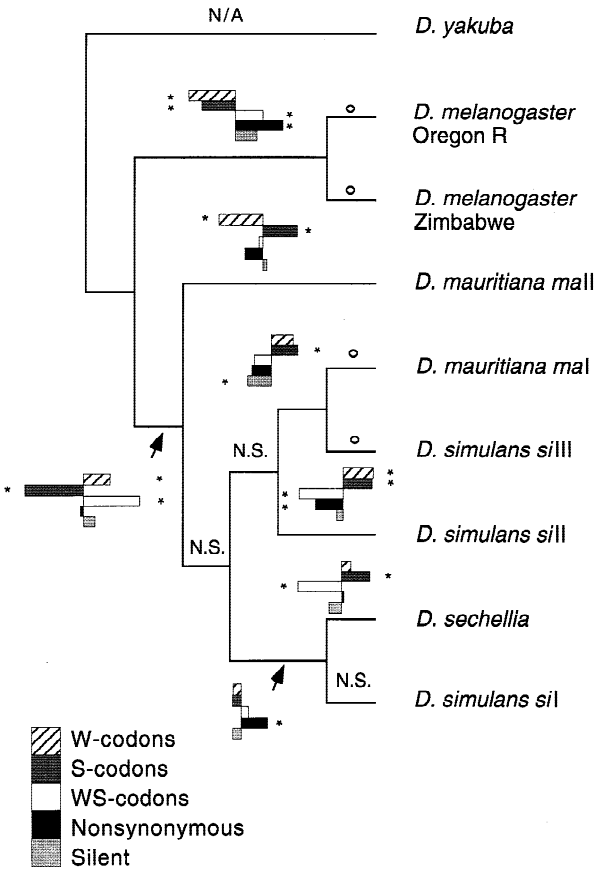

Fig. 5. Post hoc *t* values mapped onto the "true" mitochondrial genealogy. The post hoc results are placed directly above the relevant lineage except as shown with arrows. Bars to the left of each vertical line denote negative *t* values (relative deficiency). Bars to the right indicate positive *t* values (relative excess). Asterisks denote significance at $p < 0.05$. N.S. shows lineages where no category of substitutions is significantly different from expected. Lineages with an open circle had a low number of substitutions and were excluded from this analysis. *D. yakuba* was not included because this outgroup taxon was used to polarize changes.

combinations of processes intrinsic to the molecule, processes that affect the dynamics of lineages, and processes that differentially influence regions of DNA within genomes. In this study, I apportion different patterns to processes acting at these different levels. This is the first study that investigates the extent to which substitution patterns are the result of processes acting at higher levels compared with those acting at individual genes. This discussion begins by investigating the set of processes that are intrinsic to the whole molecule.

There is an A/T bias in *Drosophila* mtDNA. A/T-rich genomes may replicate more quickly than G/C-rich genomes and, if all else is equal, have a selective advantage in a heteroplasmic population. Kondo et al. (1990) observed heteroplasmy at a rate of about 6% in a natural population of *D. simulans.* There may also be a high directional nucleotide substitution rate influenced by rates of DNA damage (Martin 1995). It was hypothesized that the intervening spacer regions would be under low functional constraint and may reveal any underlying mutational biases. In this study, it was observed that the A/T bias in intervening spacer regions is less extreme than at the third position of fourfold degenerate codons. One explanation for this result is that some intervening spacer regions are subject to selection. Nakamichi et al. (1998) reported that a mitochondrial regulatory region between the phenylalanine and proline tRNAs is essential for regulation of replication and transcription of the mitochondrial genome in a variety of mammals. An alternate explanation is that there is a selection or mutation bias towards A/T-ending codons. In this study, the transcriptional hypothesis of codon usage (Xia 1996) is not supported.

Data presented here support previous observations in *Drosophila* that there is a significant bias in the patterns of nucleotide substitutions between the major and minor coding strands (Garesse 1988; Rand and Kann 1998). Genes whose reading frames lie on the major coding strand of *Drosophila* mtDNA show a preponderance of changes to or from C, while genes encoded on the minor strand experience more changes to or from G among species. This is a clear example of a process that is acting at a higher level but directly effects the patterns of substitution within individual genes. As pointed out by Rand and Kann (1998), a T $\rightarrow$ C substitution on the major strand is an A $\rightarrow$ G change if tabulated on the minor strand. Thus, although there may be one general mechanism to account for the bias in transitions on the major strand, the presence of reading frames on the complementary strand has important consequences for a putative mutation-selection balance at synonymous sites. These mutational biases may also have important consequences for reconstructing phylogenies with distinct genes, particularly under the likelihood criterion.

A second set of processes affects the evolutionary dynamics of mtDNA within lineages. These processes are expected to apply equally to all loci. The underlying theme is that slightly deleterious mutations and different effective population sizes may explain the different evolutionary patterns. However, this result is not consistent, and lineages 2 and 10 (Fig. 3) conflict with expectations. In each case the pattern of S-codon mutants conflicts with the prediction that these changes should accumulate in small populations. It is possible that S-codon mutants are not deleterious in all lineages or there was a change in the mutational bias in these lineages. One alternate explanation for these results is that selective sweeps of the mitochondrial genome have obscured any correspondence between the effective mtDNA population sizes and census population sizes (Ballard 2000).

Data from nuclear loci suggest the ancestral effective population size of *D. melanogaster* was smaller than that of *D. simulans* (Aquadro et al. 1988; Akashi 1996; Bergstrom and Prichard 1998). Consistent with expectation, there is a significant excess of nonsynonymous substitu-

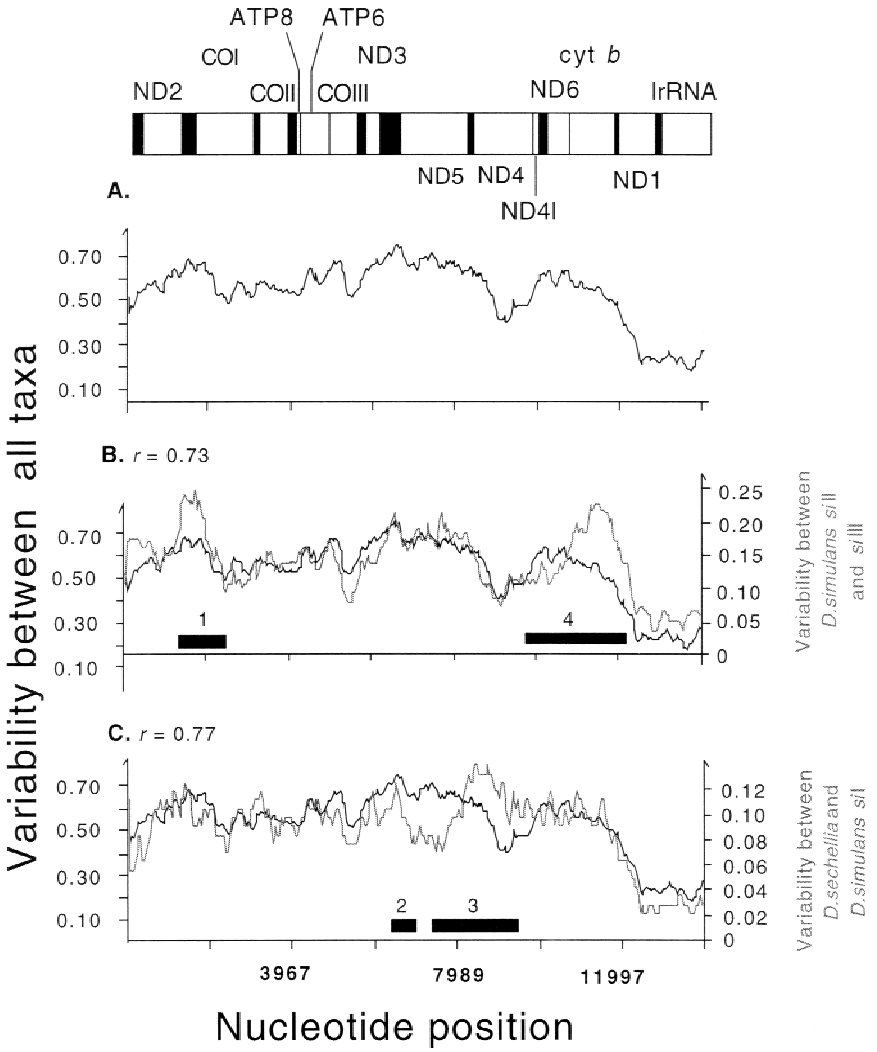

Fig. 6. Sliding-window analyses. The bar at the top illustrates the position of each protein coding locus and rRNA. The tRNAs are represented as black boxes. The window is 1,000 bp in length and slides 40 bp at a time. The formula for calculating the variability is presented in the text. Sliding-window analyses comparing the variability among **A** All nine sequences (LHS). **B** All nine sequences (LHS) and *D. simulans si*II and -III (RHS). **C** All nine sequences (LHS), and *D. sechellia* and *D. simulans si*I (RHS). Note that the Y axes are not equal in magnitude. To assess the variability between two sliding-window plots, the Pearson product moment (*r*) correlation is employed. The horizontal black bars show the four regions of statistical incongruence detected by the sliding window maximum likelihood method.

tions in the lineage to *D. melanogaster*. Ballard (2000) similarly reports a significant excess of nonsynonymous substitutions in this lineage. Kaneko et al. (1993) previously noted a high level of amino acid polymorphism at the ATPase 6 locus in *D. melanogaster*. A small ancestral population of *D. melanogaster* is, however, inconsistent with the observed deficiency of S-codon mutants in this lineage if S-codons are slightly deleterious.

The evolutionary dynamics of mtDNA in *si*II are difficult to interpret. This haplotype has a broad distribution and likely had a large historical population size. In this study, there is a relative deficiency of nonsynonymous substitutions and relative excess of S-codon mutants in the lineage leading to *si*II. However, it is clear that recent selective sweeps have influenced the evolutionary dynamics of mtDNA. At the cytochrome *b* locus, Ballard and Kreitman (1994) observed that the *si*II haplotype is segregating at only two synonymous sites. HKA segregating sites tests (Hudson et al. 1987) against either alcohol dehydrogenase (*Adh*) or *period* were significant, indicating a departure from neutrality. However, *Adh* and *period* were not significantly different from each other. This allowed Ballard and Kreitman (1994) to conclude that the synonymous polymorphism level is lower than expected for the cytochrome *b* locus, given its rate of evolution. Rand et al. (1994) observed fewer than half the number of variable sites expected under neutrality at ND5, when compared to sequences from the *period* locus. Similarly, Ballard (2000) noted a significant reduction in the diversity of seven mtDNA genomes when compared to intron 1 of *Adhr* as determined by the HKA test (Hudson et al. 1987). This suggests that there has been one or more recent selective sweeps of the mitochondrial genome in *si*II. This may result from the selective fixation of a mutation (Maynard Smith and Haigh 1974; Kaplan et al. 1989), nuclear-cytoplasmic interactions (Clark and Lyckegaard 1988; Hutter and Rand 1995), or through a maternally inherited factor, such as *Wolbachia* (Turelli et al. 1992; Ballard et al. 1996; Ballard 2000; James and Ballard 2000).

The third set of processes influence the evolutionary dynamics of mtDNA regions within the mitochondrial genome. This could be the result of mutational mechanisms or coincident selection affecting sites within the

region. Ballard (1998) proposed two possible explanations for the uncoupling of evolutionary rates within the mtDNA genome. One mechanism is lineage-specific selection affecting a subset of genes. The other, homologous recombination, is not supported in this study. Ballard (1998) suggested rate variation in distinct lineages might occur if each gene in each mtDNA haplotype (lineage) evolves by different selective rules. This may result from specific selective pressures exerted by changes in the environment or local demographic differences. For example, effective population size differences between lineages might have differential effects on rates of substitution in mitochondrial loci, the nuclear loci with which they interact, or their proteins. An alternate explanation is that there is significant variation in the intensity of purifying selection against different nonsynonymous substitutions in distinct regions of mtDNA (Xia 1998).

A primary motivation of this study was to explore the observed differences in rates of synonymous changes at two localized regions of mtDNA in *si*I and *si*II (Ballard 1998). It aimed to determine how frequently this heterogeneity in the substitution process is observed in distinct lineages of mtDNA from closely related species within the *D. melanogaster* subgroup. In this study, four regions of heterogeneity are evident. A single shift in one lineage might have caused the significantly different nucleotide substitution rates in each of Regions 1, 2, and 4. In contrast, Region 3 appears to have multiple causes. In Region 1, there is an increase in the synonymous substitution rate in the interspecific branch of the COI locus to *si*II. It defines the region that was shown by Ballard (1998) to have a high rate of substitution in the *si*II lineage. Region 2 is correlated with an increase in the number of changes in the lineage to *D. yakuba.* It spans the phenylalanine tRNA. Region 4 is correlated with cytochrome *b,* the locus shown by Ballard (1998) to exhibit an elevated rate of substitutions to *si*I. Inclusion of *D. sechellia* permitted the lineage to *si*I to be divided into two branches (lineages 11 and 13 in Fig. 3). This facilitated the delineation of the evolutionary rate increase onto the ancestral lineage 11.

The findings from this study indicate the evolutionary dynamics of *Drosophila* mtDNA can be explained by combinations of processes intrinsic to the molecule, processes that affect the dynamics of lineages and processes that differentially influence regions of DNA within genomes. A dominant force may be interpreted as selection for increased rate of replication, DNA repair, and a strand bias in the patterns of substitution among species. Processes that affect the evolutionary dynamics of mtDNA within lineages may reflect effective population size differences between lineages and weak selection acting on both nonsynonymous and synonymous substitutions. Lineage-specific selection remains a strong explanation for the spatial phylogenetic variation detected in regions of DNA. There is no evidence for homologous recombination. These data illustrate the complexity of the evolutionary process in mtDNA and showcase the power of comparative genomics in teasing apart the forces that shape evolutionary change.

Acknowledgments. I thank Kirrie Ballard and Andrew Berry for help collecting flies; ICIPE for assistance in obtaining collection permits; Jerry Coyne, Tim Karr, Rumi Kondo, Michael Solignac, and Michael Turelli for providing isofemale lines; and Nick Grassly and Todd Lowe for assistance with the analyses. Matt Dean, Dick Hudson, Avis James, Martin Kreitman, Thomas Nagylaki, and David Rand gave constructive comments. All sequencing was done in the Pritzker Laboratory of Molecular Systematics and Evolution, Field Museum, by JWOB. Funds were provided by National Science Foundation (NSF) Grant No. DEB-9702824 and the Field Museum Marshall Field Fund.

References

Akashi H (1995) Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. Genetics 139:1067–1076

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

Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori.* Nature 397:176–180

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

Awadalla P, A Eyre-Walker, J Maynard Smith (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286:2524–2525

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 (2000) Comparative genomics of mitochondrial DNA in *Drosophila simulans.* J Mol Evol 51:64–75

Ballard JWO (1998) Evolutionary rate variation within the mitochondrial DNA of *Drosophila simulans.* Current Topics of Molecular Evolution, Proceedings of the 1997 US/Germany/Japan TriNational Workshop on Molecular Evolution, Munich, Germany, pp 25–36

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

Ballard JWO, Kreitman M (1995) Is mitochondrial DNA a strictly neutral marker? TREE 10:485–488

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

Bergstrom CT, Prichard J (1998) Germline bottlenecks and the evolu-

tionary maintenance of mitochondrial genomes. Genetics 149: 2135–2146

Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1474

Bridger WA, Henderson JF (1983) Cell ATP. Wiley, New York.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effects of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303

Clark AG, Lyckegaard EMS (1988) Natural selection with nuclear and cytoplasmic transmission. III. Joint analysis of segregation and mtDNA in *Drosophila melanogaster*. Genetics 118:471–481

Clary DO, Wolstenholme DR (1983) Genes for cytochrome c oxidase I, URF2, and three tRNA's in *Drosophila* mitochondrial DNA. Nucleic Acids Res 11:6859–6872

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

de Bruijn MHL (1983) *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. Nature 304:234–241

Dykhuizen DS, Green L (1991) Recombination in *Escherichia coli* and the definition of biological species. J Bacteriol 173:7257–7268

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

Eyre-Walker A, Smith NH, Maynard Smith J (1999) How clonal are human mitochondria? Proc R Soc 266:477–483

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

Garesse R (1988) *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations. Genetics 118:649–663

Grassly NC, Holmes EC (1997) A likelihood method for the detection of selection and recombination using nucleotide sequences. Mol Biol Evol 14:239–247

Hagelberg, E. Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc B Biol Sci 266:485–492

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 21:160–174

Hey J (1994) Bridging phylogenetics and population genetics with gene tree models. In: Schierwater B, Streit B, Wagner G, DeSalle R (eds) Molecular approaches to ecology and evolution. Birkhauser, Basel, pp 435–449

Hudson RR, Kaplan N (1985) Statistical properties in the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

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

Hutter CM, Rand DM (1995) Competition between mitochondrial haplotypes in distinct nuclear genetic environments: *Drosophila pseudoobscura* vs. *D. persimilis.* Genetics 140:537–548

Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNA's and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J Mol Biol 151:389–409

Ikemura T (1982) Correlation between the abundance of yeast transfer RNA's and the occurrence of the respective codons in its protein genes: differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. J Mol Biol 2:13–34

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

James AC, Ballard JWO (2000) The expression of cytoplasmic incompatibility and its impact on population frequencies and the distribution of *Wolbachia* strains in *Drosophila simulans.* Evolution (in press)

Kaneko M, Satta Y, Matsuura ET, Chigusa SI (1993) Evolution of the mitochondrial ATPase 6 gene in *Drosophila:* evidence for an unusually high level of polymorphism in *D. melanogaster*. Genet Res 61:195–204

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

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood of the evolutionary tree topologies from DNA sequence data, and the branching order in *Hominoidea.* J Mol Evol 29:170–179

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

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

Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M (1988) Historical biogeography of the *Drosophila melanogaster* subgroup. Evol Biol 22:159–225

Lewin B (1997) Genes VI. Oxford University Press, Oxford

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

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

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Maddison WP, Maddison DR (1992) MacClade: analysis of phylogeny and character evolution, version 3.0. Sinauer Associates, Sunderland, MA

Martin AP (1995) Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. Mol Biol Evol 12:1124–1131

Maynard Smith J (1992) Analysing the mosaic structure of genes. J Mol Evol 34:126–129

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

Maynard Smith J, Smith NH (1998) Detecting recombination from gene trees. Mol Biol Evol 15:590–599

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

Milkman RIP, Crawford IP (1983) Clustered third-base substitutions among wild strains of *Escherichia coli.* Science 221:378–380

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

Moraes CT, Ciacci F, Bonilla E, Jansen C, Hirano M, Rao N, Lovelace RE, Rowland LP, Schon EA, DiMauro S (1993) Two novel pathogenetic mitochondrial mutations affecting organelle number and protein synthesis. Is the $tRNA^{LEU\ (UUR)}$ an etiologic hot spot? J Clin Invest 92:2906–2915

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

Nachman MW, Boyer SN, Aquadro CF (1994) Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proc Natl Acad Sci USA 91:6364–6368

Nakamichi N, Rhoads DD, Hayashi JI, Kagawa Y, Matsumura T (1998) Detection, localization, and sequence analysis of mitochondrial regulatory region RNAs in several mammalian species. J Biochem (Tokyo) 123:392–398

Nei M, Gojobori T (1986) Simple methods for estimating the number of synonymous and nonsynonymous substitutions. Mol Biol Evol 3:418–426

Niki Y, Chigusa SI, Matsuura T (1989) Complete replacement of mitochondrial DNA in *Drosophila.* Nature 341:551–552

Ohta T (1972) Population size and the rate of evolution. J Mol Evol 1:305–314

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. Nature 246:96–98

Proutski V, Holmes EC (1998) SWAN: sliding window analysis of nucleotide sequence variability. Bioinformatics 14:467–468

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

Robertson HM (1983) Mating behavior and the evolution of *Drosophila mauritiana.* Evolution 37:1283–1293

Rousset F, Vautin D, Solignac M (1992) Molecular identification of *Wolbachia,* the agents of cytoplasmic incompatibility in *Drosophila simulans,* and variability in relation with host mitochondrial types. Proc Roy Soc (Lond) B 247:163–168

Rozas J, Rozas R (1997) DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput Applic Biosci 13:307–311

Satta Y, Takahata N (1990) Evolution of *Drosophila* mitochondrial DNA and the history of the *melanogaster* subgroup. Proc Natl Acad Sci 87:9558–9562

Satta, Y, Ishiwa H, Chigusa SI (1987) Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. Mol Biol Evol 4:638–650

Saville, BJ, Kohli Y, Anderson J (1998) MtDNA recombination in a natural population. Proc Natl Acad Sci 95:1331–1335

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

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

Solignac M, Monnerot M, Mounolou J-C (1986) Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila.* J Mol Evol 23:31–40

Stephens J (1985) Statistical methods of DNA sequence analysis: detection of intragenetic recombination or gene conversion. Mol Biol Evol 2:539–556

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

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607

Tsacas L, Bächli B (1981) *Drosophila sechellia,* n. sp., huitiéme espéce du sous-groupe *melanogaster* des iles Séchelles (Diptera, Drosophilidae) Revue Jr Ent 3:146–150

Tsacas L, David J (1974) *Drosophila mauritiana* n sp du groupe *melanogaster* de l'île Maurice (Dipt Drosophilidae). Bull Soc Entomol Française 79:42–46

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

Xia X (1996) Maximizing transcription efficiency causes codon usage bias. Genetics 144:1309–1320

Xia X (1998) The rate heterogeneity of nonsynonymous substitutions in mammalian mitochondrial genes. Mol Biol Evol 15:336–344