Nucleotide Sequence Divergence in the A+T-rich Region of Mitochondrial DNA in Drosophila simulans and Drosophila mauritiana

Kayo Inohira, Toshiko Hara,1 and Etsuko T. Matsuura
Department of Biology, Ochanomizu University, Tokyo

We have determined the nucleotide sequences of two regions within the A+T-rich region of mitochondrial DNA (mtDNA) in the sIII type of Drosophila simulans and the maI type of D. mauritiana. The sequences of the two regions in sIII and maI are almost identical. The sequences include elements corresponding to the type I and type II repeat elements and the T-stretches as reported in D. melanogaster; an approximately 340-bp region (A region) adjacent to the tRNA^{\text{Gln}} gene includes a part of the type II repeat element, and an approximately 440-bp region (B region) includes a central portion of the A+T-rich region between the type I and type II repeat arrays. Each sequence of the two species was compared with those of D. melanogaster and D. yakuba. The sequences of the A region are relatively well conserved among the four species. The alignment of the two sequences of the B region with those of D. melanogaster and D. yakuba requires numerous insertions/deletions. For both regions, nucleotide differences between D. simulans or D. mauritiana and D. melanogaster are similar to those between the two and D. yakuba. The tendency is obvious in a subregion within the type II repeat element in the A region. These findings suggest that the rate of nucleotide substitution in the subregion is accelerated in the lineage leading to D. melanogaster. Loss of functional constraint in the stem-loop-forming sequence is proposed for this acceleration.

Introduction

Mitochondrial DNA (mtDNA) in most metazoan animals is a circular molecule which contains genes for only 13 proteins, 2 ribosomal RNAs, and 22 transfer RNAs. No intron has been observed within any of those genes, and intergenic sequences are extremely short or nonexistent (Wolstenholme 1992; Boore et al. 1995). The only large noncoding region within the mitochondrial genome is the control region. In mammals and amphibians, some signals are located within the control region for the replication initiation of one of the two strands (H-strand) and for transcription of both strands (H- and L-strands). The nuclear genome encodes all proteins that are indispensable for the replication and transcription of the mitochondrial genome. The control region provides major sites for interactions between mitochondrial and nuclear genomes. For mammalian mtDNA, such as in human and mouse, functions of the control region in mtDNA replication and transcription have been extensively studied (for a review, see Clayton 1992). Despite the important functions for mtDNA and mitochondria, the control region is known to have a higher level of sequence variation than the other regions of the genome. It has often been used for phylogenetic analysis due to the high level of variability (e.g., Aquadro and Greenberg 1983; Cann, Brown, and Wilson 1984; Foran, Hixson, and Brown 1988; Horai and Hayasaka 1990).

In Drosophila, a large single noncoding region is called the A+T-rich region, since its sequence is highly rich in adenine (A) and thymine (T). The replication origin has been mapped within the region as in the case of the control region in mammals (Goddard and Wolstenholme 1978, 1980). The size of the A+T-rich region varies in length from 1 to 5 kb between species and also varies within species (Fauron and Wolstenholme 1976, 1980a, 1980b). Most of the mtDNA length variation described so far is due to insertions/deletions and variations in the copy number of short tandemly repeated sequences within this region (Shah and Langley 1979; Fauron and Wolstenholme 1976, 1980a, 1980b; Solignac, Monnerot, and Mounolou 1983, 1986a, 1986b; Barrio et al. 1992). The nucleotide sequences of the A+T-rich region have been reported for species possessing the relatively short A+T-rich region: D. yakuba, D. virilis (Clary and Wolstenholme 1987), D. teissieri (Monnerot, Solignac, and Wolstenholme 1990), D. obscura, and D. ambigua (Monforte, Barrio, and Latorre 1993).

Recently, Lewis et al. (1994) have reported the nucleotide sequence of the 4.6-kb A+T-rich region of D. melanogaster and indicated two large arrays of tandemly repeated sequence elements. By comparing the sequence with other species, it was found that a highly conserved region among Drosophila species was also present in D. melanogaster.

Within the melanogaster species subgroup, two groups with the short and long A+T-rich regions are included. Drosophila yakuba belongs to the short group, and D. melanogaster belongs to the long group. Drosophila simulans and D. mauritiana are species closely related to D. melanogaster, and they possess the long A+T-rich region revealed from restriction analyses. Each of the two species is composed of cytoplasmic races, three for D. simulans and two for D. mauritiana (Solignac, Monnerot, and Mounolou 1986a), which provide interesting materials for the evolutionary study of mtDNA. Extensive restriction analysis was carried out by Solignac, Monnerot, and Mounolou (1986a) and length variation in the A+T-rich region within the melanogaster species subgroup was also examined (Solignac, Monnerot, and Mounolou 1986b). Sequence anal...
y's of mtDNA does not seem to be satisfactory in this subgroup, however (Satta and Takahata 1990; Kaneko et al. 1993; Ballard and Kreitman 1994; Rand, Dorfman, and Kann 1994). Especially for the A+T-rich region, sequence analysis has been virtually impossible until the nucleotide sequence of D. melanogaster is available.

To understand the evolution of the long A+T-rich region with repeated elements, we determined the nucleotide sequences of two regions within the A+T-rich region of D. simulans and D. mauritiana and compared them with corresponding regions of D. yakuba and D. melanogaster. This is the first analysis for the A+T-rich region at the nucleotide level among D. melanogaster and its closely related species. The present results revealed an unusual nucleotide divergence pattern in the four species.

Materials and Methods

Species and Strains

Isofemale strains of D. simulans (SI259) and D. mauritiana (g20), which were established from single females from wild populations, were used. The SI259 strain of D. simulans was derived from Réunion (St. Denis), and the g20 strain was derived from Mauritius (Port Louis). SI259 belongs to the siIII cytoplasmic race and g20 belongs to the maI cytoplasmic race. The two races share the nucleotide sequences of mtDNA and have extremely high homology (Satta and Takahata 1990).

Cloning and Sequencing of DNA Segments Within the A+T-rich Region

Mitochondrial DNA was extracted from adult flies as already described (Satta, Ishiwa, and Chigusa 1987), using an isofemale line newly established from an original strain for each species. By digesting intact mtDNA with HindIII, HaeIII, and XbaI, a 4.9-kb HindIII B fragment containing the entire A+T-rich region was religated through gel electrophoresis in low-melting-point agarose and a column described by Koenen (1989).

To clone a DNA segment containing a part of the A+T-rich region adjacent to the tRNAle genes, HpaI digests of the HindIII B fragment were directly ligated into a pTZ18R or pTV118N vector digested with HindIII and then Smal or HincII. The second segment, containing a central portion of the A+T-rich region, was amplified by polymerase chain reaction from the HindIII B fragment, and then the PCR product was used for cloning. PCR reactions were carried out using the oligonucleotides 5'-GTTACCATTTTTGGATTGTG-3' and 5'-CTGATAACTTTATCCCTA-3' as primers, based on the nucleotide sequences of D. yakuba mtDNA (Clary and Wolstenholme 1985). The reaction was carefully performed using a DNA thermal cycler (Perkin Elmer), in 100 µl containing 50 mM KCl, 10 mM Tris (pH8.3), 2.0 mM MgCl₂, 200 µM each of dNTP, 10 pM each of the primers, approximately 1 ng of the template HindIII B fragment DNA, and 2.5 units of AmpliTaq DNA polymerase (Perkin Elmer). Reaction conditions were 4 min at 94°C, 1 min at 37°C or 32°C, 1 min at 55°C, and 1 min at 72°C for one cycle, and 15 s at 94°C, 30 s at 42°C or 32°C, and 30 s at 72°C for 30 cycles. PCR products were treated with T4 DNA polymerase and ligated into a pTZ18R or pTV118N vector digested with SmaI.

Several deletion mutants were constructed for each DNA clone by Deletion Kit for Kilo-Sequence (TaKaRa). The DNA sequences were determined at least three times for each clone by the dyeodeoxy chain termination method (Sanger, Nicklen, and Coulson 1977), using primers commercially available and BcaBESTTM Sequencing Kit (TaKaRa). The two segments and the sequence strategies are shown in figure 1.

Sequence Analysis

Nucleotide sequences of D. simulans and D. mauritiana were aligned with the sequences of the corresponding regions of D. melanogaster (Lewis et al. 1994) and D. yakuba (Clary and Wolstenholme 1985) by GENETYX software (version 7.3) and by eye. The numbers of nucleotide substitutions were estimated by the method of Tamura (1992), taking into account an extremely high proportion of A and T within the A+T-rich region.

Results

Structure of the A+T-rich Region in D. simulans and D. mauritiana

Approximately 750 bp and 440 bp segments in the A+T-rich region were cloned and sequenced in D. simulans and D. mauritiana (fig. 1). The total numbers of nucleotides determined are 774 for D. simulans and 776 for D. mauritiana. These sequences are almost identical except for the nucleotide sites described below. The nucleotide sequences are aligned with the corresponding regions of D. melanogaster (Lewis et al. 1994) and D. yakuba (Clary and Wolstenholme 1985) (figs. 2 and 3).

In the region adjacent to the tRNAle gene (called A region), the total numbers of nucleotides determined are 338 bp in D. simulans and 341 bp in D. mauritiana. These two sequences show one nucleotide substitution and three insertions/deletions (fig. 2). Two types of conserved DNA elements were found in the two species. The first element is a T-stretch located close to the tRNAle gene, although the numbers of T's are different.

---

FIG. 1.—A part of the A+T-rich region and flanking genes of D. simulans and D. mauritiana. Black boxes denote tRNAle (I), tRNAQ (Q), and tRNAM (M) genes and a leftmost hatched box denotes a part of the NADH dehydrogenase subunit 2 (ND2) gene. Black bars indicate segments used for cloning. DNA sequencing strategy is indicated by arrows.
among the four species compared. The second is the type II repeat element reported in *D. melanogaster* by Lewis et al. (1994). A part of the highly conserved sequence element is observed within the type II element in both *D. simulans* and *D. mauritiana*.

In the region including the central portion of the A+T-rich region (called B region), *D. simulans* and *D. mauritiana* sequences differ from each other in only one insertion/deletion out of 436 nucleotide sites (fig. 3). Three types of DNA elements are identified in this region of the two species. The first element is a part of the highly conserved sequence element composed of 104 bp, which should be at the innermost position in the type II repeat array. The second element is a T-stretch present on the opposite strand to the other in the A region, also with different lengths among the four species. Only in *D. melanogaster*, a 242-bp sequence is present preceding the T-stretch, as marked in figure 3. Although numerous nucleotide differences are observed in the sequences following the T-stretch among the four species, a type I repeat element can be assigned in *D. simulans* and *D. mauritiana* as a third conserved element. By this alignment, the type I element is found even in *D. yakuba*, in which it was not detected previously (Lewis et al. 1994). An approximately 110-bp region between the T-stretch and the type I element is relatively well conserved. Figure 4 schematically represents the structure of the central portion of the A+T-rich regions for *D. simulans* and *D. mauritiana* with those for *D. melanogaster* and *D. yakuba*.

Comparisons of the Nucleotide Sequences

From the alignments (figs. 2 and 3), nucleotide differences among the four species in the A and B regions are analyzed (tables 1 and 2). In each region, a relatively well conserved region is analyzed separately from the rest of the sequence. Deletions/insertions are excluded from this analysis. From the two tables, totals of nucleotide differences observed between the two species, *D. simulans* and *D. mauritiana*, and *D. melanogaster* are almost the same as those between the two and *D. yakuba*. These results are strikingly different from the previous analyses using protein-coding genes in both mitochondrial and nuclear genomes (Satta and Takahata 1990; Caccione et al. 1996).

In the A region (table 1), the highly conserved sequence elements in *D. simulans* and *D. mauritiana* are identical and contain only 17 nucleotide substitutions
and 2 deletions out of 195 nucleotide sites when compared with *D. melanogaster*. Sixteen substitutions and five deletions were observed when these species were compared with *D. yakuba*, essentially the same as the comparison with *D. melanogaster*. This tendency is slightly different in the region other than the highly conserved sequence element; the numbers of differences between the two species and *D. melanogaster* are less than those between the two and *D. yakuba*. In contrast to the A region, many insertions/deletions are observed in the B region (fig. 3). Especially in the type I repeat element, 48 and 51 nucleotide substitutions, 0 and 13 insertions, and 42 and 42 deletions are observed when *D. simulans* is compared with *D. melanogaster* and *D. yakuba*, respectively.

The distributions of transitions and transversions are also indicated for the A region (table 1) and the B region (table 2). Transversions predominate transitions in all comparisons except for those of *D. melanogaster* with *D. simulans* and *D. mauritiana* in the highly con-

---

**Fig. 3.—Alignment of the nucleotide sequences of the B region in the A+T-rich region. Sequences are given as in figure 2. The highly conserved sequence element is bracketed. A box indicates a T-stretch. A closed triangle shows the position of a 242-bp sequence observed only in *D. melanogaster*. An arrow is located at the start site of the type I element. A boxed 19-bp sequence indicates one of the PCR primers.**
Fig. 4.—Possible structures of the central portion of the A+T-rich region in D. simulans and D. mauritiana as compared with D. melanogaster and D. yakuba. Corresponding regions such as the type I (filled box), type II (mixed filled box) including the highly conserved sequence element (hatched box), and T-stretch (open box) and a relatively well conserved region between the T-stretch and the type I element (brick filled box) are indicated by boxes marked similarly. Origins and direction of mtDNA replication are given by arrows.

Discussion

Structure of the A+T-rich Region

The present results indicate that the nucleotide sequences of the A+T-rich regions in sIII mtDNA of D. simulans and mel mtDNA of D. mauritiana are almost identical, as was observed in the coding region of mtDNA (Satta and Takahata 1990). Comparing these sequences with those of D. melanogaster and D. yakuba reveals that they possess the same repeat elements as were reported in D. melanogaster (Lewis et al. 1994).

Solignac, Monnerot, and Mounolou (1986b) have indicated the presence of tandemly repeated elements in species in the melanogaster subgroup by restriction analysis using HpaI and AccI. The type II element, one of the repeated elements, is well conserved among the four species examined (fig. 2). A part of a 281-bp highly conserved sequence element found in the Drosophila species with the short A+T-rich region (Clary and Wolstenholme 1985, 1987; Monnerot, Solignac, and Wolstenholme 1990) and D. melanogaster (Lewis et al. 1994) is also noted within the type II element of the two species examined. For the B region, insertions/deletions of short sequences could have occurred during the evolution of this portion, as suggested by the alignment (fig. 3). As both the highly conserved sequence element and the T-stretch are observed in all the species compared, the alignment is made so that the two sequence elements are firstly aligned. As a result of this alignment, the type I element was detected as the other element in D. simulans and D. mauritiana, and even in D. yakuba. The type I element has higher variability within the repeated elements than do the type II elements in D. melanogaster (Lewis et al. 1994). Partial sequences of this element of D. simulans, D. mauritiana, and D. yakuba indicate that its sequence also varies between species. These findings confirm the previous observations that the half of the A+T-rich region where the type I elements reside is variable in size in D. melanogaster and in the other

Table 1

<table>
<thead>
<tr>
<th>Species Compared</th>
<th>Transitions</th>
<th>Transversions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>sim vs. mau</td>
<td>1 (0.3)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>mel vs. sim</td>
<td>38 (11.3)</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>mel vs. mau</td>
<td>37 (10.9)</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>mel vs. yak</td>
<td>64 (19.0)</td>
<td>13</td>
<td>29</td>
</tr>
<tr>
<td>sim vs. yak</td>
<td>49 (14.3)</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>mau vs. yak</td>
<td>50 (14.8)</td>
<td>3</td>
<td>16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Highly Conserved Sequence Element</th>
<th>Transitions</th>
<th>Transversions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>sim vs. mau</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>mel vs. sim</td>
<td>10</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>mel vs. mau</td>
<td>10</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>mel vs. yak</td>
<td>13</td>
<td>16</td>
<td>29</td>
</tr>
<tr>
<td>sim vs. yak</td>
<td>3</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>mau vs. yak</td>
<td>3</td>
<td>13</td>
<td>16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Sequences</th>
<th>Transitions</th>
<th>Transversions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>sim vs. mau</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>mel vs. sim</td>
<td>8</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>mel vs. mau</td>
<td>8</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>mel vs. yak</td>
<td>10</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>sim vs. yak</td>
<td>9</td>
<td>24</td>
<td>33</td>
</tr>
<tr>
<td>mau vs. yak</td>
<td>10</td>
<td>24</td>
<td>34</td>
</tr>
</tbody>
</table>

NOTE.—sim = D. simulans, mau = D. mauritiana, mel = D. melanogaster, yak = D. yakuba. Numbers in parentheses denote percent differences in total nucleotide sites compared.
Table 2
Numbers of Nucleotide Differences in the B Region

<table>
<thead>
<tr>
<th>SPECIES COMPARED</th>
<th>TYPE I ELEMENT</th>
<th>OTHER SEQUENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transitions</td>
<td>Transversions</td>
</tr>
<tr>
<td>sim vs. mau</td>
<td>0 (0.0)</td>
<td>0</td>
</tr>
<tr>
<td>mel vs. sim</td>
<td>75 (17.6)</td>
<td>1</td>
</tr>
<tr>
<td>mel vs. mau</td>
<td>74 (17.5)</td>
<td>1</td>
</tr>
<tr>
<td>mel vs. yak</td>
<td>112 (26.9)</td>
<td>4</td>
</tr>
<tr>
<td>sim vs. yak</td>
<td>80 (20.6)</td>
<td>3</td>
</tr>
<tr>
<td>mau vs. yak</td>
<td>80 (20.6)</td>
<td>3</td>
</tr>
</tbody>
</table>

Note.—Abbreviations are as given in table 1. Numbers in parentheses denote percent differences in total nucleotide sites compared.

Table 3
Estimated Numbers of Nucleotide Substitutions per Site (d)

<table>
<thead>
<tr>
<th>SPECIES COMPARED</th>
<th>A REGION</th>
<th>B REGION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Highly Conserved Element</td>
<td>Others</td>
</tr>
<tr>
<td>mel vs. sim</td>
<td>0.147 (0.033)</td>
<td>0.017</td>
</tr>
<tr>
<td>mel vs. mau</td>
<td>0.141 (0.032)</td>
<td>0.112</td>
</tr>
<tr>
<td>mel vs. yak</td>
<td>0.261 (0.048)</td>
<td>0.198</td>
</tr>
<tr>
<td>sim vs. yak</td>
<td>0.170 (0.026)</td>
<td>0.090</td>
</tr>
<tr>
<td>mau vs. yak</td>
<td>0.173 (0.025)</td>
<td>0.090</td>
</tr>
</tbody>
</table>

Note.—Abbreviations are as given in table 1. Numbers in parentheses denote standard deviations. Data for the COI and ADH genes are from Satta and Takahata (1990) and Jeffs, Holmes, and Ashburner (1994), respectively. The third codon positions of the genes are used.
to D. melanogaster. When the sequences of the highly conserved regions of the four species are compared with that of D. virilis (Clary and Wolstenholme 1987), which belongs to a different subgenus from D. melanogaster, an extremely high rate of nucleotide substitutions is also observed between D. melanogaster and D. virilis (data not shown). Loss of functional constraint in the lineage could be a conceivable explanation for the acceleration. To examine this possibility, functional regions in the A region were analyzed by assessing stem-loop-forming sequences using the program MulFold (Zuker 1989; Jaeger, Turner, and Zuker 1989, 1990).

Figure 5 shows 75 variable sites found in the whole A region of the four species. It is worth noting that, within the subregion 2, in which a stem-loop structure has been reported in D. yakuba (Clary and Wolstenholme 1987), many nucleotide differences are observed between D. melanogaster and the other three species. The nucleotide sequences of D. simulans and D. mauritiana, which exactly correspond to the sequence forming the stem-loop structure in D. yakuba, were examined to determine whether they are capable of forming similar structures. The structure expected in the two species is indicated in figure 6 with that in D. yakuba with slight modifications from that by Clary and Wolstenholme (1987), and it is found to be very similar to those reported in D. yakuba, D. virilis (Clary and Wolstenholme 1987), D. obscura, D. ambigua, and D. teissieri (Monforte, Barrio, and Latorre 1993). In D. melanogaster, however, such a structure is not formed in the corresponding sequence. Instead, a similar secondary structure is expected in another position (nucleotides 167–222). Therefore, in the lineage to D. melanogaster, some nucleotide substitutions in the subregion 2 could change the stem-loop-forming ability to accumulate further substitutions thereafter. Another secondary structure should function equally in D. melanogaster. This, in turn, suggests that each of the highly conserved sequence elements has some function, which may partly explain the high degree of conservation among the elements reported in D. melanogaster (Lewis et al. 1994).

Similar analyses to detect the stem-loop-forming sequences were carried out for the B region, which is expected to include the origin of mtDNA replication of the long A+T-rich region. In contrast to the A region, nucleotide sites at which D. melanogaster is different from the other species are observed only in a part of the type I element (fig. 3). Within this subregion of the type I element, some stem-loop structures are possible in the four species examined. However, these structures are different from those reported in D. yakuba and other species (Clary and Wolstenholme 1987, Monforte, Barrio, and Latorre 1993) in sizes of stem and/or loop (data not shown). A stem-loop structure which is equivalent to that of D. yakuba is expected within a 242-bp region of the innermost type II element present only in D. melanogaster (data not shown). In D. simulans and D. mauritiana, it may be necessary to analyze the whole innermost element of the type II for the assessment of the functional requirement in mtDNA replication. Detection of these stem-loop structures by S1 nuclease

![Subregion 1 and Subregion 2](image-url)

**FIG. 5.—Variable sites in the A region among D. melanogaster, D. simulans, D. mauritiana, and D. yakuba. Asterisks indicate the nucleotide element are bracketed. The limits of the subregions are indicated according to Monforte, Barrio, and Latorre (1993).**

![Secondary Structures](image-url)

**FIG. 6.—Possible secondary structures formed in the highly conserved sequence element in D. simulans, D. mauritiana, and D. yakuba. The sequence of D. yakuba is from Clary and Wolstenholme (1987). The sequences shown are nucleotides 254–324 of the sequences of the three species given in figure 2.**
mapping is now in progress. At present, it is not clear that the unusual nucleotide substitutions in the B region are also explained by loss of functional constraint.

Although the function of these secondary structures and the exact position of the origin of mtDNA replication is not fully clarified yet in Drosophila, the present observation that the nucleotide differences are related to the stem-loop-forming capability should provide an important clue for a better understanding. Further sequence analyses including the rest of the repeat elements are necessary to clarify the evolution of the A+T-rich region of D. melanogaster and its sibling species.

The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases with the following accession numbers: AB003095–AB003098.

Acknowledgments

We wish to thank S. I. Chigusa for discussions and encouragement during the course of this study. We also thank N. Takahata and Y. Satta for useful comments on the manuscript and K. Tamura and T. Moriya for their help on the sequence analyses. We are grateful to K. Kurata for her contribution to the amplification of a part of the A+T-rich region. This work was supported in part by Grants-in-Aid for Scientific Research on Priority Areas 05263210 and 06256207 from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED


NAOYUKI TAKAHATA, reviewing editor

Accepted April 11, 1997