

---

*Drosophila* mitochondrial DNA: a novel gene order

---

Douglas O.Clary, Judy M.Goddard\*, Susan C.Martin, Christiane M.-R.Fauron† and David R.Wolstenholme

---

Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

---

Received 23 August 1982; Revised and Accepted 5 October 1982

---

### ABSTRACT

Part of the replication origin-containing A+T-rich region of the *Drosophila yakuba* mtDNA molecule and segments on either side of this region have been sequenced, and the genes within them identified. The data confirm that the small and large rRNA genes lie in tandem adjacent to that side of the A+T-rich region which is replicated first, and establish that a tRNA<sup>val</sup> gene lies between the two rRNA genes and that URF1 follows the large rRNA gene. The data further establish that the genes for tRNA<sup>ile</sup>, tRNA<sup>gln</sup>, tRNA<sup>f-met</sup> and URF2 lie in the order given, on the opposite side of the A+T-rich region to the rRNA genes and, except for tRNA<sup>gln</sup>, are contained in the opposite strand to the rRNA, tRNA<sup>val</sup> and URF1 genes. This is in contrast to mammalian mtDNAs where all of these genes are located on the side of the replication origin which is replicated last, within the order tRNA<sup>phe</sup> small (12S) rRNA, tRNA<sup>val</sup>, large (16S) rRNA, tRNA<sup>leu</sup>, URF1, tRNA<sup>ile</sup>, tRNA<sup>gln</sup>, tRNA<sup>f-met</sup> and URF2, and, except tRNA<sup>gln</sup>, are all contained in the same (H) strand. In *D. yakuba* URF1 and URF2, the triplet AGA appears to specify an amino acid, which is again different from the situation found in mammalian mtDNAs, where AGA is used only as a rare termination codon.

### INTRODUCTION

The mitochondrial DNA (mtDNA) molecules of all metazoa examined to date, which range from nematode worms to humans, are in the form of circular duplexes which show species specific variation in size within the range 14.5 to 19.5 kb (1). Recently, sequencing of the entire mitochondrial genome of human, bovine and mouse has been completed and much of the gene content of these DNAs has been determined (2-5). In each of these species, all of the mtDNA molecules of an individual are identical, and each mtDNA contains the genes for the two rRNAs and 22 tRNAs of the mitochondrion's distinct protein synthesizing system, and five known polypeptides (3 subunits of cytochrome c oxidase, ATPase subunit 6, and cytochrome b). In addition, there are eight presently unidentified open reading frames (URF) which are presumed also to be genes for polypeptides.

In mtDNA of human, bovine, and mouse, the gene order is essentially

identical. One strand (H) of the molecule contains all genes except eight tRNAs and URF6, and there are very few nucleotides separating individual genes (2,5-7). In many cases tRNA genes, which are found singly or in clusters of two to five, punctuate genes for rRNAs and polypeptides and appear to play a role as processing sites for long transcripts of entire strands of the mtDNA molecule (8,9). Some modifications of the standard genetic code are found in mammalian mtDNA (5,10,11). Also, it appears that the 22 tRNAs specified by mammalian mtDNA are sufficient to permit translation of all the polypeptides and presumptive polypeptides encoded by mtDNA (10,11).

The replication origin of mammalian mtDNA molecules lies within a sequence delimited by the genes for tRNA<sup>pro</sup> and tRNA<sup>phe</sup>. This region, for which transcripts are absent or rare, varies substantially in length between species (2,5,12,13). Although the remaining regions of mammalian mtDNA molecules show high levels of homology between species, only limited segments of homology are found between these mtDNAs in the replication origin-containing regions (3,14).

Much less is known concerning mtDNA from metazoa other than mammals. In mtDNA of the amphibian Xenopus laevis, the relative locations of the rRNA genes, tRNA genes, replication origin, and templates for a number of polyadenylated RNA transcripts are similar to those found in mammalian mtDNA (15,16). Among invertebrates, Drosophila mtDNA has been the most studied. Again it is known that the rRNA genes and the origin of replication occupy similar relative positions to those found among the vertebrate mtDNAs (17-20) and some transcripts have been mapped on the mtDNA molecules (21,22). In Drosophila mtDNA the replication origin lies within a region which is approximately 95% adenine and thymine (A+T) (18,19,23) and these A+T-rich regions vary in size from 1.0 kb to 5.1 kb in mtDNA molecules of different species (17,24,25). Further, it has been shown that extensive sequence divergence has occurred in these A+T-rich regions both within as well as between species (26,27) and they do not appear to be transcribed (21,22). In all metazoan mtDNAs, replication is unidirectional around the molecule. In Drosophila mtDNA, however, replication relative to the rRNA genes is in the opposite direction to that found in vertebrate mtDNAs (19,20).

In this report we present the nucleotide sequences of selected portions of the mtDNA molecule of Drosophila yakuba and our analysis of the genes found within them. Our data indicate a number of differences between Drosophila and mammalian mtDNAs, including a difference in gene order.

---

**MATERIAL AND METHODS**

Stocks of D. yakuba I.C. (2371.6, Ivory Coast), D. virilis (2375.8, Chile) and D. melanogaster, (Oregon R-Utah) used in the present experiments were the same as those used previously (19,20,25-27). D. yakuba and D. virilis were originally obtained from the Species Stock Collection of the Genetics Foundation, University of Texas at Austin.

MtDNA was obtained by cesium chloride-ethidium bromide centrifugation of lysates of hand-dissected ovaries of D. yakuba and D. virilis, and from lysates of mitochondrial fractions obtained from embryos of D. melanogaster as described previously (24).

DNAs were digested with restriction enzymes under the conditions recommended by the manufacturers (New England Biolabs, Bethesda Research Labs). EcoRI\* digestions were carried out under conditions given in ref. 28. Analytical and preparative electrophoreses of restricted DNAs were carried out as described in ref. 29.

EcoRI and HindIII fragments of D. yakuba mtDNA were cloned into pBR325 (30) and pBR322 (31) respectively, using as host E. coli K12, HB101. Inserts of D. yakuba mtDNA in plasmids were identified by gel electrophoresis of restriction enzyme digests of presumptive hybrid DNAs, and in the case of inserts of A+T-rich region-containing fragments, by electron microscope denaturation mapping (26). Cloned EcoRI or HindIII fragments of D. yakuba mtDNA, or subfragments produced by further restriction enzyme digestions, were recloned into M13mp2, M13mp8 or M13mp9 (32-34). The EcoRI-D fragments of D. virilis and D. melanogaster were cloned from EcoRI digests of mtDNAs directly into M13mp2.

Preparation and selection of viral-Drosophila recombinant DNA molecules, growth of the bacteriophage and purification of single- and double-stranded bacteriophage DNA were as described (34-36).

All DNA sequences were obtained by the extension, dideoxyribonucleotide termination procedure of Sanger et al. (36,37) using [ $\alpha$ -<sup>32</sup>P]dATP (800 Ci/mM; New England Nuclear) and either a 26 nucleotide pair primer (38) or a single stranded pentadecamer primer (35). Reagents used in sequencing reactions were obtained from Bethesda Research Laboratories, P-L Biochemicals, or New England Biolabs. The reaction products were separated on 0.4 mm x 33 cm, 8% acrylamide gels (39) at 34 watts (approximately 1.5KV and 35mA). Following electrophoresis, gels were exposed to NS-5T no screen X-ray film (Kodak) at -20°C for 6 to 96 hours.

The strategy used to determine the nucleotide sequences of various

segments of the D. yakuba mtDNA molecules is given in Fig. 1. Genes within segments of D. yakuba mtDNA outside the A+T-rich region were identified by comparing both nucleotide and predicted amino acid sequences with the corresponding sequences of previously identified genes of mouse mtDNA (5). Nucleotide sequences were analysed using the SEQ Program (40) and amino acid sequences were analysed using the TYPIN and SEARCH programs (41,42) in a Digital Equipment Corporation 20/60 computer.

### RESULTS

Nucleotide sequences within and to the left of the A+T-rich region. In order to gain further information on the structure of the A+T-rich region, and to elucidate which genes are contained in the region of the mtDNA molecule lying adjacent to the left end of the A+T-rich region, we determined the nucleotide sequence of a portion of the D. yakuba mtDNA molecule identified as segment I in Fig. 1.

The nucleotide sequence of the entire segment I is given in Fig. 2. Within this sequence are three regions which have the potential to fold into the characteristic cloverleaf secondary structure of tRNAs with anticodons indicating them to be the genes for tRNA<sup>ile</sup>, tRNA<sup>gln</sup>, and tRNA<sup>f-met</sup> (Fig. 3). The genes for tRNA<sup>ile</sup> and tRNA<sup>gln</sup> are separated by 31 nucleotides while 8 nucleotides separate the tRNA<sup>gln</sup> and tRNA<sup>f-met</sup> genes. The nucleotide sequences of these three tRNA genes are 60%, 51%, and 67% homologous to the corresponding tRNA genes of mouse mtDNA. The sequence to the left of the tRNA<sup>f-met</sup> gene contains a single open-reading frame which, beginning with the first triplet (ATT, isoleucine) would be transcribed to the left. This nucleotide sequence and the predicted amino acid sequence from it have 49% and 38% homology, respectively, to the amino terminal region of URF2 of mouse mtDNA. The region of segment I to the right of the EcoRI site which delimits the EcoRI-A and EcoRI-C fragments maps in the region defined as the A+T-rich region (see Fig. 1). The 112 nucleotide segment lying between this EcoRI site and the right end of the tRNA<sup>ile</sup> gene has a G+C content of 9.8%, compared to a G+C content of 10.3% for the region to the right of this EcoRI site, and a tRNA gene has not been detected within it. It seems reasonable, therefore, to define the left terminus of the A+T-rich region as the nucleotide to the right of the tRNA<sup>ile</sup> gene. The relative positions of the codon (sense) strands of each gene indicate that transcription of the tRNA<sup>ile</sup>, tRNA<sup>f-met</sup> and URF2 genes is in the direction away from the left boundary of the A+T-rich region, while transcription of the tRNA<sup>gln</sup> gene is towards the A+T-rich region.

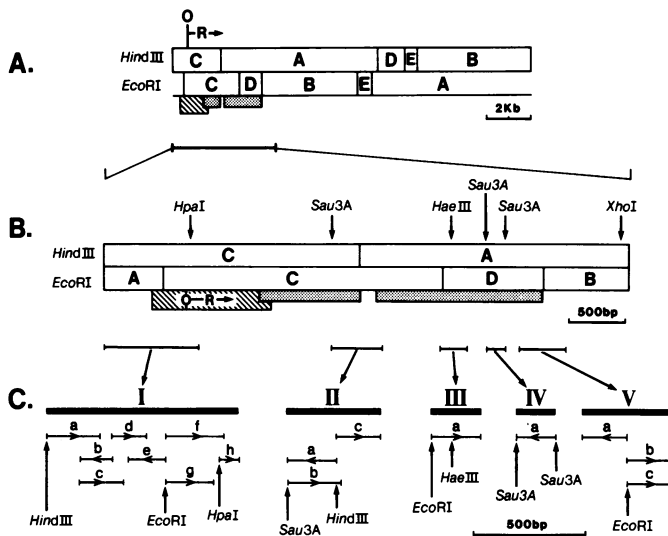


Figure 1. A map of the *D. yakuba* mtDNA molecule, and the strategy employed to obtain various sequences. Diagram A: A map of the whole *D. yakuba* mtDNA molecule showing the relative locations of the A+T-rich region (cross-hatched), the two mitochondrial rRNA genes (dotted; deduced from the homology of this region to a similarly located region of the *D. melanogaster* mtDNA molecule (27) shown (17) to contain the small and large mitochondrial rRNA genes), the origin (O) and direction (R) of replication (19) and the EcoRI and HindIII cleavage sites and fragments (A-E in each case) (27, Goddard, J. M., and Wolstenholme, D. R., unpublished). The circular molecule has been linearized for convenience at the HindIII site delimiting the HindIII-B and C fragments. Diagram B: Expanded region of the *D. yakuba* mtDNA molecule shown by a bar under Diagram A. Previously mapped Sau3A, HpaI and XhoI as well as EcoRI and HindIII restriction sites used to obtain and identify various sequences are shown. Other details are as for Diagram A. The bars under the map in Diagram B indicate the segments which have been sequenced. These segments (I-V) have been expanded in Diagram C and the sequence strategy used to assemble each is shown. The origins of the sequences were as follows. Segment I: a, the left end of the HindIII-C fragment; b, c, d, and e, EcoRI\* fragments of the HindIII-C fragment; f, the left end of the EcoRI-C fragment [identification of this end of the EcoRI-C fragment as the A+T-rich region was accomplished by determining the distance of the A+T-rich region (identified by electron microscopy of partially denatured molecules) from the M13 DNA molecules' unique Aval site, in double-stranded M13 DNA molecules separately containing the *D. yakuba* EcoRI-C fragment in the two possible orientations.]; g, an EcoRI-HindIII subfragment of the HindIII-C fragment; h, this sequence was obtained using an EcoRI-HpaI fragment of the EcoRI-C fragment as a primer for the EcoRI-C fragment. Segment II: a, the right end of the HindIII-C fragment; b, a Sau3A - EcoRI subfragment of the EcoRI-C fragment. c, a HindIII-EcoRI subfragment of the EcoRI-C fragment. Segment III: a, the left end of the EcoRI-D fragment. Segment IV: a, a Sau3A subfragment of the EcoRI-D fragment. Segment V: a, the right end of the EcoRI-D fragment; b, the left end of the EcoRI-B fragment; c, an EcoRI - XhoI subfragment of the EcoRI-B fragment. All of the above subfragments generated by two different restriction enzymes were sequenced from the site given first.

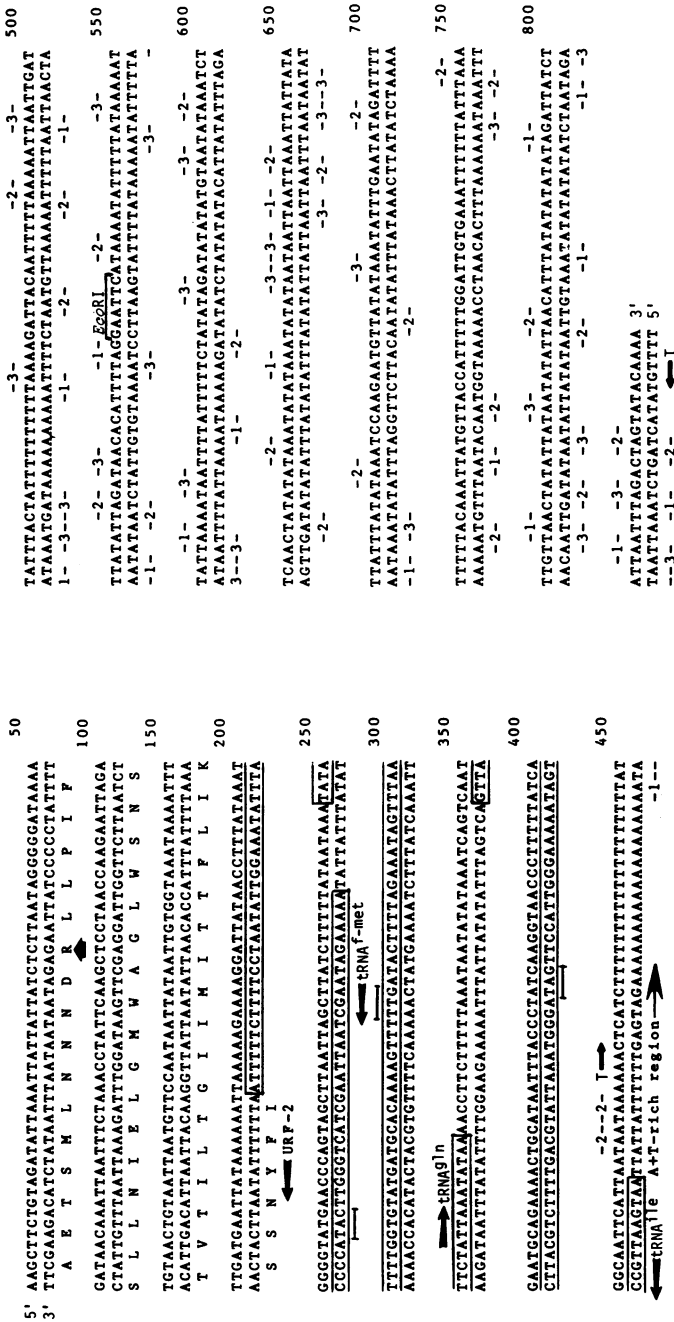


Figure 2. Nucleotide sequence of segment I (Fig. 1) of the *D. yakuba* mtDNA molecule. The sequence contains the left end of the A+I-rich region, and from considerations of nucleotide and predicted amino acid sequence homologies to mouse mtDNA (see text), the genes for tRNA<sup>Leu</sup>, tRNA<sup>Gln</sup>, and tRNA<sup>Met</sup> (boxed), and a portion of the amino terminus of URF2. The directions of transcription of URF2 and of each of the three tRNA genes (anticodons underlined) are indicated by arrows. The amino acid sequence predicted from the only open reading frame of URF2 is shown. The wide vertical arrow indicates the position of an AGA codon. The left boundary of the A+I-rich region is defined as the nucleotide to the right of the tRNA<sup>Leu</sup> gene (sense strand right to left, as indicated). Termination codons (TAA or TAG) are indicated by the number of the reading frame in which they occur in each of the two complementary strands, which are shown as the 5'-3' transcriptional (T) sense strands.

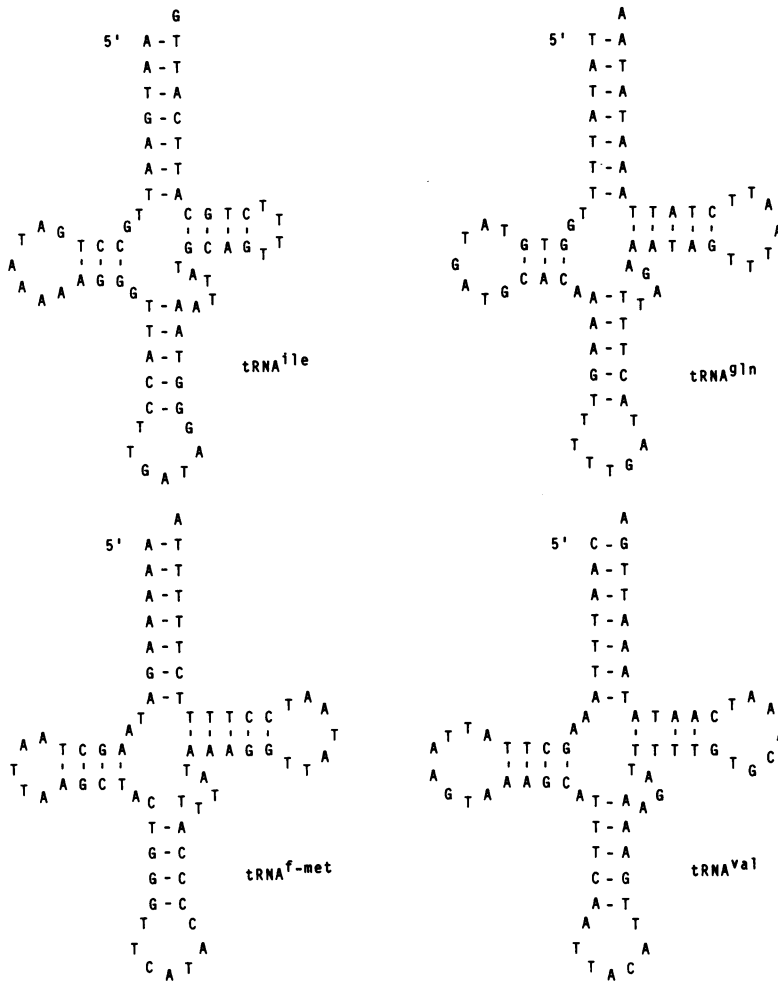


Figure 3. The genes for tRNA<sup>ile</sup>, tRNA<sup>gln</sup>, tRNA<sup>f-met</sup>, and tRNA<sup>val</sup>, of *D. yakuba* mtDNA shown in the presumed characteristic secondary structures of the corresponding tRNAs.

We have mapped the origin of replication within the A+T-rich region of *D. yakuba* mtDNA at approximately 250 to 450 nucleotides from its left boundary (19). The G+C content of this section of the A+T-rich region in our sequence (nucleotides 662-824, Fig. 2) is 14.1% compared to a G+C content of 7.6% for the remainder of the A+T-rich region to the left (nucleotides 412-661, Fig. 2). However, within the A+T-rich region, sequences homologous to sequences close to or associated with the origin of replication in mammalian

mtDNA (14,43,44) have not been found. At least seven termination codons (TAA and TAG, common to mammalian mtDNAs; 3,5,10) were found in each of the six possible reading frames of the two complementary strands of the A+T-rich region, and their spacing would prevent uninterrupted reading of a maximum of 117 nucleotides (Fig. 2).

In mammalian mtDNAs very few nucleotides occur between genes. With the exception of the 31 or 32 nucleotide sequence between the tRNA<sup>asn</sup> and tRNA<sup>cys</sup> genes, which contains the L strand origin of replication, the largest number of nucleotides found between any two tRNA genes is the seven nucleotides which separate the tRNA<sup>trp</sup> and tRNA<sup>ala</sup> genes of human mtDNA. Thus a further interesting feature of the tRNA<sup>ile</sup>, tRNA<sup>gln</sup> and tRNA<sup>f-met</sup> gene cluster of D. yakuba mtDNA is the presence of 31 nucleotides between the tRNA<sup>ile</sup> and tRNA<sup>gln</sup> genes (compared to a constant overlap of three nucleotides in mammalian mtDNAs), and of 8 nucleotides between the tRNA<sup>gln</sup> and tRNA<sup>f-met</sup> genes (compared to one to three nucleotides in mammalian mtDNAs (2,3,5)). It is noted that the octanucleotide sequence 5'TTTATTAT which separates the genes for tRNA<sup>gln</sup> and tRNA<sup>f-met</sup> also occurs close to the 5' end of the tRNA<sup>ile</sup> gene (Fig. 2). Further, the sequence 3'TTTATTAT occurs in the 31 nucleotide sequence separating the tRNA<sup>gln</sup> and tRNA<sup>ile</sup> genes.

The data presented above indicate that the genes for tRNA<sup>ile</sup>, tRNA<sup>gln</sup>, tRNA<sup>f-met</sup>, and URF2 of D. yakuba mtDNA lie on the opposite side of the replication origin to the previously deduced locations of the rRNA genes (17,26). This contrasts with the situation found in mammalian mtDNAs where all of these genes lie on the same side of the replication origin (that replicated last), within the order tRNA<sup>phe</sup>, small (12S) rRNA, tRNA<sup>val</sup>, large (16S) rRNA, tRNA<sup>leu</sup><sub>UUR</sub>, URF-1, tRNA<sup>ile</sup>, tRNA<sup>gln</sup>, tRNA<sup>f-met</sup>, and URF2 (2,3,5).

Nucleotide sequences to the right of the A+T-rich region. In order to confirm the position of the rRNA genes within the D. yakuba mtDNA molecule and to gain information on gene order on that side of the replication origin in which the rRNA genes are presumed to be located, (the right side, Fig. 1) we sequenced the various segments of the mtDNA molecule identified as II, III, IV and V in Fig. 1.

Segment II maps in a region of the D. yakuba mtDNA molecule expected to contain the junction of the two rRNA genes (Fig. 1). This segment (Fig. 4) was found to include a sequence with primary structure and potential secondary structure (Fig. 3) expected for a tRNA<sup>val</sup> gene. The sequence of 222 nucleotides to the left of the tRNA<sup>val</sup> gene is approximately 64% homologous to a region of mouse mtDNA which specifies the 3' end of the small rRNA molecule,



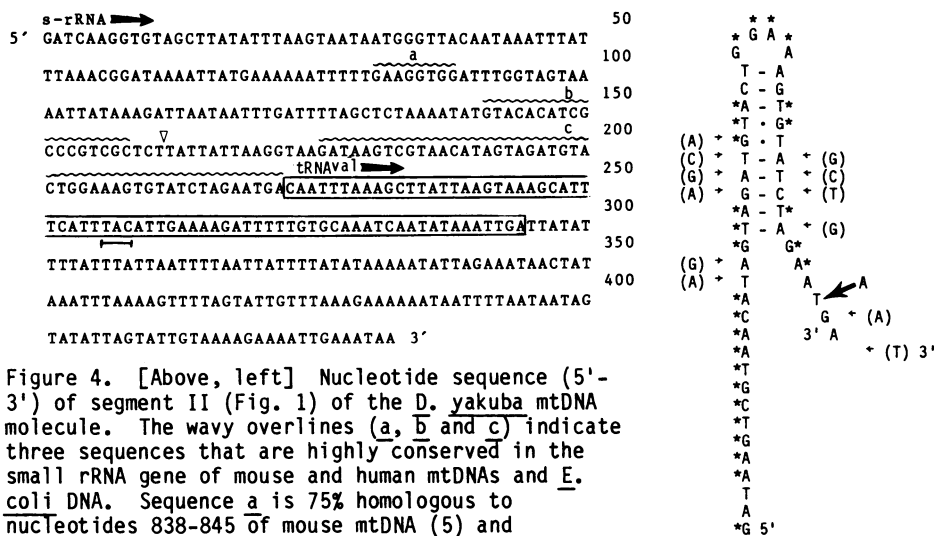


Figure 4. [Above, left] Nucleotide sequence (5'-3') of segment II (Fig. 1) of the *D. yakuba* mtDNA molecule. The wavy overlines (a, b and c) indicate three sequences that are highly conserved in the small rRNA gene of mouse and human mtDNAs and *E. coli* DNA. Sequence a is 75% homologous to nucleotides 838-845 of mouse mtDNA (5) and identical to nucleotides 1415-1422 of human mtDNA (45) and to 1117-1124 of *E. coli* DNA (46). Sequence b is 83% homologous to nucleotides 899-916 of mouse mtDNA and 89% homologous to nucleotides 1476-1493 of human mtDNA and nucleotides 1391-1408 of *E. coli* DNA. Sequence c is 75% homologous to nucleotides 977-1024 of mouse mtDNA and to nucleotides 1554-1601 of human mtDNA. Detailed comparisons of this latter sequence to the corresponding sequences of mouse mtDNA and *E. coli* DNA are given in Fig. 5. The inverted open triangle indicates the approximate position in the corresponding mouse mtDNA sequence of a unique segment of 47 nucleotides. The region to the right of the tRNA<sup>Val</sup> gene is presently unidentified (see text). The sequence shown is of the sense strand for both the small (S) rRNA gene and the tRNA<sup>Val</sup> gene, as indicated by the arrows.

Figure 5. [Above, right] A potential hairpin secondary structure consisting of a 10 base pair stem and a 4 base loop which is located immediately 5' to the tRNA<sup>Val</sup> gene of *D. yakuba* mtDNA (Fig.4). A similar hairpin structure is also found at the 3' end of the small rRNAs from mammalian and plant mitochondria and the small rRNAs of *E. coli*, *Zea mays* chloroplasts, *Bombyx mori* and rat cytoplasm (see 7, 47 for references). Differences to mouse mtDNA sequences are shown in parentheses. Asterisks indicate homologous nucleotides in *E. coli* 16S rRNA. The large arrow indicates the only difference in sequence found between the 3' terminal 20 nucleotides and the 3' terminal 20 nucleotides of the small mitochondrial rRNA of mosquito (48). In *E. coli* 16S rRNA and *Zea mays* chloroplasts 16S rRNA only, an mRNA binding site is found 3' to the hairpin structure (see 7).

except that a single sequence of 47 nucleotides located in this region of mouse mtDNA is absent from the *D. yakuba* sequence (Fig. 4). This region of the *D. yakuba* mtDNA molecule includes three sequences of 8, 18, and 47 nucleotides which are highly homologous to sequences found in the same relative order in the small rRNA gene of mouse and human mtDNA, and *E. coli* DNA (Fig. 4). The sequence of 47 nucleotides immediately to the left of the

tRNA<sup>val</sup> gene has the potential to fold into a hairpin structure characteristic of the 3' end of all small rRNAs examined to date (Fig. 5). The terminal 20 nucleotides of this sequence are identical to the recently reported (48) terminal 20 nucleotides of the major component of small rRNA molecules of mosquito (Aedes albopictus) except for nucleotide 3 from the 3' end. This indicates that the 3' terminal nucleotide of the small rRNA gene lies immediately adjacent to the 5' terminal nucleotide of the tRNA<sup>val</sup> gene in the D. yakuba mtDNA molecule. The sequence of 138 nucleotides to the right of the tRNA<sup>val</sup> gene has not been identified. The corresponding position in mouse mtDNA specifies the 5' end of the large rRNA molecule. However, this D. yakuba sequence, which is only 8.3% G+C, does not contain the sequences ACTAA and GCCCT which are present within the first 34 nucleotides of the 5' terminus of the large rRNA molecule of mammalian mtDNA and E. coli DNA (49) and is not convincingly homologous to any region of the gene for the large mitochondrial rRNA of mouse. Also, within this Drosophila sequence, we have not detected a sequence with the structural properties of any tRNA gene.

Segment III maps to the right of segment II, close to the center of the region of the D. yakuba mtDNA molecule expected to contain the large rRNA gene (Fig. 1). This sequence of 219 nucleotides (not shown) is 68% homologous to nucleotides 814 to 1032 of the large (16S) rRNA gene of mouse mtDNA (5).

Segment IV (172 nucleotides) also maps within the region expected to contain the large rRNA gene but to the right of Segment III, as shown in Fig. 1. This sequence is 68% homologous to a region of mouse mtDNA which specifies a region of the large rRNA near to its 3' end (5). The right 137 nucleotides have the potential to form a secondary structure which is highly conserved among the large mitochondrial rRNAs of mammals, yeast, Paramecium and the 23S rRNA of E. coli (Fig. 6). The sequence has been shown to encompass the binding sites for the 3' terminus of amino-acyl tRNA (53) and the nucleotides implicated in chloramphenicol resistance in yeast and mammalian mtDNA (50,52,53).

Segment V maps to the right of segment IV in the D. yakuba molecule near the region expected to specify the 3' terminus of the large rRNA. The nucleotide sequence of this segment is given in Fig. 7. The right 354 nucleotides of the sequence are 56.5% homologous to the amino terminal region of the nucleotide sequence of URF1 of mouse mtDNA, and the amino acid sequence predicted from the only open reading frame of this sequence is 48% homologous to the amino acid sequence predicted from the corresponding region of the URF1 sequence of mouse mtDNA (Fig. 8). We have also determined the nucleotide

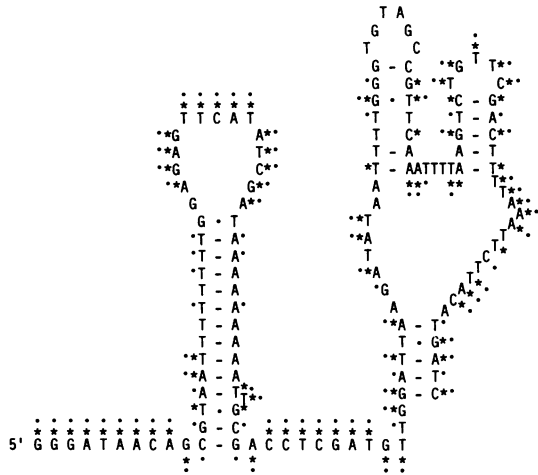


Figure 6. Nucleotide sequence and potential secondary structure of the right end of segment IV (Fig. 1) of the *D. yakuba* mtDNA molecule. A sequence with highly homologous primary and secondary structure to this *D. yakuba* sequence has been found in the gene for the large mitochondrial rRNA of mouse and other mammals (5, 45, 50), *Paramecium* (51) yeast (52) and *E. coli* (53,54). Dots indicate homologous nucleotides (66%) in the corresponding mouse mtDNA sequence (nucleotides 2373 to 2508; ref. 5). Asterisks indicate homologous nucleotides (57%) in the *E. coli* 23S rRNA sequence (nucleotides 2445 to 2579; ref. 53).

sequence of the region of the *D. melanogaster* and *D. virilis* mtDNA molecule which includes the amino terminal region of URF1 (in each species, the right end of the *EcoRI* fragment corresponding to the *EcoRI*-D fragment of *D. yakuba* mtDNA (Fig 1)). These sequences are shown aligned with the *D. yakuba* sequence in Fig. 7. In Fig. 8, the predicted amino acid sequences for the amino terminal region of URF1 of *D. melanogaster* and *D. virilis* mtDNAs, and of human and of bovine mtDNAs are shown aligned with the corresponding predicted amino acid sequences of *D. yakuba* and mouse mtDNA. URF1 of human and bovine mtDNA appears to initiate translation at a triplet (ATA and ATG, respectively) specifying methionine which in each case is three triplets 5' to the triplet ATT (isoleucine) interpreted as the initiator codon of URF1 of mouse mtDNA (2,3,5). It has been suggested that as an initiator codon, ATT may specify methionine (5). While in both *D. yakuba* and *D. melanogaster* mtDNAs an ATT codon occupies a position corresponding to the initiator codon of URF1 of mouse mtDNA, in *D. virilis* mtDNA, the corresponding triplet is GTA (valine). As there is presently no evidence to suggest that GTA can specify initiation, it seems more likely that the initiator codon of URF1 in *Drosophila* mtDNAs is

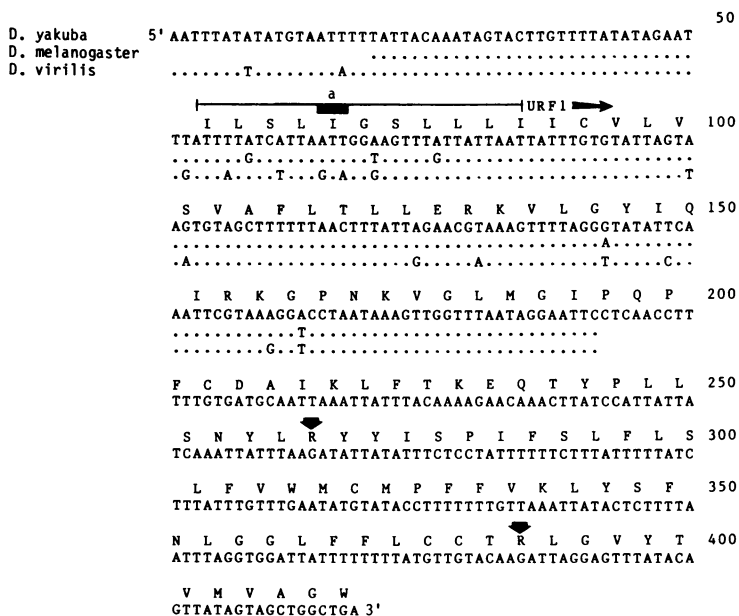


Figure 7. Nucleotide sequence (5'-3') of segment V (Fig. 1) of the *D. yakuba* mtDNA molecule. This segment includes a section of 354 nucleotides which from considerations of nucleotide and predicted amino acid sequence homologies to mouse mtDNA (see text) contains the amino terminal end of URF1. The amino acid sequence predicted from the only open reading frame, and the direction of transcription of the portion of the *D. yakuba* sequence (shown as the sense strand) corresponding to the mouse URF1 gene are given. The wide vertical arrows indicate the positions of two AGA codons. Sequences of the mtDNA molecules of *D. melanogaster* and *D. virilis* which correspond to the left end of segment V of the *D. yakuba* molecule are also shown (dots indicate homologous nucleotides). Region a includes three potential initiation codons (ATT, isoleucine). The thick bar indicates the ATT triplet similarly located to the initiator ATT of mouse mtDNA (5).

one of the two triplets, each ATT (isoleucine), which occur four to six triplets 5' and 3', respectively, away from the ATT triplet which corresponds to the initiator codon of mouse mtDNA. The unusual number of nucleotide substitutions which occurs among the *Drosophila* sequences in the region between these two ATT triplets suggests that this region may be devoid of genetic function, and it seems more likely that the triplet furthest 3' acts as the initiator codon for URF1.

In mammalian mtDNAs the gene for tRNA<sup>leu</sup><sub>UUR</sub> terminates either three nucleotides (human and bovine; 2,3) or ten nucleotides (mouse; 5) 5' to the URF1 initiator codon. In the nucleotide sequence 5' to the region expected to

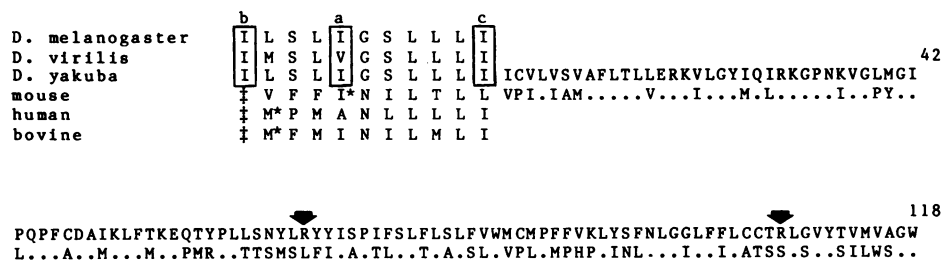


Figure 8. A comparison of the amino acid sequences of the amino terminal region of URF1 in D. yakuba and mouse mtDNAs (see Fig. 6 and ref. 5). Dots denote conserved amino acids. Amino acid sequences predicted from nucleotide sequences and expected to include the amino termini of URF1 are also shown for D. melanogaster, D. virilis (Fig. 7), human, and bovine mtDNA (2,3). The asterisks indicate the amino terminal amino acid given previously for mouse (5), human (2), and bovine (3). The crosses indicate the last nucleotide of the tRNA<sup>Leu</sup><sub>UUR</sub> gene of mouse mtDNA, and the second nucleotide following the last nucleotide of the tRNA<sup>Leu</sup><sub>UUR</sub> gene of human and bovine mtDNAs. Box a indicates the amino acid in each of the Drosophila mtDNAs corresponding to the assumed amino terminal isoleucine in mouse mtDNA. Box b and box c indicate the closest isoleucine or methionine common to all Drosophila mtDNAs, before, and following, respectively, the amino termini of URF1 of the three mammalian mtDNAs. The wide vertical arrows indicate the positions of two amino acids, tentatively identified as arginine, which are specified by AGA.

contain the initiator codon of URF1 of D. yakuba mtDNA (Fig. 7.), secondary structures which could indicate the presence of a tRNA<sup>Leu</sup><sub>UUR</sub> gene have not been found.

In the three Drosophila sequences which lie 3' to the region expected to contain the initiator codon of URF1 there are two nucleotide substitutions between the D. yakuba and D. melanogaster sequences, and eight nucleotide substitutions between the D. yakuba and D. virilis sequence. All but one of these substitutions are in the third position of a codon. The one exception is a second position change which would result in a serine/asparagine difference between the corresponding URF1 polypeptides from D. yakuba and D. virilis.

The data obtained from sequence analysis of segments II-V confirm the relative locations of the rRNA genes of D. yakuba mtDNA deduced previously (Fig. 1) and establish that a tRNA<sup>Val</sup> gene lies between the two rRNA genes, that URF1 follows the large rRNA gene and that all four of these genes are transcribed in the same direction, away from the A+T-rich region.

The structure of tRNA genes. The four D. yakuba tRNA genes we have sequenced (Fig. 3) show many of the standard features of tRNAs found in both

procaryotes and eucaryotes (2,3,5). All of these D. yakuba genes fold into the characteristic cloverleaf secondary structure which has a 7 nucleotide pair amino-acyl stem, a 5 nucleotide pair anticodon stem, an anticodon loop of 7 nucleotides, a dihydrouridine arm with a stem 3 to 4 nucleotide pairs and a loop of 5 to 8 nucleotides, a variable loop of 4 to 5 nucleotides, and a T $\Psi$ C arm with a stem of 4 to 5 nucleotide pairs and a loop of 4 to 7 nucleotides. As has been found for mammalian mitochondrial tRNAs, the Drosophila tRNAs predicted from the gene sequences lack several of the constant features found in non-organelle tRNAs. The 3' terminal trinucleotide CCA is absent, indicating that if it is present in mature tRNAs it must (as in mammalian mitochondria) be added post-transcriptionally. The dihydrouridine loop lacks the G<sub>18</sub>-G<sub>19</sub> sequence (numbering system given for yeast cytoplasmic tRNA<sup>Phe</sup>; 55), though G<sub>18</sub> appears to be conserved in tRNA<sup>Gln</sup> (which is one of the three tRNAs of mouse mtDNA in which the G<sub>18</sub>-G<sub>19</sub> sequence is found; (5)) and G<sub>19</sub> appears to be conserved in tRNA<sup>Val</sup>. The sequence T<sub>54</sub>- $\Psi$ -Pu-A is only partially conserved in all four tRNAs. T<sub>55</sub> (from which a  $\Psi$  could be predicted in the corresponding tRNA) is absent in the tRNA<sup>Val</sup> gene, and this gene alone contains C<sub>56</sub>. Pu<sub>57</sub> is found only in the tRNA<sup>Gln</sup> and tRNA<sup>Val</sup> genes. The Pu<sub>11</sub>-Py<sub>24</sub> pair is reversed in the tRNA<sup>f-met</sup> gene. In the tRNA<sup>Ile</sup> gene, the T $\Psi$ C loop is reduced from the standard 7 to 4. Further, at least one of these four tRNA genes lacks 1 or 2 of the constant T<sub>8</sub>, Pu<sub>15</sub>, Py<sub>32</sub>, or Py<sub>48</sub> nucleotides.

All four D. yakuba tRNAs have a T and a purine immediately 5' and 3' respectively to the anticodon, features noted to be constant for mammalian tRNAs (3). However, while the mammalian tRNA<sup>f-met</sup> genes are unique in having a C adjacent to the anticodon on the 5' side, in D. yakuba this position is also occupied by a T. Detailed comparisons of corresponding mitochondrial tRNA genes of Drosophila and mammals will be presented elsewhere.

### DISCUSSION

The results of the sequence analyses described above confirm that the small and large mitochondrial rRNA genes of the D. yakuba mtDNA molecule lie in tandem adjacent to that side of the replication origin-containing A+T-rich region which is replicated first, and establish that a gene for tRNA<sup>Val</sup> lies between the two rRNA genes, and that URF1 follows the large rRNA gene. Our data further establish that the genes for tRNA<sup>Ile</sup>, tRNA<sup>Gln</sup>, and tRNA<sup>f-met</sup> and URF2 occur, in the order given, on the opposite side of the A+T-rich region to the rRNA genes, that side which is replicated last (Fig. 9). This arrangement differs from that found in mouse, human, and bovine mtDNAs where all of these

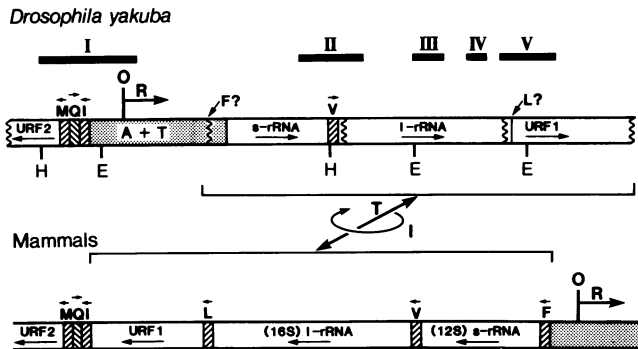


Figure 9. Difference in gene order between *D. yakuba* and mammalian mtDNA molecules as indicated by the data presented. I-V indicate the segments of *D. yakuba* mtDNA which have been sequenced (Fig. 1). The identification of each tRNA gene (hatched areas) is shown above the molecule. Arrows within and outside the molecules indicate the direction of transcription of each gene. Wavy vertical lines indicate uncertain gene termini. The origin (O) and direction (R) of replication of each molecule is shown. H and E indicate the HindIII and EcoRI sites, respectively. The main difference in gene arrangement between the *D. yakuba* and mammalian molecules could have resulted from a single translocation (T) and inversion (I) of a single segment containing the two rRNA genes, the intervening tRNA<sup>val</sup> gene and URF1 in an ancestral molecule. The question marks indicate that in the *D. yakuba* mtDNA molecule neither the tRNA<sup>leu</sup><sub>UUR</sub> or the tRNA<sup>phe</sup> gene have been located (see text).

genes are located on that side of the replication origin which is replicated last within the order tRNA<sup>phe</sup>, small (12S) rRNA, tRNA<sup>val</sup>, large (16S) rRNA, tRNA<sup>leu</sup><sub>UUR</sub>, URF-1, tRNA<sup>ile</sup>, tRNA<sup>gln</sup>, tRNA<sup>f-met</sup>, and URF2 (2,3,5). Furthermore, in contrast to the situation found in *D. yakuba* mtDNA, where the URF2, tRNA<sup>ile</sup> and tRNA<sup>f-met</sup> genes appear to be contained in the opposite strand to that containing the rRNA genes, tRNA<sup>val</sup> and URF1 genes, in mammalian mtDNAs all of these genes are contained in the same strand (the H strand) of the mtDNA molecule (Fig. 9). In both *D. yakuba* and mammalian mtDNAs, the tRNA<sup>gln</sup> gene is contained in the opposite strand to that containing the tRNA<sup>ile</sup> and tRNA<sup>f-met</sup> genes. As there is no distinct base bias between the complementary strands of *Drosophila* mtDNAs (23,25) they cannot be directly compared with the H and L strands of mammalian mtDNAs.

While we have not yet established the location of the genes for tRNA<sup>phe</sup> and tRNA<sup>leu</sup><sub>UUR</sub>, it is clear from our results that in *D. yakuba* mtDNA neither of these genes are located between the replication origin and the gene for tRNA<sup>ile</sup>.

Our previous heteroduplex analyses indicate that gene order outside the

A+T-rich region is identical among the mtDNAs of different Drosophila species (26). It seems most likely from our data that the difference in gene order between Drosophila and mammalian mtDNA molecules could have resulted in part from a major translocation and inversion of a segment containing the rRNA genes, the intervening tRNA<sup>Val</sup> gene and URF1 (Fig. 9). It is not clear from the presently available data whether the Drosophila mtDNA molecule or the mammalian mtDNA molecule contains the gene order most resembling the predicted common ancestral molecule.

The gene order in Drosophila mtDNA indicated by our data accounts for the apparent difference in direction of replication around the molecule in Drosophila and vertebrate mtDNAs described previously (20).

The replication origin-containing region of mammalian mtDNA (lying between the tRNA<sup>Pro</sup> and tRNA<sup>Phe</sup> genes (2,3,5)) is of higher G+C content but shares a number of features in common with the A+T-rich region of Drosophila mtDNA. This region of mammalian mtDNA is the most variable in length (from 879 nucleotides in mouse to 1122 nucleotides in human mtDNA), lacks extensive open reading frames, appears not to be transcribed (except for a segment of 230 nucleotides between the tRNA<sup>Phe</sup> gene and the replication origin of human mtDNA (56)) and contains segments of nucleotides which show little conservation between species (14). It seems reasonable to argue therefore that this region of mammalian mtDNA and the A+T-rich region of Drosophila mtDNA have evolved from a common ancestral sequence.

Reading of the sequences of D. yakuba URF1 and URF2 which have been determined is dependent in both cases upon the assumption that both TGA and AGA specify amino acids. In the genetic code of mammalian and fungal mitochondria (5,10,57,58), but possibly not plant mitochondria (59), TGA specifies tryptophan rather than termination as in the standard genetic code. However, while AGA specifies arginine (as in the standard genetic code) in yeast mitochondria (57), this triplet is used only as a rare termination codon in human and bovine mtDNAs (2). In mouse, AGA is not used at all (5). In mammalian mtDNAs, CGN codons specify arginine. CGT is found in the D. yakuba URF1 sequence at two different positions which both correspond to codons (CGC and CGA) interpreted to specify arginine in the mouse URF1 (5). The open reading frames of URF1 and URF2 of the D. yakuba sequences contain two and one AGA triplets respectively, at positions corresponding in mouse mtDNA to triplets specifying leucine and serine in URF1 (Fig. 8) and isoleucine in URF2 (Fig. 2). In view of this latter observation, it seems reasonable at the present time to assume that, as in the standard genetic code, AGA specifies



arginine in the Drosophila mitochondrial genetic code. The use of AGA to specify an amino acid requires the presence within the Drosophila mitochondrion of a tRNA (presumably with the anticodon UCU, as in yeast (57)) to decode it.

In comparisons of URF1 and URF2 of D. yakuba and mouse, nucleotide sequence homologies (57% and 49% respectively) are greater than amino acid sequence homologies (48% and 39% respectively). In comparisons of corresponding genes from mouse and human (5) nucleotide sequence homologies were also greater than amino acid sequence homologies in all URFs except URF1, and these were the genes with homologies of both nucleotide and amino acid sequences below 71%. In comparisons of corresponding genes of human and bovine mtDNAs (3), nucleotide homologies were again greater than amino acid homologies for genes (URFs) with both nucleotide and amino acid sequence homologies of less than 71%.

Of the nucleotide substitutions which have occurred between the portions of URF1 and URF2 of D. yakuba we have sequenced and the corresponding mouse mtDNAs, 45% in each gene fail to result in an amino acid substitution. As the ancestral lines of these organisms separated approximately  $600 \times 10^6$  years ago (60) these high frequencies of conservative nucleotide substitutions between URF genes constitute further evidence to that presented previously (2,3,5) that the URFs of mtDNA do indeed encode functional polypeptides.

#### ACKNOWLEDGEMENTS

We thank Dr. J. Messing for strains of bacteriophage M13 and their E. coli host, and Cedric Griss for assistance in setting up the computer programs. This work was supported by National Institutes of Health Grant No. GM 18375.

\*Present address: Department of Medicine, University of California, San Francisco, CA 94143, USA.

†Present address: Plant Breeding Institute, Trumpington, Cambridge, CB2 2LQ England.

#### REFERENCES

1. Altman, P.L. and Katz, D.D. (1976) Biological Handbooks. I. Cell Biology. pp. 217-219. Federation of American Societies for Experimental Biology. Bethesda, Md.
2. Anderson, S., Bankier, A.T., Barrell, B.G., deBruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457-465.
3. Anderson, S., DeBruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F., and Young, I.G. (1982) J. Mol. Biol. 156, 683-717.
4. Montoya, J., Ojala, D. and Attardi, G. (1981) Nature 290, 465-470.

5. Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W., and Clayton, D.A. (1981) *Cell* 26, 167-180.
6. Crews, S. and Attardi, G. (1980) *Cell* 19, 775-784.
7. Van Etten, R.A., Walberg, M.W. and Clayton, D.A. (1980) *Cell* 22, 157-170.
8. Ojala, D., Merkel, C., Gelfand, R. and Attardi, G. (1980) *Cell* 22, 393-403.
9. Ojala, D., Montoya, J. and Attardi, G. (1981) *Nature* 290, 470-474.
10. Barrell, B.G., Bankier, A.T. and Drouin, J. (1979) *Nature* 282, 189-194.
11. Barrell, V.G., Bankier, A.T., DeBruijn, M.H.L., Chen, E., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3164-3166.
12. Upholt, W.B. and Dawid, I.B. (1977) *Cell* 11, 571-583.
13. Ferris, S.D., Brown, W.M., Davidson, W.S., and Wilson, A.C. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6319-6323.
14. Walberg, M.W. and Clayton, D.A. (1981) *Nucl. Acids Res.* 9, 5411-5421.
15. Ramirez, J.L. and Dawid, I.B. (1978) *J. Mol. Biol.* 119, 133-146.
16. Rastl, E. and Dawid, I.B. (1980) *Cell* 18, 501-510.
17. Klukas, C.K. and Dawid, I.B. (1976) *Cell* 9, 615-625.
18. Goddard, J.M. and Wolstenholme, D.R. (1978) *Proc. Natl. Acad. Sci. USA* 75, 3886-3890.
19. Goddard, J.M. and Wolstenholme, D.R., 1980, *Nucl. Acids Res.* 8, 741-757.
20. Wolstenholme, D.R., Goddard, J.M. and Fauron, C.M.-R. (1979) in *Proc. 8th Ann. ICN-UCLA Symp. Molecular and Cellular Biology; Extrachromosomal DNA*, Cummings, D.J., Dawid, I.B., Borst, P., Weissman, S.M. and Fox, C.F., Eds., pp. 409-425, Academic Press, New York.
21. Battey, J., Rubenstein, J.L.R. and Clayton, D.A. (1979) in *Proc. 8th Ann. ICN-UCLA Symp. Molecular and Cellular Biology; Extrachromosomal DNA*, Cummings, D., Dawid, I.B., Borst, P., Weissman, S. and Fox, F., Eds., pp. 427-442, Academic Press, New York.
22. Mertens, S.H. and Pardue, M.L. (1981) *J. Mol. Biol.* 153, 1-21.
23. Wolstenholme, D.R., Fauron, C.M.-R. and Goddard, J.M. (1980) in *The Organization and Expression of the Mitochondrial Genome, Developments in Genetics* - pp. 241-250, Saccone, C. and Kroon, A.M., Eds., Vol. 2, Elsevier/North Holland Biomedical Press, Amsterdam.
24. Wolstenholme, D.R. and Fauron, C.M.-R. (1976) *J. Cell Biol.* 71, 434-448.
25. Fauron, C.M.-R. and Wolstenholme, D.R. (1976) *Proc. Natl. Acad. Sci. USA* 73, 3623-3627.
26. Fauron, C.M.-R. and Wolstenholme, D.R. (1980) *Nucl. Acids Res.* 8, 2439-2452.
27. Fauron, C.M.-R. and Wolstenholme, D.R. (1980) *Nucl. Acids Res.* 8, 5391-5410
28. Polisky, B., Greene, P., Garfin, D.E., McCarthy, B.J., Goodman, H.M. and Boyer, H. W. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3310-3314.
29. Goddard, J.M., Masters, J.N., Jones, S.S., Ashworth, W.D., Jr., and Wolstenholme, D.R. (1981) *Chromosoma (Berl.)* 82, 595-609
30. Bolivar, F. (1978) *Gene* 4, 121-136.
31. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H., Falkow, S. (1977) *Gene* 2, 95-113.
32. Gronenborn, B. and Messing, J. (1978) *Nature* 272, 375-377.
33. Messing, J. (1981) in *Proc. Third Cleveland Symp. on macromolecules: Recombinant DNA*, Walton, E.G., Ed., pp. 143-153, Elsevier Scientific Pub. Co., Amsterdam.
34. Messing, J., Gronenborn, B., Muller-Hill, B., and Hofschneider, P.H. (1977) *Proc. Natl. Acad. Sci. USA* 74, 3642-3646.
35. Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucl. Acids Res.* 9, 309-321.
36. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B. (1980) *J. Mol. Biol.* 143, 161.

37. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
38. Anderson, S., Gait, M.J., Mayol, L., and Young, I.G. (1980) *Nucl. Acids Res.* 8, 1731-1743.
39. Sanger, F. and Coulson, A.R. (1978) *FEBS Lett.* 87, 107-110.
40. Brutlag, D.L., Clayton, J., Friedland, P., and Kedes, L. (1982) *Nucl. Acids Res.* 10, 279-294.
41. Jue, R.A., Woodbury, N.W. and Doolittle, R.F. (1980) *J. Mol. Evol.* 15, 129-148.
42. Doolittle, R.F. (1981) *Science* 214, 149-159.
43. Crews, S., Ojala, D., Posakony, J., Nishiguchi, J., and Attardi, G. (1979), *Nature* 277, 192-198.
44. Sekiya, T., Kobayashi, M., Seki, T. and Koike, K. (1980) *Gene* 11, 53-62.
45. Eperon, I.C., Anderson, S. and Nierlich, D.P. (1980) *Nature* 286, 460-467.
46. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4801-4805.
47. Schnare, M.N. and Gray, M.W. (1982) *Nucl. Acids Res.* 10, 3921-3932.
48. Dubin, D.T., Hsu Chen, C.C., Timko, K.D., Azzolina, T.M., Prince, D.L. and Ranzini, J.L. (1982) in *Mitochondrial Genes*, Slonimski, P., Borst, P. and Attardi, G., eds., pp. 89-98, Cold Spring Harbor Laboratory.
49. Dubin, D.T., and Baer, R.J.D. (1980) in *The Organization and Expression of the Mitochondrial Genome*, Developments in Genetics, Saccone, C. and Kroon, A.M., Eds., Vol. 2, pp. 231-240, Elsevier/North Holland Biomedical Press, Amsterdam.
50. Blanc, H., Wright, C.T., Bibb, M.J., Wallace, D.C., and Clayton, D.A. (1981) *Proc. Natl. Acad. Sci. USA* 78, 3789-3793.
51. Seilhamer, J.J., and Cummings, D.J. (1981) *Nucl. Acids Res.* 23, 6391-6406.
52. Dujon, B. (1980) *Cell* 20, 185-197.
53. Baer, R.J., and Dubin, D.T. (1980) *Nucl. Acids Res.* 8, 3603-3610.
54. Brosius, J., Dull, T.J., and Noller, F. (1980) *Proc. Natl. Acad. Sci. USA* 77, 201-204.
55. Kim, S.-H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A., Seeman, N.C. and Rich, A. (1974) *Science* 185, 435-440.
56. Ojala, D., Crews, S., Montoya, J., Gelfand, R., and Attardi, G. (1981) *J. Mol. Biol.* 150, 303-314.
57. Bonitz, S.G., Berlani, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F.G., Nobrega, M.P., Thanlenfeld, B.E., and Tzagoloff, A. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3167-3170.
58. Heckman, J.E., Sarnoff, J., Alzner-DeWeerd, B., Yin, S. and RajBhandary, U.L. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3159-3163.
59. Fox, T.D. and Leaver, C.J. (1981) *Cell* 26, 315-323.
60. Wilson, E.O., Eisner, T., Briggs, W.R., Dickerson, R.E., Metzberg, R.L., O'Brien, R.D., Susman, M., and Boggs (1978) *Life on Earth*, 2nd edn., p. 663, Sinauer Assoc. Inc., Sunderland, Mass.,