Sequence and arrangement of the genes for cytochrome *b*, URF1, URF4L, URF4, URF5, URF6 and five tRNAs in *Drosophila* mitochondrial DNA

Douglas O.Clary, Jill A.Wahleithner and David R.Wolstenholme

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

Received 19 March 1984; Accepted 13 April 1984

ABSTRACT

The nucleotide sequence of a segment of the mtDNA molecule of *Drosophila yakuba* has been determined, within which have been identified the 3' end of the large rRNA gene and the entire genes for tRNA^{leu}, URF1, tRNA^{ser}, cytochrome *b*, URF6, tRNA^{pro}, tRNA^{thr}, URF4L, URF4, tRNA^{his} and URF5. The genes are arranged in the order given, with the large rRNA gene being closest to the A+T-rich region which contains the origin of replication. Transcription of all of these genes except those for cytochrome *b*, URF6, tRNA^{ser} and tRNA^{thr} proceeds in the same direction as replication. Differences occur in the relative arrangement and in the direction of transcription of these twelve genes between *D. yakuba* and mammalian mtDNA molecules. Internal AGA codons occur in all of the polypeptide genes except URF6. Comparisons of the positions of these AGA codons to codons in corresponding mouse genes is consistent with the view that in the *D. yakuba* mitochondrial genetic code AGA specifies serine. Genes equivalent to all of the polypeptide, tRNA and rRNA genes found in mammalian mtDNA have now been identified in *D. yakuba* mtDNA.

INTRODUCTION

We have reported nucleotide sequences of segments of the mitochondrial DNA (mtDNA) molecule of *Drosophila yakuba*, within which we have identified the genes for two rRNAs, 18 tRNAs and nine polypeptides (cytochrome *c* oxidase (CO) subunits I, II and III, ATPase subunit 6, Unidentified Reading Frames (URF) 1, 2, 3, 5 and A6L) (1-5). Sequences of segments of the *D. melanogaster* mtDNA molecule corresponding to some of the sequenced segments of the *D. yakuba* mtDNA molecule have also been determined (1,2,6). The genes found in *Drosophila* mtDNA are the same as those found in mammalian mtDNA (7-12) but their arrangement in the *Drosophila* and mammalian molecules differs considerably. Further, the triplet AGA which is used in mammalian mtDNA only as a rare termination codon (7-9) specifies an amino acid in *Drosophila* mtDNA (1) which is most probably serine (3,4,6).

In this paper we report a sequence of the *D. yakuba* mtDNA molecule that contains the 3' end of the large rRNA gene, and the entire genes for

cytochrome b, URF1, URF4L, URF4, URF5, URF6 and five tRNAs. This completes identification in D. yakuba mtDNA of the entire gene complement common to mammalian mtDNAs.

MATERIALS AND METHODS

Experimental details regarding isolation of mtDNA from D. yakuba (stock 2371.6, Ivory Coast), preparation and identification of pBR322 and pBR325 clones of D. yakuba mtDNAs, restriction enzyme digestions, electrophoresis, cloning of fragments into M13mp8 or M13mp9, purification of single-stranded and double-stranded M13 DNAs and preparation of viral DNAs containing partial deletions of cloned restriction fragments of mtDNA are given or referred to in (1,3).

All DNA sequences were obtained from M13mp8- or M13mp9-cloned fragments by the extension-dideoxyribonucleotide termination procedure (13) using [α - 32 P]dATP (600 Ci/mmol; New England Nuclear) as described (1). The sequencing strategies used are given in Fig. 1.

Computer analyses of DNA sequences and amino acid sequences were carried out using the programs referred to in (3).

RESULTS AND DISCUSSION

Gene structure and organization. The sequence of a continuous 6642 nucleotide segment of the D. yakuba mtDNA molecule, determined using the strategies summarized in Fig. 1, is shown in Fig. 2. Within this sequence are six open reading frames. From comparisons of nucleotide and amino acid sequences to the corresponding sequences of previously identified genes of mouse mtDNA (9; Table 1 and Fig. 3) five of these open reading frames have been clearly identified as the genes for URF1, cytochrome b, URF4L, URF4 and URF5. The remaining open reading frame is similar in length to the URF6 gene of mouse mtDNA (Fig. 3) and, while the overall homology of the amino acid sequences predicted from the two genes is only 16.5% (Table 1), there is a single short region of 22 amino acids (57-78, Fig. 3) with a homology of 64%. Designation of this D. yakuba open reading frame as URF6 also seems reasonable as genes corresponding to all of the other polypeptide genes found in mouse mtDNA have been identified in D. yakuba mtDNA, and only regions of the D. yakuba mtDNA molecule expected to contain portions of the A+T-rich region and the two rRNA genes remain unsequenced.

The sequence shown in Fig. 2 also contains five regions each of which can fold into the characteristic structure of a tRNA with anticodons indicating

Figure 1. A map of the *D. yakuba* mtDNA molecule showing the relative locations of the A+T-rich region (cross-hatched), the two rRNA genes (dotted), the origin (O) and direction (R) of replication, and *EcoRI* and *HindIII* sites and fragments (A-E and A-F, respectively), (see (1) for details and references). The bar under the map indicates the segment sequenced. This segment is expanded below and the arrows beneath it show the extent and direction of individual sequences obtained, either from the ends of various restriction fragments or from deletion subfragments of restriction fragments. The 5'-3' direction of the sequence of this segment shown in Fig. 2 is from left to right.

them to be the genes for $\text{tRNA}_{\text{CUN}}^{\text{leu}}$, $\text{tRNA}_{\text{UCN}}^{\text{ser}}$, tRNA^{pro} , tRNA^{thr} and tRNA^{his} (Fig. 4). The six polypeptide and five tRNA genes occur in the order $\text{tRNA}_{\text{CUN}}^{\text{leu}}$, URF1, $\text{tRNA}_{\text{UCN}}^{\text{ser}}$, cytochrome b, URF6, tRNA^{pro} , tRNA^{thr} , URF4L, URF4, tRNA^{his} and URF5. The sequence located upstream from the $\text{tRNA}_{\text{CUN}}^{\text{leu}}$ gene is highly homologous to the 3' end of the 16S rRNA gene of mouse mtDNA (9) (Fig. 2; see below), and the 3' end of the URF5 gene is located adjacent to the tRNA^{phe} gene as shown previously (5). All of these genes except cytochrome b, URF6, $\text{tRNA}_{\text{UCN}}^{\text{ser}}$ and tRNA^{thr} are transcribed in the same direction as that in which replication proceeds around the molecule (Fig. 5).

Alignment of the predicted amino acid sequences (Fig. 3) of the six polypeptide genes of *D. yakuba* and mouse mtDNAs indicates that insertion/deletions have occurred between all of the corresponding genes. The largest single difference occurs between the URF5 genes. Of the first 65 codons of the mouse URF5 gene, a total of 32 are absent from the *D. yakuba* URF5 gene (Fig. 3), which apparently includes a block of the first 26 codons.

We have previously shown that the locations of the large rRNA, URF1 and $\text{tRNA}_{\text{UCN}}^{\text{ser}}$ genes relative to other genes differ between *D. yakuba* and mammalian mtDNAs (1,2; Fig. 5). The arrangements of the other polypeptide and tRNA genes included in the presently reported sequence (Fig. 2) also differ between *D. yakuba* and mammalian mtDNA molecules (Fig. 5). The order of the URF4L,

Table 1. Nucleotide and amino acid sequence comparisons of the cytochrome b, URF1, URF4L, URF4, URF5 and URF6 genes of D. yakuba with the corresponding genes of mouse and yeast.

Gene	Number of nucleotides (<u>D. yakuba</u>)	% Nucleotide sequence homology ^{a,b} (<u>D. yakuba</u> /mouse)	% Amino acid sequence homology ^b	
			<u>D. yakuba</u> / mouse	<u>D. yakuba</u> / yeast
Cyt b	1134	65.2	66.8	52.5
URF1	972	51.8	45.6	-
URF4L	288	34.3	40.4	-
URF4	1338	50.6	42.3	-
URF5	1719	42.0	32.5	-
URF6	522	37.5	16.5	-

^aExcluding nucleotides concerned with termination.

^bThese values include all deduced insertion/deletions (see Fig. 3). The data for mouse and yeast mtDNAs are taken from (9) and (14), respectively.

URF4 and URF5 genes relative to each other is the same in D. yakuba and mammalian mtDNAs, but is reversed with respect to the location of other polypeptide genes. Furthermore, in mammalian mtDNA a cluster of three tRNA genes (tRNA^{his}, tRNA^{leu}_{CUN} and tRNA^{ser}_{AGY}) is found between URF4 and URF5, whereas in D. yakuba mtDNA only tRNA^{his} is located in this position.

The URF6 gene in mammalian mtDNA is transcribed in the opposite direction to the cytochrome b gene from which it is separated by the tRNA^{glu} gene, and the carboxyl terminal regions of the URF6 and URF5 genes overlap. In D. yakuba mtDNA the amino terminus of the URF6 gene is separated from the URF4L gene by the tRNA^{pro} and tRNA^{thr} genes, and the URF6 gene ends in a standard translation termination codon (TAA) which is separated by three nucleotides from the apparent translation initiation codon (ATG) of the cytochrome b gene (Fig. 2). This is the only case so far recorded for any metazoan mtDNA where two polypeptide genes are adjacent in the molecule, are not separated by a tRNA gene, and are translated in the same reading frame of one strand.

In mammalian mtDNAs the carboxyl terminal region of the URF4L gene, which ends with a TAA codon, overlaps the amino terminus of the URF4 gene by seven nucleotides, but these two genes would be translated in different reading frames. However, only a single mRNA containing the transcripts of both URF4L and URF4 has been identified (15). In Drosophila mtDNA the URF4L gene ends in TA, the latter nucleotide being adjacent to the ATG translation initiation codon of the URF4 gene. A similar situation is found between the ATPase 6 and COIII genes of both mammalian and Drosophila mtDNAs (3,6-9). From studies

Figure 2.

ACTAATGCAATAACTCCTCCTAATTATTAGGAATTGAACGAAGAATTGCGTAAGCAAATAAAAAATATCATTCT 1725
 V L A I V G G L K N P I S R L I A Y A F L F Y W E
 GGTGTAATGTGAGCTGGTGTACTAAAGGATTAGCAGGAATAAAGTTATCTGGGTCTCCCAATAAATTTGGTCTA 1800
 P Q I H A P T V L V L P N A P I F N D P D G L L N P S
 ATTAATACTAATGAAATTAGAATAAAAAATTATTACAATAAATCCTACAATATCCTTAAATGTGAAGTATGGGTGA 1875
 I L V L S I L I F I M V I F G V I D K F T F Y P H
 AAAGGAATTTTATCAATATTAGAATTTAAACCAATAGGGTTATTAGATCCTGTTTGATGTAAAAATAGTAGATGA 1950
 F P I K D I N S N L G I P N N S G T Q H L F L L H
 ATTATAGTTATAGCAAGAACAATAAAAGGTAAAAATAAATGAAATGTGAAAAATCGAGTTAAAGTAGCATTATCT 2025
 I M T M A L V I F P L I F H F T F F R T L T A N D
 ACAGCAAATCCTCCTCATAATCATCTACTAAGTCTATACCTAAATAAGGGATAGCTGACAATAAATTAGTAATT 2100
 V A F G G W L W Q V L D M G L Y P I A S L L N T I
 ACAGTTGCTCCTCAAAATGATATTGTCTCAAGGTAACACATAACCTATAAAAGCTGTTCCTATTACTAAAAAT 2175
 V T A G G W F S M Q G W P L V Y G M F A T G M V L F
 AAAATAATTACTCCTACTAATCAAGTTGGTGTAATAAATATGATCCGTAATAAATTCCTCGACCAATATGTAAG 2250
 L I I V G V L W T P T F L Y S G Y Y I G R G I H L
 TAAATACAAATAAAAAAAATGATGCACGGTTAGCGTGTAAGTTCGTAATAATCAACCATAATTTACATCTCGG 2325
 Y I C I F F F S A G N A H L T R L L W G Y N V D R
 CAAATATGATTAACACTATAAAAGCTAAAGTTAACATCTGCTGTGTAGTGATAGCTAAAAATAATCCAGTTAAA 2400
 C I H N V S Y F A L N V D A T Y H M A L F L G T L
 ATTTGAATAATTAACATAATCCAAGTAATGATCCAAAATTTTCATCATCTTGAATATTAATTTGGAGCTCGTAAA 2475
 I Q I I L C L G L L S G F N W W S S I N I P A P L
 TCAACTAAAGCATTATTAGCAATTTTAAATAAAGGTTGGGAATTCGTAAGGTTTATGCMATTAATTAAGATATT 2550
 D V L A N N A I K F L P H S N R L P K H M *** S M
 ← Cyt b
 ATTCCGATAGACCTTTAAATAGTTTAGTAATTTTACTACAACAATTAAGTAATTAATAAATAATTTATTAAT 2625
 M R I P G K F L K T I K V V V I L T I L L Y N M L
 AAAATGTTACAAAATTTGTTGGAAAATTTATATAATTTATTTAAAGATAAAGAATTTTCCTGTAATAAAGAATTT 2700
 L I T V F N T P F N Y L K N L S L S N E T F Y S N
 ATTTCAATAATAGATTGTATTTTCGTTATTATTAAAAATAAAGTAATAGAAGTTTATCAAGAATTATTGATAAA 2775
 M E I I S Q M E N N M L F L T I S T K D L I M S L
 ATAAATATAAAAAATAAAATAATATGGAATAAAGTTAATTTAATTGATAAATAAATATTTTCATTAGAAGCT 2850
 I F M F F L I F M S F L T L K I S L N F M E N S A
 AATGATGTAACATAAATAAATAAACAAGTATTCCTCCTAAAAAAATTAATAAATAAATGATGAGTATCAAAAA 2925
 L S T V Y I F L V L M G G L F I L F L I Y S Y W F
 CTTTAGTTATTAATCCTGAAAGTAAACATACAAAAATTGTTGAATTAATAAAGTTAATCCTAAAGCTAATGGG 3000
 S K T M L G S L L C V F I T Q I L L T L G L A L P
 TGAATTATATTAATAAATAAATAGATCTAGTAATAATTAATGAATATAATTAATTAATGAATAATTTCAACAGGT 3075
 H I M N F F I I S T T I I L S Y L M L Q I I
 ← URF6
 tRNA^{pro} →
 AGTTTATTTATAAATATTAATTTTGGGGATTAATGAAAAAGAAATTTCTTTTCTCTTGAAGTTTAAAAAGAAAT 3150
 ← tRNA^{thr} URF4L →
 AATCTTATTTTGTATTACAGACCAATGTTTTTATTAACTATTAAAACTAATGATTATAATTTTATATTGAAG 3225
 L P M I L F I L G L F C F V S N R K H L L S M L L
 TTTACCTATAATTTTATTTTATTTAGGTTTATTTGTTTGTTCATCGAAAACATTACTTTCAATACITTT 3300

S L E F I V L M L F F M L F I Y L N M L N Y E N Y
 AAGTTTGAATTTATTGTTTTAATATTATTTTTATATTATTTATTTTAAATATACTAAATTATGAAATTA 3375
 F s M M F L T F s V C E G A L G L S I L V S M I R
 TTTTGAATAATATTTTTAACATTAGACTATGTGAAGGACCTTAGGTTTATCAATTTTGGTTTCTATAATTCG 3450
 T H G N D Y F Q S F s I M ** M L K I I L F L L F L
 TACTCATGGAATGATTATTTTCAATCTTTTGAATATATATAATGTTAAAAATTATTTTATTTTATTATTTTA 3525
 T P V C F I N N M Y W M V Q I M L F F I s F I F L
 ACTCCTGTTTGTTTTATAATAATATATTGAATGGTACAAATTATACTATTTTATTAGATTATTTTTTTA 3600
 L M N N F M N Y W S E I S Y F L G C D M L S Y G L
 TTAATAAATAATTTTATAAATTATTGATCAGAAATTTCTTATTTTTCAGGATGTGATATATTATCTTATGGTTG 3675
 V L L S L W I C S L M L L A s E s I N K Y N N Y K
 GTTTTCCTTAGTTTATGAATTTGTTTCATTGATATTATTAGCTAGAGAAAGAAATTAATAATATAAATTACAAA 3750
 N L F L L N I V I L L L L V L T F S s M s L F M
 AATTTATTTTATTAAATATTGTTATTTTGTATTATTATTGTTTAACTTTTTCAGAATAAGATTATTATA 3825
 F Y L F F E s s L I P T L F L I L G W G Y Q P E R
 TTTTATTATTTTGAAGAAGATTAAATCTACATTATTTTAATTTTACGATGAGGTTATCAACCGGAACGA 3900
 L Q A G V Y L L F Y T L L V S L P M L I G I F Y V
 TTGCAAGCTGGTGTTTATTTATTATTTTATACTTTATTAGTTTCTTTACCTATATTAATTGGAATTTTATGTA 3975
 M N K T G S M N F Y L M N N F M F N Y D L L Y F C
 ATAAATAAACTGGGTCAATAAATTTTATTTAATAAATAATTTTATGTTTAAATTATGATTTATTATATTTTGT 4050
 L L C A F L V K M P M F L V H L W L P K A H V E A
 TTATTATGTGCTTTTTTAGTGAAATACCAATATTTTATGACATTATGATTACCTAAAGCTCATGTTGAAGCT 4125
 P V S G S M I L A G I M L K L G G Y G L L R V I N
 CCTGTATCTGGTTCTATAAATTTTAGCTGGTATTATATTAATAAATTAGGAGGTTATGGTTTATTACGAGTAATTAAT 4200
 F L Q L M N L K Y S F V W I s I s L V G G V L M s
 TTTTACAATTAATAAATTTAAATATAGATTGTTTGAATTAGAATTAGATTAGTAGAGGTGTGTTAATAAGA 4275
 L V C L R Q T D L K A L I A Y S S V A H M G I V L
 TTAGTATGTTTACGACAACTGATTTAAAGGCTTTAATTGCATATTCATCACTGCTCATATAGGAATTGTTTA 4350
 A G L L T M T Y W G L C G S Y T L M I A H G L C S
 GCTGTTTATTAACAATAACTTATTGAGGATTATGTGGATCTTATACATTAATAATTGCTCATGTTTATGTTCT 4425
 S G L F C L A N V S Y E R L G s R s M L I N K G L
 TCTGTTTATTTGTTTACCAATGTTTCTTATGAACGATTAGGAAGACGAAGAATTAATAATAAGGGGTTA 4500
 L N F M P A M T L W W F L L s S A N M A A P P T L
 TTAATTTTATACCTGCTATAACTTTATGATGATTTTATTAAGATCAGCTAATATAGCAGCTCCTCCAAGCTTA 4575
 N L L G E I S L L N s I V S W S W I S M I M L S F
 AATTTATTAGGAGAAATTTCTTTATTAATAAGAATTGTTCTTGGTCATGAATTTCAATAATTATATTATCTTTT 4650
 L S F F s A A Y T L Y L Y S F S Q H G K L F S G V
 TTATCTTTCTTTAGAGCAGCCTATACCTTTATATTATATTCTTTTAGTCAACATGGTAATTATTTTCTGGGGTA 4725
 Y S F s s G K I R E Y L L M L L H W L P L N L L I
 TATTCATTTAGAAGAGGTAATAATTCGGGAATATTATTAATATTATTACATTGATTACCTTTAAATTTATTAATT 4800
 L K s E S C I L W L * tRNA^{his}
 TTAATAAGAGAAATCATGTATTTTATGATTATATTTAAATAGTTTAAAAAAATACTAATTTGTCGGTGTAGTCAT 4875
 I C s I s F I N L I S I S
 ATGAAGTTATTCATTTTATAGTCGTGAATATTTTCAATTTGTAGAATTAGATTTTATTAATTTAATTTCTATTAG 4950

L T C F L L S L Y Y L L N N M V Y F I E W E V V S
TTAACATGTTTTTATTAAGTTTATATTATTATTAATAATATAGTTTATTTTATTGAATGAGAAGTAGTATC 5025

L N S M s I V M T F L F D W M S L L F M S F V L M
TTAAATTCATAGAAGATTGATAACTTTTTTATTGATTGAATAAGTTTATTATTTATATCTTTTGTCTTAT 5100

I A S L V I F Y s k E Y M E S D E N I N R F I M L
AATTGCTCTCTTAGTAATTTTTTATAGAAAAGAATATATAGAAAGTGATGAAAATATTAATCGGTTTATTATATT 5175

V L M F V L S M M L L I I s P N L V s I L L G W D
AGTATTAATATTGTTTGTCAATAATATTATTAATTATCAGACCAAATTTAGTGAGAATTCATTAGGGTGAGA 5250

G L G L V S Y C L V I Y F Q N I K S Y N A G M L T
TGGATTAGGACTTGTTCCTATTGTTTAGTAATTTATTTTCAAAATATCAAATCTTACAATGCTGGAATATTAAC 5325

A L S N R I G D V A L L L A I A W M L N Y G s W N
TGCATTATCTAATCGAATTGGAGATGTAGCTCTTCTTCTTGCTATTGCTTGAATATTAATATGCTAGATGAAA 5400

Y I F Y L E V M Q N E F S M L M I G S L V M L A A
TTATATTTTTTATTAGAAGTTATACAAAATGAATTTTCAATATTAATAATTGGTAGTTAGTTATATTAGCTGC 5475

M T K s A Q I P F S S W L P A A M A A P T P V S A
TATAACTAAAGAGCTCAAATTCCTTTTTTCTTCTGATTACCAGCTGCTATAGCTGCCCTACTCCTGCTCTCTGC 5550

L V H S S T L V T A G V Y L L I R F N I V L S T S
TTTAGTCCATTCTCTACTTTTAGTAACAGCAGGAGTTTATTTATTAATTCGATTAAATATTGTTTAACTACTTC 5625

W L G Q L L L L L S G L T M F M A G L G A N F E F
TTGATTAGGACAATTATTGTTGTTATTATCAGGTTTAACTATATTTATAGCAGGACTGGGGCTAATTTTGAGTT 5700

D L K K I I A L S T L S Q L G L M M s I L S M G F
TGATTAAAGAAAATTATGCGTTGTCAACTCTTAGTCAATTAGGATTAATAAAGAATTTTATCTATGGGATT 5775

Y K L A M F H L L T H A L F K A L L F M C A G A I
TTATAAATTAGCTATATTTCAATTATTAACATGCTTTATTTAAAGCTTTATTTATTTATATGTCAGGAGCTAT 5850

I H N M N N S Q D I R L M G G L s I H M P L T S A
TATTCATAATAAATAATTCCTCAAGATATTCGGTTAATAGGAGGGTTGAGAATTCATATACCTTTAACTTCAGC 5925

C F N V S N L A L C G M P F L A G F Y S K D M I L
TTGTTTTAATGTTTCTAATTTGGCTTTATGTGGAATACCATTTTATGCTGGATTTTATTCTAAGGATATAATTTT 6000

E I V s I s N I N M F S F F L Y F F S T G L T V S
AGAAATGTTAGAATTAGAAATATTAATATATTTTCATTTTTTTTATACTTTTTTCCACAGGTTTAAACAGTAAG 6075

Y S F R L V Y Y S M T G D L N C G S L N M L N D E
TTATTCCTTTTCGGTTAGTATATTATTCAATAACTGGAGATTTAAATTGTGGAAGTTTAAATATATTAATGATGA 6150

S W V M L R G M L G L L F M s I I G G s M L N W L
AAGTTGAGTAATACTTCGGGTATACCTAGGCTTATTATTATAAGAATTATTGGAGGTAGAATGTTAAATTGGTT 6225

I F P F P Y M I C L P G Y L K M L T L F V C I V G
AATTTTTCCITTTCTATATAATTGTTTACCTGGGTATTTAAAAATATTAACCATTTTGTGTTGATTGTTGG 6300

G L F G Y L I S I s N L Y S L N K S L L N Y N L T
GGGTTTATTGGGTATCTAATTTCTATCAGAAATTTATATTCTTTGAATAAATCTCTATTAATTAATTAATTAAC 6375

L F L G S M W F M P Y I s T Y G M I F Y P L N Y G
ATTATTTTTAGGTTCAATATGATTATACCTTATATAGAACTTATGGAATAATTTTTTATCCTTTAAATTATCG 6450

Q L V V K S F D Q G W S E Y F G G Q H L Y Y K L S
TCAATTAGTTGTAAGAAAGTTTGTATCAAGGTTGGTCAGAATATTTGGGGGTCAACATTATATTACAAGTTATC 6525

N Y S K T L F L M H N N S L K I Y L M L F V F W I
AAATTATTCTAAACTTTATTTCTGATACATAATAATAGTTTAAAAATTTATTTAATATTATTGTTTGAAT 6600

M I L F S F L L F L * tRNA^{phe} →
TATAATTTTATTAGTTTTTTATTATTTTATATTCAAAATAG

with HeLa cells there is evidence for precise cleavage between the terminal UA of the ATPase 6 transcript and the AUG initiation codon of the COIII transcript, but the recognition signal by which this is facilitated is unknown (7,15).

It appears that any ATN codon may act as a methionine specifying, translation initiation codon in polypeptide genes of mammalian mtDNA (7-9). The first ATN codon in the open reading frame interpreted as the URF1 gene (Fig. 2) following the 3' end of the *D. yakuba* tRNA^{leu}_{CUN} gene is ATA. Although it has been suggested that ATAA may serve as the initiation codon of the *Drosophila* COI gene (4,6), the initiation codon of each of the other *Drosophila* mitochondrial polypeptide genes sequenced seems to be either ATT or ATG (Fig. 2; 1,3-6). From alignment of amino acid sequences (1 and Fig. 3) the ATA codon in the *D. yakuba* URF1 gene is four codons upstream from the initiation codon of the human and bovine URF1 genes and seven codons upstream from the initiation codon of the mouse URF1 gene (Fig. 3). In the *D. yakuba* URF1 gene sequence, an ATT codon occurs at a position one codon upstream from the initiation codon of the human and bovine URF1 genes. This would be the most likely initiation codon of the *D. yakuba* URF1 gene in the event that ATA cannot specify initiation in the *D. yakuba* mitochondrial genetic code. The ATA and ATT codons of the *D. yakuba* URF1 gene mentioned above are conserved in the *D. melanogaster* and *D. virilis* mtDNA molecules (1). In *D. yakuba* mtDNA, inframe ATA codons also occur seven and 18 codons upstream from the codons interpreted (Fig. 2) as the initiation codons of URF4L (ATG) and URF6 (ATT) respectively, but in both cases in regions which overlap the sense strand of an adjacent tRNA gene. In two cases where the 5' ends of a tRNA gene and a

Figure 2. Nucleotide sequence of the segment of the *D. yakuba* mtDNA molecule identified in Fig. 1. From considerations of nucleotide and amino acid sequence homologies to mouse mtDNA (9), this sequence contains the 3' end of the large rRNA gene and the genes for cytochrome b, URF1, URF4L, URF4, URF5 and URF6. Sequences of five tRNA genes (see Fig. 4) are shown boxed and the anticodon of each is underlined. The nucleotide sequence shown is the (5'-3') sense strand for the large rRNA, URF1, URF4L, URF4, URF5, tRNA^{leu}_{CUN}, tRNA^{Pro} and tRNA^{His} genes, and the arrows indicate the direction of transcription of each gene. Asterisks indicate partial or complete termination codons. In the amino acid sequences a lower case letter s indicates the tentative assignment of serine to an AGA codon (see text). The overlines indicate nucleotides (121, 221-224) which are not found in the mouse 16S rRNA gene. The inverted open triangles indicate the approximate positions of two single nucleotides and of segments of 17, seven and 23 nucleotides present in the mouse 16S rRNA gene but not in the *D. yakuba* large rRNA gene. Nucleotides 38-172 and 329-746, and 1180-1446 are given in (1) and (2), respectively. The sequence containing the carboxyl terminal 815 nucleotides of the URF5 gene and the entire tRNA^{Phe} gene are given in (5).

Cytochrome b

D. yakuba MHKPLRNSHP LFKIANNALV DLPAPINIS_s WNWFGSLGL CLIIQILTGL FLAMHYTADV NLAFYSVNH_I 70
mouse .-TNM.KT.I.HSFI D....S....SV .MV..I... ..S.T MT..S..T..
yeast --AF.K.NV YLSLV.SYII .S.Q.SS.NYM..... .V...V..I .M....SSNI E...S.E..
 CRDENVYGLL RTLHANGASF FFCIYLHIG RGIYGSYLF -TPT-WLVGV IILFLVMCTA FMGVVLPWGQ MSFWGATVIT 150
I .YM.....MLF..V. ..L.....T. -ME..-NI.. LL..A.A..
 I...HN.YI. .Y..... .MVMFM.MA K.L.....RS PRV.LWN... ..FI.TIA.. .L.CCVY.. .H.....
 NLLSAIPYLG MDLVQWLWGG FAVDNATLTR FFTFHFILPF IVLAMTMIHL LFLHQTGSNN PIGLNSNIDK IPFHPYFTFK 230
I. TT..E.I... .S.K..... .A..... .IA.LAIV..E..... .T...DA..Y.I..
 ..F....FV. N.I.S..... .S.S.P.IQ. ..AL.YLV.. .IA..VIM.. MA..IH..S. .L.ITG.L.R .M.S..I..
 DIVGFIVMIF ILISLVLI_sP NLLGDPDNFI PANPLVTPAH IQPEWYFLFA YAILRSIPNK LGGVIALVLS IAILMILPFY 310
 ..L.ILI.FL ..MT...FF. DM.....YMN..P. .K..... ..L..I.. .L.ALM..L
 .L.TVFLFML .ALP.FYS. .T.Q...Y. .G.....S .D...L.PFD. .L...TMFAA .LV.LV...T
 NLsKFRGIQF YPINQILFWS MLTVVILLTW ICARPVEEYP VLIGQILTII YFLYYLIN-P LVTKWWDNLL N-----*
 HTS.Q.SLM. R..T...Y.I LVANLLI... .GQ...H.F II...LAS.S .SII..LM. ISGIIIE.KM. KLYP----*
 DRSVV..NT. KVLSKFF.FI FVPNFV..GQ ...CH..V.. ..M...A.F. .A.F..I-V P.ISTIE.V. FYIGRVNK*

URF1

D. yakuba MEFILSLIGS LLLIICVLVS VAFILTLLERK VLGYIQIRK PNKVGMLGIP QPFCDAIKLF TKEQTYPLLS 70
mouse -----NI .T.LVPI.IA M.....V... I...M.L... .I..PY..L .A..M... M..PMR..TT
 NYLsYYISPI FSLFLSLFVW MCMPPFFVKLY SFNLGGLFPL CCTsLGVYTV MVAGWSSNSN YALLGGLRAV AQTISYEVsL 150
 SMLFI.A.T L..T.A.SL. VPL.MPH.P.I NL...I..I. ATSS.S..SI LWS..A..K .S.F.A... ..TM
 ALIMLSFIFL IGSYNMIYFF YQIYMWFLI ILFPMSLVWL TISLAETNRT PFDFAEGESE LVSGFNVEYs sGGFALIFMA 230
 .I.L..VLLM N...SLQTLI TT.EH..L.L PAW..AMM.F IST.....A ...LT..... A.A.P...F...
 EYAsILFMsM LFCVIFLGLCD VFNLLFYVKL T-----FI SFVFIWARGT LPRFRYDKLM YLAWKCLFSF SLNYLLFFIG 310
 .TN.IL.NA .TTI....PL YYIN.PELYS .NFMMEALLL .ST.L.I.AS Y.....Q.. H.L..N..PL T.ALCMWH.S
 FKILLFSFLL WIFFSKKLME N*
 LP.FTAGVPP YM----- *

URF4L

D. yakuba MIMILYWSLP MILFILGLFC FVSNRKHLIS MLLSLEFIVL MLFPMFLFIY- LNMLNYENYF sMMFLTFsVC 70
mouse .-PSTFFN.T .-A.S.S.LC TLMF.S..M. T..C..GM.. S..I.TSVTS .SNSMSSMP IPIT.V.AA.
 EGALGLSILV SMIRTHGNDY QQSFSIM--*
 .A.V..AL.. KVSIN.Y.T.. V.NLNLQC*

URF4

D. yakuba MLKIILFLLF LTPVCFINNM --YWM-VQIM LFFIsFIFL- LMNFMNYWS EISYFLGCDM LSYGLVLLSL 70
mousePS.M .L.LTWLSSP KKT.TN.TSY S.LISLTS.T .LWQTDENYK NF.NMFSS.P .TP.II.TA
 WICSLMLLAS Es-INKYNNY KN-LFLLNIV ILLLLLVLTF SsMsLFMFYL FFEsLIPTL FLILGWGYQP ERLQAGVYLL 150
 .LLP...M.S QNHLK.D.V LQK.YISMLI S.QI..IM.. .ATE.I...I L..AT..... II.TR..NQT ...N..I.F.
 FYTLVSLPM LIGIFYVMNK TCSMNFYLMN NFMFNVDLLY F---CLLC- -APLVKMPMF LVHLWLPKHA VEAPVSGSMI 230
IG.I.L ..ALILIQ.H V.TL.LMILS FTTHTL.ASW SNNLLW.A.M M...I...LY G..... ..IA....
 LAGIMLKLG YGLLRVINFL QLMNLKysFV WisIsLVGGV LmsLVCLRQT DLKALIAYS VAHMGIVLAG LLTMTYWGCL 310
 .A.L....S .MI.ISII. DPLTKYMAP F.LLS.W.MI MTSSI..... .S..... .S..AL..I.S IMIQ.P.SFM
 GSYTLMAHG LCSSGLFCLA NVSYERLGsR sMLINKGLLN PMPAMTLWWF LLSANMAAP PTLNLLGEIS LLNsIVSWSW 390
 .ATM..... .T..L..... .SN....IHS. T.IMAR..QM VF.L.AT..L MASL..L.L. .SI..M..LF ITMSLF...N
 ISMIMLSFLS FFsAAYTLYL YSFSQHKLFL SCVYSFsGK IREYLLMLLH WLPLNLLILK sESCI-LWL*
 FTI.LMGINI IITGM.SM.M IITT.R...T NHMINLQPSH T..LT..A.. MI..I..TTS PKLITG.TM*

URF5

D. yakuba ----- ICsI sF--INLISI --SLTCFLLS -LYYLLNMV YFIE-WEVVS 70
mouse INIFTTSILL IFILLSPIL ISMSNL.KH. N.PLYTTT.. KF.FIIS..P L.MFFH...E .M.TT.HW.T

LNSMeIVMTF LFDWMSLLFM SFVLMIASLV IFYsKEYMES DENINRFIML VIMFVLSMML LIIsPNLVsI LLGWDGLGLV 150
M...ELK.S. KT.FF.I..T .VA.FVTWSI MQLSSW..H. .P.....KY LTL.LIT.LI .TSAN.MFQL FI..E.V.IM

SYCLVIYQFN IKSYNACMLT A-LSNRIGDV ALLLAIW-M LNYGsWNYIF YLEVMQNEFS MLMIGSLVML AAMTKsAQIP 230
.FL.IGWWYG RTDA.TAA.Q .I.Y.....I GFI..MV.FS ..MNS.ELQQ IMFSNN.DNL IPLM.-.-LI ..TG.S..FG

FSSWLPAAAMA APTPVSAHV SSTLVTAGVY LLIRFNIVLS TYWLGQLLLL LSG-LTMFMA GLGANFEFDL KKIIALSTLS 310
LHP...S..E G.....L. ...M.V..IF ..V..HPLTT NNNFILTM. CL.A..TLFT AIC.LTQN.IF..S.

QLGLMMsILS MGFYKLMFH LLTHALFKAL LPMCAGAIH NMNNSQDIRL MGLsIHMP L TSACFNVSNL ALCGMFFLAG 390
.....VT.G .NQPH..FL. IC...F...MS.S... SLADE....K ..NITKI..F ..S.LVIGS. ..T....T.

FYSKDMILEI VsIsNINMFS FFLYFFSTGL TVSYSFRLVY YSMTGDLNCG SLNMLNDESU VMLRGLMLL FMsIIGGmL 470
.....L.I.A INTC.T.AWA LLITLIA.SM .AM..M.II. FVTMTKPRFP P.ISI.ENDP DLMNPKR.A .GS.FA.FVI

NWLIFPPFPY ICLPGYLKML TLFVCIVGGL FGYLISIsNL YSLNKSLLNY NLTFLGSMW FMPYIsTYGM IFYPLNYGQL 550
SYNIP.TSIP VLTMPWFLKT .ALIIS.L.F LIA.ELNNLT MK.SMNKA.P YSSPSTLLGF .PSI.HRITP MKSLNLSLKT

VVKSFDQGS EYFGGQHLHY KLSNYSKTLF LMHNSLKIY LMLFVFWIM- -ILFSFLLFL *
SLTLL.LI.L .KTIPKSTST LHT.-MT..T TNQGLI.L. F.S.LIN.II I.ILYSINLE *

URF6

D. yakuba --IIQLMYS LIITTSIIF NMHPLALCL TLLIQTIFVC LLSGLMTKSF WYSYLFLIF LGGMLVLFIY 70
mouse MNNYIFV.S. .FLVGCLGLA LKPS.IYG.. G.IVSGFVG. .MVLGFGG.. L-GLMV...YV.G.

VTSLASNEMF NLSIKLTLFS MFILFFMFIL SMILDKTSIT LFLMNEMQS IIEMNSYFTE NSLSLNKLYN FPTNFVTILL 150
T.AM.TE.YP ETWGSNW.IL G.LVLGVIME VFILICVLNY DEVGVINLDC LGDWLM.EVD DVGVMLEGGI GVAAMYS CAT

MNYLLITLIV VVKITKLFKG PIRMMs*
WMMVAVGWSL FAG.FIIIEI TRD---*

Figure 3. Comparisons of the amino acid sequences of cytochrome b, URF1, URF4L, URF4, URF5 and URF6 predicted from the nucleotide sequences of the respective genes of mtDNA of D. yakuba with the corresponding amino acid sequences of mouse (9) and yeast (14). A dot indicates an amino acid which is conserved relative to D. yakuba. A dash indicates an amino acid which is absent. An asterisk indicates a terminal T or TA, or a termination codon. D. yakuba amino acids specified by AGA are represented by a lower case letter s to indicate the presumed assignment of serine (see text).

polypeptide gene are adjacent in mammalian mtDNA (7-9) the transcript of the polypeptide gene appears to be cleaved from the primary transcript at the junction of the two genes (16).

The last sense codon of both the URF4 and URF5 genes is separated by a single T from the 5' terminus of a tRNA gene (Fig. 2). A similar situation has been found for the URF2 and COII genes of D. yakuba mtDNA. Presumably, the terminal T as well as the TA dinucleotide found at the end of the ATPase 6 and URF4L genes are completed as termination codons by polyadenylation (17) following cleavage of a polycistronic primary transcript (7,15).

We did not initially recognize the tRNA^{leu}_{CUN} gene between the large rRNA and URF1 genes of D. yakuba mtDNA (1,3). In mammalian mtDNA it is the tRNA^{leu}_{UUR}

Figure 4. Four tRNA genes of *D. yakuba* mtDNA shown in the presumed secondary structures of the corresponding tRNAs. To the right of the tRNA^{pro} gene is shown an alternative pairing configuration for the amino-acyl stem. A secondary structure for the tRNA^{ser}_{UCN} gene is given in (2).

gene which is found between the 16S rRNA and URF1 genes (7-9).

The sequence upstream from the 5' end of the tRNA^{leu}_{CUN} gene, (nucleotides 1-296, Fig. 2) is 69% homologous (except as indicated in Fig. 2) to a region of mouse mtDNA which specifies the 3' end of the 16S rRNA. Also, as we have shown previously (1), nucleotides 48-199 have the potential to form a secondary structure which is highly conserved among the large mt-rRNA genes of different eukaryotic species, and the 23S gene of *Escherichia coli*.

It is not possible from comparisons of the sequences of the *D. yakuba* and mouse large rRNA genes to determine whether the 3' terminal nucleotide of the *D. yakuba* large rRNA gene is immediately adjacent to the 5' end of the tRNA^{leu}_{CUN} gene (Fig. 2). However, the 17 nucleotides upstream from the 5' end of the *D. yakuba* tRNA^{leu}_{CUN} gene have the potential to fold into a secondary structure remarkably similar to the stem and loop structure which is found at, or near,

Figure 5. The arrangement of genes in the circular mtDNA molecules of *D. yakuba* (left) and mammals (right). The *D. yakuba* molecule is derived from the results of various studies (for references see 1-5) and the present data. The mammalian mtDNA molecule is derived from that given for mouse (9). Each tRNA gene (hatched area) is identified by the one letter amino acid code and individual serine and leucine tRNA genes are identified by the codon family (in parentheses) which their transcription products recognize. Arrows within and outside the molecules indicate the direction of transcription of each gene. Wavy lines in the *D. yakuba* molecule indicate uncertain gene boundaries. O and R in the *D. yakuba* molecule indicate the origin and direction of replication, respectively. O_H and O_L in the mammalian molecule indicate the origin of heavy and light strand DNA synthesis.

the terminus of a number of mammalian 16S mt-rRNA genes (Fig. 6), and for which a transcriptional attenuator function has been proposed (18-20).

The relative arrangement and direction of transcription of genes in the *D. yakuba* and mouse mtDNA molecules are compared in Fig. 5. Overall, in the *D. yakuba* mtDNA molecule, the proportions of the two strands which serve as

Figure 6. A potential stem and loop configuration of the nucleotides 5' to the tRNA^{leu}_{CUN} gene of *D. yakuba* mtDNA, compared to a similar potential configuration of the 3' terminal nucleotides of the mouse mitochondrial 16S rRNA gene (20). In the mouse sequence, the 3' terminal nucleotide of the 16S rRNA gene is adjacent to the 5' terminal nucleotide of the tRNA^{leu}_{UUR} gene (9).

templates for transcription of genes are approximately equal. All genes in the approximate half of the molecule to the right of the A+T-rich region except for the cytochrome b, URF6, tRNA^{ser}_{UCN} and tRNA^{thr} genes are transcribed in the same direction as that in which replication proceeds around the molecule, away from the A+T-rich region. All of the genes in the other half of the D. yakuba molecule, with the exception of three tRNA genes, are transcribed in the other direction. This differs from the situation in mammalian mtDNA molecules where all of the genes except those for URF6 and eight tRNAs are transcribed in the direction which is opposite to the direction of replication (7-9). It is interesting to note that the arrangement relative to each other of the five identified polypeptide genes (COI, COII, COIII, ATPase 6, cytochrome b) is almost identical in the D. yakuba and mammalian mtDNA molecules, and that in both cases they are all transcribed in the same direction.

Transfer RNA genes. The sequence of the tRNA^{ser}_{UCN} gene of D. yakuba mtDNA has been presented and discussed previously (2). The tRNA^{leu}_{CUN}, tRNA^{pro}, tRNA^{thr} and tRNA^{his} genes are 45%, 52%, 67% and 56% homologous, respectively, to the corresponding tRNA genes of mouse mtDNA (9). The major structural characteristics of these four tRNA genes are similar to those of the other 18 D. yakuba mt-tRNA genes we have previously described (1-5), including the occurrence of the constant nucleotides Pu₂₆, T₃₃ and Pu₃₇ (numbering system in 21).

The amino-acyl and dihydrouridine stems of the tRNA^{his} gene, the dihydrouridine and anticodon stems of the tRNA^{pro} gene, and the anticodon stem of the tRNA^{leu}_{CUN} gene all lack G-C pairs (Fig. 4). Among the stems of other D. yakuba mt-tRNA genes, G-C pairs are absent only in the amino-acyl and T Ψ C stems of the tRNA^{glu} and tRNA^{asp} genes, in the amino-acyl stem of the tRNA^{gln} gene (1-5) and in the dihydrouridine arm of the tRNA^{gly} gene. There appears to be only two standard base pairs and two G-T pairs holding together the two component strands of the amino-acyl stem of the D. yakuba mt-tRNA^{pro} gene (Fig. 4). Secondary structural stability of the amino-acyl stem could be improved by the formation of four standard base pairs and a G-T pair if nucleotide T₅ is permitted to loopout (Fig. 4). Evidence that such a loopout is possible in the anticodon stem of bovine mt-tRNA^{ser}_{AGY} has been presented (22). Also, the secondary structure of the amino-acyl stem of bovine mt-tRNA^{ser}_{UCN} appears to be dependent upon only four standard base pairs and a G-T pair (23).

The 22 tRNA genes found to date in D. yakuba mtDNA correspond in their

expected codon recognition properties to the 22 tRNA genes of mammalian mtDNA (7-9). In mammalian mtDNA a tRNA gene is found between the replication origin-containing region and the small (12S) rRNA gene. As the equivalent region of the D. yakuba mtDNA molecule has not yet been sequenced, it remains possible that this mtDNA molecule contains further tRNA genes. The most likely candidates are genes for a tRNA which recognizes AGA (serine) codons and for a tRNA which recognizes internal AUR (methionine) codons.

Codon usage and the genetic code. The frequencies of codons ending in A or T in the six polypeptide genes presently described, which vary from 92.4% (cytochrome b) to 97.9% (URF6), are within the range of values (91.6% to 98.1%) found for the other seven polypeptide genes of D. yakuba mtDNA which have been analyzed (3-5).

The URF5 gene contains two codons, CTG (leu; nucleotide 5681-5683, and 6548-6550, Fig. 2) and GAG (glu; 5696-5698, Fig. 2) which have not been found in any other polypeptide gene of D. yakuba mtDNA. Only the codons CAG, CGC, AGG and TAG (used as the termination codon in the COIII gene of D. melanogaster (2)) are absent from the 13 polypeptide genes of D. yakuba mtDNA we have sequenced.

The G+C content of the cytochrome b gene (26.1%) is within the range of G+C contents (24.2% to 30.1%) of the other four identified polypeptide genes of D. yakuba mtDNA. The G+C contents of the URF1, URF4L, URF4, URF5 and URF6 genes (21.0%, 16.2%, 20.5%, 22.3% and 15.2%, respectively) are all lower than the G+C contents of the identified polypeptide genes, as was found for the D. yakuba URF2 and URF3 genes (4,5).

As is the case in all previously sequenced Drosophila polypeptide genes (1,3-6), internal TGA codons are found in each of the cytochrome b, URF1, URF4L, URF4, URF5 and URF6 genes. Thirty-one out of 47 of these TGA codons correspond in position to tryptophan-specifying codons (TGA or TGG) in the equivalent mouse genes (Fig. 3) further supporting the view that, as in the mammalian and fungal mitochondrial genetic codes (9,24-26), TGA also specifies tryptophan in the Drosophila mitochondrial genetic code.

All of the six presently described polypeptide genes except URF6 contain internal AGA codons. The 48 AGA codons found in the cytochrome b, URF1, URF4L, URF4 and URF5 genes correspond in position to codons specifying 13 different amino acids, but never arginine, in the equivalent mouse genes (Fig. 3). AGAs correspond to twice as many serine-specifying codons (a total of 14) than to codons specifying any other amino acid (seven each for alanine and asparagine), which is consistent with the view that in the Drosophila mitochondrial genetic code AGA specifies serine (3-6).

ACKNOWLEDGEMENTS

We wish to thank Felix Machuca for excellent technical assistance. This work was supported by National Institutes of Health Grant No. GM 18375.

REFERENCES

1. Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M.-R. and Wolstenholme, D.R. (1982) Nucl. Acids Res. 10, 6619-6637.
2. Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R. (1983) Nucl. Acids Res. 11, 2411-2425.
3. Clary,D.O. and Wolstenholme,D.R. (1983a) Nucl. Acids Res. 11, 4211-4227.
4. Clary,D.O. and Wolstenholme,D.R. (1983b) Nucl. Acids Res. 11, 6859-6872.
5. Clary,D.O. and Wolstenholme,D.R. (1984) Nucl. Acids Res. in press.
6. de Bruijn,M.H.L. (1983) Nature 304, 234-241.
7. Anderson,S., Bankier,A.T., Barrell,B.G., de Bruijn,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.
8. Anderson,S., de Bruijn,M.H.L., Coulson,A.R., Eperon,I.C., Sanger,F. and Young,I.G. (1982) J. Mol. Biol. 156, 683-717.
9. Bibb,M.J., Van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) Cell 26, 167-180.
10. Grosskopf,R. and Feldmann,H. (1981) Curr. Genet. 4, 151-158.
11. Koike,K., Kobayashi,M., Yaginuma,K., Taira,M., Yoshida,E. and Imai,M. (1982) Gene 20, 175-183.
12. Pepe,G., Holtrop,M., Gadaleta,G., Kroon,A.M., Cantatore,P., Gallerani,R., De Benedetto,D., Quagliariello,C., Sbisà,E. and Saccone,C. (1983) Biochem. Int. 6, 553-563.
13. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
14. Nobrega,F.G. and Tzagoloff,A. (1980) J. Biol. Chem. 255, 9828-9837.
15. Ojala,D., Montoya,J. and Attardi,G. (1981) Nature 290, 470-474.
16. Montoya,J., Ojala,D. and Attardi,G. (1981) Nature 290, 465-470.
17. Mertens,S.H. and Pardue,M.L. (1981) J. Mol. Biol. 153, 1-21.
18. Dubin,D.T., Timko,K.D. and Baer,R.J. (1981) Cell 23, 271-278.
19. Dubin,D.T., HsuChen,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.
20. Attardi,G., Cantatore,P., Chomyn,A., Crews,S., Gelfard,R., Merkel,C., Montoy,J. and Ojala,D. (1982) in Mitochondrial Genes, Slonimski,P., Borst,P. and Attardi,G., Eds., pp. 51-71, Cold Spring Harbor Laboratory.
21. Gauss,D.H. and Sprinzl,M. (1983a) Nucl. Acids Res. 11, r1-r53.
22. de Bruijn,M.H.L. and Klug,A. (1983) EMBO J. 2, 1309-1321.
23. 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., Schrier,P.H., Smith,A.J.H., Staden,R. and Young,I.G. (1982) in Mitochondrial Genes, Slonimski,P., Borst,P. and Attardi,G., Eds., pp. 5-49, Cold Spring Harbor Laboratory.
24. Heckman,J.E., Sarnoff,J., Alzner-DeWeerd,B., Yin,S. and RajBhandary,U.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3159-3163.
25. Barrell,B.G., Bankier,A.T., de Bruijn,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. and Young,I.G. (1980) Proc. Natl. Acad. Sci. USA 77, 3164-3166.
26. Bonitz,S.G., Berlanì,R., Coruzzi,G., Li,M., Macino,G., Nobrega,F.G., Nobrega,M.P., Thalenfeld,B.E. and Tzagoloff,A. (1980) Proc. Natl. Acad. Sci. USA 77, 3167-3170.