Volume 11 Number 12 1983 Nucleic Acids Research

Nucleotide sequence of a segment of *Drosophila* mitochondrial DNA that contains the genes for cytochrome *c* oxidase subunits II and III and ATPase subunit 6

Douglas O.Clary and David R.Wolstenholme

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

Received 4 April 1983; Revised and Accepted 25 May 1983

ABSTRACT

The nucleotide sequence of a segment of the mtDNA molecule of Drosophila yakuba has been determined, within which have been identified the genes for $tRNA^{Leu}_{UUR}$, cytochrome c oxidase subunit II (COII), $tRNA^{lys}$, $tRNA^{asp}$, URFA6L, ATPase subunit 6 (ATPase6), cytochrome c oxidase subunit III (COIII) and $tRNA^{gly}$. The genes are arranged in the order given and all are transcribed from the same strand of the molecule in a direction opposite to that in which replication proceeds around the molecule. The $tRNA^{lys}$ gene is unusual among mitochondrial $tRNA^{lys}$ genes in that it contains a CTT anticodon. The triplet AGA is used to specify an amino acid in all of the COII, COIII, ATPase6, and URFA6L genes. However, the AGA codons found in these four polypeptide genes correspond in position to codons which specify nine different amino acids, but never arginine, in the equivalent polypeptide genes which have been sequenced from mtDNAs of mouse, yeast and Zea mays.

INTRODUCTION

The mitochondrial genome of all metazoans examined to date, which range from platyhelminth worms to man, is in the form of a circular molecule of approximately 16.5 kb (1). Mitochondrial DNA (mtDNA) molecules from different Drosophila species range in size from 15.7 to 19.5 kb, but this variability can be accounted for by differences in size of the A+T-rich region which contains the replication origin (2-4). We have recently sequenced part of the A+T-rich region, and segments lying on either side of this region of the mtDNA molecule of Drosophila yakuba (5). The latter segments were shown to contain genes which are also found in the mtDNA molecules of various mammalian species (6-8). However, the order in which the genes are arranged differs between the Drosophila and mammalian mtDNA molecules (5). We also found that in Drosophila mtDNA the triplet AGA is used to specify an amino acid (5). This again differs from the situation in mammalian mtDNAs where AGA is found either as a rare termination codon (human and bovine; 6,7) or not at all (mouse; 8).

In this paper we report the sequence of a 2550 nucleotide segment of the mtDNA molecule of D. yakuba which contains the genes for three known poly-

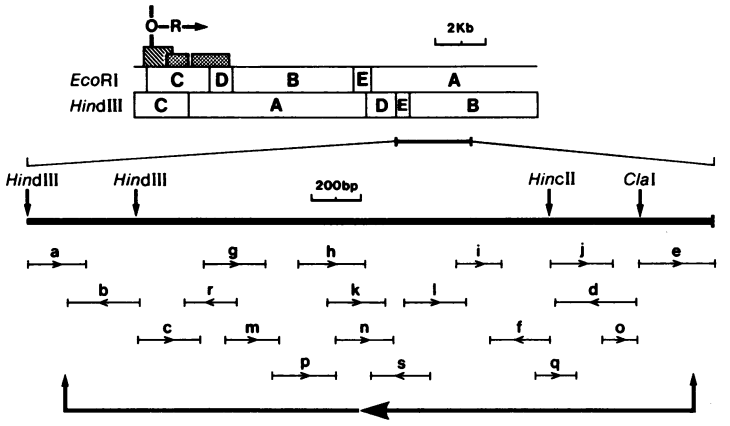


Figure 1. A map of the D. yakuba mtDNA molecule showing the relative locations of the A+T-rich region (crosshatched), the two rRNA genes (dotted), the origin (O), and direction (R) of replication, EcoRI and HindIII sites and fragments (A-E in each case) (see reference 5 for details and references). The bar under the map indicates the segment sequenced. This segment is expanded below, and the restriction sites and strategy employed to obtain the entire nucleotide sequence are shown. The origin of each sequence is as follows: a and b, the two different orientations of the HindIII-E fragment which was subcloned from the pBR325-cloned EcoRI-A fragment. c and d, the two ends of the smaller (2.1 kb) HindIII-ClaI subfragment of the pBR322-cloned HindIII-B fragment. e, the larger (2.8 kb) HindIII - ClaI subfragment of the HindIII-B fragment. f, a HincII - HindIII subfragment of the 2.1 kb HindIII - ClaI subfragment. g through q, DNaseI - ClaI deletions (cloned in M13mp8) of the 2.1 kb HindIII - ClaI subfragment. r and s, DNaseI - HindIII deletions (cloned in M13mp9) of the 2.1 kb HindIII - ClaI subfragment. The small vertical terminal arrows on the lower bar indicate the extent of the sequence shown in Fig. 2, and the solid arrow head shows the 5'-3' direction of this sequence.

peptides, cytochrome c oxidase subunits II and III, and ATPase subunit 6, an unidentified reading frame, URFA6L, and the genes for $tRNA^{leu}_{UUR}$, $tRNA^{lys}$ and $tRNA^{asp}$.

MATERIALS AND METHODS

Experimental details regarding isolation of mtDNA from Drosophila yakuba (stock 2371.6, Ivory Coast), preparation and identification of pBR322 and pBR325 clones of D. yakuba mtDNAs, restriction enzyme digestions, electrophoresis, recloning of fragments or subfragments into M13mp8 or M13mp9, and purification of M13 DNAs are given or referenced in (5).

The smaller (2.1 kb) HindIII - ClaI fragment of the pBR322-cloned HindIII-B fragment of D. yakuba mtDNA (Fig. 1) was cloned separately into

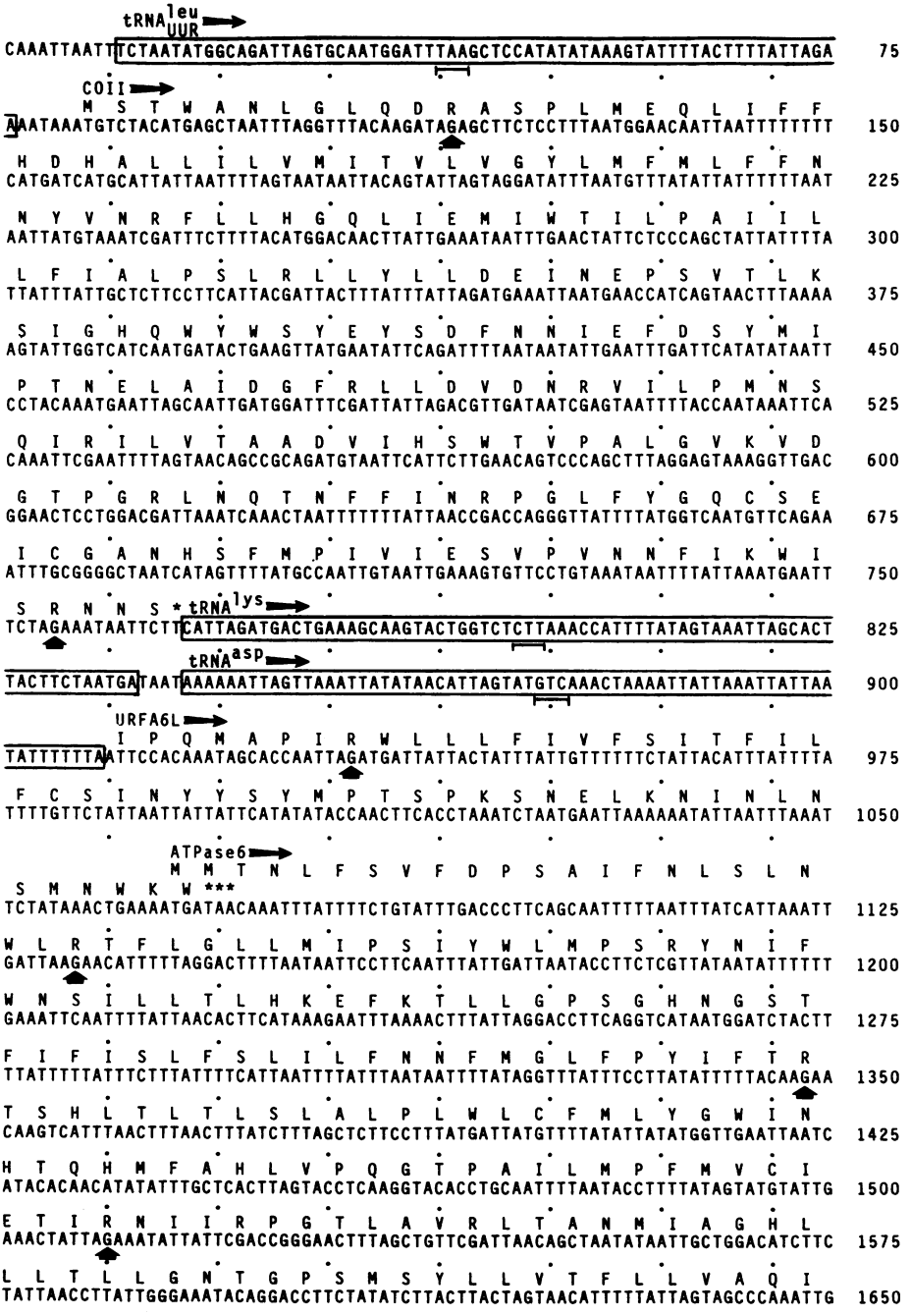
M13mp8 and M13mp9, and replicative forms (RF) of each hybrid M13 DNA molecule were prepared (9). Partial deletions of the 2.1 kb HindIII - ClaI fragment were generated using DNaseI digestion in the presence of Mn^{++} as described by Hong (10). In each case, the digestion products were religated and used directly to transfect E. coli JM103. Single-stranded DNA was prepared from virus of each of about 20 plaques, and viral DNAs containing suitable deletions of the original HindIII - ClaI fragment were selected by size using agarose gel electrophoresis.

All DNA sequences were obtained from M13mp8- or M13mp9-cloned fragments by the extension-dideoxyribonucleotide termination procedure (11) using [α-^{32}P]dATP (800 Ci/mM; New England Nuclear) as described previously (5). The sequencing strategies used are given in Fig. 1.

Sequences were stored and assembled using the computer method of Staden (12). Transfer RNA genes were identified within D. yakuba sequences from their ability to fold into the characteristic cloverleaf secondary structure of tRNAs, and from the trinucleotide in the anticodon position in such structures, either by eye, or using the TRNA computer program of Staden (13). Nucleotide sequences were analyzed by the SEQ computer program (14). Polypeptide- or presumptive polypeptide-encoding genes were identified by comparing predicted amino acid sequences with corresponding amino acid sequences of previously identified genes of mouse mtDNA (8) using the TYPIN and SEARCH computer programs (15,16).

RESULTS AND DISCUSSION

Gene organization, and comparisons to mammals and other organisms. Using the strategies summarized in Fig. 1, the entire sequence of a continuous 2330 nucleotide section of the HindIII-B fragment and of the adjacent 430 nucleotide HindIII-E fragment of D. yakuba mtDNA was determined. Within this sequence (Fig. 2) are three open reading frames which, from comparisons of nucleotide and amino acid sequences to the corresponding sequences of previously identified genes of mouse mtDNA (8; Table 1 and Fig. 3), have been identified as the genes for cytochrome c oxidase subunits II and III (COII and COIII), and ATPase subunit 6 (ATPase6). A fourth open reading frame appears to correspond to URFA6L of mouse mtDNA (Fig. 3; see below). The sequence also contains four regions which can fold into the characteristic secondary structure of tRNAs with anticodons indicating them to be the genes for $tRNA^{leu}_{UUR}$, $tRNA^{lys}$, $tRNA^{asp}$ and $tRNA^{gly}$ (Fig. 4). The eight genes occur in the order $tRNA^{leu}_{UUR}$, COII, $tRNA^{lys}$, $tRNA^{asp}$, URFA6L, ATPase6, COIII and $tRNA^{gly}$,


tRNA leu UUR
COII
tRNA lys
tRNA asp
URFA6L
ATPase6

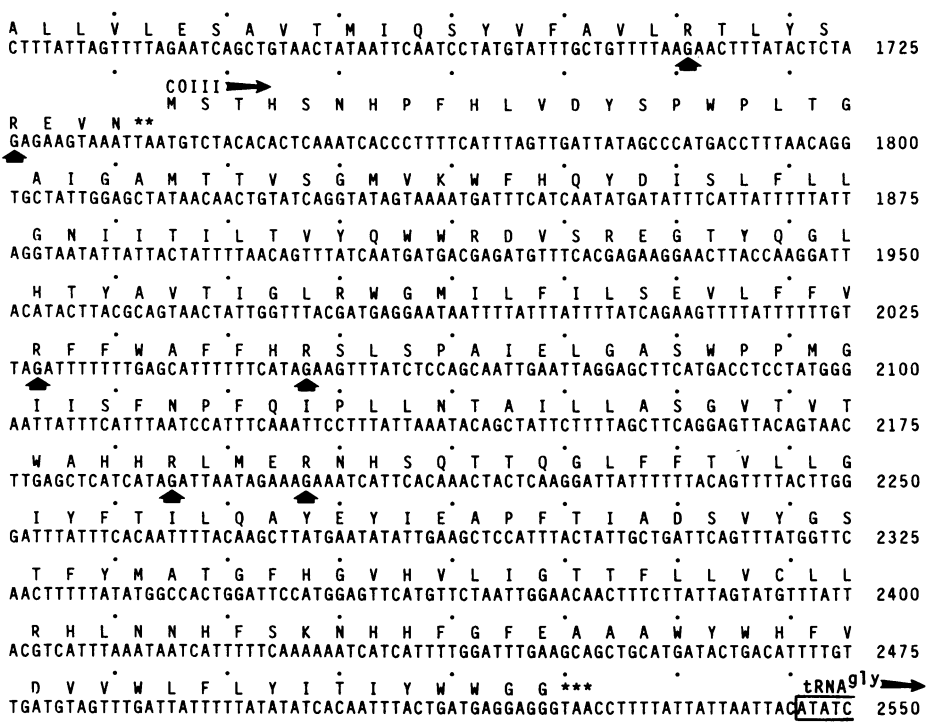
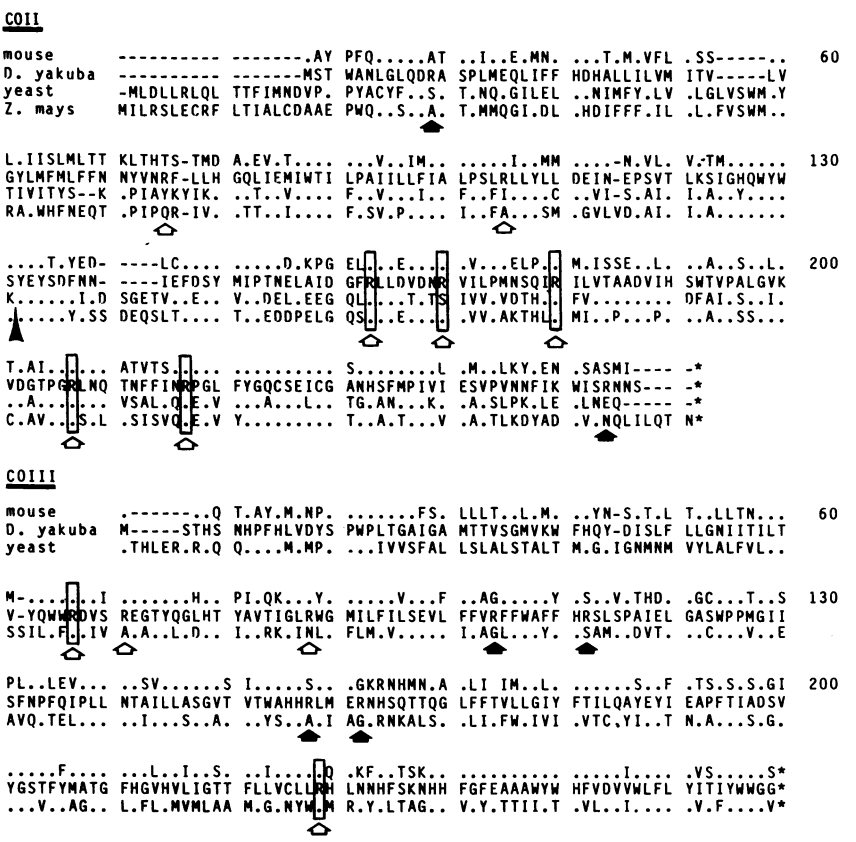


Figure 2. Nucleotide sequence of the segment of the D. yakuba mtDNA molecule identified in Fig. 1. From considerations of nucleotide and predicted amino acid sequence homologies to mouse mtDNA, this sequence contains the genes for cytochrome c oxidase subunits II and III (COII and COIII) and ATPase subunit 6 (ATPase6). Designation of the unidentified open reading frame as URFA6L is based on its amino acid sequence homology to URFA6L of mouse mtDNA (see text). The boxed nucleotide sequences fold into the characteristic cloverleaf structure of tRNAs (Fig. 4). The anticodons for $tRNA^{leu}_{UUR}$, $tRNA^{lys}$ and $tRNA^{asp}$ are underlined. The nucleotide sequence shown is the sense strand of all of the genes, and the arrows indicate the direction of transcription of each gene. Asterisks indicate partial or complete termination codons. The wide verticals arrows indicate the positions of AGA codons. The sequence to the left of the $tRNA^{leu}_{UUR}$ gene is presently unidentified. The sequence containing the carboxyl terminal 254 nucleotides of the COIII gene, and the entire $tRNA^{gly}$ gene has been published elsewhere (17).

and, beginning with the $tRNA^{leu}_{UUR}$ gene, are all transcribed in the same direction, that opposite to the direction in which replication proceeds around the molecule (Figs. 1 and 5).

The distribution of the COII, URFA6L, ATPase6 and COIII genes within the D. yakuba mtDNA molecule is similar to the distribution of these genes in mammalian mtDNA (6-8) in regard to their location relative to each other, to


COII
mouse ---------- --------.AY PFQ.....AT ..I..E.MN. ...T.M.VFL .SS-----.. 60
D. yakuba ---------- -------MST WANLGLQDRA SPLMEQLIFF HDHALLILVM ITV-----LV
yeast -MLDLLRLQL TTFIMNDVP. PYACYF..S. T.NQ.GILEL ..NIMFY.LV .LGLVSWM.Y
Z. mays MILRSLECRF LTIALCDAAE PWQ..S..A. T.MMQGI.DL .HDIFFF.IL .L.FVSWM..
L.IISLMLTT KLTHTS-TMD A.EV.T.... ...V..IM.. .....I..MM .....-N.VL. V.TM..... 130
GYLMFMLFFN NYVNRF-LLH GQLIEMIWTI LPAIILLFIA LPSLRLLYLL DEIN-EPSVT LKSIGHQWYW
TIVITYS--K .PIAYKYIK. ..T..V.... F..V...I.. F..FI....C ..VI-S.AI. I.A..Y....
RA.WHFNEQT .PIPQR-IV. .TT..I.... F.SV.P.... I..FA...SM .GVLVD.AI. I.A.......
....T.YED- ----LC.... .....D.KPG EL...E.... .V...ELP.. M.ISSE..L. ..A..S..L. 200
SYEYSDFNN- ----IEFDSY MIPTNELAID GFRLLDVDNR VILPMNSQIR ILVTAADVIH SWTVPALGVK
K......I.D SGETV..E.. V..DEL.EEG QL....T.TS IVV.VDTH.. FV..... DFAI.S..I.
.....Y.SS DEQSLT.... T..EDDPELG QS...E.... .VV.AKTHL. MI..P....P. ..A..SS...
T.AI.....A ATVTS.....T ..... S.....L .M..LKY.EN .SASMI---- -*
VDGTPGRLNQ TNFFINRPGL FYGQCSEICG ANHSFMPIVI ESVPVNNFIK WISRNNS--- -*
..A..... VSAL.Q.E.V ...A...L.. TG.AN...K. .A.SLPK.LE .LNEQ----- -*
C.AV...S.L .SISVQ.E.V Y..... T..A.T...V .A.TLKDYAD .V.NQLILQT N*
COIII
mouse .--------Q T.AY.M.NP. .....FS. LLLT..L.M. ..YN-S.T.L T..LLTN... 60
D. yakuba M-----STHS NHPFHLVDYS PWPLTGAIGA MTTVSGMVKW FHQY-DISLF LLGNIITILT
yeast .THLER.R.Q Q....M.MP. ...IVVSFAL LSLALSTALT M.G.IGNMNM VYLALFVL..
M-....I..I .....H.. PI.QK...Y. .....V...F ..AG.....Y .S..V.THD. .GC...T..S 130
V-YQWWRDVS REGTYQGLHT YAVTIGLRWG MILFILSEVL FFVRFFWAFF HRSLSPAIEL GASWPPMGII
SSIL.F..IV A.A..L.D.. I..RK.INL. FLM.V..... I.AGL...Y. .SAM..DVT. ..C...V..E
PL..LEV... ..SV......S I.....S.. .GKRNHMN.A .LI IM..L. ......S..F .TS.S.S.GI 200
SFNPFQIPLL NTAILLASGVT VTWAHHRLM ERNHSQTTQG LFFTVLLGIY FTILQAYEYI EAPFTIADSV
AVQ.TEL... ..I...S..A. ..YS..A.I AG.RNKALS. .LI.FW.IVI .VTC.YI..T N.A...S.G.
.....F.... ...L..I..S. ..I....Q .KF..TSK.. ....I.... .VS.....S*
YGSTFYMATG FHGVHVLIGTT FLLVCLLRH LNNHFSKNHH FGFEAAAWYW HFVDVVWLFL YITIYWWGG*
...V..AG.. L.FL.MVMLAA M.G.NYW.M R.Y.LTAG.. V.Y.TTII.T .VL..I.... .V.F....V*
ATPase6
mouse .NE..--.-- AS----.ITP TMMGFPIVVA ---------- -IIMFPSILF ..SKR-.I-N 60
D. yakuba MMTNL--F-- SV----FDPS AIFNLSLNWL ---------- -RTFLGLLMI PSIYW-LM-P
yeast .-F..LNTYI TSPLDQ.EIR TL.G.QSSFI DLSCLNLTTF SLYTIIV.LV ITSLYT.TNN
N.LHS.QHWL VK-----LII .Q--MMLIHT .K.RTW-.LM IV..IMF.GS T.LL..L.HT ..P.TQ.SMN 130
SRYNIFWNSI LL-----TLH KE--FKTLLG PSGHNGSTFI FISLFSLILF NNFMGLFPYI FTRTSHLTLT
NNKI.GSRWL ISQEAIYDTI INMTKGQIG. KNWGLY-FPM IFT..MF.FI A.LISMI..S .ALSA..VFI
..M.I...AG AVIT.FRHKL KSSL..FL.. ...IS.I.ML II....SLF. Q.MA.....T .IT.....M 200
LSLALPLWLC FMLYGWINHT QHMFAHLVPQ GTPAILMPFM VCIETIRNII RPGTLAVRLT ANMIAGHLLL
I..SIVI..G NTIL.LYK.G WVF.SLF..A ...LP.V.LL .IM.TLSY.A .AIS.GL..G S.IL.....M
H.I.GA.LVL .NISPP.ATI TFI.L..LT- --I..F..AL ..A...TL.V S..LHDNT-- -*
TLLGN-TGPS MSYLLVTFLL VAQIALL--- --VLESAVTM IQSYVFAVLR TLYSREVN-- -*
VI.AGL.FNF .LIN.F.LVF GFVPLAMILA IMI..F.IGI .....WTI.T AS.LKDTLYL H*
URFA6L
mouse M..LDTST.F ITI.SSM..L ....QLKVSS QTF.LAP... .LTTMKVKTP ---.EL..TK 60
D. yakuba IPQMAPIRWL LLFIVFSIT- FILFCSINYY SYMPT--SPK SNELKNINLN SMNW--KW--
IYLPHSLPQQ*
----------*

the presence of intervening tRNA genes and to the relationships of amino terminal and carboxyl terminal regions of genes not separated by tRNA genes (Fig. 5). However there are some differences as to which tRNA genes occur between the polypeptide coding genes, and in the number of noncoding (spacer) nucleotides separating some genes.

In D. yakuba mtDNA the $tRNA^{leu}_{UUR}$ gene precedes the COII gene, rather than the $tRNA^{asp}$ gene found in this position in mouse mtDNA. In mouse mtDNA the $tRNA^{leu}_{UUR}$ gene is located between the large (16S) rRNA and URF1 genes. Consistent with the present observation, we have shown that a tRNA gene does not occur between the large rRNA and URF1 genes of D. yakuba mtDNA (5; Wahleithner, Clary and Wolstenholme, unpublished, see Fig. 5).

From considerations of both nucleotide (Fig. 2) and amino acid sequence comparisons (Fig. 3), it appears that internal insertion/deletions between the COII genes of D. yakuba and mouse have not occurred. The D. yakuba COII gene extends one sense codon beyond the mouse COII gene and is separated by a single T from the 5' terminal nucleotide of a $tRNA^{lys}$ gene. The single T following the COII gene could be completed as a termination codon (UAA) by polyadenylation of the transcript following excision from a multicistronic primary transcript, as has been suggested for transcripts of some mammalian mtDNA genes (6,25). The $tRNA^{lys}$ gene is separated from a $tRNA^{asp}$ gene by four apparently non-coding nucleotides (Fig. 2).

Figure 3. A comparison of the amino acid sequences of cytochrome c oxidase subunits II and III (COII and COIII), ATPase subunit 6 (ATPase6), and URFA6L predicted from the nucleotide sequences of the respective genes of mtDNAs of D. yakuba, with the corresponding amino acid sequences of mouse (8), yeast (18-21) and Zea mays (22). 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 partial or a complete termination codon. Wide vertical solid arrows indicate D. yakuba amino acids specified by AGA (tentatively shown as arginine). Wide vertical open arrows indicate D. yakuba amino acids (arginine) specified by a CGN codon. Boxes indicate positions in the sequences where arginines are conserved in all four species, or in D. yakuba and yeast (position 181, ATPase6), or in D. yakuba, mouse and Z. mays (position 170, COII). In all of these cases the arginine is specified by an AGA codon in yeast, by a CGN codon in D. yakuba and mouse, and by either a CGA, CGT or AGA (position 170) codon in Z. mays. Alignment of COII amino acid sequences follows the comparisons of the COII amino acid sequences of Z. mays, yeast (each predicted from nucleotide sequences (18,19,22)) and bovine (23) given in reference 22. The solid thin arrow head indicates the position of the 794 nucleotide intron in the Z. mays COII gene. Alignment of ATPase6 amino acid sequences follows the comparisons of ATPase6 sequences of yeast, human and Aspergillus nidulans (all predicted from nucleotide sequences) given in reference 24. Alignment of URFA6L amino acid sequences is based on three blocks of apparently homologous sequences, positions 21-24, 38-41 and the terminal lysine and tryptophan residues.

Table 1. Nucleotide and amino acid sequence comparisons of the genes for cytochrome c oxidase subunits II and III (COII, COIII) and ATPase subunit 6 (ATPase6) of D. yakuba with the corresponding genes of mouse, yeast and Zea mays.

Gene	Number of nucleotides[a] (D. yakuba)	% Nucleotide sequence homology[a,b] (D. yakuba/mouse)	% Silent nucleotide substitutions[a,b] (D. yakuba/mouse)	% Amino acid sequence homology[b]: D. yakuba/ mouse	D. yakuba/ yeast	D. yakuba/ Z. mays
COIII	786	66.5	47.2	64.1	42.8	-
COII	684	63.9	41.8	56.1	40.6	40.3
ATPase6	672	49.1	31.4	35.7	23.0	-

[a]Excluding nucleotides concerned with termination.
[b]These values include all deduced insertion/deletions (See Fig. 3). The data for mouse, yeast and Z. mays mt-DNAs are taken from references (8,18-22).

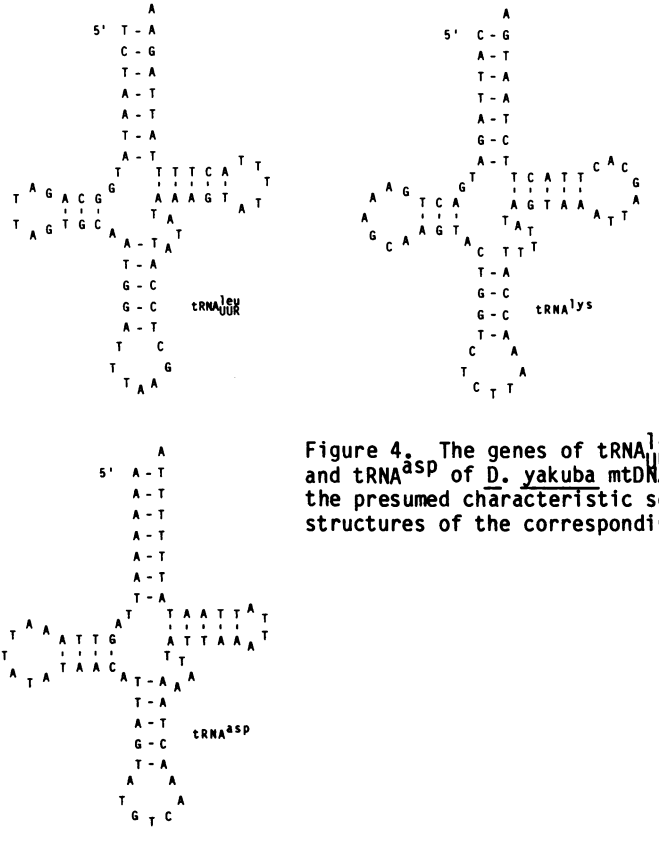


Figure 4. The genes of $tRNA^{leu}_{UUR}$, $tRNA^{lys}$ and $tRNA^{asp}$ of D. yakuba mtDNA shown in the presumed characteristic secondary structures of the corresponding tRNAs.

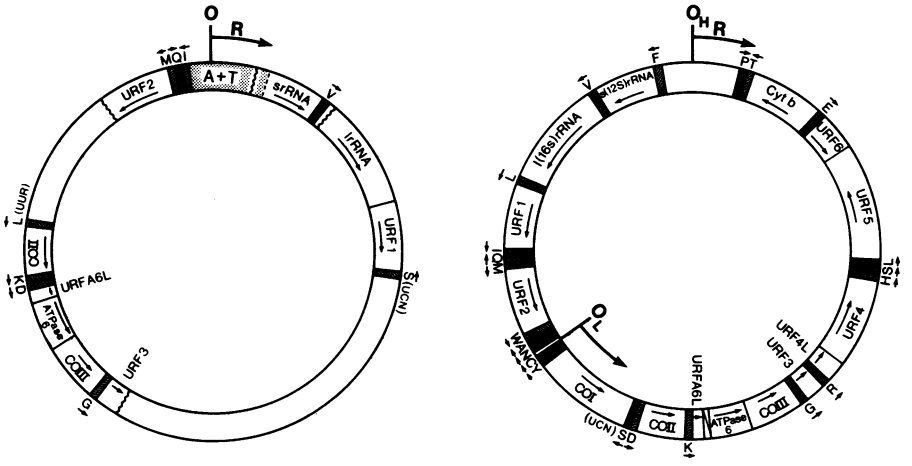


Figure 5. A comparison of gene order in the circular mtDNA molecule of D. yakuba (left) and mammal (right). The D. yakuba molecule is derived from the results of various studies (for references see (5)) and the present data. The mammalian mtDNA molecule is derived from that given for mouse (8). Each tRNA gene (hatched areas) is identified by the one letter amino acid code. Arrows within and outside the molecules indicate the direction of transcription of each gene. Wavy lines in the D. yakuba molecule indicate uncertain gene termini. O and R indicate the origin and direction of replication, respectively, in the D. yakuba molecule. O_H and O_L indicate the origins of heavy and light strand DNA synthesis in the mammalian molecule.

In the D. yakuba sequence (Fig. 2) the $tRNA^{asp}$ gene is followed by an open reading frame, the predicted amino acid sequence of which is 30% homologous to the major portion of the amino acid sequence predicted from URFA6L of mouse mtDNA. This homology value is dependent upon the alignment of three segments of amino acids, two internal and one at the carboxyl terminus of the D. yakuba gene, which necessitates the assumption that insertion/deletions have occurred between the two sequences (Fig. 3). It has been noted by others (7,8) that this is the least conserved open reading frame among mammalian mtDNAs.

In mouse mtDNA, only the gene for $tRNA^{lys}$ separates the COII and URFA6L genes. In both D. yakuba and mouse mtDNA the URFA6L gene begins with the first triplet following the tRNA gene which precedes it. In D. yakuba mtDNA the carboxyl terminal region of URFA6L overlaps the amino terminal region of the ATPase6 gene by seven nucleotides, but these two genes are translated in different reading frames. A similar situation is found in mouse mtDNA except that in that case the URFA6L gene overlaps the ATPase6 gene by 43 nucleotides

which apparently code for 12 carboxyl terminal amino acids for which the D. yakuba URFA6L gene product has no equivalent.

Alignment of the nucleotide and amino acid sequences (Fig. 3) of the ATPase6 genes of D. yakuba and mouse mtDNAs indicates that insertion/deletions involving four internal codons (12 nucleotides) have occurred between these genes. In D. yakuba, as in mouse and other mammals (6,8), the ATPase6 gene ends with TA, the latter nucleotide being adjacent to the initiation codon of the COIII gene. While there is evidence that in human mtDNA there is precise cleavage of the primary transcript between the terminal UA of URFA6L and the AUG initiation codon of COIII, the recognition signal by which this is facilitated is unknown (6,25).

Only a single codon insertion/deletion is found between the D. yakuba and mouse COIII genes. In both species the COIII gene is followed by the $tRNA^{gly}$ gene. However, as noted previously (17), the D. yakuba COIII gene terminates with a TAA codon which is separated by 18 apparently noncoding nucleotides from the $tRNA^{gly}$ gene, whereas the mouse COIII gene terminates with a single T which is immediately adjacent to the $tRNA^{gly}$ gene.

Comparisons of nucleotide and amino acid sequences of the COII, COIII and ATPase6 genes of D. yakuba with those of the corresponding genes of mouse are shown in Table 1 and Fig. 3. The nucleotide sequence homologies of the COII and COIII genes between D. yakuba and mouse are similar, 66.5% and 63.9% respectively, while conservation of the ATPase6 gene in the two species is considerably less, 49.1%. The amino acid sequence homology for each gene comparison between D. yakuba and mouse is lower than the corresponding nucleotide sequence homology, and again is highest for the COIII gene (64.1%) and lowest for the ATPase6 gene (31.4%). The percentage of nucleotide substitutions which would not result in a corresponding amino acid substitution shows a positive correlation with the nucleotide and amino acid sequence homologies of the three genes. Amino acid sequence homologies of the COII, COIII and ATPase6 genes between D. yakuba and yeast mtDNAs are all lower than the corresponding values for comparisons of D. yakuba and mouse. However, as in the latter comparisons, the COIII genes have the highest homology and the ATPase6 genes have the least. Homology between the D. yakuba and Zea mays COII genes is similar to the homology between the D. yakuba and yeast COII genes.

In D. yakuba mtDNA as in mouse (8) and other mammalian mtDNAs (6,7), the initiation codon of COII, COIII and ATPase6 is ATG. However, while in mammalian mtDNA the initiation codon of URFA6L is also ATG, in D. yakuba mtDNA

ATT appears to serve this function. ATT, which as an initiation codon might specify methionine in mammalian mtDNAs (8), has also been interpreted as the initiation codon of URF1 and URF2 of D. yakuba mtDNA (5).

The relative locations in the D. yakuba mtDNA molecule of all of the genes determined to date, and the direction in which each gene is transcribed are shown in Fig. 5. In mouse mtDNA, the templates for all transcripts except those for eight tRNA genes and URF6 are contained in one strand (the H strand) of the molecule. In contrast, of the genes which have been mapped to date in the D. yakuba mtDNA molecule, all of those to the left of the A+T-rich region (as shown in Fig. 5), except for one tRNA gene, are transcribed from one strand of the molecule, while all of those to the right of the A+T-rich region, again with the exception of one tRNA gene, are transcribed from the other strand.

Transfer RNA genes. The $tRNA^{leu}_{UUR}$, $tRNA^{lys}$ and $tRNA^{asp}$ genes of D. yakuba mtDNA are 41%, 41% and 70% homologous, respectively, to the corresponding tRNA genes of mouse mtDNA (8). These three D. yakuba tRNA genes (Fig. 4) show the major secondary structural characteristics of the six other D. yakuba mt-tRNA genes we have described previously (including $tRNA^{gly}$; 5,17). The number of nucleotide pairs found in the amino-acyl and anticodon stems, and the number of nucleotides in the anticodon loop are constant and the same as those found in mammalian mt-tRNAs (8,26) and prokaryotic and non-organelle eukaryotic tRNAs (27,28). The numbers of nucleotides found in the remaining stems and loops are within the limits of variability found for other mt-tRNAs (5,8,26). Also, as found for other D. yakuba tRNA genes (5,17), among these three tRNA genes only eleven ($tRNA^{leu}_{UUR}$), fourteen ($tRNA^{lys}$) and eleven ($tRNA^{asp}$) of the 18 nucleotides which are constant in prokaryotic and non-organelle eukaryotic tRNAs (27), are present. The common occurrence among D. yakuba tRNA genes of the constant Pu_{26}, T_{33} and Pu_{37} nucleotides (numbering system in (27)) is maintained in all three of the $tRNA^{leu}_{UUR}$, $tRNA^{asp}$ and $tRNA^{lys}$ genes. Among the $tRNA^{leu}_{UUR}$, $tRNA^{lys}$ and $tRNA^{asp}$ genes only a single mismatched nucleotide pair is present in the stem regions ($tRNA^{lys}$, Fig. 3). This is consistent with the previous observation of a low occurrence of mismatches in the stems of D. yakuba mt-tRNA genes (17).

The CUU anticodon of the $tRNA^{lys}$ transcribed from D. yakuba mtDNA would be expected to recognize the codon AAG which is utilized in D. yakuba mtDNA (Table 2). However, the codon AAA, also expected to specify lysine, is used more frequently than AAG. Although nuclear $tRNA^{lys}$ genes with a CTT anticodon have been reported for a number of organisms (27-28) including Drosophila

(29), a tRNA gene with this anticodon has not been found in mammalian or fungal mtDNAs (6-8,24,30,31). Mammalian and fungal mtDNAs each contain a single $tRNA^{lys}$ gene and the UUU anticodon of the tRNA transcribed from it can recognize both AAA and AAG codons (6-8,30,31). If there is only a single $tRNA^{lys}$ gene in D. yakuba mtDNA, then the CUU anticodon of the corresponding $tRNA^{lys}$ must also be able to recognize both codons specifying lysine. Among mammalian, Drosophila and fungal mtDNAs the only other tRNA gene known which contains an anticodon with a C in the 5' (wobble) position is the $tRNA^{f\text{-}met}$ gene (5-8,17,24,30,31). It appears that in mammalian mitochondria the CAU anticodon of $tRNA^{f\text{-}met}$ can recognize both AUG and AUA as methionine specifying codons when they occur internally, and can recognize all four of the AUN codon family as specifying methionine when they occur as initiation codons (6-8). It has been suggested that the C residue of the anticodon must be modified to permit it to read all four initiation codons (8).

The $tRNA^{asp}$ gene has the lowest G+C content (8.8%) of any of the D. yakuba mt-tRNA genes described to date. Both the amino-acyl stem and the TψC stem contain only A-T pairs, a situation which among Drosophila mt-tRNA genes so far described is otherwise limited to the amino-acyl stem of the $tRNA^{gln}$ gene (5,17). The amino-acyl stems of the $tRNA^{asp}$ and $tRNA^{gln}$ genes contain runs of 5 and 6 As (and Ts), respectively, in one strand, the stacking effect of which would be expected to add stability to the secondary structure of this region (7).

We have recently reported (17) that the nucleotide sequence 5'TTTATTAT which occurs in the apparently noncoding region between the terminal codon of the COIII gene and the 5' terminal nucleotide of the $tRNA^{gly}$ gene (nucleotides 2531-2538, Fig. 2), or a sequence different from 5'TTTATTAT by one nucleotide substitution, also occurs in the 5' flanking, noncoding regions of four other tRNA genes. In three of these cases (including $tRNA^{gly}$, nucleotides 2514-2519, Fig. 2) the sequence 5'GATGAG is found upstream from the A+T-octanucleotide sequence. Neither of these related sequences is found close to the 5' terminus of the $tRNA^{lys}$ or $tRNA^{asp}$ genes (Fig. 2) which is consistent with the occurrence of zero ($tRNA^{lys}$) and four ($tRNA^{asp}$) noncoding nucleotides 5' to these tRNA genes (17). Furthermore, neither of the related sequences are found within 50 nucleotides of the unidentified sequence upstream from the 5' terminus of the $tRNA^{leu}_{UUR}$ gene (sequence not shown).

Codon usage and the genetic code. Codon usage among the COII, COIII, ATPase6 and URFA6L genes of D. yakuba mtDNA is summarized in Table 2. The most striking observation is that 94.4% of all codons found in these genes end

Table 2. Codon usage in the genes for cytochrome c oxidase subunits II and III (COII and COIII), ATPase subunit 6 (ATPase6) and URFA6L in mtDNA of D. yakuba.

Phe-TTT	58	Ser-TCT	19	Tyr-TAT	24	Cys-TGT	5
TTC	3	TCC	1	TAC	7	TGC	1
Leu-TTA	94	TCA	27	TER-TAA	3	Trp-TGA	27
TTG	1	TCG	0	TAG	0	TGG	0
Leu-CTT	10	Pro-CCT	21	His-CAT	26	Arg-CGT	2
CTC	1	CCC	0	CAC	3	CGC	0
CTA	4	CCA	13	Gln-CAA	19	CGA	12
CTG	0	CCG	1	CAG	0	CGG	0
Ile-ATT	68	Thr-ACT	25	Asn-AAT	43	Ser-AGT	6
ATC	1	ACC	1	AAC	3	AGC	1
Met-ATA	23	ACA	26	Lys-AAA	9	(Arg)-AGA	12
ATG	8	ACG	0	AAG	1	AGG	0
Val-GTT	19	Ala-GCT	24	Asp-GAT	13	Gly-GGT	12
GTC	1	GCC	2	GAC	3	GGC	0
GTA	25	GCA	11	Glu-GAA	21	GGA	26
GTG	0	GCG	0	GAG	0	GGG	4

Codon ending in:	D. yakuba (n=782)	Mouse (n=784)	yeast (n=779)
T	48.6%	26.5%	48.3%
C	3.6%	24.0%	4.6%
A	45.8%	46.8%	42.8%
G	2.0%	2.7%	4.4%
A or T	94.4%	73.3%	91.1%

As is the case in mammalian mtDNA, ATA is assumed to specify methionine and TGA is assumed to specify tryptophan. AGA, which in mouse mtDNA is absent, and is used only as a termination codon in human mtDNA (COI) and bovine mtDNA (cytochrome b), is used to specify an amino acid in D. yakuba mtDNA (5), and is tentatively shown as arginine (but see text). AGG has been interpreted as the termination codon of URF6 in human mtDNA, but has not been found in other mammalian mtDNAs (6,8) or Drosophila mtDNAs (5,17). TAG has been interpreted as the termination codon for the COIII gene of D. melanogaster mtDNA (17). In the lower portion of the table, the frequencies of nucleotides in the third position of codons of COII, COIII, ATPase6 and URFA6L genes for D. yakuba and mouse (8), and of COII, COIII and ATPase6 genes of yeast (18-21) are compared.

in A or T. However, only 13 expected sense codons and TAG are not used at all, and each of these sense codons contains only a single A or T. Of these, four are used at least once in the segments of the URF1 and URF2 genes which we have reported (5,17). Further, TAG is used as the termination codon of the COIII gene of D. melanogaster mtDNA (17). From these observations, there seems no reason to expect that any one codon is excluded from use in Drosophila mtDNA.

Table 3. Nucleotide composition (as percentages) of the sense strands of the genes for cytochrome c oxidase subunits II and III (COII and COIII), ATPase subunit 6 (ATPase6) and URFA6L of D. yakuba mtDNA. Also given are the mean nucleotide compositions of the sense strands of the COII, COIII, ATPase6 and URFA6L genes of mouse mtDNA (8) and of the COII, COIII and ATPase6 genes of yeast mtDNA (18-21).

	COII (n=685)	URFA6L (n=162)	ATPase6 (n=675)	COIII (n=789)	D. yakuba mean (n=2346)	Mouse mean (n=2355)	Yeast mean (n=2346)
Thymine	40.1	43.8	44.0	40.4	41.6	29.6	41.7
Cytosine	12.4	12.3	14.1	14.2	13.5	25.6	12.1
Adenine	33.7	38.9	31.9	30.8	32.5	33.1	31.9
Guanine	13.7	4.9	10.1	14.6	12.3	11.7	14.3

The extremely infrequent use of C and G nucleotides in the third position of codons of the four D. yakuba genes corresponds to the very low content of C and G nucleotides in the sense strands of these genes (Table 3). A similar situation occurs in yeast mtDNA (Tables 2 and 3). Differences in frequencies of nucleotides in the third position of codons between the COII, COIII, ATPase6 and URFA6L genes of D. yakuba and mouse (Table 2) reflect differences in average base composition of the sense strands of these genes within the two mtDNAs (Table 3).

The predictions of amino acid sequences from the four polypeptide encoding genes shown in Fig. 2 assume that ATA specifies methionine rather than isoleucine, as is the case in the mammalian mitochondrial genetic code (32,33). We have shown previously (5) that the reading of URF1 and URF2 sequences of D. yakuba mtDNA depends upon the assumption that TGA and AGA each specify an amino acid. In the D. yakuba genes for COII, COIII, ATPase6 and URFA6L, TGA occurs as a sense codon a total of 27 times (Fig. 2). Twenty-one of these TGA codons correspond in position to TGA codons in the corresponding four genes of mouse (Fig. 3; 8) indicating that in D. yakuba mtDNA as in mammalian and fungal mtDNAs (8,24,30,32,34), TGA specifies tryptophan.

In mammalian mtDNAs only CGN codons specify arginine. Inframe AGA triplets do not occur, except as rare termination codons in human and bovine mtDNAs, and a gene for a tRNA which might be expected to recognize AGA has not been located (6-8). In yeast mtDNA, AGA is the only triplet which specifies arginine in the genes for the three cytochrome c oxidase subunits, ATPase6 and cytochrome b. However, in unidentified reading frames, as well as in the var1 gene of yeast mtDNA, both AGA and CGN codons are utilized (35-37). In the COII gene of Zea mays, CGA, CGY and AGR codons have all been interpreted as

specifying arginine (22). As twelve of the 14 CGN codons found in D. yakuba genes for COII, COIII, ATPase6 and URFA6L correspond in position to CGN codons in mouse mtDNA (Fig. 3) it seems reasonable to conclude that CGN codons of D. yakuba mtDNA also specify arginine. In contrast, none of the twelve AGA codons found in these four D. yakuba genes correspond in position to arginine-specifying codons (CGN) in mouse mtDNA. This was also found to be the case for the three AGA codons present in D. yakuba URF1 and URF2 (5). The 15 AGA codons of D. yakuba mtDNA found to date correspond to codons specifying nine different amino acids in mouse mtDNA; serine (five), glycine (two), isoleucine (two), alanine, leucine, valine, threonine, histidine and proline.

Comparisons of the amino acid sequences of the COII, COIII and ATPase6 genes of D. yakuba with those of yeast indicate again that none of the AGA codons in the D. yakuba genes correspond to AGA codons in yeast mtDNA (Fig. 4). However, within these three genes a total of seven of the D. yakuba CGN codons do correspond to AGA codons in yeast as well as CGN codons in mouse mtDNA. Five of the CGN codons in the D. yakuba COII gene also correspond to five arginine-specifying codons (three CGT, one CGA and one AGA) in the Z. mays COII gene. It is clearly not possible from these observations to conclude which amino acid is specified by AGA in the D. yakuba mitochondrial genetic code. If, in fact, AGA specifies arginine in D. yakuba, then this codon usage is indicated to have evolved separately from the use of AGA to specify arginine by yeast (or plant) mtDNA.

The codon AGA requires the presence of a tRNA, presumably with the anti-codon UCU, to decode it (5). From the data presented it appears that AGA codons in D. yakuba mtDNA do not correspond in position to arginine-specifying codons in mouse, yeast or plant mtDNAs, but do most frequently correspond in position to serine-specifying codons in mouse mtDNA. In view of these observations and the fact that AGY codons specify serine in all genetic codes known (Table 2), it seems reasonable to suggest that in the D. yakuba mitochondrial genetic code, AGA may also specify serine. Because in mammalian and presumably D. yakuba mitochondria a single tRNA with a U in the wobble position of the anticodon can apparently decode all four codons of a single family (Table 2), it seems plausible that all AGN codons of D. yakuba mtDNA could be decoded by a tRNA with a UCU anticodon. Such a tRNA might occur in place of the tRNA of mammalian mitochondria which has a GCU anticodon and decodes only AGY codons.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health Grants Nos. GM 18375 and RR 07092.

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. Fauron,C.M.-R. and Wolstenholme,D.R. (1976) Proc. Natl. Acad. Sci. USA 73, 3623-3627.
3. Fauron,C.M.-R. and Wolstenholme,D.R. (1980) Nucl. Acids. Res. 8, 2439-2452.
4. Goddard,J.M. and Wolstenholme,D.R. (1978) Proc. Natl. Acad. Sci. USA 75, 3886-3890.
5. Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M.-R. and Wolstenholme, D.R. (1982) Nucl. Acids Res. 10, 6619-6637.
6. 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.
7. 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.
8. Bibb,M.J., Van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) Cell 26, 167-180.
9. Holmes,D.S. and Quigley,M. (1981) J. Biochem. 114, 193-197.
10. Hong,G.F. (1982) J. Mol. Biol. 158, 539-549.
11. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
12. Staden,R. (1982) Nucl. Acids Res. 10, 4731-4751.
13. Staden,R. (1980) Nucl. Acids Res. 8, 817-825.
14. Brutlag,D.L., Clayton,J., Friedland,P. and Kedes,L. (1982) Nucl. Acids Res. 10, 279-294.
15. Jue,R.A., Woodbury,N.W. and Doolittle,R.F. (1980) J. Mol. Evol. 15, 129-148.
16. Doolittle,R.F. (1981) Science 214, 149-159.
17. Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R. (1983) Nucl. Acids Res. 11, 2411-2425.
18. Coruzzi,G. and Tzagoloff,A. (1979) J. Biol. Chem. 254, 9324-9330.
19. Fox,T.D. (1979) Proc. Natl. Acad. Sci. USA 76, 6534-6538.
20. Thalenfeld,B.E. and Tzagoloff,A. (1980) J. Biol. Chem. 255, 6173-6180.
21. Macino,G. and Tzagoloff,A. (1980) Cell 20, 507-517.
22. Fox,T.D. and Leaver,C.J. (1981) Cell 26, 315-323.
23. Steffens,G.J. and Buse,G. (1979) Hoppe-Seyler's Z. Physiol. Chem. 360, 613-619.
24. Netzker,R., Kochel,H.G., Basak,N. and Kunzel,H. (1982) Nucl. Acids Res. 10, 4783-4794.
25. Ojala,D., Montoya,J. and Attardi,G. (1981) Nature 290, 470-474.
26. 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.
27. Sprinzl,M. and Gaus,D.H. (1983) Nucl. Acids Res. 11, r1-r53.
28. Sprinzl,M. and Gaus,D.H. (1983) Nucl. Acids Res. 11, r55-r103.
29. DeFranco,D., Schmidt,O. and Soll,D. (1980) Proc. Natl. Acad. Sci. USA 77, 3365-3368.

30. Bonitz,S.G., Berlani,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.
31. Kochel,H., Lazarus,C.M., Basak,N. and Kunzel,H. (1981) Cell 23, 625-633.
32. Barrell,B.G., Bankier,A.T. and Drouin,J. (1979) Nature 282, 189-194.
33. 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. and Young,I.G. (1980) Proc. Natl. Acad. Sci. USA 77, 3164-3166.
34. Heckman,J.E., Sarnoff,J., Alzner-DeWeerd,B., Yin,S. and RajBhandary,U.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3159-3163.
35. Bonitz,S.G., Coruzzi,G., Thalenfeld,B.E., Tzagoloff,A. and Macino,G. (1980) J. Biol. Chem. 255, 11927-11941.
36. Coruzzi,G., Bonitz,S.G., Thalenfeld,B.E. and Tzagoloff,A. (1981) J. Biol. Chem. 256, 12780-12787.
37. Hudspath,M.E.S., Ainley,W.M., Schumard,D.S., Butow,R.A. and Grossman,L.I. (1982) Cell 30, 617-626.