J Mol Evol (1985) 22:252–271

Journal of
Molecular Evolution

The Mitochondrial DNA Molecule of *Drosophila yakuba*: Nucleotide Sequence, Gene Organization, and Genetic Code

Douglas O. Clary* and David R. Wolstenholme

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

Summary. The sequence of the 16,019 nucleotide-pair mitochondrial DNA (mtDNA) molecule of *Drosophila yakuba* is presented. This molecule contains the genes for two rRNAs, 22 tRNAs, six identified proteins [cytochrome b, cytochrome c oxidase subunits I, II, and III (COI–III), and ATPase subunits 6 and 8] and seven presumptive proteins (URF1–6 and URF4L). Replication originates within a region of 1077 nucleotides that is 92.8% A + T and lacks any open reading frame larger than 123 nucleotides. An equivalent to the sequence found in all mammalian mtDNAs that is associated with initiation of second-strand DNA synthesis is not present in *D. yakuba* mtDNA. Introns are absent from *D. yakuba* mitochondrial genes and there are few (0–31) intergenic nucleotides. The genes found in *D. yakuba* and mammalian mtDNAs are the same, but there are differences in their arrangement and in the relative proportions of the complementary strands of the molecule that serve as templates for transcription. Although the *D. yakuba* small and large mitochondrial rRNA genes are exceptionally low in G and C and are shorter than any other metazoan rRNA genes reported, they can be folded into secondary structures remarkably similar to the secondary structures proposed for mammalian mitochondrial rRNAs. *D. yakuba* mitochondrial tRNA genes, like their mammalian counterparts, are more variable in sequence than nonorganelle tRNAs. In mitochrondrial protein genes ATG, ATT, ATA, and in one case (COI) ATAA appear to be used as translation initiation codons. The only termination codon found in these genes is TAA. In the *D. yakuba* mitochondrial genetic code, AGA, ATA, and TGA specify serine, isoleucine, and tryptophan, respectively. Fifty-nine types of sense codon are used in the *D. yakuba* mitochondrial protein genes, but 93.8% of all codons end in A or T. Codon–anticodon interactions may include both G–A and C–A pairing in the wobble position. Evidence is summarized that supports the hypothesis that A and T nucleotides are favored at all locations in the *D. yakuba* mtDNA molecule where these nucleotides are compatible with function.

Key words: *Drosophila* — Mitochondrial DNA — Mitochondrial genes — Nucleotide sequence — Gene arrangement — Genetic code — Codon–anticodon interaction — Ribosomal RNA genes — Transfer RNA genes

Introduction

All metazoa, from platyhelminthes to mammals, possess a mitochondrial genome that consists of a single circular molecule ranging in size from 14.5 to 19.5 kb (Altman and Katz 1976). Complete nucleotide sequences and gene contents have been determined for the mitochondrial DNA (mtDNA) molecules of human (Crews and Attardi 1980; Anderson et al. 1981; Ojala et al. 1981; Montoya et al. 1981), mouse (Van Etten et al. 1980, 1982; Bibb et al. 1981), and cow (Anderson et al. 1982b). Also various portions of the sequences of mtDNA molecules of other mammalian species have been determined (Grosskopf and Feldmann 1981; Brown et al. 1982; Pepe et al. 1983; Taira et al. 1983). All mammalian mtDNA molecules contain the genes for two rRNAs and 22 tRNAs of the mitochondri-

Offprint requests to: D.R. Wolstenholme
**Present address:* Department of Biochemistry, Stanford School of Medicine, Stanford, California 94305, USA

on's own protein-synthesizing system, and for six identified proteins [cytochrome b, cytochrome c oxidase subunits I, II, and III, and ATPase subunits 6 and 8 (formerly URFA6L)]. In addition, there are seven unidentified open reading frames (URF) that also appear to code for proteins (Chomyn et al. 1983; Mariottini et al. 1983; Michael et al. 1984). The arrangement of genes in mammalian mtDNAs is totally conserved. Introns have not been identified in any of the genes and there are few or no nucleotides separating individual genes. The only exception to this is a region of approximately 1 kb that lies between the $tRNA^{phe}$ and $tRNA^{pro}$ genes and contains the replication origin.

The relative locations of the rRNA genes, tRNA genes, replication origin, and a number of polyadenylated RNAs have been mapped on the mtDNA molecule of the amphibian *Xenopus laevis* (Ohi et al. 1978; Ramirez and Dawid 1978; Rastl and Dawid 1978), and nucleotide sequences of some segments of the molecule have been determined (Wong et al. 1983). The gene content and arrangement appear to be the same as in mammalian mtDNA.

In this paper, we present the nucleotide sequence of the mtDNA molecule of *Drosophila yakuba*. This is the first invertebrate mtDNA molecule to be sequenced completely. Detailed descriptions and discussions of the various sections of the *D. yakuba* mtDNA molecule and some discussion of the complete sequence have been presented elsewhere (Clary et al. 1982, 1983, 1984; Clary and Wolstenholme 1983a,b, 1984a,b, 1985). Here we summarize what we have learned from analysis of the whole sequence regarding gene structure, arrangement, and transcription; novel features of the genetic code; and codon–anticodon interactions.

Materials and Methods

The *D. yakuba* strain (2371.6, Ivory Coast) used in this study was originally obtained from the species stock collection of the Genetic Foundation, University of Texas at Austin. Mitochondrial DNA was obtained by cesium chloride–ethidium bromide centrifugation of sodium dodecyl sulfate (SDS) lysates of ovaries obtained by hand dissection of yeast-fed *D. yakuba*, as described in Fauron and Wolstenholme (1976). EcoRI and HindIII restriction fragments of *D. yakuba* mtDNA were cloned into pBR325 and pBR322, respectively, using *Escherichia coli* K12.HB101 as host. These cloned fragments, or subfragments produced by further restriction enzyme digestion, were recloned into bacteriophage M13mp2, M13mp8, or M13mp9 (Gronenborn and Messing 1978; Messing and Vieira 1982). Using the replicative forms of M13 DNA molecules containing various restriction fragments of *D. yakuba* mtDNA and DNaseI digestion (Hong 1982), partial deletions of the mtDNA fragments were generated. Viral DNAs containing deletions of the original mtDNA restriction fragments were selected by size using agarose gel electrophoresis.

Experimental details concerning restriction enzyme digestions, electrophoresis, cloning, and purification of M13 DNAs are given or referred to in Clary et al. (1982) and Clary and Wolstenholme (1983a).

DNA sequences were obtained from M13-cloned fragments by the extension–dideoxyribonucleotide termination procedure of Sanger et al. (1977) using [α-^{32}P]dATP.

Individual sequences were assembled into a circular consensus sequence using the computer program of Staden (1982) in a Digital Equipment Corporation 20/60 computer. Transfer RNA genes were identified from their ability to fold into the characteristic cloverleaf structure of tRNAs and from the trinucleotide in the anticodon position in these structures either by eye or using the TRNA program of Staden (1980). Ribosomal RNA genes were identified by nucleotide sequence homologies to mouse mitochondrial rRNA genes (Bibb et al. 1981). Nucleotide sequences were analyzed by the SEQ program (Brutlag et al. 1982). Protein and presumptive protein genes (URFs) were identified by comparing predicted amino acid sequences with amino acid sequences of previously identified mouse mitochondrial protein genes (Bibb et al. 1981) using the TYPIN and SEARCH programs (Jue et al. 1980; Doolittle 1981).

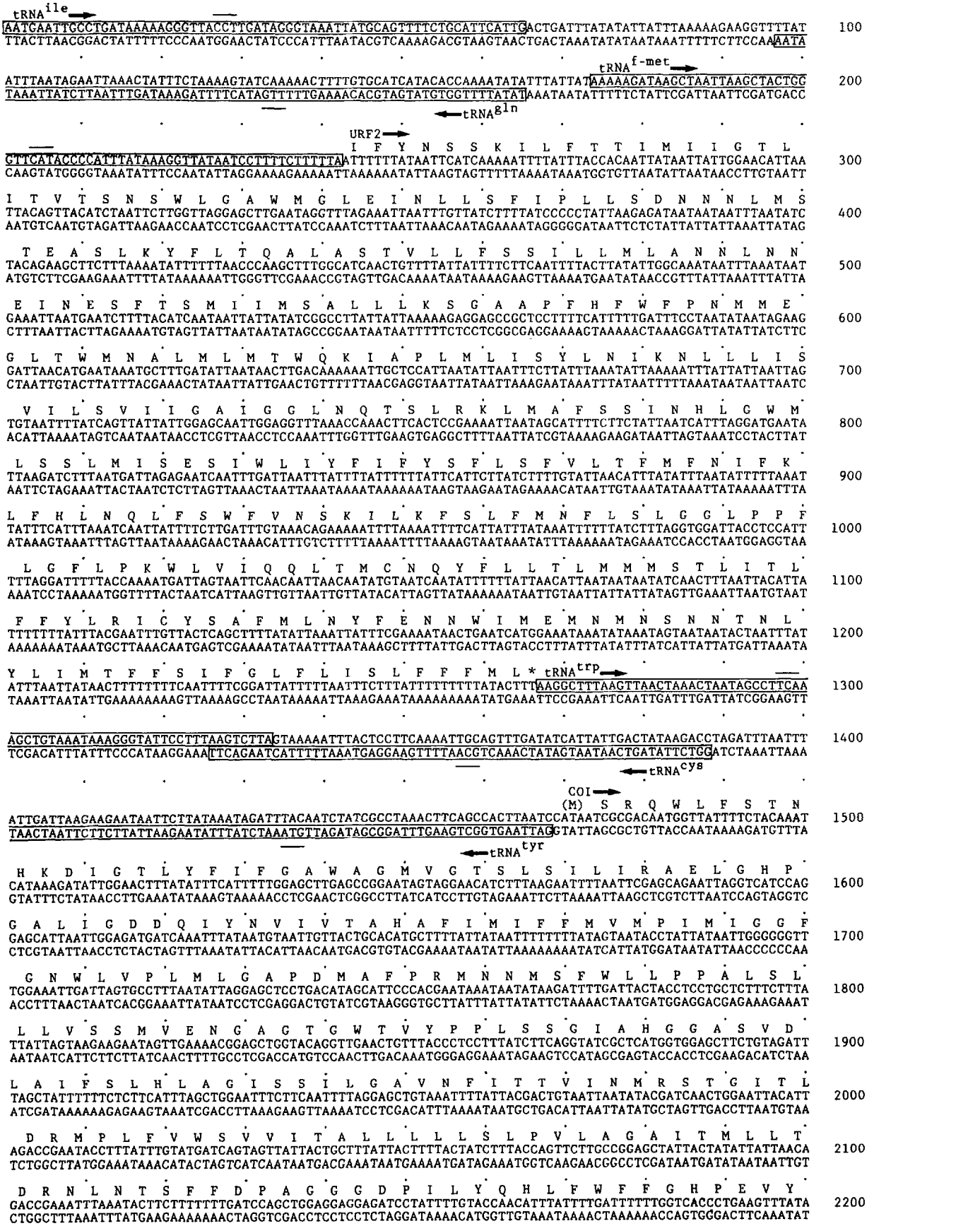
Results and Discussion

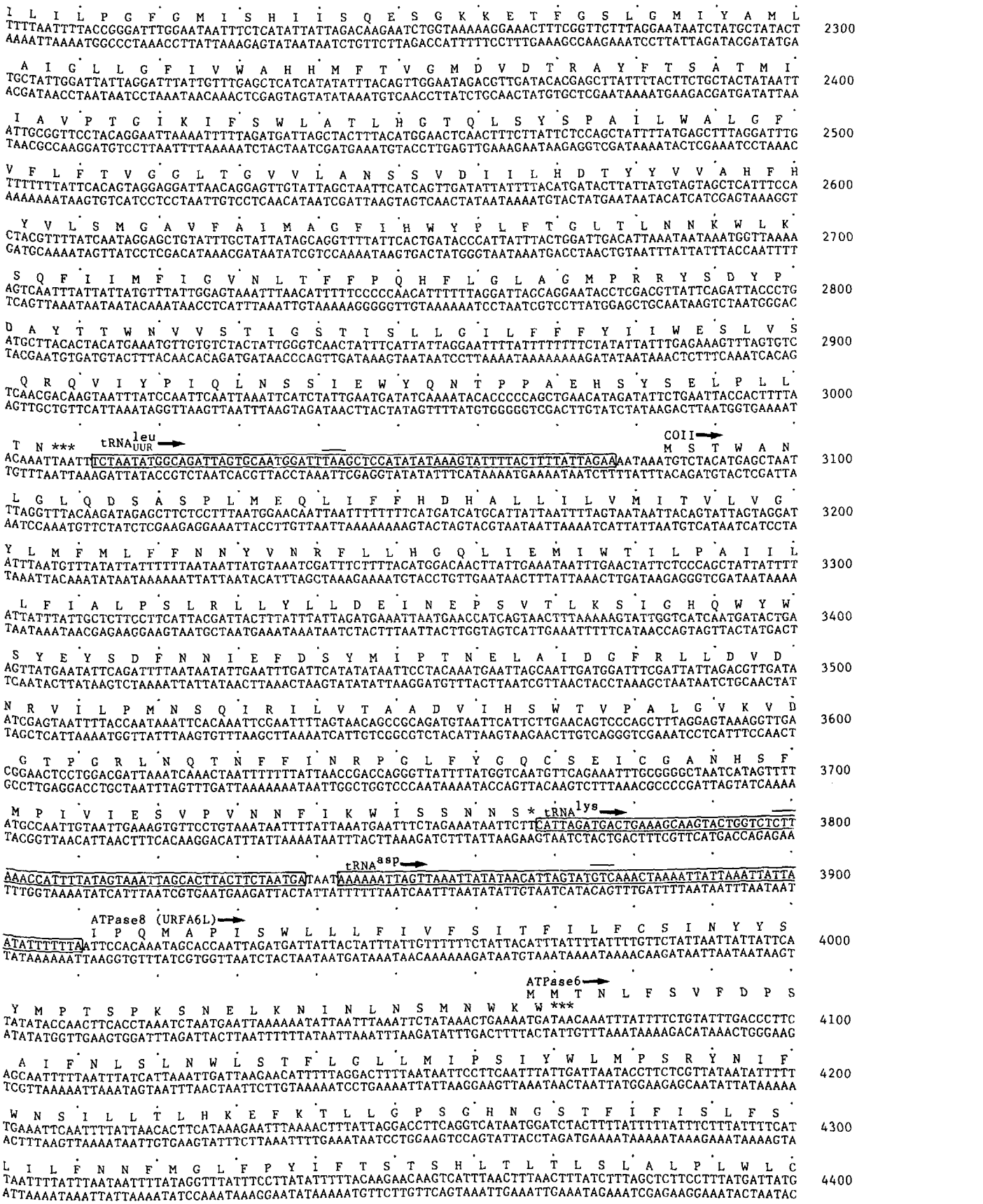
Genome Organization and Transcription

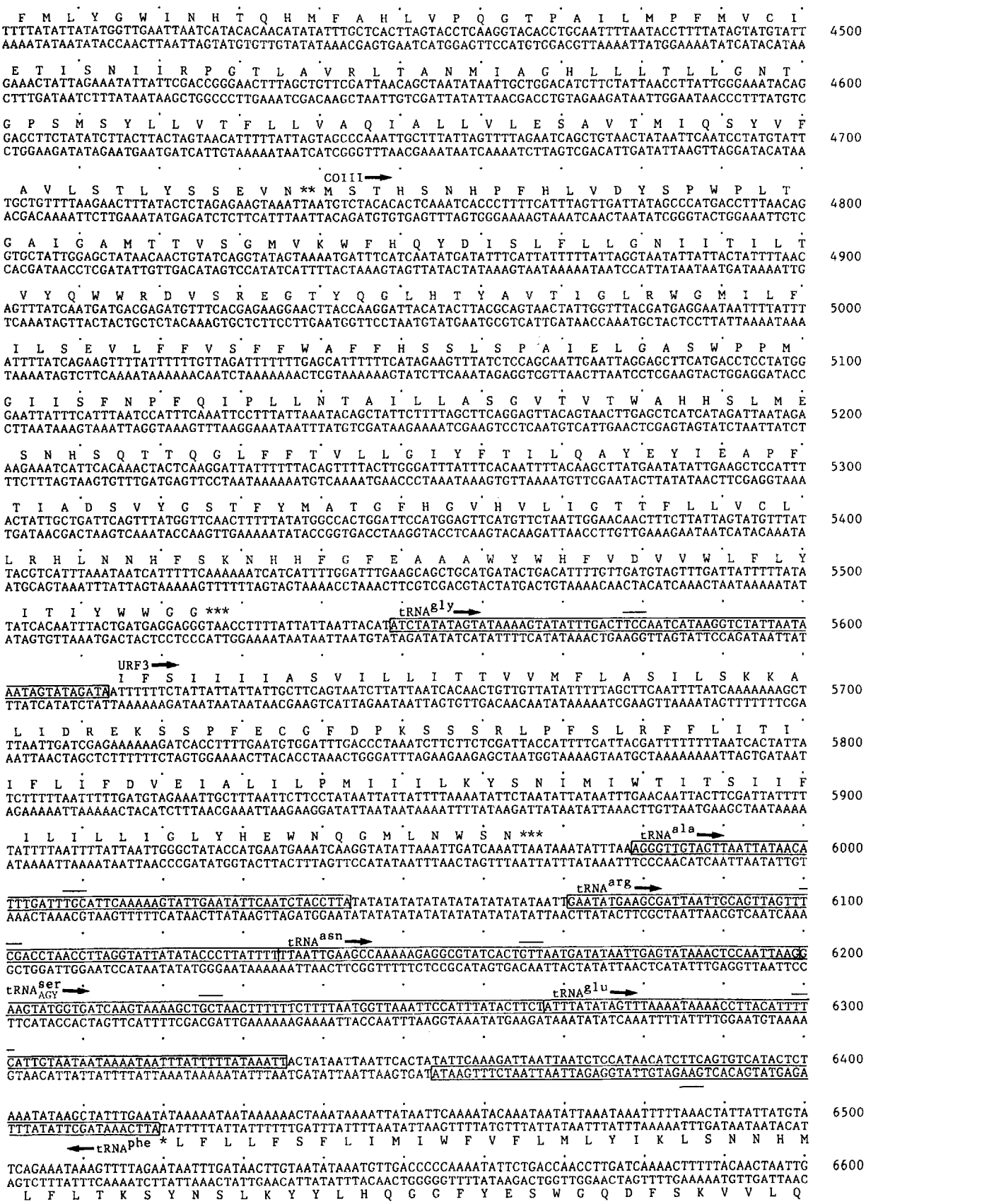
The 16,019 nucleotide-pair sequence of the *D. yakuba* mtDNA molecule is presented in Fig. 1, and the relative arrangement of genes in the molecule is summarized in Fig. 2. The molecule contains 13 open reading frames, which, based on comparisons of nucleotide sequences and predicted amino acid sequences with sequences of mouse mtDNA (Table 1), have been shown to be the genes for cytochrome b, cytochrome c oxidase subunits I, II, and III (COI–III), ATPase subunits 6 and 8 (ATPase 6 and ATPase 8: the latter was formerly URFA6L, see below), and the seven presumptive protein genes that have been designated URF1, 2, 3, 4L, 4, 5, and 6. Two genes for mitochondrial rRNAs (mt-rRNAs) have been identified within the *D. yakuba* sequence. Also, there are 22 tRNA genes, which are found either singly or in clusters of two to six genes between the protein and rRNA genes. As in mammalian mtDNAs, none of the genes found in *D. yakuba* mtDNA contain introns.

Between the $tRNA^{ile}$ gene and the small rRNA gene occurs a sequence of 1077 nucleotides that is 92.8% A + T. This region constitutes the major portion of the A + T-rich region of the *D. yakuba* mtDNA molecule previously defined from electron microscope studies (Fauron and Wolstenholme 1976, 1980) and shown to contain the molecule's origin of replication (Goddard and Wolstenholme 1980).

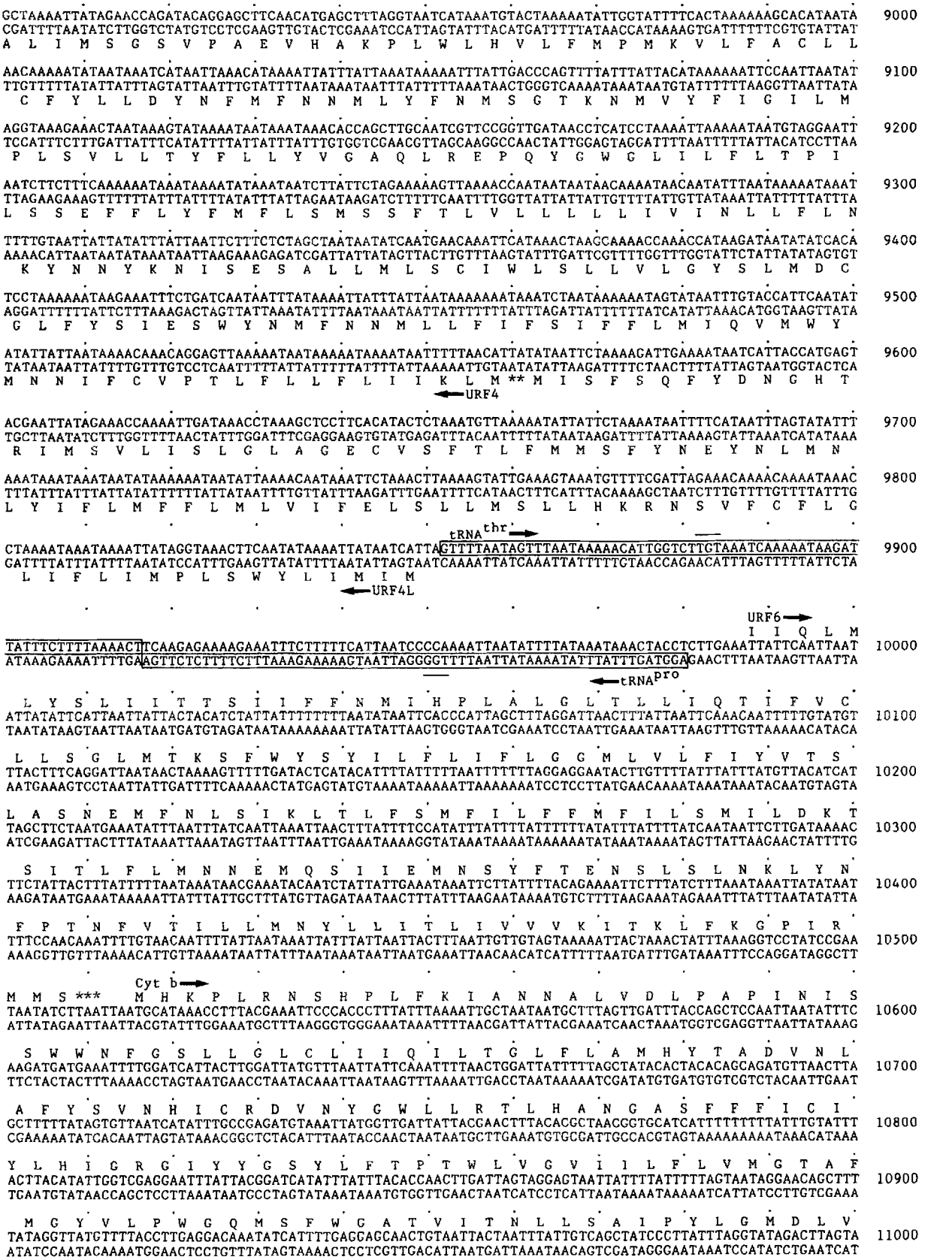
Except for the replication origin-containing region, nucleotides between genes in *D. yakuba* mtDNA are either absent or occur in small numbers (1–31). The total number of such nucleotides is 183, which is greater than the total number of intergenic nucleotides in the mtDNAs of mouse (64), human


tRNA ile
tRNA f-met
tRNA gln
URF2
tRNA trp
tRNA cys
COI
tRNA tyr


tRNA leu UUR
COII
tRNA lys
tRNA asp
ATPase8 (URFA6L)
ATPase6


COIII
tRNA gly
URF3
tRNA ala
tRNA arg
tRNA asn
tRNA ser AGY
tRNA glu
tRNA phe

```
ACCATAATTTAAAGGATAAAAAATTATTCCATAAGTTCTAATATAAGGTATAAATCATATTGAACCTAAAAATAATGTTAAATTATAATTTAATAGAGAT  6700
TGGTATTAAATTTCCTATTTTTTAATAAGGTATTCAAGATTATATTCCATATTTAGTATAACTTGGATTTTTATTACAATTTAATATTAAATTATCTCTA
  G  Y  N  L  P  Y  F  I  M  G  Y  T  S  I  Y  P  M  F  W  M  S  G  L  F  L  T  L  N  Y  N  L  L  S

TTATTCAAAGAATATAAATTTCTGATAGAAATTAGATACCCAAATAAACCCCCCAACAATACAAACAAATAATGTTAATATTTTTAAATACCCAGGTAAAC  6800
AATAAGTTTCTTATATTTAAAGACTATCTTTAATCTATGGGTTTATTTGGGGGTTGTTATGTTTGTTTATTACAATTATAAAAATTTATGGGTCCATTTG
  K  N  L  S  Y  L  N  S  I  S  I  L  Y  G  F  L  G  G  V  I  C  V  F  L  T  L  M  K  L  Y  G  P  L  C

AAATTATATAAGGAAAAGGAAAAATTAACCAATTTAACATTCTACCTCCAATAATTCTTATAAATAATAAGCCTAGTATACCCCGAAGTATTACTCAACT  6900
TTTAATATATTCCTTTTCCTTTTTAATTGGTTAAATTGTAAGATGGAGGTTATTAAGAATATTTATTATTCGGATCATATGGGGCTTCATAATGAGTTGA
    I  M  Y  P  F  P  F  I  L  W  N  L  M  S  G  G  I  I  S  M  F  L  L  G  L  M  G  R  L  M  V  W  S

TTCATCATTTAATATATTTAAACTTCCACAATTTAAATCTCCAGTTATTGAATAATATACTAACCGAAAAGAATAACTTACTGTTAAACCTGTGGAAAAA  7000
AAGTAGTAAATTATATAAATTTGAAGGTGTTAAATTTAGAGGTCAATAACTTATTATATGATTGGCTTTCTTATTGAATGACAATTTGGACACCTTTTT
  E  D  N  L  M  N  L  S  G  C  N  L  D  G  T  M  S  Y  Y  V  L  R  F  S  Y  S  V  T  L  G  T  S  F

AAGTATAAAAAAAATGAAAATATATTAATATTTCTAATTCTAACAATTTCTAAAATTATATCCTTAGAATAAAATCCAGCTAAAAATGGTATTCCACATA  7100
TTCATATTTTTTTTACTTTTATATAATTATAAAGATTAAGATTGTTAAAGATTTTAATATAGGAATCTTATTTTAGGTCGATTTTTACCATAAGGTGTAT
  F  Y  L  F  F  S  F  M  N  I  N  S  I  S  V  I  E  L  I  M  D  K  S  Y  F  G  A  L  F  P  M  G  C  L

AAGCCAAATTAGAAACATTAAAACAAGCTGAAGTTAAAGGTATATGAATTCTCAACCCTCCTATTAACCGAATATCTTGAGAATTATTTATATTATGAAT  7200
TTCGGTTTAATCTTTGTAATTTTGTTCGACTTCAATTTCCATATACTTAAGAGTTGGGAGGATAATTGGCTTATAGAACTCTTAATAAATATAATACTTA
    A  L  N  S  V  N  F  C  A  S  T  L  P  M  H  I  S  L  G  G  M  L  R  I  D  Q  S  N  N  M  N  H  I

AATAGCTCCTGCACATATAAATAATAAAGCTTTAAATAAAGCATGAGTTAATAAATGAAATATAGCTAATTTATAAAATCCCATAGATAAAATTCTTATT  7300
TTATCGAGGACGTGTATATTTATTATTTCGAAATTTATTTCGTACTCAATTATTTACTTTATATCGATTAAATATTTTAGGGTATCTATTTTAAGAATAA
    I  A  G  A  C  M  F  L  L  A  K  F  L  A  H  T  L  L  H  F  M  A  L  K  Y  F  G  M  S  L  I  S  M

ATTAATCCTAATTGACTAAGAGTTGACAACGCAATAATTTTCTTTAAATCAAACTCAAAATTAGCCCCCAGTCCTGCTATAAATATAGTTAAACCTGATA  7400
TAATTAGGATTAACTGATTCTCAACTGTTGCGTTATTAAAAGAAATTTAGTTTGAGTTTTAATCGGGGGTCAGGACGATATTTATATCAATTTGGACTAT
  M  L  G  L  Q  S  L  T  S  L  A  I  I  K  K  L  D  F  E  F  N  A  G  L  G  A  M  F  M  T  L  G  S  L

ATAACAACAATAATTGTCCTAATCAAGAAGTACTTAAAACAATATTAAATCGAATTAATAAATAAACTCCTGCTGTTACTAAAGTAGAAGAATGGACTAA  7500
TATTGTTGTTATTAACAGGATTAGTTCTTCATGAATTTTGTTATAATTTAGCTTAATTATTTATTTGAGGACGACAATGATTTCATCTTCTTACCTGATT
    L  L  L  L  Q  G  L  W  S  T  S  L  V  I  N  F  R  I  L  L  Y  V  G  A  T  V  L  T  S  S  H  V  L

AGCAGAGACAGGAGTAGGGGCAGCTATAGCAGCTGGTAATCAAGAAGAAAAAGGAATTTGAGCTCTTTTAGTTATAGCAGCTAATATAACTAAACTACCA  7600
TCGTCTCTGTCCTCATCCCCGTCGATATCGTCGACCATTAGTTCTTCTTTTTCCTTAAACTCGAGAAAATCAATATCGTCGATTATATTGATTTGATGGT
  A  S  V  P  T  P  A  A  M  A  A  P  L  W  S  S  F  P  I  Q  A  S  K  T  M  A  A  L  M  V  L  S  G

ATTATTAATATTGAAAATTCATTTTGTATAACTTCTAAATAAAAAATATAATTTCATCTACCATAATTTAATATTCAAGCAATAGCAAGAAGAAGAGCTA  7700
TAATAATTATAACTTTTAAGTAAAACATATTGAAGATTTATTTTTTATATTAAAGTAGATGGTATTAAATTATAAGTTCGTTATCGTTCTTCTTCTCGAT
  I  M  L  M  S  F  E  N  Q  M  V  E  L  Y  F  I  Y  N  W  S  G  Y  N  L  M  W  A  I  A  L  L  L  A  V

CATCTCCAATTCGATTAGATAATGCAGTTAATATTCCAGCATTGTAAGATTTGATATTTTGAAAATAAATTACTAAACAATAAGAAACAAGTCCTAATCC  7800
GTAGAGGTTAAGCTAATCTATTACGTCAATTATAAGGTCGTAACATTCTAAACTATAAAACTTTTATTTAATGATTTGTTATTCTTTGTTCAGGATTAGG
    D  G  I  R  N  S  L  A  T  L  M  G  A  N  Y  S  K  I  N  Q  F  Y  I  V  L  C  Y  S  V  L  G  L  G

ATCTCACCCTAATAGAATTCTCACTAAATTTGGTCTGATAATTAATAATATTATTGACAAAACAAATATTAATACTAATATAATAAACCGATTAATATTT  7900
TAGAGTGGGATTATCTTAAGAGTGATTTAAACCAGACTATTAATTATAATAACTGTTTTGTTTATAATTATGATTATATTATTTGGCTAATTATAAA
  D  W  G  L  L  I  S  V  L  N  P  S  I  I  L  L  M  M  S  L  V  F  M  L  V  L  M  I  F  R  N  I  N

TCATCACTTTCTATATATTCTTTTCTATAAAAAATTACTAAAGAAGCAATTATAAGAACAAAAGATATAAATAATAAACTTATTCAATCAAATAAAAAAG  8000
AGTAGTGAAAGATATATAAGAAAAGATATTTTTTAATGATTTCTTCGTTAATATTCTTGTTTTCTATATTTATTATTTGAATAAGTTAGTTTATTTTTTC
  E  D  S  E  M  Y  E  K  S  Y  F  I  V  L  S  A  I  M  L  V  F  S  M  F  L  L  S  M  W  D  F  L  F  T

TTATCACAATTCTTATAGAATTTAAAGATACTACTTCTCATTCAATAAAATAAACTATATTATTTAATAAATAATATAAACTTAATAAAAAACATGTTAA  8100
AATAGTGTTAAGAATATCTTAAATTTCTATGATGAAGAGTAAGTTATTTTATTTGATATAATAAATTATTTATTATATTTGAATTATTTTTTGTACAATT
    M  V  I  S  M  S  N  L  S  V  V  E  W  E  I  F  Y  V  M  N  N  L  L  Y  Y  L  S  L  L  F  C  T  L

ACTAATAGAAATTAAATTAATAAATCTAATTCTACAAATTGATAAATATTTCACGATCTAAAATGAATAACTTCATATCACTAACACCACAAATTAGTAT  8200
TGATTATCTTTAATTTAATTATTTAGATTAAGATGTTTAACTATTTATAAAGTGCTAGATTTTACTTATTGAAGTATAGTGATTGTGGTGTTTAATCATA
  S  I  S  I  L  N  I  F  S  I  S  C  I
                                    ←URF5

TTTTTTTAAACTATTTAAATATAATCATAAAATACATGATTCTCTTTTTAAAATTAATAAATTTAAAGGTAATCAATGTAATAATATTAATAAATATTCC  8300
AAAAAAATTTGATAAATTTATATTAGTATTTTATGTACTAAGAGAAAAATTTTAATTATTTAAATTTCCATTAGTTACATTATTATAATTATTTATAAGG
          ←tRNA^his  *  L  W  L  I  C  S  E  S  K  L  I  L  L  N  L  P  L  W  H  L  L  M  L  L  Y  E

CGAATTTTACCTCTTCTAAATGAATATACCCCAGAAAATAATTTACCATGTTGACTAAAAGAATATAAATATAAAGTATAGGCTGCTCTAAAGAAAGATA  8400
GCTTAAAATGGAGAAGATTTACTTATATGGGGTCTTTTATTAAATGGTACAACTGATTTTCTTATATTTATATTTCATATCCGACGAGATTTCTTTCTAT
  R  I  K  G  S  S  F  S  Y  V  G  S  F  L  K  G  H  Q  S  F  S  Y  L  Y  L  T  Y  A  A  S  F  F  S  L

AAAAAGATAATATAATTATTGAAATTCATGACCAAGAAACAATTCTATTTAATAAAGAAATTTCTCCTAATAAATTTAACGTTGGAGGAGCTGCTATATT  8500
TTTTTCTATTATATTAATAACTTTAAGTACTGGTTCTTTGTTAAGATAAATTATTTCTTTAAAGAGGATTATTTAAATTGCAACCTCCTCGACGATATAA
    F  S  L  M  I  M  S  I  W  S  W  S  V  I  S  N  L  L  S  I  E  G  L  L  N  L  T  P  P  A  A  M  N

AGCTGATCTTAATAAAAATCATCATAAAGTTATAGCAGGTATAAAATTTAATAACCCCTTATTAATTAATATTCTTCGTCTTCCTAATCGTTCATAAGAA  8600
TCGACTAGAATTATTTTTAGTAGTATTTCAATATCGTCCATATTTTAAATTATTGGGGAATAATTAATTATAAGAAGCAGAAGGATTAGCAAGTATTCTT
  A  S  S  L  L  F  W  W  L  T  M  A  P  M  F  N  L  L  G  K  N  I  L  M  S  R  S  G  L  R  E  Y  S

ACATTTGCTAAACAAAATAAACCAGAAGAACATAAACCATGAGCAATTATTAATGTATAAGATCCACATAATCCTCAATAAGTTATTGTTAATAAACCAG  8700
TGTAAACGATTTGTTTTATTTGGTCTTCTTGTATTTGGTACTCGTTAATAATTACATATTCTAGGTGTATTAGGAGTTATTCAATAACAATTATTTGGTC
  V  N  A  L  C  F  L  G  S  S  C  L  G  H  A  I  M  L  T  Y  S  G  C  L  G  W  Y  T  M  T  L  L  G  A

CTAAAACAATTCCTATATGAGCAACTGATGAATATGCAATTAAAGCCTTTAAATCAGTTTGTCGTAAACATACTAATCTTATTAACACACCTCCTACTAA  8800
GATTTTGTTAAGGATATACTCGTTGACTACTTATACGTTAATTTCGGAAATTTAGTCAAACAGCATTTGTATGATTAGAATAATTGTGTGGAGGATGATT
  L  V  I  G  M  H  A  V  S  S  Y  A  I  L  A  K  L  D  T  Q  R  L  C  V  L  S  M  L  V  G  G  V  L

TCTAATTCTAATTCAAACAAATCTATATTTTAAATTTATTAATTGTAAAAAATTAATTACTCGTAATAAACCATAACCTCCTAATTTTAATATAATACCA  8900
AGATTAAGATTAAGTTTGTTTAGATATAAAATTTAAATAATTAACATTTTTTAATTAATGAGCATTATTTGGTATTGGAGGATTAAAATTATATTATGGT
  S  I  S  I  W  V  F  S  Y  K  L  N  M  L  Q  L  F  N  I  V  R  L  L  G  Y  G  G  L  K  L  M  I  G
```


URF4
tRNA thr
URF4L
URF6
tRNA pro
Cyt b

```
  Q  W  L  W  G  G  F  A  V  D  N  A  T  L  T  R  F  F  T  F  H  F  I  L  P  F  I  V  L  A  M  T  M
CAATGATTATGAGGAGGATTTGCTGTAGATAATGCTACTTTAACTCGATTTTTCACATTTCATTTTATTTTACCTTTTATTGTTCTTGCTATAACTATAA  11100
GTTACTAATACTCCTCCTAAACGACATCTATTACGATGAAATTGAGCTAAAAAGTGTAAAGTAAAATAAAATGGAAAATAACAAGAACGATATTGATATT

  I  H  L  L  F  L  H  Q  T  G  S  N  N  P  I  G  L  N  S  N  I  D  K  I  P  F  H  P  Y  F  T  F  K  D
TTCATCTACTATTTTTACATCAAACAGGATCTAATAACCCTATTGGTTTAAATTCTAATATTGATAAAATTCCTTTTCACCCATACTTCACATTTAAGGA  11200
AAGTAGATGATAAAAATGTAGTTTGTCCTAGATTATTGGGATAACCAAATTTAAGATTATAACTATTTTAAGGAAAAGTGGGTATGAAGTGTAAATTCCT

    I  V  G  F  I  V  M  I  F  I  L  I  S  L  V  L  I  S  P  N  L  L  G  D  P  D  N  F  I  P  A  N  P
TATTGTAGGATTTATTGTAATAATTTTTATTCTAATTTCATTAGTTTTAATTAGACCAAATTTATTGGGAGACCCAGATAACTTTATTCCTGCTAATCCT  11300
ATAACATCCTAAATAACATTATTAAAAATAAGATTAAAGTAATCAAAATTAATCTGGTTTAAATAACCCTCTGGGTCTATTGAAATAAGGACGATTAGGA

  L  V  T  P  A  H  I  Q  P  E  W  Y  F  L  F  A  Y  A  I  L  R  S  I  P  N  K  L  G  G  V  I  A  L
TTAGTAACACCAGCTCACATTCAACCAGAATGATATTTTTTATTTGCTTACGCAATTCTTCGTTCAATTCCTAATAAATTAGGAGGAGTTATTGCATTAG  11400
AATCATTGTGGTCGAGTGTAAGTTGGTCTTACTATAAAAAATAAACGAATGCGTTAAGAAGCAAGTTAAGGATTATTTAATCCTCCTCAATAACGTAATC

  V  L  S  I  A  I  L  M  I  L  P  F  Y  N  L  S  K  F  R  G  I  Q  F  Y  P  I  N  Q  I  L  F  W  S  M
TTTTATCAATTGCAATTTTAATAATTTTACCTTTTTATAATTTAAGAAAATTCCGAGGAATCCAATTTTATCCAATTAACCAAATTTTATTTTGATCTAT  11500
AAAATAGTTAACGTTAAAATTATTAAAATGGAAAAATATTAAATTCTTTTAAGGCTCCTTAGGTTAAAATAGGTTAATTGGTTTAAAATAAAACTAGATA

  L  V  T  V  I  L  L  T  W  I  G  A  R  P  V  E  E  P  Y  V  L  I  G  Q  I  L  T  I  I  Y  F  L  Y
ATTAGTTACAGTAATTTTATTAACATGAATTGGAGCTCGACCAGTTGAAGAACCTTATGTATTAATTGGACAAATTTTAACTATTATTTATTTTTTATAT  11600
TAATCAATGTCATTAAAATAATTGTACTTAACCTCGAGCTGGTCAACTTCTTGGAATACATAATTAACCTGTTTAAAATTGATAATAAATAAAAAATATA

  Y  L  I  N  P  L  V  T  K  W  W  D  N  L  L  N  ***          tRNA ser UCN
TATTTAATTAACCCACTAGTTACAAAATGATGAGATAATTTATTAAATTAATTAATTAGTTAATGAGCTTGAACAAGCGTATGTTTTGAAAACATAAGAT  11700
ATAAATTAATTGGGTGATCAATGTTTTACTACTCTATTAAATAATTTAATTAATTAATCAATTACTCGAACTTGTTCGCATACAAAACTTTTGTATTCTA

AGAATTTAATTTTCTATTAACTTTTTACTAAAAAAAATTCACAATAAAAAAGAAAATAATAAAATTTTAAACCCAATAAAAAATAATAAATAATTTAAAG  11800
TCTTAAATTAAAAGATAATTGAAAAATGATTTTTTTTAAGTGTTATTTTTTCTTTTATTATTTAAAATTTGGGTTATTTTTTATTATTTATTAAATTTC
    *** N  E  M  L  K  K  S  F  F  I  W  L  L  F  S  F  L  L  I  K  F  G  I  F  F  L  L  Y  N  L  S

AAAATGATAAAAAACATTTTCAAGCTAAATATATTAATTTATCATAACGAAACCGAGGTAATGTACCTCGAGCTCAAATAAATACAAATGAAATAAAAGT  11900
TTTTACTATTTTTTGTAAAAGTTCGATTTATATAATTAAATAGTATTGCTTTGGCTCCATTACATGGAGCTCGAGTTTATTTATGTTTACTTTATTTTCA
  F  S  L  F  C  K  W  A  L  Y  M  L  K  D  Y  R  F  R  P  L  T  G  R  A  W  I  F  V  F  S  I  F  T

TAATTTTACATAAAATAATAAATTAAATACATCACAACCTAAAAAAATTACACAAAACAATATTCTTATAAATAAAATTCTAGCATATTCTGCTATAAAA  12000
ATTAAAATGTATTTTATTATTTAATTTATGTAGTGTTGGATTTTTTTAATGTGTTTTGTTATAAGAATATTTATTTTAAGATCGTATAAGACGATATTTT
  L  K  V  Y  F  L  L  N  F  V  D  C  G  L  F  I  V  C  F  L  M  S  M  F  L  I  S  A  Y  E  A  M  F

ATTAAAGCAAAACCACCTCTTCTATATTCTACATTAAATCCTGAAACTAATTCTGATTCACCTTCAGCAAAATCAAAAGGAGTTCGATTAGTTTCAGCTA  12100
TAATTTCGTTTTGGTCGAGAAGATATAAGATGTAATTTAGGACTTTGATTAAGACTAAGTGGAAGTCGTTTTAGTTTTCCTCAAGCTAATCAAAGTCGAT
I  L  A  F  G  G  S  S  Y  E  V  N  F  G  S  V  L  E  S  E  G  E  A  F  D  F  P  T  R  N  T  E  A  L

ATGAAATTGTTAATCAAACTAAACTTATAGGAAATAAAATAATTAAAAATCATATATAAATTTGGTAATAAAAAAAATAAATTATATTATAACTCCCAAT  12200
TACTTTAACAATTAGTTTGATTTGAATATCCTTTATTTTATTAATTTTTAGTATATATTTAAACCATTATTTTTTTTATTTAATATAATATTGAGGGTTA
  S  I  T  L  W  V  L  S  M  P  F  L  I  I  L  F  W  M  Y  I  Q  Y  Y  F  F  Y  I  M  N  Y  S  G  I

TAAAAAAATAAATGATAATATAATTAAAGCTAATCTAACTTCATAAGAAATAGTTTGAGCCACAGCTCGTAAACCCCCTAATAAAGCATAATTAGAATTA  12300
ATTTTTTTATTTACTATTATATTAATTTCGATTAGATTGAAGTATTCTTTATCAAACTCGGTGTCGAGCATTTGGGGGATTATTTCGTATTAATCTTAAT
  L  F  I  F  S  L  M  I  L  A  L  S  V  E  Y  S  I  T  Q  A  V  A  R  L  G  G  L  L  A  Y  N  S  N

GAAGATCAGCCAGCTACTATAACTGTATAAACTCCTAATCTTGTACAACATAAAAAAAATAATCCACCTAAATTAAAAGAGTATAATTTAACAAAAAAAG  12400
CTTCTAGTCGGTCGATGATATTGACATATTTGAGGATTAGAACATGTTGTATTTTTTTTATTAGGTGGATTTAATTTTCTCATATTAAATTGTTTTTTTC
S  S  W  G  A  V  M  V  T  Y  V  G  L  S  T  C  C  L  F  F  L  G  G  L  N  F  S  Y  L  K  V  F  F  P

GTATACATATTCAAACAAATAAAGATAAAAATAAAGAAAAAATAGGAGAAATATAATATCTTAAATAATTTGATAATAATGGATAAGTTTGTTCTTTTGT  12500
CATATGTATAAGTTTGTTTATTTCTATTTTTATTTCTTTTTTATCCTCTTTATATTATAGAATTTATTAAACTATTATTACCTATTCAAACAAGAAAACA
  M  C  M  W  V  F  L  S  L  F  L  S  F  I  P  S  I  Y  Y  S  L  Y  N  S  L  L  P  Y  T  Q  E  K  T

AAATAATTTAATTGCATCACAAAAAGGTTGAGGAATTCCTATTAAACCAACTTTATTAGGTCCTTTACGAATTTGAATATACCCTAAAACTTTACGTTCT  12600
TTTATTAAATTAACGTAGTGTTTTTCCAACTCCTTAAGGATAATTTGGTTGAAATAATCCAGGAAATGCTTAAACTTATATGGGATTTTGAAATGCAAGA
  F  L  K  I  A  D  C  F  P  Q  P  I  G  M  L  G  V  K  N  P  G  K  R  I  Q  I  Y  G  L  V  K  R  E

AATAAAGTTAAAAAAGCTACACTTACTAATACACAAATAATTAATAATAAACTTCCAATTAATGATAAAATAAATTCTATATAAAACAAGTACTATTTGT  12700
TTATTTCAATTTTTTCGATGTGAATGATTATGTGTTTATTAATTATTATTTGAAGGTTAATTACTATTTTATTTAAGATATATTTTGTTGTATGATAAACA
L  L  T  L  F  A  V  S  V  L  V  C  I  I  L  L  L  S  G  I  L  S  L  I  F  E  M
                                                                                  URF1

AATAAAAATTACATATATAAATTCTAAATTTATTGCACTAATCTGCCAAAATAGTTTATTATATATTAATAATATTCTTATATAAAAATATAATTATTTT  12800
TTATTTTTAATGTATATATTTAAGATTTAAATAACGTGATTAGACGGTTTTATCAAATAATATATAATTATTATAAGAATATATTTTTATATTAATAAAA
                                  tRNA leu CUN

AATATTTGGTCCTTTCGTACTAAAATATTATAATTTTTTAAAGATAGAAACCAACCTGGCTTACACCGGTTTGAACTCAGATCATGTAAGAATTTAAAAG  12900
TTATAAACCAGGAAAGCATGATTTTATAATATTAAAAAATTTCTATCTTTGGTTGGACCGAATGTGGCCAAACTTGAGTCTAGTACATTCTTAAATTTTC

TCGAACAGACTTAAAATTTGAACGGCTACACCCAAAATTATATCTTAATCCAACATCGAGGTCGCAATCTTTTTTATCGATATGAACTCTCCAAAAAAAT  13000
AGCTTGTCTGAATTTTAAACTTGCCGATGTGGGTTTTAATATAGAATTAGGTTGTAGCTCCAGCGTTAGAAAAAATAGCTATACTTGAGAGGTTTTTTTA

TACGCTGTTATCCCTAAAGTAACTTAATTTTTTAATCATTATTAATGGATCAATTATTCATAAATTAATGTTTTTTAAAATTAAAAGTTTTTTAAATTTT  13100
ATGCGACAATAGGGATTTCATTGAATTAAAAAATTAGTAATAATTACCTAGTTAATAAGTATTTAATTACAAAAAATTTTAATTTTCAAAAAATTTAAAA

AATATCACCCCAATAAAATATTTTAATTTATTAAAATTAAATTAATCTTTATAATTAAAATAAATAAAATATAAAGATTTATAGGGTCTTCTCGTCTTTT  13200
TTATAGTGGGGTTATTTTATAAAATTAAATAATTTTAATTTAATTAGAAATATTAATTTTATTTATTTTATATTTCTAAATATCCCAGAAGAGCAGAAAA

AAATTAATTTTAGCTTTTTGACTAAAAAATAAAATTCTATTTTAAATTTAAATGAAACAGTTAATATTTCGTCCAACCATTCATTCCAGCCTTCAATTAA  13300
TTTAATTAAAATCGAAAAACTGATTTTTTATTTTAAGATAAAATTTAAATTTACTTTGTCAATTATAAAGCAGGTTGGTAAGTAAGGTCGGAAGTTAATT
```

```
AAGACTAATGATTATGCTACCTTTGCACAGTCAAAATACTGCGGCCATTTAAAATTTTCAGTGGGCAGGTTAGACTTTATATATAATTCAAAAAGACATG  13400
TTCTGATTACTAATACGATGGAAACGTGTCAGTTTTATGACGCCGGTAAATTTTAAAAGTCACCCGTCCAATCTGAAATATATATTAAGTTTTTCTGTAC

TTTTTGTTAAACAGGCGAACATTATTTTTGCCGAATTCTTTATTTAAACTTTTCATAAAAATTTATTTTAACAATATTATATACTAATTCTATCATTATT  13500
AAAAACAATTTGTCCGCTTGTAATAAAAACGGCTTAAGAAATAAATTTGAAAAGTATTTTTAAATAAAATTGTTATAATATATGATTAAGATAGTAATAA

ACTTAATTTTAAATATTAAAATTAATATTTTAATAAATAATTAAAATTTAATAAATAATTTAATTTATAAAATAAATTATAACACATTTTTTAATAATTG  13600
TGAATTAAAATTTATAATTTTAATTATAAAATTATTTATTAATTTTAAATTATTTATTAAATTAAATATTTTATTTAATATTGTGTAAAAAATTATTAAC

CTAATTCTAAGCATATATTTATTAAATTTATTTATTATTTTTAAAAATTTATTTTATAGCTTATCCCATAAAATATTAAAATTATAAATTATTTAATTAA  13700
GATTAAGATTCGTATATAAATAATTTAAATAAATAATAAAAATTTTTAAATAAAATATCGAATAGGGTATTTTATAATTTTAATATTTAATAAATTAATT

ATAAATAATTAAATAAATTTATAATTTCTAAATTAAATTTATTTCTTAAAAAACTAGATACCTTTAAAAACGAATAACATTTCATTTCTAATATAATATT  13800
TATTTATTAATTTATTTAAATATTAAAGATTTAATTTAAATAAAGAATTTTTTGATCTATGGAAATTTTTGCTTATTGTAAAGTAAAGATTATATTATAA

ATAAATAATTTTGTCACATTAACTTAAATATTATATTAACTCTTTTAAAATCGAGAAAAATAAATATTTATTTTTTATTTAATAAACGCTGATACACAAG  13900
TATTTATTAAAACAGTGTAATTGAATTTATAATATAATTGAGAAAATTTTAGCTCTTTTTATTTATAAATAAAAAATAAATTATTTGCGACTATGTGTTC

GTACAATAAATTAAATTTTCTTTTAAAATAAAATTTTTTCAAATTATTTCAATTTTCTTTTACAATACTAATATACTATTATTAAAATTATTTTTTCTTT  14000
CATGTTATTTAATTTAAAAGAAAATTTTATTTTAAAAAAGTTTAATAAAGTTAAAAGAAAATGTTATGATTATATGATAATAATTTTAATAAAAAAGAAA

AAACAATACTAAAACTTTTAAATTTATAGTTATTTCTAATATTTTTATATAAAATAATTAAAATTAATAAATAAAATATAATCAATTTATATTGATTTGC  14100
TTTGTTATGATTTTGAAAATTTAAATATCAATAAAGATTATAAAAATATATTTTATTAATTTTAATTATTTATTTATTTATTAGTTAAATATAACTAAACG
                                                                                    ←large rRNA

ACAAAAATCTTTTCAATGTAAATGAAATGCTTTACTTAATAAGCTTTAAATTGTCATTCTAGATACACTTTCCAGTACATCTACTATGTTACGACTTATC  14200
TGTTTTTAGAAAAGTTACATTTACTTTACGAAATGAATTATTCGAAATTTAACAGTAAGATCTATGTGAAAGGTCATGTAGATGATACAATGCTGAATAG
                                                    ←tRNA^val

TTACCTTAATAATAAGAGCGACGGGCGATGTGTACATATTTTAGAGCTAAAATCAAATTATTAATCTTTATAATTTTACTACCAAATCCACCTTCAAAAA  14300
AATGGAATTATTATTCTCGCTGCCCGCTACACATGTATAAAATCTCGATTTTAGTTTAATAATTAGAAATATTAAAATGATGGTTTAGGTGGAAGTTTTT

TTTTTTCATAATTTTATCCGTTTAAATAAATTTATTGTAACCCATTATTACTTAAATATAAGCTACACCTTGATCTGATATAAATTTTTATTAAAATTAT  14400
AAAAAAGTATTAAAATAGGCAAATTTATTTAAATAACATTGGGTAATAATGAATTTATATTCGATGTGGAACTAGACTATATTTAAAAATAATTTTAATA

TGAATATTATTATTCTTATAAAATATTCTGATAACGACGGTATATAAACTGATTACAAATTTAAGTAAGGTCCATCGTGGATTATCGATTACAAAACAGG  14500
ACTTATAATAATAAGAATATTTTATAAGACTATTGCTGCCATATATTTGACTAATGTTTAAATTCATTCCAGGTAGCACCTAATAGCTAATGTTTTGTCC

TTCCTCTGGATAGACTAAAATACCGCCAAATTTTTTAAGTTTCAAGAACATAACTATTACTACTTTAGCAATTTATTTACATTTTAAATAATAGGGTATC  14600
AAGGAGACCTATCTGATTTTATGGCGGTTTAAAAAATTCAAAGTTCTTGTATTGATAATGATGAAATCGTTAAATAAATGTAAAATTTATTATCCCATAG

TAATCCTAGTTTTTTATTAAAATTTTTTAACTTCAATTATTATTTTATAAAATAATTTAAATATAAAATTTCACTTAATATATTTAATTTTATTTTTAAA  14700
ATTAGGATCAAAAAATAATTTTAAAAAATTGAAGTTAATAATAAAATATTTTATTAAATTTATATTTTAAAGTGAATTATATAAATTAAAATAAAAATTT

TAAATCAATTTAATTCATACTAAAAAAATTTATTTGTATTATTGGTATAACCGCGACTGCTGGCACCAATTTGGTCAATACTTTTTAATATTGCTATTTC  14800
ATTTAGTTAAATTAAGTATGATTTTTTTAAATAAACATAATAACCATATTGGCGCTGACGACCGTGGTTAAACCAGTTATGAAAAATTATAACGATAAAG

TAAATTTCTTTAATTAATAATATTAATTACTGCGAATAAATAATTTATAATATATTTATTTTTTAAATAAATATAAATTCACACAAAAATTTACATATAA  14900
ATTTAAAGAAATTAATTATTATAATTAATGACGCTTATTTATTAAATATTATATAAATAAAAAATTTATTTATATTTAAGTGTGTTTTTAAATGTATATT

                                    A+T rich region
ATCAAACTAATAACAAATTTTTAAGCCAAAATAAAACTTTAAAAATATACAATAAAAAAATTTATTTATAATATAAATTTAATAAAATTTATCAAAATAA  15000
TAGTTTGATTATTGTTTAAAAATTCGGTTTTATTTTGAAATTTTTATATGTTATTTTTTTAAATAAATATTATATTTAAATTATTTTAAATAGTTTTATT
                          ←small rRNA

TTTTTATAAATAATAAATAAAATAAAAAATAATAGATAATTTAAATAATAATTAATAAATAAAAACTTTATAAACAAAAATAATTTAATTTATATATATA  15100
AAAAATATTTATTATTTATTTTATTTTTTATTATCTATTAAATTTATTATTAATTATTTATTTTTGAAATATTTGTTTTTATTAAATTAAATATATATAT

TATATATATTAGTATATTAAATAAAATTCAAATTATTTTTTATCATAAATAATTAATAATTAAAATTAGTAAAATTTTTAAAAAAATAATAAAATTTGTA  15200
ATATATATAATCATATAATTTATTTTAAGTTTAATAAAAAATAGTATTTATTAATTATTAATTTTAATCATTTTAAAAATTTTTTTATTATTTTAAACAT

ACTTATAATAAACTATTTAAATAATTATTTTTAATCACTAAATCTGATAACTTATTCCCCTATAATAAAATATAATAATTTCTTAAATTAAATACCTTAT  15300
TGAATATTATTTGATAAATTTATTAATAAAAAATTAGTGATTTAGACTATTGAATAAGGGGATATTATTTTATATTATTAAAGAATTTAATTTATGGAATA

TTACATATTTATTTTTTAATATTTATACAAATAAAATTATTTTTTTTTATACATTAAAATAATAAAAATTAAAATTAATAAAATTTACTATATTTATTAAA  15400
AATGTATAAATAAAAAATTATAAATATGTTTATTTTAATAAAAAAAATATGTAATTTTATTATTTTTAATTTTAATTATTTTAAATGATATAAATAATTT

CAAAATAATTATTTATAAATAAAAATTATAAAACAATATTTGTTAATTAAATAAATTAAAAAATTTTATAAATTAATTAATAAATTATAATAATTTTTAA  15500
GTTTTATTAATAAATATTTATTTTTAATATTTTGTTATAAACAATTAATTTATTTAATTTTTTAAAATATTTAATTAATTATTTAATATTATTAAAAATT

AATATTAATTTTAAAATTATATTAATTAAATAAATTTTTCTAATTAATAATTATTTATAAAATAATAAAAAACCTATTCTTTTTTTTTTTTTTAAAAAAAA  15600
TTATAATTAAAATTTTAATATAATTAATTTATTTAAAAAGATTAATTATTAATAAATATTTTATTATTTTTTGGATAAGAAAAAAAAAAAAATTTTTTTT
```

```
AATTTTTGTATACTAAGTTCTAAATTAATAGATAATCTATATATATATAAATGTTTAATATATTATAATATAGTTAACAATTTAAATAAAAAATTTCACA  15700
TTAAAAACATATGATTCAAGATTTAATTATCTATTAGATATATATATATTTACAAATTATATAATATTATATCAATTGTTAAATTTATTTTTTAAAGTGT
     .         .         .         .         .         .         .         .         .         .

ATCCAAAAATGGTAACATAATTTGTAAAAAAAAATCTATATTCAAATATTTATATAACATTCTTGGATTTATATAAATAATATAATAATTTAATTAATTA  15800
TAGGTTTTTACCATTGTATTAAACATTTTTTTTTAGATATAAGTTTATAAATATATTGTAAGAACCTAAATATATTTATTATATTATTAAATTAATTAAT
     .         .         .         .         .         .         .         .         .         .

TTATATATTTATATATTTATATATAGTTGAAGATTTATATTACATATATATCTATATAGAAAAATAAAATTATTTTAATAATTTTTATAAAAATATTTTA  15900
AATATATAAATATATAAATATATATCAACTTCTAAATATAATGTATATATAGATATATCTTTTTATTTTAATAAAATTATTAAAAATATTTTTATAAAAT
     .         .         .         .         .         .         .         .         .         .

TGAATTCCTAAAATGTGTTATCTAATATAAATCAATTAATTTTTAAAAATTGTAATCTTTTAAAAAAAAAATAGTAAATAATAAAAAAAAAAAAAAAAA  16000
ACTTAAGGATTTTACACAATAGATTATATTTAGTTAATTAAAAATTTTTAACATTAGAAAATTTTTTTTTTATCATTTATTATTTTTTTTTTTTTTTT
     .         .         .         .         .         .         .         .         .         .

AGATGAGTTTTTTATTATT
TCTACTCAAAAAATAATAA
```

Fig. 1. The 16,019 nucleotide sequence of the circular *D. yakuba* mtDNA molecule. The numbering shown begins with the first nucleotide of the $tRNA^{ile}$ gene, proceeds through that gene continuously around the entire molecule, and terminates with the last nucleotide of the A + T-rich region adjacent to the $tRNA^{ile}$ gene (see Fig. 2). Transfer RNA genes are boxed and the anticodon of each is underlined. The terminal regions of the two rRNA genes are marked with broken lines above and below. The predicted amino acid sequence (one-letter code) of each protein gene is shown above the sense strand of the gene. An arrow indicates the direction of transcription of each gene. Asterisks indicate partial or complete termination codons. AGA codons are shown corresponding to serine. The "(M)" at the beginning of the COI sequence reflects uncertainty regarding translation initiation of this gene (see text). Cyt b, cytochrome b

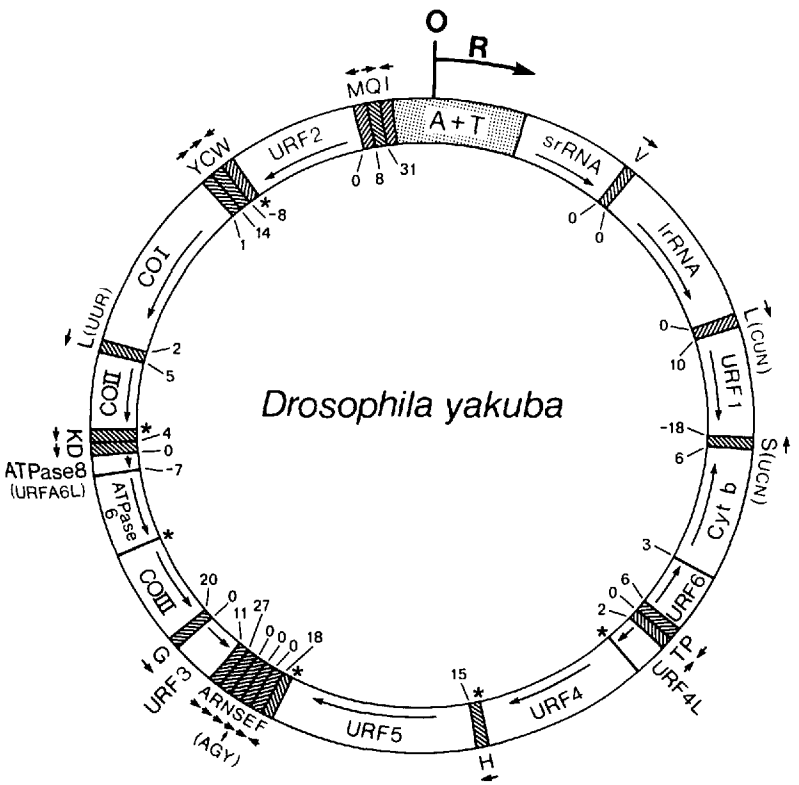


Fig. 2. Gene map of the *D. yakuba* mtDNA molecule. The location of the origin of replication (O) within the A + T-rich region (shaded) and the direction of replication (R) were determined by electron microscope studies (Goddard and Wolstenholme 1980). The identities and arrangement of the various genes were determined from nucleotide sequence studies (Clary et al. 1982, 1983, 1984; Clary and Wolstenholme 1983a,b, 1984a, 1985). Each tRNA gene (hatched areas) is identified by the one-letter amino acid code, and individual serine and leucine tRNA genes are identified by the codon family (in parentheses) that their transcription products recognize. Arrows within and outside the molecule indicate the direction of transcription of each gene. The numbers of apparently noncoding nucleotides that occur between the genes are shown at the gene boundaries on the inner side of the map. Negative numbers indicate overlapping nucleotides of adjacent genes. An asterisk indicates an incomplete termination codon (T or TA)

(87), and cow (57) (Anderson et al. 1981, 1982b; Bibb et al. 1981).

The relative arrangements and transcriptions of the genes within the *D. yakuba* and mouse mtDNA molecules are compared in Fig. 3. Five protein genes, the two rRNA genes, and 11 tRNA genes are arranged differently relative to adjacent genes in the *D. yakuba* and mammalian mtDNA molecules (Fig. 3). One major difference in gene order between the two molecules could have resulted from translocation and inversion of a single segment containing the URF1, small rRNA, $tRNA^{val}$, and large rRNA genes in a molecule ancestral to either the present-day *D. yakuba* or the present-day mouse mtDNA molecule. Alternatively, this difference in gene order could have resulted from a single inversion of the segment containing the URF1, small rRNA, $tRNA^{val}$, and large rRNA genes and the adjacent replication origin-containing region. However, the latter explanation requires that following the inversion there occurred a change in the direction of replication around the molecule in either the line lead-

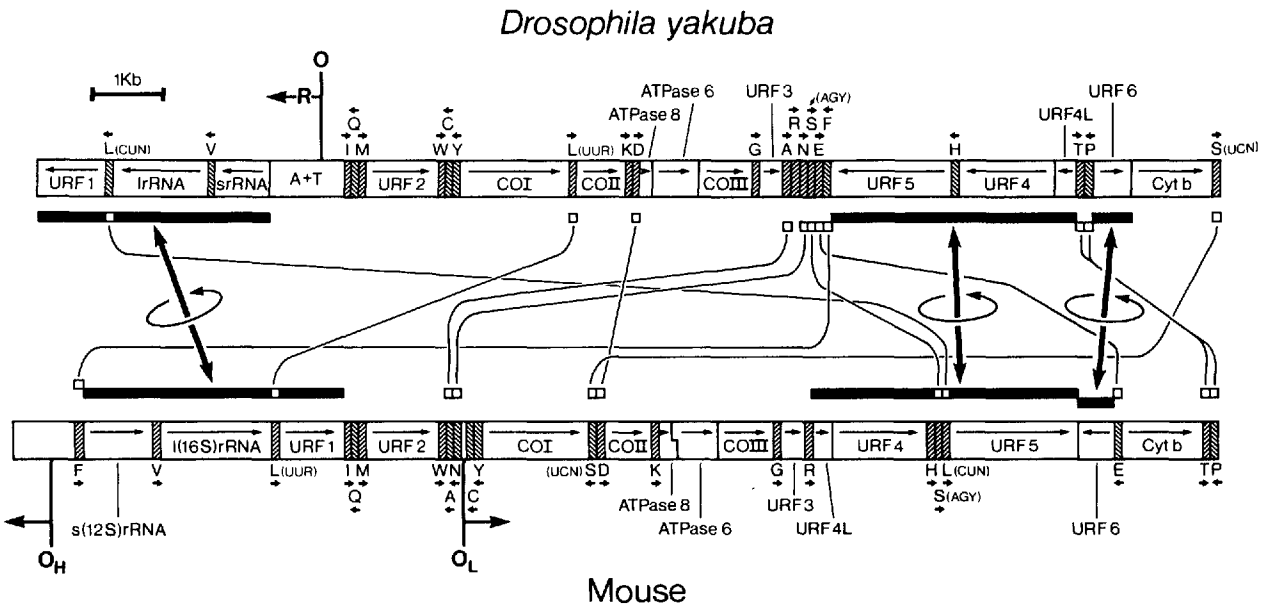


Fig. 3. Comparison of gene arrangements in the mtDNA molecules of *D. yakuba* and mouse. In the diagram the circular molecules have been linearized and aligned at the 3′ ends of the $tRNA^{ile}$ genes. The solid bars linked by two-headed arrows indicate three segments comprising the URF1, $tRNA^{val}$, and the two rRNA genes; the URF4L, URF4, URF5, and $tRNA^{his}$ genes; and the URF6 gene, respectively, each of which has been translocated and/or inverted (circular arrows) between the two molecules. Transfer RNA genes that are in different locations in the two molecules relative to adjacent protein genes are shown as open squares and are connected by lines. In the mouse map [derived from that given by Bibb et al. (1981)], "O_H" and "O_L" indicate the origins of heavy- and light-strand synthesis, respectively. All other details are as given in Fig. 2

ing to *D. yakuba* or that leading to mammals. A further major difference in gene order between *D. yakuba* and mammalian mtDNA molecules seems to have resulted from a single inversion of a segment containing the URF4L, URF4, $tRNA^{his}$, and URF5 genes. An independent inversion involving only the URF6 gene must have occurred in order to explain why this gene is transcribed in opposite directions in the two molecules. Comparison of the *D. yakuba* and mammalian mtDNA gene maps (Fig. 3) indicates that nine tRNA genes have been independently translocated during evolution. Also, the $tRNA^{thr}$ and $tRNA^{pro}$ genes appear to have been translocated as a single segment.

All of the genes in that half of the *D. yakuba* mtDNA molecule to the right of the A + T-rich region, except for cytochrome b, URF6, and two tRNA genes [ser (UCN) and thr], are transcribed in the same direction as that in which replication proceeds around the molecule (Fig. 2). All of the genes in the other half of the molecule, with the exception of three tRNA genes (gly, cys, and tyr), are transcribed in the other direction. In contrast, in mammalian mtDNAs the rRNA genes, 14 tRNA genes, and all of the protein genes except URF6 are transcribed in the same direction, which is opposite to the direction of replication (Anderson et al. 1981, 1982b; Bibb et al. 1981).

There are three cases of genes overlapping in the *D. yakuba* mtDNA molecule. Two of these involve overlapping of the 3′ ends of adjacent genes ($tRNA^{trp}$ and $tRNA^{cys}$ overlap by eight nucleotides; URF1 and $tRNA^{ser}_{UCN}$ by 18 nucleotides; Fig. 1). The third case is the out-of-phase overlapping of the 5′ end of the ATPase6 gene by the 3′-terminal seven nucleotides of the ATPase8 gene. Small overlaps also occur between some genes in mammalian mtDNAs, the largest of which is a 40- to 46-nucleotide overlap of the 3′ end of the ATPase8 gene and the 5′ end of the ATPase6 gene (Anderson et al. 1981, 1982b; Bibb et al. 1981).

Protein Genes

Eleven of the *D. yakuba* mitochondrial protein genes differ in size by less than 3% from the corresponding mouse genes (Table 1). A deficiency in codons in the 5′ end of the *D. yakuba* URF5 gene relative to the mouse URF5 gene is the main component of an overall difference of 6% between these genes.

The open reading frame originally designated URFA6L in mammalian mtDNAs has recently been shown to have amino acid sequence homology and similarity in transcriptional control to the ATPase8 gene of *Saccharomyces cerevisiae* (Macreadie et al. 1983). Because URFA6L in *D. yakuba* mtDNA has considerable nucleotide and amino acid sequence homology to the mouse URFA6L gene (Clary and Wolstenholme 1983a), it has been renamed the

Table 1. Comparisons of protein genes of *D. yakuba* and mouse

Gene	Number of amino acids		% Difference	% Nucleotide sequence homology[a,b] (*D. yakuba*/mouse)	% Amino acid sequence homology[b] (*D. yakuba*/mouse)
	D. yakuba	Mouse			
Cyt b	378	381	0.79	65	67
COI	512	514	0.39	72	75
COII	228	227	0.44	64	56
COIII	262	261	0.38	67	64
ATPase6	224	226	0.89	49	36
ATPase8	53	67	20.90	59	26
URF1	324	315	2.86	52	46
URF2	341	345	1.16	47	35
URF3	117	114	2.63	52	42
URF4L	96	97	1.03	34	40
URF4	446	459	2.83	51	42
URF5	573	607	5.60	42	33
URF6	174	172	1.16	38	17

[a] These values do not include nucleotides concerned with termination

[b] These values include all deduced insertions/deletions. The mouse data are taken from Bibb et al. (1981)

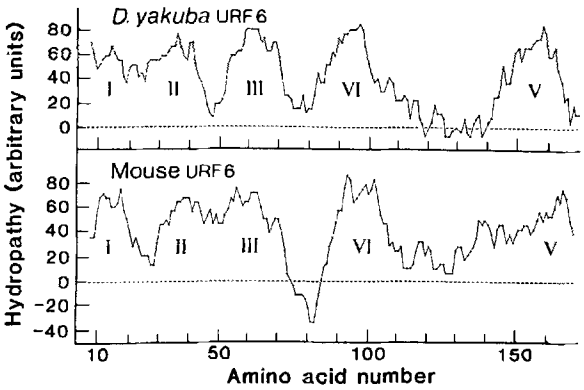


Fig. 4. Comparison of the hydropathy profiles of the proteins predicted from the URF6 genes of *D. yakuba* and mouse. Each profile was calculated by the method of Kyte and Doolittle (1982) using an 11-amino-acid window. Hydrophobic regions have hydropathy of >0, and hydrophilic regions have hydropathy of <0. Corresponding hydrophobic domains in the two polypeptides are indicated as I–V

ATPase8 gene. The ATPase8 genes of *D. yakuba* and mouse show a 20% difference in size, which results mainly from the *D. yakuba* gene lacking the equivalent of a segment of 12 codons at the 3′ end of the mouse ATPase8 gene that overlaps the 5′ end of the ATPase6 gene.

The amino acid sequence homology between the *D. yakuba* and mouse URF6 genes is less than that between any of the other corresponding *D. yakuba* and mouse mitochondrial protein genes. However, the *D. yakuba* and mouse URF6 genes are almost identical in length (Table 1) and there is a striking similarity in the hydropathy profiles of their amino acid sequences (Fig. 4). The latter observation suggests that in spite of extensive divergence in the primary sequences of the *D. yakuba* and mouse URF6 genes, the function of protein products of these genes has been conserved.

Translation Initiation and Termination Signals

An ATG, ATT, or ATA codon occurs at the beginning of all of the *D. yakuba* mitochondrial protein genes except the COI gene. In each case this codon either immediately follows the previous gene or is separated from it by no more than 15 nucleotides. The view that, as in mammalian mtDNAs, ATT and ATA are used in addition to ATG as initiation codons is supported by the findings that the URF1, URF3, URF6, and ATPase8 genes totally lack ATG codons and that the first ATG codon in both the URF2 and URF5 genes occurs near the center of the gene.

In *D. yakuba* mtDNA, as in mammalian mtDNA, only one gene for a tRNA (anticodon CAT) expected to recognize methionine-specifying codons (AUG and AUA) has been located, and there is evidence that in mammals both mt-tRNA^{met} and mt-tRNA$^{f\text{-}met}$ are specified by that gene (discussed in Van Etten et al. 1982).

In both *D. melanogaster* and *D. yakuba* mtDNAs (Fig. 1), the COI gene lacks an ATN translation initiation codon and it has been reasoned that the tetranucleotide ATAA may serve this function (Clary and Wolstenholme 1983b; de Bruijn 1983).

In a small number of prokaryotic genes, GTG functions as a translation initiation codon (Stormo et al. 1982). GTG occurs as the first in-frame triplet of the open reading frames of both the URF5 gene of *D. yakuba* (Fig. 1) and the URF1 gene of mouse (Bibb et al. 1981). Since the first ATN codon in both of these genes is separated from the preceding tRNA gene by a greater number of nucleotides than is found for any other mitochondrial protein gene (15 for *D. yakuba* URF5 and nine for mouse URF1), it seems worth considering the possibility that in these genes GTG serves as a translation initiation codon.

TAA is the only termination codon found in mouse mitochondrial protein genes. Some bovine and human protein genes also end in TAA but others end in TAG or AGA, and AGG appears to be the termination codon of the human URF6 gene. A number of mammalian mitochondrial protein genes end in a T rather than a termination codon, and this

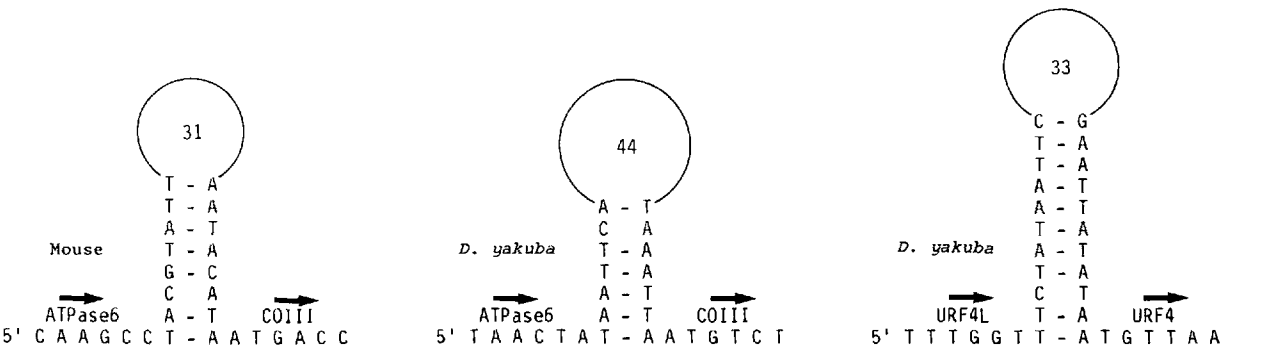


Fig. 5. Potential stem and loop configurations of the nucleotides of the junctional sequences of the ATPase6 and COIII genes of mouse (Bibb et al. 1981) and *D. yakuba*, and of the URF4L and URF4 genes of *D. yakuba*. The number of nucleotides in each loop is shown

T is immediately adjacent to the 5'-terminal nucleotide of the sense strand of a tRNA gene. It has been shown that the primary transcription products of mammalian mtDNAs are multicistronic RNA molecules and that following precise cleavage to yield individual gene transcripts, those protein gene transcripts ending in U acquire a complete termination codon by polyadenylation (Ojala et al. 1980, 1981; Anderson et al. 1981). The only termination codon found in *D. yakuba* mtDNA is TAA (seven genes), although TAG is the termination codon of the *D. melanogaster* COIII gene (Clary et al. 1983). Four *D. yakuba* mitochondrial protein genes end in a T (Fig. 1). It seems plausible that a cleavage–polyadenylation mechanism also operates in *Drosophila* mitochondria, as both polyadenylated mRNAs and mRNAs that could contain multiple gene transcripts have been obtained from *D. melanogaster* and *D. virilis* mitochondria (Spradling et al. 1977; Battey et al. 1979; Merten and Pardue 1981).

The presence or absence of complete termination codons in corresponding mitochondrial protein genes of different mammals and *D. yakuba* is poorly conserved. The only protein genes that have the same terminal nucleotide(s) in mtDNAs of *D. yakuba*, mouse, human, and cow are URF2 and URF4 (a T in each case) and ATPase6 (a TA).

The dinucleotide TA at the end of the *D. yakuba* ATPase6 and URF4L genes is in each case adjacent to the ATG codon of the following protein gene (COIII and URF4, respectively). An identical situation is found for the ATPase6 and COIII genes in mouse and other mammalian mtDNAs, but in these molecules there is a seven-nucleotide out-of-phase overlap between the 3' end of the URF4L gene and the 5' end of the URF4 gene. In HeLa cells there appears to be precise cleavage between the terminal UA of the ATPase6 transcript and the AUG initiation codon of the COIII transcript, but the regulatory signal by which this is facilitated has not been established (Anderson et al. 1981; Ojala et al. 1981). However, as first pointed out by Bibb et al. (1981), the 3' end region of the ATPase6 genes of mouse, human, and bovine mtDNAs has the potential to form a hairpin structure with a stem of eight nucleotide pairs (with one or two mismatches) and a loop of 31 nucleotides, with the 5' terminus of the COIII gene lying at the base of the stem on the 3' side. A similar hairpin can be formed from the 3' end region of the *D. melanogaster* (de Bruijn 1983) and *D. yakuba* (Fig. 5) ATPase6 genes, although in each of these species the stem would be expected to be less stable than in the proposed mammalian structures and the loop is 44 nucleotides. Interestingly, however, the 3' end region of the *D. yakuba* URF4L gene can also form a hairpin structure with an 11-nucleotide-pair stem (with one mismatch) and a 33 nucleotide loop (Fig. 5). These observations are consistent with the view that the potential to form a secondary structure is in some way involved in cleavage site recognition in transcripts of genes with a TAATG junctional sequence.

Ribosomal RNA Genes

Unlike in mammalian mtDNAs, a tRNA gene is not located at the 5' end of the *D. yakuba* small rRNA gene. The 5' end of this gene was located by S1 protection analysis (Clary and Wolstenholme 1985). The 3' end of the small rRNA gene and both the 5' and 3' ends of the large rRNA genes (Fig. 1) have been identified by sequence homologies to the sequenced end regions of mosquito mt-rRNAs (Dubin et al. 1982; HsuChen et al. 1984) and, in the case of the 3' end of the large rRNA gene, from secondary structure comparisons (Clary et al. 1984).

The *D. yakuba* small and large rRNA genes contain only 789 and 1326 nucleotides, respectively, and are the smallest metazoan rRNA genes reported to date. As first shown by Klukas and Dawid (1976) for *D. melanogaster* mt-rRNAs, the *D. yakuba*

small and large mt-rRNA genes are unusually low in G + C content (21% and 17%, respectively). The 3′ 418 nucleotides of the *D. yakuba* small mt-rRNA gene are 60.1% homologous to the corresponding region of the mouse small (12S) mt-rRNA gene. Also, within the 3′ 683 nucleotides of the *D. yakuba* large mt-rRNA gene are two sequences of 272 and 302 nucleotides that are 67.7% and 67.9% homologous, respectively, to correspondingly located sequences in the mouse large (16S) mt-rRNA gene. However, the remaining 5′ portions of the *D. yakuba* small and large mt-rRNA genes are exceptionally low in G + C content (14.3% and 9.5%, respectively) and convincing sequence homologies between these regions and corresponding regions of mouse mt-rRNA genes were found for only a few short segments (Clary and Wolstenholme 1985).

In spite of the small size and low G + C content of the *D. yakuba* mt-rRNA genes, segments throughout the entire lengths of both of these rRNA genes can be folded into secondary structures that show remarkable resemblance to the secondary structures proposed by Zweib et al. (1981) and Glotz et al. (1981) for the corresponding segments of mouse mt-rRNA genes. Also, the few convincing homologous sequences in the 5′ region of each of the corresponding *D. yakuba* and mouse rRNA genes occur at similar structural locations. Of particular interest is that many of the predicted helices in both *D. yakuba* genes, particularly in the 5′ regions, are composed either totally of A–T pairs or of mixtures of A–T and G·T pairs (Clary and Wolstenholme 1985).

Differences in size between corresponding *D. yakuba* and mouse mt-rRNA genes result from the absence in the *D. yakuba* genes of specific blocks of contiguous nucleotides that in the mouse genes form distinct loops, or stem and loop structures. These findings are consistent with data from comparisons of rRNA genes from a variety of organelle and nonorganelle sources that indicate that variations in size are accounted for by deletion/insertion of blocks of contiguous nucleotides (Chan et al. 1983; Ware et al. 1983; Woese et al. 1983; Clark et al. 1984; Gray et al. 1984).

The high degree of conservation of secondary structure in the *D. yakuba* and mouse mt-rRNA genes, particularly in the regions of low sequence homology, provides strong support in favor of distinct functional significance for each of the component stems, loops, and interconnecting regions that are conserved in the secondary structure models of these genes. The secondary structures of the mouse mt-rRNA genes were based on a general model of secondary structure derived from comparisons of various organelle and nonorganelle rRNA genes (Woese et al. 1983; Noller 1984). The secondary structures proposed for *Drosophila* rRNA genes fit the general model and in turn contribute to its validity.

Transfer RNA Genes

Interpretation of sequences that lie between protein genes as tRNA genes is based primarily on the possibility of folding these sequences into the characteristic secondary structure of tRNAs. Further strong support for the interpretation of these sequences as tRNA genes is the finding that six of them [ile, lys, arg, ser (AGY), asp, and f-met] are between 84% and 96% homologous to corresponding mitochondrial tRNAs (mt-tRNAs) from mosquito (*Aedes albopictus*) tissue culture cells (HsuChen et al. 1983a, b; Dubin and HsuChen 1984; Dubin et al. 1984). *D. yakuba* mt-tRNA genes are between 32% ($tRNA^{arg}$) and 70% ($tRNA^{asp}$, $tRNA^{thr}$) homologous to their mouse counterparts.

The general structure shared by 21 of the *D. yakuba* mt-tRNA genes resembles that shared by the 21 corresponding mammalian mt-tRNA genes (Barrell et al. 1979; Crews and Attardi 1980; Van Etten et al. 1980, 1982; Anderson et al. 1982a; Roe et al. 1982). The *D. yakuba* mt-tRNA genes vary in size from 63 nucleotides ($tRNA^{cys}$) to 72 nucleotides ($tRNA^{val}$). There is strict conservation of the size of the aminoacyl stem (seven nucleotide pairs), the anticodon stem (five nucleotide pairs), and the anticodon loop (seven nucleotides). The stem of the dihydrouridine arm is either three or four nucleotide pairs and the loop is between three and eight nucleotides. The variable loop that lies between the anticodon and TψC arms is always four or five nucleotides. Within the TψC arm, the stem varies in length from four to five nucleotide pairs and the loop from three to eight nucleotides. The trinucleotide sequence CCA, which occurs at the 3′ end of prokaryotic and eukaryotic nuclear-encoded tRNA genes, is absent in this position from the *D. yakuba* mt-tRNA genes, as it is from mammalian mt-tRNA genes.

There is considerable variation among *D. yakuba* mt-tRNA genes in regard to the presence of various nucleotides that are conserved in prokaryotic and eukaryotic nuclear-encoded tRNAs (see Sprinzl and Gauss 1984a,b). Only the conserved T_{33} and Pu_{37} nucleotides that lie immediately 5′ and 3′, respectively, to the anticodons are found in all the *D. yakuba* mt-tRNA genes. Pu_{26} is found in all but the *D. yakuba* $tRNA^{glu}$ gene, where it is a C. Also, the conserved nucleotide pair Py_{11}–Pu_{24} occurs in all *D. yakuba* mt-tRNA genes except $tRNA^{f\text{-}met}$ and $tRNA^{trp}$, where a G–C pair is found in this position.

Some of the conserved nucleotides in the dihydrouridine and TψC arms of prokaryotic and eukaryotic nuclear-encoded tRNAs are involved in

tertiary bonding (Kim 1979). The lack of conservation of nucleotides in the dihydrouridine and TψC arms in *D. yakuba* mt-tRNA genes indicates that in the tRNAs transcribed from these genes, tertiary bonding is different and possibly weaker than in prokaryotic and eukaryotic nuclear-encoded tRNAs. Nucleotides that are conserved in the dihydrouridine and TψC arms of eukaryotic nuclear-encoded tRNAs have been shown to function as recognition sites in the control of transcription of individual tRNA genes (see Rajput et al. 1982 for references). Little is known concerning transcriptional control of *D. yakuba* mtDNA. However, as noted above, in mammalian mtDNAs transcripts of individual tRNAs (in which the nucleotide contents of dihydrouridine and TψC arms are also variable) are cleaved from primary multicistronic transcripts that are initiated in or close to the heavy-strand replication origin-containing region of the molecule (Montoya et al. 1982, 1983; Bogenhagen et al. 1984; Chang and Clayton 1984). Thus in mammalian mtDNAs, at least, the lack of conservation of nucleotides in the dihydrouridine and TψC arms of the tRNA genes can be correlated with the absence of primary transcripts of individual tRNA genes.

The $tRNA^{ser}_{AGY}$ encoded by *D. yakuba* mtDNA is structurally different from the other tRNAs encoded in this genome (Clary and Wolstenholme 1984a). The dihydrouridine arm is replaced by an 11-nucleotide loop within which secondary structure formation seems unlikely. Also, both the variable loop (six nucleotides) and the TψC loop (nine nucleotides) are larger than in other *D. yakuba* mt-tRNAs. The identification of the *D. yakuba* mt-$tRNA^{ser}_{AGY}$ gene is supported by the isolation of a mt-tRNA from mosquito tissue culture cells that is 91% homologous to the *D. yakuba* mt-$tRNA^{ser}_{AGY}$ gene, and can be folded into a structure similar to the one we have proposed, with a GCU anticodon (Dubin et al. 1984). Mammalian mtDNAs also contain unusual $tRNA^{ser}_{AGY}$ genes, in which the dihydrouridine arm is replaced by a five-nucleotide loop, and the corresponding tRNAs from bovine and human mitochondria have been identified (Arcari and Brownlee 1980; de Bruijn et al. 1980). Proposed tertiary interactions between nucleotides of the bovine mt-$tRNA^{ser}_{AGY}$ (de Bruijn and Klug 1983) could also occur in the tRNA predicted from the *D. yakuba* mt-$tRNA^{ser}_{AGY}$ gene (Clary and Wolstenholme 1984a).

Codon Usage and the Genetic Code

Codon usage among the 13 protein genes of *D. yakuba* mtDNA is shown in Table 2; 93.8% of all codons end in A or T. Among the individual protein genes the percentage of codons ending in A or T ranges from 91.6% for the COI gene to 98.1% for the ATPase8 gene.

In spite of the very high use of codons ending in A or T, all codons except CAG, CGC, AGG, and TAG are found among the protein genes, and it has been shown, as mentioned above, that TAG is utilized as the termination codon of the *D. melanogaster* COIII gene (Clary et al. 1983). Peculiar to *Drosophila* mtDNA is the use of AGA codons to specify serine. The triplets AGA and AGG specify arginine in the standard genetic code and are used only as rare termination codons (in human and bovine) or not at all (in mouse) in mammalian mtDNA. In *D. yakuba* mtDNA, internal AGA codons (a total of 73) are present in all protein genes except URF6. It is clear that these AGA codons do not specify arginine. None of the AGA codons in the *D. yakuba* mitochondrial protein genes correspond in position to arginine-specifying codons (CGN) in the equivalent genes of mouse mtDNAs. Also, none of the AGA codons of the *D. yakuba* cytochrome b, COI, COII, COIII, and ATPase6 genes correspond to arginine-specifying codons (AGA) in the equivalent mitochondrial genes of yeast (Clary and Wolstenholme 1983b).

Three related observations support the view that AGA specifies serine. Of the seven AGA codons found in the *D. yakuba* COI gene (which has the greatest homology to an equivalent mouse gene; Table 1) five correspond in position to serine-specifying codons in the mouse COI gene. Also, five of these seven *D. yakuba* AGA codons correspond in position to serine-specifying codons in the yeast COI gene (Clary and Wolstenholme 1983b). Although the 73 AGA codons in *D. yakuba* mtDNA correspond in position to codons specifying 14 different amino acids in the equivalent mouse genes, AGA codons correspond to more than twice as many serine-specifying codons (23 total) than to codons specifying any other amino acid (10 for alanine). Similar arguments have been made by de Bruijn (1983) for corresponding nucleotide sequences of four complete and two partial genes of mtDNAs of *D. melanogaster* and human. The third and strongest line of evidence that AGA specifies serine in the *D. yakuba* mitochondrial genetic code was obtained from consideration of frequencies of nucleotide substitutions between AGA codons and AGT codons in six mitochondrial protein genes (URF2, COI, COII, COIII, ATPase6, and ATPase8) of *D. yakuba* and *D. melanogaster* (Wolstenholme and Clary 1985). As in all other known genetic codes, AGT is expected to specify serine in the *Drosophila* mitochondrial genetic code. The frequency of third-position substitutions between AGA and AGT codons was 21.2%, which was close to the mean frequency (20.6%) of A ↔ T third-position substitu-

Table 2. Codon usage in the 13 protein genes of *D. yakuba* mtDNA

Phe	TTT	313	(GAA)	Ser	TCT	120	(UGA)	Tyr	TAT	142	(GUA)	Cys	TGT	40	(GCA)
	TTC	17			TCC	4			TAC	28			TGC	2	
Leu	TTA	542	(UAA)		TCA	102		TER	TAA	7		Trp	TGA	96	(UCA)
	TTG	25			TCG	3			TAG	0			TGG	6	
Leu	CTT	36		Pro	CCT	79		His	CAT	65	(GUG)	Arg	CGT	8	
	CTC	2	(UAG)		CCC	3	(UGG)		CAC	12			CGC	0	(UCG)
	CTA	19			CCA	45		Gln	CAA	70	(UUG)		CGA	45	
	CTG	2			CCG	3			CAG	0			CGG	6	
Ile	ATT	345	(GAU)	Thr	ACT	97	(UGU)	Asn	AAT	193	(GUU)	Ser	AGT	34	(GCU)
	ATC	15			ACC	3			AAC	13			AGC	1	
Met	ATA	195	(CAU)		ACA	85		Lys	AAA	76	(CUU)		AGA	73	
	ATG	18			ACG	2			AAG	9			AGG	0	
Val	GTT	90		Ala	GCT	125	(UGC)	Asp	GAT	54	(GUC)	Gly	GGT	67	(UCC)
	GTC	3	(UAC)		GCC	9			GAC	10			GGC	2	
	GTA	93			GCA	37		Glu	GAA	82	(UUC)		GGA	129	
	GTG	8			GCG	2			GAG	1			GGG	22	

Table shows the number of occurrences of each codon. The total number of codons in the 13 *D. yakuba* protein genes is 3735, which includes 7 TAA codons. TGA, AGA, and ATA are assumed to specify tryptophan, serine, and methionine, respectively (see text). AGG, CAG, CGC, and TAG have not been found in *D. yakuba* mtDNA (Clary et al. 1984). The anticodon corresponding to each two- and four-codon family is shown in parentheses

tions that did not result in an amino acid replacement (silent substitutions) among codons in the six corresponding genes. If in *Drosophila* mtDNA AGA coded for a different amino acid than AGT, then a frequency of third-position A ↔ T substitutions between AGA and AGT codons of approximately 0.5% would have been expected; that value is the mean frequency found for third-position replacement A ↔ T substitutions among all codons in the six corresponding genes of *D. yakuba* and *D. melanogaster*. These data indicate, therefore, that AGA and AGT specify the same amino acid.

Similarly, the frequencies of interchange of ATA and ATG codons in corresponding *D. yakuba* and *D. melanogaster* genes are fully consistent with the interpretation that in the *Drosophila* mitochondrial genetic code, as in the mammalian and fungal mitochondrial genetic codes (Barrell et al. 1979, 1980; Hudspeth et al. 1982), ATA specifies methionine rather than isoleucine (Wolstenholme and Clary 1985).

Internal TGA codons are found in all 13 protein genes of *D. yakuba* mtDNA. Of a total of 96 TGA codons, 70 correspond in position to tryptophan-specifying codons (TGA or TGG) in mouse mitochondrial genes, indicating that in *D. yakuba* mtDNA, as in mammalian and fungal mtDNAs (Barrell et al. 1979, 1980; Fox 1979; Bonitz et al. 1980; Heckman et al. 1980), TGA specifies tryptophan.

Codon–Anticodon Interactions

The 22 tRNAs encoded by mammalian and *Drosophila* mtDNAs appear to be sufficient to decode the mitochondrial protein genes (Barrell et al. 1980; Anderson et al. 1981, 1982b; Bibb et al. 1981; Clary et al. 1984). When all four codons in a box of the genetic code (Table 2) specify the same amino acid (four-codon families) these codons are recognized by a single tRNA. Codons of each of the two-codon families are also recognized by a single tRNA. Transfer RNAs that recognize codons of four-codon families have a U in the wobble position. However, $tRNA^{f\text{-}met}$ contains a CAU anticodon. It appears that this anticodon can recognize both AUG and AUA as methionine-specifying codons when they occur internally, and all four AUN codons as specifying methionine when they occur as initiation codons (Anderson et al. 1981; 1982b; Bibb et al. 1981). It is not known whether ATT or ATC initiation codons specify isoleucine or methionine, since no metazoan mitochondrial protein has been sequenced that is encoded by a gene beginning with these codons.

A gene for a tRNA with a UCU anticodon, which would specifically recognize AGA codons, has not been identified in *D. yakuba* mtDNA. It seems reasonable to assume therefore that the GCU anticodon of $tRNA^{ser}_{AGY}$, which is expected to recognize AGU and AGC codons, also recognizes AGA codons. This would necessitate either selective two out of three nucleotide pair recognition (Lagerkvist 1981) or effective pairing of the G in the wobble position of the anticodon with C, U, or A.

In both mammalian and *D. yakuba* mtDNAs, for codons of two-codon families ending in a pyrimidine, the corresponding tRNA has a G in the wobble position (Table 2), whereas for those ending in a purine the corresponding tRNA has a U in the wob-

ble position. There is one exception to this latter rule in *Drosophila* mtDNA. Although protein genes contain both AAA and AGA codons, only a $tRNA^{lys}$ gene wtih a CTT anticodon has been located (Clary and Wolstenholme 1983a; de Bruijn 1983). Thus if all three nucleotides in the anticodons of this tRNA and $tRNA^{f\text{-}met}$ are contributing to codon–anticodon recognition, then C–A pairing is indicated. In this regard it is interesting to note that the C in the wobble position was found to be unmodified in the $tRNA^{f\text{-}met}$ of bovine mitochondria (Roe et al. 1982) and in both a $tRNA^{f\text{-}met}$ and a $tRNA^{lys}$ isolated from mitochondria of mosquito tissue culture cells (HsuChen et al. 1983b; Dubin and HsuChen 1984).

Replication Origin-Containing Regions

Replication of the *D. yakuba* mtDNA molecule has been shown to originate in the half of the A + T-rich region closest to the $tRNA^{ile}$ gene and to proceed unidirectionally toward the rRNA genes (Figs. 1 and 2; Goddard and Wolstenholme 1980). However, the nucleotide location at which nascent DNA synthesis begins has not been determined. The A + T-rich region lacks open reading frames longer than 123 nucleotides. Also, RNA–DNA hybridization studies have failed to demonstrate a transcript complementary to any portion of the A + T-rich region (Bonner et al. 1978; Battey et al. 1979; Merten and Pardue 1981).

In both *Drosophila* and mammalian mtDNAs replication of the complementary strands is highly asymmetrical (Robberson et al. 1972; Wolstenholme et al. 1974; Goddard and Wolstenholme 1978, 1980). In mtDNAs of cultured mammalian cells replication begins with synthesis of the heavy-strand, which proceeds for two-thirds of the molecule length before synthesis on the second (light) strand commences (Robberson et al. 1972; Kasamatsu and Vinograd 1974; Nass 1980). Initiation of light-strand synthesis in all mouse L cell and most human KB cell mtDNA molecules occurs within a specific sequence of 32 (mouse) or 31 (human) nucleotides that has the potential to fold into a hairpin loop and is located between the $tRNA^{asn}$ and $tRNA^{cys}$ genes (see Clayton 1982). This 31- or 32-nucleotide sequence is highly conserved in other mammalian and amphibian mtDNAs (Anderson et al. 1982b; Taira et al. 1983; Wong et al. 1983). However, such a sequence is not found in the corresponding region of *D. melanogaster* and *D. yakuba* mtDNA (Clary and Wolstenholme 1983b; de Bruijn 1983) or in any other intergenic region of the *D. yakuba* mtDNA molecule. From data obtained by electron microscopy evidence for a highly preferred site of initiation of second-strand synthesis in *D. melanogaster* mtDNA molecules was found (Goddard and Wolstenholme 1978), but this site maps close to the expected boundary of the $tRNA^{ile}$ gene and the A + T-rich region of the molecule.

Nucleotide Composition

Drosophila mtDNAs are distinguished by a high (74–80%) A + T content (Bultmann and Laird 1973; Polan et al. 1973; Peacock et al. 1974; Fauron and Wolstenholme 1976). The average nucleotide compositions of the sense strands of the protein genes encoded by *D. yakuba,* mouse, and human mtDNAs are compared in Table 3. The frequency of G and A nucleotides is similar in all three species, but *D. yakuba* mtDNA has a much higher T and much lower C content than mouse and human mtDNAs. Most of these differences clearly result from different frequencies of nucleotides in the third position of codons (Table 3). Of codons in *D. yakuba* mitochondrial protein genes, 48.4% end in T and 45% end in A. In the third position of codons in human and mouse mitochondrial protein genes, C and A nucleotides are the most frequent, but Cs predominate (43.1%) in human and As predominate (46.6%) in mouse. In *D. yakuba,* mouse, and human mtDNA-encoded proteins, leucine accounts for 16.8%, 15.5%, and 17.3%, respectively, of all amino acids. However, there are striking differences among the three species in regard to the use of TTPu and CTN codons that correlate with the differential use of T and C nucleotides in the third position of codons (Table 3).

Both tRNA genes and rRNA genes in the *D. yakuba* mtDNA molecule have high A + T contents. The 22 tRNA genes range from 66% A + T (in $tRNA^{ile}$) to 91% A + T (in $tRNA^{asp}$ and $tRNA^{glu}$), with a mean of 76%. Also, the dihydrouridine loops and the TψC loops, the primary sequence functions of which may have been lost, average 88.9% A + T. As discussed above, large portions of the *D. yakuba* small and large mt-rRNA genes are extremely A + T-rich, particularly in the 5′ regions. Although these nucleotide sequences show little homology to sequences of corresponding regions of mouse mtDNA, there is a high degree of conservation of secondary structure. The apparently nontranscribed, replication origin-containing region of the *D. yakuba* mtDNA molecule has a higher A + T content (92.8%) than any other region of the molecule.

The above observations are consistent with the suggestion that there has been selection in the *D. yakuba* mitochondrial genome for A + T nucleotides in all positions where primary nucleotide sequence is of secondary or little importance. We have obtained strong evidence in further support of this hypothesis from analyses of patterns of substitution

Table 3. Nucleotide composition data for the protein genes of *D. yakuba*, mouse, and human mtDNAs[a]

	Percentage nucleotide composition of the sense strand				Percentage of codons ending in:				Percentage of codons that specify leucine	Percentage of leucine codons beginning with:	
	T	C	A	G	T	C	A	G		T	C
D. yakuba	44.4	11.1	32.3	12.2	48.4	3.3	45.4	2.9	16.8	90.6	9.4
Mouse[b]	29.4	26.0	33.2	11.4	22.9	27.4	46.6	3.2	15.5	21.9	78.1
Human[c]	25.6	33.1	29.3	12.0	15.3	43.1	36.5	5.1	17.3	12.0	88.0

[a] Data for mammalian mtDNAs are derived from all protein genes except URF6. The URF6 gene in these mtDNAs is the only gene transcribed from the light strand, and both the nucleotide composition of this strand and the frequencies of nucleotides in the third positions of codons are complementary to those of the other 12 genes. The *D. yakuba* data are derived from all protein genes, as there is no clear distinction between the nucleotide compositions of the sense strands of genes transcribed from the two complementary strands

[b] Data from Bibb et al. (1981)

[c] Data from Anderson et al. (1981)

in protein-coding genes in corresponding segments of *D. yakuba* and *D. melanogaster* mtDNA molecules. The observed frequency of third-position silent A ↔ T substitutions was too high to be accounted for simply by the high frequency of codons ending in A or T. However, it could not be determined whether the anomalously high frequency of A ↔ T substitutions was due to a high A ↔ T mutation rate or to selection in favor of fixation of the products of A ↔ T mutations. We have suggested as a possible basis for selection in favor of A + T nucleotides in *Drosophila* mtDNAs that once DNA has become A + T rich during long-term evolution (for reasons that are obscure), the various enzymes responsible for transcription and replication of the DNA become adapted to function optimally on that DNA, and therefore less than optimally on a more G + C-rich DNA (Wolstenholme and Clary 1985).

Acknowledgments. We thank John F. Atkins, W. Joseph Dickinson, Raymond F. Gesteland, John R. Roth, Martin C. Rechsteiner, John S. Parkinson, and Jill A. Wahleithner for discussions during the course of this work, and Felix Machuca for excellent technical assistance. This work was supported by National Institutes of Health Grants GM 18375 and RR 07092. Douglas O. Clary's contribution was carried out to fulfill partially the requirements for the degree of Bachelor of Science (Biology) University of Utah.

References

Altman PL, Katz DD (eds) (1976) Biological handbooks I: cell biology. Federation of American Societies for Experimental Biology, Bethesda, Maryland, pp 217–219

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1982a) Comparison of the human and bovine mitochondrial genomes. In: Slonimski P, Borst P, Attardi G (eds) Mitochondrial genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 5–43

Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG (1982b) The complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717

Arcari P, Brownlee GG (1980) The nucleotide sequence of a small (3S) seryl-tRNA (anticodon GCU) from beef heart mitochondria. Nucleic Acids Res 8:5207–5212

Barrell BG, Bankier AT, Drouin J (1979) A different genetic code in human mitochondria. Nature 282: 189–194

Barrell BG, Anderson S, Bankier AT, de Bruijn MHL, Chen E, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1980) Different pattern of codon recognition by mammalian mitochondrial tRNAs. Proc Natl Acad Sci USA 77:3164–3166

Battey J, Rubenstein JLR, Clayton DA (1979) Transcription patterns of *Drosophila melanogaster* mitochondrial DNA. In: Cummings DJ, Dawid IB, Borst P, Weissman SM, Fox CF (eds) Extrachromosomal DNA. Academic Press, New York, pp 427–442 (ICN–UCLA symposia on molecular and cellular biology, vol XV)

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180

Bogenhagen DF, Applegate EF, Yoza BK (1984) Identification of a promoter for transcription of the heavy strand of human mtDNA: *in vitro* transcription and deletion mutagenesis. Cell 36:1105–1113

Bonitz SG, Berlani R, Coruzzi G, Li M, Macino G, Nobrega FG, Nobrega MP, Thalenfeld BE, Tzagoloff A (1980) Codon recognition rules in yeast mitochondria. Proc Natl Acad Sci USA 77:3167–3170

Bonner JJ, Berninger M, Pardue ML (1978) Transcription of polytene chromosomes and of the mitochondrial genome in *Drosophila melanogaster*. Cold Spring Harbor Symp Quant Biol 42:803–814

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences in primates: tempo and mode of evolution. J Mol Evol 18:225–239

Brutlag DL, Clayton J, Frieland P, Kedes LH (1982) SEQ: a nucleotide sequence analysis and recombination system. Nucleic Acids Res 10:279–304

Bultmann H, Laird CD (1973) Mitochondrial DNA from *Drosophila melanogaster*. Biochim Biophys Acta 299:196–209

Chan Y-L, Olvera J, Wood IG (1983) The structure of rat 28S

ribosomal ribonucleic acid inferred from the sequence of nucleotides in a gene. Nucleic Acids Res 11:7819–7830

Chang DD, Clayton DA (1984) Precise identification of mitochondrial promoters for transcription of each strand of human mitochondrial DNA. Cell 36:635–643

Chomyn A, Mariottini P, Gonzalez-Cadavid N, Attardi G, Strong DD, Trovato D, Riley M, Doolittle R (1983) Identification of the polypeptides encoded in the ATPase 6 gene and in the unassigned reading frames 1 and 3 of human mtDNA. Proc Natl Acad Sci USA 80:5535–5539

Clark CG, Tague BW, Ware VC, Gerbi SA (1984) *Xenopus laevis* 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res 12:6197–6220

Clary DO, Wolstenholme DR (1983a) 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. Nucleic Acids Res 11:4211–4227

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

Clary DO, Wolstenholme DR (1984a) A cluster of six tRNA genes in *Drosophila* mitochondrial DNA that includes a gene for an unusual $tRNA^{ser}_{AGY}$. Nucleic Acids Res 12:2367–2379

Clary DO, Wolstenholme DR (1984b) The *Drosophila* mitochondrial genome. In: Maclean N (ed) Oxford surveys on eukaryotic genes, vol 1. Oxford University Press, Oxford, pp 1–35

Clary DO, Wolstenholme DR (1985) The ribosomal genes of *Drosophila* mitochondrial DNA. Nucleic Acids Res 113:4029–4045

Clary DO, Goddard JM, Martin SC, Fauron CMR, Wolstenholme DR (1982) *Drosophila* mitochondrial DNA: a novel gene order. Nucleic Acids Res 10:6619–6637

Clary DO, Wahleithner JA, Wolstenholme DR (1983) Transfer RNA genes in *Drosophila* mitochondrial DNA: related 5′ flanking sequences and comparisons to mammalian mitochondrial tRNA genes. Nucleic Acids Res 11:2411–2425

Clary DO, Wahleithner JA, Wolstenholme DR (1984) Sequence and arrangement of the genes for cytochrome b, URF1, URF4L, URF4, URF5, URF6 and five tRNAs in *Drosophila* mitochondrial DNA. Nucleic Acids Res 12:3747–3762

Clayton DA (1982) Replication of animal mitochondrial DNA. Cell 28:693–705

Crews S, Attardi G (1980) The sequences of the small ribosomal RNA gene and the phenylalanine tRNA gene are joined end to end in human mitochondrial DNA. Cell 19:775–784

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

de Bruijn MHL, Klug A (1983) A model for the tertiary structure of mammalian mitochondrial transfer RNAs lacking the entire dihydrouridine loop and stem. EMBO J 2:1309–1321

de Bruijn MHL, Schreier PH, Eperon IC, Barrell BG, Chen EY, Armstrong PW, Wong JFH, Roe BA (1980) A mammalian mitochondrial serine transfer RNA lacking the "dihydrouridine" loop and stem. Nucleic Acids Res 8:5213–5222

Doolittle RF (1981) Similar amino acid sequences: chance or common ancestry. Science 214:149–159

Dubin DT, HsuChen CC (1984) Sequence and structure of a methionine transfer RNA from mosquito mitochondria. Nucleic Acids Res 12:4185–4189

Dubin DT, HsuChen CC, Timko KD, Azzolina TM, Prince DL, Ranzini JL (1982) 3′ termini of mammalian and insect mitochondrial rRNAs. In: Slonimski P, Borst P, Attardi G (eds) Mitochondrial genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 89–98

Dubin DT, HsuChen CC, Cleaves GR, Timko KD (1984) Sequence and structure of a serine transfer RNA with GCU anticodon from mosquito mitochondria. J Mol Biol 176:251–260

Fauron CMR, Wolstenholme DR (1976) Structural heterogeneity of mitochondrial DNA molecules within the genus *Drosophila*. Proc Natl Acad Sci USA 73:3623–3627

Fauron CMR, Wolstenholme DR (1980) Extensive diversity among *Drosophila* species with respect to nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules. Nucleic Acids Res 8:2439–2452

Fox TD (1979) Five TGA "stop" codons occur within the translated sequence of the yeast mitochondrial gene for cytochrome *c* oxidase subunit II. Proc Natl Acad Sci USA 76: 6534–6538

Glotz C, Zweib C, Brimacombe R (1981) Secondary structure of the large subunit ribosomal RNA from *Escherichia coli, Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. Nucleic Acids Res 9:3287–3306

Goddard JM, Wolstenholme DR (1978) Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*. Proc Natl Acad Sci USA 75:3886–3890

Goddard JM, Wolstenholme DR (1980) Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. Nucleic Acids Res 8:741–757

Gray MW, Sankoff D, Cedergren RJ (1984) On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucleic Acids Res 12:5837–5852

Gronenborn B, Messing J (1978) Methylation of single stranded DNA *in vitro* introduces new restriction endonuclease cleavage sites. Nature 272:375–377

Grosskopf R, Feldmann H (1981) Analysis of a DNA segment from rat liver mitochondria containing the genes for cytochrome oxidase subunits I, II and III, ATPase6 and several tRNA genes. Curr Genet 4:151–158

Heckman JE, Sarnoff J, Alzner-De Weerd B, Yin S, RajBhandary UL (1980) Novel features in the genetic code and codon reading patterns in *Neurospora crassa* mitochondria based on sequences of six mitochondrial tRNAs. Proc Natl Acad Sci USA 77:3159–3163

Hong GF (1982) A systematic DNA sequencing strategy. J Mol Biol 158:539–549

HsuChen CC, Cleaves GR, Dubin DT (1983a) Sequences of three transfer RNAs from mosquito mitochondria. Plasmid 10:55–65

HsuChen CC, Cleaves GR, Dubin DT (1983b) A major lysine tRNA with a CUU anticodon in insect mitochondria. Nucleic Acids Res 11:8659–8662

HsuChen CC, Kotin RM, Dubin DT (1984) Sequence of the coding and flanking regions of the large ribosomal subunit RNA gene of mosquito mitochondria. Nucleic Acids Res 12: 7771–7785

Hudspeth MES, Ainley WM, Shumard DS, Butow RA, Grossman LI (1982) Location and structure of the varl gene of yeast mitochondrial DNA: nucleotide sequence of the 40.0 allele. Cell 30:617–626

Jue RA, Woodbury NW, Doolittle RF (1980) Sequence homologies among *E. coli* ribosomal proteins: evidence for evolutionarily related groupings and internal duplications. J Mol Evol 15:129–148

Kasamatsu H, Vinograd J (1974) Replication of circular DNA in eukaryotic cells. Annu Rev Biochem 43:695–719

Kim SH (1979) Crystal structure of yeast $tRNA^{phe}$ and general structural features of other tRNAs. In: Schimmel PR, Soll D, Abelson JN (eds) Transfer RNA: structure, properties and recognition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 83–100

Klukas CK, Dawid IB (1976) Characterization and mapping of mitochondrial ribsomal RNA and mitochondrial DNA in *Drosophila melanogaster*. Cell 9:615–625

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:333–348

Lagerkvist U (1981) Unorthodox codon reading and the evolution of the genetic code. Cell 23:305–306

Macreadie IG, Novitski CE, Maxwell RJ, John LL, Goi B-G, McMullan L, Lukins HB, Linnane AW, Nagley P (1983) Biogenesis of mitochondria: the mitochondrial gene (aap1) coding for mitochondrial ATPase subunit 8 in *Saccharomyces cerevisiae*. Nucleic Acids Res 11:4435–4451

Mariottini P, Chomyn A, Attardi G (1983) Antibodies against synthetic peptides reveal that the unidentified reading frame A6L, overlapping the ATPase 6 gene, is expressed in human mitochondria. Cell 32:1269–1277

Merten SH, Pardue ML (1981) Mitochondrial DNA in *Drosophila*. An analysis of genome organization and transcription in *Drosophila melanogaster* and *Drosophila virilis*. J Mol Biol 153:1–23

Messing J, Vieira J (1982) A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Gene 19:269–276

Michael NL, Rothbard JB, Shiurba RA, Linke HK, Schoolnik GK, Clayton DA (1984) All eight unassigned reading frames of mouse mitochondrial DNA are expressed. EMBO J 3:3165–3175

Montoya J, Ojala D, Attardi G (1981) Distictive features of the 5′ terminal sequences of the human mitochondrial mRNAs. Nature 290:465–470

Montoya J, Christianson T, Levens D, Rabinowitz M, Attardi G (1982) Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proc Natl Acad Sci USA 79:7195–7199

Montoya J, Gaines GL, Attardi G (1983) The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell 34:151–159

Nass MMK (1980) Pulse-label analysis and mapping of the two terminal regions of asynchronous complementary strand replication of mitochondrial DNA in transformed hamster cells. J Mol Biol 140:257–281

Noller HF (1984) Structure of ribosomal RNA. Annu Rev Biochem 53:119–162

Ohi S, Ramirez JL, Upholt WB, Dawid IB (1978) Mapping of mitochondrial 4S RNA genes in *Xenopus laevis* by electron microscopy. J Mol Biol 121:299–310

Ojala D, Merkel C, Gelfand R, Attardi G (1980) The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell 22:393–403

Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474

Peacock WJ, Brutlag D, Goldring E, Appels R, Hinton CW, Lindsley DL (1974) The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. Cold Spring Harbor Symp Quant Biol 38:405–416

Pepe G, Holtrop M, Gadaleta G, Kroon AM, Cantatore P, Gallerani R, De Benedetto C, Quagliariello E, Sbisa E, Saccone C (1983) Non random patterns of nucleotide substitutions and codon strategy in the mammalian mitochondrial genes coding for identified and unidentified reading frames. Biochem Int 6:553–563

Polan ML, Friedman S, Gall JG, Gehring W (1973) Isolation and characterization of mitochondrial DNA from *Drosophila melanogaster*. J Cell Biol 56:580–589

Rajput B, Duncan L, DeMille D, Miller RC Jr, Spiegelman G (1982) Transcription of cloned transfer RNA genes from *Drosophila melanogaster* in a homologous cell free extract. Nucleic Acids Res 10:6541–6550

Ramirez JL, Dawid IB (1978) Mapping of mitochondrial DNA in *Xenopus laevis* and *X. borealis*: the positions of ribosomal genes and D-loops. J Mol Biol 119:133–146

Rastl E, Dawid IB (1978) Expression of the mitochondrial genome in *Xenopus laevis*: a map of transcripts. Cell 18:501–510

Robberson DL, Kasamatsu H, Vinograd J (1972) Replication of mitochondrial DNA. Circular replicative intermediates in mouse L cells. Proc Natl Acad Sci USA 69:737–741

Roe BA, Wong JFH, Chen EY, Armstrong PW, Stankiewicz A, Ma DP, McDonough J (1982) Mammalian mitochondrial tRNAs: A modified nucleotide 3′ to the anticodon may modulate their codon response. In: Slonimski P, Borst P, Attardi G (eds) Mitochondrial genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 45-49

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Spradling A, Pardue ML, Penman S (1977) Messenger RNA in heat-shocked *Drosophila* cells. J Mol Biol 109:559–587

Sprinzl M, Gauss DH (1984a) Compilation of tRNA sequences. Nucleic Acids Res 12:r1–r57

Sprinzl M, Gauss DH (1984b) Compilation of sequences of tRNA genes. Nucleic Acids Res 12:r59–r131

Staden R (1980) A computer program to search for tRNA genes. Nucleic Acids Res 8:817–825

Staden R (1982) Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res 10:4731–4751

Stormo GD, Schneider TD, Gold LM (1982) Characterization of translation initiation sites in *E. coli*. Nucleic Acids Res 10: 2971–2996

Taira M, Yoshida E, Kobayashi M, Yaginuma K, Koike K (1983) Tumor-associated mutations of rat mitochondrial transfer RNA genes. Nucleic Acids Res 11:1635–1643

Van Etten RA, Walberg MW, Clayton DA (1980) Precise localization and nucleotide sequence of the two mouse mitochondrial rRNA genes and three immediately adjacent novel tRNA genes. Cell 22:157–170

Van Etten RA, Michael NL, Bibb MJ, Brennicke A, Clayton DA (1982) Expression of the mouse mitochondrial DNA genome. In: Slonimski P, Borst P, Attardi G (eds) Mitochondrial genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 73–88

Ware VC, Tague BW, Clark CG, Gourse RL, Brand RC, Gerbi SA (1983) Sequence analysis of 28S ribosomal DNA from the amphibian *Xenopus laevis*. Nucleic Acids Res 11:7795–7817

Woese CR, Gutnell R, Gupta R, Noller HF (1983) Detailed analysis of the higher order structure of 16S-like ribosomal ribonucleic acids. Microbiol Rev 47:621–669

Wolstenholme DR, Clary DO (1985) Sequence evolution of *Drosophila* mitochondrial DNA. Genetics 109:725–744

Wolstenholme DR, Koike K, Cochran-Fouts P (1974) Replication of mitochondrial DNA: replicative forms of molecules from rat tissues and evidence for discontinuous replication. Cold Spring Harbor Symp Quant Biol 38:267–280

Wong JFH, Ma DP, Wilson RK, Roe BA (1983) DNA sequence of the *Xenopus laevis* mitochondrial heavy and light strand replication origins and flanking tRNA genes. Nucleic Acids Res 11:4977–4995

Zweib C, Glotz C, Brimacombe R (1981) Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Res 9:3621–3640

Received April 5, 1985/Revised June 25, 1985