
Genes for cytochrome c oxidase subunit I, URF2, and three tRNAs in *Drosophila* mitochondrial DNA

Douglas O. Clary and David R. Wolstenholme

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

Received 27 July 1983; Revised and Accepted 12 September 1983

ABSTRACT

Genes for URF2, tRNA^{trp}, tRNA^{cys}, tRNA^{tyr} and cytochrome c oxidase subunit I (COI) have been identified within a sequenced segment of the *Drosophila yakuba* mtDNA molecule. The five genes are arranged in the order given. Transcription of the tRNA^{cys} and tRNA^{tyr} genes is in the same direction as replication, while transcription of the URF2, tRNA^{trp} and COI genes is in the opposite direction. A similar arrangement of these genes is found in mammalian mtDNA except that in the latter, the tRNA^{ala} and tRNA^{asn} genes are located between the tRNA^{trp} and tRNA^{cys} genes. Also, a sequence found between the tRNA^{asn} and tRNA^{cys} genes in mammalian mtDNA, which is associated with the initiation of second strand DNA synthesis, is not found in this region of the *D. yakuba* mtDNA molecule. As the *D. yakuba* COI gene lacks a standard translation initiation codon, we consider the possibility that the quadruplet ATAA may serve this function. As in other *D. yakuba* mitochondrial polypeptide genes, AGA codons in the URF2 and COI genes do not correspond in position to arginine-specifying codons in the equivalent genes of mouse and yeast mtDNAs, but do most frequently correspond to serine-specifying codons.

INTRODUCTION

We have recently reported the sequences of various segments of the mitochondrial DNA (mtDNA) molecule of the fly, *Drosophila yakuba*, within which we have identified the genes for the two ribosomal RNAs, nine tRNAs, and seven polypeptides (cytochrome c oxidase (CO) subunits II and III, ATPase subunit 6, Unidentified Reading Frames (URF) I, II, III and A6L) (1-3). While these genes are similar to the genes found within a number of mammalian mtDNAs (4-6), the order in which they are arranged differs considerably between the *D. yakuba* and mammalian mtDNA molecules (1-3). A difference in the genetic code of *Drosophila* and mammalian mtDNA is also indicated by our data. In mammalian mtDNA the triplet AGA (arginine in the standard genetic code) is used only as a rare termination codon. In contrast, in *D. yakuba* AGA is used to specify an amino acid in the seven polypeptide genes we have identified. However, these AGA codons never correspond in position to codons which specify arginine in

the equivalent polypeptide genes which have been sequenced from mtDNAs of mouse (4), yeast (7-10) and a plant (Zea mays; 11). We have presented arguments that rather than specifying arginine, AGA may specify serine in the Drosophila mitochondrial genetic code.

In this paper we report a new sequence of D. yakuba mtDNA which contains the genes for URF2, tRNA^{TRP}, tRNA^{CYS}, tRNA^{TYR} and COI. Further interesting differences between D. yakuba and mammalian mtDNAs are noted, including the absence in the D. yakuba molecule of a sequence associated with initiation of second strand DNA synthesis in mammalian molecules, and the lack of a standard translation initiation codon in the D. yakuba COI gene. We suggest that initiation may be achieved in response to the four nucleotide codon ATAA.

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 digestion, electrophoresis, cloning of fragments into M13mp8 or M13mp9 DNA and purification of M13 DNAs are given or referenced in (1).

The larger (2.8 kb) HindIII - ClaI fragment of the pBR322-cloned HindIII-B fragment of D. yakuba mtDNA (Fig. 1) was cloned separately into M13mp8 and M13mp9. From replicative forms (RF) of the two hybrid M13 DNA molecules, viral DNAs containing partial deletions of the 2.8 kb HindIII - ClaI mtDNA fragment were prepared as previously described (3). In order to confirm the orientation of the HindIII-F fragment, the 1.0 kb BglII - EcoRI subfragment was cleaved from the pBR325-cloned EcoRI-A fragment, cloned into M13mp8, and partial deletions were prepared.

All DNA sequences were obtained from M13mp8- or M13mp9- cloned fragments by the extension-dideoxyribonucleotide termination procedure (12) using [α -³²P]dATP (800 Ci/mM; New England Nuclear) as described (1). The sequencing strategies used are given in Fig. 1.

Computer analyses of DNA sequences were carried out using the programs referenced in (3).

RESULTS AND DISCUSSION

Sequence organization. The entire sequence of a continuous 2789 nucleotide section of the D. yakuba mtDNA molecule, determined using the strategies summarized in Fig. 1, is shown in Fig. 2. Within this sequence are two open reading frames which, from comparisons of nucleotide and amino acid

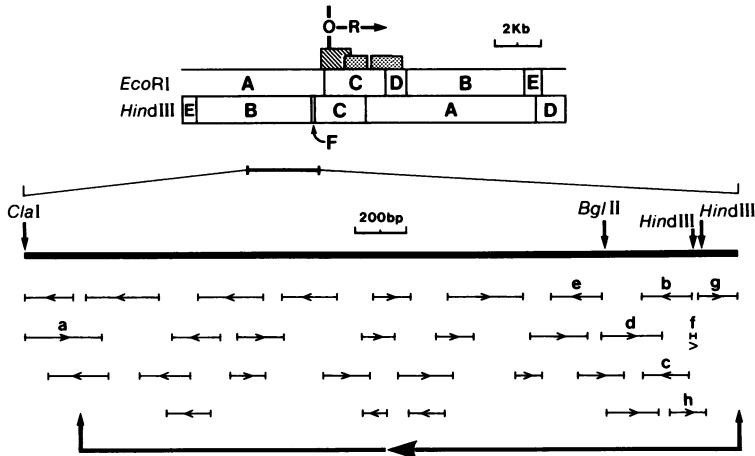


Figure 1. A map of the *D. yakuba* mtDNA molecule showing the relative locations of the A+T-rich region (cross hatched), the two rRNA genes (dotted), the origin (O) and direction (R) of replication, and *EcoRI* and *HindIII* sites and fragments (A-E and A-F respectively) (see (1) for details and references). The bar under the map indicates the segment sequenced. This segment is expanded below and the 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 ends of the larger (2.8 kb) *HindIII* - *ClaI* subfragment of the pBR322-cloned *HindIII*-B fragment. c and d, the two different orientations of the smaller (0.4 kb) *HindIII* - *BglII* subfragment of the *HindIII*-B fragment. e, the larger (4.4 kb) *HindIII* - *BglII* subfragment of the *HindIII*-B fragment. f, the *HindIII*-F fragment. g, the *HindIII*-C fragment. h, a *DNaseI* - *BglII* deletion (cloned in M13mp8) of the 1.0 kb *BglII* - *EcoRI* subfragment of the pBR325-cloned *EcoRI*-A fragment. Unlabeled sequences with arrows pointing to the left are *DNaseI* - *HindIII* deletions (cloned in M13mp8) of the 2.8 kb *HindIII* - *ClaI* subfragment. Unlabeled sequences with arrows pointing to the right are *DNaseI* - *ClaI* deletions (cloned in M13mp9) of the 2.8 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.

sequences to the corresponding sequences of previously identified genes of mouse mtDNA (4; Table 1 and Fig. 3), have been identified as the genes for cytochrome c oxidase subunit I (COI) and URF2. The sequence also contains three regions each of which can fold into the characteristic secondary structure of a tRNA, with anticodons indicating them to be the genes for tRNA^{trp}, tRNA^{cys}, and tRNA^{tyr} (Fig. 4). The five genes occur in the order URF2, tRNA^{trp}, tRNA^{cys}, tRNA^{tyr}, and COI. The 5' end of the URF2 gene is located adjacent to the tRNA^{f-met} gene as shown previously (1), and the 3' end of the COI gene is next to the previously described (3) tRNA^{leu}_{UUR} gene (Fig.

tRNA^{f-met} → URF2 →
 TTTCTTTTAAATTTTATAATTCATCAAAAATTTTATTACCACAATTATAATATTGGAACATTAATTACAGT 75
 T S N S W L G A W M G L E I N L L S F I P L L s D
 TACATCTAAATTCWTGGTTAGGAGCTTGAATAGGTTTAGAAATTAATTTGTTATCTTTATCCCCCTTAAGAGA 150
 N N N L M S T E A S L K Y F L T Q A L A S T V L L L
 TAATAATAATTTAATATCTACAGAAGCTTCTTTAAAATATTTTTAAACCAAGCTTGGCATCAACTGTTTATT 225
 F S S I L L M L A N N L N N E I N E S F T S M I I
 ATTTTCTTCAATTTACTTATATTGGCAAATAATTTAAATAATGAAATTAATGAATCTTTTACATCAATAATTAT 300
 M S A L L L K s G A A P F H F W F P N M M E G L T
 TATATCGGCCTTATTATTA AAAAGAGGAGCGCTCCTTTTCATTTTGTATTTCCTAATATAATAGAAGGATTAAC 375
 W M N A L M L M T W Q K I A P L M L I S Y L N I K
 ATGAATAAATGCTTTGATATTAATAACTTGACAAAAAATGCTCCATTAATATTAATTTCTTATTTAAATATTA 450
 N L L L L I S V I L S V I I G A I G G L N Q T S L R
 AAATTTATTAATAGTCTAATTTATCAGTTATTATTTGGACCAATTTAAACCAAACTTCACTCCG 525
 K L M A F S S I N H L G W M L s S L M I s E S I W
 AAAATTAATAGCATTCTTCTATTAATCATTAGGATGAATATTAAGATCTTAAATGATTAGAGAATCAATTTG 600
 L I Y F I F Y S F L S F V L T F M F N I F K L F H
 ATTAATTTATTTATTTTATTCATTCTTATCTTTGTATTAAACATTTATTTAATATTTTAAATATTTC 675
 L N Q L F S W F V N s K I L K F S L F M N F L S L
 TTTAAATCAATTTTCTTGATTGTAAACAGAAAAATTTAAAAATTTTCATTATTTATAAATTTTTTATCTTT 750
 G G L P P P F L G F L P K W L V I Q Q L T M C N Q Y
 AGGTGATTACCTCCATTTTAGGATTTTACCAAAATGATTAGTAATCAACAATTAACAATATGTAATCAATA 825
 F L L T L M M M S T L I T L F F Y L R I C Y S A F
 TTTTTATTAACATTAATAATAATCAACTTAAATTAGATTATTTTTTTATTACGAATTTGTTACTCAGCTTT 900
 M L N Y F E N N W I M E M N M N S N N T N L Y L I
 TATATTAATTTATTCGAAAATAACTGAATCATGAAATAAATAAATAGTAATAATACTAATTTATATTTAAT 975
 M T F F S I F G L F L I S L F F F M L * tRNA^{trp} →
 TATAACTTTTTTTCAATTTTCGGATTAATTTTAAATTTCTTATTTTTTTTTTATACTTTAAGGCTTTAAGTTAAC 1050
 TAACTAATAGCCTTCAAAGCTGTAATAAAGGGTATTCCTTTAAGTCTTAATAAAAATTTACTCCTTCAAATTT 1125
 GCAGTTTGATATCATTATTGACTATAAGACCTAGATTAAATTTATGATTAAGAAGAATAATTCTTATAAATAGA 1200
 TTTACAATCTATCGCCTAAACTTCAGCCACTTAATCATAATCGGACAATGGTTATTTTCTACAAATCATAAAG 1275
 D I G T L Y F I F G A W A G M V G T S L s I L I R
 ATATTGGAACTTTATATTTTCTTTGGAGCTTGAGCCGAATAGTAGGAACATCTTAAAGAATTTAATTCGAG 1350
 A E L G H P G A L I I G G D D Q I Y N V I V T A H A F
 CAGAATTAGGTCACCGGACATTAATTTGGAGATGATCAAAATTTATAATGTAATTTACTGACATGCTTTTA 1425
 I M I F F M V M P I M I G G F G N W L V P L M L G
 TTATAATTTTTTTATAGTAATACCTATTATAATTTGGGGGTTTGGAAATGATTAGTGCCTTTAATATTAGGAG 1500
 A P D M A F P R M N N M s F W L L P P A L S L L L
 CTCCTGACATAGCATTCCCACGAATAAATAATATAAGATTTTGATTACTACCTCCTGCTCTTTCTTTATTATTAG 1575
 V s s M V E N G A G T G W T V Y P P L S S G I A H
 TAAGAAGAATAGTTGAAAACGGAGCTGGTACAGGTTGAAGTGTACCCTCCTTTATCTTCAGGTATCGCTCATG 1650
 G G A S V D L A I F S L H L A G I S S I L G A V N
 GTGGAGCTTCTGTAGATTAGCTATTTTTCTTCTTCAATTTAGCTGGAATTTCTTCAATTTTAGGAGCTGTAATTT 1725

```

F I T T V I N M R S T G I T L D R M P L F V W S V
TTATTAGACTGTAATTAATATACGATCAACTGGAATTACATTAGACCGAATACCTTTATTGTATGATCAGTAG 1800
V I T A L L L L L S L P V L A G A I T M L L T D R
TTATTACTGCTTTATTACTTTTACTACTCTTTACCAGTTCCTGCGGAGCTATTACTATATTATTAAACAGACGGAA 1875
N L N T S F F D P A G G G D P I L Y Q H L F W F F
ATTTAAATACTTCTTTTTTGTATCCAGCTGGAGGAGGAGATCCTATTTTGTACCAACATTATTTTGTATTTTTG 1950
G H P E V Y I L I L P G F G M I S H I I S Q E S G
GTCACCTGAAGTTTATATTTAATTTTACCAGGATTGGAATAATTTCTCATATTATTAGACAAGAATCTGTA 2025
K K E T F G S L G M I Y A M L A I G L L G F I V W
AAAAGGAAACTTTCGGTTCTTTAGGAATAATCTATGCTATACTTGCTATTGGATTATTAGGATTATTGTTTGAG 2100
A H H M F T V G M D V D T R A Y F T S A T M I I A
CTCATCATATATTTACAGTTGGAATAGACGTTGATACAGGAGCTTATTTTACTTCTGCTACTATAATTATTGGCG 2175
V P T G I K I F s W L A T L H G T Q L S Y S P A I
TTCCACAGGAATAAAAATTTTAGATGATTAGCTACTTTACATGGAACTCAACTTTCTTATTCTCCAGCTATTT 2250
L W A L G F V F L F T V G G L T G V V L A N S S V
TATGAGCTTTAGGATTGTTTTTTTATTACAGTAGGAGGATTAACAGGAGTTGATTAGCTAATTCATCAGTTG 2325
D I I L H D T Y Y V V A H F H Y V L S M G A V F A
ATATTATTTACATGATACTTATTATGTAGTAGCTCATTCCACTACGTTTTATCAATAGGAGCTGATTTGGCTA 2400
I M A G F I H W Y P L F T G L T L N N K W L K S Q
TTATAGCAGTTTTATTCACTGATACCCATTATTTACTGGATTGACATTAATAATAAATGCTAAAAAGTCAAT 2475
F I I M F I G V N L T F F P Q H F L G L A G M P R
TTATTATTAGTTTATTGGAGTAAATTTAACATTTTCCCCCAACATTTTTTAGGATTAGCAGGAATACCTCGAC 2550
R Y S D Y P D A Y T T W N V V S T I G S T I S L L
GTTATTACAGATTACCTGATGCTTACACTACATGAAATGTTGTGTCTACTATTGGGTCAACTATTTTCATTATTAG 2625
G I L F F F Y I I W E S L V S Q R Q V I Y P I Q L
GAATTTTATTTTTTCTATATTATTTGAGAAAGTTTAGTGTCTCAACGACAAGTAATTTATCCAATTCAATTA 2700
N S S I E W Y Q N T P P A E H s Y S E A L P L L T N *
ATTCATCTATGAATGATACAAAATACACCCCGCTGAACATAGATATTCTGAATTACCACTTTTAAACAATT 2775

```

** $\xrightarrow{\text{leu}}$
AATTCTAATATGG

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 (4), this sequence contains the URF2 and cytochrome c oxidase subunit I (COI) genes. The boxed nucleotide sequences fold into the characteristic cloverleaf structures of tRNAs with anticodons (underlined) indicating them to be the genes for tRNA^{Leu}, tRNA^{Cys} and tRNA^{Tyr} (Fig. 4). The nucleotide sequence shown is the (5'-3') sense strand of the URF2, COI and tRNA^{Leu} genes, and the arrows indicate the direction of transcription of each gene. Asterisks indicate partial or complete termination codons. In the amino acid sequences a lower case letter s indicates the tentative assignment of serine to an AGA codon (see text). The question mark at the beginning of the COI sequence indicates uncertainty as to which nucleotides constitute the translation initiation sequence (see text). The sequence containing the amino-terminal 169 nucleotides of the URF2 gene and the entire tRNA^{Leu}-met gene are published in (1). The entire tRNA^{Leu} gene is given in (3).

URF2	mouse	MNPITLA.IY	F..FL-.PV.	.MS.TNLMLM	.V...FS..A	I..M.INKK.	PR...AT..	60
	<i>D. yakuba</i>	IFYNSSKILF	TTIMIIGTLI	TVTSNSWLGA	WMCLEINLLS	FIPLL [↓] DN [↓] NN	LMSTEASLKY	
		.V...T..MI	I.LAIV.NYK	QLGTW [↓] MFQQQ	TNGLILN.TL	M..SM.L.L.L.EVT	Q.IPLHMG.I
		FLTQALASTV	LLFSSILLML	ANNLNNEINE	SFTSMIIMS-	-ALLLK [↓] sGAA	PFHFWFPNMM	EGLTWMNALM
		.L.....	SI..QI.P.L	NSTII.MLA.	T.IFM..W..QM..I.	.Y...A.M..	..AI.PYNP.
		LMTWQKIAPL	M-LIS-Y-LN	IKNLLLISVI	LSVIIIGAIGG	LNQTSLRKLM	AFSSINHLGW	M [↓] LSL [↓] IES
		LT.LNLMI.I	I.TAPMFMAL	MLNNSMTI.S	ISLLW [↓] NKTPA	MLTHISLML.LT.II.TE
		IMLIYFIFYS	FLSFVLTFMP	NIFKLFHLNQ	LFSWFVN [↓] sKI	LKFSLFM [↓] NFL	SLGG [↓] LPFLG	FLPKWLV [↓] IQQ
		.MKN.CLIMA	...A.MA.LNI.LI..	TSLTMFPT..	NSKM.THQTK	TKPNLMFS.L	AIMSTMT.PL
		LTMCNQYFLL	TLM [↓] MMSTLIT	LFFYLRCISY	AFMLNYFENN	WIMEMNNSN	NTNLYLIMTF	FSIFGLFLIS
		APQLIT*						
		LFFFML*						
COI	mouse	MFINR.....L	L.....	.A.S.....	..Q....L..	60
	<i>D. yakuba</i>	?-SRQWL [↓] FST	NHKDIGTLYF	IFGAWAGMVG	TSL [↓] sILIRAE	LGHGALIGD	DQIYNVIVTA	
	yeast	M-VQR..Y..	.A...AV...	MLAIFS..A.	.AMSLI..L.	.AA..SQYLH	GNSQLFL.VG	
		...V.....	..M.....I..SSFL	..ASS...A
		HAFIMIFFMV	MPIMIGGFGN	WLVPLMLGAP	DMAFPRMNM	sFWLLPPALS	LLLvsMVEN	GAGTGWTVYP
		..VL...CL.	..AL.....	Y.L..II..T	.T...I..I	A..V..MG.V	C.VTSTL.S
		..AGNPV.A.T....	...V.....	.I.....I..	.KPPAM.QYQ	T.....L.	..V.....
		PLSSGIAHGC	ASVDLAI [↓] FSL	HLAGISSILG	AVNFITTVIN	MRSTGITLDR	MPLFVWSVVI	TALLLLLSL [↓] P
		...IQ..S.	P.....A.	..TS...L.	.I...V.TL.	..TN.MTMHK	L.....IF.	..F.....
		...AG.....T....I...VVTY	Y.....P..Y
		VLGAI [↓] TMLL	TDRNLNTSFF	DPAGGGDPII	YQHLFWFFGH	PEVYILILPG	FGMISHII [↓] sQ	ESGKKE [↓] TFGS
		..SAG.....	L...F.....	EV.....	.E.....I..	..I...VVST	Y...P.VF.E
		M..VW..MS.	.F.....L...C.....	...I..V.V	.S.....G	NIKW...M.
		LGM [↓] YAMLA [↓] I	GLLGPFVWAH	HMFTVGDVD	TRAYFTSATM	IIAVPTGIKI	FsWLATLHGT	QLSYS [↓] PAILW
		IS.V...AS.L.S.	..YI..L.A.I.....	.S...IY.G	SIRLATPM.Y
	I.....I..S.	..L.VL...V..F..	S.F..DDT.A
		ALGFVFLPTV	GCLTGVVLAN	SSVDIILHDT	YYVVAHPHYV	LSMGAVFAIM	AGFIHWYPLF	TGLTLN [↓] NKWL
		.IA.L...MA..	A.L.VAF...G.Y..I.SLF	..YYY.S.QI	L..NY.E.LA
		.AH.A...V.	..M.....	...S.....T..SM.F.	..TAV.IMIF	M...AFA.K.
		KSQFIIMFIG	VNLTFPPQHF	LGLAGMPRRY	SDYPDAYTTW	NVSTIGSTI	SLLGILFFFY	IIWESLVSQR
		QI..WLI...	A.VI...M..	..IN.....I	P.....FAG.	.Y.AS...F.	AT.SLFL.I.	.LYDQ..NNK
		E.MSVSYAST	NL..LHGC..	PY.TFE.PTY	VKVK*			
		QVIYPIQLNS	SIEWYQNTTP	AEH [↓] sYSELPL	LTN-*			
		S...AKAPS.	...PLLTS..	.V.SFNT-.A	VQS-*			

Figure 3. Comparisons of the amino acid sequences of URF2 and cytochrome c oxidase subunit I (COI) predicted from the nucleotide sequences of the respective genes of mtDNA of *D. yakuba*, with the corresponding amino acid sequences of mouse (4) and yeast (13). 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 complete termination codon. Wide vertical solid arrows indicate *D. yakuba* amino acids specified by AGA which are represented by a lower case letter s to indicate the tentative assignment of serine (see text). The question mark at the beginning of the *D. yakuba* COI sequence indicates uncertainty regarding translation initiation (see text).

Table 1. Nucleotide and amino acid sequence comparisons of the URF2 and cytochrome c oxidase subunit I (COI) genes of D. yakuba with the corresponding genes of mouse and yeast.

Gene	Number of nucleotides ^a (<u>D. yakuba</u>)	% Nucleotide sequence homology ^{a,b} (<u>D. yakuba</u> /mouse)	% Silent nucleotide substitutions ^{a,b} (<u>D. yakuba</u> /mouse)	% Amino acid sequence homology ^b	
				<u>D. yakuba</u> /mouse	<u>D. yakuba</u> /yeast
URF2	1023	47.0	30.2	34.7	--
COI	1537	71.6	55.7	74.7	59.2

^aExcluding nucleotides concerned with termination.

^bThese values include all deduced insertion/deletions (see Fig. 3).

The data for mouse and yeast mtDNAs are taken from references (4,13).

2). The tRNA^{CYS} and tRNA^{TYR} genes are transcribed in the same direction as that in which replication proceeds around the molecule, while the URF2, COI and tRNA^{TRP} genes are transcribed in the opposite direction.

The arrangement of these five genes relative to each other and to adjacent genes within the D. yakuba mtDNA molecule is similar to what is found in mouse and other mammalian mtDNAs (4-6,14) except that in mammalian mtDNAs, the genes for tRNA^{ALA} and tRNA^{ASN} separate the tRNA^{TRP} and tRNA^{CYS} genes, and the tRNA^{SER}_{UCN} gene is next to the 3' end of the COI gene. The relative location of the genes for URF2, tRNA^{TRP}, tRNA^{CYS}, tRNA^{TYR} and COI and all of the other D. yakuba mitochondrial genes which have been determined to date are shown in Fig. 5.

Alignment of the nucleotide and amino acid sequences (Fig. 3) of the URF2 genes of D. yakuba and mouse indicates that insertion-deletions involving six internal codons have occurred between these two genes resulting in a D. yakuba URF2 which is 12 nucleotides shorter. The last codon of both the D. yakuba and the mouse URF2 genes is separated by a single T from the 5' terminal nucleotide of the tRNA^{TRP} gene. This observation provides further support for the view that polyadenylation of some gene transcripts (15), following their excision from multicistronic primary transcripts, generates the UAA termination codon (3), as first suggested for transcripts of some mammalian mitochondrial genes (5,16).

Replication of the two strands of the mtDNA molecules of both mammals and Drosophila is highly asymmetrical (17-24). In mtDNAs of cultured cells of mouse, human, and hamster, synthesis of one strand (H strand) proceeds for two-thirds of the genome length before synthesis of the second strand (L strand) begins (17,19,22). It has been clearly demonstrated that initiation of L strand synthesis in all mouse L-cell and most human KB cell mtDNA

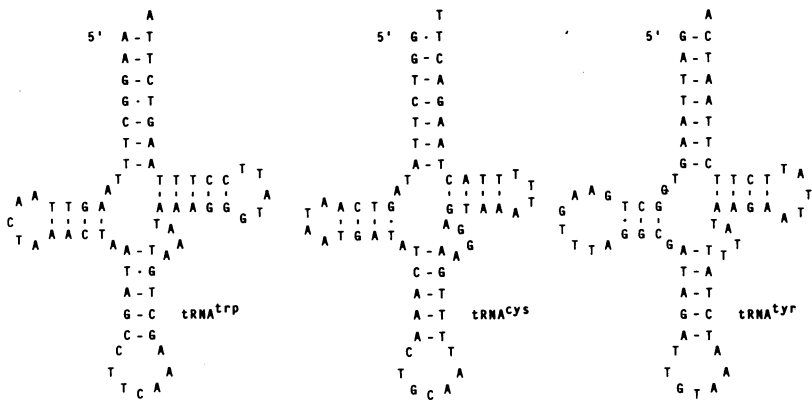


Figure 4. The genes for tRNA^{trp}, tRNA^{cys} and tRNA^{tyr} of D. yakuba mtDNA shown in the presumed characteristic secondary structures of the corresponding tRNAs.

molecules occurs within a specific sequence of 32 (mouse) or 31 (human) nucleotides which can fold into a hairpin loop with a perfect twelve base pair stem, and is located between the 5' end of the tRNA^{asn} gene and the 3' end of the tRNA^{cys} gene (4,25,26). This 31-32 nucleotide sequence is highly conserved in other mammalian mtDNAs (6,14). However, such a sequence is not found in the corresponding region of D. yakuba mtDNA molecules (Fig. 2). As noted above, a tRNA^{asn} gene does not occur in the region between the URF2 and COI genes, and the 3' end of the tRNA^{trp} gene overlaps by eight nucleotides the 5' end of the tRNA^{cys} gene (Fig. 2). The tRNA^{cys} and tRNA^{tyr} genes are separated by 14 apparently non-coding nucleotides, but this sequence does not have the potential to form a hairpin loop. These observations are consistent with the results of our previous electron microscope studies on replicating forms of Drosophila mtDNA molecules (20,21) which failed to provide evidence for a preferred site of initiation of second strand synthesis that would map in the region between the URF2 and COI genes. However, evidence for a highly preferred site of initiation of second strand synthesis was obtained for D. melanogaster mtDNA molecules (20) which, from consideration of mapping and sequencing data (1,27), lies close to the boundary of the tRNA^{ile} gene and the A+T-rich region of these mtDNA molecules (Fig. 5).

The COI gene lacks a standard initiation codon. Homology of both nucleotide and predicted amino acid sequences of the COI genes of D. yakuba and mouse are higher than has been found for any other mitochondrial gene compared between these species (Table 1; Fig. 3; 3). However, the D. yakuba

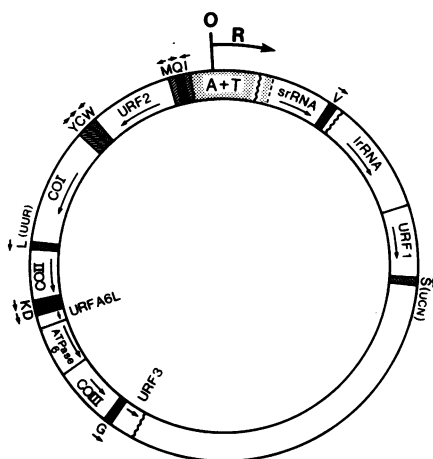


Figure 5. The arrangement of genes determined to date in the circular mtDNA molecule of *D. yakuba*, from the results of various studies (for references see 1,3) and the present data. Each tRNA gene is identified by the one letter amino acid code. Arrows within and outside the molecule indicate the direction of transcription of each gene. Wavy lines indicate uncertain gene termini. O and R indicate the origin and direction of replication, respectively.

COI gene does not appear to have a standard translation initiation codon.

Each of the three other identified polypeptide genes (COII, COIII, and ATPase6) of *D. yakuba* mtDNA which we have sequenced has an ATG initiation codon, and the initiation codon for all four of the URFs (1, 2, 3 and A6L) for which complete or partial sequences have been obtained appears to be ATT (1,3). These codons as well as ATA and ATC are recognized as initiation codons in mammalian mtDNAs (4,5). Also we have identified in *D. yakuba* mtDNA a putative tRNA^{f-met} gene with a 5' CAT anticodon (1) which presumably could recognize ATN triplets as initiation codons as in mammalian mitochondria (4,6). The initiation codon of the COIII gene of *Aspergillus nidulans* is unique among mitochondrial genes in that it appears to be GUG (28), which is also the initiation codon of the bacteriophage MS2 A-protein gene (29) and of some *Escherchia coli* genes (30).

The first sense codon in the *D. yakuba* COI gene is TCG which would be expected to specify serine. This codon is two codons downstream from the mouse COI initiation codon and one codon downstream from the yeast COI initiation codon (ATG in both cases; Fig. 3). TCG is only used one other time as an internal codon in the *D. yakuba* mitochondrial genes we have sequenced (Table 2), and it is not clear how it might be recognized as an initiation codon by the *D. yakuba* tRNA^{f-met} gene. There is not an obvious candidate for an initiation codon among the inframe codons which lie downstream from the TCG codon (Fig. 2); beginning with the fourth sense codon (TGA) this region is highly conserved in the *D. yakuba*, mouse and yeast genes (Fig. 3). Also, the possibility that initiation occurs upstream in the region overlapping the

sense strand of the tRNA^{tyr} gene is unlikely due to the presence of a TAA codon immediately 5' to the TCG codon. As the D. yakuba COI gene is highly conserved relative to the mouse COI gene (Table 1, Fig. 3) it seems extremely unlikely that this D. yakuba mitochondrial COI gene is nonfunctional.

Because of these difficulties of interpretation, we have considered the possibility that initiation of translation of the COI gene involves an unusual initiation sequence. The four nucleotides immediately 5' to the first sense codon of the COI gene are ATAA. If a +1 ribosomal frameshift occurred in this region, it would allow the triplet ATA to function as an initiation codon for the open reading frame of the COI gene. Ribosomal frameshifting has frequently been invoked as a mechanism to correct frameshift mutations (see 31,32 for references), in one case in the mitochondrial COII gene of yeast (33). More important to the present discussion is evidence of ribosomal frameshifting in some wild type viruses which is essential to the production of some proteins, and which may have a regulatory function (34,35).

Alternatively there may be a tRNA in D. yakuba mitochondria with an anticodon loop with unusual properties that permit it to read the complete four letter ATAA sequence as a single codon. Suppressor tRNAs have been described from bacteria and yeast which correct frameshift mutations by reading four nucleotides as a codon (36-38). Such tRNAs have been found which have an extra nucleotide in their anticodon loop (36,38; Bossi, L., unpublished). A tRNA^{thr} gene with an eight rather than a seven nucleotide anticodon loop has been found in yeast mtDNA (39).

The only potential tRNA^{f-met} gene we have sequenced from D. yakuba mtDNA has a normal anticodon loop of seven nucleotides (1). However, 5' to the CAT anticodon is a T. This differs from the situation found in mammalian mtDNA where a C is located in this position (4-6). The T residue seen in D. yakuba mtDNA may be functional in allowing the tRNA^{f-met} to read the four letter sequence ATAA as a single codon and thus permit a continuation of reading in the correct frame. This suggestion is consistent with evidence from studies with bacteriophage Q β coat cistron, that an A following an AUA (as well as an AUG) initiation codon might interact with a complementary U in the anticodon loop to increase the efficiency of codon/anticodon interaction (40). Also, it has been shown that in Salmonella typhimurium interaction of an A adjacent to the 3' side of an amber codon with a U located 5' to the anticodon of a suppressor tRNA increases the efficiency of translation of the codon (41). It is interesting to note that the sequence ATAA precedes the initiation codon (ATG) of the D. yakuba COII gene (3) and is separated by one sense codon (AAA)

from the putative GUG initiation codon of the A. nidulans COIII gene (28).

Transfer RNA genes. The tRNA^{trp}, tRNA^{cys} and tRNA^{tyr} genes of D. yakuba mtDNA are 57%, 67% and 60% homologous, respectively, to the corresponding tRNA genes of mouse mtDNA (4). The major structural characteristics of these three tRNA genes are similar to those of the nine other D. yakuba and D. melanogaster mt-tRNA genes we have described previously (1-3). All these tRNA genes lack a number of the nucleotides which are constant in prokaryotic and non-organelle eukaryotic tRNAs (42,43). However, the common occurrence among other Drosophila mt-tRNA genes of the constant Pu₂₆, T₃₃ and Pu₃₇ nucleotides (numbering system in 42) is maintained in each of the tRNA^{trp}, tRNA^{cys} and tRNA^{tyr} genes. The Py₁₁-Pu₂₄ nucleotide pair, which is conserved in all other Drosophila tRNA genes except tRNA^{f-met} where it is a G-C pair, is also a G-C pair in tRNA^{trp}.

Codon usage and the genetic code. The frequencies of codons in the URF2 and COI genes are shown in Table 2. An exceptionally high frequency of codons ending in A or T is found in both the URF2 gene (93.3%) and the COI gene (91.6%). These values are similar to the mean frequency (94.4%) of codons ending in A or T in the COII, COIII, ATPase6 and URFA6L genes of D. yakuba mtDNA reported previously (3). Among the completed sequences of six polypeptide genes of D. yakuba mtDNA (3; and Fig. 2) and the partial URF1 sequence (1,2) only 5 expected sense codons are not used at all, and each of these contains at most a single A or T.

The G+C content of the URF2 gene (18.6%) is similar to that of the URFA6L gene (17.2%), both of which are considerably lower than the G+C content of the COI gene (30.1%) and of the other three identified D. yakuba polypeptide genes (range 24.2% to 28.8%) sequenced to date (3 and the present data).

In comparisons of nucleotide sequences of corresponding genes of D. yakuba and mouse, the frequency of nucleotide substitutions which would not result in a corresponding amino acid substitution (silent substitutions, Table 1) is considerably higher for the more homologous COI genes than for the URF2 genes, which is consistent with our previous observations (3).

In comparisons to the nucleotide sequences of mouse genes, TGA codons of D. yakuba mtDNA are conserved in the URF2 gene (five out of nine), which has a nucleotide sequence homology to the mouse URF2 gene of only 47.0% (Table 1), as well as in the more homologous (71.6%; Table 1) COI gene (13 out of 13). The finding of a tRNA gene with an anticodon (5' TCA; Figs. 2 and 4) expected to be able to decode TGA and TGG adds to earlier evidence that in D. yakuba mtDNA as in mammalian (44) and fungal mtDNAs (8,45,46), TGA specifies

Table 2. Codon usage in the URF2 and COI genes of D. yakuba mtDNA.

	URF2	COI	Others ^a		URF2	COI	Others ^a
Phe-TTT	34	32	58	Ser-TCT	15	19	19
TTC	3	7	3	TCC	0	0	1
Leu-TTA	57	54	94	TCA	13	11	27
TTG	4	2	1	TCG	1	1	0
Leu-CTT	2	7	10	Pro-CCT	3	13	21
CTC	1	0	1	GCC	1	2	0
CTA	1	2	4	CCA	3	9	13
CTG	0	0	0	CCG	0	1	1
Ile-ATT	32	47	68	Thr-ACT	6	18	25
ATC	2	2	1	ACC	2	0	1
Met-ATA	25	22	23	ACA	11	15	26
ATG	2	1	8	ACG	0	1	0
Val-GTT	3	15	19	Ala-GCT	7	27	24
GTC	0	0	1	GCC	2	2	2
GTA	4	16	25	GCA	4	6	11
GTG	0	3	0	GCG	0	1	0
Tyr-TAT	9	13	24	Cys-TGT	2	0	5
TAC	1	6	7	TGC	0	0	1
TER-TAA	0	1	3	Trp-TGA	9	13	27
TAG	0	0	0	TGG	1	2	0
His-CAT	3	14	26	Arg-CGT	0	1	2
CAC	0	3	3	CGC	0	0	0
Gln-CAA	7	11	19	GCA	2	9	12
CAG	0	0	0	CGG	0	0	0
Asn-AAT	27	17	43	Ser-AGT	2	2	6
AAC	3	1	3	AGC	0	0	1
Lys-AAA	10	5	9	(Ser)-AGA	5	7	12
AAG	0	1	1	AGG	0	0	0
Asp-GAT	1	11	13	Gly-GGT	3	9	12
GAC	0	4	3	GGC	0	0	0
Glu-GAA	8	9	21	GGA	10	35	26
GAG	0	0	0	GGG	0	3	4

^aThese values are the sums of the codon frequencies in the COII, COIII, ATPase6 and URFA6L genes of D. yakuba mtDNA (3).

AGA is tentatively shown as serine (see text). AGG has not been found in Drosophila mtDNA, and among mammalian mtDNAs has been found only as the termination codon of human URF6 (5). TGA is interpreted as tryptophan (see text). TAG has been interpreted as the termination codon for the D. melanogaster COIII gene (2). As in mammalian mtDNA, ATA is assumed to specify methionine (4,5).

tryptophan (1,3).

Among metazoan mtDNAs sequenced so far, the occurrence of AGA as an amino acid-specifying codon is unique to Drosophila mtDNA (1,3) where it has been found in every polypeptide gene for which sequences have been obtained. In comparisons of the sequences of D. yakuba genes to the corresponding genes of mouse, yeast and corn, we found that D. yakuba AGA codons never correspond in position to codons specifying arginine in the other species (3). Further, in comparisons to mouse mtDNA it was noted that while AGAs do correspond in

position to codons specifying nine different amino acids, the most frequent were those specifying serine. From these observations and considerations of the characteristic features of mitochondrial genetic codes, we have argued that AGA may specify serine rather than arginine in the Drosophila mitochondrial genetic code. AGA occurs five times in the URF2 gene and seven times in the COI gene. As in previous D. yakuba/mouse gene comparisons, none of the twelve AGA codons correspond in position to arginine specifying codons in URF2 or COI genes of mouse and yeast (Fig. 2). In the D. yakuba URF2 gene the five AGA codons correspond in position to codons specifying five different amino acids (isoleucine, leucine, alanine, asparagine, and threonine) in the mouse URF2 gene (Fig. 3). However, the seven AGA codons in the D. yakuba COI gene (which has a higher homology to mouse and yeast COI genes (Table 1) than has been found to date for any other D. yakuba/mouse or D. yakuba/yeast mitochondrial gene comparison) show a high degree of correspondence to serine-specifying codons in both the mouse COI gene and the yeast COI gene (five in each case; Fig. 3). This latter observation adds further support to the arguments (3) that AGA may specify serine rather than arginine in the Drosophila mitochondrial genetic code.

ACKNOWLEDGEMENTS

We wish to thank John F. Atkins for valuable discussions, and Felix Machuca for excellent technical assistance. This work was supported by the National Institutes of Health Grants GM 18375 and RR 07092.

REFERENCES

1. Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M.-R. and Wolstenholme, D.R. (1982) Nucl. Acids Res. 10, 6619-6637.
2. Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R. (1983) Nucl. Acids Res. 11, 2411-2425.
3. Clary,D.O. and Wolstenholme,D.R. (1983) Nucl. Acids Res. 11, 4211-4227.
4. Bibb,M.J., Van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) Cell 26, 167-180.
5. 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.
6. 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.
7. Coruzzi,G. and Tzagoloff,A. (1979) J. Biol. Chem. 254, 9324-9330.
8. Fox,T.D. (1979) Proc. Natl. Acad. Sci. USA 76, 6534-6538.
9. Thalenfeld,B.E. and Tzagoloff,A. (1980) J. Biol. Chem. 255, 6173-6180.
10. Macino,G. and Tzagoloff,A. (1980) Cell 20, 507-517.
11. Fox,T.D. and Leaver,C.J. (1981) Cell 26, 315-323.
12. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
13. Bonitz,S.G., Coruzzi,G., Thalenfeld,B.E., Tzagoloff,A. and Macino,G.

- (1980) *J. Biol. Chem.* 255, 11927-11941.
14. Taira, M., Yoshida, E., Kobayashi, M., Yaginuma, K. and Koike, K. (1983) *Nucl. Acids Res.* 11, 1635-1643.
 15. Mertens, S.H. and Pardue, M.L. (1981) *J. Mol. Biol.* 153, 1-21.
 16. Ojala, D., Montoya, J. and Attardi, G. (1981) *Nature* 290, 470-474.
 17. Robberson, D.L., Kasamatsu, H. and Vinograd, J. (1972) *Proc. Natl. Acad. Sci. USA* 69, 737-741.
 18. Wolstenholme, D.R., Koike, K. and Cochran-Fouts, P. (1973) *J. Cell Biol.* 56, 230-245.
 19. Kasamatsu, H. and Vinograd, J. (1974) *Ann. Rev. Biochem.* 43, 695-719.
 20. Goddard, J.M. and Wolstenholme, D.R. (1978) *Proc. Natl. Acad. Sci. USA* 75, 3886-3890.
 21. Goddard, J.M. and Wolstenholme, D.R. (1980) *Nucl. Acids Res.* 8, 741-757.
 22. Nass, M.M.K. (1980) *J. Mol. Biol.* 140, 257-281.
 23. Clayton, D.A. (1982) *Cell* 28, 693-705.
 24. Wolstenholme, D.R., Goddard, J.M. and Fauron, C.M.-R. (1983) in *Developments in Molecular Virology. Vol. 2. Molecular Events in the Replication of Viral and Cellular Genomes*, Becker, Y., Ed., Martinus Nijhoff B.V. The Hague. in press.
 25. Martens, P.A. and Clayton, D.A. (1979) *J. Mol. Biol.* 135, 327-351.
 26. Tapper, D.P. and Clayton, D.A. (1981) *J. Biol. Chem.* 256, 5109-5115.
 27. Fauron, C.M.-R. and Wolstenholme, D.R. (1980) *Nucl. Acids Res.* 8, 2439-2452.
 28. Netzker, R., Kochel, H.G., Basak, N. and Kuntzel, H. (1982) *Nucl. Acids Res.* 10, 4783-4794.
 29. Fiers, W., Contreras, R., Duerinck, F., Haegmean, G., Merregaert, J., Min Jou, W., Raeymakers, A., Volckaert, G., Ysebaert, M., Van de Kerckhove, J., Nolf, F., and Van Montagu, M. (1975) *Nature* 256, 273-278.
 30. Gay, N.J. and Walker, J.E. (1981) *Nucl. Acids Res.* 9, 3919-3926.
 31. Atkins, J.F., Nichols, B.P. and Thompson, S. (1983) *The EMBO Journal* 2, in press.
 32. Atkins, J.F. (1980) *tRNA: Biological Aspects*, Albelson, J., Schimell, P. and Soll, D. Eds., in *Cold Spring Harbor Laboratory, New York* pp. 439-449.
 33. Fox, T.D. and Weiss-Brummer, B. (1980) *Nature* 288, 60-63.
 34. Beremand, M.N. and Blumenthal, T. (1979) *Cell* 18, 257-266.
 35. Kastelein, R.A., Remant, E., Fiers, W. and Van Duin, J. (1982) *Nature*, 295, 35-41.
 36. Riddle, D.L. and Carbon, J. (1973) *Nature New Biology* 242, 230-234.
 37. Bossi, L. and Roth, J.R. (1981) *Cell* 25, 489-496.
 38. Cummins, C.M., Donahue, T.F. and Culbertson, M.R. (1982) *Proc. Natl. Acad. Sci. USA* 79, 3565-3659.
 39. Tzagoloff, A., Macino, G., Nobrega, M.P. and Li, M. (1979) in *Proc. 8th Ann. ICM-UCLA Symp. Molecular and Cellular Biology; Extra chromosomal DNA*, Cummings, D.J., Dawid, I.B., Bert, P., Weissman, S.M. and Fox, C.F. Eds., pp. 339-355, Academic Press, New York.
 40. Taniguchi, T. and Weissmann, C. (1978) *J. Mol. Biol.* 118, 533-565.
 41. Bossi, L. and Roth, J.R. (1980) *Nature* 286, 123-127.
 42. Gauss, D.H. and Sprinzl, M. (1983) *Nucl. Acids Res.* 11, r1-r53.
 43. Gauss, D.H. and Sprinzl, M. (1983) *Nucl. Acids Res.* 11, r55-r103.
 44. Barrell, B.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.
 45. 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.
 46. Heckman, J.E., Sarnoff, J., Alzner-DeWeerd, B., Yin, S. and RajBhandary, U.L. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3159-3163.
-