Volume 11 Number 8 1983 Nucleic Acids Research

Transfer RNA genes in *Drosophila* mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes

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

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

Received 29 December 1982; Revised and Accepted 18 March 1983

ABSTRACT

Genes for $tRNA^{gly}$ and $tRNA^{ser}_{UCN}$ have been identified within sequences of mtDNA of Drosophila yakuba. The $tRNA^{gly}$ gene lies between the genes for cytochrome c oxidase subunit III and URF3, and all three of these genes are contained in the same strand of the mtDNA molecule. The $tRNA^{ser}_{UCN}$ gene is adjacent to the URF1 gene. These genes are contained in opposite strands of the mtDNA molecule and their 3' ends overlap. The structures of the $tRNA^{gly}$ and $tRNA^{ser}_{UCN}$ genes, and of the four tRNA genes of D. yakuba mtDNA reported earlier ($tRNA^{ile}$, $tRNA^{gln}$, $tRNA^{f-met}$ and $tRNA^{val}$) are compared to each other, to non-organelle tRNAs, and to corresponding mammalian mitochondrial tRNA genes. Within 19 nucleotides upstream from the 5' terminal nucleotide of each of the Drosophila mitochondrial $tRNA^{gly}$, $tRNA^{ser}_{UCN}$, $tRNA^{ile}$, $tRNA^{gln}$ and $tRNA^{f-met}$ genes occurs the sequence 5'TTTATTAT, or a sequence differing from it by one nucleotide substitution. Upstream from this octanucleotide sequence, and separated from it by 3, 4 and 11 nucleotides, respectively, in the 5' flanking regions of the $tRNA^{ile}$, $tRNA^{ser}_{UCN}$ and $tRNA^{gly}$ genes occurs the sequence 5'GATGAG.

INTRODUCTION

The nucleotide sequence of the whole mitochondrial DNA (mtDNA) molecule of each of three mammals, human, bovine and mouse, has been completed (1-3). Each of these mtDNAs contains the genes for two rRNAs, 22 tRNAs and five polypeptides (three subunits of cytochrome c oxidase, ATPase subunit 6 and cytochrome b). In addition, each molecule contains eight presently unidentified open reading frames (URF) and evidence has been obtained indicating that they too encode polypeptides (3,4). The gene order is essentially identical in all three of these mammalian mtDNAs (1-3).

We have recently reported the results of studies of nucleotide sequences within and on either side of the replication-origin containing A+T-rich region of the mtDNA molecule of the fly Drosophila yakuba (5). Our data indicate that while the genes found in these Drosophila sequences correspond to genes contained in mammalian mtDNAs, the order in which they are arranged in Drosophila and mammalian mtDNAs differs. Specifically, we showed that in

D. yakuba mtDNA, the small and large rRNA genes lie in tandem adjacent to that side of the A+T-rich region which is replicated first, that a $tRNA^{val}$ gene lies between the two rRNA genes, that URF1 follows the large rRNA gene, and that all of these genes are contained in the same strand of the mtDNA molecule. The genes for $tRNA^{ile}$, $tRNA^{gln}$, $tRNA^{f\text{-}met}$ and URF2 lie in the order given on the opposite side of the A+T-rich region to the rRNA genes, and except for $tRNA^{gln}$, are contained in the opposite strand of the molecule to the rRNA genes. These arrangements are in contrast to the situation found in mammalian mtDNAs, where all of these genes are located on the side of the replication origin which is replicated last, within the order $tRNA^{phe}$, small (12S) rRNA, $tRNA^{val}$, large (16S) rRNA, $tRNA^{leu}_{UUR}$, URF1, $tRNA^{ile}$, $tRNA^{gln}$, $tRNA^{f\text{-}met}$ and URF2, and except for $tRNA^{gln}$ are all contained in the same strand of the molecule.

We have also provided evidence (5) that in Drosophila mtDNA, the triplet AGA is used to specify an amino acid. This is again different from mammalian mtDNA in which AGA is used only as a rare termination codon (1-3).

In the present paper we report the nucleotide sequences of two further tRNA genes of D. yakuba mtDNA, their locations in the molecule, and the occurrence of related oligonucleotide sequences in the 5' flanking regions of D. yakuba tRNA genes. We also compare the structure of all of the Drosophila mitochondrial tRNA (mt-tRNA) genes we have sequenced to date with the structure of corresponding mammalian mt-tRNA genes.

MATERIALS AND METHODS

Stocks of D. yakuba I.C. (2371.6, Ivory Coast) and D. melanogaster (Oregon R-Utah) used in the present experiments were those used previously (5).

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

D. yakuba mtDNA was found to contain a single XhoI site which, from the results of electrophoresis of the products of single and double EcoRI, HindIII and XhoI digests of native D. yakuba mtDNA, was mapped within the EcoRI-B fragment, approximately 650 bp to the right of the EcoRI site delimiting the EcoRI-B and EcoRI-D fragments (Fig. 1).

All DNA sequences were obtained from M13mp8- or M13mp9-cloned fragments

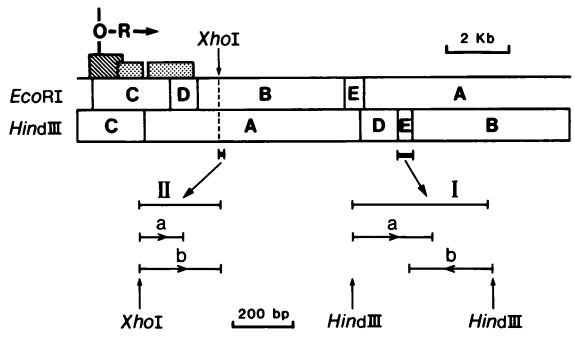

Figure 1. A map of the D. yakuba mtDNA molecule showing the relative locations of the A+T-rich region (cross-hatched), the two rRNA genes (dotted), the origin (O), and direction (R) of replication, EcoRI and HindIII sites and fragments (A-E in each case) (see (5) for details and references), and the molecule's single XhoI site (see Material and Methods). The bars under the map indicate the two segments which have been sequenced. Each of these segments (I and II) is expanded below, and the strategy employed to obtain the nucleotide sequence of each is shown. The origin of each sequence is as follows. Segment I: a and b, the two different orientations of the HindIII-E fragment, which was subcloned from the pBR325-cloned EcoRI-A fragment. Segment II: a, the larger (approximately 4,200 nucleotide pairs) XhoI-EcoRI subfragment of the pBR325-cloned EcoRI-B fragment; b, an XhoI-Sau3A subfragment of the EcoRI-B fragment. Each of the above subfragments generated by two different restriction enzymes was sequenced from the site given first.

by the extension-dideoxyribonucleotide termination procedure (6) using [α-^{32}P]dATP (800 Ci/mM; New England Nuclear) as described (5). The sequencing strategies used are given in Fig. 1.

The HindIII-D fragment of D. melanogaster mtDNA used for sequencing was cloned directly into the HindIII site of M13mp8 in each of the two possible orientations from the HindIII digestion products of native mtDNA.

Transfer RNA genes within Drosophila mtDNA sequences were identified from their ability to fold into the characteristic cloverleaf secondary structure of tRNAs, and from the trinucleotide in the anticodon position in such structures. Other genes were identified by comparing both nucleotide and predicted amino acid sequences with the corresponding sequences of previously identified genes of mouse mtDNA (3). Nucleotide sequences and amino acid sequences were analyzed by the SEQ Program (7) and the TYPIN and SEARCH Programs (8-9), respectively.

RESULTS

A gene for tRNA^{gly}, and adjacent sequences. The nucleotide sequence of the entire HindIII-E fragment (Segment I, Fig. 1) of D. yakuba mtDNA is given in Fig. 2. Within this sequence is a region with the potential to fold into the characteristic cloverleaf secondary structure of a tRNA, with an anticodon triplet indicating it to be a tRNA^{gly} gene (Fig. 3). The nucleotide sequence

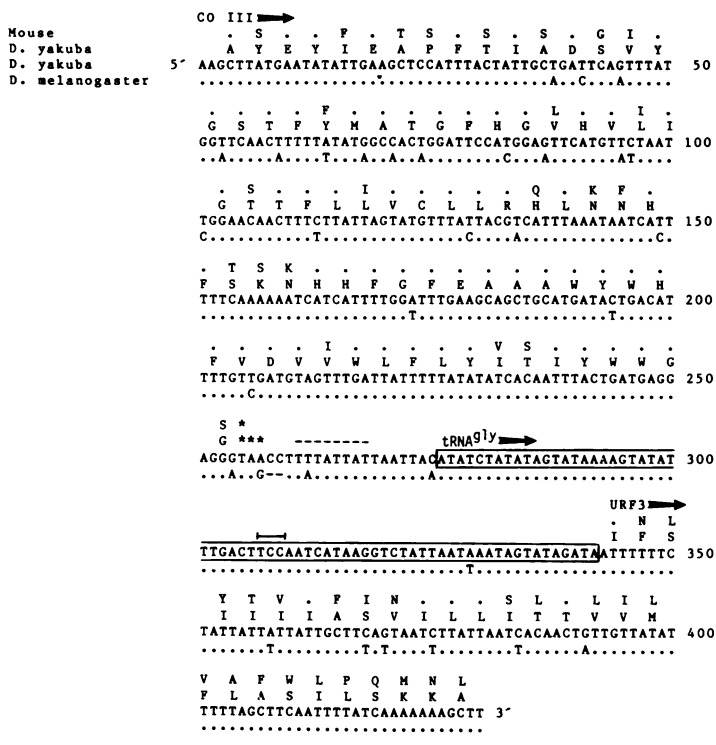

Figure 2. Nucleotide sequence (5'-3') of Segment I (Fig. 1) of the D. yakuba mtDNA molecule. The boxed segment folds into the characteristic cloverleaf structure of a tRNA (Fig. 3) with an anticodon (overlined) indicating it to be a $tRNA^{gly}$ gene. From considerations of nucleotide and predicted amino acid sequence homologies to mouse mtDNA (3), the first 254 nucleotides of the sequence shown encode the carboxyl terminal region of COIII and the last 88 nucleotides encode the amino terminal region of URF3 (see text). The corresponding amino acid sequences of D. yakuba and mouse (3) are shown. The D. yakuba COIII gene terminates at a TAA codon (***) while the mouse COIII gene terminates at a single T (*) (see text). The sequence overlined by the broken bar is the A+T-octanucleotide referred to in Fig. 5. The nucleotide sequence shown is the sense strand for COIII, $tRNA^{gly}$ and URF3, and the arrows indicate the direction of transcription of each gene. The sequence of the mtDNA molecule of D. melanogaster which corresponds to this D. yakuba sequence is also shown. Dots indicate homologous nucleotides, and a dash indicates the absence of a nucleotide.

to one side of the $tRNA^{gly}$ gene (the left as shown in Fig. 2) contains a single open reading frame which would be transcribed in the same direction as the $tRNA^{gly}$ gene. This nucleotide sequence is 71.8% homologous to the nucleotide sequence of mouse mtDNA (3) which encodes the carboxyl terminal region of cytochrome c oxidase subunit III (COIII), and the amino acid sequence

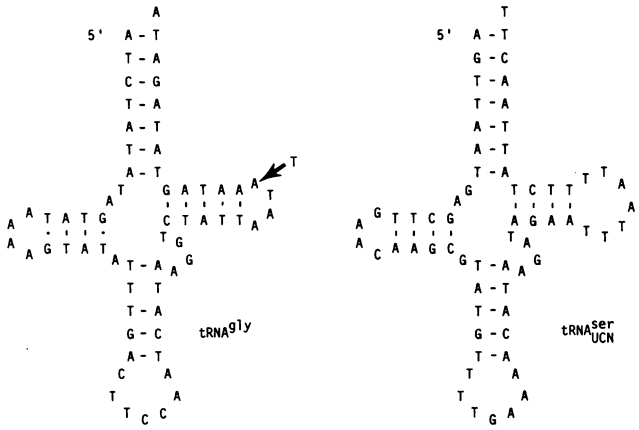

Figure 3. The genes for $tRNA^{gly}$ and $tRNA^{ser}_{UCN}$ of D. yakuba mtDNA shown in the presumed characteristic secondary structures of the corresponding tRNAs. The arrow indicates the only difference in sequence found in the $tRNA^{gly}$ gene of D. melanogaster.

predicted from it is 73% homologous to the amino acid sequence of COIII of mouse (Fig. 2). The mouse COIII gene terminates with a single T which is immediately adjacent to the 5' terminal nucleotide of the $tRNA^{gly}$ gene. It has been argued that genes on each of the complementary strands of mammalian mtDNAs are initially transcribed as a long RNA molecule and that upon cleavage of these primary transcripts, individual gene transcripts which terminate with a single U acquire a translational termination codon by polyadenylation (10). In contrast to the mouse COIII gene, the D. yakuba COIII gene terminates with a TAA codon which is separated by a sequence of 18 nucleotides from the 5' end of the $tRNA^{gly}$ gene (Fig. 2). This spacer sequence includes the octanucleotide sequence 5'TTTATTAT which was noted to occur close to the 5' end of the $tRNA^{ile}$ and $tRNA^{f\text{-}met}$ genes of D. yakuba mtDNA (5).

The nucleotide sequence of the strand which continues beyond the sense strand of the $tRNA^{gly}$ gene is 42% homologous to the region of the mouse mtDNA molecule which contains the amino terminus of URF3. The amino acid sequence predicted from the open reading frame of this nucleotide sequence which begins with the first triplet (ATT, a possible initiator codon (3)) is 21% homologous to the amino acid sequence of the corresponding region of URF3 of mouse. The interpretation of this segment of the D. yakuba HindIII-E fragment as URF3 has been confirmed by sequence analysis of the 213 nucleotides at the left end of the HindIII-D fragment (Fig. 1). This sequence (not shown) contains only one

open reading frame which proceeds in the direction away from the HindIII site common to the HindIII-E and -D fragments. The amino acid sequence predicted from this open reading frame is 48% homologous to the amino acid sequence of the section of mouse URF3 which is continuous with the amino acid sequence of the amino terminal region of the mouse URF3 gene shown in Fig 2. This observation also determines that within the D. yakuba mtDNA molecule as shown in Fig. 1, the gene order CoIII-tRNA^{gly}-URF3 is from right to left. Transcription of all of these genes (Fig. 2) is therefore in the direction opposite to that in which replication proceeds around the molecule. The arrangement of these three genes and the direction in which they are transcribed are the same as have been found for the CoIII, tRNA^{gly} and URF3 genes in mouse mt-DNA (3).

We have also obtained the nucleotide sequence of the entire HindIII-D fragment of D. melanogaster mtDNA which corresponds in length and map position (11) to the D. yakuba HindIII-E fragment. This D. melanogaster sequence is shown aligned with the D. yakuba HindIII-E nucleotide sequence in Fig. 2. There is a single nucleotide substitution between the mt-tRNA^{gly} genes of the two species, located in the TψC loop (Fig. 3). The 254 nucleotides of the 3' end of the COIII gene of the two species are 91.3% homologous. Of the 22 nucleotide substitutions between these sequences, only three have resulted in amino acid substitutions. A TAG codon is found at the end of the D. melanogaster COIII gene rather than a TAA codon which is found at this position in the D. yakuba COIII gene. In D. melanogaster, compared to D. yakuba, there are two less nucleotides (C-C) separating the 3' terminus of the COIII gene from the 5' terminus of the tRNA^{gly} gene. These observations argue against the importance of conserving one termination codon over another for a specific gene, and against the absolute length conservation of a spacer region. Also, within the D. melanogaster spacer sequence, the second nucleotide of the octanucleotide 5'TTTATTAT, found in this position of the D. yakuba sequence, is substituted by an A.

A gene for tRNA$^{ser}_{UCN}$, and adjacent sequences. The nucleotide sequence of Segment II (Fig. 1) is shown in Fig. 4. This sequence also contains a region with primary and secondary structure expected for a tRNA gene, and the triplet present in the anticodon position indicates it to be a gene for tRNA$^{ser}_{UCN}$ (Fig. 3). The sequence of the first 138 nucleotides of segment II, to the left of the tRNA$^{ser}_{UCN}$ gene (Fig. 4), is 50% homologous to the region of mouse mtDNA which contains the carboxyl terminal region of URF1. The predicted amino acid sequence of the only open reading frame of this D. yakuba nucleotide sequence and of the corresponding region of the mouse mtDNA molecule are 38% homolo-

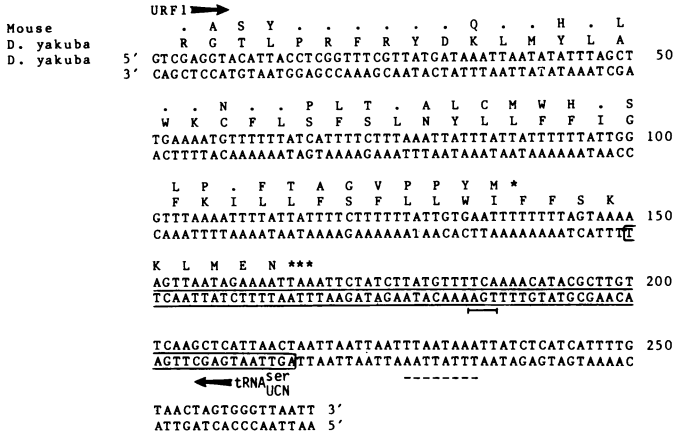

Figure 4. Nucleotide sequence of Segment II of the D. yakuba mtDNA molecule shown in Fig. 1. The boxed sequence folds into the characteristic cloverleaf structure of a tRNA (Fig. 3) with an anticodon (underlined) indicating it to be a $tRNA^{ser}_{UCN}$ gene. From considerations of nucleotide and predicted amino acid sequence homologies to mouse mtDNA (3), the sequence to the left of the $tRNA^{ser}_{UCN}$ gene encodes the carboxyl terminal region of URF1. The amino acid sequences predicted from the only open reading frame of the D. yakuba sequence and the corresponding amino acid sequence of URF1 of mouse are shown. Termination of mouse URF1 is at a single T (*) (see text). The first termination codon which occurs in the corresponding D. yakuba URF1 sequence is in a region overlapping the $tRNA^{ser}_{UCN}$ gene (TAA, indicated by ***). The sequence to the right of the $tRNA^{ser}_{UCN}$ gene has not been identified. The sequence underlined by the broken bar is the A+T-octanucleotide sequence referred to in Fig. 5. The direction of transcription of each gene is indicated by an arrow above or below the sense strand of the respective gene.

gous. Identification of this D. yakuba sequence as that which encodes the carboxyl terminal region of URF1 is consistent with the map position and orientation of segment II (Fig. 1) relative to the previously mapped position of the sequence encoding the amino terminal region of URF1 (5), assuming that the total size of the URF1 gene in mouse and D. yakuba is approximately the same.

The nucleotide sequence of the mouse URF1 gene terminates with a T which is immediately adjacent to the 5' end of the $tRNA^{ile}$ gene (3). The corresponding nucleotide in the D. yakuba sequence is also a T, but this nucleotide is separated by eleven nucleotides from the $tRNA^{ser}_{UCN}$ gene (Fig. 4) and an inframe termination codon (TAA or TAG) is not present within this sequence. A TAA codon does occur, however, in the D. yakuba URF1 sequence beginning 27 nucleotides 3' to the terminal T of the mouse URF1 gene, in the region which overlaps the $tRNA^{ser}_{UCN}$ gene. This suggests that the carboxyl terminus of the polypeptide encoded by D. yakuba URF1 extends nine amino acids beyond that of

the corresponding mouse polypeptide. This is particularly interesting as we have presented evidence (5) suggesting that the amino terminal region of the URF1 polypeptide is six amino acids shorter in Drosophila than in mouse.

The sequence of 52 nucleotides to the right of the $tRNA^{ser}_{UCN}$ gene does not contain a continuous open reading frame, and secondary structures which would indicate the presence of a tRNA gene have not been found within it. These observations, together with the fact that this region has a G+C content of 8.6%, suggest that at least the portion adjacent to the $tRNA^{ser}_{UCN}$ gene is a spacer region.

Related oligonucleotide sequences in 5' flanking regions of tRNA genes. We have noted previously that in D. yakuba mtDNA the octanucleotide sequence 5'TTTATTAT, which lies immediately adjacent to the 5' end of the sense strand of the $tRNA^{f\text{-}met}$ gene and separates this gene from the transcribed strand of the $tRNA^{gln}$ gene, also occurs at the end of the A+T-rich region and is separated by only one nucleotide from the 5' end of the sense strand of the $tRNA^{ile}$ gene (5). We have noted above that this octanucleotide sequence (5'TATATTAT in D. melanogaster) also occurs in the spacer region between the COIII gene and the $tRNA^{gly}$ gene, separated from the 5' end of the sense strand of the latter by nine nucleotides. The 5' flanking sequences of the sense strands of all six D. yakuba mt-tRNA genes which we have sequenced are shown in Fig. 5. Within the sequences upstream from the 5' ends of the sense strands of the genes for $tRNA^{ser}_{UCN}$ and $tRNA^{gln}$ there occurs an octanucleotide sequence differing by only one nucleotide from 5'TTTATTAT, and separated from the $tRNA^{ser}_{UCN}$ and $tRNA^{gln}$ genes by eleven and three nucleotides respectively. The sequence preceding the $tRNA^{gln}$ gene partially overlaps the sense strand of the $tRNA^{f\text{-}met}$ gene while the sequence preceding the $tRNA^{ser}_{UCN}$ gene is within a presently unidentified (presumptive spacer) region.

The $tRNA^{val}$ gene differs from the other five tRNA genes sequenced in that its 5' terminal nucleotide is immediately adjacent to the 3' terminal nucleotide of the gene (small rRNA) which precedes it (5). Examination of the 222 nucleotides of the 3' end of the small rRNA gene (Fig. 5 and Fig. 4 of ref. 5) reveals that neither the sequence 5'TTTATTAT nor a sequence differing from it by one nucleotide occurs in this entire sequence.

In each of the 5' flanking sequences of the genes for $tRNA^{gly}$, $tRNA^{ser}_{UCN}$ and $tRNA^{ile}$, upstream from the A+T octanucleotide and separated from it by 11, 4 and 3 nucleotides, respectively, occurs the sequence 5'GATGAG (Fig. 5). The 5'GATGAG sequence which precedes the $tRNA^{gly}$ gene (in both D. yakuba and D. melanogaster) is included in the sense strand of the COIII gene. Neither

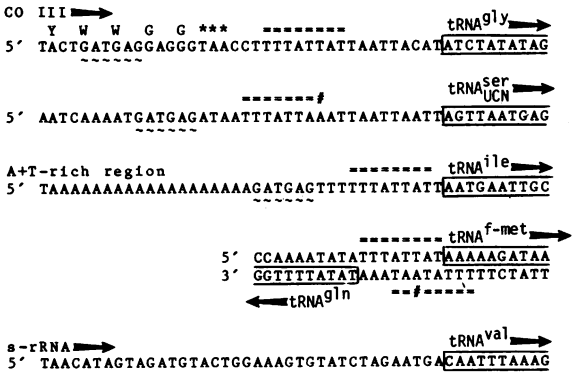

Figure 5. Nucleotide sequences of the sense (non-transcribed) strands of the D. yakuba mtDNA molecule which include and precede the 5' end of six tRNA genes. Above the left end of the uppermost sequence the five terminal amino acids and the termination codon (***) of COIII (see Fig. 2) are indicated. The broken double bars above or below each sequence indicate either the sequence 5'TTTATTAT or a sequence differing from this octanucleotide by one nucleotide substitution (cross lines). The wavy underlines indicate the sequence 5'GATGAG.

of the 5' flanking sequences of the genes for $tRNA^{f\text{-}met}$ or $tRNA^{gln}$ contains 5'GATGAG. These flanking sequences include the entire antisense strand of the $tRNA^{gln}$ gene and of the $tRNA^{f\text{-}met}$ gene, respectively (5). Also, 5'GATGAG does not occur in the sense strand of the 3' end of the small rRNA gene which precedes the $tRNA^{val}$ gene.

Structure of D. yakuba mitochondrial tRNA genes and comparisons to corresponding mammalian tRNA genes. The D. yakuba mitochondrial genes for $tRNA^{gly}$ and $tRNA^{ser}_{UCN}$ have a similar general structure to the mitochondrial genes for $tRNA^{ile}$, $tRNA^{gln}$, $tRNA^{f\text{-}met}$ and $tRNA^{val}$ reported earlier (5). All of the D. yakuba mt-tRNA genes (and the tRNAs predicted from them) are similar in size and structure to their mammalian counterparts. The D. yakuba genes vary in size from 65 nucleotides (ile and gly) to 72 nucleotides (val) and, as in mammalian mt-tRNA genes, the sizes of the amino-acyl stem (seven nucleotides), the anticodon stem (five nucleotide pairs) and the anticodon loop (seven nucleotides) are strictly conserved. The dihydrouridine stem varies in size from three to four nucleotide pairs and the dihydrouridine loop varies from four to eight nucleotides. The so-called variable loop is either four or five nucleotides in D. yakuba mt-tRNA genes, and this number is always the same as that found in the corresponding mammalian mt-tRNA gene. In the D. yakuba mt-tRNA genes the TψC stem is four or five nucleotide pairs, and the

TψC loop varies from four to seven nucleotides. As in mammalian mt-tRNA genes, the trinucleotide sequence CCA, which occurs at the 3' terminus of all non-organelle tRNA genes, is absent from D. yakuba mt-tRNA genes.

As is the case in mammals, the D. yakuba mt-tRNA genes differ among themselves in regard to the presence or absence of various nucleotides which are conserved in prokaryotic and non-organelle eukaryotic tRNAs (12, 13). Only the conserved Pu_{26}, T_{33} and Pu_{37} nucleotides (numbering system in (14)) are present in all six D. yakuba tRNA genes. None of the conserved nucleotides are absent from all of the Drosophila tRNA genes, although C_{56} in the TψC loop is present only in the $tRNA^{val}$ gene. Also, while the conserved G_{18}-G_{19} pair is not found in the dihydrouridine loop of any of these tRNA genes, it appears likely that G_{18} is conserved in both $tRNA^{gly}$ and $tRNA^{gln}$, and that G_{19} is conserved in $tRNA^{val}$. All of the D. yakuba mt-tRNA genes contain the conserved Py_{11}-Pu_{24} pair except $tRNA^{f\text{-}met}$ where a G-C pair is found in this position. In prokaryotic $tRNA^{f\text{-}met}$ molecules, nucleotides 11 and 24 are also a Pu-Py pair, but always A-U (15).

In non-organelle tRNAs, tertiary bonding occurs between a number of nucleotides including three sets of conserved nucleotides, T_8 and A_{14}, G_{18} and T_{55}, and G_{19} and C_{56} (14,16). The absence of one or more of these nucleotides in each of the D. yakuba mt-tRNAs predicted from their genes clearly suggests that tertiary bonding is different, and is possibly weaker, in all of these mt-tRNAs than in prokaryotic and non-organelle eukaryotic tRNAs.

Comparisons were made of homologies of the nucleotide sequences of each of the six D. yakuba mt-tRNA genes and the corresponding mt-tRNA genes of mouse, rat, human and bovine (1-3, 17-19). Overall homologies are in the ranges, 62-67% ($tRNA^{f\text{-}met}$), 56-61% ($tRNA^{ile}$), 51-57% ($tRNA^{gln}$), 49-60% ($tRNA^{val}$), 48-57% ($tRNA^{gly}$) and 38-44% ($tRNA^{ser}_{UCN}$). Homologies between the different stems and loops of the mt-tRNAs of D. yakuba and mammals show considerable variation among the different genes. Overall, the greatest homology is found among the dihydrouridine stems and the anticodon loops. The least overall homology is found among amino-acyl stems and TψC stems.

In view of the very low homologies (38-44%) found between the D. yakuba mt-$tRNA^{ser}_{UCN}$ gene and the mt-$tRNA^{ser}_{UCN}$ genes of mouse, rat, human and bovine, we compared the D. yakuba mt-$tRNA^{ser}_{UCN}$ gene to each of the 21 other mt-tRNA genes of mouse. With one exception, homologies resulting from these comparisons were within the range 29%-46%. The one exception resulted from the comparison of the D. yakuba mt-$tRNA^{ser}_{UCN}$ gene to the mouse mt-$tRNA^{tyr}$ gene, which indicated a homology of 54%. Comparable homologies (49% to 56%) occur between the

D. yakuba mt-tRNA$^{ser}_{UCN}$ gene and the mt-tRNA^{tyr} genes of rat, human and bovine (1-3, 17-19). As noted above for other corresponding mt-tRNA genes of D. yakuba and mammals, the dihydrouridine stems of the D. yakuba mt-tRNA$^{ser}_{UCN}$ gene and the mammalian mt-tRNA^{tyr} gene are highly homologous.

It has been noted that the mammalian mt-tRNAs have a strikingly low G+C content compared to non-organelle tRNAs, and an unusually high content of mismatched base pairs in the stem regions, both of which must contribute to a weakening of the overall tertiary (and secondary) structure of the tRNAs (2). The G+C content of the D. yakuba mt-tRNA genes we have reported range from 18.8% to 33.8%, with a mean of 24.2%, which is close to the mean G+C content (approximately 25%) of the D. yakuba (and other Drosophila) mtDNA molecules outside of the A+T-rich region (20,21). With one exception (tRNA^{ile}), the G+C content and the frequency of G-C nucleotide pairs are less in each D. yakuba mt-tRNA gene than in the corresponding mt-tRNA gene of mouse. [Mouse mtDNA is chosen for comparisons in all of our studies, because its average G+C content of 36.7% (3) is the lowest of the three mammalian mtDNAs sequenced to date.]

The tRNAs which would be predicted from the mitochondrial tRNA^{gly}, tRNA^{ile}, tRNA^{gln}, tRNA$^{f\text{-}met}$ and tRNA^{val} genes of D. yakuba contain an average of one mismatched nucleotide pair (Fig. 3; (5)), while the corresponding mammalian tRNAs contain an average of 2.1 mismatched nucleotide pairs (3,22,23). In Drosophila and mammalian mt-tRNAs the majority of these mismatched nucleotide pairs (45%) are G·U pairs, as has been found in prokaryotic and non-organelle eukaryotic tRNAs (15). The tRNA predicted from the D. yakuba mt-tRNA$^{ser}_{UCN}$ gene (Fig. 3) does not contain any mismatched nucleotide pairs, a situation which again more closely resembles that found in the four mammalian mt-tRNA^{tyr} molecules which contain an average of only 0.5 mismatched nucleotide pairs, than that found in the four mammalian mt-tRNA$^{ser}_{UCN}$ molecules which contain an average of 6.0 mismatched nucleotide pairs (3,22,23).

Between the corresponding mitochondrial tRNA^{gly}, tRNA^{val}, tRNA^{ile}, tRNA^{gln} and tRNA$^{f\text{-}met}$ genes of D. yakuba and mouse, there is a total of 139 nucleotide substitutions (Table 1). Overall, there are approximately equal numbers of transitions and transversions. However, while there are approximately equal frequencies of the two kinds of transitions, one kind of transversion (A↔T) predominates, and is the most frequent kind of nucleotide substitution. This contrasts with the distribution of nucleotide substitutions found between mt-tRNA genes of human and bovine, where transitions predominate, and the most frequent transversion is an A↔C substitution.

Table 1: Frequency distributions of various nucleotide substitutions (within the sense strand) among five tRNA genes of mtDNAs of D. yakuba and mouse.

	tRNA gene					Totals	
Transitions	gly	val	ile	gln	f-met	Whole gene	Stems only
A↔G	8	4	7	10	4	33 (23.7%)	28 (26.7%)
C↔T	8	7	7	5	8	35 (25.2%)	26 (24.8%)
Total	16	11	14	15	12	68 (48.9%)	54 (51.4%)
Transversions							
A↔C	3	5	0	1	1	10 (7.2%)	7 (6.7%)
A↔T	11	9	6	15	9	50 (36.0%)	35 (33.3%)
G↔C	0	2	2	0	0	4 (2.9%)	4 (3.8%)
G↔T	2	3	0	1	1	7 (5.0%)	5 (4.8%)
Total	16	19	8	17	11	71 (51.1%)	51 (48.6%)
Total Substitutions	32	30	22	32	23	139 (100%)	105 (100%)

The data were assembled from sequences given in Figs. 2 and 4 and references 3 and 5. The frequencies (as percentages) of each of the six different nucleotide substitutions among whole genes and among stem regions of these genes are given in parentheses.

A↔C substitutions are almost as rare between D. yakuba and mouse tRNA genes as are G↔C and G↔T substitutions, which are the rarest substitutions occurring between human and bovine mt-tRNA genes (2). Between D. yakuba and mouse mt-tRNA genes, 75.5% of all nucleotide substitutions are in stems, and their distributions among the different kinds of transitions and transversions very clearly reflect the overall distributions (Table 1). Among the five mt-tRNA genes compared, nucleotide substitutions occur within 55 nucleotide pairs. Among these nucleotide pairs, nucleotide substitutions are compensated to retain pairing in 42 (76.4%) cases. Of the remaining 13, five are changes between a standard nucleotide pair to a G·T pair, all involving a transversion.

DISCUSSION

The tRNAs predicted from the mt-tRNA genes of D. yakuba which we have described to date show similar general properties of size and structure to the corresponding mt-tRNAs of mammals. The characteristic cloverleaf secondary structure is maintained, and within it only the sizes of the dihydrouridine loop and the TψC loop show extensive variations. Also, as has been noted for

mammalian mt-tRNAs (2,3), the occurrence among Drosophila mt-tRNAs of the nucleotides which are conserved in prokaryotic and non-organelle eukaryotic tRNAs is highly variable. The stability of the various stems of most of the Drosophila mt-tRNAs is less dependent on G-C bonds than is the case in mammals. This is somewhat compensated, however, by the lower frequency of mismatched nucleotide pairs in the stem regions of Drosophila mt-tRNAs. In the absence of information regarding tertiary interaction in Drosophila mt-tRNAs, it is not possible to draw conclusions regarding either the stability or flexibility of these tRNAs relative to their non-organelle counterparts.

Five of the six D. yakuba mt-tRNA genes sequenced to date are in the same location relative to at least one other gene in both D. yakuba and mammals. The one exception is $tRNA^{ser}_{UCN}$. In mammals this gene is found between the genes for COI and $tRNA^{asp}$ (1-3), while in D. yakuba it is located at the carboxyl terminus of URF1. Although the nucleotide sequence on the other side of $tRNA^{ser}_{UCN}$ has not been identified, this sequence shows no convincing homology to either the COI gene or the $tRNA^{asp}$ gene. The mt-$tRNA^{ser}_{UCN}$ gene of D. yakuba is of further interest as its overall homology and the homologies of specific regions of it are greater to the mouse mt-$tRNA^{tyr}$ gene than to any other mouse mt-tRNA gene, including $tRNA^{ser}_{UCN}$.

Our data establish that within 19 nucleotides upstream from the 5' terminal nucleotide of five of the mt-tRNA genes of D. yakuba and the mt-$tRNA^{gly}$ gene of D. melanogaster (none of which are immediately preceded by another gene) there occurs the octanucleotide sequence 5'TTTATTAT or a sequence differing from it by one nucleotide substitution. In the 5' flanking sequences of three of the mt-tRNA genes, the hexanucleotide 5'GATGAG is also found upstream from the A+T-octanucleotide and separated from the latter by a maximum of eleven nucleotides. In mammalian mtDNA molecules (1-3) there are very few nucleotides in the regions which lie between genes, and related sequences in different intergenic regions have not been reported. It seems plausible that one or both of the two sets of related oligonucleotides in the 5' flanking regions of D. yakuba tRNA genes could have a promoter function in regard to initiation of transcription of the segments of the molecule which follow them. Both A+T-rich and G+C-rich homologous, oligonucleotide sequences are included in the upstream promoter regions of a large number of prokaryotic and eukaryotic genes (24-26). Alternatively, as has been reported for homologous sequences preceding different nuclear tRNA genes of D. melanogaster (27), the related 5' flanking sequences of Drosophila mt-tRNA genes may be

involved in determining the level at which the genes they precede are transcribed.

ACKNOWLEDGEMENTS

We thank Raymond F. Gestland and George B. Spiegelman for valuable discussions. This work was supported by National Institutes of Health Grants Nos. GM 18375 and RR 07092.

REFERENCES

1. 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.
2. 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.
3. Bibb,M.J., Van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) Cell 26, 167-180.
4. Montoya,J., Ojala,D. and Attardi,G. (1981) Nature 290, 465-470.
5. Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M.-R. and Wolstenholme, D.R. (1982) Nucl. Acids Res. 10, 6619-6637.
6. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
7. Brutlag,D.L., Clayton,J., Friedland,P. and Kedes,L. (1982) Nucl. Acids Res. 10, 279-294.
8. Jue,R.A., Woodbury,N.W. and Doolittle,R.F. (1980) J. Mol. Evol. 15, 129-148.
9. Doolittle,R.F. (1981) Science 214, 149-159.
10. Ojala,D., Montoya,J. and Attardi,G. (1981) Nature 290, 470-474.
11. Fauron,C.M.-R. and Wolstenholme,D.R. (1980) Nucl. Acids Res. 8, 2439-2452.
12. Sprinzl,M. and Gauss,D.H. (1982) Nucl. Acids Res. 10, r1-r55.
13. Sprinzl,M. and Gauss,D.H. (1982) Nucl. Acids Res. 10, r57-r81.
14. Kim,S.-H., Suddath,F.L., Quigley,G.J., McPherson,A., Sussman,J.L., Wang,A., Seeman,N.C. and Rich,A. (1974) Science 185, 435-440.
15. Clark,B.F.C. (1978) in Transfer RNA, Altman,S., Ed., pp. 14-47, The MIT Press, Cambridge.
16. Kim,S.H. (1979) in Transfer RNA: Structure, properties and recognition, Schimmel,P.R., Soll,D. and Abelson,J.N., Eds., pp.83-100, Cold Spring Harbor Laboratory.
17. Wolstenholme,D.R., Fauron,C.M.-R. and Goddard,J.M. (1982) Gene, 20, 63-39.
18. Saccone,C., Cantatore,P., Galadeta,G., Gallerani,R., Lanave,C., Pepe,G. and Kroon,A.M. (1981) Nucl. Acids Res. 9, 4139-4148.
19. Cantatore,P., DeBenedetto,C., Galadeta,G., Gallerani,R., Kroon,A.M., Holtrop,M., Lanave,C., Pepe,G., Quagliariello,C., Saccone,C. and Sbisa,E. (1982) Nucl. Acids Res. 10, 3279-3289.
20. Fauron,C.M.-R. and Wolstenholme,D.R. (1976) Proc. Natl. Acad. Sci. USA 73, 3623-3627.
21. Wolstenholme,D.R., Fauron,C.M.-R. and Goddard,J.M. (1980) in The Organization and Expression of the Mitochondrial Genome, Developments in Genetics, Saccone,C. and Kroon,A.M., Eds., Vol. 2, pp. 241-250, Elsevier/North Holland Biomedical Press, Amsterdam.

22. Van Etten,R.A., Michael,N.L., Bibb,M.J., Brennicke,A. and Clayton,D.A. (1982) in Mitochondrial Genes, Slonimski,P., Borst,P. and Attardi,G., Eds., pp. 73-88, Cold Spring Harbor Laboratory.
23. Anderson,S., Bankier,A.T., Barrell,B.G., DeBruijn,M.H.L., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., Schrier,P.H., Smith,A.J.H., Staden,R. and Young,I.G. (1982) in Mitochondrial Genes, Slonimski,P., Borst,P. and Attardi,G., Eds., pp. 5-49, Cold Spring Harbor Laboratory.
24. Rosenberg,M. and Court,D. (1979) Ann. Rev. Genet. 13, 319-353.
25. Breathnach,R. and Chambon,D. (1981) Ann. Rev. Biochem. 50, 349-383.
26. Seeburg,P.H., Nusslein,C. and Schaller,H. (1977) Eur. J. Biochem. 74, 107-113.
27. DeFranco,D., Sharp,S. and Soll,D. (1981) J. Biol. Chem. 256, 12424-12429.