Mytilus Mitochondrial DNA Contains a Functional Gene for a tRNASer(UCN) With a Dihydrouridine Arm-Replacement Loop and a pseudo-tRNASer(UCN) Gene

C. Timothy Beagley, Ronald Okimoto¹ and David R. Wolstenholme

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

Manuscript received September 21, 1998
Accepted for publication March 5, 1999

ABSTRACT

A 2500-nucleotide pair (ntp) sequence of F-type mitochondrial (mt) DNA of the Pacific Rim mussel Mytilus californianus (class Bivalve, phylum Mollusca) that contains two complete (ND2 and ND3) and two partial (COI and COIII) protein genes and nine tRNA genes is presented. Seven of the encoded tRNAs (Ala, Arg, His, Met(AUA), Pro, Ser(UCN), and Trp) have the potential to fold into the orthodox four-armed tRNA secondary structure, while two [tRNAs(AGN) and a second tRNAs(UCN)] will fold only into tRNAs with a dihydrouridine (DHU) arm-replacement loop. Comparison of these mt-tRNA gene sequences with previously published, corresponding M. edulis F-type mtDNA indicates that similarity between the four-armed tRNAs(UCN) genes is only 63.8% compared with an average of 92.1% (range 86.2-98.5%) for the remaining eight tRNA genes. Northern blot analysis indicated that mature tRNAs encoded by the DHU arm-replacement loop-containing tRNAs(UCN), tRNAs(AGN), tRNAs(AUA), tRNAs(TG), and tRNAs(TG) genes occur in M. californianus mitochondria, strengthening the view that all of these genes are functional. However, Northern blot and 5' RACE (rapid amplification of cDNA ends) analyses indicated that the four-armed tRNAs(UCN) gene is transcribed into a stable RNA that includes the downstream COI sequence and is not processed into a mature tRNA. On the basis of these observations the M. californianus and M. edulis four-armed tRNAs(UCN) sequences are interpreted as pseudo-tRNAs(UCN) genes.

The mitochondrial (mt) genomes of almost all multicellular animals (metazoans) comprise a single circular molecule of 14-42 kb (Wolstenholme 1992a; Wolstenholme and Faún 1995). The mtDNA molecules of ~75 metazoans have been completely sequenced. Approximately two-thirds of these are from chordates (Krettek et al. 1995; Arndson et al. 1996; Zar doya and Meyer 1997) but the remainder represent a variety of invertebrate phyla, including Cnidaria, Nematoda, Arthropoda, Mollusca, Annelida, and Echinodermata (see, for example, Clary and Wolstenholme 1985; Jacobs et al. 1988; Cantatore et al. 1989; Garay and Wolstenholme 1989; Okimoto et al. 1991, 1992; Crozier and Crozier 1993; Boore and Brown 1994, 1995; Asakawa et al. 1995; Beagley et al. 1995, 1998; Hatzoglou et al. 1995; Terrett et al. 1996).

Most metazoan mtDNAs contain the same set of genes for 13 energy-pathway proteins: cytochrome b (Cyt b); subunits I-III of cytochrome c oxidase (COI-COIII); subunits 6 and 8 of F0 ATP synthetase (ATPase6 and ATPase8); subunits 1-6 and 4L of the respiratory chain NADH dehydrogenase (ND1-ND6 and ND4L); and two rRNAs that are homologous to the 16S and 23S ribosomal RNAs (rRNAs) of Escherichia coli (but smaller and usually referred to as s-rRNA and l-rRNA) and 22 transfer RNAs (tRNAs). The arrangement of genes within metazoan mtDNAs is very compact. Differences in gene arrangement occur among metazoan mtDNAs, and the extent of these differences appears to be a function of evolutionary distances (Wolstenholme 1992a; Boore and Brown 1994, 1995).

Metazoan mt-genomes are characterized by a number of unusual features. These include genetic code modifications, the use of unorthodox translation initiation codons, post-transcriptional generation of translation termination codons, translation by only 22 tRNAs that are mtDNA encoded and of diverse size and structural form, and replication by a unique asymmetrical mode (Clyton 1992; Wolstenholme 1992a, b; Wolstenholme and Faún 1995).

Hoffmann et al. (1992) determined nucleotide sequences that collectively represent ~80% of the ~17.1-kb mtDNA molecule of the blue mussel Mytilus edulis (phylum Mollusca, class Bivalvia). Because these sequences contained the terminal 5' and 3' ends of all protein and rRNA genes and all sequences between these genes including tRNA genes, they provided sufficient information to construct a complete gene map.

Corresponding author: David R. Wolstenholme, Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112-0840. E-mail: wolstenholme@biology.utah.edu

¹Present address: Department of Poultry Science, University of Arkansas, Fayetteville, AK 72701.

Genetics 152: 641-652 (June 1999)
The *M. edulis* mtDNA molecule was found to contain the gene complement common to fully mapped metazoan mtDNAs with two exceptions. The first is that there are two genes for methionine-specifying tRNAs: one with a CAU anticodon (tRNA$^{Met}$), as has been reported for all other metazoan mtDNAs so far examined, and a second with a UAU anticodon that might be expected to specifically recognize AUA codons, and is so far unique among metazoan mtDNAs. The second exception to gene content is that a gene for ATPase8 was not identified, a situation previously known among Metazoa only for nematode mtDNAs (Wolstenholme et al. 1987; Oki moto et al. 1991, 1992). Relative to the corresponding genes of Drosophila yakuba and mouse, considerable extensions were postulated at the 5' ends of the COI and ND1 genes, and at the 3' ends of the COI and COIII genes of *M. edulis*. The COIII gene 3' extension was of particular interest as the overall size of this gene and the C-terminal proximal amino acid sequence of the protein it encodes have been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined. Gene arrangement in the mtDNA molecule of *M. edulis* was found to be greatly different from gene arrangement in other sequenced metazoan mtDNAs, including those of the more recently sequenced polyplacophoran mollusk *Katherine tunicata* and the gastropod mollusks *Albinaria coerules* and *Cepaea nemoralis* (Boore and Brown 1994; Hatzoglou et al. 1997). One mtDNA type (F) is found in both females and males and, at the 3' end of the COI gene, the C-terminal proximal amino acid sequence of the protein it encodes has been highly conserved in all organisms previously examined.
Double-stranded cDNA that included the 5' sequence within the 5' oligonucleotide F4 [which is complementary to a nucleotide primer TB17-2 (5'9mCOI transcript was accomplished in a 100-mM Kingdon Bgl transcription conditions but using as primer 20 p by incubation in a 50-

mM formaldehyde-containing 3% agarose gel, transferred to a nitrocellulose filter, and hybridized with 32P-labeled gene sequence-specific oligonucleotides. A similar procedure was used to determine and compare the sizes of RNAs containing transcripts of the ND3, COI, and \( \omega \) tRNA\(_{Ser}(UCN) \) genes, except that electrophoresis was carried out in a 2.2 M formaldehyde-containing 1% agarose gel.

For precise determination of RNA sizes, M. californianus mtRNA (10 \( \mu \)g/lane) was electrophoresed in polyacrylamide gels, electroblotted, and probed with gene sequence-specific oligonucleotides as described in Kimoto and Wolstenholme (1990), with the following exceptions: the 15% polyacrylamide gel (containing 8 \( \mu \)m urea) was 17 cm wide by 15 cm long; electrophoresis was continued until the cyano dye ran completely off the gel; the blots were washed twice at 25°C in 2\( \times \) SSC, 0.1% SDS for 10 min, followed by two washes at hybridization temperature (37°C rather than 35°C) in 0.1\( \times \) SSC, 0.1% SDS for 30 min each.

**Location of the 5' end of the COI transcript:** 5' RACE (rapid amplification of cDNA ends) analysis (Frohman 1990) was carried out to locate the 5' end of COI transcript-containing RNA molecules as follows: M. californianus whole mtRNA (10 \( \mu \)g) was reverse transcribed in a 100-\( \mu \)l reaction using 1.5 \( \mu \)l random hexamers (BRL) and 1000 units of Moloney-MLV reverse transcriptase under conditions recommended by the supplier (Gibco BRL, Gaithersburg, MD). Incubation was for 10 min at 25°C followed by 30 min at 42°C.

Following extraction of the incubation product with phenol, chloroform, cDNAs were purified using a Centricon 10 microconcentrator (Amicon, Beverly, MA). Deoxynucleosine monophosphates (dAs) were added to the 3' ends of cDNAs by incubation in a 50-\( \mu \)l reaction containing cDNAs, 2.0 \( \mu \)l of 10 mm dATP, and 34 units of terminal deoxynucleotidyld transferase (Amersham) under the supplier's recommended reaction conditions for 5 min at 37°C. Following a further phenol/chloroform extraction and Centricon purification, a second-strand synthesis reaction was carried out under the reverse transcription conditions but using as primer 20 pm of BamHI, BglII, and PstI site-containing adaptor oligonucleotide (TB17-1: 5' CCAGATCGGATCCGTCTGACGATTGTTTTTTTTTTTTTTTTTT; Pont-Kingdon et al. 1994). The product was phenol/chloroform extracted, Centricon purified, and taken up in 50 \( \mu \)l of Tris-EDTA (TE) to provide a 5' RACE pool. Amplification of the double-stranded cDNA that included the 5' end region of the COI transcript was accomplished in a 100-\( \mu \)l PCR (in a DNA thermal cycler (Perkin Elmer-Cetus, Norwalk, CT) using 0.5 \( \mu \)l of the RACE pool, 25 pm of a BamHI, BglII, PstI site-containing primer TB17-2 (5' CCAGATCGGATCCGTCTGACG), 25 pm of oligonucleotide F4 which is complementary to a nucleotide sequence within the 5' end region of the M. californianus COI gene (Figure 1A), and an annealing temperature of 58°C. The PCR product was digested with BamHI, inserted into BamHI-cleaved M13mp19, and sequenced.

**RESULTS AND DISCUSSION**

Complete 2500-ntp sequences were obtained from two reversely oriented inserts of a PCR-generated 3.2-kb M. californianus mtDNA segment, cloned in M13mp19. As the two 2500-ntp sequences obtained were identical, we concluded that they both originated from the same mtDNA type. That this was the F-type gene sequence with the equivalent COI F-type and COI M-type sequences identified by Beagley et al. (1997).

The entire, continuous 2500-ntp sequence of M. californianus F-type mtDNA is given in Figure 1A, and is compared with two sequences of M. edulis F-type mtDNA reported by Hoffmann et al. (1992). We interpret the M. californianus sequence to contain the following genes in the order given: COIII (3' end region), tRNA\(_{Ser}(UCN)\), tRNA\(_{Met}(AUA)\), ND2, tRNA\(_{Arg}\), tRNA\(_{Lys}\), tRNA\(_{His}\), tRNA\(_{Pro}\), ND3, tRNA\(_{Ser}(UCN);\) (see below), and COI (5' end region). The two M. edulis sequences, collectively, have the same gene content, but the sequence of a central region of the ND2 gene was not determined (Figure 1B).

**Protein genes:** Hoffmann et al. (1992) found that the M. edulis COI gene-containing open reading frame extended upstream beyond the codon corresponding to the translation initiation codons of the mouse and D. yakuba COI genes, for 21 and 22 codons, respectively. They suggested that, based on available information, a number of different triplets might act as the M. edulis COI translation initiation codon, and tentatively assigned this function to the ATA codon located 8 and 9 codons, respectively, upstream from the mouse and D. yakuba COI initiation codons, and at the same relative location as the Caenorhabditis elegans and Ascaris suum COI initiation codons. However, neither this ATA codon nor the ATT and ATA codons at the beginning of the M. edulis COI-gene containing open reading frame are conserved in the M. californianus sequence (Figure 1A). Therefore, from considerations of nucleotide sequence and predicted amino acid sequence conservation, it seems most likely that translation initiation of the COI gene occurs at the ATG codon in the open reading frame of both Mytilus species, located 17 and 18 codons upstream from the mouse and D. yakuba translation initiation codons (Figure 1A).

Data from nucleotide and amino acid sequence comparisons of the complete M. californianus and M. edulis ND3 genes, corresponding segments of the complete M. californianus ND2 gene and the two sequences of the M. edulis ND2 gene, and the portions of the M. californianus and M. edulis COIII genes that extend 3' relative to other metazoan COIII genes are given in
Figure 1.—(A) Comparison of a continuous 2500-ntp sequence of the F-type mtDNA molecule of M. californianus (M.c) and two corresponding, partial F-type mtDNA sequences of M. edulis (M.e) taken from Hoffmann et al. (1992) (GenBank accession nos. M83760 and M83761). The numbering to the right is shown only for the M. californianus sequence. All of the partial and complete protein genes shown (COI and COIII, cytochrome c oxidase subunits I and III; ND2 and ND3, NADH dehydrogenase...
Mytilus Mitochondrial DNA

Table 1. For the ND3 and ND2 genes, nucleotide and amino acid sequence similarities are 80.2 and 73.2%, and 94 and 80.7%, respectively. The corresponding values for comparisons of the M. californianus and M. edulis COIII 3' extensions are intermediate to those obtained for the ND3 and ND2 genes: 79.9 and 91.7%, respectively. Between the ND3 and ND2 genes of the two species, silent substitutions are in excess of replacement substitutions, and in silent positions, transitions exceed transversions, as has been reported for mt-protein genes of closely related species of other invertebrates and vertebrates (Brown et al. 1982; Brown and Simpson 1982; Desalle et al. 1987). Between the M. californianus and M. edulis COIII extension sequences the proportion of substitutions that are silent, and the proportion of silent substitutions that are transitions are both greater than found for the ND3 and ND2 gene comparisons (Table 1). Taken together, similarities between patterns of nucleotide substitution among the COIII gene extensions and the ND3 and ND2 genes suggest that the COIII extension sequences have been subjected to selective pressures similar to those that have operated on the ND3 and ND2 gene sequences. Therefore, these data are consistent with the view (Hoffmann et al. 1992) that the extra 46 codons of the M. californianus and M. edulis COIII genes encode a unique functional segment of the Mytilus COIII protein.

Transfer RNA genes: Comparisons of the sequences of the M. californianus and M. edulis tRNA genes are given in Figure 2 and Table 2. The sequences of eight of the tRNA genes[Ala, Arg, His, Met, Pro, Trp, Ser(AGN), and Ser(UCN)] are highly conserved in regard to primary structure (mean, 92.1%; range, 86.2–98.5%; Table 2) and their potential to fold into secondary structures with features characteristic of mt-tRNAs (Figure 2). Compared with the M. edulis sequences, the M. californianus sequences have an average of only 4.5 nucleotide substitutions (range 1–9), 0.4 nucleotide deletions

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. of nucleotides compared</th>
<th>Nucleotide sequence similarity (%)</th>
<th>Amino acid sequence similarity (%)</th>
<th>Silent substitutions (%)</th>
<th>Transition:transversion ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND3</td>
<td>348</td>
<td>80.2</td>
<td>94.0</td>
<td>82.6</td>
<td>2.4:1</td>
</tr>
<tr>
<td>ND2</td>
<td>642</td>
<td>73.2</td>
<td>80.7</td>
<td>71.5</td>
<td>1.8:1</td>
</tr>
<tr>
<td>COIII</td>
<td>144</td>
<td>79.9</td>
<td>91.7</td>
<td>86.2</td>
<td>5.3:1</td>
</tr>
</tbody>
</table>

Calculated from data in Figure 1. Data for M. edulis are from Hoffmann et al. (1992).

a Excluding termination codons.

b Extension only.
The eight mt-tRNA genes and the $\psi$tRNA$^{Ser}$(UCN) gene of $M$. californianus contained in the sequence in Figure 1, and the $\psi$tRNA$^{Ser}$(UCN) gene of $M$. edulis (Hoffmann et al. 1992), shown in the predicted secondary structural forms of the corresponding tRNAs. Each solid line on the outside of some $M$. californianus tRNA genes, and the solid line (F5) on the outside of the $M$. californianus $\psi$tRNA$^{Ser}$(UCN) gene, identifies a segment to which a complementary oligonucleotide was made and used in hybridization experiments to test for the presence of the various corresponding tRNAs. The identity (derived from the anticodon) and size (in nucleotides, nt) of each tRNA gene are shown. In the DHU arm-replacement loop-containing tRNA$^{Ser}$(AGN) gene, possible secondary interactions between nucleotides in the DHU arm-replacement loop and the variable loop are shown by dashed lines. Reconstruction of the $\psi$C arm of the $M$. edulis tRNAArg gene relative to that given by Hoffmann et al. (1992), based on comparisons of the $M$. californianus and $M$. edulis tRNA$^{Arg}$ gene sequences, is shown. In each functional (or presumed functional) $M$. californianus tRNA gene, nucleotide substitutions that occur in the corresponding $M$. edulis tRNA gene (Hoffmann et al. 1992) are shown circled outside the sequence. Nucleotides in the $M$. californianus sequence that are circled with an arrow pointing out are not found in the $M$. edulis sequence. Arrows attached to circled nucleotides outside the $M$. californianus sequences indicate the locations of nucleotides found in the $M$. edulis tRNA sequences but not in the $M$. californianus sequences. Circled nucleotides in the $M$. californianus $\psi$tRNA$^{Ser}$(UCN) gene and the $M$. edulis $\psi$tRNA$^{Ser}$(UCN) gene are those that are substituted in the corresponding sequence of the other species.

Experiments were carried out to test for the presence in $M$. californianus mitochondria of RNAs corresponding to the tRNA$^{Ser}$(UCN), tRNA$^{Ser}$(AGN), tRNA$^{Trp}$, and tRNA$^{Pro}$ genes and the $\psi$tRNA$^{Ser}$(UCN) sequence. Oligonucleotides complementary to each of these genes and the $\psi$tRNA$^{Ser}$(UCN) sequence (Figure 2) were synthesized, $^{32}$P-labeled, and hybridized to a nitrocellulose blot.
TABLE 2
Comparisons of nucleotide sequences in eight tRNA genes and a ψtRNA gene in *M. californianus* F-type mtDNA with the nucleotide sequences of the corresponding tRNA genes of *M. edulis* F-type mtDNA

<table>
<thead>
<tr>
<th>M. californianus F-type tRNA gene</th>
<th>Total nucleotides</th>
<th>Number of nucleotide changes</th>
<th>Percentage identity (divergence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>64</td>
<td>5</td>
<td>92.2 (7.8)</td>
</tr>
<tr>
<td>Arg</td>
<td>65</td>
<td>9</td>
<td>86.2 (13.8)</td>
</tr>
<tr>
<td>His</td>
<td>65</td>
<td>3</td>
<td>91.0 (9.0)</td>
</tr>
<tr>
<td>Met(AUA)</td>
<td>65</td>
<td>6</td>
<td>90.8 (9.2)</td>
</tr>
<tr>
<td>Pro</td>
<td>64</td>
<td>6</td>
<td>89.1 (10.9)</td>
</tr>
<tr>
<td>Trp</td>
<td>67</td>
<td>3</td>
<td>94.0 (6.0)</td>
</tr>
<tr>
<td>Ser(AGN)</td>
<td>66</td>
<td>1</td>
<td>98.5 (1.5)</td>
</tr>
<tr>
<td>Ser(UCN)</td>
<td>63</td>
<td>3</td>
<td>95.2 (4.8)</td>
</tr>
<tr>
<td>Mean of eight tRNA genes</td>
<td>64.9</td>
<td>4.5</td>
<td>92.1 (7.9)</td>
</tr>
<tr>
<td>ψSer(UCN)²</td>
<td>69</td>
<td>23</td>
<td>63.8 (36.2)</td>
</tr>
</tbody>
</table>

Data for *M. edulis* are from Hoffmann et al. (1992).

²The comparison is for the entire *M. californianus* ψtRNA*Ser*(UCN) sequence (Figures 1 and 2) and the *M. edulis* sequence corresponding to it, which includes the entire shorter *M. edulis* ψtRNA*Ser*(UCN) sequence (Figure 2).

of whole mtRNA that had been electrophoresed in a 3% denaturing agarose gel. Examination of an autoradiograph of the product (Figure 3) revealed that the probes for each of the tRNA genes, but not for the ψtRNA*Ser*(UCN) sequence, had hybridized to an RNA of an approximate size (70 nt) expected for a tRNA. In contrast, the ψtRNA*Ser*(UCN) probe hybridized to RNA molecules averaging ~2500 nt, indicating that although the ψtRNA*Ser*(UCN) sequence is transcribed, it is rarely or never processed into a tRNA-sized molecule.

Two further Northern hybridization experiments were carried out to determine the absolute sizes of the various *M. californianus* mt-tRNAs. Both experiments involved hybridizing *3P*-labeled tRNA gene-specific oligonucleotides (Figure 2) to blots of mtRNA electrophoretically separated on 15% acrylamide gels (Okimoto and Wolstenholme 1990). In the first experiment, to provide a size marker, *E. coli* tRNA*Trp* (85 nt including the DNA-encoded 3' end-located CCA) was coelectrophoresed in a separate lane in the gel. Oligonucleotides specific for tRNA*Trp*, tRNA*Ser*(UCN), and tRNA*Met*(AUA) sequences and the *E. coli* tRNA*Trp* sequence were hybridized separately to blots of the gel. In the second experiment, the above-mentioned three oligonucleotides, together with oligonucleotides complementary to the mt-tRNA*Ser*(AGN) and mt-tRNA*Ser*(UCN) genes (Figure 2), were hybridized in various pairs. Examination of autoradiographs of the hybridization products (Figure 4, A and B) and of a plot (Figure 4C) of relative migration distances of the various bands observed in Figure 4A confirmed that for each mt-tRNA gene tested there occurs in *M. californianus* mitochondria a tRNA of the size predicted from the respective gene, but to which three nucleotides, presumably CCA, have been added. In the lanes probed for tRNA*Met*(AUA) and for tRNA*Ser*(UCN) there is also visible a second, lighter band indicating the presence of an RNA either 2 or 3 nt smaller than the main band, a size expected for tRNAs lacking either CA or CCA at their 3' ends.

ψtRNA*Ser*(UCN)-containing transcripts: The observation that an oligonucleotide complementary to the ψtRNA*Ser*(UCN) gene sequence hybridized only to RNAs

![Figure 3](image-url)

Figure 3.—Northern blot hybridization experiment to detect and determine the relative sizes of RNAs containing transcripts of genes for the DHU arm-replacement loop-containing tRNA*Ser*(UCN), the DHU arm-replacement loop-containing tRNA*Ser*(AGN), the four-armed tRNA*Ser*(UCN), tRNA*Trp*, and tRNA*Pro* predicted from gene sequences of *M. californianus* mtDNA. In the autoradiograph shown, each lane contains 3 μg of *M. californianus* mtRNA separated by electrophoresis through a 3% denaturing (2.2 M formaldehyde) agarose gel. The locations of the tRNA*Ser*(UCN), tRNA*Ser*(AGN), tRNA*Trp*, tRNA*Pro*, and ψtRNA*Ser*(UCN) gene-specific probes used are shown in Figure 2. Approximate band sizes (in kilobases to the right) were determined by comparison to a methylene blue-stained 0.24- to 7.46-kb ladder (Gibco-BRL; 5 μg/ lane) produced by coelectrophoresis.
of ~2500 nt indicates that the ψtRNA\textsuperscript{Ser}(UCN) transcript remains linked to a transcript that includes the downstream COI gene and/or the upstream ND3 gene. To resolve the relationship of the ψtRNA\textsuperscript{Ser}(UCN) transcript and adjacent gene transcripts, Northern blot hybridizations were carried out using M. californianus whole mtRNA and probes consisting of an oligonucleotide complementary to the 3' end of the ND3 gene (F3, Figures 2 and 5A), an oligonucleotide complementary to the ψtRNA\textsuperscript{Ser}(UCN) gene (F5, Figures 2 and 5A), and an oligonucleotide complementary to the 5' end region of the COI gene (F4, Figures 2 and 5A). The results are shown in Figure 5B. The ND3 probe hybridized to a single band of RNA of ~350 nt, the size (351 nt) expected for an RNA containing only a transcript of this gene. However, the ψtRNA\textsuperscript{Ser}(UCN) probe and the COI probe both hybridized to RNA of ~2400 nt. Therefore, these data indicate that transcripts of the ψtRNA\textsuperscript{Ser}(UCN) sequence and the COI gene, and probably the downstream ATPase6 gene (assuming identical gene order in this region of M. californianus and M. edulis mtDNAs), are contained in the same RNA molecule.

5' RACE analysis was used to locate the 5' terminus of the M. californianus COI gene-containing transcript. cDNAs were generated from whole mtRNA using random hexamers and reverse transcriptase. A poly(A) sequence was added to the 3' end of the cDNAs using terminal deoxynucleotidyl transferase, and cDNA complementary strands were synthesized from a 35-nt primer (TB 17-1) containing a 19-nt sequence that includes BglII, BamHI, and PstI sites followed by 16 T's (Pott-Kingdon \textit{et al.} 1994). Double-stranded cDNAs of the 5' end region of the COI-gene-containing transcript were selectively amplified in a polymerase chain reaction (PCR) using as primer oligonucleotide F4 (Figures 2 and 5B) that is complementary to a sequence within the 5' end region of the COI transcript and a 19-nt sequence (TB 17-2) comprising the BamHI-BglII-PstI sequence of primer TB 17-1. The PCR product was inserted into M13mp19 DNA, and five clones of the COI transcript 5' end region were sequenced. All five sequences obtained were identical and comparison of this sequence with the mtDNA sequence indicated that the 5' end nucleotide of the COI gene-containing transcript is seven nucleotides downstream from the TAA termination codon of the ND3 gene (Figure 6).

These data indicate that the entire ψtRNA\textsuperscript{Ser}(UCN) transcript [which is separated from the 5' terminus of the COI gene by 126 nt (Figures 1 and 5A)] remains

---

**Figure 4.**—(A and B) Northern blot hybridization experiments to precisely determine the sizes of tRNAs predicted from tRNA\textsuperscript{Trp}, tRNA\textsuperscript{Ser}(UCN), tRNA\textsuperscript{Ser}(AGN), tRNA\textsuperscript{Met(AUA)}, and tRNA\textsuperscript{Pro} gene sequences of M. californianus mtDNA. In the autoradiographs shown, each lane contains 10 μg of M. californianus mtRNA (M.c) or (A only), as a standard, 200 ng of E. coli (E.c) tRNA\textsuperscript{Trp} (85 nt including the 3' CCA) that have been electrophoresed through a 15% polyacrylamide gel containing 8 M urea. In A, the arrowheads that bracket the top of the gels identify the common 17-cm migration position. At the top of each lane containing M. californianus (M.c) mtRNA is shown the \textsuperscript{32}P-labeled tRNA probe or mixture of probes used (Figure 2). The expected number of nucleotides in the tRNA (or tRNAs) predicted from each gene sequence, plus CCA, are shown above each lane in A, and below each lane in B. The location of E. coli tRNA\textsuperscript{Trp} was also determined by hybridization of a \textsuperscript{32}P-labeled oligonucleotide complementary to that tRNA. (C) A plot of the expected size (log nucleotides) of each of the three mt-tRNAs (based on the size of the respective gene and the assumption that CCA is added post-transcriptionally at the 3' end of each tRNA) and E. coli tRNA\textsuperscript{Trp} (including the 3' terminal CCA) vs. the measured migration distance of the tRNAs in A.
Figure 5.—Northern blot hybridization experiment to detect and determine the relative sizes of the transcripts containing the COI gene, the ψtRNA\textsuperscript{Ser}(UCN) gene, and the ND3 gene of M. californianus mtDNA. (A) Gene map to show the locations and directions of oligonucleotide probes used (arrows labeled F3 and F4, Figure 1B; and F5, Figure 2). The number of nucleotide pairs in the ND3 and ψtRNA\textsuperscript{Ser}(UCN) genes, and in the intergenic spaces, and the estimated number of nucleotide pairs (1635) in the COI gene are shown. (B) Autoradiographs of Northern blots. Each lane contains 3 μg M. californianus mtRNA separated by electrophoresis through a 1% denaturing (2.2 M formaldehyde) agarose gel. The COI, ND3, and ψtRNA\textsuperscript{Ser}(UCN) gene-specific oligonucleotides used to probe each lane are shown. Sizes of bands (arrow a, 0.35 kb; arrows b, 2.5 kb) were determined by comparisons to a methylene blue-stained, 0.24- to 7.46-kb ladder (5 μg/lane; Gibco-BRL; represented in kilobases to the right).

Figure 6.—Autoradiograph showing the nucleotide sequence of a clone of the amplification product of a 3' polyadenylated (using terminal deoxynucleotidyl transferase) cDNA of the 5' end-proximal region of the M. californianus mitochondrial transcript that contains the ψtRNA\textsuperscript{Ser}(UCN) and COI genes. As the sequence shown is the transcript equivalent, the terminal reiterated nucleotide is T rather than A. The upper small arrowhead indicates the 5' terminal nucleotide of the transcript. To the right is shown the sequence downstream from the poly-T sequence (equivalent to the sequence between the two small arrowheads). The written sequence above the large arrowhead is that between the termination codon of the ND3 gene (ter ND3) and the 5' end of the ψtRNA\textsuperscript{Ser}(UCN)-COI gene transcript.

gene in mtDNAs of a number of metazoan species including mouse, D. yakuba, C. elegans, and A. suum, and it has been suggested that these sequences may be important for precise transcript cleavage (Bibb et al. 1981; Clary and Wolstenholme 1985; Okimoto et al. 1992). Hoffmann et al. (1992) noted secondary structure potential of both the sequence between the ND3 and ψtRNA\textsuperscript{Ser}(UCN) genes and the sequence between the ψtRNA\textsuperscript{Ser}(UCN) and COI genes in M. edulis mtDNA, but they concluded that the secondary structure of the first of these sequences was nonsignificant.

CONCLUDING REMARKS

In metazoan mtDNAs, although most adjacent genes are separated by few or no nucleotides, there is a single region of apparently noncoding nucleotides that varies in size between 125 ntp and ~8000 ntp. As this region in both mammalian and Drosophila mtDNAs has been linked to the COI transcript. The finding that the 5' end of the COI-gene-containing transcript is 81 nt upstream from the 5' end of the ψtRNA\textsuperscript{Ser}(UCN) sequence makes it unlikely that there has been selection for retention of the tRNA\textsuperscript{Ser}(UCN) sequence so that its secondary structure can act as a transcript processing site, a role suggested for interprotein gene tRNA sequences in mammalian multicistronic mtDNA transcripts (Ojala et al. 1981). Interestingly, however, in this regard the 5' end region of the M. californianus COI gene transcript can be folded into a stable stem and loop structure that includes the transcript cleavage site. This stable secondary structure potential is conserved in the corresponding sequence of M. edulis (F-type) mtDNA (Figure 7). Similar (but smaller) potential single stem and loop structures have been identified at the 3' ends of protein genes that are immediately followed by another protein...
Figure 7.—Potential stem and loop configuration of the nucleotide sequence of the 5′ end region of the COI transcript and presumed 3′ end region of the ND3 transcript of F-type M. californianus mtDNA. The 5′ end of the COI transcript determined by 5′ RACE analysis (Figure 6) is indicated by an arrow. Transcription is from left to right. The ND3 translation termination codon is boxed. The location of the tRNA^{Ser}(UCN) sequence within the COI transcript is shown. Circled nucleotides are substitutions in the corresponding transcripts of F-type M. edulis, predicted from the mtDNA. A nucleotide present in the M. edulis sequence but not in the M. californianus sequence is indicated by an arrow attached to a circled U. The predicted free energy changes shown were calculated using the free energy increments given by Freier et al. (1986).

The presently provided evidence that a sequence that appears to have acted as a tRNA^{Ser}(UCN) gene at some time in the past is now nonfunctional (at least in its original capacity) is the first direct demonstration of a pseudogene in metazoan mtDNA.

The single functional tRNA^{Ser}(UCN) gene of the two Mytilus species appears to be a 63-nt sequence located between the COIII and tRNA^{Met}(AUA) genes, which encodes a highly conserved DHU-replacement loop-containing tRNA. The interrelationship of the functional tRNA^{Ser}(UCN) gene and the ωtRNA^{Ser}(UCN) gene remains unclear. Sequence comparisons do not suggest that one is a duplicate of the other. As there is considerable divergence between the M. californianus and M. edulis mt-tRNA^{Ser}(UCN) gene sequences it remains possible that these sequences are the remnants of a gene for a tRNA other than one that recognizes UCN codons. However, this is not supported by sequence comparisons to other M. edulis mt-tRNA genes. It has been suggested that the tRNA^{Ser}(UCN) genes of the gastropods Euhadra herklotsi and Cepaea nemoralis (but not Albineria coerulea), the polyplophoran mollusk Katherina tunicata, and the annelid Lumbricus terrestris encode DHU arm-replacement loop-containing structures (Boore and Brown 1994, 1995; Hatzoglou et al. 1995; Terrett et al. 1996; Yamazaki et al. 1997). However, the K. tunicata and L. terrestris predicted tRNAs can be folded to give the usual U, A, or G between the aminoacyl and DHU arms, and a DHU arm with a stem of 3 ntp and a loop of 3 nt, but with three (rather than the usual one) nucleotides between the DHU and anticodon arms. Comparisons of the DHU arm-replacement loop-containing tRNA^{Ser}(UCN) genes of the two Mytilus species and the tRNA^{Ser}(UCN) gene of K. tunicata reveal conservation of 4/5 ntp in the T\(\omega\)C stems, 3/5 ntp in the anticodon stems, and 7/7 nt in the anticodon loops.

The mt-tRNA^{Ser}(UCN) genes of arthropods, echinoderms, and chordates so far examined encode a four-armed tRNA and the primary sequences of these tRNAs of the latter two phyla are close to those of standard tRNAs. However, as in mtDNAs of mollusks and annelids, genes for tRNAs expected to recognize UCN codons that have a DHU arm-replacement loop have also been found in nematode mtDNAs (Okimoto and Wolstenholme 1990; Okimoto et al. 1992). Considerations of the various proposed relationships of invertebrate phyla (see, for example, Wilson et al. 1978; Nielsen 1995) do not lend support to the view that the observed distribution of the genes for the two structural kinds of mt-tRNA^{Ser}(UCN) could have resulted from a single loss of a DHU arm-forming sequence from a mt-tRNA^{Ser}(UCN) gene during invertebrate evolution. Because it seems improbable that, once lost, a DHU arm-forming sequence of standard tRNAs would be regenerated (discussed in Wolstenholme et al. 1987), then the most likely explanation for the presence of a DHU arm-replacement loop-containing mt-tRNA^{Ser}(UCN) in nem-
atodes, some (but not all) mollusks, and an annelid, is that the gene for this kind of tRNA has been generated by multiple independent sequence losses. This would represent an unusual case of convergent structural evolution in metazoan mt-genomes.

The mtDNA sequence interpreted by Hoffmann et al. (1992) as an M. edulis tRNA\textsuperscript{Met}(AUA) gene is highly conserved in M. californianus mtDNA. Our finding that an RNA molecule of the sequence and size expected for a mature tRNA encoded by the M. californianus tRNA\textsuperscript{Met}(AUA) gene occurs in M. californianus mitochondria provides strong evidence that this gene is functional.

Hoffmann et al. (1992) stated that AUA specifies methionine in the M. edulis mt-genetic code, but they did not explain how they reached this conclusion. In fact, comparisons of ATA codons in the partial sequences of M. edulis Cyt b, COI, and COII genes (more highly conserved genes in metazoan mtDNAs), with similarly located codons in C. elegans, D. yakuba, and mouse mtDNAs, do not provide compelling evidence that AUA is more likely to encode methionine than isoleucine or leucine. However, interchange of ATA and ATG codons in the partial sequences of M. californianus F-type and M-type COI and ND5 genes (Beagley et al. 1997) does strongly support the coding specificity of AUA as methionine. Of 15 ATA codons that occur in these M-type sequences, 6 have ATA and 7 have ATG in the corresponding locations of the F-type sequences; of 8 ATG codons that occur in the M-type sequences, 3 have ATG and 4 have ATA in the corresponding locations of the F-type sequences. Therefore, assuming that the tRNA\textsuperscript{Met} (AUG) gene is functional, which seems likely, as suggested by Hoffmann et al. (1992), Mytilus mtDNAs are unique among known metazoan mtDNA sequences in that they contain two methionine-specifying tRNAs. Based on sequence similarities in the aminoacetyl and DHU arms, Hoffmann et al. (1992) suggested that the tRNA\textsuperscript{Met}(AUA) gene may have arisen as a duplication of the tRNA\textsuperscript{Met}(AUG) gene.

It is uncertain why Mytilus alone among Metazoa examined to date should utilize two tRNAs to recognize methionine codons. In M. edulis, the tRNA\textsuperscript{Met}(AUG) would be expected to recognize the AUG initiation codons of 10 of the 12 mt-protein genes and internal AUG codons. In regard to its initiation function it is noteworthy that the predicted tRNA would have three consecutive G-C base pairs at the base of the anticodon stem. This is a characteristic feature of the E. coli initiator tRNA\textsuperscript{Met} and it has been shown that these three G-C base pairs are essential to initiation function (Rajbhandary and Chow 1995). Also, three, two, or one of these G-C nucleotide pairs occur in the single mt-tRNA\textsuperscript{Met} of various other Metazoa (Anderson et al. 1982; Clary et al. 1982; Wolstenholme et al. 1994; Beagley et al. 1995). All but one of the ATA codons in the M. edulis mtDNA molecule are internal. Only in the ND5 gene does it appear likely that the initiation codon is ATA. We have been unable to detect any convincing similarity between the Mytilus tRNA\textsuperscript{Met}(AUA) gene and the E. coli elongator tRNA\textsuperscript{Met} that recognizes only internal AUG codons.

We thank Jane L. Macfarlane for excellent technical assistance and for discussions, and Robert Schackmann for oligonucleotides (partially subsidized by National Institutes of Health (NIH) grant CA-42014). This work was supported by NIH grant GM-18375.