A Novel Mitochondrial Genome Organization for the Blue Mussel, *Mytilus edulis*

Richard J. Hoffmann,*[,1] Jeffrey L. Boore† and Wesley M. Brown†

**Department of Zoology and Genetics, Iowa State University, Ames, Iowa 50011-3223, and †Department of Biology, University of Michigan, Ann Arbor, Michigan 48109-1048*

Manuscript received December 11, 1991
Accepted for publication February 7, 1992

ABSTRACT

The sequence of 13.9 kilobases (kb) of the 17.1-kb mitochondrial genome of *Mytilus edulis* has been determined, and the arrangement of all genes has been deduced. Mytilus mitochondrial DNA (mtDNA) contains 37 genes, all of which are transcribed from the same DNA strand. The gene content of Mytilus is typically metazoan in that it includes genes for large and small ribosomal RNAs, for a complete set of transfer RNAs and for 12 proteins. The protein genes encode the cytochrome *b* apoenzyme, cytochrome *c* oxidase (CO) subunits I–III, NADH dehydrogenase (ND) subunits 1–6 and 4L, and ATP synthetase (ATPase) subunit 6. No gene for ATPase subunit 8 could be found. The reading frames for the ND1, COI, and COIII genes contain long extensions relative to those genes in other metazoan mtDNAs. There are 23 tRNA genes, one more than previously found in any metazoan mtDNA. The additional tRNA appears to specify methionine, making Mytilus mtDNA unique in having two $tRNA^{Met}$ genes. Five lengthy unassigned intergenic sequences are present, four of which vary in length from 79 to 119 nucleotides and the largest of which is 1.2 kb. The base compositions of these are unremarkable and do not differ significantly from that of the remainder of the mtDNA. The arrangement of genes in Mytilus mtDNA is remarkably unlike that found in any other known metazoan mtDNA.

THE mitochondrial genome of metazoans is a closed circular DNA; only among hydrazoans (*Hydra* spp.) is this DNA linear (WARRIOR and GALL 1985). Although metazoan mitochondrial DNA (mtDNA) ranges in size from *ca.* 14–40 kb (SEDEROFF 1984; MORITZ, DOWLING, and BROWN 1987; SNYDER *et al.* 1987), its gene content is highly conserved. The same set of genes is usually present: 12 or 13 for proteins, one each for the small and large subunit ribosomal RNAs (s-rRNA and l-rRNA), and 22 for transfer RNAs (tRNAs; a sufficient number to translate the simplified mitochondrial genetic code). In addition, at least one extensive noncoding sequence is present which, in vertebrates (CLAYTON 1982, 1984) and in insects (CLARY and WOLSTENHOLME 1985), is known to contain elements controlling the initiation of replication and transcription. Size variation in mtDNA is usually due to differences in this noncoding sequence (BROWN 1985; HARRISON 1989), but is occasionally due to duplications of other portions of the genome (MORITZ and BROWN 1986, 1987; ZEVERING *et al.* 1991).

Although the nucleotide sequence of mtDNA changes at a rate as much as ten times higher than that of nuclear DNA (BROWN, GEORGE and WILSON 1979), the order in which the genes are arranged appears to change at a much slower rate (BROWN 1985). For example, among chordate mtDNAs there appears to be one basic arrangement (ANDERSON *et al.* 1981, 1982; BIBB *et al.* 1981; GADALETA *et al.* 1989; ROE *et al.* 1985; JOHANSEN, GUDDAL and JOHANSEN 1990) with only minor variations (PÄÄBO *et al.* 1991; DESJARDINS and MORAIS 1990, 1991; L. SZURA and W. BROWN, unpublished data); this within-phylum stability also appears to hold true among echinoderm (JACOBS *et al.* 1988; CANTATORE *et al.* 1989; DE GIORGI *et al.* 1991; HIMENO *et al.* 1987; SMITH *et al.* 1989, 1990; M. SMITH, personal communication) and arthropod mtDNAs (CLARY and WOLSTENHOLME 1985; BATUECAS *et al.* 1988; HSUCHEN, KOTIN and DUBIN 1984; DUBIN, HSUCHEN and TILLOTSON 1986; MCCRACKEN, UHLENBUSCH and GELLISSEN 1987; UHLENBUSCH, MCCRACKEN and GELLISSEN 1987; D. STANTON, L. SZURA and W. BROWN, unpublished data), even though the chordate, echinoderm and arthropod arrangements differ substantially from each other.

One exception to the conservation of gene arrangement within a phylum occurs in nematodes. *Ascaris suum* and *Caenorhabditis elegans* have gene arrangements that are identical except for the location of an A + T-rich noncoding region (WOLSTENHOLME *et al.* 1987; OKIMOTO and WOLSTENHOLME 1990; OKIMOTO, MACFARLANE and WOLSTENHOLME 1990), but that are radically different from that of a third ne-

[1] To whom correspondence should be addressed.

matode, *Meloidogyne javanica* (OKIMOTO *et al.* 1991). Whether this difference is due to a change in the rearrangement rate or to a very ancient split in the lineages leading to these taxa is unknown, although such a split in nematodes has been postulated (POINAR 1983). Neither of the nematode gene arrangements shows an affinity with that of any other metazoan mtDNA. At present, nematodes are the only phylum known in which representatives of different subgroups fail to share a basic mitochondrial gene arrangement, and they are further distinguished from other metazoans by their lack of a mitochondrial gene for ATPase8 (OKIMOTO *et al.* 1991).

The slow but perceptible rate of rearrangement and the very large number of arrangements possible suggest that mitochondrial gene order may provide useful information about the phylogeny of distantly related metazoan groups. Although present data provide encouragement for this approach, and methods for analyzing the evolutionary relationships among gene rearrangements are being developed (SANKOFF, CEDERGREN and ABEL 1990), only a few groups have been investigated and data from many more are needed for even a preliminary assessment. To this end, we have determined the mitochondrial gene arrangement for a bivalve mollusk, the blue mussel *Mytilus edulis*, the first such reported for any mollusk. Mytilus mtDNA is especially interesting in that it is commonly inherited biparentally (HOEH, BLAKLEY and BROWN 1991), providing a possible opportunity for recombination. Its gene arrangement is unlike that reported for any other metazoan mtDNA. Like nematodes, Mytilus apparently lacks a mitochondrial ATPase8 gene. Further, Mytilus mtDNA contains a unique tRNA gene which appears to specify methionine and, thus, has two tRNA genes for this amino acid.

MATERIALS AND METHODS

A λ EMBL3 vector containing an entire Mytilus mtDNA was a gift of D. O. F. SKIBINSKI. The mtDNA had been inserted into λ EMBL3 DNA using a unique *Bam*HI site, as described by SKIBINSKI and EDWARDS (1987). Although the mtDNA was isolated from *Mytilus galloprovincialis*-like mussels, it appears to be *M. edulis* mtDNA, perhaps acquired by hybridization between the two species (see below). Fragments produced by digesting DNA from the λ EMBL3 clone with combinations of *Bam*HI, *Xba*I, *Eco*RI, *Xho*I and *Nhe*I endonucleases were subcloned into pBluescript II KS(−) plasmids (Stratagene), producing five nonoverlapping subclones that jointly comprise the entire mtDNA, as well as additional overlapping subclones. Further subclones of the largest pBluescript II subclone were made using unique *Nhe*I, *Nco*I, *Bst*XI and *Pst*I restriction sites found within it. Detailed cleavage maps of each of the five original nonoverlapping subclones were constructed using combination digests of restriction endonucleases.

Double stranded DNA sequencing reactions employed modified T7 DNA polymerase (Sequenase 2.0, United States Biochemical Corp.), used according to supplier's instructions. [^{35}S]dATP (Amersham) was used to label the products, which were separated on 5–6% polyacrylamide gels containing 8 M urea and cast with wedge spacers. Initially, sequences from each end of each subcloned fragment were obtained using primers complementary to the T7 and T3 priming sites in pBluescript II; sequences internal to those were obtained using synthetic 17 nucleotide (nt) primers designed on the basis of the sequences obtained. Some automated sequencing, primarily to confirm sequences manually determined from only one strand, was performed by the Iowa State University Nucleic Acid Facility on an Applied Biosystems model 373A DNA sequencing system, following manufacturer's protocols. The subcloning and sequencing strategies are illustrated in Figure 1.

Computer analysis was performed with the University of Wisconsin Genetics Computer Group package of programs (DEVEREUX, HAEBERLI and SMITHIES 1984). Tentative identifications of protein coding genes were made using FastA searches of the animal mtDNA nucleotide sequences in GenBank and confirmed using BestFit comparisons of inferred amino acid sequences. Because both its nucleotide and amino acid sequences are known to be poorly conserved, ND6 gene identification was further confirmed by hydropathy profile comparisons (KYTE and DOOLITTLE 1982) with the ND6 proteins of Drosophila and the sea urchin Paracentrotus. Final amino acid alignments were made using Gap with program defaults (gap weight = 3.0, gap length weight = 0.1). The s-rRNA gene was identified by its sequence similarity with known s-rRNA genes, and ends were assigned by adjacency to inferred tRNA genes. The l-rRNA gene has no directly adjacent tRNA gene to define its 3′ end and, therefore, its approximate 3′ end was designated as the average position of the end of its sequence similarity with several known metazoan mitochondrial l-rRNA genes. Sequences encoding tRNA genes were identified generically by their ability to fold into typical metazoan mitochondrial tRNA structures, and specifically by their anticodon sequences.

RESULTS AND DISCUSSION

The total length of this Mytilus mtDNA is 17.1 kb, as determined by the sequence data combined with the restriction fragment size estimates for the few unsequenced portions. The cleavage map generated for this cloned mtDNA appears to be identical to that for the *M. edulis* mtDNA that was designated A in the report by HOEH, BLAKLEY and BROWN (1991). We did not map *Bcl*I sites in this study, but all other cleavage sites reported in that study appear identical to those we found. Comparison of 588 nt from the 5′ and 3′ ends of the cloned ND4 gene with ND4 sequence data obtained by S. MCCAFFERTY (personal communication) from two mtDNAs representing ostensibly pure populations of *M. edulis* and *M. galloprovincialis* revealed differences at 32 positions; the cloned DNA sequence was identical with *M. edulis* at 20, with *M. galloprovincialis* at five, and differed from both at seven. Although the original λ clone was produced using DNA from *M. galloprovincialis*-like mussels (SKIBINSKI and EDWARDS 1987), the ND4 sequence and the cleavage map similarities suggest that the cloned mtDNA is derived from *M. edulis*, and

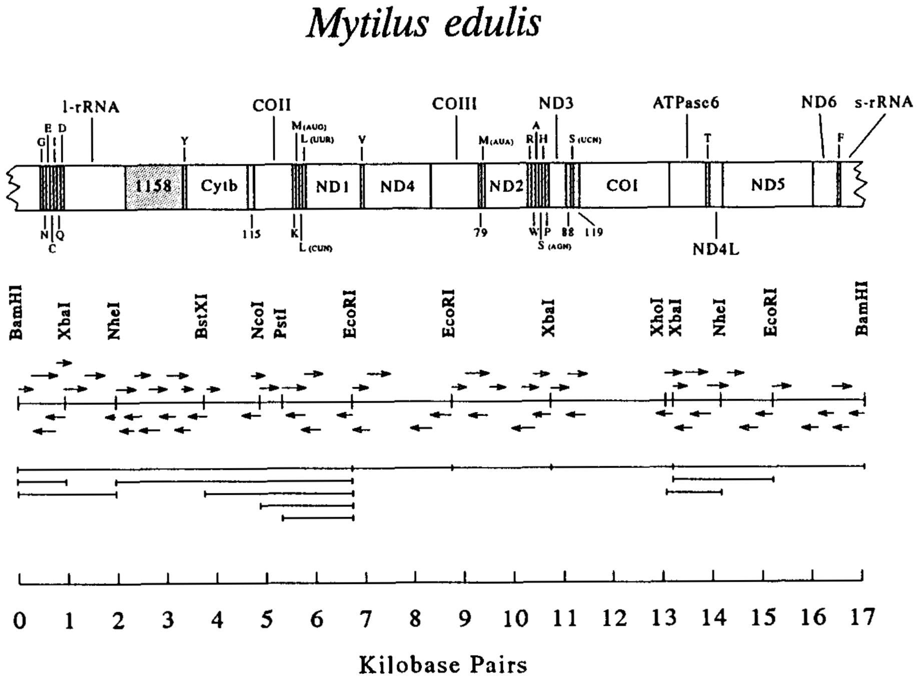


FIGURE 1.—Map of the genes contained in the *M. edulis* mitochondrial genome, together with subcloning and sequencing strategies. The cleavage map shows only those sites employed in subcloning. The original genome was cloned into λ EMBL3 using a unique *Bam*HI site, which appears at both ends of the linearized map illustrated here. Ranges encompassed by the pBluescript II subclones appear immediately above the scale line, with the five nonoverlapping subclones shown as a single line marked by the cleavage sites used to make them. Arrows show the locations and directions of individual sequencing runs. Locations of tRNA genes are marked by their single-letter amino acid codes. Alternative tRNA genes for methionine, leucine and serine are noted by subscripts designating the codons they recognize. Numbers indicate the quantity of nucleotides contained within five lengthy unassigned intergenic regions. The direction of transcription of all genes is left to right.

it is so identified in this study. This observation is not especially surprising, since hybridization between the two species is frequent (SKIBINSKI, AHMED and BEARDMORE 1978) and could easily result in interspecific exchange of mtDNA. A caution in this assignment to *M. edulis* is that on-going analysis of l-rRNA and ATPase6 sequence has so far failed to distinguish *M. edulis* from *M. galloprovinicialis* (R. KOEHN, personal communication).

Although 3.2 kb of the mtDNA remain unsequenced, the information from the 13.9-kb sequence obtained is sufficient to identify unambiguously all genes, to locate all gene boundaries and to sequence fully on both DNA strands 21 of 23 tRNA genes, all noncoding regions, plus unusual protein gene termini. The few nucleotides that were not determined by sequencing both strands of the two remaining tRNA genes are at their 3′ ends; thus, even if there are errors at these positions, the correct identification of the amino acid specificity of the two tRNAs is uncompromised. Aside from the five noncoding regions discussed in a following section, Mytilus mitochondrial genes either abut directly or are separated by short (1–15 nt) sequences, as in other metazoan mtDNAs.

Gene content: *M. edulis* mtDNA (Figure 1, APPENDIX I) contains 12 open reading frames (ORFs) that correspond to genes for cytochrome *b* (Cyt *b*), cytochrome *c* oxidase subunits I–III (COI–COIII), NADH dehydrogenase subunits 1–6 and 4L (ND1–6 and ND4L), and ATPase6. The ATPase8 gene appears to be missing (see below), as is also the case in nematodes (WOLSTENHOLME *et al.* 1987; OKIMOTO and WOLSTENHOLME 1990; OKIMOTO, MACFARLANE and WOLSTENHOLME 1990; OKIMOTO *et al.* 1991). Both rRNA genes are also present. There are 23 regions that can be folded into structures that are characteristic for tRNA genes, one more than is present in other metazoan mtDNAs. The additional tRNA gene employs an anticodon (TAT) for methionine that is unique among animal mtDNAs, giving Mytilus two mitochondrial $tRNA^{Met}$ genes rather than the single one found in other metazoans. (This unique tRNA gene is herein designated $tRNA^{Met}_{AUA}$, see below.) All of the genes are transcribed from the same DNA strand.

Unassigned DNA: A 1.2-kb region with no readily assignable function lies between the l-rRNA and the $tRNA^{Tyr}$ genes. This region contains several poly(A) tracts near the $tRNA^{Tyr}$ gene and one G + C-rich tract (27 of 35 nt; see APPENDIX I). These sequence features and the size of the region suggest it as a candidate for a role in initiating replication and/or transcription, but experimental data are needed before a functional assignment can be made. If the region does contain the origin of replication, it is much less A + T-rich (61%) than the control region in Drosophila, which is >90% A + T (CLARY and WOLSTENHOLME 1985).

This 1.2-kb region plus the adjacent 3′ end of the l-rRNA gene contain six ORFs that begin with the metazoan mitochondrial initiation codons, ATG, ATA or ATT, end with a termination codon and are >40 codons long (Segment #1, APPENDIX I). One ORF, of 111 codons, begins 13 nt beyond the 3′ end of the l-rRNA gene, on the same mtDNA strand as the known protein coding genes. The remaining five are on the opposite DNA strand. They range in size from 47 to 170 codons and overlap the 111 codon

TABLE 1

Stabilities (kcal/mol) of predicted RNA secondary structures (ZUKER and STIEGER 1981) for five unassigned intergenic sequences and a 5′ extension to the ND1 reading frame from *M. edulis* mitochondrial DNA

	Energy	
Region	True sequence	Randomized sequences
1.2-kb unassigned sequence	−277.3*	−253.5 ± 4.5
Cyt b to COII	−19.1*	−16.4 ± 1.3
ND3 to $tRNA^{Ser}_{UCN}$	−9.8	−11.0 ± 1.2
$tRNA^{Ser}_{UCN}$ to COI	−20.6*	−16.6 ± 1.9
COIII to $tRNA^{Met}_{AUA}$	−48.8*	−39.1 ± 2.1
ND1 5′ extension	−30.8	−32.7 ± 2.0

Stabilities of the true sequences are compared to the means ± 95% confidence intervals of 10 random sequences of nucleotides with the same nucleotide composition to aid in assessing significance. The 5′ extension to the COI and the 3′ extension to the COIII reading frames are included with their flanking unassigned sequence.

* Significantly more stable than random, in that they fall outside of the 95% confidence intervals for 10 randomized sequences with the same nucleotide composition as the native sequence.

ORF, each other, and/or the 3′ end of the l-rRNA gene extensively. None of the ORFs or their inferred protein products show any similarity to sequences in the databases, and none appears to correspond to the ATPase8 gene (see below). Because of this, and because of the extensive overlaps with each other and the l-rRNA gene, we doubt that any of these ORFs represent functional genes. However, without direct experimentation no strong conclusion is possible.

Optimal folding of the 1.2-kb region by the method of ZUKER and STIEGER (1981) yields a stable secondary structure (Table 1). Comparison of its free energy value with the mean energy value for 10 randomized sequences with the same nucleotide composition and size suggest that the true sequence is significantly more stable than a random collection of nucleotides having the same nucleotide composition (Table 1). The predicted secondary structure of the 1.2-kb region contains several long potential stem and loop structures. There is evidence that one such structure mediates initiation of lagging strand synthesis during vertebrate mtDNA replication (WONG and CLAYTON 1985), and it has been speculated that others might serve as recognition signals for replication, transcription, and RNA processing (BROWN 1985; FORAN, HIXSON and BROWN 1988; JACOBS *et al.* 1988).

There are four shorter noncoding regions in addition to the 1.2-kb region just discussed. One, of 115 nt, is between Cyt *b* and COII; another, of 79 nt, lies between the genes for COIII and $tRNA^{Met}_{AUA}$; and two, of 88 and 119 nt, flank the $tRNA^{Ser}_{UCN}$ gene at its 5′ and 3′ ends, respectively (Figure 1, APPENDIX I). Like the 1.2-kb region, each contains sequence elements capable of secondary structure formation (Table 1) and, thus, could be potentially important for processing or transcription. The sequences between the Cyt *b* and COII genes and between the $tRNA^{Ser}_{UCN}$ and COI genes fold into structures that are significantly more stable than random. The predicted structure at the 5′ end of the $tRNA^{Ser}_{UCN}$ gene, however, is not significantly different in stability than random. Finally, the 79-nt sequence between the end of the COIII reading frame and the beginning of the $tRNA^{Met}_{AUA}$ gene could be included in a stable structure including a 3′ extension of the COIII gene (discussed below). Assessing biological significance of these structures, if any, requires further investigation. In the aggregate, these four smaller unassigned regions represent more noncoding nucleotides outside of the control region than has been found previously. Finally, because of uncertainties about the exact lengths of some of the protein genes (discussed below), there may be additional noncoding regions in Mytilus mtDNA.

A missing ATPase8 gene: Release 17 of the Swiss-Prot database includes 13 entries for ATPase8 genes of metazoan mtDNAs: eight from vertebrates, two from echinoids and three from insects. There is very little sequence conservation among these, even at the amino acid level, which, when combined with the small size of the gene (159–204 nt), makes its identification exceedingly difficult. A few features aid in its recognition, however. In 12 of the 13 entries, ATPase8 is inferred to begin with Met-Pro-Gln (in some, the initial codon used for Met is a nonstandard one), and Leu follows Gln in all but the three insect proteins, which have Met instead. All 13 are identical in having Trp as the sixth amino acid following Pro. Additionally, all but the five mammals have the sequence Trp-*X*-Trp at or within one amino acid of the C terminus. Examination of hydropathy profiles (KYTE and DOOLITTLE 1982) indicates that all of the proteins are hydrophobic over the amino terminal half of their length and hydrophilic over the C terminal half. Finally, for all metazoan mtDNAs in which an ATPase8 gene has been identified, it immediately precedes (and sometimes overlaps in another reading frame) the gene for ATPase6. In human cells, both gene products are translated from a single mRNA (OJALA, MONTOYA and ATTARDI 1981).

Translation of all portions of the sequence from Mytilus mtDNA in all possible reading frames and examination of all inferred polypeptides, including those with nonstandard initiation codons, yielded one with an initial Met-Pro-Gln-Leu sequence. This reading frame is 34 codons long, lacks any Trp codons, has a hydropathy profile that is different from those of all other ATPase8 proteins, and lies entirely within the ND2 gene but on the opposite DNA strand. The initial sequence Met-Pro-Pro-Gln-$(X)_5$-Trp occurred once, in an ORF of only 21 codons that also lacks the characteristic hydropathy profile of ATPase8. Allowing for the possibility of a nonstandard initiation codon by searching for Pro-Gln-$(X)_5$-Trp failed to reveal additional candidates. Two open reading frames with the sequence Trp-*X*-Trp within one codon of a termination codon were found, but neither had the other characteristic features of ATPase8. None of the six ORFs in the 1.2-kb unassigned region discussed above has the characteristics of ATPase8. Finally, there is no reading frame with any of these characteristics in the vicinity of the ATPase6 gene. In summary, we have been unable to find evidence of an ATPase8 gene in the sequence obtained from Mytilus mtDNA and, barring the unlikely possibility that it is entirely within an unsequenced portion of one of the other protein genes, we conclude that, like nematodes, Mytilus lacks a mitochondrial ATPase8 gene.

Although not resolvable with present data, at least two interesting and testable evolutionary questions are worth noting in this regard. One concerns the fate of the missing ATPase8 gene. Has it been transferred to the nucleus, or has its function been subsumed by a different nuclear-encoded protein that is transported into the mitochondrion, or is that function somehow dispensable for some organisms, such as Mytilus and nematodes? The other concerns the shared absence of this gene among the bivalve and nematode species examined. Does this reflect two independent losses from two separate lineages or one loss from a taxon that was ancestral to both bivalves and nematodes? Both questions can be addressed by further investigation.

Comparisons of gene order: As Figure 2 shows, Mytilus mtDNA is organized in a radically different fashion from that of *Drosophila yakuba*, the only protostome for which a complete mtDNA sequence is available (CLARY and WOLSTENHOLME 1985). Few genes share the same boundaries in both; if tRNA genes are ignored, only the s- and l-rRNA genes are found in the same relative order and transcriptional polarity. However, in Drosophila (and in other arthropods and chordates as well) the two rRNA genes are separated by one tRNA gene, for valine, whereas in Mytilus they are separated by seven tRNA genes, none of which specifies valine. With regard to mitochondrial gene order, there is substantially less similarity between Mytilus and other representative protostomes than between those of protostomes and deuterostomes, in marked contradiction to expectations based on an assumption of monophyly among protostomes.

The only other metazoans that have a spiral pattern of cleavage during development and for which a complete mitochondrial gene order is published are two nematodes, *Ascaris suum* and *Caenorhabditis elegans* (WOLSTENHOLME *et al.* 1987; OKIMOTO and WOLSTENHOLME 1990; OKIMOTO, MACFARLANE and WOLSTENHOLME 1990). Comparing Mytilus with these, only the ND5 and ND6 genes have the same relative orientation, but there are three intervening tRNA genes in nematodes and none in Mytilus. Although these two genes are also adjacent in chordates and echinoderms, their relative polarities are opposite, whereas they are the same in Mytilus.

Partial gene orders are known for a few other taxa with spiral cleavage. Although all of its tRNA genes have not been mapped, a third nematode mtDNA, from *Meloidogyne javanica* (OKIMOTO *et al.* 1991), shares little of its order with the other two nematodes or with Mytilus. The ND2 and COIII genes are adjacent in both Mytilus and Meloidogyne, but their relative polarities are different. Although its full mtDNA organization is unknown, in the flatworm *Fasciola hepatica* the genes for ND3 and COI are separated by those for $tRNA^{Ser}_{AGN}$ and $tRNA^{Tyr}$ (GAREY and WOLSTENHOLME 1989); these two protein genes are in the same order and orientation in Mytilus, but with a $tRNA^{Ser}_{UCN}$ gene and two noncoding regions between them. Finally, no two mitochondrial protein or rRNA genes share the same boundaries in Mytilus and the sea anemone *Metridium senile* (D. WOLSTENHOLME, personal communication).

If tRNA genes are included in the comparisons, a few additional common organizational features are revealed among the mtDNAs of Mytilus and other metazoans. The COII and $tRNA^{Lys}$ genes are directly adjacent and have the same relative polarity in Mytilus, in Drosophila and several other arthropods, in echinoderms, and in chordates. Also, the 5′ end of the ND1 gene is preceded by the gene for $tRNA^{Leu}_{UUR}$ in Mytilus, in chordates and in echinoderms; regarding this, we note that the gene for $tRNA^{Leu}_{CUN}$ is upstream of the ND1 gene in several arthropod mtDNAs, which suggests the possibility that the $tRNA^{Leu}_{CUN}$ and $tRNA^{Leu}_{UUR}$ genes might be homologous in some lineages and simply have "exchanged" anticodons reciprocally. The $tRNA^{Thr}$ gene is upstream of the ND4L gene in both Mytilus and Drosophila, albeit with reversed polarity in Drosophila. The 5′ end of the s-rRNA gene is bounded by the $tRNA^{Phe}$ gene in Mytilus, in chordates, and in echinoderms. In Mytilus and echinoderms, the $tRNA^{Trp}$ gene is followed by that for $tRNA^{Ala}$, but with their relative polarities switched; in chordates, these two tRNA genes are also adjacent with opposite polarities, but in the reverse order. The $tRNA^{Ser}_{AGN}$ and $tRNA^{His}$ genes are adjacent and have the same polarity in Mytilus, chordates and echinoderms, but the order is 5′-$tRNA^{Ser}$-$tRNA^{His}$-3′ in Mytilus and 5′-$tRNA^{His}$-$tRNA^{Ser}$-3′ in the others. The $tRNA^{Gln}$ gene follows the $tRNA^{Ile}$ gene in Mytilus, Drosophila and chordates, but the genes are on opposite strands in the

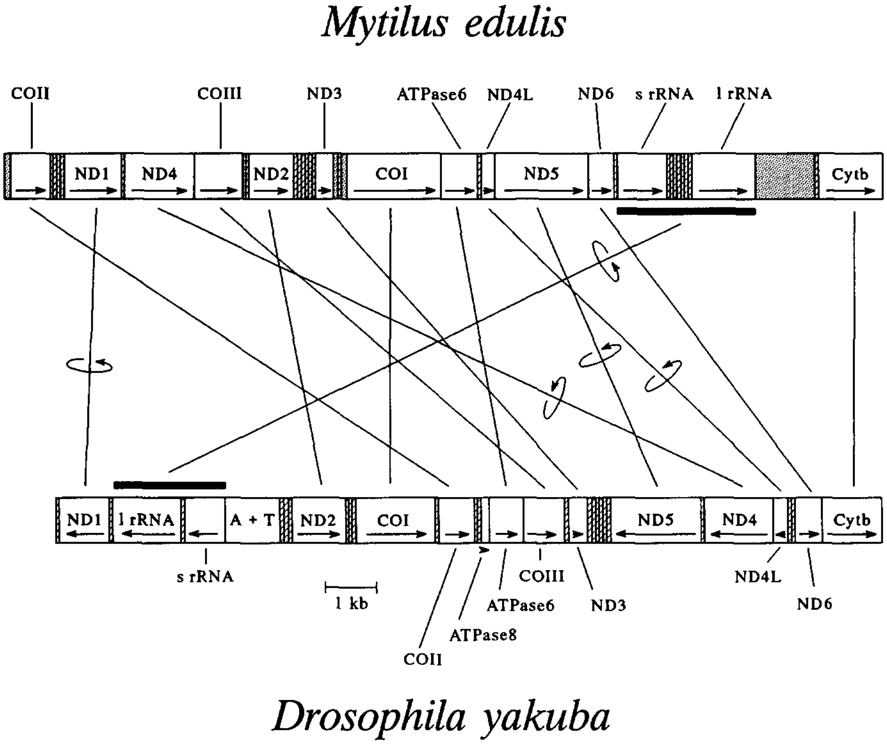


FIGURE 2.—Comparison of mitochondrial gene organization between *M. edulis* and *D. yakuba*. The circular genomes were linearized and aligned at the Cyt *b* gene. Lines connect identical large genes, and ellipses with arrows highlight reversals in polarity. Positions of tRNA genes, indicated by cross hatching, have been ignored in this comparison. Arrows under gene labels show direction of transcription. The bars spanning the s-rRNA and the l-rRNA genes show the only two large genes that share the same orientation and transcriptional polarity between the genomes. Stippled areas in the Mytilus genome show regions of unassigned function. The *D. yakuba* gene order is from CLARY and WOLSTENHOLME (1985).

latter two. Finally, the genes for $tRNA^{Met}$ and ND2 are directly adjacent in Mytilus, chordates, and Drosophila (but not locusts); however, the $tRNA^{Met}$ gene in Mytilus is the one with the anticodon TAT, which has not been found in other animal mtDNAs.

Initiation and termination of translation: Although uncertainty about the exact positions of the 5′ ends of the COI and ND1 genes makes a reasonable assignment of their initiation codons difficult, it is less so for the remaining protein genes. Initiation of translation at ATG codons seems clearly indicated for the COII, COIII, Cyt *b*, ATPase6, ND2-ND4, ND4L and ND6 genes. However, ND5 appears to initiate at ATA, because (1) this codon begins 11 nt immediately 3′ of the termination codon for ND4L, (2) this results in a predicted protein of the same length as in Drosophila, and (3) the first in-frame ATG is 98 codons downstream. Initiation of COI and ND1, discussed in detail below, could also begin at ATG, although other in-frame codons are also available.

The mitochondrial termination codon TAA is found at the 3′ ends of the reading frames for COI, COIII, Cyt *b*, ND3, ND4, ND4L and ND6, and the alternative termination codon TAG occurs at the ends of the genes for COII, ATPase6 and ND2. Two genes, ND1 and ND5, may end with an abbreviated termination codon, TA, which is presumably completed by polyadenylation of their mRNAs. For ND1 this must be so, because the nucleotide following the 3′-TA is T, which is the first nucleotide of the $tRNA^{Val}$ gene. The case for ND5 is less clear; the nucleotide following the 3′-TA is A, which would complete the ND5 stop codon, but this same A must also participate in the presumptive ATG initiation codon for ND6. While it is possible that ND5 and ND6 do not overlap and that their transcripts are processed by cleavage between these adjacent As, several alternatives that allow ND5 and ND6 to overlap and share this A also exist, such as the initiation of translation from alternate start codons on a common, digenic mRNA, alternative (semidestructive) RNA processing, or initiation of ND6 translation from a codon further downstream. The present data do not allow us to choose among these alternatives. (Depending upon the status of a 3′ extension of the COIII gene, discussed below, it may terminate with a single T.)

Protein genes: Those portions of the protein genes that were sequenced were translated and their amino acid sequences compared with those of other animals. Most are similar in sequence and size to their homologs in other species. However, some Mytilus proteins appear to be longer or shorter than their Drosophila counterparts, and uncertainties about initiation sites of some genes confound inferences about their lengths.

Based upon alignment of conserved amino acids near its ends, the COI gene of Mytilus is potentially longer at both its 5′ and 3′ ends than it is in Drosophila (APPENDIX I). If the amino terminus of COI is the same in both species, translation in Mytilus must start at a TGG codon, which normally specifies tryptophan. The nearest in-frame methionine codon (ATA) occurs nine codons upstream from this TGG; we note in this regard that the inferred start of COI in nematodes is also 9–10 amino acids upstream from that in Drosophila (OKIMOTO, MACFARLANE and WOLSTENHOLME 1990). However, the ORF for COI in Mytilus actually begins with an ATT codon that is 22 codons upstream of the TGG codon just mentioned. ATT, which ordinarily specifies isoleucine, is an inferred initiation codon in Drosophila mtDNA (CLARY and WOLSTENHOLME 1985) and also in mouse mtDNA, where any ATN codon apparently can initiate translation (BIBB *et al.* 1981). In Mytilus, three methionine codons (both ATA and ATG) are just downstream from this ATT, and each could serve as a start site. As a final complication, there is evidence that GTT is an initiation codon for one of the protein genes in nematodes (OKIMOTO, MACFARLANE and WOLSTENHOLME 1990), and in Mytilus there is an in-frame GTT three codons upstream of the TGG codon that corresponds to the position of the start codon for COI in Drosophila.

Given the above uncertainties, we are unwilling to assign this sequence to the Mytilus COI gene. However, we are also unable to suggest any alternative assignment for it. The sequence is adjacent to the 119-nt unassigned sequence that follows the $tRNA^{Ser}_{UCN}$ gene, and it may simply be a part of that sequence that, purely by chance, is open and in-frame with the following COI gene. It is possible to form a stable RNA secondary structure that includes both the 119-nt unassigned sequence and all of the 5′ extension of COI (Table 1). We note that it includes all but the last nucleotide before the first methionine codon (ATA) upstream of the start of the Drosophila protein in the COI reading frame, coincident with the length of COI in nematodes. This structure or another like it might serve as an RNA processing signal and thus result in the elimination of this region before translation.

The COI gene also appears to be longer at its 3′ end in Mytilus than in Drosophila (APPENDIX I). The reading frame remains open to a termination (TAA) codon that is located 17 codons beyond the 3′ end in Drosophila. There are 9 nt between this TAA and the initiation codon of the ATPase6 gene. It may be worth noting that the second codon following the asparagine at the end of the COI alignment between the Mytilus and Drosophila is TAT, part of which could serve as an abbreviated termination signal, but there is no obvious reason to favor such an inference over one in which COI is 16 amino acids longer at its C terminus in Mytilus than in Drosophila. Moreover, no candidate stem-loop processing signals are predicted to be formed in this region.

The apparent 3′ extension of COI may be related to the absence of an ATPase8 gene in Mytilus mtDNA. As noted, the start of ATPase6 follows the termination (TAA) codon of COI by 9 nt. In all other metazoans, the ATPase8 gene precedes and sometimes extensively overlaps the 3′ end of the ATPase6 gene (BIBB *et al.* 1981). If an ATPase8 gene was once present at this location, it is possible that part of it became attached to the COI gene after the ATPase8 gene became dispensable.

The Mytilus COIII gene may also be longer at its 3′ end than that of Drosophila. The first termination codon (TAA) in this gene is 48 codons beyond the C terminus of Drosophila (APPENDIX I). In Mytilus, a TGA (Trp) codon follows the last alignable residue (Gly) between Mytilus and Drosophila, the T of which could serve as an abbreviated termination signal. In this regard, two things are worth noting: One is that this region and the unassigned 79 nt following it in Mytilus have the potential to fold into stable stem-loop structures (Table 1) and thus may be eliminated during RNA processing. The other is that the COIII gene in Lasaea, another bivalve, lacks this 3′ extension (D. O'FOIGHIL and M. SMITH, personal communication), as do all of the other COIII proteins in the databases. That the 3′ ends of other COIII proteins are highly conserved in length makes it seem unlikely that the 3′ extension in Mytilus is actually translated.

In contrast to the ambiguities just described, the ND1 gene of Mytilus is almost certainly truncated at its 3′ end. Although overlap with the downstream $tRNA^{Val}$ gene cannot be ruled out absolutely, the dinucleotide TA, which could serve as an abbreviated stop codon for ND1, immediately precedes the first nucleotide of the $tRNA^{Val}$ gene. Moreover, even if overlap occurred, an in-frame termination (TAA) codon is only four codons downstream from the start of the $tRNA^{Val}$ gene. This 3′ truncation of ND1 results in an inferred protein that is 20 amino acids shorter than its Drosophila counterpart.

The converse may be true at the 5′ end of ND1, where a substantial addition, as compared with Drosophila, is possible. If the two ND1 proteins had identical lengths at their N-terminal ends, translation in Mytilus would have to initiate with GTG (APPENDIX I). GTG is occasionally used as an initiation codon in bacterial systems (STORMO, SCHNEIDER and GOLD 1982), and a similar use in animal mtDNA has been suggested (CLARY and WOLSTENHOLME 1985; JACOBS *et al.* 1988; GADALETA *et al.* 1989; JOHANSEN, GUDDAL and JOHANSEN 1990; DESJARDINS and MORAIS 1990, 1991). However, if GTG can serve as a start signal in Mytilus, the ND1 reading frame is open back to another GTG codon, which is the first codon downstream of the $tRNA^{Leu}_{UUR}$ gene. In this context, GADALETA *et al.* (1989) have observed that GTG appears to serve as an initiation codon only when it is directly preceded by the last nucleotide of the preceding gene. In Mytilus, this GTG is located at a position that is 54 codons upstream of the inferred 5′ end of the Drosophila ND1 gene.

However, RNA sequences in this 54-codon region have the potential to form a stable secondary structure (Table 1) that includes all of the nucleotides before the first GTG mentioned, and this could result in its complete or partial elimination from the final processed transcript. However, the predicted structure is not significantly more stable than random. There are one ATG, two ATTs, two GTTs and another GTG in this region, any of which could serve as start sites if a portion of the region were eliminated. Also, there are three TTGs among the 54 codons, which might function as start codons in Mytilus as well as in nematodes (OKIMOTO, MACFARLANE and WOLSTENHOLME 1990). Although we are reluctant to assign this region to the ND1 gene, we have been unable to assign an alternative function to this DNA sequence.

The remaining differences in lengths of gene products between Drosophila and Mytilus are minor. The most notable extension among these is at the 5′ end of the Cyt *b* gene, which is somewhat longer than other Cyt *b* genes. There are no alternative start codons in the Cyt *b* reading frame, however. In addition to truncations or extensions at gene termini, there are a few cases of internal sequence deletions/additions. The longest of these, equivalent to 21 codons, is near the middle of the ND6 gene.

In conclusion, we have found evidence of relative size changes in several Mytilus genes, among which those involving internal deletions or additions are less ambiguous than those suggesting truncations or additions to the ends of the genes. As a caveat, it must be remembered that all such assignments depend on the validity of inferences drawn by comparative sequence analysis and that any final conclusions about gene lengths must await independent confirmation by direct analysis of the gene products themselves.

Ribosomal RNA genes: The region between the genes for $tRNA^{Phe}$ and $tRNA^{Gly}$ shows extensive sequence similarity to other metazoan mitochondrial s-rRNA genes. The Mytilus s-rRNA gene is 945 nt, similar in size to mammalian mtDNAs and considerably larger than in Xenopus (819 nt; ROE *et al.* 1985), echinoderms (*ca.* 880 nt; JACOBS *et al.* 1988; CANTATORE *et al.* 1989; DE GIORGI *et al.* 1991), or Drosophila (*ca.* 790 nt; CLARY and WOLSTENHOLME 1985).

The boundaries of the l-rRNA gene are less certain, especially at the 3′ end where there is no directly adjacent tRNA or protein gene. The computer program Gap was used to compare the final 300 nt at the 3′ ends of six different l-rRNA genes (those of Drosophila, Xenopus, Paracentrotus, mouse, human, and cow) against the region in Mytilus mtDNA bounded by positions 2222 and 3800 of Segment #1 in APPENDIX I. This region includes the entire l-rRNA gene, plus approximately 300 nt of the downstream unassigned sequence. The nt positions where the individual alignments ended were averaged to estimate the approximate 3′ end of the gene. By this procedure, the average position of the end of similarity was found to lie at position 3465 of Segment #1 (APPENDIX I). From this, we estimate the size of the Mytilus l-rRNA gene to be approximately 1245 nt, which, if accurate, makes it the smallest on record (Drosophila's is 1326 nt). However, the largest (1.2 kb) unassigned region is adjacent to the l-rRNA gene in Mytilus, and it is possible that 100–400 nt of this region are part of the mt-l-rRNA gene. If so, however, sequence similarity between Mytilus and other species must be weak.

Transfer RNA genes: The 23 inferred Mytilus tRNA genes can be folded into the typical cloverleaf secondary structures shown in Figure 3. Structurally, these genes are similar to most other metazoan mt-tRNAs. Mismatches are present in a few stems, G-T pairs (= G-U in RNA) are common in stems, and all genes lack a 3′ terminal CCA. The aminoacyl-accepting stem is usually of seven nt pairs, although there is the potential for 8 to 10 pairs in a few genes (those for Arg, Leu(CUN), Leu(UUR), Lys, and Ser(AGN); extra nucleotides not shown in Figure 3), and a few also have aminoacyl stems with mismatched termini (those for Cys, Lys, Ser(UCN) and Val; Figure 3). The dihydrouridine (DHU) arm consists of a stem of 3 or 4 paired and a loop of three to 9 unpaired nt. The TψC arm has a stem of 3–5 paired and a loop of 3–7 unpaired nt. The variable arm is of either 4 or 5 nt. The anticodon stem is of 5 nt pairs, with a few genes having one of these pairs mismatched. The anticodon loop is of 7 nt, with the possibility for pairing across the loop in some cases. Anticodons are preceded by T, except in the gene for $tRNA^{Met}_{AUA}$, which has a C at this position. Anticodons are followed by either A or G with approximately equal likelihood.

There are two leucine tRNA genes ($tRNA^{Leu}_{CUN}$ and $tRNA^{Leu}_{UUR}$), as in other animal mtDNAs. These are directly adjacent in Mytilus, in a cluster of tRNA genes that follows the COII gene. There are also two serine tRNA genes, $tRNA^{Ser}_{AGN}$ and $tRNA^{Ser}_{UCN}$. As in other metazoan mtDNAs, $tRNA^{Ser}_{AGN}$ has an unpaired loop that replaces the DHU arm. In Mytilus, this loop is of 13 nt, the longest so far reported. There is potential for some nucleotide pairing across the loop and also between its 3′ end and the variable arm. Aside from the novel $tRNA^{Met}$ gene, the anticodons are the same as in Drosophila mtDNA, except that the anticodon for $tRNA^{Lys}$ is TTT in Mytilus, as it is in chordate and nematode mtDNAs.

A second tRNA gene specifying methionine? In all previously characterized metazoan mtDNAs there is a single $tRNA^{Met}$ gene that presumably can be charged with either *N*-formyl-methionine and used for translation initiation, or with methionine and used at internal codons. In mouse, its anticodon (CAT) is

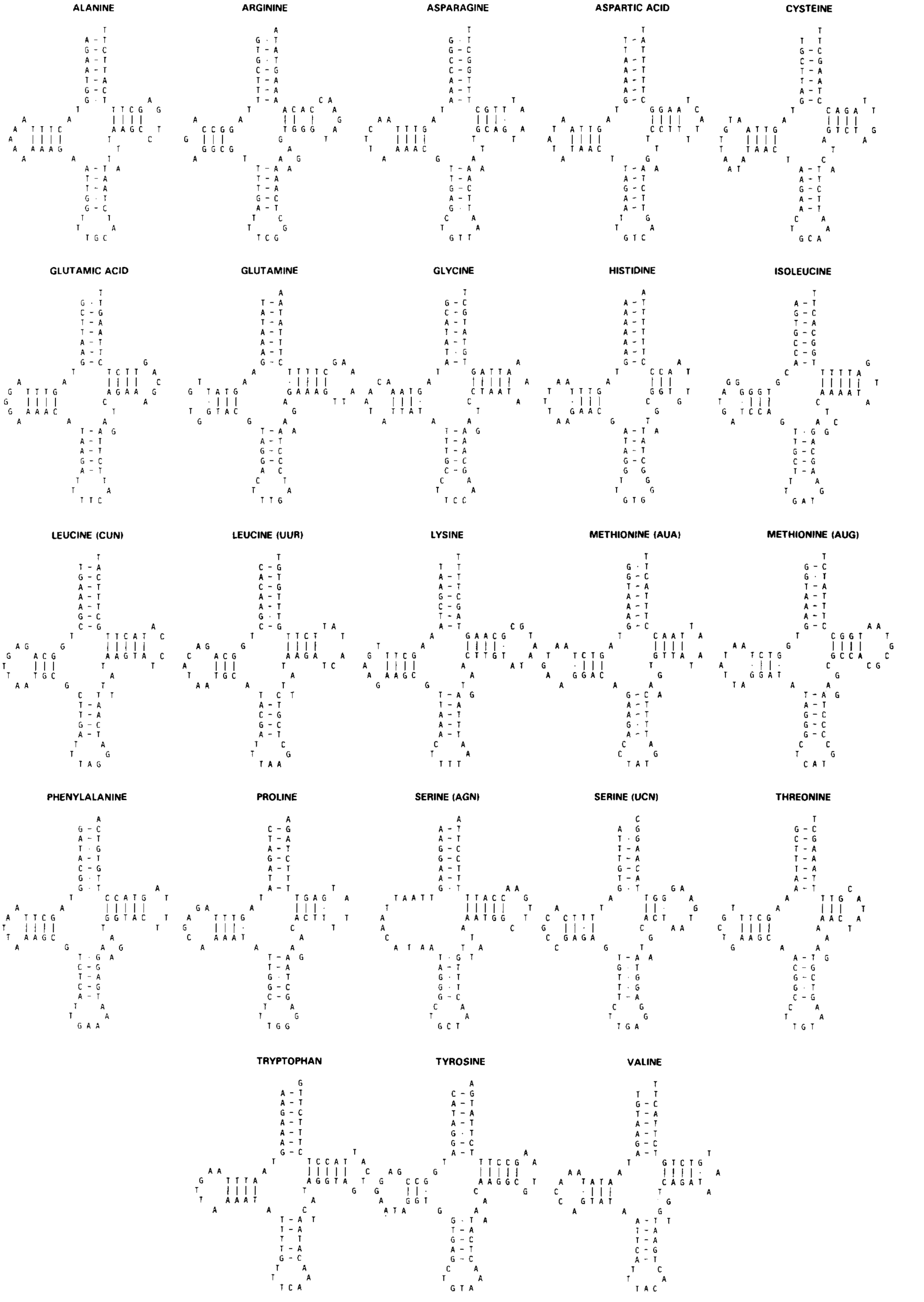


FIGURE 3.—Sequences of 23 tRNA genes from *M. edulis* folded into cloverleaf forms expected for their gene products. Amino acid identities are given above each sequence. Watson-Crick pairing is indicated by solid lines; G-T pairs are shown with dots.

inferred to recognize any initial ATN codon (BIBB *et al.* 1981). In Mytilus mtDNA, a $tRNA^{Met}$ gene with a CAT anticodon is present and located in a cluster of four tRNA genes, between $tRNA^{Lys}$ and $tRNA^{Leu}_{CUN}$ (Figure 1, APPENDIX I).

Unlike other metazoan mtDNAs, that of Mytilus appears to encode a second $tRNA^{Met}$ gene. Located immediately 5′ of the ND2 gene, it has the anticodon TAT (Figures 1 and 3), which presumably recognizes ATA codons most efficiently. ATA is a normal Met codon only in mitochondria (PEEBLES, GEGENHEIMER and ABELSON 1983); it specifies isoleucine in the cytoplasm. However, it is unlikely that ATA specifies isoleucine in Mytilus mtDNA, since both ATA and ATG codons are employed at conserved methionine positions of mitochondrial proteins in both Drosophila (CLARY and WOLSTENHOLME 1985) and Mytilus.

The sequence and predicted structure of this novel tRNA are unexceptional; a C precedes the anticodon, but this is also the case in various mt-tRNA genes of other metazoans. Although the conclusion that this gene is transcribed and that its product is functional must await experimental verification, there is no sequence-based reason to doubt that this is true. We therefore refer to this gene as $tRNA^{Met}_{AUA}$, to distinguish it from the second $tRNA^{Met}$ gene in Mytilus, which we call $tRNA^{Met}_{AUG}$. Sequence comparison suggests that the two Mytilus $tRNA^{Met}$ genes arose by duplication and subsequent divergence. At their 5′ ends the genes are identical at 15 of the first 16 positions and, if the genes are aligned by regions (excluding the TψC loop, which cannot be aligned), only 23 positions are different between them. Gene duplication has been thought to be involved in high rearrangement rates for mt-tRNA genes (*e.g.*, JACOBS *et al.* 1988; PÄÄBO *et al.* 1991).

A $tRNA^{Met}_{AUA}$ gene is not only unusual among animal mtDNAs, but among all organisms. The only other tRNA with this anticodon (UAU) listed in SPRINZL *et al.*'s (1991) compilation of all known tRNAs occurs in yeast, where it specifies isoleucine. It is possible that a few directly sequenced tRNAs have UAU anticodons, but in all cases examined the nucleotide in the anticodon wobble position is modified and of uncertain derivation. For example, it is either U or C in *Halobacterium volcanii* (GUPTA 1984). Other tRNAs with NAU anticodons include those from phage T4 (GUTHRIE and MCCLAIN 1979), *Escherichia coli* (HARADA and NISHIMURA 1974), spinach chloroplasts (FRANCIS and DUDOCK 1982; KASHDAN and DUDOCK 1982), and potato mitochondria (WEBER *et al.* 1991). However, the sequences of the tRNA genes in the latter two cases show that CAT is the anticodon and, thus, rule out U as the original nucleotide at the modified wobble position in those cases.

Even if $tRNA^{Met}_{AUA}$ is expressed in Mytilus, it is unclear how or whether its function would differ from that of $tRNA^{Met}_{AUG}$. Both ATA and ATG codons are present at internal positions in Mytilus protein genes, and presumably each would be most efficiently recognized by its cognate tRNA. Similarly, a differential function of the two tRNAs in initiation seems unlikely, since both ATG and ATA appear to be used as initiation codons in Mytilus.

Genetic code and codon usage: The mitochondrial genetic codes of Mytilus and Drosophila appear to be identical, with TGA specifying tryptophan, ATA methionine, and AGA and AGG serine. With the possible exception of the ND1 gene, for which an initiation codon cannot presently be assigned with certainty, it appears that ATT is not used as an initiation codon in Mytilus. Although in vertebrate mtDNAs, AGA and AGG are unused and may be termination codons (*e.g.*, see BIBB *et al.* 1981), both are used in Mytilus and Drosophila mtDNAs, and the positional similarity of this usage argues that AGA and AGG are serine codons in both cases.

A summary of usage for the 2824 codons sequenced (Table 2) indicates no pronounced codon bias in Mytilus. No unused codons occur in any codon family, and no third position bias is evident. C, G, A and T occur in the third position in 13.3, 23.6, 28.0 and 35.1% of the codons, respectively, and these amounts hardly differ from those (14.0, 23.8, 28.0 and 34.2%) in the 13.9-kb portion of Mytilus mtDNA sequenced. Excluding termination codons, those used 10 or fewer times are ACG (Thr), TCC (Ser), CGC (Arg) and CCC (Pro).

CONCLUSIONS

The size, sequence, cleavage map data and, as inferred from them, the structure and genetic organization of Mytilus mtDNA furnish yet another example of both the unity and diversity that appear to be characteristic of metazoan mtDNAs. With respect to unity: (1) The gene complement is identical to that of other metazoans, except for the presence of two $tRNA^{Met}$ genes (even the absence of ATPase8 has a precedent in nematodes). (2) The genome organization is very compact; no introns were found and, with the exception of a 1.2-kb noncoding region and four other unassigned regions of 79–119 nt, intergenic sequences are minimal or absent. (3) The tRNA genes exhibit the relaxed sequence and structural constraints that have become a hallmark of metazoan mt-tRNAs.

With respect to diversity, the genetic organization of Mytilus mtDNA furnishes another example of the nonconservation of mitochondrial gene order among metazoans. Based on the assumption that coelomate metazoans are monophyletic, we anticipated that the Mytilus gene order would differ, but still be a recog-

TABLE 2

Codon usage in sequenced portions of 12 protein genes from *M. edulis* mitochondrial DNA

Amino acid	Codon	*N*	Amino acid	Codon	*N*	Amino acid	Codon	*N*	Amino acid	Codon	*N*
Phe	TTT	168	Ser	TCT	58	Tyr	TAT	81	Cys	TGT	43
	TTC	52		TCC	9		TAC	40		TGC	25
Leu	TTA	135		TCA	18	Term	TAA	7	Trp	TGA	35
	TTG	85		TCG	20		TAG	3		TGG	41
Leu	CTT	52	Pro	CCT	43	His	CAT	29	Arg	CGT	25
	CTC	13		CCC	10		CAC	20		CGC	10
	CTA	62		CCA	21	Gln	CAA	33		CGA	20
	CTG	38		CCG	20		CAG	14		CGG	12
Ile	ATT	122	Thr	ACT	46	Asn	AAT	59	Ser	AGT	48
	ATC	37		ACC	12		AAC	30		AGC	21
Met	ATA	102		ACA	26	Lys	AAA	54		AGA	69
	ATG	65		ACG	8		AAG	34		AGG	58
Val	GTT	79	Ala	GCT	55	Asp	GAT	33	Gly	GGT	49
	GTC	27		GCC	18		GAC	22		GGC	29
	GTA	93		GCA	42	Glu	GAA	36		GGA	39
	GTG	96		GCG	19		GAG	51		GGG	103

Entries include extensions to the reading frames of COI, COIII and ND1.

nizable variant on the relatively conserved arrangements seen among other coelomates (*e.g.*, chordates, echinoderms and arthropods). Further, based on present phylogenetic hypotheses, it was also reasonable to expect that traces of a shared ancestry might be preserved most strongly between mollusk and arthropod mtDNAs. Neither of these expectations is met. The degree to which the arrangement in Mytilus is unique was not only unexpected, but, given the currently accepted placement of mollusks with other coelomate metazoans, is much more surprising than the highly divergent arrangements that have been found in pseudocoelomate (*e.g.*, nematode) and acoelomate (*e.g.*, flatworm and sea anemone) metazoans. Understanding the evolutionary derivation of the Mytilus gene arrangement, for example by the method of SANKOFF, CEDERGREN and ABEL (1990), will require information about other metazoan groups.

Previous mechanistic speculations about the reasons for conservation of mitochondrial gene organization have included a suggestion that arrangement of genes, particularly rRNA and tRNA genes, is important for regulation of gene expression (*e.g.*, OJALA, MONTOYA and ATTARDI 1981; MONTOYA, GAINES and ATTARDI 1983). The plethora of unique and radically different arrangements, among which that of Mytilus is but the latest example, argue strongly against such a conclusion and, instead, suggest that metazoan mitochondrial gene order is a relatively neutral characteristic with regard to selection.

Although the reasons for the scrambled gene order in Mytilus mtDNA are unknown, it may be germane that mtDNA heteroplasmy is widespread in this genus, apparently as a result of occasional biparental inheritance of mtDNA (FISHER and SKIBINSKI 1990; HOEH, BLAKLEY and BROWN 1991). Heteroplasmy may provide an opportunity for recombination among different mtDNAs that is usually absent. HOEH, BLAKLEY and BROWN (1991) have presented cleavage maps from two Mytilus mtDNAs that were found in heteroplasmy. The map for one of the mtDNAs is identical to that for the genome characterized in the present study, and it is, therefore, most probably also identical in gene order. However, the map for the other mtDNA differs extensively from this map, and that mtDNA is also longer by about 1.6 kb. Further, while the two mtDNAs share many cleavage sites mapped in the section of the genome containing ND3, COI, ATPase6, ND4L, ND5 and the two rRNA genes (Figure 1), they share essentially no sites in the remainder of the genome, raising the possibility that multiple gene arrangements might be found among individual mussels.

Little is known about the mtDNA of other bivalves. S. MCCAFFERTY (personal communication) has enzymatically amplified and sequenced a region of the mtDNA from four different species of Mytilus that spans the ND4 and COIII genes, and he found that these two genes have the same relative order and polarity in all four species. The ND4 and COIII genes are in a region that appears, by cleavage map comparisons (HOEH, BLAKLEY and BROWN 1991), to be greatly dissimilar among individuals, thus suggesting that the dissimilarity may not necessarily reflect a dissimilar gene arrangement. D. O'FOIGHIL and M. SMITH (personal communication) have data from another bivalve (genus *Lasaea*) indicating that, as in Mytilus, the ND2 gene is downstream from the COIII gene, but with one unidentified tRNA gene intervening. In Mytilus these genes are separated by a 79 nt noncoding region and the $tRNA^{Met}_{AUA}$ gene. Their data also suggest that the ND4 gene is not upstream of the COIII gene as it is in Mytilus. In the Japanese scallop, *Patinopecten yessoensis*, the COIII gene precedes and is proximal to the ATPase6 gene, but it is not yet known whether an ATPase8 gene is located between them (E. BOULDING and A. BECKENBACH, personal communication). COIII and ATPase6 thus have different relative positions in Patinopecten than in Mytilus. If confirmed, these results would indicate that bivalves are nonidentical in their mtDNA organization and that the great dissimilarity of Mytilus mitochondrial gene order from those of other metazoans may stem from an accelerated rate of intragenomic rearrangement as well as from phylogenetic distance.

These results emphasize the need for a much broader and more detailed survey of mtDNA gene order. Without data for other protostome mtDNAs, including data for other phyla as well as for other molluscan classes, it will be impossible to determine whether Mytilus represents an aberration or whether some protostome groups will simply turn out to be much less conservative in mtDNA gene order than deuterostome groups and arthropods appear to be. It is important in this regard that representatives of two other molluscan classes (a chiton, *Katharina tunicata*, and a gastropod, *Plicopurpura* sp.) have gene arrangements that are different from that of Mytilus (J. BOORE, T. COLLINS and W. BROWN, unpublished data), demonstrating that gene order is not a conserved feature among all molluscan classes, at least as they are presently constituted.

This work was supported by grants from the National Science Foundation (BSR-8904633 and BSR-8916466, to R.J.H.; BSR-8613902 and BSR-8517830, to W.M.B.), and the National Institutes of Health (GM30144, to W.M.B.). We thank LYNNE SZURA and CAROL MANTHEY for expert assistance and DAVID SKIBINSKI for providing cloned Mytilus mtDNA. The DNA sequences presented in this report have been deposited in GenBank (accession numbers M83756, M83757, M83758, M83759, M83760, M83761 and M83762, corresponding to Segments 1–7, respectively, of APPENDIX I).

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. L. H. DE BRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADEN and I. G. YOUNG, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

ANDERSON, S., M. H. L. DE BRUIJN, A. R. COULSON, I. C. EPERON, F. SANGER and I. G. YOUNG, 1982 Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J. Mol. Biol. **156:** 683–717.

BATUECAS, B., R. GARRESSE, M. CALLEJA, J. R. VALVERDE and R. MARCO, 1988 Genome organization of *Artemia* mitochondrial DNA. Nucleic Acids Res. **16:** 6515–6529.

BIBB, M. J., R. A. VAN ETTEN, C. T. WRIGHT, M. W. WALBERG and D. A. CLAYTON, 1981 Sequence and gene organization of mouse mitochondrial DNA. Cell **26:** 167–180.

BROWN, W. M., 1985 The mitochondrial genome of animals, pp. 95–130 in *Molecular Evolutionary Genetics*, edited by R. J. MACINTYRE. Plenum Press, New York.

BROWN, W. M., M. GEORGE, JR., and A. C. WILSON, 1979 Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971.

CANTATORE, P., M. ROBERTI, G. RAINALDI, M. N. GADALETA and C. SACCONE, 1989 The complete nucleotide sequence, gene order, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J. Biol. Chem. **264:** 10965–10975.

CLARY, D. O., and D. R. WOLSTENHOLME, 1985 The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. **22:** 252–271.

CLAYTON, D. A., 1982 Replication of mammalian mitochondrial DNA. Cell **28:** 693–705.

CLAYTON, D. A., 1984 Transcription of the mammalian mitochondrial genome. Annu. Rev. Biochem. **53:** 573–594.

DE GIORGI, C., C. LANAVE, M. D. MUSCI and C. SACCONE, 1991 Mitochondrial DNA in the sea urchin *Arbacia lixula*: evolutionary inferences from nucleotide sequence analysis. Mol. Biol. Evol. **8:** 515–529.

DESJARDINS, P., and R. MORAIS, 1990 Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. J. Mol. Biol. **212:** 599–634.

DESJARDINS, P., and R. MORAIS, 1991 Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. J. Mol. Evol. **32:** 153–161.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DUBIN, D. T., C.-C. HSUCHEN and L. E. TILLOTSON, 1986 Mosquito mitochondrial tRNAs for valine, glycine and glutamate: RNA and gene sequences and vicinal genome organization. Curr. Genet. **10:** 701–707.

FISHER, C., and D. O. F. SKIBINSKI, 1990 Sex-biased mitochondrial DNA heteroplasmy in the marine mussel *Mytilus*. Proc. R. Soc. Lond. B **242:** 149–156.

FORAN, D. R., J. E. HIXSON and W. M. BROWN, 1988 Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res. **16:** 5841–5861.

FRANCIS, M. A., and B. DUDOCK, 1982 Nucleotide sequence of a spinach chloroplast isoleucine tRNA. J. Biol. Chem. **257:** 11195–11198.

GADALETA, G., G. PEPE, G. DE CANDIA, C. QUAGLIARIELLO, E. SBISA and C. SACCONE, 1989 The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J. Mol. Evol. **28:** 497–516.

GAREY, J. R., and D. R. WOLSTENHOLME, 1989 Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of a $tRNA^{ser}AGN$ that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. J. Mol. Evol. **28:** 374–387.

GUPTA, R., 1984 *Halobacterium volcanii* tRNAs: identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNAs. J. Biol. Chem. **259:** 9461–9471.

GUTHRIE, C., and W. H. MCCLAIN, 1979 Rare transfer ribonucleic acid essential for phage growth. Nucleotide sequence comparison of normal and mutant T4 isoleucine-accepting transfer ribonucleic acid. Biochemistry **18:** 3786–3795.

HARADA, F., and S. NISHIMURA, 1974 Purification and characterization of AUA specific isoleucine transfer ribonucleic acid from *Escherichia coli* B. Biochemistry **13:** 300–307.

HARRISON, R. G., 1989 Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol. Evol. **4:** 6–11.

HIMENO, H., H. MASAKI, T. KAWAI, T. OHTA, I. KUMAGAI, K. MIURA and K. WATANABE, 1987 Unusual genetic codes and a novel gene structure for tRNA$^{ser}_{AGY}$ in starfish mitochondrial DNA. Gene **56:** 219–230.

HOEH, W. R., K. H. BLAKLEY and W. M. BROWN, 1991 Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. Science **251:** 1488–1490.

HSUCHEN, C.-C., R. M. KOTIN and D. T. DUBIN, 1984 Sequences of the coding and flanking regions of the large ribosomal subunit RNA gene of mosquito mitochondria. Nucleic Acids Res. **12:** 7771–7785.

JACOBS, H. T., D. J. ELLIOTT, V. B. MATH and A. FARQUARSON, 1988 Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. **202:** 185–217.

JOHANSEN, J., P. H. GUDDAL and T. JOHANSEN, 1990 Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. Nucleic Acids Res. **18:** 411–419.

KASHDAN, M. A., and B. DUDOCK, 1982 The gene for spinach chloroplast isoleucine tRNA has a methionine anticodon. J. Biol. Chem. **257:** 11191–11194.

KYTE, J., and R. F. DOOLITTLE, 1982 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 333–348.

MCCRACKEN, A., I. UHLENBUSCH and G. GELLISSEN, 1987 Structure of the cloned *Locusta migratoria* mitochondrial genome: restriction mapping and sequence of its ND-1 (URF-1) gene. Curr. Genet. **11:** 625–630.

MONTOYA, J., G. L. GAINES and G. ATTARDI, 1983 The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell **34:** 151–159.

MORITZ, C., and W. M. BROWN, 1986 Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science **233:** 1425–1427.

MORITZ, C., and W. M. BROWN, 1987 Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA **84:** 7183–7187.

MORITZ, C., T. E. DOWLING and W. M. BROWN, 1987 Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18:** 269–292.

OJALA, D., J. MONTOYA and G. ATTARDI, 1981 tRNA punctuation model of RNA processing in human mitochondria. Nature **290:** 470–474.

OKIMOTO, R., J. L. MACFARLANE and D. R. WOLSTENHOLME, 1990 Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans*. Nucleic Acids Res. **18:** 6113–6118.

OKIMOTO, R., and D. R. WOLSTENHOLME, 1990 A set of tRNAs that lack either the TψC arm or the dihydrouridine arm: towards a minimal tRNA adaptor. EMBO J. **9:** 3405–3411.

OKIMOTO, R., H. M. CHAMBERLIN, J. L. MACFARLANE and D. R. WOLSTENHOLME, 1991 Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host race identification. Nucleic Acids Res. **19:** 1619–1626.

PÄÄBO, S., W. K. THOMAS, K. M. WHITFIELD, Y. KUMAZAWA and A. C. WILSON, 1991 Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. **33:** 426–430.

PEEBLES, C. L., P. GEGENHEIMER and J. ABELSON, 1983 Precise excision of intervening sequences from precursor tRNAs by a membrane-associated yeast exonuclease. Cell **32:** 525–536.

POINAR, G. O., JR., 1983 *The Natural History of Nematodes*. Prentice-Hall, Englewood Cliffs, N.J.

ROE, B. A., D.-P. MA, R. K. WILSON and J. F.-H. WONG, 1985 The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. **260:** 9759–9774.

SANKOFF, D., R. CEDERGREN and Y. ABEL, 1990 Genomic divergence through gene rearrangement. Methods Enzymol. **183:** 426–438.

SEDEROFF, R. R., 1984 Structural variation in mitochondrial DNA. Adv. Genet. **22:** 1–108.

SKIBINSKI, D. O. F., M. AHMAD and J. A. BEARDMORE, 1978 Genetic evidence for naturally occurring hybrids between *Mytilus edulis* and *Mytilus galloprovincialis*. Evolution **32:** 354–364.

SKIBINSKI, D. O. F., and C. A. EDWARDS, 1987 Mitochondrial DNA variation in marine mussels (*Mytilus*), pp. 209–226 in *Proceedings of the World Symposium on Selection, Hybridization, and Genetic Engineering in Aquaculture*, Vol. I. Heeneman, Berlin.

SMITH, M. J., D. K. BANFIELD, K. DOTEVAL, S. GORSKI and D. J. KOWBEL, 1989 Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. Gene **76:** 181–185.

SMITH, M. J., D. K. BANFIELD, K. DOTEVAL, S. GORSKI and D. J. KOWBEL, 1990 Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Pisaster ochraceus*. J. Mol. Evol. **31:** 195–204.

SNYDER, M., A. R. FRASER, J. LAROCHE, K. E. GARTNER-KEPKAY and E. ZOUROS, 1987 Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. Proc. Natl. Acad. Sci. USA **84:** 7595–7599.

SPRINZL, M., N. DANK, S. NOCK and A. SCHÖN, 1991 Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. **19:** 2127–2171.

STORMO, G. D., T. D. SCHNEIDER and L. M. GOLD, 1982 Characterization of translation initiation sites in *E. coli*. Nucleic Acids Res. **10:** 2971–2996.

UHLENBUSCH, I., A. MCCRACKEN and G. GELLISSEN, 1987 The gene for the large (16S) ribosomal RNA from the *Locusta migratoria* mitochondrial genome. Curr. Genet. **11:** 621–638.

WARRIOR, R., and J. GALL, 1985 The mitochondrial DNA of *Hydra attenuata* and *Hydra littoralis* consists of two linear molecules. Arch. Sci. **38:** 439–445.

WEBER, F., A. DIETRICH, J.-H. WEIL and L. MARÉCHAL-DROUARD, 1990 A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine codon. Nucleic Acids Res. **18:** 5027–5030.

WOLSTENHOLME, D. R., J. L. MACFARLANE, R. OKIMOTO, D. O. CLARY and J. A. WAHLEITHNER, 1987 Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. USA **84:** 1324–1328.

WONG, T. W., and D. A. CLAYTON, 1985 *In vitro* replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. Cell **42:** 951–958.

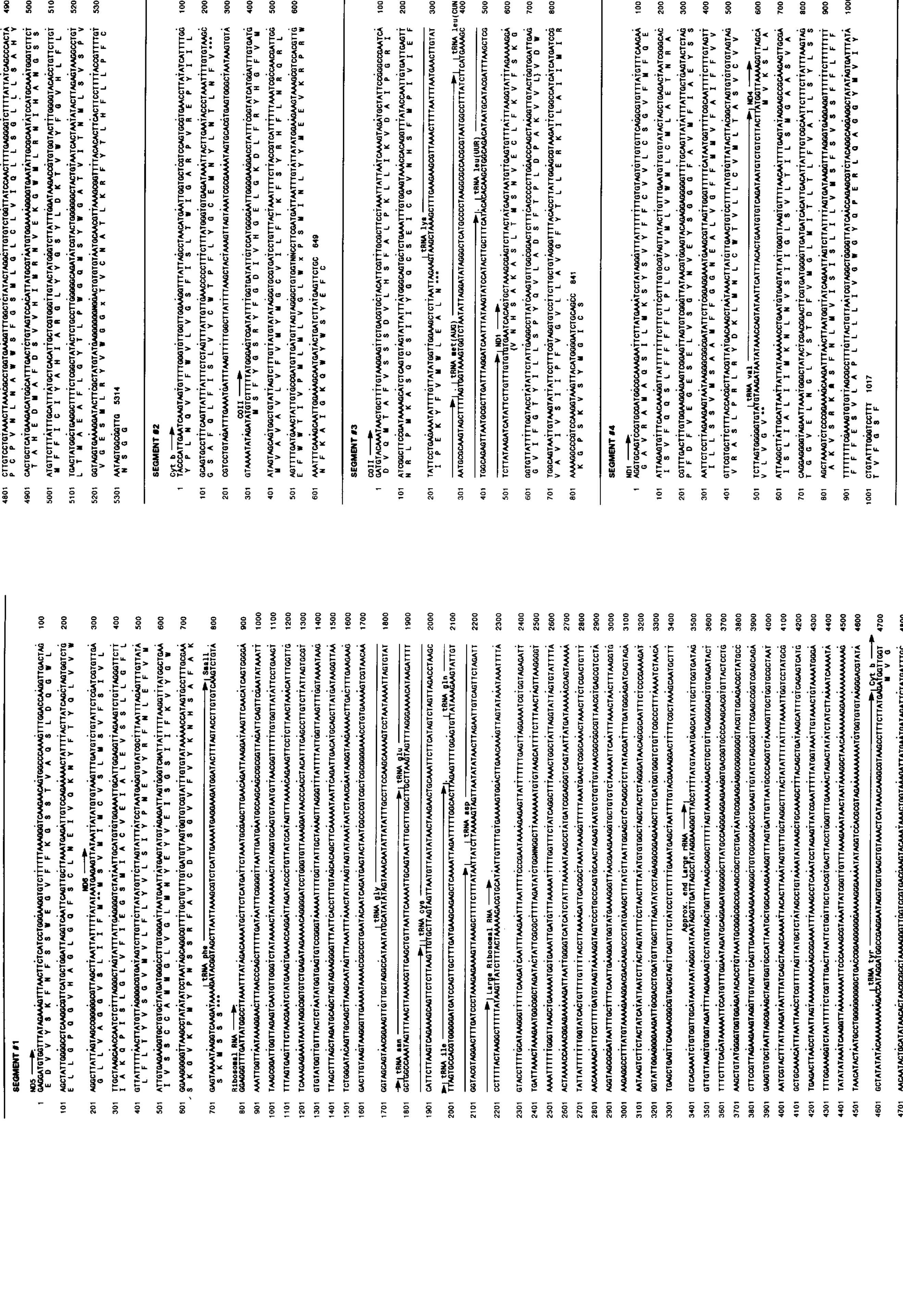
ZEVERING, C. E., C. MORITZ, A. HEIDEMAN and R. A. STURM, 1991 Parallel origins of duplications and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei*; Gekkonidae). J. Mol. Evol. **33:** 431–441.

ZUKER, M., and P. STIEGER, 1981 Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9:** 133–148.

Communicating editor: A. G. CLARK

APPENDIX I

The nucleotide sequences of seven contiguous segments of the *Mytilus edulis* mitochondrial genome are



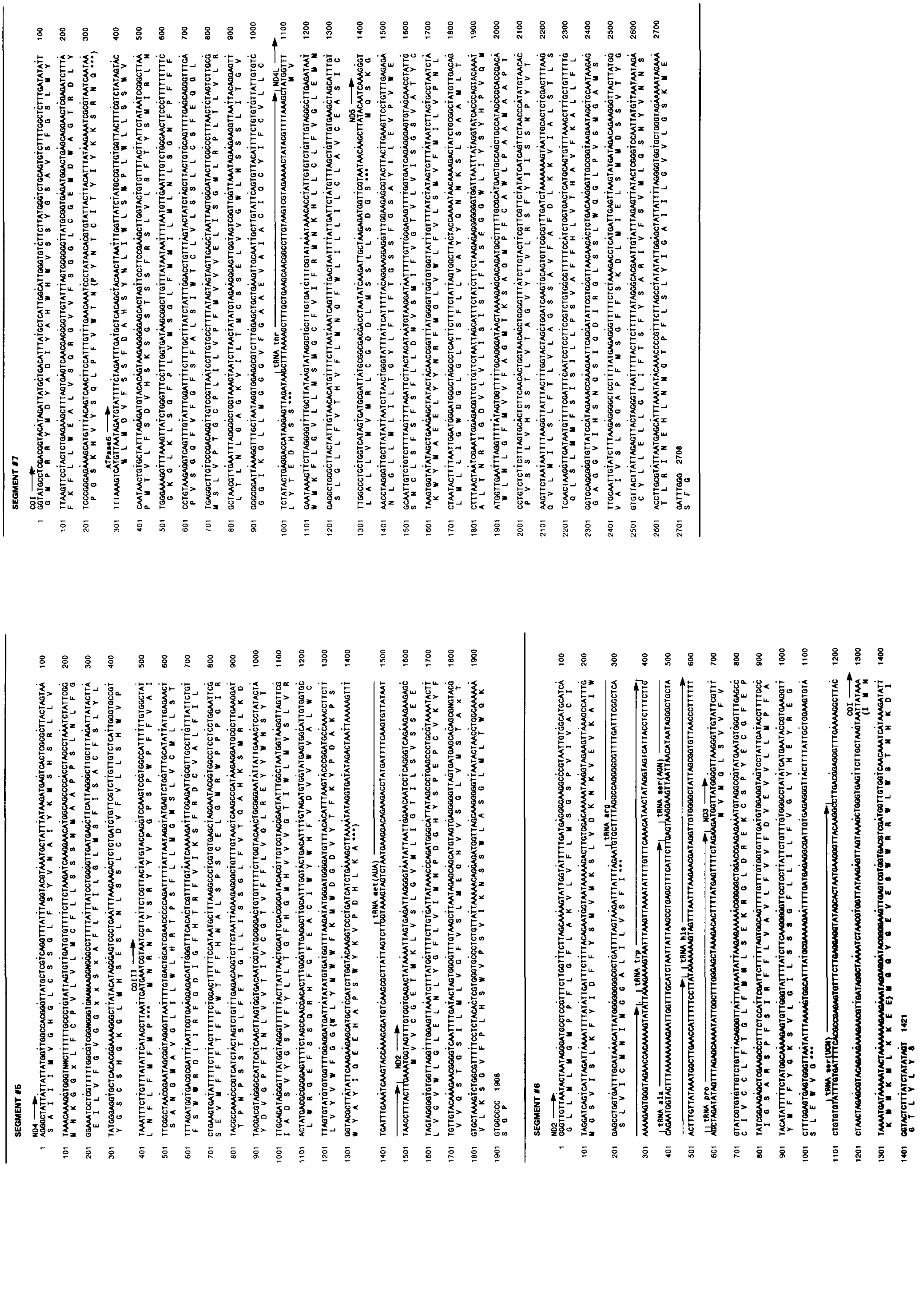


FIGURE 4.—Nucleotide sequences of seven contiguous segments of the *M. edulis* mitochondrial genome.

shown in Figure 4. Together these sequences fix the locations of all of the genes and unassigned sequences in the genome. Gene products are identified above the sequences. Translations of protein genes are shown with single letter amino acid codes centered below the second nucleotide of their corresponding codons; they include the extensions of reading frames for COI, COIII and ND1, which are marked with curly brackets. Unsequenced regions between segments all fall within protein genes, so that the 5′ portions of incompletely sequenced genes are followed in one segment by their 3′ portions in the following segment. The *Bam*HI site used to make the original λ EMBL3 clone begins at nucleotide 1243 in Segment #1, within the s-rRNA gene. *N* = unidentified nucleotides. *x* = unidentified amino acids.