Evidence That the Large Noncoding Sequence Is the Main Control Region of Maternally and Paternally Transmitted Mitochondrial Genomes of the Marine Mussel (Mytilus spp.)

Li Qin Cao,* Ellen Kenchington,*† Eleftherios Zouros‡ and George C. Rodakis§,1

*Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada, †Department of Fisheries and Oceans, Dartmouth, Nova Scotia B2Y 4A2, Canada, ‡Department of Biology, University of Crete, 71409 Heraklion, Crete, Greece and §Department of Biochemistry and Molecular Biology, National and Kapodistrian University of Athens, 15701 Athens, Greece

Manuscript received December 30, 2003
Accepted for publication February 20, 2004

ABSTRACT

Both the maternal (F-type) and paternal (M-type) mitochondrial genomes of the Mytilus species complex M. edulis/ galloprovincialis contain a noncoding sequence between the tRNA and the rRNA genes, here called the large unassigned region (LUR). The LUR, which is shorter in M genomes, is capable of forming secondary structures and contains motifs of significant sequence similarity with elements known to have specific functions in the sea urchin and the mammalian control region. Such features are not present in other noncoding regions of the F or M Mytilus mtDNA. The LUR can be divided on the basis of indels and nucleotide variation in three domains, which is reminiscent of the tripartite structure of the mammalian control region. These features suggest that the LUR is the main control region of the Mytilus mitochondrial genome. The middle domain has diverged by only 1.5% between F and M genomes, while the average divergence over the whole molecule is ~20%. In contrast, the first domain is among the most divergent parts of the genome. This suggests that different parts of the LUR are under different selection constraints that are also different from those acting on the coding parts of the molecule.

Several species of the mollusk bivalve families Mytilidae, Unionidae, and Veneridae are known to have a mitochondrial DNA (mtDNA) system that differs radically from that of other animal species (Skrbinski et al. 1994a,b; Zouros et al. 1994a,b; Liu et al. 1996; Hoeh et al. 1997; Passamonti and Scali 2001; Serb and Lydeard 2003). The system, known as doubly uniparental inheritance (DUI; Zouros et al. 1994a), is characterized by the presence in the same species of two mtDNA genomes, one that is transmitted through the egg (the F genome) and another that is transmitted through the sperm (the M genome). In the Mytilus species M. edulis, M. galloprovincialis, and M. trossulus, in which DUI has been studied more extensively, it is known that females are normally homoplasmic for the F genome and males are heteroplasmic, with their somatic tissues dominated by the F and their gonad by the M genome (Skrbinski et al. 1994a,b; Zouros et al. 1994a,b; Stewart et al. 1995; Garrido-Ramos et al. 1998).

Hoffmann et al. (1992) published the sequence of 13.9 kb of the total of 17.1 kb of the M. edulis mitochondrial genome. The most striking feature of the genome is its gene order, which has undergone many rearrangements compared to the mtDNA of most metazoans. It contains the full complement of genes of the metazoan mtDNA and an extra tRNA for methionine, but lacks the ATPase8 subunit gene. It also contains five intergenic sequences of no apparent function, four of which are relatively small (79–119 bp) and one of which is large (~1.2 kb). The latter sequence, which we call here the large unassigned region (LUR) to distinguish it from the smaller sequences of unassigned function, is the focus of this study.

The Hoffmann et al. (1992) sequence was published before the discovery of DUI. Subsequent sequencing of selected mtDNA regions from female and male gonads showed that the Hoffmann et al. (1992) sequence was of the F-type (Skrbinski et al. 1994b; Rawson and Hilbish 1995; Stewart et al. 1995). These and subsequent studies (Hoeh et al. 1997; Quesada et al. 1998, 1999) showed that the two sex-specific sequences differ by >20%, a difference normally found among mitochondrial genomes from species of distant taxonomic units. The same studies showed that the M genome evolves faster than the F genome. The faster evolution of the M genome was subsequently observed in freshwater unionid mussels (Liu et al. 1996; Hoeh et al. 1997) and venerid clams (Passamonti and Scali 2001). Hoeh et al. (1996) extended the divergence comparison to species from several animal phyla and concluded that, for at least the COI and the COIII genes that they examined, both
the M and the F mussel mtDNA evolve at higher rates than the mtDNA of other animals.

Another conclusion from the above studies is that, in spite of their large differences in nucleotide substitutions and amino acid replacements, the F and M genomes do not differ in gene content and arrangement. This is confirmed from the full sequence of the F and M genomes of M. galloprovincialis (GenBank accession nos. AV497292 and AT363687; A. Mzii, E. Zouros, N. Moschonas and G. C. Rodakis, unpublished results), which shows that the molecules are identical in gene content and order, including the possession of an additional rRNA for methionine and the lack of the ATPase8 subunit gene. In view of the different transmission mode, it is important to know whether this similarity extends to regions that do not code for a gene but may have regulatory functions. Because it is the largest segment of the mussel genome of unknown function, the LUR has been considered as the primary candidate for the site of regulatory elements of replication and transcription of mussel mtDNA (Burzynski et al. 2003).

Here we consider evidence for the hypothesis that the LUR is the main control region of the mussel mtDNA. In the absence of an in vivo replication/transcription analysis, we have searched for sequences that show sufficient similarity with motifs that are known to play crucial roles in the replication or transcription of well-characterized invertebrate and vertebrate mitochondrial genomes. We have identified several of these motifs in the LUR of both the F and M genomes. In contrast, no such motifs were found in other noncoding regions of these genomes. We also noted that the LUR consists of three distinct domains, of which the middle is highly conserved and the flanking ones are highly divergent within and between genomes. This tripartite structure and pattern of variation, which have not been reported in any other invertebrate species, have strong parallels with the vertebrate mtDNA control region.

MATERIALS AND METHODS

Sample collection and identification: Mussels were collected from Canada (Lunenburg Bay, Nova Scotia; individuals w22, w24, w128, and w143), the United States (Totten Inlet, Puget Sound, Washington; individual 6), and France (Morgat, Bay of Douarnenez, West Brittany; individuals 1 and 12), and transported alive to the laboratory where they were sexed by examining the gonad under a light microscope for the presence of sperm or eggs (Table 1). Total DNA was extracted from gonad tissue using a modified salt extraction procedure (Miller et al. 1988). Species identification was performed by PCR amplification of the internal transcribed spacer of the nuclear ribosomal RNA genes and Hha restriction of the PCR product to produce species-specific restriction patterns (Health et al. 1995). As a second assay, we used the restriction patterns of the polyphenolic adhesive protein Glu gene, as described by Rawson et al. (1996).

DNA amplification, cloning, and sequencing: Total DNA from gonad tissue was used as the template to amplify the LUR with primers UNFOR1, 5'-TTG CGA CCT CGA TGT TGG C-3' and UNREV1, 5'-AGC TCA CCA CCT ATT CCT C-3', which correspond to nucleotide positions 3218–3236 and 4654–4636 of segment 5 of the M. edulis mitochondrial F genome (Hoffmann et al. 1992). One band was obtained in females and two in males. Of the two male bands, one was weaker and of the same size as the female band, and the other was stronger and of smaller size. To enhance the larger band in males, we used the primer UNFOR1, given above in combination with primer UNREV2, 5'-GGG TTA GTG TTA TAT GCA G-3', which corresponds to position 4726–4708 of segment 5 in Hoffmann et al. (1992). The PCR mixture consisted of 2 µl of template DNA, 0.8 mm of each primer, 1 mm dNTP, 2.5 mm MgCl₂, and 1 unit Tag polymerase (MBI Fermentas) in the buffer supplied by MBI Fermentas in a total 25-µl volume. The cocktail was heated initially at 94°C for 3 min and then incubated at 94°C for 1 min, 53°C for 1.5 min, and 72°C for 1 min for 40 cycles and 72°C for 6 min for a final extension.

The amplified products were visualized on a 1% agarose gel, and the corresponding bands were excised. DNA was recovered using the UltraClean DNA purification kit (Mo Bio) and cloned on pGEM-T vector (Promega, Madison, WI) following the procedure provided by the supplier. Positive clones from each individual were confirmed by PCR amplification with the same primers and conditions as above. One randomly chosen clone from each female and two from each male, one with the small and another with the large band, were sequenced commercially from both directions with 80–85% overlap using either the LICOR 4200 or the ABI 373 automated sequencer.

Sequence analysis: Sequences were aligned with the aid of the computer program ClustalX (Thompson et al. 1997). Selection of the optimal parameters (opening and extension gap penalties) was performed according to Douris et al. (1998). Further corrections of the alignment were performed manually to maximize the sequence similarity, particularly for the two size-variable regions that flank the relatively conserved central region (Appendix). Kimura two-parameter genetic distances (Kimura 1980) and neighbor-joining (Saitou and Nei 1987) trees were calculated using the computer program MEGA, version 2.1 (Kumar et al. 2001). The Goss and Lewontin (1996) statistical test was performed using the DNA-slider computer program (McDonald 1998) for 10,000 simulation replicates; the estimation of recombination parameter (R) was performed according to the technique suggested by the authors (M-type sequences, R = 20; F-type sequences, R = 12). The program DnaSP version 3.99 (Roza et al. 2003) was used to generate the input file for the DNA-slider program. Prediction of potential secondary structures was performed by the online version of the mfold software, version 3.1 (Zuker 2003).

RESULTS

We have included in this study sequences from seven individual mussels—three females, two of which belonged to species M. edulis and one to its sibling species M. galloprovincialis, and four males, two of which belonged to M. edulis and two to M. galloprovincialis (Table 1). We focused on a segment of the mitochondrial genome that includes part of the l-rRNA gene, the LUR, and a small part of the tRNA⁹ gene. Each sequence was characterized as F-type or M-type on the basis of the part of the l-rRNA gene, whose F-type and M-type sequences are known from previous studies (Rawson and Hilbish 1995). Each female yielded an F-type sequence,
TABLE 1
Origin, type, and the parts of 11 mtDNA sequences studied

<table>
<thead>
<tr>
<th>Species</th>
<th>Gender</th>
<th>Individual code</th>
<th>Sequence code</th>
<th>mtDNA type</th>
<th>Length (in bp)</th>
<th>LUR</th>
<th>Total</th>
<th>F/M (SE)</th>
<th>Diversity within F (SE)</th>
<th>Diversity within M (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M. edulis</strong></td>
<td>♀</td>
<td>w22</td>
<td>ef.w22-F</td>
<td>F</td>
<td>227 654 368 137 1159 12 1398</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>w24</td>
<td>ef.w24-F</td>
<td>F</td>
<td>227 654 366 136 1156 12 1395</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>♂</td>
<td>w128</td>
<td>em.w128-F</td>
<td>F</td>
<td>227 654 368 136 1158 12 1397</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>w143</td>
<td>em.w143-M</td>
<td>M</td>
<td>227 490 368 70 928 12 1168</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>em.w143-M</td>
<td></td>
<td></td>
<td>227 490 367 71 928 12 1167</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M. galloprovincialis</strong></td>
<td>♀</td>
<td>#1</td>
<td>gf.1-F</td>
<td>F</td>
<td>227 654 368 138 1160 12 1399</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>♂</td>
<td>#6</td>
<td>gm.6-F</td>
<td>F</td>
<td>227 690 368 136 1194 12 1433</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>#12</td>
<td>gm.12-F</td>
<td>F</td>
<td>227 654 368 136 1158 12 1397</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>gm.12-M</td>
<td></td>
<td></td>
<td>227 500 368 69 937 12 1176</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Divergence 0.197 0.496 0.017 0.156 0.237 0.229
F/M (SE) (0.034) (0.047) (0.005) (0.053) (0.017) — (0.015)
Diversity 0.017 0.016 0.008 0.021 0.016 — 0.016
within F (SE) (0.005) (0.003) (0.002) (0.011) (0.002) — (0.002)
Diversity within M (SE) 0.007 0.054 0.007 0.036 0.033 0.028

*a* Contains a 36-bp insert.

*b* Contains an additional 10-base-long A-stretch.
as is the case with most female mussels. Male mussels normally contain an M-type and an F-type sequence (“typical” males), but exceptional males that lack an M-type sequence at the examined part of the molecule (“atypical” males) can be found in varying frequencies in natural populations (see Ladoukakis et al. 2002 for review and terminology). All four males used in this study were typical; i.e., they yielded an F-type and an M-type sequence. Thus, the whole set of data consisted of 4 M-type and 7 F-type sequences. The 11 nucleotide sequences are given in the appendix with the corresponding sequence from Hoffmann et al. (1992).

**The tripartite structure of the LUR:** The large unassigned region of the mussel mtDNA was first identified by Hoffmann et al. (1992) as the noncoding region between the lrsRNA and the tRNA^Sp gene. The 3′-end of the LUR can be easily defined as the nucleotide preceding the first nucleotide of the tRNA^Sp gene, which itself can be identified from the tRNA folding pattern. In contrast, the 3′-end of the lrsRNA gene cannot be readily identified because the length of the gene as well as the sequence of its 3′-end vary among metazoan species (e.g., Hatzoglou et al. 1995). This introduces an arbitrariness regarding the 5′ starting point of the LUR. For consistency, we have used the same starting point as Hoffmann et al. (1992).

Visual inspection of the 12 aligned LUR sequences (appendix) suggests that there is a high degree of differentiation between the set of F sequences and the set of M sequences. The first and most marked difference is the presence of a large number of indels of varying length. Gaps are more common in the M sequences with the net result that M sequences are shorter than F by ~250 bp. Among the 4 M sequences examined, the length varied from 925 to 937 bp, whereas among the 8 F sequences the length varied from 1156 to 1194 bp (Table 1). Indels are not randomly distributed over the LUR. Starting from the 5′-end, they are commonly found for about half of the length, are practically absent in the next ~360 nucleotide positions, and become prominent again in the remaining part of LUR. This indel-based division of LUR in three parts is further divergence from M sequences before the region of low sequence variation corresponds to alignment position 897, which is also the position of the last indel that differentiates F from M sequences before the region of low sequence divergence (appendix and Figure 1). To identify the other end of the middle domain, we followed the same method for the part of LUR downstream from position 897. Exempting the large terminal indel, the length from position 897 to position 1343 (where the large indel starts) is maximally divided at position 1265, which is also the last point before the reappearance of indels. We refer to the three parts of LUR defined by these nucleotide positions as variable domain 1 (VD1), conserved domain (CD), and variable domain 2 (VD2). It must be emphasized that the division points are mostly for reference. Burzynski et al. (2003) have also noted that the large unassigned region of the *M. edulis* species complex can be divided in terms of polymorphism in three regions, which correspond closely to ones described here.

**Figure 1.—** Nucleotide identity in steps of 30 alignment positions. A position was considered identical if occupied by the same nucleotide in all 12 sequences and if not an indel. Dots correspond to the midpoint of the sliding window. The four segments that comprise the aligned sequences are shown schematically.
VD1 is the most variable region. Its length in the F genomes is 654 bp (Table 1). An exception occurs in one *M. galloprovincialis* sequence (sequence gm.6-F), which carries a 36-bp insert (APPENDIX) that is an almost perfect repeat of the preceding 36 bp and has most probably resulted from replication slippage (LEVINSON and GUTMAN 1987). The length of VD1 is 490 bp in two M genomes, 488 bp in a third, and 500 bp in the fourth. The latter size was found again in a *M. galloprovincialis* individual (sequence gm.12-M) and is due to an insert of 10 adenines that is flanked by strings of adenines, also a possible result of replication slippage. The length of the central domain (CD) is similar in F and M genomes, varying between 366 and 368 bp. The third domain (VD2) is dominated by strings of purines (73.8% A + G, compared to 55.2% for the first region and 53.9% for the second) and contains a 67-bp gap that is present in all M sequences and absent in all F sequences. This is the largest indel for the whole LUR. Mean intergenomic divergence and mean intragenomic diversities are listed in Table 1 for the three LUR domains and for the sequenced part of *l*-rRNA. The values for *l*-rRNA are typical of the whole genome (RAWSON and HILBRISH 1995; STEWART et al. 1995; HOEH et al. 1996; QUESADA et al. 1998). Divergence in the central domain of LUR is lower than is typical for the genome by an order of magnitude, but higher by ~2.5 times in the first domain. The third domain is not different from *l*-rRNA or the rest of the genome. As expected for these values, phylogenetic trees (see online figure “NJ-trees” available at http://www.genetics.org/supplemental/) based on Kimura two-parameter distances produced a clear separation of the F and M sequences for all four parts, except for one F sequence (gm.6-F), which clustered with the M sequences in the tree based on the CD region.

The recognition of three parts in the LUR, of which the central part is conservative and the two flanking parts are variable, is suggestive of a structural similarity with the control region of vertebrate mtDNA. This tripartite structure is well established in mammals (SACCONE et al. 1991, 1999; SBISÀ et al. 1997; STONEKING 2000) and has also been observed in other vertebrates (SACCONE et al. 1987; MARSHALL and BAKER 1997; RANDI and LUCCHINI 1998; DELPORT et al. 2002; RAY and DENSMORE 2002), but presently there is no evidence for its presence in other invertebrate species.

Evidence that the LUR is the main control region of the mussel mtDNA: Most animal mitochondrial genomes have a “major” noncoding region, which is distinctly longer than other “minor” noncoding regions and contains elements involved in the regulation of replication and transcription (WOLSTENHOLME 1992; SHADEL and CLAYTON 1997; BOORE 1999; SACCONE et al. 2002). Such elements are also found in smaller noncoding regions (see DISCUSSION). Because the LUR is, as noted, the largest noncoding segment of the mussel mtDNA, it was considered as the prime candidate for the main control region (BURZYNSKI et al. 2003). To produce evidence for this hypothesis, we compared the LUR with well-characterized control regions of other animal mitochondrial genomes, namely the fruit fly *Drosophila melanogaster* (GODDARD and WOLSTENHOLME 1978, 1980; CLARY and WOLSTENHOLME 1985, 1987; LEWIS et al. 1994), the sea urchins *Strongylocentrotus purpuratus* and *Paracentrotus lividus* (JACOBS et al. 1988, 1989; CANTATORE et al. 1989, 1990; ELLIOTT and JACOBS 1989; MAYHOOK et al. 1992; LOGUERCIO POLOSA et al. 1999; ROBERTI et al. 1999), and the mammal *Homo sapiens* (reviewed by SHADEL and CLAYTON 1997; SACCOME et al. 1999, 2002; TAANMAN 1999; CLAYTON 2000). In addition, we searched the LUR for the presence of palindromes that have the ability to form secondary structures and could, therefore, be involved in the regulation of replication and transcription (BROWN et al. 1986) or of motifs that have sufficient sequence similarity with known cis-acting elements. The fruit fly control region is dominated by strings of A and T residues and, therefore, provides no suitable background against which to search for regional similarities with the LUR. Tracks of TA are found in the LUR, but a direct correspondence with those of the fruit fly control region cannot be made.

The sea urchin’s control region is only 121–136 bp long (*S. purpuratus*, 121 bp; JACOBS et al. 1988; *P. lividus*, 132 bp, CANTATORE et al. 1989; ARBOCIA LIXULA, 136 bp, DE GIORGI et al. 1996) and, as a whole, it shows little sequence similarity with the LUR or the human control region. However, JACOBS et al. (1988, 1989) and CANTATORE et al. (1989, 1990) have suggested several analogies between the human and the sea urchin control region, which they characterized as a “condensed version” of the vertebrate mtDNA replication origin. More specifically, direct experiments (JACOBS et al. 1989; CANTATORE et al. 1990) imply that the mechanisms of replication initiation of sea urchin and vertebrate mtDNA are similar and that the major noncoding region of the sea urchin mtDNA contains sequence motifs that are homologous to the mammalian conserved sequence blocks (CSBs; CANTATORE et al. 1989). Analogies of this type also exist between the sea urchin control region and the LUR (Figure 2). The *S. purpuratus* control region contains a string of 20 Gs (nucleotide positions from 75 to 94) that divides it into a proximal and a distant domain. A string of similar length (varying from 21 to 26 bp) with 80% in G is found in the third domain of LUR. The proximal domain of the sea urchin (nucleotide positions 1–74) shows loose correspondence with a region in the conserved domain of LUR. The distal domain (nucleotide positions 95–121) is more interesting because it contains the motif TATATATAA, which is the consensus sequence found in the other four noncoding regions of echinoid mtDNA (DE GIORGI et al. 1996) and may represent a bidirectional promoter (ELLIOtT and JACOBS 1989; CANTATORE et al. 1990; ROBERTI et al.
Figure 2.—Sequence matching of the sea urchin (A) and five functional elements of the human (B) mtDNA control region with sequences of the F- and M-type LUR. See text for explanation of symbols used for human elements. The mussel sequences are from ef.w22-F (F-type) and em.w143-M (M-type; appendix). The sea urchin sequence used is from *S. purpuratus*, GenBank accession no. NC001453, region 1085–1205 (Sp1-3). The human sequence used is the revised Cambridge reference sequence, GenBank accession no. NC001807, corrected according to Andrews et al. (1999) and annotated by Wallace and Lott (2003). Numbers in parentheses correspond to the probability that the sequence match is accidental, as obtained from the online program PRDF (Pearson and Lipman 1988) for 1000 replicates (ktup = 1).
Figure 3.—Potential tRNA-like structures in the first variable domain of the human mtDNA control region and the first domain of the F- and M-type LUR. The mussel sequences are from ef.w22-F (F-type) and em.w143-M (M-type; Appendix).

The Control Region of Mussel mtDNA

The search for folding sequences in the LUR identified one tRNA-like structure in the first domain of the F genome and two in the first domain of the M genome (Figure 3). In the human control region there is also a similar tRNA-like secondary structure extending from the beginning of the first region to the beginning of TAS (Brown et al. 1986). We have also noted that about half of the second domain of the LUR can be folded into a stem-like secondary structure (Figure 4). This induced us to examine if a similar structure can be formed in the human control region. Indeed, about two-thirds of the second domain can fold to produce a stem-like secondary structure similar to the one of the Mytilus LUR. No secondary structures of any comparable size can be found in the much smaller sea urchin control region.
The arrangement of these elements along the lengths of the LUR and the human control region invites comment. With the exception of mtTF1, which in the LUR is before rather than after CSB1, all elements and the folding sequences are colinearly arranged. Three elements (mt4, CBS1, mtTF1) are closely located with each other in all three genomes and found in the second domain of LUR, but in the third domain of the human control region. Element mt3(L) is located farther upstream and within the stem-like structure in the second domain of all three genomes. TAS is located even farther up in the first domain of the genomes and after the sequence that forms a tRNA-like structure.

**DISCUSSION**

In this study we examine the hypothesis that the large noncoding region of the Mytilus mtDNA (LUR) is the main control region of this genome. We consider this to be an important issue, given the exceptional mtDNA transmission system of this species, i.e., its DUI. A central question about DUI is whether there is a systematic difference in information content between maternally and paternally transmitted genomes. From the nearly full sequence of the F genome of *M. edulis* (Hoffmann et al. 1992) and the complete sequence of the M and F genome of its sibling species *M. galloprovincialis* (GenBank accession nos. AY363687 and AY497292, respectively; A. Mizi, E. Zouros, N. Moschonas and G. C. Rodakis, unpublished results) we know that the two molecules are identical in terms of coding genes and gene order, even though highly divergent in terms of nucleotide substitutions and amino acid replacements. Consideration of the LUR as a possible control region required the examination of a collection of F and M sequences, first because sequences of noncoding regions are known to foster large amounts of variation and second because only the examination of several LURs from each type of mtDNA could help identify F-specific and M-specific differences.

Examination of the LUR from several M and F genomes shows that this region has the same organization in both types of genomes. The LUR can be clearly divided into three domains, which differ in the amount of indels and in the rate of divergence among themselves and differ from the other parts of the genome for which there is information on variation in natural populations (Rawson and Hilbish 1995; Stewart et al. 1995; Hoeh et al. 1996; Quesada et al. 1998). The LUR contains one of the slowest and one of the fastest-evolving parts of the mussel mtDNA. Given that these parts are exposed to the same stochastic forces (and barring the unlikely possibility that the mutation rate could vary drastically among adjacent segments), there can be little doubt that the pronounced difference in sequence similarity in the LUR and the human control region are shown above the line in Figure 5. The arrangement of these elements along the lengths of the LUR and the human control region invites comment. With the exception of mtTF1, which in the LUR is before rather than after CSB1, all elements and the folding sequences are colinearly arranged. Three elements (mt4, CBS1, mtTF1) are closely located with each other in all three genomes and found in the second domain of LUR, but in the third domain of the human control region. Element mt3(L) is located farther upstream and within the stem-like structure in the second domain of all three genomes. TAS is located even farther up in the first domain of the genomes and after the sequence that forms a tRNA-like structure.

**Figure 4.**—Stem-and-loop structure in the central domain of the human control region and the second domain of the F and M LUR. The mussel sequences are from ef.w22-F (F genome) and em.w143-M (M genome; **APPENDIX**).
rate of divergence is the result of different selection pressures acting upon the three domains of the LUR. This in turn suggests that the LUR, or parts of it, have an important function in the mitochondrial genome.

Most animal mitochondrial genomes contain a major noncoding region and several smaller ones. As a rule, the major noncoding region is found to contain elements that are involved in the regulation of replication and transcription (Shadel and Clayton 1997). It is noteworthy, however, that particular motifs and palindromes that function as promoters [such as the origin of replication of the light strand (O_L); Chang et al. 1985; Roe et al. 1985; Clary and Wolstenholme 1987; Okimoto et al. 1992; Clayton 2000] or as recognition signals for enzymes involved in transcription or processing (Cantatore et al. 1989, 1990; Roberti et al. 1991, 1999; Mayhook et al. 1992; Valverde et al. 1994) are located outside of the major noncoding region. For this reason the major noncoding region in vertebrates is more accurately referred to as the main control region. The complete sequence of several mitochondrial genomes has also shown that they contain a major noncoding region (Shadel and Clayton 1997; Boore 1999; Saccone et al. 2002). Exceptions are the mtDNA molecules of the snake Dinodon semicarinatus and the cephalopod Loligo bleekeri, which, respectively, contain two (1018 bp) and three (507–515 bp) nearly identical noncoding sequences (Kumazawa et al. 1998; Tomita et al. 2002). Another exception is the presence of only short noncoding regions. In the land snail Albinaria coerulae the largest noncoding region is 42 bp and contains, together with another noncoding region of 16 bp, A + T-rich palindrome sequences (Hatzoglou et al. 1995). In Amphioxus, the longest noncoding segment is 57 bp and does not contain palindromes or any motif known to be implicated in the replication of the mtDNA (Spruit et al. 1998). No fewer than 28 small (from 2 to 282 bp) noncoding regions were found in the mtDNA of the fresh water bivalve Lampsilis ornata, but they were not reported to contain motifs with notable similarity to known regulatory elements, except an increased A + T content in a 136-bp noncoding region (Serb and Lydeard 2003).

The evidence that the LUR is the main control region of the mussel mtDNA can be summarized as follows. First, it is capable of producing characteristic secondary structures. Second, a motif of the sea urchin mtDNA molecule for which there is wide consensus that it plays a crucial role in replication and transcription (Elliott and Jacobs 1989; Cantatore et al. 1990; Roberti et al. 1991, 1999; De Giorgi et al. 1996; Loguercio Polosa et al. 1999; Fernandez-Silva et al. 2001) was found in the central, most conservative, domain of the LUR. Third, there are several similarities between the LUR and the mammalian control region. These include the tripartite structure, the presence of five motifs of significant sequence similarity and same relative position, and the presence of similar secondary structures in corresponding domains. Whereas each of these “matches” can be fortuitous when taken in isolation, it is difficult to maintain this claim for the entire set of them. Fourth, the complete sequence of the F and M genomes (GenBank
accession nos. AY363687 and AY497292, respectively; A. Mizi, E. Zouros, N. Moschonas and G. C. Rodakis, unpublished results) failed to identify sequences in other noncoding regions with the ability to form tRNA-like structures or with sufficient similarity to motifs that are characteristic of main control regions of other mitochondrial genomes. The only exception is the presence of a stem loop and a tRNA-like structure [pseudo-tRNA \textsuperscript{uc} (UCN)] in the second largest noncoding region, which is located between the NDIII and COI genes and might be related to transcript processing mechanisms (Beagley et al. 1999).

These results provide substantial support that the LUR is the main control region of the mussel mtDNA, as proposed by Hoffmann et al. (1992) and Burzynski et al. (2003). These latter authors examined the LUR from several female and male individuals of *M. trossulus* from the Baltic Sea. In this population an introgression of the *M. edulis* mtDNA has resulted in a nuclear-mtDNA mosaic, in which the nuclear genome is of the *M. trossulus* and the mtDNA of the *M. edulis* type. The majority of LURs obtained by Burzynski et al. (2003) were of the F or M *M. edulis* type, as expected. But they also found two recombinant types in male gonads that were of the F-type but contained a segment of the M-type sequence at the first domain of LUR. This observation prompted Burzynski et al. (2003) to suggest that the first part of the large unassigned region, i.e., the first domain of LUR, may contain sequences that determine whether a genome will be maternally or paternally transmitted. This is consistent with our observation that the first domain is the most differentiated among the two types of genomes and warrants further focusing on the LUR, both as the control region and as the region that may determine whether a mitochondrial genome will be transmitted by the female or the male gamete.

We are grateful to two anonymous reviewers for comments and suggestions. This work was supported by the Natural Sciences and Engineering Research Council of Canada (to E.K.), by Dalhousie University and Patrick Lett Scholarships and Fisheries and Oceans, Canada (to L.C.), by the Greek General Secretariat for Research and Technology (grant PENED-01ED42 to E.Z. and G.C.R.), and by the University of Athens (to G.C.R.).

LITERATURE CITED


Beagley, C. T., R. Okimoto and D. R. Wolstenholme, 1999 *Mylitus* mitochondrial DNA contains a functional gene for a tRNA\textsuperscript{uci} (UCN) with a dihydrooridine arm-replacement loop and a pseudo-tRNA\textsuperscript{uc} (UCN) gene. Genetics 152: 641–652.


Heath, D. D., P. D. Rawson and T. J. Hildish, 1995 PCF-based nuclear markers identify alien blue mussel (*Mylitus spp.*) geno-


Taamsen, J. W., 1999 The mitochondrial genome: structure, trans-
1410: 103–123.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and
D. G. Higgins, 1997 The ClustalX windows interface: flexible
strategies for multiple sequence alignment aided by quality analy-

Tomita, K., S. Yokobori, T. Oshima, T. Ueda and K. Watanabe,
2002 The cephalopod Loligo bleekeri mitochondrial genome:
multiplied noncoding regions and transposition of tRNA genes.

Valverde, J. R., R. Marko and R. Garesse, 1994 A conserved hep-
tamer motif for ribosomal RNA transcription termination in ani-

Wallace, D. C., and M. T. Lott, 2003 MITOMAP: a human mito-
ochondrial genome database (http://www.mitomap.org).

Wolstenholme, D. R., 1992 Animal mitochondrial DNA: structure

Zouros, E., A. O. Ball, C. Saavedra and K. R. Freeman, 1994a Mi-

Zouros, E., A. O. Ball, C. Saavedra and K. R. Freeman, 1994b A
new type of mitochondrial DNA inheritance in the blue mussel

Zuker, M., 2003 Mfold web server for nucleic acid folding and

Communicating editor: T. H. Eickbush
**Figure A1.** Multiple alignment of the 12 nucleotide sequences used in this study. The boxed sequence in gm.6-F is found in tandem repeat with two changes (T instead of A in the last position and T instead of C in the seventh position from the end). Numbers in parentheses (far right column) indicate the actual length of sequence gm.6-F.
Figure A1.—Continued.
The Control Region of Mussel mtDNA

<table>
<thead>
<tr>
<th>Figure A1.—Continued.</th>
</tr>
</thead>
<tbody>
<tr>
<td>M83756-F</td>
</tr>
</tbody>
</table>
Figure A1.—Continued.