DOI: 10.1534/genetics.103.026187

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

Liqin 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 *l-rRNA* and the *tRNA^{Tyr}* 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 (Skibinski *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 (Skibinski *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 (Skibinski *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

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY350784–AY350794.

[1]*Corresponding author:* Department of Biochemistry and Molecular Biology, National and Kapodistrian University of Athens, Panepistimioupolis, 15701 Athens, Greece. E-mail: grodakis@biol.uoa.gr

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. AY497292 and AY363687; A. Mizi, 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 *tRNA* 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*I restriction of the PCR product to produce species-specific restriction patterns (Heath *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′-GCG 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 $MgCl_2$, and 1 unit *Taq* polymerase (MBI Fermentas) in the buffer supplied by MBI Fermentas in a total 25-μl volume. The cocktail was heated initially at 94° for 3 min and then incubated at 94° for 1 min, 53° for 1.5 min, and 72° for 1 min for 40 cycles and 72° 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 (Rozas *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 $tRNA^{Tyr}$ 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

Species	Gender	Individual code	Sequence code	mtDNA type	Length (in bp)						
						LUR					
					l-rRNA	VD1	CD	VD2	Total	*tRNA^{Tyr}*	Total
M. edulis	♀	w22	ef.w22-F	F	227	654	368	137	1159	12	1398
		w24	ef.w24-F	F	227	654	366	136	1156	12	1395
	♂	w128	em.w128-F	F	227	654	368	136	1158	12	1397
			em.w128-M	M	228	490	368	70	928	12	1168
		w143	em.w143-F	F	227	654	368	137	1159	12	1398
			em.w143-M	M	227	490	367	71	928	12	1167
M. galloprovincialis	♀	#1	gf.1-F	F	227	654	368	138	1160	12	1399
	♂	#6	gm.6-F	F	227	690[a]	368	136	1194	12	1433
			gm.6-M	M	227	488	367	70	925	12	1164
		#12	gm.12-F	F	227	654	368	136	1158	12	1397
			gm.12-M	M	227	500[b]	368	69	937	12	1176
	Divergence F/M (SE)				0.197 (0.034)	0.496 (0.047)	0.017 (0.005)	0.156 (0.053)	0.237 (0.017)	—	0.229 (0.015)
	Diversity within F (SE)				0.017 (0.005)	0.016 (0.003)	0.008 (0.002)	0.021 (0.011)	0.016 (0.002)	—	0.016 (0.002)
	Diversity within M (SE)				0.007 (0.004)	0.054 (0.008)	0.007 (0.003)	0.036 (0.019)	0.033 (0.004)	—	0.028 (0.004)

[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 *l-rRNA* and the *tRNA*^{Tyr} gene. The 3′-end of the LUR can be easily defined as the nucleotide preceding the first nucleotide of the *tRNA*^{Tyr} gene, which itself can be identified from the tRNA folding pattern. In contrast, the 3′-end of the *l-rRNA* 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 supported by the degree of inter- and intragenomic variation along the region. Figure 1 gives the percentage of sites, in consecutive lengths of 30 bp, that are fixed for the same nucleotide in all 12 sequences. Presence of a different nucleotide or an indel in a nucleotide position is counted as a difference. Identity is consistently low in the first part of the LUR, rises dramatically in the second part, and drops again in the third.

Statistical tests of the distribution of variability across a sequence in a collection of sequences normally ignore indels, which, as noted, constitute a prominent distinguishing feature of F from M LURs. Even so, the Goss and Lewontin (1996) test yields significant results. For the four M sequences (using the F sequence em.w143-F as outgroup), the interval length variance (V_{IL}) and the modified internal length variance (Q_{IL}) are 0.0008 (P = 0.018) and 0.0014 (P = 0.044), respectively. For the eight F sequences (using the M sequence em.w143-M as outgroup), the corresponding figures are V_{IL} = 0.0009 (P = 0.084) and Q_{IL} = 0.0024 (P = 0.036). Goss and Lewontin (1996) suggest that the null hypothesis of homogeneity should be rejected even if only one of the two tests is significant. We conclude, therefore, that there is substantial heterogeneity of divergence across the LUR, even if distribution of indels is ignored.

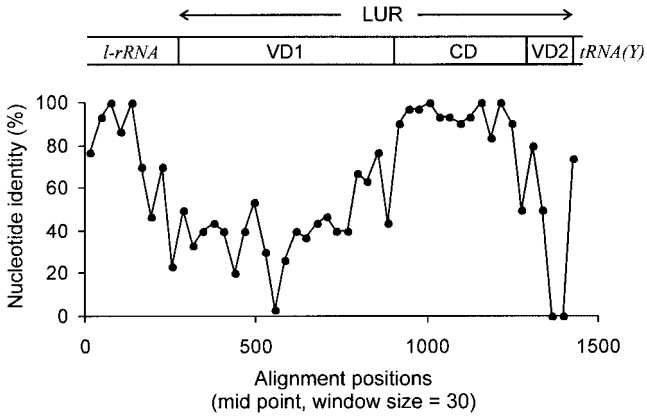

Figure 1.—Nucleotide identity in steps of 30 alignment positions. A position was considered identical if occupied by the same nucleotide in all 12 sequences and if not an indel. Dots correspond to the midpoint of the sliding window. The four segments that comprise the aligned sequences are shown schematically.

To provide statistical support for the internal points that define the middle region of LUR, we first searched for the nucleotide position that divides the LUR in two parts (first domain *vs.* second and third domains) in such a way that heterogeneity of the two parts in the distribution of fixed *vs.* variable sites is maximized (all 12 sequences and gap positions considered). The 2 × 2 chi-square homogeneity test that produces the highest value 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.

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 HILBISH 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; SACCONE *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; *Arbacia 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.*

A

```
                ** * *     *   *  ****  *     *** *    * *  *** **** ***  ** * ******   **** * *
Mytilus-F  829 AAAA-TGGGATTTGGAAGGGTCTAATTTTTCTCGTTTGACTTAATTCTGGTTGCTCACGTGACTTACCTGGGTTTGAAAAC 908
Sp1          1 AAGAATCCCCATC--ACAGCTCCCACCCCTCTC-TCCTTCGTCGTTCGGGTTTCTCG-GT-A-TTACCTCT-TTTGCATAT 74
Mytilus-M  664 AAAAATGGGATTTGGAAGG-TCTAATTTTTCTCGTTTGACTTAATTCTGGTTGCTCACGTGATTTACCTGGGTTTGAAAAC 743
                ** ***     *   *  * **  *    **** *    * *  *** **** ***  ** * ******   **** * *

                *******  *    ** ******
Mytilus-F 1054 GGGGGGGCTGACC-GGAGGGGGG 1075
Sp2         75 GGGGGGGG-GGG--GGGGGGGGG 94
Mytilus-M  889 GGGGGGGCTGGCCTGG-GGGGGG 911
                *******  **   ** ******

                * **  **** ******** * ***
Mytilus-F  928 AATCAGAATATATATATAAATCAAGGTTA 956
Sp3         95 ACTCTAAATA-ATATATAA-TGAAGAACC 121
Mytilus-M  763 AATCAGAATATATATATAAATCAAGGTTT 791
                * **  **** ******** * ***
```

B

TAS

```
                 *********  * **
Mytilus-F  143 CACATAAAAATTCCAT 158 (P=0.013)
Human      133 TACATAAAAACCCAAT 149
Mytilus-M  346 TACGTAAAAGTCTACT 362 (P=0.026)
                *** *****  * * *
```

CSB1

```
                      *** ***  * ****   * ** ****
Mytilus-F  975 TAAATTATCGGTTGTTTAAAGAAATAA 1001 (P=0.049)
Human      759 TAA-TTAATGCTTGT---AGGACATAA  781
Mytilus-M  809 TAAATTATCGGTTGTTCAAAGAAATAA  836 (P=0.032)
                *** ***  * ****   * ** ****
```

mtTF1

```
                 **   * * ******** ***   *  ***
Mytilus-F  948 TCAAGGTTAAAAAAAATTCCCAAAGCGTAAAT 979 (P=0.032)
Human      822 ACATCATAACAAAAAATTTCCA---C-CAAAC 849
Mytilus-M  782 TCAAGGTTTAAAAAAATTCCCAAAGCGTAAAT 814 (P=0.032)
                 **   *   ******** ***   *  ***
```

mt3(L)

```
                      ** ******
Mytilus-F  871 ATTCTGGTT 879 (P=0.003)
Human      476 AT-CTGGTT 483
Mytilus-M  706 ATTCTGGTT 714 (P=0.003)
                      ** ******
```

mt4

```
                 *****  *
Mytilus-F  1038 GTAACATAC- 1046 (P=0.063)
Human       917 CTAACAC-CA  925
Mytilus-M   870 GTAACACACA  879 (P=0.044)
                 ****** **
```

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).

1999) or a binding site for transcription termination factors (ROBERTI *et al.* 1991, 1999; LOGUERCIO POLOSA *et al.* 1999; FERNANDEZ-SILVA *et al.* 2001). The same motif is found in the conserved domain of the Mytilus LUR. Interestingly, the sequence around this motif also shows similarity between the two molecules, which drops the farther one moves away from the motif (Figure 2).

The control region of the human mitochondrial genome contains several elements for which there is firm evidence that they are involved in the initiation or termination of replication or transcription of the genome (see WALLACE and LOTT 2003 for review). Five of these elements show statistically significant sequence similarity with an equal number of motifs in the LUR (Figure 2). These are the termination-associated sequence (TAS) element, which is considered as the termination site of heavy-strand synthesis (DODA *et al.* 1981); the CSB1 (conserved sequence block) element, which is one of the three CSBs assumed to play a role in replication and/or transcription (CHANG and CLAYTON 1985); and three binding sites of transcription factors (mTF1, FISHER *et al.* 1987; mt3 and mt4, SUZUKI *et al.* 1991). For other elements of the human control region, sequence similarity with parts of the LUR was below the conven-

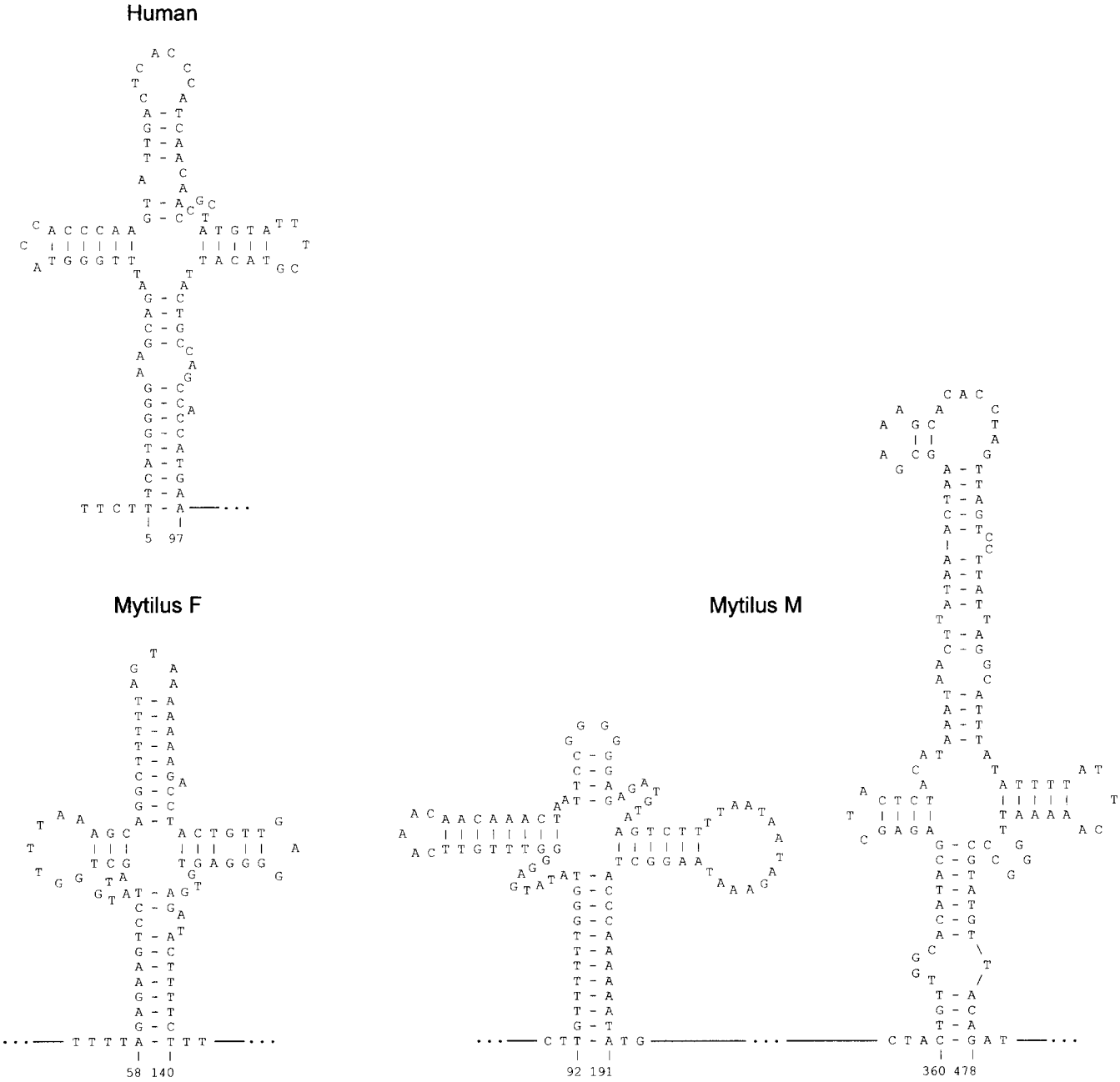

FIGURE 3.—Potential tRNA-like structures in the first variable domain of the human mtDNA control region and the first domain of the F- and M-type LUR. The mussel sequences are from ef.w22-F (F-type) and em.w143-M (M-type; APPENDIX).

tional 5% level of significance (data not shown). We note that low sequence similarity of homologous elements (or even complete absence of an element) is often observed in comparisons of control regions from different mammalian species (SBISÀ *et al.* 1997; SACCONE *et al.* 2002).

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.

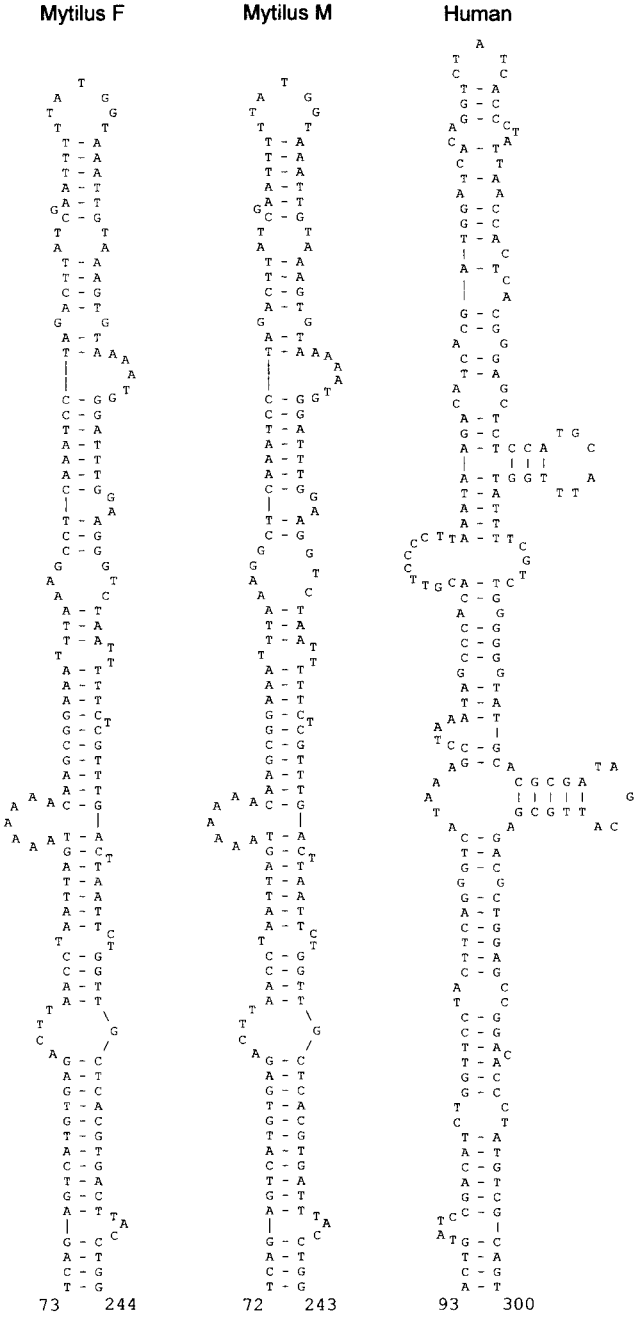

FIGURE 4.—Stem-and-loop structure in the central domain of the human control region and the second domain of the F and M LUR. The mussel sequences are from ef.w22-F (F genome) and em.w143-M (M genome; APPENDIX).

Figure 5 summarizes the information from the comparison of the LUR with the sea urchin and human control regions. The two Mytilus and the human control regions are presented as tripartite structures. The three domains of the sea urchin control region, as defined above, are shown in Figure 5 below the line of the two LURs in the positions with which they show highest affinity. The three regions that can fold into secondary structure and the five motifs that show statistical 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.

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* (Gen Bank 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

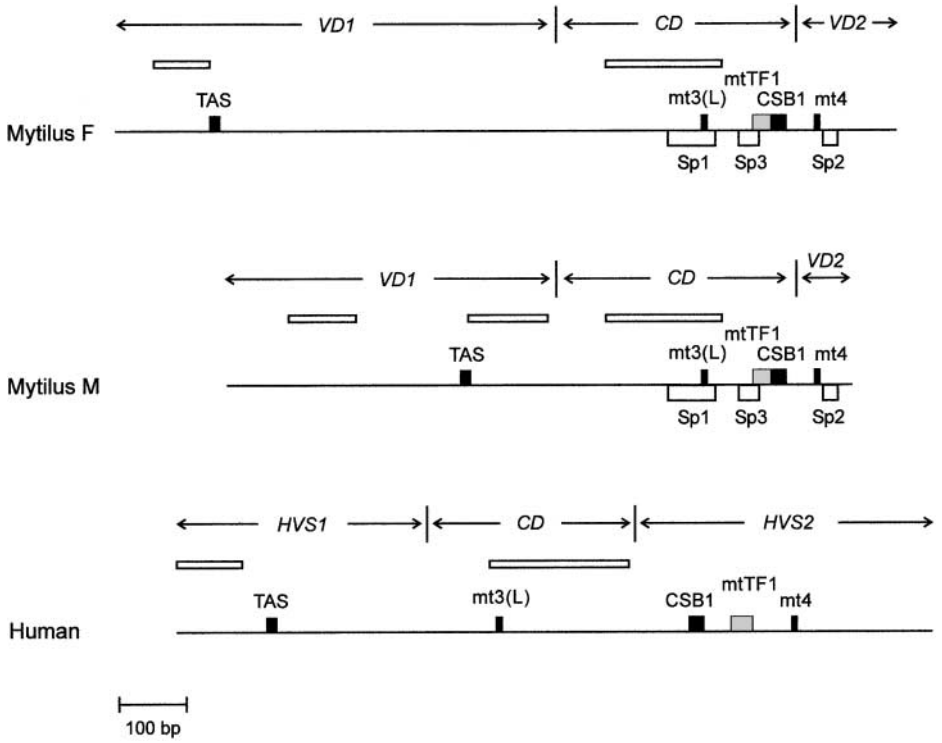

FIGURE 5.—Apparent correspondence between elements and folding sequences of the F and M LUR and the human (above the line) and sea urchin (below the line) control regions. See text and Figure 2 for explanation of symbols.

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 coerulea* 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^{Ser} (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

Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull *et al.*, 1999 Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. **23:** 147.

Beagley, C. T., R. Okimoto and D. R. Wolstenholme, 1999 Mytilus mitochondrial DNA contains a functional gene for a tRNA^{Ser} (UCN) with a dihydrouridine arm-replacement loop and a pseudo-tRNA^{Ser}(UCN) gene. Genetics **152:** 641–652.

Boore, J. L., 1999 Animal mitochondrial genomes. Nucleic Acids Res. **27:** 1767–1780.

Brown, G. G., G. Gadaleta, G. Pepe, C. Saccone and E. Sbisà, 1986 Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J. Mol. Biol. **192:** 503–511.

Burzynski, A., M. Zbawicka, D. O. Skibinski and R. Wenne, 2003 Evidence for recombination of mtDNA in the marine mussel *Mytilus trossulus* from the Baltic. Mol. Biol. Evol. **20:** 388–392.

Cantatore, P., M. Roberti, G. Rainaldi, M. N. Gadaleta and C. Saccone, 1989 The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J. Biol. Chem. **264:** 10965–10975.

Cantatore, P., M. Roberti, P. Loguercio Polosa, A. Mustich and M. N. Gadaleta, 1990 Mapping and characterization of *Paracentrotus lividus* mitochondrial transcripts: multiple and overlapping transcription units. Curr. Genet. **17:** 235–245.

Chang, D. D., and D. A. Clayton, 1985 Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. USA **82:** 351–355.

Chang, D. D., T. W. Wong, J. E. Hixon and D. A. Clayton, 1985 Regulatory sequences for mammalian mitochondrial transcription and replication, pp. 135–144 in *Achievements and Perspectives of Mitochondrial Research, Vol. II: Biogenesis*, edited by E. Quagliariello, E. C. Slator, F. Palmieri, C. Saccone and A. M. Kroon. Elsevier, Amsterdam.

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

Clary, D. O., and D. R. Wolstenholme, 1987 *Drosophila* mitochondrial DNA: conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J. Mol. Evol. **25:** 116–125.

Clayton, D. A., 2000 Transcription and replication of mitochondrial DNA. Hum. Reprod. **2** (Suppl.): 11–17.

De Giorgi, C., A. Martiradonna, C. Lavane and C. Saccone, 1996 Complete sequence of the mitochondrial DNA in the sea urchin *Arbacia lixula*: conserved features of the echinoid mitochondrial genome. J. Biol. Chem. **264:** 10965–10975.

Delport, W., J. W. Ferguson and P. Bloomer, 2002 Characterization and evolution of the mitochondrial DNA control region in hornbills (Bucerotiformes). J. Mol. Evol. **54:** 794–806.

Doda, J. N., C. T. Wright and D. A. Clayton, 1981 Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78:** 6116–6120.

Douris, V., R. A. D. Cameron, G. C. Rodakis and R. Lecanidou, 1998 Mitochondrial phylogeography of the land snail *Albinaria* in Crete: long-term geologic and short-term vicariance effects. Evolution **52:** 116–125.

Elliott, D. J., and H. T. Jacobs, 1989 Mutually exclusive synthetic pathways for sea urchin mitochondrial rRNA and mRNA. Mol. Cell. Biol. **9:** 1069–1082.

Fernandez-Silva, P., P. Loguercio Polosa, M. Roberti, B. Di Ponzio, M. N. Gadaleta *et al.*, 2001 Sea urchin mtDBP is a two-faced transcription termination factor with a biased polarity depending on the RNA polymerase. Nucleic Acids Res. **29:** 4736–4743.

Fisher, R. P., J. N. Topper and D. A. Clayton, 1987 Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. Cell **50:** 247–258.

Garrido-Ramos, M. A., D. T. Stewart, B. W. Sutherland and E. Zouros, 1998 The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels: implications for the operation of doubly uniparental inheritance of mitochondrial DNA. Genome **41:** 818–824.

Goddard, J. M., and D. R. Wolstenholme, 1978 Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **75:** 3886–3890.

Goddard, J. M., and D. R. Wolstenholme, 1980 Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. Nucleic Acids Res. **8:** 741–757.

Goss, P. J. E., and R. C. Lewontin, 1996 Detecting heterogeneity of substitution along DNA and protein sequences. Genetics **143:** 589–602.

Hatzoglou, E., G. C. Rodakis and R. Lecanidou, 1995 Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. Genetics **140:** 1353–1366.

Heath, D. D., P. D. Rawson and T. J. Hilbish, 1995 PCR-based nuclear markers identify alien blue mussel (*Mytilus spp.*) geno-

types on the west coast of Canada. Can. J. Fish. Aquat. Sci. **52:** 2621–2627.

Hoeh, W. R., D. T. Stewart, B. W. Sutherland and E. Zouros, 1996 Cytochrome c oxidase sequence comparisons suggest an unusually high rate of mitochondrial DNA evolution in *Mytilus* (Mollusca: Bivalvia). Mol. Biol. Evol. **13:** 418–421.

Hoeh, W. R., D. T. Stewart, C. Saavedra, B. W. Sutherland and E. Zouros, 1997 Phylogenetic evidence for role-reversals of gender-associated mitochondrial DNA in *Mytilus* (Bivalvia: Mytilidae). Mol. Biol. Evol. **14:** 959–967.

Hoffmann, R. J., J. L. Boore and W. M. Brown, 1992 A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics **131:** 397–412.

Jacobs, H., D. Elliot, V. Math and A. Farguharson, 1988 Nucleotide sequence and gene organization of sea urchin mtDNA. J. Mol. Biol. **201:** 185–217.

Jacobs, H. T., E. R. Herbert and J. Rankine, 1989 Sea urchin egg mitochondrial DNA contains a short displacement loop (D-loop) in the replication origin region. Nucleic Acids Res. **17:** 8949–8965.

Kimura, M., 1980 A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

Kumar, S., K. Tamura, I. B. Jakobsen and M. Nei, 2001 MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:** 1244–1245.

Kumazawa, Y., H. Ota, M. Nishida and T. Ozawa, 1998 The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. Genetics **150:** 313–329.

Ladoukakis, E. D., C. Saavedra, A. Magoulas and E. Zouros, 2002 Mitochondrial DNA variation in a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. Mol. Ecol. **11:** 755–769.

Levinson, G., and G. A. Gutman, 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:** 203–221.

Lewis, D. L., C. L. Farr, A. L. Farquhar and L. S. Kaguni, 1994 Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. Mol. Biol. Evol. **11:** 523–538.

Liu, H. P., J. B. Mitton and S. K. Wu, 1996 Paternal mitochondrial DNA differentiation far exceeds maternal DNA and allozyme differentiation in the freshwater mussel, *Anodonta grandis grandis*. Evolution **50:** 952–957.

Loguercio Polosa, P., M. Roberti, C. Musicco, M. N. Gadaleta, E. Quagliariello *et al.*, 1999 Cloning and characterisation of mtDBP, a DNA-binding protein which binds two distinct regions of sea urchin mitochondrial DNA. Nucleic Acids Res. **27:** 1890–1899.

Marshall, H. D., and A. J. Baker, 1997 Structural conservation and variation in the mitochondrial control region of fringilline finches (*Fringilla* spp.) and the greenfinch (*Carduelis chloris*). Mol. Biol. Evol. **14:** 173–184.

Mayhook, A. G., A. M. Rinaldi and H. T. Jacobs, 1992 Replication origins and pause sites in sea urchin mitochondrial DNA. Proc. R. Soc. Lond. Ser. B **248:** 85–94.

McDonald, J. H., 1998 Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **15:** 377–384.

Miller, S. A., D. D. Dykes and H. F. Polesky, 1988 A simple salting out procedure for extraction DNA from human nucleated cells. Nucleic Acids Res. **16:** 1215.

Okimoto, R., J. L. Macfarlane, D. O. Clary and D. R. Wolstenholme, 1992 The mitochondrial genomes of two nematodes *Caenorhabditis elegans* and *Ascaris suum*. Genetics **130:** 471–498.

Passamonti, M., and V. Scali, 2001 Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca Bivalvia). Curr. Genet. **39:** 117–124.

Pearson, W. R., and D. J. Lipman, 1988 Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:** 2444–2448.

Quesada, H., M. Warren and D. O. F. Skibinski, 1998 Nonneutral evolution and differential mutation rate of gender-associated mitochondrial DNA lineages in the marine mussel *Mytilus*. Genetics **149:** 1511–1526.

Quesada, H., R. Wenne and D. O. F. Skibinski, 1999 Interspecies transfer of female mitochondrial DNA is coupled with role-reversals and departure from neutrality in the mussel *Mytilus trossulus*. Mol. Biol. Evol. **16:** 655–665.

Randi, E., and V. Lucchini, 1998 Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. J. Mol. Evol. **47:** 449–462.

Rawson, P. D., and T. J. Hilbish, 1995 Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. Mol. Biol. Evol. **12:** 893–901.

Rawson, P. D., C. L. Secor and T. J. Hilbish, 1996 The effects of natural hybridization on the regulation of doubly uniparental mtDNA inheritance in blue mussels (*Mytilus* spp.) Genetics **144:** 241–248.

Ray, D. A., and L. Densmore, 2002 The crocodilian mitochondrial control region: general structure, conserved sequences, and evolutionary implications. J. Exp. Zool. **294:** 334–345.

Roberti, M., A. Mustich, M. N. Gadaleta and P. Cantatore, 1991 Identification of two homologous mitochondrial DNA sequences, which bind strongly and specifically to a mitochondrial protein of *Paracentrotus lividus*. Nucleic Acids Res. **19:** 6249–6254.

Roberti, M., P. Loguercio Polosa, C. Musicco, F. Milella, S. A. Qureshi *et al.*, 1999 In vivo mitochondrial DNA-protein interactions in sea urchin eggs and embryos. Curr. Genet. **34:** 449–458.

Roe, B. A., D.-P. Ma, R. K. Wilson and J. F.-H. Wong, 1985 The complete nucleotide sequence of the Xenopus mitochondrial genome. J. Biol. Chem. **260:** 9759–9774.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Saccone, C., M. Attimonelli and E. Sbisà, 1987 Structural elements highly preserved during the evolution of the D-loop containing regions in vertebrate mitochondrial DNA. J. Mol. Evol. **26:** 205–211.

Saccone, C., G. Pesole and E. Sbisà, 1991 The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol. **33:** 83–91.

Saccone, C., C. De Giorgi, C. Gissi, G. Pesole and A. Reyes, 1999 Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene **238:** 195–209.

Saccone, C., C. Gissi, A. Reyes, A. Larizza, E. Sbisà *et al.*, 2002 Mitochondrial DNA in metazoa: degree of freedom in a frozen event. Gene **286:** 3–12.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Sbisà, E., F. Tanzariello, A. Reyes, G. Pesole and C. Saccone, 1997 Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene **205:** 125–140.

Serb, J. M., and C. Lydeard, 2003 Complete mtDNA sequence of the North American freshwater mussel, *Lampsilis ornata* (Unionidae): an examination of the evolution and phylogenetic utility of mitochondrial genome organization in Bivalvia (Mollusca). Mol. Biol. Evol. **20:** 1854–1866.

Shadel, G. S., and D. A. Clayton, 1997 Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. **66:** 409–435.

Skibinski, D. O. F., C. Gallagher and C. M. Beynon, 1994a Mitochondrial DNA inheritance. Nature **368:** 817–818.

Skibinski, D. O. F., C. Gallagher and C. M. Beynon, 1994b Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. Genetics **138:** 801–809.

Spruit, N., C. Delabre, G. Gachelin and V. Laudet, 1998 Complete sequence of the amphioxus (*Branchiostoma lanceolatum*) mitochondrial genome: relations to vertebrates. Nucleic Acids Res. **26:** 3279–3285.

Stewart, D. T., C. Saavedra, R. R. Stanwood, A. O. Ball and E. Zouros, 1995 Male and female mitochondrial lineages in the blue mussel (*Mytilus edulis*) species group. Mol. Biol. Evol. **12:** 735–747.

Stoneking, M., 2000 Hypervariable sites in the mtDNA control region are mutational hotspots. Am. J. Hum. Genet. **67:** 1029–1032.

Suzuki, H., Y. Hosokawa, M. Nishikimi and T. Ozawa, 1991 Existence of common homologous elements in the transcriptional regulatory regions of human nuclear genes and mitochondrial gene for the oxidative phosphorylation system. J. Biol. Chem. **266:** 2333–2338.

Taanman, J. W., 1999 The mitochondrial genome: structure, tran-

scription, translation and replication. Biochim. Biophys. Acta **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 analysis tools. Nucleic Acids Res. **24:** 4876–4882.

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. J. Mol. Evol. **54:** 486–500.

Valverde, J. R., R. Marko and R. Garesse, 1994 A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. Proc. Natl. Acad. USA **91:** 5368–5371.

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

Wolstenholme, D. R., 1992 Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. **141:** 173–216.

Zouros, E., A. O. Ball, C. Saavedra and K. R. Freeman, 1994a Mitochondrial DNA inheritance. Nature **368:** 817–818.

Zouros, E., A. O. Ball, C. Saavedra and K. R. Freeman, 1994b A new type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proc. Natl. Acad. Sci. USA **91:** 7463–7467.

Zuker, M., 2003 Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:** 3406–3415.

Communicating editor: T. H. Eickbush

APPENDIX

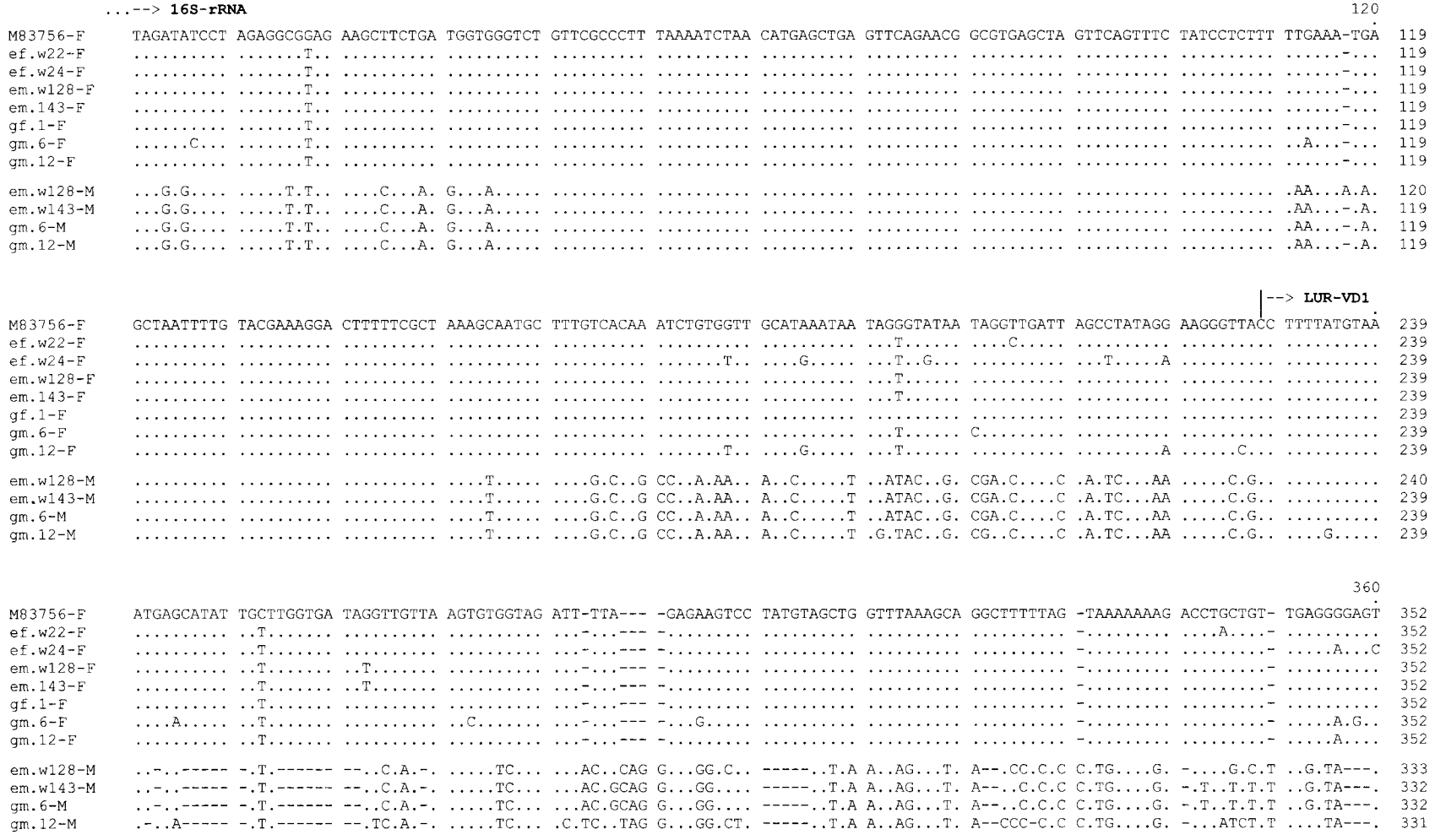

FIGURE A1.—Multiple alignment of the 12 nucleotide sequences used in this study. The boxed sequence in gm.6-F is found in tandem repeat with two changes (T instead of A in the last position and T instead of C in the seventh position from the end). Numbers in parentheses (far right column) indicate the actual length of sequence gm.6-F.

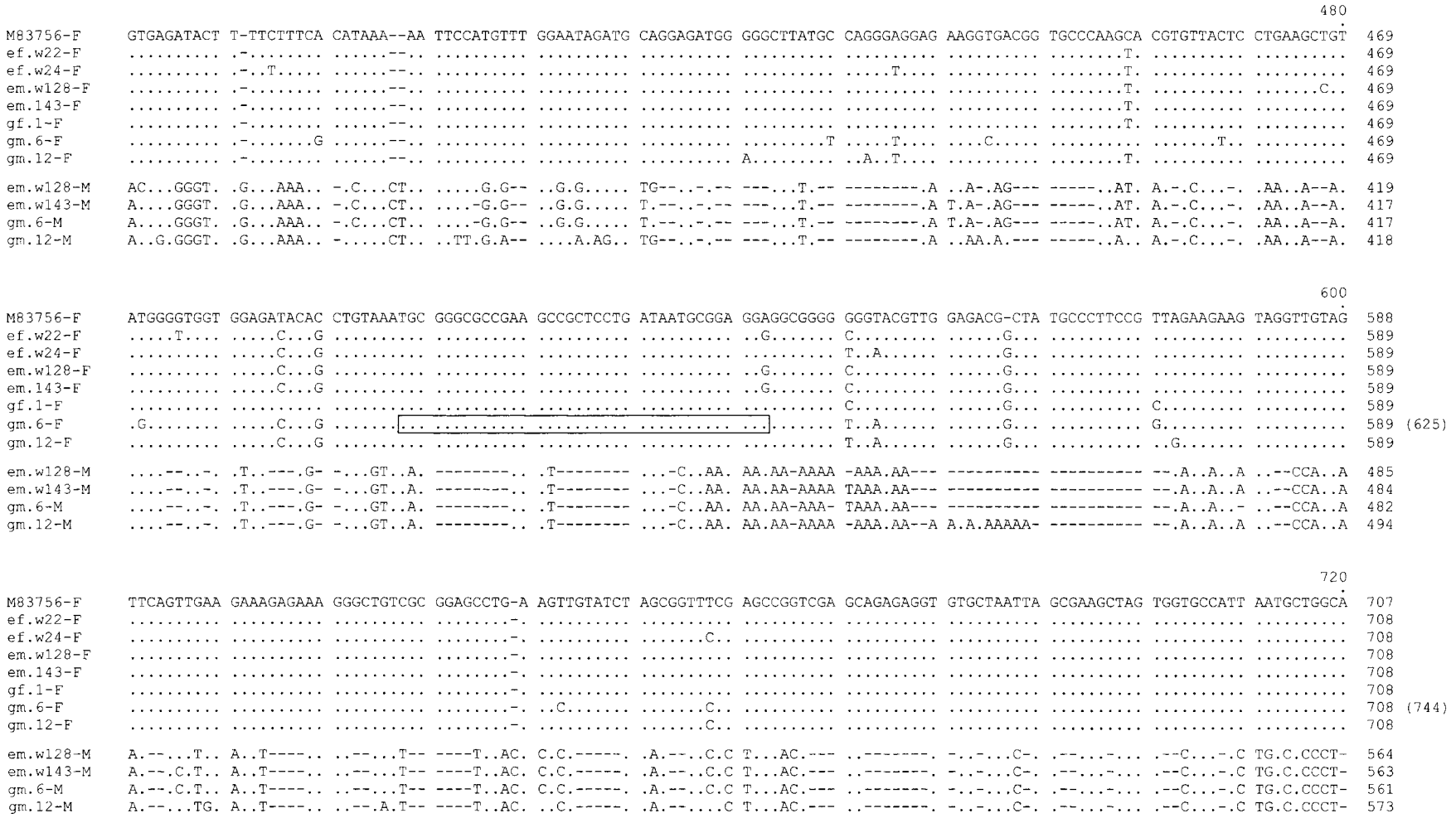

FIGURE A1.—*Continued.*

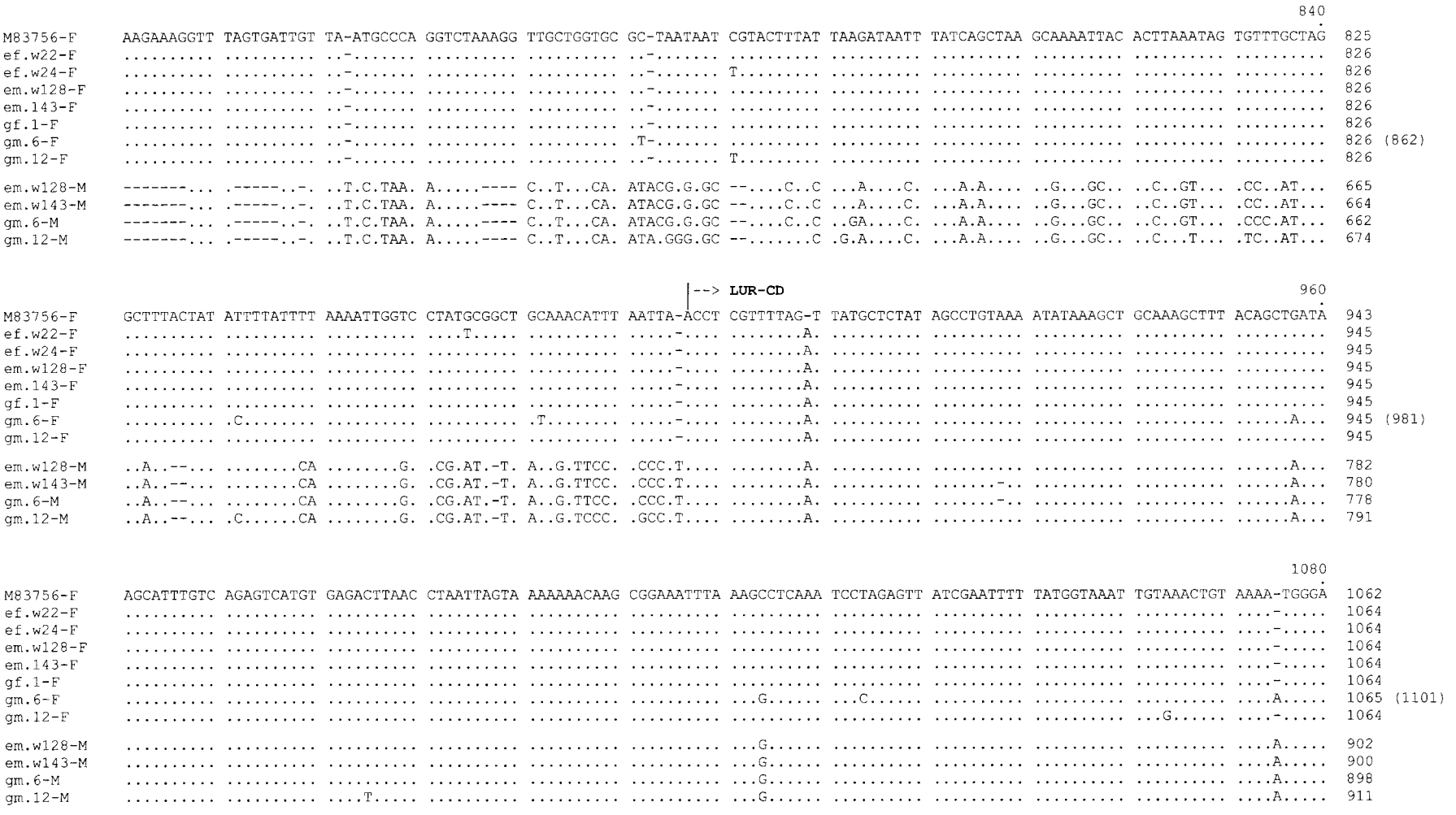

FIGURE A1.—*Continued.*

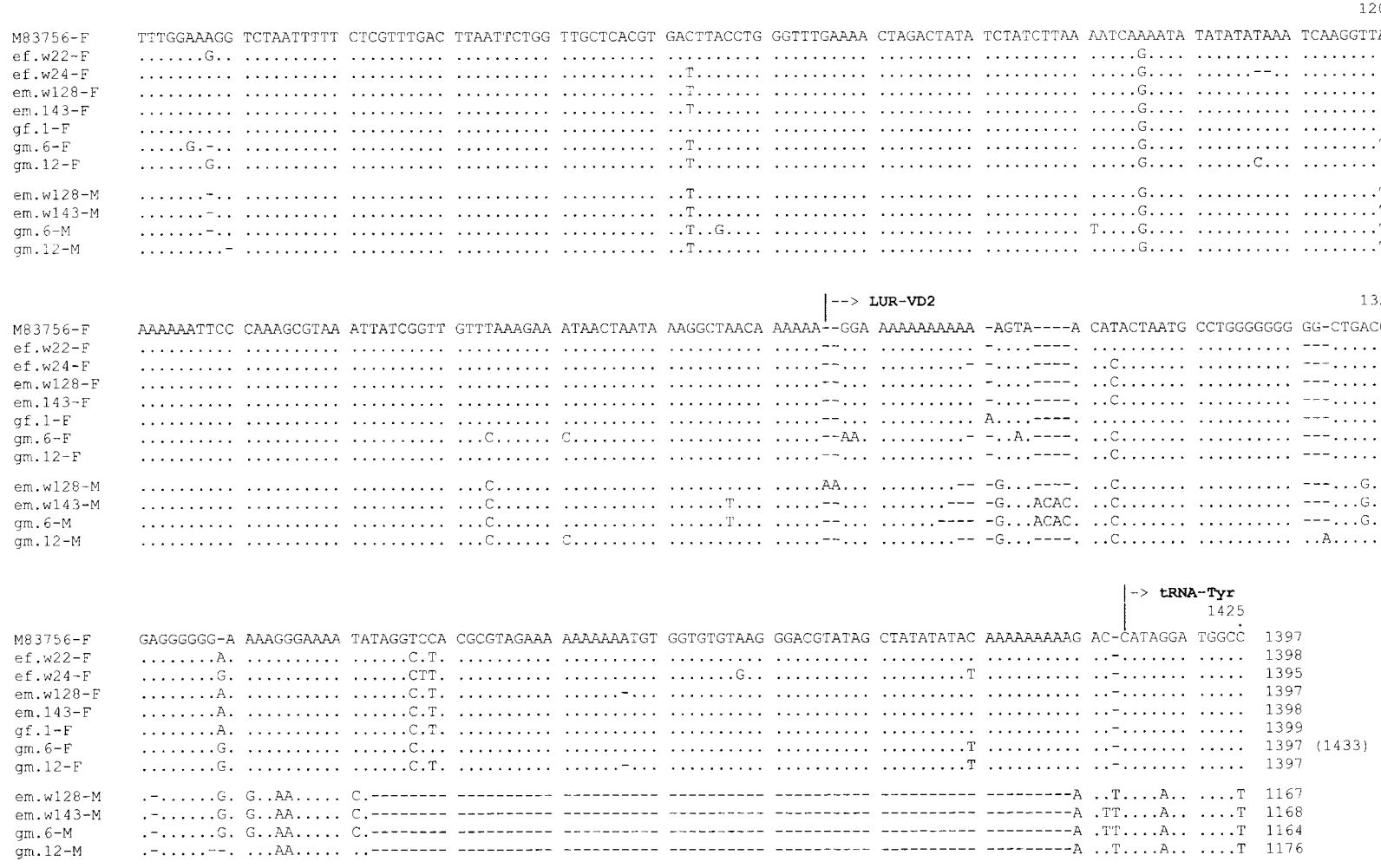

FIGURE A1.—*Continued.*