J Mol Evol (1991) 33:556–568

Journal of
Molecular Evolution

The Complete Nucleotide Sequence of the Mitochondrial DNA of the Fin Whale, *Balaenoptera physalus*

Ulfur Arnason, Anette Gullberg, and Bengt Widegren

Department of Molecular Genetics, University of Lund, The Wallenberg Laboratory, Box 7031, S-220 07 Lund, Sweden

Summary. The composition of the mitochondrial DNA (mtDNA) of the fin whale, *Balaenoptera physalus,* was determined. The length of the molecule is 16,398 bp, and its organization conforms with that of other mammals. The general similarity between the mtDNA of the fin whale and the cow is greater than the similarity between the fin whale and other species (human, mouse, rat) in which the composition of the entire molecule has been described. The D-loop region of the mtDNA of the fin whale is 81% identical to the D-loop of dolphin DNA, and the central portion of the D-loop is similar to the bovine D-loop. The accumulation of transversions and gaps in the 12S and 16S rRNA genes was assessed by comparing the fin whale, cow, and human. The sequence difference between human and the whale and human and the cow was at the same level, indicating that the rate of evolution of the mtDNA rRNA genes is about the same in artiodactyls and cetaceans. In the 12S rRNA gene an accumulation rate of 0.05% per million years places the separation of cetaceans and artiodactyls at about 55 million years ago. The corresponding figure for human and either the whale or the cow is about 80 million years. In the 16S rRNA gene a 0.08% accumulation rate of transversions and gaps per million years yields concurring figures. A comparison between the cytochrome *b* gene of the fin whale and cytochrome *b* sequences in the literature, including dolphin (*Stenella*) sequences, identified the cetaceans as monophyletic and the artiodactyls as their closest relatives. The comparison between the cytochrome *b* sequences of the fin whale and *Stenella* showed that differences in codon positions one or two were frequently associated with a change in another codon position.

Key words: Mitochondrial DNA — Molecular phylogeny — Cetaceans

Introduction

In recent years the mitochondrial DNA (mtDNA) molecule has attracted great interest for determining evolutionary relationships at various levels. Thus, restriction analysis of mtDNA has proved to be very useful in defining relationships between individuals and populations (e.g., Wilson et al. 1989). Sequence comparisons of portions of mtDNA have demonstrated the applicability of such analyses for resolving both primate relationships (Hixson and Brown 1986; Hayasaka et al. 1988; Hasegawa et al. 1990) and different relationships among ungulates (Irwin et al. 1991). In a comprehensive study of the cytochrome *b* gene, Irwin et al. (1991) showed that this particular gene can be used for resolving phylogenetic relationships at levels ranging from those within a genus to those between different mammalian orders, covering a divergence time of about 70 million years (Myr). Also other portions of the mtDNA, notably the 12S and 16S rRNA genes, have been shown to be useful for determining relationships at different taxonomic levels among artiodactyls (Miyamoto and Boyle 1989; Miyamoto et al. 1990; Kraus and Miyamoto 1991).

The characteristics and particulars of the mitochondrial genome of animals have been comprehensively surveyed by Brown (1985). It is conceivable that accumulating sequence data on the mtDNA molecule will become an increasingly valuable tool

---

Offprint requests to: U. Arnason

for elucidating various higher order relationships when information on more species and groups becomes available. Until now, the complete mtDNA sequences of four mammals had been determined. These species are human (Anderson et al. 1981), mouse (Bibb et al. 1981), cow (Anderson et al. 1982), and rat (Gadaleta et al. 1989). We report the sequence for a fifth species, the fin whale, *Balaenoptera physalus,* of the suborder Mysticeti (baleen whales), in this context.

Materials and Methods

An enriched mtDNA fraction was obtained from the liver of a female fin whale (no. 27, 1987), taken west of Iceland. The tissue was collected about 24 h postmortem and the material homogenized, using a Waring blender, in 10 volumes of 0.25 M sucrose, 10 mM Tris-HCl, 2 mM EDTA, pH 7.4. The suspension was centrifuged for 5 min at 2500 rpm rotor in SS34 (Sorvall) and the pelleted material removed. This step was repeated four times. The supernatant was then centrifuged for 15 min at 10,000 rpm, and the pellet containing the mitochondria was collected and suspended in homogenization buffer and the centrifugation step repeated. The pellet was suspended in 50 mM glucose, 10 mM EDTA, 100 mM Tris, pH 7.9, using 0.5 ml buffer for each initial gram of tissue. Fresh SDS solution (1% SDS, 0.2 M NaOH), 1 ml/g tissue, was added and the sample was kept on ice for 5–10 min with occasional agitation. A solution of 3 M potassium-acetate, 1.8 M formic acid, pH 5.6 was added (0.75 ml/g tissue) and the sample held at −70°C until it started to solidify. The sample was centrifuged for 10 min at 5500 rpm. The supernatant was collected and the centrifugation step repeated. The supernatant was extracted with phenol–chloroform. The aqueous phase was collected and a 0.6-volume of isopropanol was added. After precipitation the sample was centrifuged at 10,000 rpm for 15 min. The precipitated DNA was washed twice with 70% ethanol and dried.

The enriched mtDNA was cleaved with various restriction enzymes and hybridized (Southern 1975) to mtDNA of the harbor seal in order to identify suitable restriction enzymes for further analysis. BamHI and HindIII were found to yield fragments of convenient sizes. The BamHI fragments were ≈8000, ≈4500, ≈3000, and ≈1000 bp long, respectively. The fragments were isolated from a preparative agarose gel. The 3000- and 1000-bp fragments were cloned directly in M13mp19. The 8000- and 4500-bp fragments and also the 3000-bp fragment were recleaved with BclI or BglII and the fragments ligated into the BamHI site of M13, using the complementarity of the ends created by the three enzymes. The 4500-bp fragment was also digested with Sau3A and ligated into the same site. The M13 clones obtained covered the whole mtDNA molecule except an ≈800-bp portion, which was amplified and reamplified using asymmetric polymerase chain reaction (PCR) amplification (Gyllensten and Erlich 1988) prior to sequencing.

Sequencing was performed applying the dideoxy termination technique (Sanger 1981) with ^{35}S-adenine. Sequencing was carried out from 31 cloning sites plus the two PCR ends. Version 6:2 of the GCG program package (Deveraux et al. 1984) was used for the DNA analyses and comparisons.

Results and Discussion

The localizations of the HindIII, BamHI, BclI, BglII, and the four Sau3A sites used for sequencing of the fin whale mtDNA molecule are shown in Fig. 1. The direction of sequencing in the various clones is also shown. The sequencing at the different cloning sites was supplemented with sequencing initiated by the use of 28 internal primers. The localizations of the

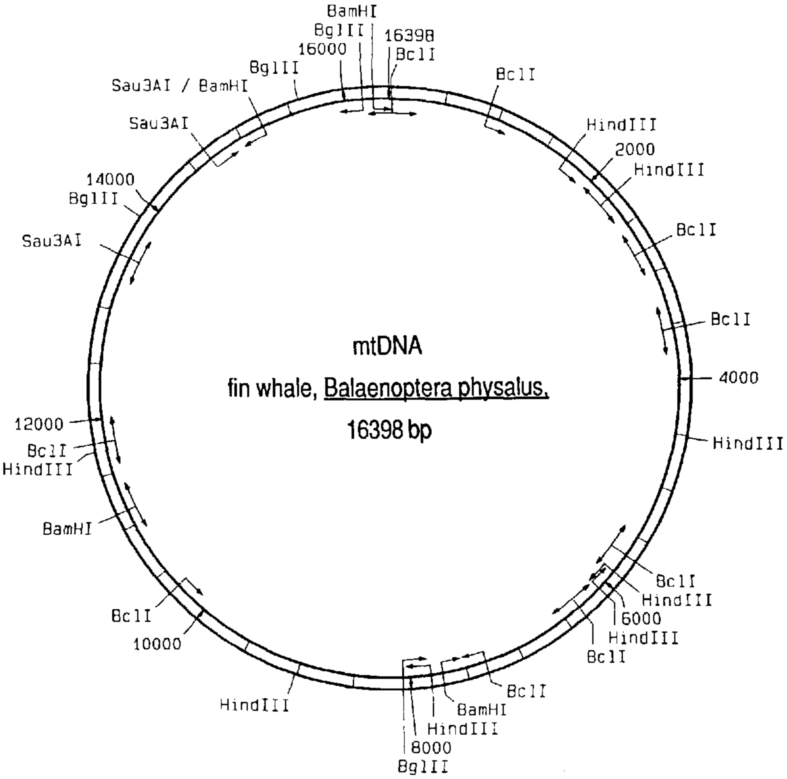


Fig. 1. Localization of the restriction sites BamHI, BclI, BglII, and HindIII in the mtDNA of the fin whale. The position of three Sau3A sites that were used for sequencing is also shown. Direction of sequencing in the clones used is shown by arrows. Sequencing of the region 13,870–14,668 was based upon PCR amplification.

Table 1. Localization of features in the mtDNA molecule of the fin whale

Feature	From bp	To bp
tRNA-Phe	422	494
12S rRNA	496	1471
tRNA-Val	1472	1538
16S rRNA	1539	3112
tRNA-Leu (UUR)	3113	3187
NADH subunit 1	3190	4146
tRNA-Ile	4146	4214
tRNA-Gln	4284	4212 (L)
tRNA-Met	4286	4354
NADH subunit 2	4355	5398
tRNA-Trp	5397	5465
tRNA-Ala	5539	5471 (L)
tRNA-Asn	5614	5541 (L)
Origin L-strain replication	5614	5650
tRNA-Cys	5714	5647 (L)
tRNA-Tyr	5780	5715 (L)
Cytochrome *c* oxidase subunit I	5782	7332
tRNA-Ser (UCN)	7398	7328 (L)
tRNA-Asp	7404	7471
Cytochrome *c* oxidase subunit II	7472	8155
tRNA-Lys	8159	8226
ATPase subunit 8	8228	8419
ATPase subunit 6	8389	9069
Cytochrome *c* oxidase subunit III	9069	9854
tRNA-Gly	9854	9922
NADH subunit 3	9923	10,267
tRNA-Arg	10,270	10,338
NADH subunit 4L	10,339	10,635
NADH subunit 4	10,629	12,005
tRNA-His	12,007	12,075
tRNA-Ser (AGY)	12,076	12,136
tRNA-Leu (CUN)	12,138	12,207
NADH subunit 5	12,208	14,028
NADH subunit 6	14,539	14,012 (L)
tRNA-Glu	14,608	14,540 (L)
Cytochrome *b*	14,613	15,752
tRNA-Thr	15,753	15,824
tRNA-Pro	15,890	15,824 (L)
D-loop	15,891	421

Gene nomenclature is according to Attardi et al. (1986). (L) denotes light strand sense. The length of the molecule is 16,398 bp. Position 1 was determined by analogy with bovine mtDNA (Anderson et al. 1982)

Table 2. Similarity in percent between 12S rRNA, 16S rRNA, and peptide coding genes in mtDNA of the fin whale, cow, and human

Gene	I	II	III	IV
12S rRNA	83 (94.3)	80 (92.1)	80 (92.2)	85 (73)
16S rRNA	84 (91.5)	80 (87.1)	79 (86.6)	75 (64)
NADH1	81	74	74	73
NADH2	77	67	68	71
COI	80	78	80	95
COII	82	73	71	64
ATPase8	72	64	64	78
ATPase6	73	73	73	100
COIII	81	78	79	88
NADH3	79	68	73	71
NADH4	79	72	72	75
NADH4L	80	73	75	77
NADH5	77	72	72	82
NADH6	81	67	66	57
Cytb	80	74	73	75

I: whale–cow; II: whale–human; III: cow–human; IV: difference between I and II expressed as percentage of the mean of the figures for whale–human and cow–human. The bracketed figures in the comparison of the 12S and 16S rRNA genes are based only on differences with respect to transversions and gaps. Abbreviations are defined in the text

various features of the molecule are given in Table 1.

The complete sequence of the L-strand of the fin whale mtDNA is presented in Fig. 2. The length of the molecule is 16,398 bp, and the organization of the molecule is the same as in other mammalian mtDNAs studied. The base composition of the L-strand is A, 32.7%; C, 27.3%; T, 26.7%; and G, 13.3%. Position 1 in the sequence corresponds to position 1 in bovine mtDNA, and genic sequences were defined by analogy with human and bovine mtDNA (Anderson et al. 1981, 1982; Attardi et al. 1986) and the occurrence of stop codons. In Fig. 2, the 13 long open reading frames of mammalian mtDNA (Attardi et al. 1986) are shown translated into the one-letter amino acid code with the letter placed above position 1 of each coding triplet. Termination codons are indicated by an asterisk. Of the 13 reading frames only one, NADH subunit 4 (NADH4), does not have a termination codon. The number of open reading frames with a termination codon is thus considerably larger in the fin whale mtDNA than in human and bovine mtDNA.

The overall homology between the 12S rRNA, 16S rRNA, and peptide coding genes in the mtDNA of the fin whale, cow, and human was compared. The generation length of these species is much longer than that of the two species, the mouse and the rat, that were not included in the comparison. Optimal alignments were obtained by using the best-fit local homology algorithm (Smith and Waterman 1981) of the GCG program package. The comparison does not distinguish between transitions and transversions or differences at different codon positions in the peptide coding genes. The results (Table 2) showed an overall greater identity between the whale and the cow than between the whale and human.

The cytochrome *c* oxidase subunit II (COII) gene shows the highest (82%) and the ATPase subunit 8 (ATPase8) gene the lowest (72%) degree of identity in the comparison between the fin whale and the cow. The figures for human/whale and human/cow are almost the same for each individual gene with the highest degree of identity registered in the COI and COIII genes. The figures on the identity between cow and human agree with those of Brown (1985) who also presented figures including the mouse.

```
      GTTAATTACTAATCAGCCCATGATCATAACATAACTGAGGTTTCATACATTTGGTATTTTTTTATTTTTTTGGGGGGCTTGCACGGACTCCCCTATGACCCTAAAGGGTCTCGTCGCAG
    1 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  120

      TCAGATAAATTGTAGCTGGGCCTGGATGTATTTGTTATTTGACTAGCACAACCAACATGTGCAGTTAAATTAATGGTTACAGGACATAGTACTCCACTATTCCCCCCGGGCTCAAAAAAC
  121 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  240

      TGTATGTCTTAGAGGACCAAACCCCCCTCCTTCCATACAATACTAACCCTCTGCTTAGATATTCACCACCCCCCTAGACAGGCTCGTCCCTAGATTTAAAAGCCATTTTATTTATAAATC
  241 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  360

                                                             D-loop   tRNA-Phe →
      AATACTAAATCTGACACAAGCCCAATAATGAAAATACATGAACGCCATCCCTATCCAATACGTTGATGTAGCTTAAACACTTACAAAGCAAGACACTGAAAATGTCTAGATGGGTCTAGC
  361 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  480

                     12S rRNA →
      CAACCCCATTGACATTAAAGGTTTGGTCCCAGCCTTTCTATTAGTTCTTAACAGACTTACACATGCAAGTATCCACATCCCAGTGAGAACGCCCTCTAAATCATAAAGATTAAAAGGAGC
  481 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  600

      GGGTATCAAGCACGCTAGCACTAGCAGCTCACAACGCCTCGCTTAGCCACGCCCCCACGGGACACAGCAGTGATAAAAATTAAGCTATAAACGAAAGTTCGACTAAGTCATGTTAATTTA
  601 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  720

      AGGGTTGGTAAACTTCGTGCCAGCCACCGCGGTCATACGATCGACCCAAATTAATAGAAGCACGGCGTAAAGAGTGTTAAGGAGCCACATGAAATAAAGTCAAACCTTAATTAAGCTGTA
  721 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  840

      AAAAGCCCTAATTAAAATTAAGCCAAACTACGAAAGTGACTTTAATATAATCTGATCACACGACAGCTAAGATCCAAACTGGGATTAGATACCCCACTATGCTTAGTCGTAAACCCCAAT
  841 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  960

      AGTCACAAAACAAGACTATTCGCCAGAGTACTACTAGCAACAGCCTAAAACTCAAAGGACTTGGCGGTGCCTCATACCCATCTAGAGGAGCCTGTTCTGTAACCGATAAACCCCGATCAA
  961 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  1080

      CCTCACCAACCCTTGCTACTTCAGTCTATATACCGCCATCTTCAGCAAACCCTAAAGGGAGAAAAGTAAGCATAACCATCCTACATAAAAACGTTAGGTCAAGGTGTAACCCATGGGTTG
 1081 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  1200

      GGAAGTAATGGGCTACATTTTCTAAGCTAAGAACATCCCCTATACTCACACGAAAGTTTTTATGAAACTTAAAAACTAAAGGAGGATTTAGTAGTAAATCAAGAGCAGAGTGCTTGATTG
 1201 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  1320

      AATAAGGCCATGAGGGCACGCACACACCGCCCGTCACCCTCCTCAAGTACCCCAGCTATAAACCCCAGTTCGTTAACTCAGGCCAAGCAATTATACGAGAGGAGACAAGTCGTAACAAGG
 1321 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  1440

                                     tRNA-Val →                                                          16S rRNA →
      TAAGCATACCGGAAGGTGTGCTTGGACAAAACAAGATATAGCTTAAACAAAGCATGTAGTTTACACCTAGAAGATTCCACAGCCCGTGTATATCTTGAACTAGCCCTAGCCCACACCCTC
 1441 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  1560

      CCCACCTCTACTACCACAAATCAATCAAATAAAACATTTACCATCCCTTCAAAGTATAGGAGATAGAAATTTAAATATCAGTGGCGCTATAGAGATAGTACCGTAAGGAAAGATGAAAGA
 1561 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  1680

      AAAACCTAAAAGTAATAAAAAGCAAAGCTTACCACTTGTACCTTTTGCATAATGACTTAACTAGTAATAAATTAGCAAAGAGACCTTAAGTTAAATTACCCGAAACCAGACGAGCTACTT
 1681 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  1800

      ATGAGCAGCACCTAGAACGAACTCATCTATGTGGCAAAATAGTGAGAAGACTTATAAGTAGAGGTGAAAAGCCTAACGAGCCTGGTGATAGCTGGTTGTCCCTGAAAAGAATCTCAGTTC
 1801 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  1920

      AACATTAAATAATACTAAAAGCCCATGCCAAGCCTTAACGTATATTTAACTGTTAATCTAAAAAGGTACAGCTTTTTAGAAATGGGTACAACCTTGACTAGAGAGTAAAATCAAACATAA
 1921 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  2040

      ACATAGTTGGCCTAAAAGCAGCCATCAATTAAGAAAGCGTTCAAGCTCGACAACAAAATAATGTTTTAATTCCAACATTAAGTAAATCAACTCCTAGCCTGACTATTGGACTAATCTATA
 2041 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  2160

      CAAATATAGAAGCAATACTGTTAATATGAGTAACAAGAAATTTTTCTCCTAGCACAAGCTTACACCAGTAACTGATAATATACTGATAATTAACAGCAAATAAATAAAACCCAACACTAA
 2161 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  2280

      ATTATTTATTAAAATACTGTTAACCCAACACAGGCGTGCATTAAGGAAAGATTAAAAAAAGTAAAAGGAACTCGGCAAACACAAACCCCGCCTGTTTACCAAAAACATCACCTCTAGCAT
 2281 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  2400

      AACCAGTATTAGAGCACTGCCTGCCCGGTGACTAATCGTTAAACGGCCGCGGTATCCTGACCGTGCAAAGGTAGCATAATCACTTGTTCTCTAATTAGGGACTTGTATGAATGGCCACAC
 2401 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  2520

      GAGGGTTTTACTGTCTCTTACTTTTAATCAGTGAAATTGACCTCTCCGTGAAGAGGCGGAGATAACAAAATAAGACGAGAAGACCCTATGGAGCTTCAATTAATCAACCCAAAAACCATA
 2521 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  2640

      ACCTTAAACCACCAAGGGATAACAAAACCTTATATGGGCTGACAATTTCGGTTGGGGTGACCTCGGAGTACAAAAAACCCTCCGAGTGATTAAAACTTAGGCCCACTAGCCAAAGTACAA
 2641 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  2760

      TATCACTTATTGATCCAATCCTTTGATCAACGGAACAAGTTACCCTAGGGATAACAGCGCAATCCTATTCTAGAGTCCATATCGACAATAGGGTTTACGACCTCGATGTTGGATCAGGAC
 2761 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  2880

      ATCCTAATGGTGCAGCTGCTATTAAGGGTTCGTTTGTTCAACGATTAAAGTCCTACGTGATCTGAGTTCAGACCGGAGTAATCCAGGTCGGTTTCTATCTATTACGCATTTCTCCCAGTA
 2881 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3000

                                                                                                                        tRNA-Leu(UUR)→
      CGAAAGGACAAGAGAAATAAGGCCAACTTCAAACAAGCGCCTTCAAACAATTAATGACCTAGTCTCAACTTAATAATTAAGCGCAAACAAACCTGCCCAAGACCAGGGCCTTGTTGAGGT
 3001 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3120

                                                                          NADH1 →
                                                                          M  F  M  I  N  I  L  T  L  I  L  P  I  L  L  A  V
      GGCAGAGTTCGGTAATTGCATAAAACTTAAACTTTTACACCCAGAGGTTCAAATCCTCTCCCCAACAAAATGTTTATAATTAACATTCTAACACTCATTCTCCCCATCCTCCTAGCCGTA
 3121 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3240

      A  F  L  T  L  V  E  R  K  I  L  G  Y  M  Q  F  R  K  G  P  N  I  V  G  P  H  G  L  L  Q  P  F  A  D  A  I  K  L  F  T
      GCATTCCTAACGCTAGTAGAACGCAAAATTCTAGGCTATATGCAGTTCCGAAAGGGGCCAAACATCGTAGGCCCACATGGCTTACTCCAACCCTTTGCCGATGCAATTAAATTATTCACT
 3241 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3360

      K  E  P  L  R  P  A  T  S  S  T  T  M  F  I  I  A  P  V  L  A  L  T  L  A  L  T  M  W  S  P  L  P  M  P  Y  P  L  I  N
      AAAGAACCCCTACGGCCAGCTACATCCTCAACTACTATGTTTATCATTGCACCAGTACTAGCCCTAACCCTGGCCCTCACTATATGAAGCCCCCTACCCATACCATACCCCCTCATTAAC
 3361 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3480

      M  N  L  G  V  L  F  M  L  A  M  S  S  L  A  V  Y  S  I  L  W  S  G  W  A  S  N  S  K  Y  A  L  I  G  A  L  R  A  V  A
      ATAAACCTAGGAGTATTATTCATATTAGCAATATCCAGCCTAGCCGTCTACTCCATCCTATGATCAGGCTGAGCCTCCAACTCAAAATACGCACTAATTGGAGCCCTACGAGCAGTAGCA
 3481 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3600

      Q  T  I  S  Y  E  V  T  L  A  I  I  L  L  S  V  L  L  M  N  G  S  Y  T  L  S  T  L  A  T  T  Q  E  Q  L  W  L  L  F  P
      CAAACAATCTCATATGAGGTAACACTAGCCATTATCCTCCTATCAGTACTCCTAATAAACGGCTCCTACACCTTATCAACATTAGCCACAACACAAGAACAACTATGATTACTATTCCCA
 3601 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3720

      S  W  P  L  A  M  M  W  F  I  S  T  L  A  E  T  N  R  A  P  F  D  L  T  E  G  E  S  E  L  V  S  G  F  N  V  E  Y  A  A
      TCATGACCCTTAGCCATAATGTGATTCATCTCCACCCTAGCAGAAACTAATCGAGCTCCTTTTGATCTAACAGAGGGAGAATCAGAACTCGTATCAGGCTTCAACGTAGAATATGCAGCA
 3721 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3840

      G  P  F  A  L  F  F  L  A  E  Y  A  N  I  I  M  M  N  M  L  T  A  I  L  F  L  G  T  F  H  N  P  H  N  P  E  L  Y  T  A
      GGCCCTTTCGCCCTATTCTTCCTGGCAGAATACGCCAACATCATTATAATGAATATACTCACAGCCATTTTATTCCTAGGAACATTCCACAACCCTCATAACCCAGAATTGTACACAGCA
 3841 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  3960

      N  L  I  I  K  T  L  L  L  T  M  S  F  L  W  I  R  A  S  Y  P  R  F  R  Y  D  Q  L  M  H  L  L  W  K  N  F  L  P  L  T
      AACCTTATTATCAAGACACTACTACTCACAATATCCTTCCTATGAATTCGAGCATCCTACCCCCGATTCCGATATGACCAACTAATACACCTACTCTGAAAAAATTTCCTTCCCTTAACA
 3961 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  4080

      L  A  L  C  M  W  H  I  S  L  P  I  M  T  A  S  I  P  P  Q  T  * tRNA-Ile →
      CTAGCCCTTTGCATATGACATATCTCATTACCAATCATAACTGCAAGTATCCCCCCTCAAACATAAGAAATATGTCTGATAAAAGAGTTACTTTGATAGAGTAAATAATAGAGGCCCAAA
 4081 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+  4200
```

Fig. 2. The sequence of the mitochondrial DNA molecule of the fin whale. The composition of the L-strand is shown. Position 1 corresponds to position 1 in bovine mtDNA (Anderson et al. 1982). The sense of the different genes is shown by an arrow. Nomenclature of peptide genes is according to Attardi et al. (1986).

```
                                                                              ← tRNA-Gln    tRNA-Met →
      TCCTCTTATTTCTAGAATAATAGGAATCGAACCTACCCTTAAGAATTCAAAGTTCTTCGTGCTACCATATTACACTACAATCTACAGTAAGGTCAGCTAAACAAGCTATCGGGCCCATAC
4201  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 4320
                                        NADH2 →
                                         M  N  P  L  I  L  I  I  L  L  T  T  L  I  L  G  T  M  M  V  V  T  S  S  H  W  L  L  A
      CCCGAAAATGTTGGTTCATATCCTTCCCATACTAATAAACCCATTAATCCTTATTATCCTCCTGACAACCCTTATCCTAGGTACAATAATGGTAGTCACCAGCTCTCACTGACTATTAGC
4321  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 4440

        W  I  G  F  E  M  N  M  M  A  F  I  P  I  M  M  K  N  P  T  P  R  A  T  E  A  S  T  K  Y  L  L  T  Q  A  T  A  S  A  L
      CTGAATTGGCTTCGAAATGAACATAATAGCCTTCATCCCTATCATAATAAAAAATCCTACTCCCCGGGCCACAGAAGCTTCTACCAAGTACCTCCTAACACAAGCCACTGCTTCCGCACT
4441  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 4560

        L  M  M  A  V  I  I  N  L  M  H  S  G  Q  W  T  I  T  K  L  F  N  P  T  A  S  T  L  M  T  V  A  L  A  I  K  L  G  L  A
      CCTCATAATAGCAGTCATCATTAACTTAATGCACTCAGGCCAATGAACTATTACAAAACTATTTAACCCAACAGCATCCACACTCATAACAGTAGCCCTAGCCATCAAACTGGGATTAGC
4561  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 4680

        P  F  H  F  W  V  P  E  V  T  Q  G  I  P  L  T  T  G  L  I  L  L  T  W  Q  K  L  A  P  L  S  I  L  Y  Q  I  S  P  S  I
      CCCCTTCCACTTTTGAGTTCCAGAAGTAACACAAGGTATCCCCCTAACCACAGGCCTAATCCTATTAACATGACAAAAACTAGCTCCCTTATCAATCCTATACCAAATTTCACCATCAAT
4681  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 4800

        N  L  H  L  M  L  I  M  S  L  L  S  I  L  M  G  G  W  G  G  L  N  Q  T  Q  L  R  K  I  M  A  Y  S  S  I  A  H  M  G  W
      TAACCTACACCTAATATTAATCATATCCTTACTTTCCATCTTAATAGGAGGCTGAGGTGGACTAAACCAAACACAACTTCGAAAAATCATAGCCTACTCATCAATTGCCCACATAGGATG
4801  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 4920

        M  T  A  I  L  L  Y  N  P  T  L  T  L  L  N  L  L  I  Y  I  T  M  T  F  T  M  F  M  L  F  I  Q  N  S  T  T  T  T  L  S
      AATAACGGCCATTCTACTATATAATCCAACCCTAACATTACTAAATCTACTAATCTACATCACAATAACCTTCACCATATTTATATTATTTATCCAAAACTCAACTACCACTACATTGTC
4921  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 5040

        L  S  Q  T  W  N  K  M  P  V  I  T  T  L  T  M  L  T  L  L  S  M  G  G  L  P  P  L  S  G  F  M  P  K  W  M  I  I  Q  E
      ACTGTCTCAAACCTGAAATAAAATACCCGTCATCACAACCCTTACCATACTCACTTTACTCTCGATAGGAGGACTCCCACCACTATCGGGGTTTATACCCAAATGAATAATTATTCAAGA
5041  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 5160

        L  T  K  N  D  M  L  I  V  P  T  F  M  A  I  T  A  L  L  N  L  Y  F  Y  M  R  L  T  Y  S  T  A  L  T  L  F  P  S  T  N
      ACTAACAAAAAATGATATACTCATTGTACCAACATTCATAGCCATTACAGCATTACTCAACCTATACTTTTATATACGTCTTACTTACTCCACAGCACTAACACTATTTCCCTCCACAAA
5161  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 5280

        N  M  K  M  K  W  Q  F  N  S  T  K  R  T  P  L  L  P  T  A  I  V  I  S  T  M  L  L  P  L  T  P  M  L  S  I  L  L  *tRNA-Trp →
      TAATATAAAAATAAAATGACAATTCAACTCCACAAAACGAACTCCTCTCCTACCAACAGCAATCGTAATTTCCACTATGCTACTACCCCTCACACCAATACTCTCAATCCTACTATAGGA
5281  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 5400

      GTTTAGGTTAAACCCAGACCAAGAGCCTTCAAAGCCCTAAGCAAGTATAATTTACTTAACTCCTGCCCAATAAGGACTGCAAGACTATATCTTACATCAATTGAATGCAAGTCAAACGCT
5401  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 5520

              ← tRNA-Ala                                                               ← tRNA-Asn         Or of L-strand repl
      TTAATTAAGCTAAATCCTCCACTAGATTGGAGGGATACATCTCCCACGAATTTTTAGTTAACAGCTAAATACCCTAATCAACTGGCTTCAATCTACTTCTCCCGCCGCGAGAAAAAAAAGG
5521  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 5640

                                                                     ← tRNA-Cys
      CGGGAGAAGTCCCGGCAGGATTTGAAGCTGCTTCCTTGAATTTGCAATTCAAAATGATCATTCACCACAGGACTTGGTAAAAAGAGGACTCAACCTCTGTCTTTAGATTTACAGTCTAAT
5641  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 5760
                           COI →
                 ← tRNA-Tyr M  F  M  N  R  W  L  F  S  T  N  H  K  D  I  G  T  L  Y  L  L  F  G  A  W  A  G  M  V  G  T  G  L
      ACCTACTCGGCCATTTTACCTATGTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATCGGCACCCTATACTTATTATTCGGTGCTTGAGCAGGAATAGTAGGCACTGGCCTA
5761  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 5880

       S  L  L  I  R  A  E  L  G  Q  P  G  T  L  I  G  D  D  Q  V  Y  N  V  L  V  T  A  H  A  F  V  M  I  F  F  M  V  M  P  I
      AGCTTATTAATCCGCGCTGAGCTAGGTCAACCTGGCACACTAATCGGAGATGACCAAGTCTACAACGTATTAGTAACAGCCCATGCCTTCGTGATAATTTTCTTCATGGTTATACCTATC
5881  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6000

       M  I  G  G  F  G  N  W  L  V  P  L  M  I  G  A  P  D  M  A  F  P  R  M  N  N  M  S  F  W  L  L  P  P  S  F  L  L  L  M
      ATAATTGGCGGATTCGGAAACTGACTAGTCCCACTAATAATTGGAGCACCTGACATAGCTTTCCCTCGTATAAATAATATAAGCTTCTGACTACTCCCTCCTTCTTTCTTACTGTTAATA
6001  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6120

       A  S  S  M  I  E  A  G  A  G  T  G  W  T  V  Y  P  P  L  A  G  N  L  A  H  A  G  A  S  V  D  L  T  I  F  S  L  H  L  A
      GCATCCTCAATAATCGAAGCCGGTGCAGGCACAGGCTGAACTGTATATCCCCCCTTAGCCGGAAATTTAGCACATGCAGGAGCCTCAGTTGACCTTACCATCTTCTCCCTACACTTAGCC
6121  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6240

       G  V  S  S  I  L  G  A  I  N  F  I  T  T  I  I  N  M  K  P  P  A  M  T  Q  Y  Q  T  P  L  F  V  W  S  V  L  V  T  A  V
      GGCGTATCCTCAATCCTCGGAGCCATCAACTTTATCACAACCATTATCAACATAAAACCACCCGCCATAACCCAGTATCAAACTCCCCTTTTCGTATGATCAGTCCTAGTCACAGCAGTA
6241  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6360

       L  L  L  L  S  L  P  V  L  A  A  G  I  T  M  L  L  T  D  R  N  L  N  T  T  F  F  D  P  A  G  G  G  D  P  I  L  Y  Q  H
      CTACTCCTATTATCACTACCTGTTTTAGCAGCCGGAATTACAATGCTACTTACTGACCGAAACCTAAATACAACTTTCTTCGACCCTGCAGGTGGAGGAGACCCAATCCTGTACCAACAC
6361  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6480

       L  F  W  F  F  G  H  P  E  V  Y  I  L  I  L  P  G  F  G  M  I  S  H  I  V  T  Y  Y  S  G  K  K  E  P  F  G  Y  M  G  M
      CTATTCTGATTCTTTGGCCACCCTGAGGTATATATCCTAATTCTTCCTGGGTTCGGAATAATTTCACACATTGTGACTTATTACTCAGGAAAAAAAGAACCCTTTGGCTACATAGGAATA
6481  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6600

       V  W  A  M  V  S  I  G  F  L  G  F  I  V  W  A  H  H  M  F  T  V  G  M  D  V  D  T  R  A  Y  F  T  S  A  T  M  I  I  A
      GTTTGAGCTATGGTGTCCATTGGGTTTTTAGGCTTTATCGTATGAGCCCATCATATGTTTACAGTAGGTATAGACGTTGATACACGAGCATACTTCACATCCGCCACTATAATTATCGCT
6601  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6720

       I  P  T  G  V  K  V  F  S  W  L  A  T  L  H  G  G  N  I  K  W  S  P  A  L  M  W  A  L  G  F  I  F  L  F  T  V  G  G  L
      ATTCCCACAGGAGTAAAAGTATTCAGTTGATTAGCAACACTACACGGGGGTAACATTAAATGATCTCCTGCCCTAATATGAGCTCTAGGATTCATCTTCCTTTTCACAGTAGGCGGTCTA
6721  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6840

       T  G  I  V  L  A  N  S  S  L  D  I  V  L  H  D  T  Y  Y  V  V  A  H  F  H  Y  V  L  S  M  G  A  V  F  A  I  M  G  G  F
      ACTGGTATCGTCCTAGCCAACTCATCATTAGACATTGTCCTACACGATACTTACTATGTAGTCGCTCACTTCCACTATGTCCTATCAATAGGGGCAGTCTTCGCCATTATAGGAGGCTTT
6841  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 6960

       V  H  W  F  P  L  F  S  G  Y  T  L  N  T  T  W  A  K  I  H  F  M  I  M  F  V  G  V  N  L  T  F  F  P  Q  H  F  L  G  L
      GTCCACTGATTTCCACTATTCTCAGGGTATACACTTAACACAACATGAGCAAAAATTCACTTCATAATCATGTTCGTAGGTGTAAACCTAACATTCTTTCCACAACACTTCTTAGGCCTA
6961  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 7080

       S  G  M  P  R  R  Y  S  D  Y  P  D  A  Y  T  T  W  N  T  I  S  S  M  G  S  F  I  S  L  T  A  V  M  L  M  I  F  I  I  W
      TCCGGCATGCCTCGACGATATTCCGATTACCCAGACGCCTACACAACATGAAATACTATTTCATCTATAGGCTCATTTATCTCACTGACAGCAGTTATACTAATAATCTTCATTATCTGA
7081  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 7200

       E  A  F  T  S  K  R  E  V  L  A  V  D  L  T  S  T  N  L  E  W  L  N  G  C  P  P  P  Y  H  T  F  E  E  P  A  F  V  N  P
      GAAGCATTTACATCCAAACGAGAGGTACTGGCAGTAGACCTCACCTCTACCAACCTTGAATGACTAAATGGGTGTCCTCCACCATACCATACATTCGAAGAACCAGCATTCGTCAACCCA
7201  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 7320

       K  W  S  *                                                  ← tRNA-Ser(UCN)       tRNA-Asp →
      AAGTGATCAAGAAAGGAAGGAATCGAACCCTCTCCCATTGGTTTCAAGCCAACATCATAATTACTATGTCTTTCTTTATAAACGAGATATTAGTAAAACCTTATATAACTTTGTCAAAGT
7321  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 7440
                                     COII →
                                      M  A  Y  P  F  Q  L  G  F  Q  D  A  A  S  P  I  M  E  E  L  L  H  F  H  D  H  T  L  M  I
      TAAGTTACAAGTGAAAATCCTGTATATCTCCATGGCATATCCATTCCAACTAGGTTTCCAAGATGCAGCATCACCCATCATAGAAGAGCTCCTACACTTTCACGATCATACACTAATAAT
7441  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 7560

        V  F  L  I  S  S  L  V  L  Y  I  I  T  L  M  L  T  T  K  L  T  H  T  S  T  M  D  A  Q  E  V  E  T  V  W  T  I  L  P  A
      CGTTTTTCTAATTAGCTCTTTAGTTCTCTACATTATTACCCTAATGCTTACAACCAAATTAACACATACTAGTACAATAGACGCCCAAGAAGTAGAAACTGTCTGAACTATCCTCCCAGC
7561  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 7680

        I  I  L  I  L  I  A  L  P  S  L  R  I  L  Y  M  M  D  E  V  N  N  P  S  L  T  V  K  T  M  G  H  Q  W  Y  W  S  Y  E  Y
      CATTATCTTAATTTTAATTGCCTTGCCTTCATTACGGATCCTTTACATAATAGACGAAGTCAATAACCCCTCCCTCACTGTAAAAACAATAGGTCACCAATGATATTGAAGCTATGAGTA
7681  ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 7800
```

Fig. 2. Continued

```
         T  D  Y  E  D  L  S  F  D  S  Y  M  I  P  T  S  D  L  K  P  G  E  L  R  L  L  E  V  D  N  R  V  V  L  P  M  E  M  T  I
       TACCGACTACGAAGACCTAAGCTTCGACTCCTATATAATCCCAACATCAGACCTAAAGCCAGGAGAACTACGATTATTAGAAGTAGATAACCGAGTTGTCTTACCTATAGAAATAACAAT
  7801 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 7920

         R  M  L  V  S  S  E  D  V  L  H  S  W  A  V  P  S  L  G  L  K  T  D  A  I  P  G  R  L  N  Q  T  T  L  M  S  T  R  P  G
       CCGAATATTAGTCTCATCAGAAGACGTACTCCACTCATGGGCCGTACCCTCCTTGGGCCTAAAAACAGATGCAATCCCAGGACGCCTAAACCAAACAACCTTAATATCAACACGACCAGG
  7921 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 8040

         L  F  Y  G  Q  C  S  E  I  C  G  S  N  H  S  F  M  P  I  V  L  E  L  V  P  L  E  V  F  E  K  W  S  V  S  M  L  *
       CCTATTTTATGGACAATGCTCAGAGATCTGCGGCTCAAACCACAGTTTCATACCAATTGTCCTAGAACTAGTACCCCTAGAAGTCTTTGAAAAATGATCTGTATCAATACTATAACTTCA
  8041 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 8160

                                                                           ATPase8 →
       tRNA-Lys →                                                          M  P  Q  L  D  T  S  M  W  L  L  T  I  L  S  M  L  L
       TTAAGAAGCTAAATTAGCATTAACCTTTTAAGTTAAAGATTGAGAGTCTATAACTCTCCTTAATGATATGCCACAATTAGATACATCAATATGACTCCTTACTATTCTATCAATACTCTT
  8161 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 8280

                                                                                                               ATPase6 →
                                                                                                               M  N  E  N
        T  L  F  V  L  F  Q  L  K  I  S  K  H  S  Y  S  P  N  P  K  L  A  H  T  K  T  Q  K  Q  Q  A  P  W  N  T  T  W  T  K  I
       AACCCTCTTTGTATTATTCCAATTAAAAATCTCAAAGCACTCCTATTCCCCTAACCCCAAACTAGCACATACCAAAACACAAAAACAACAAGCTCCTTGAAACACCACATGAACGAAAAT
  8281 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 8400

        L  F  A  P  F  M  I  P  V  M  L  G  I  P  I  T  T  L  I  I  I  L  P  S  M  L  F  P  A  P  N  R  L  I  N  N  R  T  I  A
         Y  L  P  L  L  *
       TTATTTGCCCCTTTTATAATCCCAGTTATGCTAGGTATCCCTATTACCACTCTAATCATTATCCTCCCATCTATATTATTCCCTGCACCAAATCGACTAATCAACAACCGCACGATCGCC
  8401 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 8520

        I  Q  Q  W  L  T  K  L  T  S  K  Q  L  M  N  V  H  S  P  K  G  Q  T  W  S  L  M  L  I  S  L  F  L  F  I  A  S  T  N  L
       ATCCAACAATGATTAACCAAACTCACATCAAAACAACTAATAAACGTACACAGTCCTAAAGGACAAACTTGATCTCTAATGCTCATCTCACTATTCCTATTTATTGCCTCTACCAATCTC
  8521 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 8640

        L  G  M  L  P  H  S  F  T  P  T  T  Q  L  S  M  N  V  G  M  A  I  P  L  W  A  G  T  V  T  T  G  F  R  N  K  T  K  M  S
       CTTGGAATATTACCTCACTCATTTACACCCACCACACAACTCTCAATAAACGTAGGAATGGCTATCCCCTTATGAGCTGGCACCGTTACCACAGGCTTCCGCAACAAAACAAAAATATCC
  8641 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 8760

        L  A  H  L  L  P  Q  G  T  P  T  F  L  I  P  M  L  V  I  I  E  T  I  S  L  F  I  Q  P  V  A  W  A  V  R  L  T  A  N  I
       TTAGCACATCTACTACCACAGGGCACACCCACTTTTCTTATTCCCATATTAGTAATTATTGAGACTATCAGTCTATTCATTCAACCGGTAGCATGAGCTGTACGGCTAACTGCCAACATC
  8761 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 8880

        T  A  G  H  L  L  M  H  L  I  G  E  T  T  L  A  L  M  N  I  N  L  F  S  A  F  I  T  F  T  I  L  A  L  L  T  I  L  E  F
       ACAGCAGGCCACTTATTAATGCATCTAATTGGAGAAACAACCCTCGCACTAATAAATATTAATCTATTTTCAGCCTTTATCACTTTCACAATCCTTGCTCTATTAACTATCCTTGAATTT
  8881 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9000

                                                                            COIII →
        A  V  A  L  I  Q  A  Y  V  F  T  L  L  V  S  L  Y  L  H  D  N  T  *M  T  H  Q  T  H  S  Y  H  M  V  N  P  S  P  W  P  L
       GCTGTCGCCCTAATTCAAGCTTACGTATTTACCCTTCTAGTAAGCCTATACCTTCATGACAACACATAATGACCCACCAAACCCACTCATACCACATAGTAAACCCCAGCCCTTGACCCC
  9001 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9120

         T  G  A  L  S  A  L  L  M  T  S  G  L  I  M  W  F  H  F  N  S  M  L  L  L  T  L  G  L  S  T  N  I  L  T  M  Y  Q  W  W
       TCACCGGAGCTCTATCAGCCCTTCTAATAACATCAGGCCTAATTATATGATTTCACTTCAACTCAATACTCCTACTAACTCTAGGCTTGTCAACAAATATTTTAACAATATACCAATGAT
  9121 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9240

         R  D  I  I  R  E  S  T  F  Q  G  H  H  T  P  T  V  Q  K  G  L  R  Y  G  M  I  L  F  I  V  S  E  V  L  F  F  T  G  F  F
       GGCGAGATATCATCCGAGAAAGCACCTTCCAAGGCCATCATACACCAACCGTCCAAAAGGGCTTACGATACGGAATAATTCTATTTATCGTCTCAGAAGTCCTATTTTTCACAGGCTTCT
  9241 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9360

         W  A  F  Y  H  S  S  L  A  P  T  P  E  L  G  G  C  W  P  P  T  G  I  R  P  L  N  P  L  E  V  P  L  L  N  T  S  V  L  L
       TCTGAGCCTTCTACCACTCAAGCCTTGCCCCTACTCCAGAACTAGGCGGATGTTGACCACCAACAGGCATCCGCCCTTTAAATCCCTTAGAAGTTCCCCTTCTCAACACCTCCGTATTAC
  9361 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9480

         A  S  G  V  S  I  T  W  A  H  H  S  L  M  E  G  N  R  K  H  M  L  Q  A  L  F  I  T  I  A  L  G  L  Y  F  T  L  L  Q  A
       TAGCCTCTGGCGTATCTATTACCTGAGCCCACCATAGCTTGATAGAAGGAAACCGCAAACACATACTTCAAGCACTCTTCATCACAATTGCACTAGGCCTCTACTTCACCTTACTACAAG
  9481 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9600

         S  E  Y  Y  E  A  P  F  T  I  S  D  G  I  Y  G  S  T  F  F  V  A  T  G  F  H  G  L  H  V  I  I  G  S  T  F  L  I  V  C
       CATCAGAGTACTACGAAGCCCCTTTCACAATCTCAGACGGAATCTACGGCTCCACCTTCTTTGTAGCCACAGGCTTTCATGGGTTACATGTAATCATTGGATCTACTTTCCTTATCGTCT
  9601 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9720

         F  L  R  Q  V  K  F  H  F  T  S  N  H  H  F  G  F  E  R  A  A  W  Y  W  H  F  V  D  V  V  W  L  F  L  Y  V  S  I  Y  W
       GTTTCCTACGTCAAGTAAAATTCCACTTCACATCAAACCACCACTTTGGCTTTGAACGTGCCGCTTGATACTGACATTTCGTAGACGTCGTATGACTATTTCTTTATGTATCTATCTATT
  9721 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9840

                                                                                                   NADH3 →
         W  G  S  *  tRNA-Gly →                                                                    M  N  L  L  L  T  L  L  T  N  T  T  L
       GATGAGGTTCCTAGTCCTTTTAGTATTAACAAGTACAACTGACTTCCAATCAGTTAGTTTCGGTGTACCCCGAAAAAGAACAATAAACCTTCTACTAACACTACTAACAAATACAACACT
  9841 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9960

         A  L  L  L  V  F  I  A  F  W  L  P  Q  L  N  V  Y  A  E  K  T  S  P  Y  E  C  G  F  D  P  M  G  S  A  R  L  P  F  S  M
       AGCCCTACTACTCGTATTCATCGCCTTCTGACTTCCACAACTAAACGTATACGCAGAAAAAACAAGCCCATATGAATGCGGATTTGACCCCATAGGATCAGCCCGCCTACCCTTCTCCAT
  9961 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 10080

         K  F  F  L  V  A  I  T  F  L  L  F  D  L  E  I  A  L  L  L  P  L  P  W  A  I  Q  S  N  N  L  N  T  M  L  T  M  A  L  F
       AAAATTCTTCTTGGTGGCCATTACTTTCCTTCTCTTTGACTTAGAAATCGCTCTCTTACTCCCCCTTCCTTGAGCAATTCAGTCAAACAACCTAAACACAATACTCACAATAGCCTTATT
 10081 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 10200

         L  I  S  L  L  A  A  S  L  A  Y  E  W  T  Q  E  G  L  E  W  A  E     tRNA-Arg →
       CTTAATCTCCCTACTAGCAGCCAGCCTAGCTTATGAATGAACTCAAGAAGGCCTAGAATGAGCTGAATATGGTATTTAGTTTAAGATAAAACAAGTGATTTCGACCCACTAGACTGTGAT
 10201 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 10320

                         NADH4L →
                         M  T  L  I  H  M  N  I  L  M  A  F  S  M  S  L  M  G  L  L  M  Y  R  S  H  L  M  S  A  L  L  C  L  E
       CAAATTCACAATTACCAAATGACCTTAATCCATATAAACATTCTCATAGCCTTCAGCATGTCCCTTATAGGCCTATTAATATATCGATCCCACCTAATATCCGCATTACTCTGCCTAGAA
 10321 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 10440

        G  M  M  L  S  L  F  V  L  A  A  L  T  I  L  S  S  H  F  T  L  A  N  M  M  P  I  I  L  L  V  F  A  A  C  E  A  A  I  G
       GGTATAATACTATCACTTTTCGTCCTAGCAGCCCTTACAATCCTAAGCTCACACTTCACTTTAGCTAACATGATGCCTATTATCCTCCTAGTTTTCGCAGCTTGTGAGGCAGCCATCGGA
 10441 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 10560

                                                                            NADH4 →
                                                                            M  L  K  F  I  I  P  T  I  M  L  M  P  L  T  W  L  S
        L  A  L  L  V  M  V  S  N  T  Y  G  T  D  Y  V  Q  N  L  N  L  L  Q  C  *
       CTAGCCTTGCTAGTTATAGTCTCTAACACATATGGTACCGATTACGTACAAAACCTCAACCTTCTCCAATGCTAAAATTTATTATTCCTACAATCATACTAATACCCCTAACCTGATTAT
 10561 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 10680

         K  N  N  L  I  W  I  N  S  T  A  H  S  L  L  I  S  F  S  S  L  L  L  L  N  Q  L  N  D  N  S  L  N  Y  S  L  M  F  F  S
       CAAAAAATAACCTAATCTGAATTAACTCCACAGCCCACAGTCTATTAATTAGCTTCTCAAGCCTTCTCCTCCTCAATCAACTCAACGACAACAGCCTTAACTACTCACTAATATTCTTCT
 10681 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 10800

         D  P  L  S  T  P  L  L  I  L  T  M  W  L  L  P  L  M  L  M  A  S  Q  S  H  L  I  K  E  P  P  V  R  K  K  L  Y  I  T  M
       CCGACCCCCTTTCTACCCCACTCCTGATCCTAACAATATGACTCCTTCCCTTAATACTAATAGCAAGTCAATCCCATCTCATCAAAGAACCACCAGTCCGAAAAAAACTCTACATTACGA
 10801 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 10920

         L  I  T  L  Q  A  L  L  I  M  T  F  T  A  T  E  L  I  L  F  Y  I  M  F  E  A  T  L  I  P  T  L  I  I  I  T  R  W  G  N
       TACTAATCACACTACAAGCCCTCCTAATTATAACATTTACTGCCACTGAATTAATCCTATTTTATATCATATTTGAAGCCACACTAATCCCTACCCTTATCATTATCACTCGCTGGGGCA
 10921 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 11040

         Q  T  E  R  L  N  A  G  L  Y  F  L  F  Y  T  L  V  G  S  L  P  L  L  V  A  L  V  Y  L  Q  N  T  T  G  S  L  N  F  L  L
       ACCAAACAGAACGACTCAATGCCGGACTATACTTCTTATTCTATACACTAGTTGGATCTCTCCCACTACTAGTAGCACTAGTATATTTACAAAATACAACAGGATCCCTAAACTTTCTAC
 11041 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 11160

         L  Q  H  W  A  Q  P  L  S  T  S  W  S  N  I  F  M  W  L  A  C  M  M  A  F  L  V  K  M  P  L  Y  G  L  Y  L  W  L  P  K
       TCCTACAACACTGAGCTCAACCATTATCTACGTCCTGATCCAACATCTTCATATGACTAGCCTGCATAATAGCCTTCCTAGTAAAAATACCTCTCTATGGACTATACCTTTGACTGCCCA
 11161 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 11280
```

Fig. 2. Continued

```
       A  H  V  E  A  P  I  A  G  S  M  V  L  A  A  V  L  L  K  L  G  G  Y  G  M  L  R  I  T  S  M  L  N  P  L  T  E  H  M  A
      AAGCACACGTAGAAGCCCCCATTGCAGGCTCCATAGTCCTTGCAGCCGTACTACTAAAACTTGGAGGCTATGGCATACTACGAATTACATCCATACTCAATCCCTTAACAGAACACATAG
11281 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 11400

       Y  P  F  L  M  L  S  L  W  G  M  I  M  T  S  S  I  C  L  R  Q  T  D  L  K  S  L  I  A  Y  S  S  V  S  H  M  A  L  V  I
      CATACCCATTTCTTATACTCTCTCTTTGAGGAATAATCATAACCAGCTCTATCTGTTTACGTCAAACAGACCTAAAATCACTAATTGCATATTCCTCAGTTAGTCACATAGCACTCGTCA
11401 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 11520

       A  A  I  L  I  Q  T  P  W  S  Y  M  G  A  T  A  L  M  I  A  H  G  L  T  S  S  M  L  F  C  L  A  N  S  N  Y  E  R  I  H
      TCGCAGCTATCCTCATCCAAACCCCCTGAAGCTATATAGGGGCCACTGCCCTAATAATTGCCCACGGCCTCACATCCTCCATACTATTCTGTTTGGCAAACTCGAACTACGAACGCATTC
11521 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 11640

       S  R  T  M  I  L  P  G  G  L  Q  V  F  L  P  L  M  A  S  W  W  L  L  A  S  L  T  N  L  A  L  P  P  T  I  N  L  I  G  E
      ATAGCCGAACCATAATTCTGCCCGGAGGCCTACAAGTCTTTCTACCACTAATAGCCAGTTGATGACTATTAGCAAGCTTAACAAATCTTGCACTACCCCCAACCATCAACCTAATCGGAG
11641 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 11760

       L  L  V  V  M  S  V  F  S  W  S  N  P  T  I  L  L  M  G  T  N  I  V  I  T  A  L  Y  S  L  Y  M  L  I  M  T  Q  R  G  K
      AACTACTCGTAGTCATGTCGGTCTTCTCATGATCAAATCCCACTATTCTCCTAATAGGAACAAATATTGTAATTACTGCTCTCTACTCTCTATATATACTAATCATAACACAACGTGGCA
11761 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 11880

       H  T  H  H  I  N  N  I  T  P  S  F  T  R  E  H  A  L  M  A  L  H  I  I  P  L  L  L  L  S  L  N  P  K  I  I  L  G  P  L
      AACACACACACCACATCAATAATATCACCCCTTCCTTCACACGAGAGCATGCCTTAATAGCCCTACACATTATTCCCCTCTTGCTCCTATCACTAAACCCTAAAATCATCTTAGGCCCTC
11881 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12000

       Y  tRNA-His →                                                         tRNA-Ser(AGY) →
      TCTACTGTAAGTATAGTTTAAAAAAGACGCTAGTTTGTGAAACTAACAATAGAAGATCAAAACTTCTTACTTACCGAAAAAGTACTGCAAGAACTGCTAATTCATGCTCCCACACCTAAC
12001 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12120
                                                                                                NADH5 →
                           tRNA-Leu(CUN) →                                                      M  N  L  F  T  S  F  T  L  L  T
      AGCTGTGGCTTTTTCAAACTTTTACAGGATAGTAGTTATCCATTGGTCTTAGGAGCCAAAAAATTGGTGCAACTCCAAATAAAAGTAATAAACCTATTTACCTCTTTCACCCTACTTACA
12121 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12240

       L  L  I  L  T  T  P  I  M  M  S  H  T  G  S  H  V  N  N  K  Y  Q  S  Y  V  K  N  I  V  F  C  A  F  I  T  S  L  V  P  A
      CTACTAATTCTGACCACCCCCATTATAATATCACATACAGGTTCCCATGTAAACAACAAATATCAATCATATGTAAAAAACATTGTCTTCTGCGCCTTCATCACTAGTCTAGTTCCCGCA
12241 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12360

      M  V  Y  L  H  T  N  Q  E  T  L  I  S  N  W  H  W  I  T  I  Q  T  L  K  L  T  L  S  F  K  M  D  Y  F  S  L  M  F  M  P
      ATAGTATATCTTCACACAAACCAAGAAACACTCATCTCAAACTGACACTGAATCACAATCCAAACCCTCAAACTAACACTTAGCTTTAAAATAGATTACTTTTCACTTATATTTATACCA
12361 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12480

      V  A  L  F  I  T  W  S  I  M  E  F  S  M  W  Y  M  H  S  D  P  Y  I  N  Q  F  F  K  Y  L  L  L  F  L  I  T  M  L  I  L
      GTAGCACTATTCATTACATGATCCATCATAGAATTCTCAATATGATATATGCACTCCGACCCCTACATCAACCAATTTTTTAAATACTTACTCCTCTTCCTCATCACCATACTAATCCTT
12481 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12600

      V  T  A  N  N  L  F  Q  L  F  I  G  W  E  G  V  G  I  M  S  F  L  L  I  G  W  W  F  G  R  T  D  A  N  T  A  A  L  Q  A
      GTTACAGCTAACAATCTCTTCCAACTTTTCATCGGATGAGAAGGAGTAGGAATTATATCCTTCTTACTAATTGGCTGATGATTCGGACGAACAGATGCAAATACAGCCGCCCTCCAAGCA
12601 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12720

      I  L  Y  N  R  I  G  D  I  G  L  L  A  S  M  A  W  F  L  S  N  M  N  T  W  D  L  E  Q  I  F  M  L  N  Q  N  P  L  N  F
      ATCCTATACAATCGTATCGGAGACATTGGACTCCTTGCATCAATAGCATGATTTCTCTCTAATATAAACACATGAGACCTAGAACAAATCTTTATACTCAACCAAAACCCCTTAAATTTC
12721 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12840

      P  L  M  G  L  V  L  A  A  A  G  K  S  A  Q  F  G  L  H  P  W  L  P  S  A  M  E  G  P  T  P  V  S  A  L  L  H  S  S  T
      CCCCTCATAGGACTCGTACTAGCCGCAGCAGGAAAATCGGCTCAATTCGGACTCCACCCTTGACTCCCATCAGCAATAGAAGGTCCTACCCCAGTCTCAGCCCTACTCCACTCAAGCACA
12841 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 12960

      M  V  V  A  G  I  F  L  L  V  R  F  Y  P  L  M  E  N  N  K  L  I  Q  T  V  T  L  C  L  G  A  I  T  T  L  F  T  A  I  C
      ATAGTTGTAGCAGGAATCTTCTTGCTTGTCCGCTTCTACCCATTAATAGAAAATAACAAGCTAATCCAAACAGTAACCCTCTGCTTAGGCGCTATCACAACTCTATTTACAGCCATCTGT
12961 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 13080

      A  L  T  Q  N  D  I  K  K  I  I  A  F  S  T  S  S  Q  L  G  L  M  M  V  T  I  G  L  N  Q  P  Y  L  A  F  L  H  I  C  T
      GCCCTCACCCAAAACGACATCAAAAAAATTATTGCTTTCTCCACCTCCAGCCAGCTAGGCCTAATAATAGTAACAATCGGCCTTAACCAACCTTACCTAGCATTCCTACACATTTGCACA
13081 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 13200

      H  A  F  F  K  A  M  L  F  L  C  S  G  S  I  I  H  N  L  N  N  E  Q  D  I  R  K  M  G  G  L  F  K  A  L  P  F  T  T  T
      CACGCCTTCTTTAAAGCTATACTATTCCTATGTTCTGGCTCCATCATCCATAACCTAAACAACGAACAAGATATCCGAAAAATAGGAGGGCTATTTAAGGCCCTCCCATTCACCACAACC
13201 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 13320

      A  L  I  I  G  C  L  A  L  T  G  M  P  F  L  T  G  F  Y  S  K  D  P  I  I  E  A  A  T  S  S  Y  T  N  A  W  A  L  L  L
      GCCCTTATCATCGGATGTCTTGCACTAACAGGAATGCCATTCCTGACCGGATTCTACTCCAAAGATCCCATTATTGAAGCCGCCACTTCGTCTTATACCAACGCCTGAGCCCTATTACTG
13321 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 13440

      T  L  I  A  T  S  L  T  A  V  Y  S  T  R  I  I  F  F  A  L  L  G  Q  P  R  F  P  P  S  T  T  I  N  E  N  N  P  L  L  I
      ACCTTAATCGCCACCTCCCTTACGGCCGTCTATAGCACCCGCATCATTTTCTTTGCACTACTAGGACAACCCCGCTTCCCTCCCTCCACAACCATTAACGAAAATAATCCACTGTTAATC
13441 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 13560

      N  P  I  K  R  L  L  V  G  S  I  F  A  G  F  I  L  S  N  S  I  P  P  M  T  T  P  L  M  T  M  P  L  H  L  K  L  T  A  L
      AACCCTATCAAACGACTACTCGTCGGAAGTATCTTCGCTGGCTTCATCCTATCCAACAGTATTCCCCCAATAACTACACCTTTAATAACCATACCCCTGCACTTAAAATTAACCGCCCTT
13561 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 13680

      A  M  T  T  L  G  F  I  I  A  F  E  I  N  L  D  T  Q  N  L  K  H  K  H  P  S  N  S  F  K  F  S  T  L  L  G  Y  F  P  T
      GCAATAACAACCCTAGGCTTCATCATCGCATTCGAAATTAACCTTGACACACAAAATCTAAAGCACAAGCACCCATCAAACTCCTTTAAATTCTCCACCTTACTAGGTTATTTCCCCACA
13681 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 13800

      I  M  H  R  L  P  P  H  L  D  L  L  M  S  Q  K  L  A  T  S  L  L  D  L  T  W  L  E  T  I  L  P  K  T  T  A  L  I  Q  L
      ATCATACATCGCCTACCCCCTCACCTTGACCTGTTAATAAGCCAAAAACTAGCAACTTCCCTACTAGATCTAACTTGACTAGAAACTATTTTACCAAAAACCACAGCCCTTATCCAACTA
13801 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 13920

                                                                                                  *  S  G  R  T  I  E  M  I
      K  A  S  T  L  T  S  N  Q  Q  G  L  I  K  L  Y  F  L  S  F  L  I  T  I  T  L  S  M  I  L  F  N  Y  P  E  *
      AAAGCCTCTACACTAACCTCTAACCAACAAGGCCTCATCAAACTCTACTTCTTATCTTTCCTCATCACCATCACCCTCAGCATAATCTTATTTAACTACCCCGAGTAATCTCCATAATAA
13921 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 14040

      I  V  V  S  I  F  L  S  W  G  T  V  I  V  L  W  V  G  Y  S  Y  L  A  A  I  G  T  A  E  E  S  F  V  G  S  G  G  T  D  Y
      TTACAACACTAATAAATAAAGACCAACCCGTAACAATCACCAACCAAACACCATAACTATATAATGCCGCAATCCCTGTAGCCTCCTCACTAAAAACCCCAGAACCCCCAGTATCATAAA
14041 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 14160

      V  V  W  D  G  L  G  D  F  E  F  M  I  K  V  E  G  S  K  L  A  Y  I  V  I  L  F  E  V  V  L  G  L  V  F  A  G  L  V  V
      CAACCCAGTCCCCTAGTCCATCAAACTCAAACATAATCTTCACCTCCCCACTCTTCAAAGCATAAATCACAATTAAAAACTCCACCACCAACCCTAAAACAAATGCTCCTAGTACAACTT
14161 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 14280

      K  N  S  V  W  V  E  P  Y  Q  E  T  A  M  A  T  T  Y  G  F  V  V  L  M  G  G  L  Y  I  L  F  V  M  L  G  L  F  S  G  G
      TATTAGAAACCCAAACCTCAGGATACTGTTCAGTAGCCATAGCTGTTGTATAACCAAATACTACCAGCATTCCCCCCAAATAAATCAAAAACACCATTAACCCCAAAAACGAACCACCAA
14281 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 14400

      F  S  L  I  V  G  C  G  V  G  G  G  V  I  L  G  L  G  G  Y  I  P  S  P  K  S  S  V  G  V  F  S  I  V  F  I  I  S  L  I
      AACTCAAAATAACTCCACATCCAACACCACCACCCACAATCAACCCTAAACCCCCATAAATAGGTGAAGGCTTTGAAGAAACCCCCACAAAACTAATTACAAAAATAATACTTAAAATGA
14401 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 14520
            ← NADH6                                                                             Cyt b →
      F  V  I  Y  M  M  M                                                       ← tRNA-Glu      M  T  N  I  R  K  T  H  P  L
      AAACAATATACATTATCATTATTCTCACATGGACTTCAACCATGACCAATGACATGAAAAATCATCGTTGTTATTCAACTACAAGAACACCAATGACCAACATCCGAAAAACACACCCAC
14521 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 14640

       M  K  I  V  N  D  A  F  V  D  L  P  T  P  S  N  I  S  S  W  W  N  F  G  S  L  L  G  L  C  L  I  M  Q  I  L  T  G  L  F
      TAATAAAAATCGTCAACGACGCATTCGTCGATCTCCCCACCCCATCAAATATCTCTTCATGATGGAACTTCGGCTCCCTACTCGGCCTCTGCTTAATTATACAAATCCTAACAGGCCTAT
14641 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 14760

       L  A  M  H  Y  T  P  D  T  T  T  A  F  S  S  V  T  H  I  C  R  D  V  N  Y  G  W  I  I  R  Y  L  H  A  N  G  A  S  M  F
      TCCTAGCAATACACTACACACCAGACACAACAACCGCCTTCTCATCAGTCACACACATCTGCCGAGACGTGAATTACGGCTGAATTATCCGATACCTACATGCAAATGGGGCTTCTATAT
14761 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 14880
```

Fig. 2. Continued

```
       F  I  C  L  Y  A  H  M  G  R  G  L  Y  Y  G  S  Y  A  F  R  E  T  W  N  I  G  V  I  L  L  F  T  V  M  A  T  A  F  V  G
      TCTTCATCTGCCTCTACGCTCACATAGGACGAGGCCTATACTACGGCTCCTACGCCTTCCGAGAAACATGAAATATTGGAGTTATTCTACTATTCACAGTTATAGCCACCGCATTCGTAG
14881 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 15000

       Y  V  L  P  W  G  Q  M  S  F  W  G  A  T  V  I  T  N  L  L  S  A  I  P  Y  I  G  T  T  L  V  E  W  I  W  G  G  F  S  V
      GCTACGTCCTGCCCTGAGGACAAATATCATTCTGAGGCGCAACTGTAATCACTAACCTCCTATCAGCAATCCCATACATTGGTACCACCCTAGTCGAATGAATCTGAGGCGGTTTCTCTG
15001 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 15120

       D  K  A  T  L  T  R  F  F  A  F  H  F  I  L  P  F  I  I  L  A  L  A  I  V  H  L  I  F  L  H  E  T  G  S  N  N  P  T  G
      TAGATAAAGCAACACTAACACGCTTTTTTGCCTTTCACTTTATCCTCCCCTTCATCATCCTAGCATTAGCAATTGTCCACCTTATTTTCCTTCACGAAACAGGATCCAACAACCCCACAG
15121 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 15240

       I  P  S  D  M  D  K  I  P  F  H  P  Y  H  T  I  K  D  I  L  G  A  L  L  L  I  L  I  L  L  M  L  T  L  F  A  P  D  L  L
      GCATCCCATCCGACATAGATAAAATCCCATTCCACCCCTACCACACAATTAAAGACATTCTAGGTGCCCTATTACTAATCCTAATCCTACTAATACTAACCCTATTCGCACCCGACCTAC
15241 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 1536

       G  D  P  D  N  Y  T  P  A  N  P  L  S  T  P  A  H  I  K  P  E  W  Y  F  L  F  A  Y  A  I  L  R  S  I  P  N  K  L  G  G
      TTGGAGACCCAGACAACTATACCCCAGCAAACCCACTCAGTACCCCAGCACACATTAAACCAGAATGGTATTTTCTATTCGCATACGCAATCCTACGATCAATCCCCAACAAACTAGGCG
15361 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 15480

       V  L  A  L  L  L  S  I  L  I  L  A  F  I  P  M  L  H  T  S  N  Q  R  S  M  M  F  R  P  F  S  Q  F  L  F  W  V  L  V  A
      GAGTCTTAGCCCTACTACTCTCAATCCTAATCCTAGCCTTCATCCCAATACTCCACACATCCAATCAACGAAGCATAATATTTCGACCCTTTAGCCAGTTCTTGTTCTGAGTCCTAGTCG
15481 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 15600

       D  L  L  T  L  T  W  I  G  G  Q  P  V  E  H  P  Y  M  I  V  G  Q  L  A  S  I  L  Y  F  L  L  I  L  V  L  M  P  V  T  S
      CAGATCTACTAACCCTAACATGGATCGGCGGCCAACCAGTAGAACACCCCTACATAATTGTAGGCCAACTCGCATCCATCCTCTATTTCCTCTTAATTCTAGTATTAATACCAGTAACTA
15601 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 15720

       L  I  E  N  K  L  M  K  W  *  tRNA-Thr →
      GTCTTATCGAGAACAAACTTATAAAATGAAGAGTCTTTGTAGTATAATTAAATACCCCGGTTTTGTAAACCGGAAAAGGAGACAAGACACACCTCCCTAAGACTCAAGGAAGAAGTATTA
15721 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 15840

                                        ← tRNA-Pro  D-loop
      CACTCCACCATCAGCACCCAAAGCTGAAGTTCTACATAAACTATTCCCTGAAAAAGTATATTGTACAATAACCACAGGACCACAGTACTATGTCCGTATTGAAAATAACTTGCCTTATTA
15841 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 15960

      GATATTATTATGTAACTCGTGCATGCATGTACTTCCACATAATTAATAGCGTCTTTCCATGGGTATGAACAGATATACATGCTATGTATAATTGTGCATTCAATTATTTTCACCACGAGC
15961 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 16080

      AGTTGAAGCTCGTATTAAATTTTATTAATTTTACATATTACATAATATGTATTAATAGTACAATAGCGCATGTTCTTATGCATCCCCAGATCTATTTAAATCAAATGATTCCTATGGCCG
16081 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 16200

      CTCCATTAGATCACGAGCTTAGTCAGCATGCCGCGTGAAACCAGCAACCCGCTTGGCAGGGATCCCTCTTCTCGCACCGGGCCCATCACTCGTGGGGGTAGCTATTTAATGATCTTTATA
16201 ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 16320

      AGACATCTGGTTCTTACTTCAGGACCATATTAACTTAAAATCGCCCACTCGTTCCCCTTAAATAAGACATCTCGATGG
16321 ---------+---------+---------+---------+---------+---------+---------+-------- 16398
```

Fig. 2. Continued

The gene for ATPase6 shows the same degree of identity in the three combinations whale/cow, whale/human, and cow/human. This indicates that this gene has become saturated by the divergence time of the whale from the cow. The difference between the whale and the cow, expressed as percent of the mean of the human/whale and human/cow figures, is shown in Table 2, column IV. Besides ATPase6 (100%) the highest figures were registered in COI (95%), COIII (88%), and NADH5 (82%). It is likely that all these sequences have become largely saturated by the divergence time of the whale from the cow. Among the peptide coding genes the three lowest figures were obtained for NADH6 (57%), COII (64%), and NADH2 (71%). Of these three genes COII shows the highest degree of identity between the whale and the cow. The figure 64% for the difference in COII between the whale and the cow relative to the difference between human and either the whale or the cow is the same as the corresponding figure for the accumulation of transversions and gaps in the 16S rRNA gene. The differences between the whale, cow, and human in the two rRNA genes appears to correspond reasonably well to the evolutionary distances involving the three species. The agreement, however, between the COII gene and the 16S rRNA gene is presumably fortuitous and due to fast evolution of the primate COII gene (Ramharack and Deeley 1987, Wilson et al. 1985).

The NADH6 gene exhibits the lowest figure (57%) in column IV in Table 2. This might be interpreted to mean that the resolution in the NADH6 gene is higher than in the COII gene. This is not so. The first ⅓ of the NADH6 gene is similar in the whale, the cow, and human, but the conformity is limited in the remaining portion of the gene. At the peptide level the identity between the NADH6 gene of the whale and human is only 56% and the similarity, based upon the evolutionary distance between amino acids (Gribskov and Burgess 1986), is 79%.

The differences in rRNA and peptide coding genes are approximately the same between human and either the whale or the cow, indicating a similar rate of evolution in the artiodactyl and cetacean lineages. Hasegawa and Kishino (1989) reported that the rate of evolution in bovine mtDNA is slower than in human mtDNA. If this is so it is conceivable that the rate of evolution in cetacean mtDNA is also slower than that of human mtDNA.

Miyamoto and Boyle (1989) have reported that conservative changes accumulate linearly with time in artiodactyl mitochondrial rRNA genes and adjacent tRNA genes. In their materials they calculated a constant divergence rate of 0.2% per Myr for a span of at least 75 Myr. The approach was based upon registering transversion differences and gaps. Each transversion mismatch was counted as 1.0 as was also each gap irrespective of its length. Kraus and Miyamoto (1991) have extended the analysis by comparing the mtDNA rRNA genes and adjacent tRNA genes in pecoran ruminants. The rate of divergence for transversion calculated by Kraus and Miyamoto (1991) was 0.14% per Myr.

The present comparison of the 12S and 16S rRNA

Table 3. Sequence differences between two portions of the mtDNA 16S rRNA gene of the fin whale (B.p.), Commerson dolphin (C.c), cow (B.t.), and human (H.s.)

	B.p.	C.c.	B.t.	H.s.
B.p.	—	8.0	16.5	23.4
C.c.	2.7	—	15.9	24.2
B.t.	7.2	5.3	—	24.5
H.s.	11.8	10.9	11.0	—

The values of the two sequences have been combined. The dolphin sequences (Southern et al. 1988) correspond to regions 2103–2441 and 2513–2827, respectively, in the mtDNA of the fin whale. The figures above the diagonal show total difference (transitions, transversions, gaps). The figures below the diagonal show differences limited to transversions and gaps

Table 4. Comparison between the cytochrome *b* gene of the fin whale (B.p.), dolphin (Ste.), cow (B.t.), camel (C.d.), pig (S.s.), rhinoceros (D.b.), and human (H.s.)

	B.p.	Ste.	B.t.	C.d.	S.s.	D.b.	H.s.
B.p.	—	83.2	80.2	77.9	81.0	80.2	74.1
Ste.	90.8	—	78.6	77.0	79.5	79.5	73.2
	95.0						
B.t.	88.1	86.5	—	79.1	80.4	81.2	73.1
	94.7	92.9					
C.d.	86.8	85.8	88.9	—	79.0	79.5	71.6
	93.7	92.9	95.5				
S.s.	87.1	85.5	90.0	89.5	—	81.1	75.1
	95.0	93.4	96.3	96.0			
D.b.	86.0	85.8	88.7	87.3	87.9	—	75.5
	92.6	91.6	93.7	93.4	93.7		
H.s.	77.6	77.8	78.9	77.8	77.9	80.7	—
	87.1	87.3	87.3	87.3	87.9	88.4	

The figures above the diagonal line show percent identity at the DNA level. Corresponding figures for the peptide level are shown below the diagonal. The figures for both percent identity (above) and similarity (below) are presented

genes of the fin whale, cow, and human included both a comparison based upon overall identity as well as a comparison of transversion and gap differences between the three species (Table 2, column IV, bracketed figures). The difference between human and either the whale or the cow is highly similar in the different genes. The rate of evolution of the rRNA genes of the artiodactyl and cetacean representatives thus appears to be the same. In both the 12S and 16S sequences the percent difference between human and either the fin whale or the cow was about 1.5 times greater than the difference between the whale and the cow. This indicates that within each gene, transversions (and gaps) accumulate at an approximately constant rate, although the rate is considerably higher in the 16S rRNA gene than in the 12S rRNA gene.

The 12S rRNA genes of the whale and the cow differ by 5.7% with respect to transversions and gaps. An accumulation rate of 0.05% per Myr places the artiodactyl/cetacean divergence at $5.7 \div (0.05 \times 2) = 57$ Myr ago. The corresponding figure for human and either the whale or the cow is about 80 Myr. In the 16S rRNA gene an accumulation rate of 0.08% for transversions and gaps per Myr gives similar figures for both the artiodactyl/cetacean separation and the distance between human and either the whale or the cow.

The figures 0.05% and 0.08%, respectively, for the 12S and 16S rRNA genes conform with the figures obtained by Kraus and Miyamoto (1991) as the rate of divergence, 0.14% per Myr, calculated by them corresponds to an accumulation rate of 0.07% for transversions per lineage and Myr.

The limited information on cetacean mitochondrial rRNA genes does not permit a detailed analysis within the order Cetacea. Southern et al. (1988) reported two portions of the 16S rRNA of the piebald dolphin, *Cephalorhynchus commersonii.* Table 3 shows the results of a comparison based upon the two portions and corresponding portions of the mtDNA of the fin whale, cow, and human. The evolutionary rate, 0.08% per Myr, for the 16S rRNA gene in its entirety conforms reasonably well with the evolutionary separation of the more distantly related lineages (whale, cow, human), but the transversion and gap difference between the whale and the dolphin is unexpectedly limited as it indicates a separation of the two lineages less than 20 Myr ago. The total difference between the whale and the dolphin is 8%. Also this figure is low compared with the results of Miyamoto and Boyle (1989) who found in artiodactyls that during the first 7 Myr of evolution the divergence rate approximates 0.8% per Myr.

In their comparison of the mitochondrial cytochrome *b* gene, Irwin et al. (1991) found that the similarities between the cetacean representatives of genus *Stenella,* of the odontocete family Delphinidae, and of the artiodactyls were greater than the similarities between *Stenella* and other ungulates included. The comparison was based upon conservative changes, i.e., changes in codon positions 1 and 2 and transversions in position 3. A comparison at the DNA level between the fin whale and some selected sequences from Irwin et al. (1991) gives similar values for different whale/artiodactyl relationships (Table 4). A comparison at the peptide level yields similar figures for the relationship between either the fin whale or *Stenella* and the cow, camel, or pig. Table 4 shows the figures for both percent identity and percent similarity based upon the evolutionary distance between the amino acids (Gribskov and Burgess 1986). A noteworthy feature of the percent similarity values is the limited difference between the fin whale and any of the species of dolphin, pig, camel, or cow. This indicates that changes in amino acid composition have been predominantly limited to residues with similar properties.

The differences in percent identity at the peptide level between the species included in Table 4 were analyzed by applying the neighbor-joining method of Saitou and Nei (1987). The tree obtained by the comparison is shown in Fig. 3a. The cytochrome *b* gene of the fin whale was also included in a parallel comparison based upon conservative changes in some of the cytochrome *b* sequences presented by Irwin et al. (1991). The shortest parsimony tree of this comparison (645 steps), obtained by using version 3.0g of the PAUP program (Swofford 1989), is shown in Fig. 3b. The topology of the two trees is the same. The results including the fin whale concur with the findings of Irwin et al. (1991), showing closer relationships between *Stenella* and artiodactyls than between the dolphin and any other ungulates. The findings conform with protein data, which group cetaceans with the ungulates, preferably the artiodactyls (de Jong and Goodman 1982; Goodman 1989; Czelusniak et al. 1990; Sarich 1991), but molecular and anatomical data do not provide entirely concurring evidence in this respect (Novacek 1989).

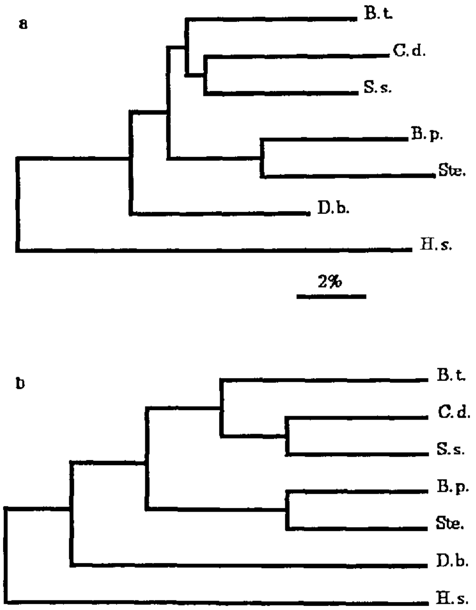
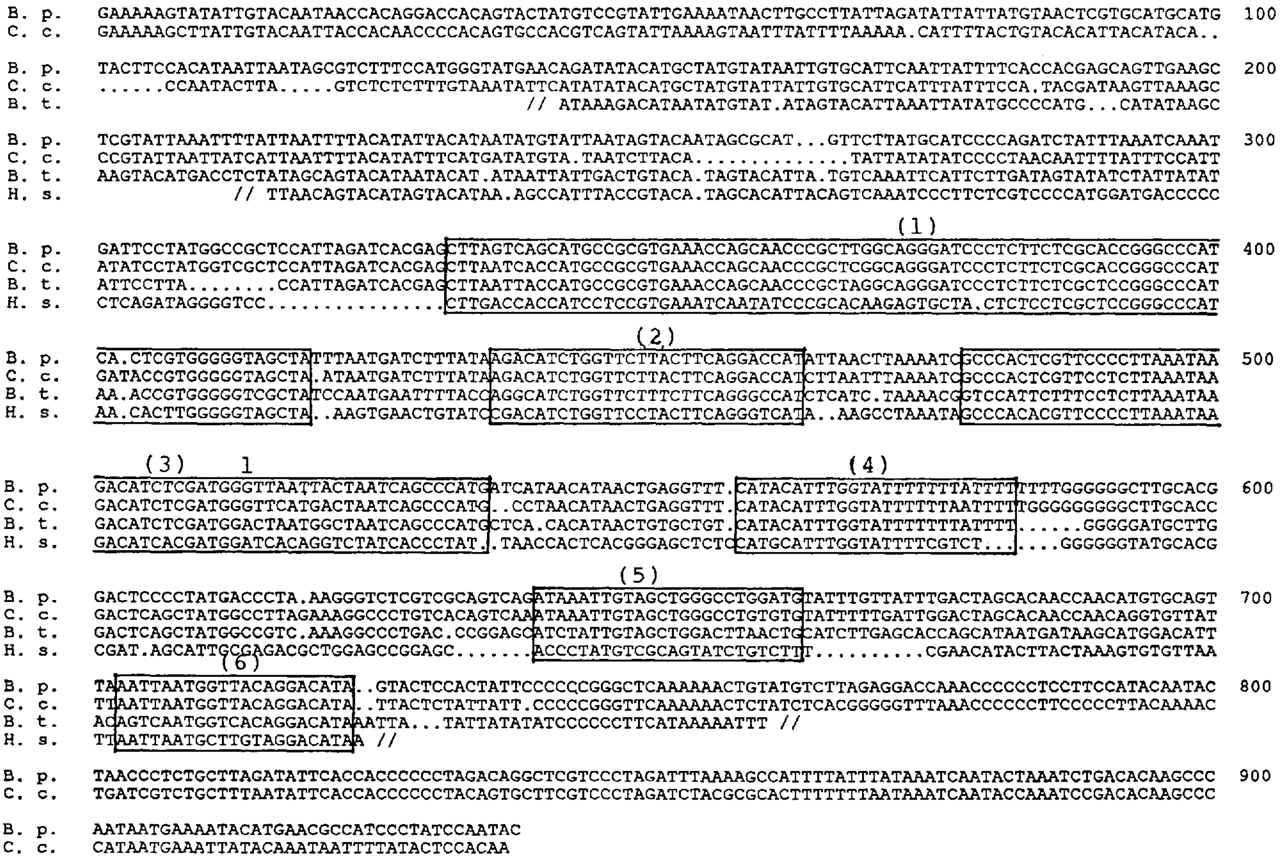


Fig. 3. Neighbor-joining tree of percent amino acid difference (a) and shortest parsimony tree (b) based upon the cytochrome *b* gene. The topology of the two trees is the same. B.t.: cow; C.d.: camel; S.s.: pig; B.p.: fin whale; Ste.: *Stenella* dolphin; D.d.: rhinoceros; H.s.: human. The bar in a represents a difference of 2%.

Table 5. Differences with respect to codon positions in the cytochrome *b* gene (1140 bp long) of the fin whale and *Stenella*

	Number of substitutions						
	1		2		3		Replacement substitutions
Type of substitutions	Ti	Tv	Ti	Tv	Ti	Tv	
	4						
		4					
Single replacements			2				17
				2			
						5	
	5				5		
	4		4				
	1					1	
		3		3			
Multiple replacements		1	1				18
		1			1		
		1				1	
	1		1		1		
	1		1			1	
Silent differences					66	26	
Total differences excluding leucine	16	10	9	5	73	34	35
	26		14		107		
Leucine	16				6	15	

At the DNA level the two species differ in 184 positions. Comparison at the peptide level shows 35 differences: 17 of the differences concur with single replacement substitutions; 8 of these (4 transitions and 4 transversions) occur in codon position 1; 4 (2 transitions and 2 transversions) are in codon position 2; and 5 (all transversions) are in codon position 3. The 18 remaining differences coincide with multiple replacement substitutions at two or three (two cases) codon positions. Ti: transitions; Tv: transversions



```
B. p.    GAAAAAGTATATTGTACAATAACCACAGGACCACAGTACTATGTCCGTATTGAAAATAACTTGCCTTATTAGATATTATTATGTAACTCGTGCATGCATG  100
C. c.    GAAAAAGCTTATTGTACAATTACCACAACCCCACAGTGCCACGTCAGTATTAAAAGTAATTTATTTTAAAAA.CATTTTACTGTACACATTACATACA..

B. p.    TACTTCCACATAATTAATAGCGTCTTTCCATGGGTATGAACAGATATACATGCTATGTATAATTGTGCATTCAATTATTTTCACCACGAGCAGTTGAAGC  200
C. c.    ......CCAATACTTA.....GTCTCTCTTTGTAAATATTCATATATACATGCTATGTATTATTGTGCATTCATTTATTTCCA.TACGATAAGTTAAAGC
B. t.                                          // ATAAAGACATAATATGTAT.ATAGTACATTAAATTATATGCCCCATG...CATATAAGC

B. p.    TCGTATTAAATTTTATTAATTTTACATATTACATAATATGTATTAATAGTACAATAGCGCAT...GTTCTTATGCATCCCCAGATCTATTTAAATCAAAT  300
C. c.    CCGTATTAATTATCATTAATTTTACATATTTCATGATATGTA.TAATCTTACA..............TATTATATATCCCCTAACAATTTTATTTCCATT
B. t.    AAGTACATGACCTCTATAGCAGTACATAATACAT.ATAATTATTGACTGTACA.TAGTACATTA.TGTCAAATTCATTCTTGATAGTATATCTATTATAT
H. s.                // TTAACAGTACATAGTACATAA.AGCCATTTACCGTACA.TAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCC

                                                                                    (1)
B. p.    GATTCCTATGGCCGCTCCATTAGATCACGAGCTTAGTCAGCATGCCGCGTGAAACCAGCAACCCGCTTGGCAGGGATCCCTCTTCTCGCACCGGGCCCAT  400
C. c.    ATATCCTATGGTCGCTCCATTAGATCACGAGCTTAATCACCATGCCGCGTGAAACCAGCAACCCGCTCGGCAGGGATCCCTCTTCTCGCACCGGGCCCAT
B. t.    ATTCCTTA........CCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTAGGCAGGGATCCCTCTTCTCGCTCCGGGCCCAT
H. s.    CTCAGATAGGGGTCC................CTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTA.CTCTCCTCGCTCCGGGCCCAT

                                                          (2)
B. p.    CA.CTCGTGGGGGTAGCTATTTAATGATCTTTATAAGACATCTGGTTCTTACTTCAGGACCATATTAACTTAAAATCGCCCACTCGTTCCCCTTAAATAA  500
C. c.    GATACCGTGGGGGTAGCTA.ATAATGATCTTTATAAGACATCTGGTTCTTACTTCAGGACCATCTTAATTTAAAATCGCCCACTCGTTCCTCTTAAATAA
B. t.    AA.ACCGTGGGGGTCGCTATCCAATGAATTTTACCAGGCATCTGGTTCTTTCTTCAGGGCCATCTCATC.TAAAACGGTCCATTCTTTCCTCTTAAATAA
H. s.    AA.CACTTGGGGGTAGCTA..AAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATA..AAGCCTAAATAGCCCACACGTTCCCCTTAAATAA

              (3)    1                                                      (4)
B. p.    GACATCTCGATGGGTTAATTACTAATCAGCCCATGATCATAACATAACTGAGGTTT.CATACATTTGGTATTTTTTTATTTTTTTTGGGGGGCTTGCACG  600
C. c.    GACATCTCGATGGGTTCATGACTAATCAGCCCATG..CCTAACATAACTGAGGTTT.CATACATTTGGTATTTTTTAATTTTTGGGGGGGGGCTTGCACC
B. t.    GACATCTCGATGGACTAATGGCTAATCAGCCCATGCTCA.CACATAACTGTGCTGT.CATACATTTGGTATTTTTTTATTTT......GGGGGATGCTTG
H. s.    GACATCACGATGGATCACAGGTCTATCACCCTAT..TAACCACTCACGGGAGCTCTCCATGCATTTGGTATTTTCGTCT.......GGGGGGTATGCACG

                                                          (5)
B. p.    GACTCCCCTATGACCCTA.AAGGGTCTCGTCGCAGTCAGATAAATTGTAGCTGGGCCTGGATGTATTTGTTATTTGACTAGCACAACCAACATGTGCAGT  700
C. c.    GACTCAGCTATGGCCTTAGAAAGGCCCTGTCACAGTCAAATAAATTGTAGCTGGGCCTGTGTGTATTTTTGATTGGACTAGCACAACCAACAGGTGTTAT
B. t.    GACTCAGCTATGGCCGTC.AAAGGCCCTGAC.CCGGAGCATCTATTGTAGCTGGACTTAACTGCATCTTGAGCACCAGCATAATGATAAGCATGGACATT
H. s.    CGAT.AGCATTGCGAGACGCTGGAGCCGGAGC.......ACCCTATGTCGCAGTATCTGTCTTT..........CGAACATACTTACTAAAGTGTGTTAA
                   (6)
B. p.    TAAATTAATGGTTACAGGACATA..GTACTCCACTATTCCCCCGGGCTCAAAAAACTGTATGTCTTAGAGGACCAAACCCCCCTCCTTCCATACAATAC  800
C. c.    TTAATTAATGGTTACAGGACATA..TTACTCTATTATT.CCCCCGGGTTCAAAAAACTCTATCTCACGGGGGTTTAAACCCCCCTTCCCCCTTACAAAAC
B. t.    ACAGTCAATGGTCACAGGACATAAATTA...TATTATATATCCCCCCTTCATAAAAATTT //
H. s.    TTAATTAATGCTTGTAGGACATAA //

B. p.    TAACCCTCTGCTTAGATATTCACCACCCCCCTAGACAGGCTCGTCCCTAGATTTAAAAGCCATTTTATTTATAAATCAATACTAAATCTGACACAAGCCC  900
C. c.    TGATCGTCTGCTTTAATATTCACCACCCCCCTACAGTGCTTCGTCCCTAGATCTACGCGCACTTTTTTTAATAAATCAATACCAAATCCGACACAAGCCC

B. p.    AATAATGAAAATACATGAACGCCATCCCTATCCAATAC
C. c.    CATAATGAAATTATACAAATAATTTTATACTCCACAA
```

Fig. 4. Alignment of the D-loop regions of the fin whale (B.p.) and the piebald dolphin (C.c.). The average conformity between the two species is 81%. Truncated sequences of the cow (B.t.) and human (H.s.) are also included. The boxed areas show a high degree of conformity between the two cetaceans and the cow. The conformity with human is less conspicuous. In the large box (1) the conformity between either the whale, the dolphin, or the cow is 91–94% and the conformity with human is ≈77%. In the second largest box (3), which includes position 1 of the molecule, the conformity between the cetaceans is 93%, between the cetaceans and the cow 86%, and between the cetaceans and human 74%.

Irwin et al. (1991) included the cytochrome *b* sequences of three dolphin specimens of the genus *Stenella* in their comparisons. Two of the specimens were spinner dolphins, *S. longirostris,* and one bridled dolphin, *S. attenuata.* We have used the consensus sequence of the three specimens in a comparison with the fin whale in which particular attention was paid to codons that differed between the fin whale and *Stenella* (Table 5). The two sequences (1140 bp) differ at 184 positions. The nucleotide differences concur with 35 differences at the peptide level. Of the 35 differences, 17 are concomitant with single replacement substitutions. The 18 remaining differences coincide with a difference at two or three (two cases) codon positions. In position 1 there are 26 differences between the fin whale and *Stenella,* 8 of which (4 transitions and 4 transversions) constitute single replacement substitutions. Of the 14 differences in position 2, 4 (2 transitions and 2 transversions) constitute single replacement substitutions. The remaining 10 differences concur with a difference in position 1. Thus, out of a total of 40 differences in codon positions 1 and 2, only 12 constitute single differences. A survey of conservative differences (transitions or transversions in codon positions 1 or 2 and transversions in position 3) between the fin whale and *Stenella* shows that conservative differences frequently occur together in nonsilent changes. However, as is evident from comparing Fig. 3a and 3b, the occurrence of double (or triple) substitutions within a codon does not result in DNA and protein data being at variance with each other in phylogenetic comparisons.

The D-loop of the fin whale is 928 bp long. The D-loop of the fin whale and that of *C. commersonii* have an average identity of 81%. However, in the central portion of the D-loop the agreement is greater. Figure 4 shows an alignment of the D-loop of the two species. Truncated portions of the D-loop of the cow and human are also included. The boxed areas comprise regions that show the highest degree of conformity between the fin whale and the dolphin. In most instances the conformity also extends to the sequences of cow and human. In box 1 the agreement between the fin whale and the dolphin is 93%, between the whale and the cow 91%, and be-

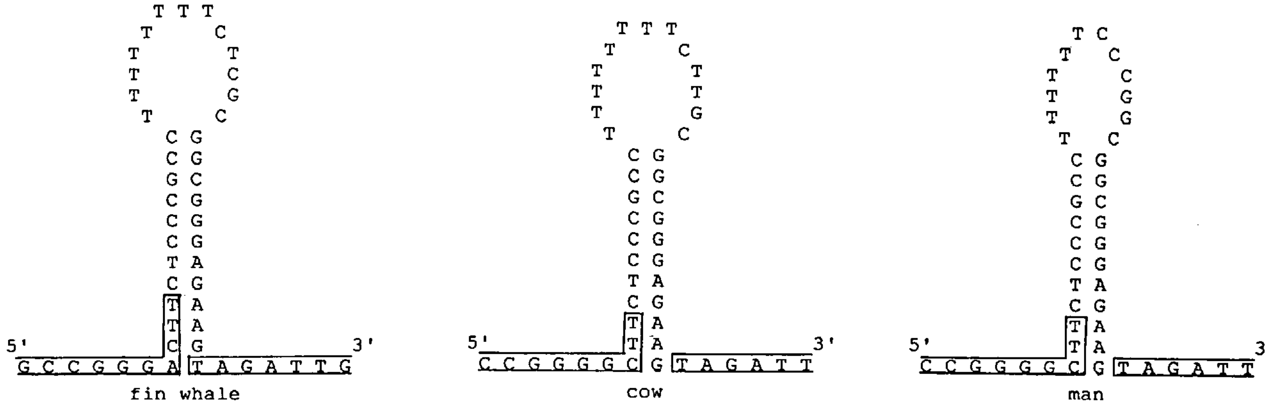


Fig. 5. L-strand origin of replication of the fin whale, cow, and human. The fin whale sequence was identified by analogy with that of the cow (Anderson et al. 1982). The intrastrand complementarity is shown as a secondary structure. The complementarity in the stem sequence of the whale is 1 bp longer than in the cow. In the loop the two species differ at one position. The flanking sequences of tRNA-Cys and tRNA-Asn are shown boxed.

tween the dolphin and the cow 94%. The conformity between any of the three sequences and the human sequence was ≈77%. Box 2 is identical in the whale and the dolphin, and the agreement with the cow and human sequences is 89%. Box 3 includes position 1 of the mtDNA molecule. In this region the homology between the whale and the dolphin is 93%, between either of these and the cow 86%, and between the cetaceans and human 74%. Boxes 4, 5, and 6 are very similar (96–100%) in the whale and the dolphin but differ considerably from the corresponding regions of cow and human.

The features of the mitochondrial L-strand origin of replication are highly conserved in mammals (Anderson et al. 1982). The composition of this region in the fin whale mtDNA together with that of the cow and human is shown in Fig. 5. With the exception of one position in the loop, the conformity between the fin whale and the cow is complete in the entire region. The stems in the three species are identical, but the intrastrand complementarity in the stem region of the fin whale extends 1 bp further than in either cow or human.

Acknowledgments. We thank Drs. David M. Irwin and Allan C. Wilson for various suggestions and for generously providing us with Fig. 3. We also thank Drs. Marc W. Allard and Michael M. Miyamoto for valuable comments on the manuscript. The financial support provided by the Erik Philip-Sörensen Foundation, the Carl Trygger Foundation, and The Swedish Natural Sciences Research Council is gratefully acknowledged. The nucleotide sequence reported in this paper has been submitted to the EMBL Data Bank with accession number X61145.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IC (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717

Attardi G, Chomyn A, Doolittle RF, Mariottini P, Ragan CI (1986) Seven unidentified reading frames of human mitochondrial DNA encode subunits of the respiratory chain NADH dehydrogenase. Cold Spring Harbor Symp Quant Biol LI:103–114

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180

Brown WM (1985) The mitochondrial genome of animals. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 95–130

Czelusniak J, Goodman M, Koop BF, Tagle DA, Shoshani J, Braunitzer G, Kleinschmidt TK, de Jong WW, Matsuda G (1990) Perspectives from amino acid and nucleotide sequences on cladistic relationships among higher taxa of eutheria. In: Genoways HH (ed) Current mammalogy, Vol 2. Plenum, New York, pp 545–572

de Jong WW, Goodman M (1982) Mammalian phylogeny studied by sequence analysis of the eye lens protein α-crystallin. Z Säugetierkd 47:257–276

Deveraux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sibisa E, Saccone C (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 28:497–516

Goodman M (1989) Emerging alliance of phylogenetic systematics and molecular biology: a new age of exploration. In: Fernholm B, Bremer B, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 43–61

Gribskov M, Burgess RR (1986) Sigma factors from *E. coli, B. subtilis,* phage SPO1, and phage T4 are homologous proteins. Nucleic Acids Res 14(16):6745–6763

Gyllensten UB, Ehrlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA–DQA* locus. Proc Natl Acad Sci USA 85:7652–7656

Hasegawa M, Kishino H (1989) Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. Jpn J Genet 64:243–258

Hasegawa M, Kishino H, Hayasaka K, Horai S (1990) Mitochondrial DNA evolution in primates: transition rate has been extremely low in the lemur. J Mol Evol 31:113–121

Hayasaka K, Gojobori T, Horai S (1988) Molecular phylogeny and evolution of primate mitochondrial DNA. Mol Biol Evol 5:626–644

Hixson JE, Brown WM (1986) A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution and phylogenetic implications. Mol Biol Evol 3:1–18

Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome b gene of mammals. J Mol Evol 32:128–144

Kraus F, Miyamoto MM (1991) Rapid cladogenesis among the pecoran ruminants: evidence from mitochondrial DNA sequences. Syst Zool 40:117–130

Miyamoto MM, Boyle SM (1989) The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny. In: Fernholm B, Bremer B, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 437–450

Miyamoto MM, Kraus F, Ryder OA (1990) Phylogeny and evolution of antlered deer determined from mitochondrial DNA sequences. Proc Natl Acad Sci USA 87:6127–6131

Novacek MJ (1989) Higher mammal phylogeny: the morphological–molecular synthesis. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 421–435

Ramharack R, Deeley RG (1987) Structure and evolution of primate cytochrome c oxidase subunit II gene. J Biol Chem 262:14014–14021

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–525

Sanger F (1981) Determination of nucleotide sequences in DNA. Science 214:1205–1210

Sarich VM (1991) Mammalian systematics: 25 years among their albumins and transferrins. In: Szalay FS, Novacek MJ, McKenna MC (eds) American Museum of Natural History symposium on mammalian phylogeny. Princeton University Press, Princeton NJ (in press)

Smith TF, Waterman MS (1981) Comparison of biosequences. Adv Appl Math 2:482–489

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Southern SO, Southern PJ, Dizon AE (1988) Molecular characterization of a cloned dolphin mitochondrial genome. J Mol Evol 28:32–42

Swofford DL (1989) PAUP: phylogenetic analysis using parsimony, version 3.0g. Illinois Natural History Survey, Champaign IL

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc 26:375–400

Wilson AC, Zimmer EA, Prager EM, Kocher TD (1989) Restriction mapping in the molecular systematics of mammals: a retrospective salute. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 407–419

Received May 21, 1991/Revised July 9, 1991