J Mol Evol (1991) 33:83–91

Journal of Molecular Evolution

© Springer-Verlag New York Inc. 1991

The Main Regulatory Region of Mammalian Mitochondrial DNA: Structure–Function Model and Evolutionary Pattern

Cecilia Saccone,[1] Graziano Pesole,[1] and Elisabetta Sbisá[2]

[1] Dipartimento di Biochimica e Biologia Molecolare, Universitá di Bari, Italy
[2] Centro di Studio sui Mitocondri e Metabolismo Energetico, CNR, Bari, Italy

Summary. The evolution of the main regulatory region (D-loop) of the mammalian mitochondrial genome was analyzed by comparing the sequences of eight mammalian species: human, common chimpanzee, pygmy chimpanzee, dolphin, cow, rat, mouse, and rabbit. The best alignment of the sequences was obtained by optimization of the sequence similarities common to all these species.

The two peripheral left and right D-loop domains, which contain the main regulatory elements so far discovered, evolved rapidly in a species-specific manner generating heterogeneity in both length and base composition. They are prone to the insertion and deletion of elements and to the generation of short repeats by replication slippage. However, the preservation of some sequence blocks and similar cloverleaf-like structures in these regions, indicates a basic similarity in the regulatory mechanisms of the mitochondrial genome in all mammalian species.

We found, particularly in the right domain, significant similarities to the telomeric sequences of the mitochondrial (mt) and nuclear DNA of *Tetrahymena thermophila.* These sequences may be interpreted as relics of telomeres present in ancestral linear forms of mtDNA or may simply represent efficient templates of RNA primase-like enzymes.

Due to their peculiar evolution, the two peripheral domains cannot be used to estimate in a quantitative way the genetic distances between mammalian species. On the other hand the central domain, highly conserved during evolution, behaves as a good molecular clock.

Reliable estimates of the times of divergence between closely and distantly related species were obtained from the central domain using a Markov model and assuming nonhomogeneous evolution of nucleotide sites.

Key words: Mammalian mitochondrial DNA — Origin of replication — Mitochondrial DNA evolution — Stationary Markov model — Phylogenetic tree — Telomeres — D-loop — Regulatory region

Introduction

The presence of only one major noncoding segment in the mitochondrial genome is a feature common to all metazoa. In vertebrates this region, spanning between the Phe- and Pro-tRNA genes, is called the D-loop-containing region because of the three-stranded displacement (D) loop structure created by the nascent heavy (H) strand at the level of the H-strand replication origin (O_H). It also contains promoters for the transcription of both the heavy strand (HSP) and the light strand (LSP). This region is the target site for numerous proteins and enzymes, such as DNA and RNA polymerases and transcription and regulatory factors and is thus subjected to various evolutionary pressures. Because all these proteins are coded for by nuclear DNA, the study of the D-loop-containing region is also extremely important for shedding light on the processes inherent in nucleus–mitochondrion coevolution.

In order to gain deeper insight into the evolutionary dynamics of the noncoding region of the mammalian mitochondrial genome, we undertook a detailed investigation of its evolution at the molecular level. In previous papers we have identified several well-preserved features in the evolution of

Offprint requests to: C. Saccone

```
                10        20        30        40        50        60        70        80        90       100       110       120
PYG     -----------TTCTTTCATGGGGAAGCAAATTTAAGTGCCACCCAAGTATTGG-------------------------------------------------------------.---C
                   :::::::::::::::::::::::: :: ::::: ::::: :::                                                                 :
COM     -----------TTCTTTCATGGGGAAGCAAATTTAGGTACCACCTAAGTACTGG-------------------------------------------------------------.---C
                   :::::::::::::::::: :::: ::::::::: ::::: ::                                                                  :
MAN     -----------TTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGA-------------------------------------------------------------.---C
                   ::  : ::      :  :       :: ::::   :::::: :
DOL     AAAAAAGCTTATT-GTACAATTACCACAACCCCACAGTGCCACGTCAGTATTAAAAGTAATTTATTTTAAAAACATTTTACTGTACACATTACATACACCAATAC----------.TTAG
            ::   ::::  :: : :: ::: ::    :  : :::    ::::::::::     :::::   :::  :   : :::  :::::  :  : ::::   ::              :
COW     ----AACACTATTAATATAGTT-CCATAAATACAAAGAGCCTTATCAGTATTAAA-----TTTATCAAAAATCCCAATAACTCAACACAGAATTTGCACCCTAACCAAATATTAC_AATG
                                                                                                                               :
RAT     -------------------------------------------------------------------------------------------------------------------.TCAG
                                                                                                                              ::
MUS     -------------------------------------------------------------------------------------------------------------------.--AG
                    .  . .. ..   .....   ...  . . ....... . . .   . .. .. .   ...  .   ..   .   .   ... .   .... .         :
RAB     -----------CTCTTTACTTTA-ATAAAACTCAAGTACTTCATCAGTACTGACAAATTACTAACACACTATGTAATT-CCGTGCATTAATGCTCGCCCCCATTAAAATGTATT.--AT
                                                                                                                        IS <A

               130       140       150       160       170       180       190       200       210       220       230       240
PYG     T-C-ATTCACTA---------------TAAC-CGCTATGTATT-TCGTACATTACTG-CCA--GCCACCATGAATA--TTACATAGTACTATAATCATTTAACCACCTATAACACATAAA
        : | ::::: ::                :|| :::|||||:|: |:||:|||::::: :::  ::::|||:::::|  :   : :||:: :|: :::   ::: :::::::  :::::::
COM     T-C-ATTCATTA---------------CAAC-CGCTATGTATT-TCGTACATTACTG-CCA--GCCACCATGAATA--TCGTACAGTACCATA-TCACCCAACTACCTATAGTACATAAA
        : | :  ::: :               ::|| :::|||||:|: |:||:|||::::: :::  ::::|||:::::|  : :::: ||::::|:.  ::   :: :::: :::::::::::
MAN     T-C-ACCCATCAA--------------CAAC-CGCTATGTATT-TCGTACATTACTG-CCA--GCCACCATGAATA--TTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAA
        : |  : : :                   ||  ::||||||:|: | || |||:  :   :     |||:    |          ||    |:::  :  :       : :: : ::::::
DOL     T-C-TCTCTTTGTAAATATTCATATA--TACATGCTATGTATTATTGTGCATTCATTT--ATTT---CCATACG-A-TAA-----GT----TAAAG-CCCGTATTAAT-TA-T-CATTAA
        : |         :    :    :   ::: :    ||::   |||||:|  :| || |||: : ::   :: ::  :: : . |||: : .| :::. .  ||.. .| :    :  ::
COW     TAC-ATAACATTA-AT-GTA-ATAAA--GACATAATATGTAT-ATAGTACATTAAATT--ATATGCCCCATGCATA-TAAGCAA-GTAC-ATGACC-TCTATAG----------------
        ::| :::: :: : :: : :.:: ::    ||::   |||||:|  :| ||:|||:::::   :::: :: : ::||| ::::| :::::: ||: :|       :::: :   .
RAT     TAC-ATAAAATGATATGG-ACATTAA--AACATT-TATGTAT-ATCGTACATTAAATT--ATTTTCCCCAAGCATA-TAAGCAT-GTA--ATATATATCTAATGATTT------------
        ::| ::::: : : :: :.::: :      ||:::  |||||:|  :|:||:|||:::: :  ::: :::::::|||:::::| :::::  ||:. :|  : ::: :::: ::
MUS     TAC-ATAAATTTACATAGTACAACAG--TACATT-TATGTAT-ATCGTACATTAAACT--ATTTTCCCCAAGCATA-TAAGCTA-GTAC-ATTAA-ATC-AATGGTTC------------
         :| :::::: : ::::    :::::     ||::  ||||| |  :|:|| ||| ::: :..    : ::||| : ::| ::::::: ||:: :|:: .    . :: :  : .. . .. .
RAB     AACAATAAAT-T-CATAA-CCAACATTTAACATACTATGTTTAATCGTGCAT-AAATTCCTCATCCCCCATGAATAATAAGCTA-GTAC-ATTACTGCTTGATTGGACATAATCCACT--
        ........................................................................................................................

               250       260       270       280       290       300       310       320       330       340       350       360
PYG     _CAGTACATAGCACATACAATTATATACCGTACATAGCACATTACAGTCAAATCCATCCTCGCCCCCACGGATG----------------------------------------------
         ::: :||::: |:|::::::  ::: |::::|:||:::|::::::::::|:: :: |::|::::::::::::::
COM     _CAGAACATAGTACATACAACCATACACCGTACATAGCACATTACAGTCAAACCCCTCCTCGCCCCCACGGATG----------------------------------------------
         ::: :||::::|:|:: : ::::  |::::|:||:::|::::::::::|:: :::| :|:: ::::: :::::
MAN     _CAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATG----------------------------------------------
            ::||:: :|:|: :         : | : :|:||:  | ::  :      |:      |: |    :    ::    :
DOL     .TTTTACATATTACATGATATGTATAATCTTACATATTATATATCCCCTAACAATTTTATTTCCATTATACCTATGGTCGCT------CCATTAGATCACGAG-----------------
            ::||:: :|:|:    ::   :: |    :|:||: :| ::         |:      |: |:    : :::: ::::  :   :         ::::::::::::::
COW     .CAGTACATAATACATATAATTATTGACTGTACATAGTACATT-ATGTCAAATTCATTCTTGATAGTATATCTATTATATATTCCTTACCATTAGATCACGAG-----------------
           : :||    :|:|: :::    :   |:: :| ||    |::          :|   :  :|::|      : :: ::        :   ::        ::::::
RAT     .AGG-ACA---TACATTTAAAC-TCAACTATAAATTC-ACAA--CAA-CATGTCTATTCTC--AAATACATT----AAG-ATAATGCTT-ATTAGACATATCTGTGTTATTAGACATG--
         :::. ||    :| |   ::: :.:::|: ::|:|| : | :   .   : :|:: ::|: |:   ::: :::::      ::   ::::: ::       ::::::::::::::::::  :::::
MUS     .AGGT-CA---TAAAA-TAATCATCAAC-ATAAAT-CAATATATATACCATGAATATTATCTTAAACACATT----AAA-CTAATG-TTATAAGGACATATCTGTGTTATCTGACATA--
            :.||. . | |   :::: ::|: : |:||.  |          : :|::::::|: |:::  ::        :::.:::::: ::       :::::           ::    ::::
RAB     .TAATACATCACACAT--AATC--CAACAAAAAATTG-ACC---CAAACATGAATATTCTCACCAA----------AAATCTAATGATTGACTTGACATCAGACATCAATTC--CATAAT
        IS ......................................................................A>

               370       380       390       400       410       420       430       440       450       460       470       480
PYG     ---------------------------------------------------------CCCCCCCTCAGATAGGAATCCCTTGGC-CACCATCCTCCGTGAAATCAATATCCCGCACAAGA
                                                                 : ::::::::::::::::::::::::  :||||||:|::|||||||:||::|:||||||::::::
COM     ---------------------------------------------------------CTCCCCCTCAGATAGGAATCCCTTGGT-CACCATCCTCCGTGAAATCAATATCCCGCACAAGA
                                                                   :::::::::::::::  :::::::   :||||:|::||||||||:||::|:||||||::::::
MAN     ---------------------------------------------------------ACCCCCCTCAGATAGGGGTCCCTTGA-CCACCATCCTCCGTGAAATCAATATCCCGCACAAGA
                                                                                      ::: :. :|||||  |  ||||||| ||  | ||||| :   :
DOL     ------------------------------------------------------------------------------CTTAAT-CACCATGCCGCGTGAAACCAGCAACCCGCTCGGCA
                                                                                      ::::::  |||||:|::||||||||:||::|:||||||: ::::
COW     ------------------------------------------------------------------------------CTTAAT-TACCATGCCGCGTGAAACCAGCAACCCGCTAGGCA
                                                                .  .. .. . .. ..  ::  :.:||||| |  ||||||| ||  :|:|||||
RAT     --------CACCATTAAGTCATAA-------ACCTTTCTCTT--CCATATGACTATCCCTGTCCCCAA-TTGGTCTCTATT--TCTACCATCCTCCGTGAAATCAACAACCCGCCCACTC
                ::::::  ::::::::       ::  :::::::  ::::::::::::::::  : :::  : ::: : ::::  ::::::..:::||||:|::||||||| ||::|:||||||::::
MUS     --------CACCATACAGTCATAA-------ACTCTTCTCTT--CCATATGACTATCCCCTT-CCCCATTTGG--TCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCA
                 :::::  : :: ::        ::::  :::::  ::::: :::::::::  :.::::::   :  .:::   ::  :|||||:|::||||||||:||::|:||||||::::::
RAB     TAAACATAGACCATCAAATC-TACACACACCACTCAACTCTTACCCATACGACTATCCCTCTCCCCCA---GTCCTCTCACAACTTACCATCCTCCGTGAAACCAACAACCCGCCCACCA
                                                                                       <C.........................................

               490       500       510       520       530       540       550       560       570       580       590       600
PYG     GTG--TACTCTCCTC-GCTCCGGGCCCAT-AACACTTGGGGGTAGCTAA--ACTGAA-CTGTATCCGACATCTGGTTCCTACCTCAGGG-CCATGAAG-TTCAAA-GGACTCCCACACGT
        :::    :||||:||| ||:|||||||||| ::::  |||||||::|||:   : |||: :|:||:::|:|||::||||||:|:|:||||||: :|||:::: ::::::   ::::::||::|:|
COM     GTG---ACTCTCCTC-GCTCCGGGCCCAT-AACATCTGGGGGTAGCTAA--AGTGAA-CTGTATCCGACATCTGGTTCCTACCTCAGGG-CCATGAAG-TTCAAA-AGACTCCCACACGT
        :::    :|||||:||| ||:|||||||||| ::::  |||||||::|||:   ::|||: :|:||:::|:|||::||||||:|:| ||||||:   ||| :::     :::     :   :||::|:|
MAN     GTGCT-ACTCTCCTC-GCTCCGGGCCCAT-AACACTTGGGGGTAGCTAA--AGTGAA-CTGTATCCGACATCTGGTTCCTACTTCAGGG-TCATAAAG-CCTAAA-TAGC--CCACACGT
        :    :  |||| ||| || ||||||||||| :     : ||||||||::|||:   : |||   :| ||:   |:|||: ||||| |:|:|||||     |||         :::: :  ::   :||: |:|
DOL     GG-ATCCCTCTTCTC-GCACCGGGCCCATGATACCGTGGGGGTAGCTAA-TAATGA-TCTTTATAAGACATC-GGTTCTTACTTCAGGA-CCATCTTAATTTAAAATCGC--CCACTCGT
        ::.::::||||:||| || ||||||||||| : ::::||||||'|| :|||    ::|||.  |:||   :| |||:.|||||:| |:|||||    :|||:: :    ::::: :    :|| :| |
COW     GGGATCCCTCTTCTC-GCTCCGGGCCCAT-AAACCGTGGGGGTCGCTATCCAATGAA-TTTTACCAGGCATCTGGTTCTTTCTTCAGGG-CCATCTCATC-TAAAACGGT--CCATTCTT
        :     ::||||:||| ||:|||||||||||         : ||||||||   |||: :   |||:   |:|| ::|:|||::|||||:| |:|||||: :|||:     :        :  ::   :||: | |
RAT     GTCC-CCCTCTTCTC-GCTCCGGGCCCAT-TCGTCCTGGGGGTGACTATAC--TGAAACTTTA-CAGGCATCTGGTTCTTACTTCAGGGGCCATCAATTG-GTTCATCGT--CCATACGT
         : : ::||||:||| ||:|||||||||| :    : |||||||   ||| ::   |||::|:|| ::| |||::||||||:|:|:||||||: :|||::: :: ::: ::::      ||::|:|
MUS     ATGC-CCCTCTTCTC-GCTCCGGGCCCAT-TAAACTTGGGGGTAGCTAAAC--TGAAACTTTATCAGACATCTGGTTCTTACTTCAGGG-CCATCAAATGCGTT-ATCGCC-TCATACGT
        : : .::||||:||| ||:||||||||||    ::::::||||||   |||:      |||:::| || : | ||| :||||||:|:| |||||: :||| ::     :     ::::::   || :|:|
RAB     AGGATCCCTCTTCTCCGCTCCGGGCCCAT-AAAACTTGGGGGTTTCTAATA--TGAAACTATAACTGGCAT-TGGTTCTTACCTCAGGG-CCATGAA--CCTAAGATCGCC--CACACGT
        ........................................................................................................................
```

Fig. 1. The best alignment of the eight mtDNA D-loop-containing regions. Regions A, B, and C, the CSBs, and the IS (_), SR, and LR, the O_H (⇐), LSP (←), and HSP (→) are indicated. Continued on pages 85–86.

```
           610       620       630       640       650       660       670       680       690       700       710       720
PYG   TCCCCTTAAATAAGGCATTCACGATGGA-TCACAGGTCTA-TCA--CCCTATTAACCACTCACGGGAGCGC--TCCATGCATTTGGTAT-TTT---CGTCTGGGGGGTG---TGCACGCG
      |||:||||||||||| || ||:||||||: :::::::::||| |||  |||:||::::|: :||:::::|:     :::|::||:|||:||| |||    ::|:|:::||:::   |:|:::::
COM   TCCCCTTAAATAAGACA-TCACGATGGA-TCACAGGTCTA-TCA--CCCTATTAACCAGTCACGGGAGCCT--TCCATGCATTTGGTAT-TTT---CGTCTGGGGGGTG---TGCACGCG
      |||:|||||||||||:|| ||:||||||: :::::::::||| |||  |||:||::::|: :||:::::|:     :::|::||:|||:||| |||    ::|:|:::||::    |:|:::::
MAN   TCCCCTTAAATAAGACA-TCACGATGGA-TCACAGGTCTA-TCA--CCCTATTAACCACTCACGGGAGCTC--TCCATGCATTTGGTAT-TTT---CGTCTGGGGGGTA---TGCACGCG
      ||| ||||||||||:|| || ||||||  :::        ||| |||  ||| ||    :|    ||      |          : :|: ||:|||:||| |||      | |:::||:      |:|:: ::
DOL   TCCTCTTAAATAAGACA-TCTCGATGGT-TCATGA--CTAATCAG-CCC-ATG--CCTAACATAACTGAGG-TTTCATACATTTGGTAT-TTTTTAATTTTGGGGGGGGCTTGCAC-CG
      |||:||||||||||||:|| ||:|||||||  : :::   |||:|||: ||| ||:     |   :||:::::| :   : ::|::||:|||:||| |||:: ::|:|:::||         |:|     :
COW   TCCTCTTAAATAAGACA-TCTCGATGGACTAATGG--CTAATCAG-CCC-ATGC-TCACACATAACTGTGC-TGTCATACATTTGGTAT-TTTTTTATTTTGGGGA-----TGC-TTGG
      ||| ||||||||||:|| ||:|||||||  :::::::..|||:|||: ||| ||:   :|: :||:::::|::  ::   | :|| |||:||| |||:: ::|:|      ||:         |:| ::
RAT   TCCCCTTAAATAAGACAATCTCGATGG--TAACGGGTCTAATCAGACCC-ATGA-TCA-ACATAACTGTGG-TGATACACA-TTGGTAT-TTTTTAATTTTC--GGA-----TGCCTTC-
      |||:||||||||||:|| ||:|||||||  :: :::::|||:||||: ||| ||::   |: :||:::::|::: ::   |   ||.|||:||| |||:: ::|:|      ||          | | :::
MUS   TCCCCTTAAATAAGACA-TCTCGATGG--TATCGGGTCTAATCAG-CCC-ATGA-CCA-ACATAACTGTGG-TGTCATGCATTTGGTATCTTTTT-ATTTT---GGCC----TACTTTC-
      ||| ||||||||||:|| ||:|||||||        :    : |||:|||: ||| ||:    |:.:||:::::|::: ::::|::||:||| ||| ||| :.::|:|      ||          | | :: .
RAB   TCCTCTTAAATAAGACA-TCTCGATGG--ACTAATGACTAATCAG-CCC-ATGC-TCACACATAACTGTGGATGTCATGCATTTG-TAT-TTT-TAATTTTTTTGGGTTA--TGC-TTGG
      ...........................................................................................C>   <B......................

           730       740       750       760       770       780       790       800       810       820       830       840
PYG   A-T-AGC-ATTGC-G---AAACGC--TGGCCCCGG--AG-CACCC-TA-TGTCGC----AGTATCTGTCTT-----TGATTCCTGCCCCATTACGTTATTTATCGCACCTACGTTCAATA
      : | |:| ||:|| :    ::::|| ||:::::::  |: |:::: :| :||:::    ::::::::::::     |:|:::::::::::::   ::::::::::::::::::::::::
COM   A-T-AGC-ATTGC-G---AAACGC--TGGCCCCGG--AG-CACCC-TA-TGTCGC----AGTATCTGTCTT-----TGATTCCTGCCCCATTGTATTATTTATCGCACCTACGTTCAATA
      : | |:| ||:|| :    : ::|| ||:  ::::  |: |:::: :| :||:::    ::::::::::::     |:|:::::::: :::  ::::::::::::::::::::::::::
MAN   A-T-AGC-ATTGC-G---AGACGC--TGGAGCCGG--AG-CACCC-TA-TGTCGC----AGTATCTGTCTT-----TGATTCCTGCCTCATCCTATTATTTATCGCACCTACGTTCAATA
      : | |:| || ||      : : || ||          |: |:      | :|| ::    :    : : ::: ::     |:|::         ::
DOL   ACTCAGCTATGGCCTTAGAAAGGCCCTGTC-----ACAGTCAAATAAATTGTAGC-GGGCCTGTGTGTATTTT---TGATTGGACTAGCA----------------------------
      ::|:|:|:||:||: :  ::::||::|| : ....  |: |:      |::||:::    :    :    :     |:|  : :: :::
COW   ACTCAGCTATGGCCGTC-AAAGGCCCTGAC-CCGG--AG-CATC--TATTGTAGC--TGGA--CTTAACTGCATCTTGA--GCACCAGC-----------------------------
       :|:| | || ||::::  :::||  ||:     :: .|: |:      |  ||    .::::  :::        | |   :   :     .....  ..
RAT   -CTCAAC-ATAGCCGTC--AAGGCA-TGAA---GGTCAG-CA----CAAAGTCCT-GTGGAACCTTTTAGT-----TAAGGG-TCATTTATCCTCATAGAC-------------------
      . |:|:| ||:||::::  :::||: ||::    :: :|: |:     :| :||:    :  : ::::     :     | | :  :::::  ::: :: :  :
MUS   A-TCAAC-ATAGCCGTC--AAGGCA-TGAAA--GGACAG-CA----CACAGTCTA-GACGCACCTAC-GG------TGAAGAATCATTAGTCCGCAAAACC-------------------
      :.|:|:| || ||::   .    :|| .||:   . ::::|   |        |  ||   : :: :::::::     |:|:::   : :   ::: :: ::
RAB   ACTCAAC-ATGGCCGCGGTG-GGCCCTGACCCGGGACA--CT----TATTGTAGACGA-GCACCTAA---------TGAAGA-CCCTCCATCCTCATAATT-------------------
      ...........B>

          850       860       870       880       890       900       910       920       930       940       950       960
PYG   TTATTACCTAGCATGATTTACTAAAGCGTG-TTAATTAATTAATGCTTGTAGGACATAA-CAATAG-CAGCAAAATAC-CACGT.-AACTGCTTTCCACACCAAC-ATCATAACAAAAAA
      :::  ::::::::: :  :::::::: |:: |::::: :|::|||:|:: ::||||||: ::: :: ::::::::: : ::: :  :::::::::::::::::::: ::::::::::::::
COM   TTACGACCTAGCAT-ACCTACTAAAGTGTG-TTAATTGATTAATGCTTGCAGGACATAA-CAACAG-CAGCAAAATGCTCACAT.-AACTGCTTTCCACACCAAC-ATCATAACAAAAAA
      ::::   : : ::: :: :::::::::|:: |::::: :|::|||:|:: ::||||||:  :: :  ::     ::::     ::   : :  :::::::::  :: ::::::::::::::
MAN   TTACAGGCGAACAT-ACTTACTAAAGTGTG-TTAATTAATTAATGCTTGTAGGACATAA-TAATAA-CAATTGAATGTCTGCAC.-AGCCACTTTCCACACAGAC-ATCATAACAAAAAA
                      :  : : :: |:: |:: ::::|::||| |:  ::||||||| :: :      :::        : :            :: : : :  : ::: : ::
DOL   ----------------CAACCAACAG-GTG-TTATTTAATTAATGGTTACAGGACATAT-TACTCTATTATT---CCCCCGGGT.---------TCAAAAAACTCTATC-TCACGGGGG-
                     .  :  : ::  |    |::    : | :||||:| ::::||||||  :    ::::::   :::::           ::: :::: : :
COW   ------------ATAATGATAAGCATGGACATTA--CAGTCAATGGTCACAGGACATAAATTA--TATTATATATCCCCCCT--.---------TCATAAAAATTT-------------
                      ::::   :  |:: | :..    |  ||| |     : ||||||::::     : ::: :::                   :::: : :
RAT   ----------------AAAGCTCGAAAGAC-T-ATTTTATTCATGTTTGTAAGACATAAATAT--TTATAAATACTG-------.-------------AAAACTCT-------------
                      :: : :  ::| : | :::   :|::||| |::: :||||||:::     :  : :::::        .                ::::::
MUS   ----------------CAATCACCTAAGGC-TAATT--ATTCATGCTTGTTAGACATAAATGC--TACTCAATACCAAATTTT-.---------------AACTCT-------------
                       :: : :   | :.|  ::   |: |||:|:::  ||||||:: :              :::::      . .    ...    :: : :..      . .
RAB   ------------------ATGAGCCGGGACATTCTT---TTAATGCTTGTCGGACATAAA-GAG--------------ATTTTA_ACGACTA-TTCC---AATTATATA--GATGAATTT
                                     <.......... CSB 1 ........>                                     S̄R
                   <=                <=  <=
                     MAN,COW           MUS RAT

           970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080
PYG   TTTCCGCCAAACCCCCCCTCCCCCACTCCTGG--------------------CTACAGCACT---CAAATTCATCT----------CTGCCAAACCCCAAAAACA-------AAGAACCC
      :: ::  :::::::::::: ||||     : ::                    : :::::::   ::::   ::::          ::::::::::::::::::::       ::::::::
COM   TTCCCA-CAAACCCCCCCTTCCCC----CCGG--------------------CCACAGCACT---CAAACAAATCT----------CTGCCAAACCCCAAAAACA-------AAGAACCC
      :: :::.:::::::::::: |||| .. : ::                    :::::::::     ::::: ::::          ::::::::::::::::::::       ::::::::
MAN   TTTCCACCAAACCCCCCCTCCCCCGCTTCTGG--------------------CCACAGCACT---TAAACACATCT----------CTGCCAAACCCCAAAAACA-------AAGAACCC
      :::     ::::::::: ::||||
DOL   TTT-----AAACCCCCCTTCCCCCT-------------------------------------------------------------------------------------------------
                         :||||:
COW   -------------------CCCCCT-------------------------------------------------------------------------------------------------
           .. ..........  :||||:    .  .                                                     ..................    .....
RAT   --GTCAACAAACCCCCCCACCCCCTACACCTGAAACTTCAA------------------------------------------TGCCAAACCCCAAAAACATTAAAGCAAGAA-TT
              :::::::::: ::||||: :  :: ::                                                 :::::::::: ::::::: :    ::::: ::
MUS   ------CCAAACCCCCC-ACCCCCTCCT-CTTAA---------------------------------------------TGCCAAACCCAAAAAACACT-----AAGAACTT
          ..:::::::::::.  |||| ::  :: :                       ...  ... .  .....     .     .:::::::::: ::::::       .:::::
RAB   A-TCCACCAAACCCCCCCTACCCCCCCA-CTAAGTA-----AATTCATTGCCCCAAGGCAATGAATAAACTATGCATATTAATTTCCTGCCAAACCCCAAAAAC-------CAAGAATCA
              <.... CSB 2  ...>                                                              <.... CSB 3   ...>

          1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
PYG   AGATACCAGCCTAACCAGACCTCAAATTT-.-CATCTTTTGGCGGTATGCATTTTTAACAGT----CACCCCTCAACTAACATGTCCTCC--CCCCTCA-ACT-CCCATTCCACTAGCCC
      :::  :::::::: ::::::: ::::::::   ::::::: ::::::::::: :::::::::    ::::::::::: :::::::: :::::  ::::::: ::: ::::::: ::::::::
COM   AGACGCCAGCCTAGCCAGACTTCAAATTT-.-CATCTTTAGGCGGTATGCACTTTTAACAGT----CACCCCTCAATTAACATGCCCTCC--CCCCTCA-ACT-CCCATTCTACTAGCCC
        ::  :::::::: ::::: ::::::::    :::::: ::::::::::::::::::::::    :::::: ::: :::::       :    :::::: ::: ::::: :::::
MAN   TAACACCAGCCTAACCAGATTTCAAATTT-.-TATCTTTTGGCGGTATGCACTTTTAACAGT----CACCCCCCAACTAACACATTATTT-TCCCCTCCCACT-CCCATACTACT-----
      ::: : : :                          :: : :::  ::::: : :    :::::::: ::         :  ::     ::::      :: : :  ::: :
DOL   TAAAAACTGA--------------------.----------TCG-TCTGC---TTTAATA-TTCACCACCCCCCTACA-----GTGCTTC-GTCCCTAGATCTACGCGCACTTTTTT---
      ::::                                                :  ::    ::   ::::::      :  :.   :  :::   :::::::::         :  :::::
COW   TAAA--------------------------.-----------------------TATCTA--CCACCAC----TTTTAACA--GA-CTTT--TCCCTAGATACTTATTTAAATTTTT---
       ::: .     ..   . .          ..  ..    ..     ::::..  :::::    :::::::         ::    ::   :::   :  ::   :  ::::::
RAT   -AAATAAAACAAAAAGCTACT---------.TAATTCTTAAAAGG--CTTCTCCATTCTAGTAGACCACAAAATTTTAAC-------TTAAATCT-TAG-CA-TTGGTAAAATTTCCCGA
       ::: : : : :: :   ::          ::          : :..:: :: :::::::::  : :::::: :        .. .::: :: : ::: :.::: :::::::
MUS   GAAAGACATATAATATTAAC----------.TATCAAACCCTATGTCCTGATCAATTCTAGTAGTTCCCAAAATAT-------GA-CTTATATTT-TAG-TACTTGTAAAAATTTT----
               :::    :::    . .....
RAB   CCGCACAGTATTTACTTAGACT-AAATT-- ------------------------------------------------------------------------------------------
       <-                         L̄R      <-                   <-  <-
        COW                                 MAN                  RAT MUS
```

Fig. 1. Continued

```
           1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320
PYG  C-----AACAACAT--AACCCCCTGCCCACCCCACTCAGCACATAT--ACCGCTGCTAACCCTATACCCTAAGCCAAC-CAAACCCCAAAGAT-ATCCCCACACA---------------
     :     : :::: :  :::::::: : ::::: ::::: :::::::  :::::::::::::: ::::::: : ::::: :::::::::::::  : :|: :::::
COM  C-----AGCAACGT--AACCCCCTACTCACCCTACTCAACACATAT--ACCGCTGCTAACCCCATACCCTGAACCAAC-CAAACCCCAAAGAC-ACCCCTACACA---------------
           : :::     :::::::  : :: ::::: :: :::: :   ::::::::::::::::::::: :::::::: :::::::::::::: :::|:  ::::
MAN  AATCTCATCAATAC--AACCCCCGCC-CATCCTACCCAGCACACACACACCGCTGCTAACCCCATACCCCGAACCAAC-CAAACCCCAAAGAC-ACCCCC-CACA---------------
     :::   ::::::::  ::                               :::   : : :::::::    :::   :  ::::           :: |: .
DOL  AATAA-ATCAATAC-CAA-----------------------------ATCCGACACAAGCCCCATA--ATGAAATTATACAAATAATTTTAT--ACTCCAAA------------------
      :     ::::::: :::                                                                      ::::   :::|::::  .
COW  CACGCTTTCAATACTCAA---------------------------------------------------------------------TTTAG-CACTCCAAACAAAGTCAATATATAAAC
     :::    ::  : :::                                                    .   ..     ::::    : | :  :  ::  ::  :   :
RAT  CACAAAATCTTTCCTCCT---------------------------------------------------------AACTAAACCCTCTTTACTTGC-CTACCCTCAGAAAA--TTCCACA
      ::::::::  :  :::                                                        :::  ::  ::::   : : :.|::   :  : ::    :  :::
MUS  -ACAAAATCAATGTTCCG-------------------------------------------------------TGAAC-CAAAACTCTAATCATACTCTA--TTACGCAA----TAAACA
                                                                                     ::: ::::: :: ::  |::
RAB  -----------------------------------------------------------------------------TAGAAATCTCTAGTC-TAGGCTAAA------------------
                                                                                          ->                     ->  ->    ->
                                                                                        MAN                    MUS COW   RAT
```

```
           1330      1340
PYG  --------------------

COM  --------------------

MAN  --------------------

DOL  --------------------

COW  GCAGGCCCCCCCCCCCC---
          :
RAT  TACACCAAA-----------
     :  : :::
MUS  TT-AACAA------------

RAB  --------------------
```

```
            D-LOOP INTERVENING SEQUENCES (IS)
            _________________________________

MAN         AACCCAATCCACATCAAAACCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCCT
            CAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAA
            ACCTACCCACCCTTAA

COW         AAACACCACTAGCTAACATAACACGCCCATACACAGACCACAGAATGAATTACCTACGCA
            AGGGGT

RABBIT     (GCACGTACACCCGTACGCAC)10 GCACGTACACCCGTACACCCGTACACCCGTACGCAC
SR          GCACGTACACCCGTACACCCGTAC

RABBIT     (TAAACCCCCTTTCCCACCCCAAGTCAGACAGCTCAGGGCATCTAAATTTTGAAATTTAAA
LR          ACGCACCTTTACAATACTGACATAGCACTCTAGCCCTTTTTTTCCTTTTAACAGGTTTAA
            CTCAATTAAATACAAATTGTATAATATTTGGAC)4
```

Fig. 1. Continued

Table 1. Sequence similarities for sequence blocks A, B, and C and CSB1, 2, and 3

	Similarity %					
Organisms compared	Region A	Region B	Region C	CSB1	CSB2	CSB3
Pygmy chimp–common chimp	88	100	96	93	94	100
Human–chimps	86	97	97	96	97	100
Rat–mouse	83	88	90	86	94	94
Rat/mouse–rabbit	70	83	82	71	85	94
Cow–dolphin	59	90	87	75	[a]	[c]
Cow/dolphin–rodents	61	78	80	61	82[b]	[c]
Cow/dolphin–primates	58	77	72	77	86[b]	[c]
Primates–rodents	57	59	74	79	91	96

[a] In cow the CSB2 sequence (17 nt) is reduced to a run of 5 C

[b] Similarity calculated with respect to the dolphin sequence

[c] In dolphin and cow the CSB3 sequence is absent

the mammalian D-loop and we have also shown that the evolutionary behavior of this region varies in interspecies and intraspecies comparisons (Brown et al. 1986; Saccone et al. 1987). This report extends our previous observations to more mammalian species, namely common chimpanzee, pygmy chimpanzee (Foran et al. 1988), dolphin (Southern et al. 1988), and rabbit (Mignotte et al. 1990). Furthermore, because mitochondrial DNA (mtDNA) has become a molecule used commonly for investigating molecular phylogeny problems, we also show that the central domain of the D-loop behaves as a reliable molecular clock and thus may be suitable for determining the branching order of mammals.

Results and Discussion

In contrast to the high degree of conservation of the rest of the genome, the D-loop region shows great variability in length and base composition in mammals. The eight sequences of mammals can be aligned easily only in the central region (about 200 bp long, Fig. 1). On the basis of this, the D-loop-containing region can be divided into three domains: the right (R) domain containing the O_H, the central (C) domain, and the left (L) domain where the nascent H strand pauses in the resting molecules. The novelty of our alignment compared to those previously presented consists in the optimization of the similarities in the R and L domains in all eight species and in the identification of insertion sequences and short repeated motifs. In Fig. 2 the dashed boxes represent regions with degrees of similarity spanning from 100 to 57% as reported in Table 1. Other segments display a significant degree of similarity only within some groups (between dolphin and cow and between rat–mouse and rabbit). The sequences denoted IS, SR, and LR and the blank boxes are unique for the species considered.

Evolution of the Right Domain

The R domain immediately adjacent to the Phe-tRNA gene is probably the most important functional part of the regulatory region as it contains the O_H and the two promoters HSP and LSP. It has different lengths in mammals as shown in Table 2 and its primary structure greatly diverges except for a short sequence of ~30 bases (block B) immediately downstream from the central domain, and the conserved sequence blocks (CSBs, Walberg and Clayton 1981) that have been suggested to act as processing signals for the enzymes involved in the generation of RNA primers for heavy strand replication. However, the degree of similarity between the CSBs is variable as shown in Table 1. The CSB1, in particular, is conserved in all the organisms considered whereas the flanking regions are very divergent even between closely related species (rat and mouse). It is striking to note that CSB1 differs from the rest of the D-loop region and is more similar in human and rabbit (86% of similarity) than in mouse and rabbit (75%) or than in cow and dolphin (75%). Though the significance of the data is weak for the small number of differences, a possible explanation could be coevolution due to an interaction between the CSB1 and a nuclear-coded protein. The branching of primates, artiodactyls, murids, and lagomorphs inferred from nuclear genes is highly controversial but in some cases infers that lagomorphs are more closely related to primates than to murids (Easteal 1990). In cow the CSB2 is restricted to a run of only 5 Cs whereas the CSB3, which is most highly conserved in primates and rodents, is completely absent in cow and dolphin.

Clayton's group has purified, from mouse and human, a site-specific endoribonuclease, called RNase MRP, which specifically recognizes CSB2 and CSB3. This enzyme, well characterized in mouse and human, contains a nuclear-encoded 5S RNA that has a region complementary to both CSB2 and CSB3. These two latter conserved sequence blocks have been found to constitute a critical bipartite recognition signal for accurate and efficient cleavage by RNase MRP. Heterologous assays with human enzyme and mouse mtRNA and analysis of mutations within CSB2 in vitro indicate that the essential components for substrate recognition are conserved in mouse and human in spite of the natural heterogeneity of CSB2 in vivo (Bennet and Clayton 1990). Further experiments are however necessary to verify this in the other mammalian species, particularly in cow and dolphin where the CSB2 and CSB3 are practically absent.

The promoters for the heavy (HSP) and the light (LSP) strand have been well characterized in mouse and human; in other species only partial information is available (Sbisá et al. 1990). The location of the two promoters with respect to other elements, like CSBs and cloverleaf-like structures (see below) seems maintained at least in the species in which they have been characterized. Recent studies using transcription factor I indicate once more than the transcriptional machinery and the overall mechanisms of transcriptional control and regulation are basically conserved in mammals in spite of the very divergent primary structures. According to Fisher et al. (1989), strict species specificity of mitochondrial transcription could be determined by RNA polymerase itself or by some as yet undetected transcription factor. These data lead us to suggest that the sequence of the R domain contains superimposed codes that determine its peculiar evolution in the various species.

We have previously reported (Saccone et al. 1985; Brown et al. 1986) that at the level of the D-loop termini the sequences of rat, mouse, cow, human, and *Xenopus*, despite their high primary structural divergence, are capable of assuming similar cloverleaf secondary structural configurations. These structures have also been identified in dolphin and rabbit (data not shown). The conservation of basically similar structures in mammals suggests that these are of principal importance for the regulation processes occurring in the D-loop-containing region.

In rabbits the R domain displays unique properties: stretches of short [SR = 20 nucleotides (nt)] and of long (LR = 153 nt) tandem repeats, present in variable copy number, are located between the CSB1 and CSB2 (SR) and downstream CSB3 (LR), generating intra- and interspecies length heterogeneity (Mignotte et al. 1990). SRs and LRs do not seem to be of mitochondrial origin because with hybridization experiments no sequence identity has been found with other parts of the mitochondrial genome. The SRs, which consist of a sequence of 20 nt (GCACGTACACCCGTACGCAC) exactly repeated in tandem 10 times, are followed by two sequence elements that are rearrangements of the SR 20-mer. For the SR-containing region Mignotte et al. (1990) have proposed a very stable secondary structure involving four repeats. We have found that alternative shorter secondary structures can be generated by three or two SRs (Fig. 3 A and B). This scheme involves the conservation of some kind of recombination mechanisms which should not be lost completely in Metazoa, as generally assumed. A closer inspection of such structures reveals that the deletion of one or two stem and loop structures, from type A and B molecules, produces the two types of rearranged repeats found in rabbit (Fig. 3A1 and B1). If this is the case the cleavage should involve the GT:AC paired palindromic region.

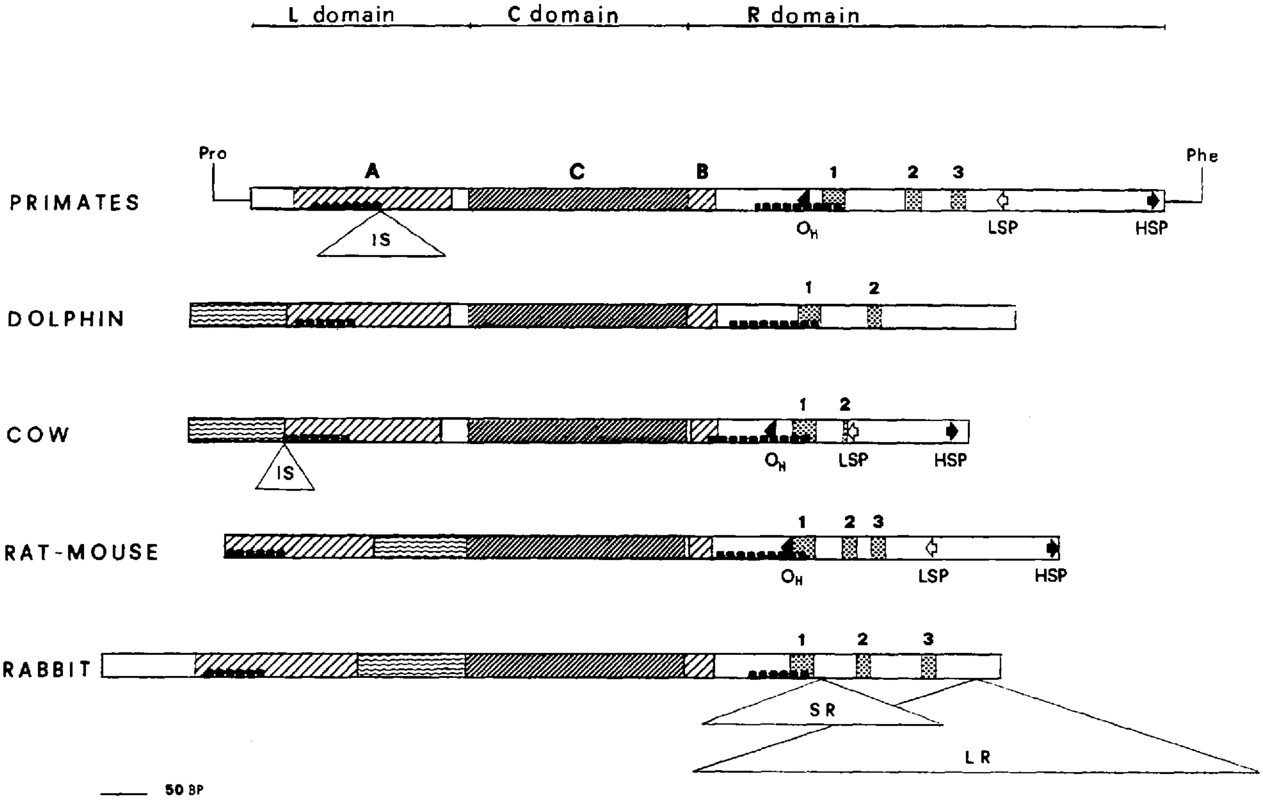

Fig. 2. General scheme of the organization of the D-loop-containing region in various mammals. The shaded segments correspond to regions of sequence similarities shown as blocks A, B, C and 1, 2, 3 (CSBs) (see also Table 1). The region upstream of block A is similar in dolphin and cow and the sequences downstream from block A are similar in rat–mouse and rabbit. IS indicates insertions unique to the organism(s). SR and LR indicate the short and long repeats found in rabbit. The location of the secondary structures is indicated by black squares. The H-strand replication origin (O_H) and the L- and H-strand promoters (HSP, LSP), where known, are also reported.

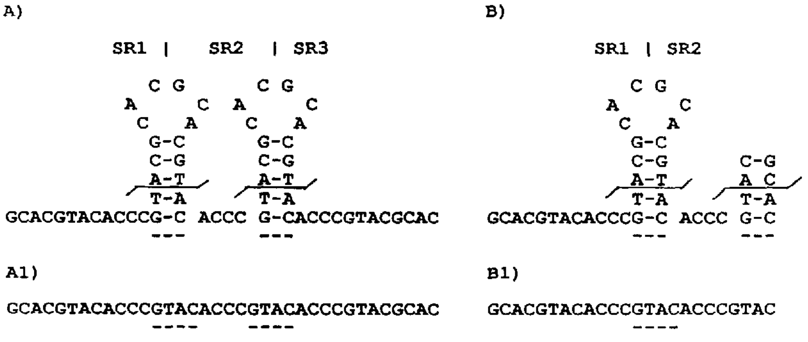

Fig. 3. Secondary structures formed by three (A) or two (B) rabbit short repeats (SR). The rearranged repeats found in rabbit (A1, B1) could be the product of a specific cleavage at the GT: AC paired palindromic regions.

The Left Domain

The L domain, downstream from the Pro-tRNA, is of crucial importance because it is here that whether the synthesis of the third strand should be arrested or whether it should continue is determined. Moreover, transcriptional analyses have demonstrated that at this level there is an active processing of the H and L polycistronic transcripts (Sbisá et al. 1990). Although this domain is highly divergent in mammals with regard to both nucleotide and sequence length, a block of about 160 nt (block A in Fig. 2), located at a variable distance (20 nt in human, 116 in rabbit) upstream from the central domain can still be aligned in all mammalian species. In the sequence block A we found short mirror symmetries (TACAT, ATGTA) repeated several times.

In particular two of these palindromic sequences form a very conserved stable hairpin-loop (nt 156–172 in Fig. 1), which are present in the cloverleaf-like structures (black squares in Fig. 2) located in this region. These sequences, present also in the pig D-loop (MacKay et al. 1986) and as part of some TAS elements [experimentally found by Doda et al. (1981) in mouse], may be involved as a recognition site for the arrest of H-strand synthesis.

In human and chimpanzee block A is interrupted by a 136-nt element we call insertion sequence (IS).

```
(A)
                     .         .         .         .         .         .
MITTTRA     -TACACATACCATAA---CAAG-CCTAAACAACACTAATAGCTAACATACCTCTAAGA--
               ** * * *** *   ****   **** ** ** ** ***** * **  *** * *
RAT         CCCCAAAAA-CATTAAAGCAAGAATTAAATAAAACAAAAAGCTA-CTTAATTCTTAAAAG
            ....csb3...>


(B)
                     .         .         .         .         .
MITTTRA     TACACATACCATAACAAGCCTAAACAACACTAATAGCT-AACATACCTCTAAGA
            * * * **  **  *** ** *** ******** * **  * * ** * *** *
MOUSE       TCCTCTTA--ATGCCAAACCCAAAAAACACTAAGAACTTGAAAGACATATAATA
                       <......csb3......>


(C)
                     .         .         .         .         .         .
MITTRA      TACACATA-------CCATAAC-AAGCCTAAACAA-CACTAATAGCTAACATACCTCTAA
             ***** *       ** **** **   *  **** *** * *********** * *
COW         AACACAGAATTTGCACC-TAACCAAATATT-ACAAACACCACTAGCTAACATAACACGCC
                                             <... cow IS ...............


(D)
                     .         .         .         .         .         .
TTRSTELC    AACTAAGAAATTATAAATTTTCTTTAAATATTTATAGTA--AATTTTATCAAAACCACCC
            ****** * *****   **   * *   *     *** **   *** * ***** ** ****
HUMAN       AACTAACACATTATT--TTCCCCTCCCACTCCCATACTACTAATCTCATCAATACAACCC

                     .         .         .         .         .         .
TTRSTELC    CCCCCCCCCCAAACCCAAC-CTCAACCTCCAACCCCAACCCCA-ACCCC-AACC---CCA
            ** ** *  *  ***** * * **  * **    * ******* ***** ****   * *
HUMAN       CCGCC-CATCCTACCCAGCACACACACACCGCTGCTAACCCCATACCCCGAACCAACCAA

                     .         .
TTRSTELC    ACCCCAACTCCAACCCCAAC
            *******   ** **** **
HUMAN       ACCCCAAAGACACCCCCCAC
```

Fig. 4. Nucleotide similarity between (A) *T. thermophila* mitochondrial telomeric region (EMBL entry: MITTRA, nt 53–1) and rat D-loop (nt 1055–1124 in Fig. 1), (B) *T. thermophila* mitochondrial telomeric region (EMBL entry: MITTRA, nt 53–1) and mouse D-loop (nt 985–1095 in Fig. 1), (C) *T. thermophila* mitochondrial telomeric region (EMBL entry: MITTRA, nt 51–1) and cow D-loop (nt 84–IS in Fig. 1), (D) *T. thermophila* nuclear telomeric region (EMBL entry: TTRSTELC, nt 330–200) and the human D-loop (nt 1155–1304 in Fig. 1).

In cow the insertion is 66 nt long. Both the IS of human and cow and the SR elements present in the R domain of the rabbit are flanked by the small repeat, TACA(T). In humans the direct repeat flanking IS becomes 11 nt long: TAGTACATAAA. It is notable, moreover, that all these expansion elements are located near the cloverleaf-like structures (Fig. 2).

The TACAT pentanucleotide is present in the TAS elements of *Xenopus* (Dunon-Bluteau and Brun 1987). Interestingly, a TACA element is found within the SR of rabbit.

Relics of Telomeric Structures?

The asymmetry of GC distribution in the two DNA strands, there being a low G content in the L-strand, is already recognized as a peculiar feature of the mitochondrial genome of mammals (Gadaleta et al. 1989) and is particularly relevant in the R and L domains. This overall strand composition asymmetry, resulting in G-rich versus C-rich strands is characteristic of the telomere structures found in the linear chromosome of eukaryotes (Blackburn 1990). Unlike nuclear DNA, mtDNA is usually circular. However linear mtDNAs are found in two strains of yeast, ciliates, *Hydra, Chlamydomonas reinhardtii,* and in some plants. In the latter two cases the linear molecules seem to be derived from circular forms. In the ciliated protozoa *Tetrahymena* and *Paramecium* the linear mtDNA molecules (about 50 kb in size) have, like nuclear chromosomes, telomeric structures at their ends. In *Tetrahymena thermophila,* BVII mtDNA telomeres consist of a 53-bp sequence tandemly repeated from 4 to 30 times (Morin and Cech 1986).

We searched the D-loop of all mammalian species regions homologous to the mitochondrial telomeres of *Tetrahymena.* A significant degree of similarity was found in the sequences of rat, mouse, and cow (Fig. 4). In rat and mouse the similarity lies near CSB3 whereas in cow it is located in and around the IS element. Primates, instead, show extensive nucleotide similarity with the nuclear telomeric sequences common to various organisms (*Tetrahymena, Paramecium,* yeast, human) in a region spanning from the 5′ end of the right domain to the CSBs (Fig. 4D).

It is well known that telomere length is variable: in the nuclei of *Tetrahymena* length variability is governed by the action of the telomerase; in budding yeast telomere length fluctuates around a constant level during long-term culture, suggesting that there is a balance of additions and losses of repeats. Morin and Cech (1986) have suggested a recombination model for the maintenance of the telomeric repeats at the mtDNA linear ends in the mitochondria of *T. thermophila.* Recently, it has been reported that telomeric sequences can be lost in human tissues with age and in human tumors thus suggesting that telomerase is inactive in somatic cells. The loss of telomeric sequences can lead to end-to-end chromosome fusions in two ways: nontelomeric sequences can be exposed and then ligated to each

Table 2. Sequence length (L) and nucleotide frequencies of the left, central, and right domain of the D-loop-containing region

	Left domain					Central domain					Right domain				
	L	A	C	G	T	L	A	C	G	T	L	A	C	G	T
Pygmy chimp	370	34	34	10	22	237	25	31	19	25	514	30	35	12	23
Common chimp	369	33	35	11	21	235	26	30	19	25	511	30	35	13	22
Human	371	32	33	13	22	235	26	30	18	26	516	30	34	13	23
Dolphin	300	35	19	9	37	236	26	27	18	29	354	29	25	16	30
Cow	367	41	20	10	29	239	24	28	19	29	304	30	27	14	29
Rat	260	37	18	10	35	236	24	29	18	29	406	37	25	11	27
Mouse	260	37	20	9	34	239	26	29	17	28	380	37	25	11	27
Rabbit	389	37	27	7	29	238	28	30	16	26	1211	31	29	14	26

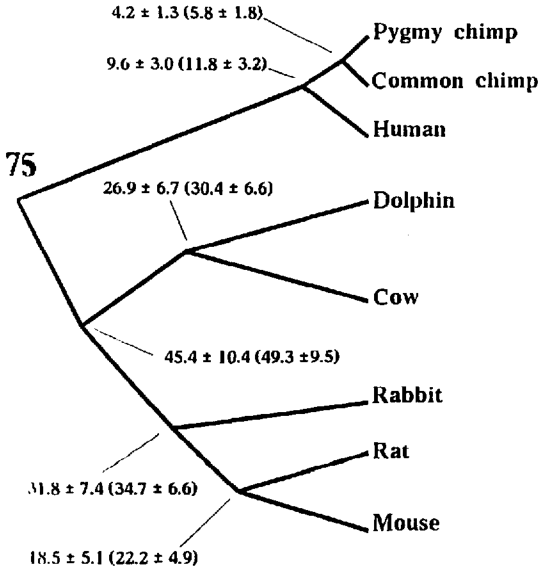

Fig. 5. Phylogenetic tree of the central domain of the D-loop region. The times of divergence (million of years) adopting the corrected and the uncorrected (in brackets) methodology are shown.

other or to telomeric sequences, or telomeric sequences may become ligated to each other (telomere–telomere fusion) (Murray 1990). With regard to the mtDNA of Metazoa, it could be envisaged that probably with the loss of recombination capacity and of the telomerase enzymes there was a progressive loss of telomeres followed by the circularization of the molecule. However, it has been suggested that the TnGm sequences at the 3′ ends of the chromosomes must be very efficient templates for DNA primase. According to Murray (1990), this consideration raises the heretical possibility that the telomeric repeat sequences were initially determined by the sequence preferences of DNA primase. In this case, the similarities found between telomeric sequences and particular regions of the mammalian D-loop could be due to convergent evolution.

Table 3. Expected Poisson distributions of nucleotide substitutions

Number of hit(s)	Observed	Expected[a]	Expected[b]
0	125	110.6	126.4
1	47	72.0	48.5
2	29	23.4	25.5
3+	11	6.0	11.7
χ^2		16.6[c]	0.6[c]

[a] All sites are assumed variable

[b] One hundred thirty two out of 212 sites are assumed variable

[c] The minimum χ^2 value between the observed and the expected distribution of nucleotide substitutions is calculated when 132 out of 212 sites are considered variable

Evolution of the Central Domain: Branching Order of Mammals

The central domain is one of the most conserved regions of the mammalian genome. It is characterized by a G content higher than in the other part of the molecule (Table 2), probably indicative of a functional constraint.

The best alignment of the central domain in the eight mammalian species, shown in Fig. 1, reveals that blocks of invariant sequences are separated by nucleotide stretches that display high similarity within the three groups: chimpanzee–human; dolphin–cow; rat–mouse–rabbit. The percentage of similarity in the mammalian species (Table 1) also shows that rabbit is closely related to rodents and dolphin to cow.

Unlike the L and R domains, which display base composition heterogeneity, the central domain obeys the stationarity conditions necessary for applying the stationary Markov model of Saccone et al. (1990). With our method, using as input the time of divergence between two species, we can calculate the time of divergence of all other species. In our case we fixed 75 million years between human and rodents as a reference time, this being the most reliable value obtained from paleontological data.

Assuming a model of homogeneous evolution of nucleotide sites, a satisfactory qualitative tree is obtained, but the estimates of the times of divergence between closely related species (like primates) are not highly consistent with data previously reported from other sources (Holmes et al. 1989). This discrepancy could be due to the presence of invariant sites, which invalidate our model of homogeneous evolution.

We have thus developed a correction method based on the following criterion. When the nucleotide substitutions are randomly distributed along the sequence, the number of observed substitutions for each site should follow a Poisson distribution. If as shown in Table 3 the observed distribution does not fit the Poisson distribution, using a χ^2 method we can determine the number of variable sites that produce the Poisson distribution that best fits the observed one. By using this fraction of variable sites we recalculate the times of divergence.

Figure 5 shows the tree constructed with and without (in brackets) this correction. The corrected estimates of the times of divergence between primates approaches the expected values obtained with other methods. We note that the time of divergence of rabbit from the other rodents is about 32 million years. The cow is more closely related to rodents than to human, confirming a result consistently found with other mitochondrial gene sequences. It is well known that the phylogeny of arctiodactyls, humans, and rodents is still a matter of controversy (Easteal 1987; Li and Wu 1987). In contrast with the highly consistent results obtained on mitochondrial genes, nuclear genes give with our method results that are statistically not significant (C. Saccone et al., unpublished). Thus the problem of the cow–rodent–human trichotomy remains an open question.

Acknowledgments. This work has been supported by Progetto Genoma Umano, Progetto finalizzato Biotecnologie e Biostrumentazione, and MURST, Italy.

References

Bennet JL, Clayton DA (1990) Efficient site-specific cleavage by RNase MRP requires interaction with two evolutionarily conserved mitochondrial RNA sequences. Mol Cell Biol 10: 2191–2201

Blackburn EH (1990) Telomeres: structure and synthesis. J Mol Biol 265:5919–5921

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

Doda JN, Wright CT, Clayton DA (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

Dunon-Bluteau DC, Brun GM (1987) Mapping at the nucleotide level of *Xenopus laevis* mitochondrial D-loop H strand: structural features of the 3' region. Biochem Int 14:643–645

Easteal S (1987) The rates of nucleotide substitution in the human and rodent lineages: a reply to Li and Wu. Mol Biol Evol 4:78–80

Easteal S (1990) The pattern of mammalian evolution and the relative rate of molecular evolution. Genetics 124:165–173

Fisher RP, Parisi MA, Clayton DA (1989) Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1. Genes & Dev 3:2202–2217

Foran DR, Hixson JE, Brown WM (1988) Comparison of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res 16: 5841–5861

Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisá 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

Holmes EC, Pesole G, Saccone C (1989) Stochastic models of molecular evolution and the estimation of phylogeny and rates of nucleotide substitution in the hominoid primates. J Hum Evol 18:775–794

Li W-H, Wu C-I (1987) Rates of nucleotide substitution are evidently higher in rodents than in man. Mol Biol Evol 4: 74–77

MacKay SD, Olivo PD, Laipis PJ, Hauswirth WW (1986) Template-directed arrest of mammalian mitochondrial DNA synthesis. Mol Cell Biol 6:1261–1267

Mignotte F, Gueride F, Champagne AM, Mounolou JC (1990) Direct repeats in the noncoding region of rabbit mitochondrial DNA: involvement in the generation of intra- and inter-individual heterogeneity. Eur J Biochem 194:561–571

Morin GB, Cech TR (1986) The telomeres of the linear mitochondrial DNA of *Tetrahymena thermophila* consist of 53 bp tandem repeats. Cell 46:873–883

Murray A (1990) All's well that ends well. Nature 346:797–798

Saccone C, Attimonelli M, Sbisá E (1985) Primary and higher order structural analysis of animal mitochondrial DNA. In: Quagliariello E, Slater EC, Palmieri F, Saccone C, Kroon AM (eds) Achievements and perspectives of mitochondrial research, vol II. Biogenesis. Elsevier, Amsterdam, p 37

Saccone C, Attimonelli M, Sbisá E (1987) Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. J Mol Evol 25: 205–211

Saccone C, Lanave C, Pesole G, Preparata G (1990) Influence of base composition on quantitative estimates of gene evolution. Meth Enzymol 183:570–583

Sbisá E, Nardelli M, Tanzariello F, Tullo A, Saccone C (1990) The complete and symmetric transcription of the main noncoding region of rat mitochondrial genome: in vivo mapping of heavy and light transcripts. Curr Genet 17:247–253

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

Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421