Volume 247, number 2, 317–322 FEB 07050 April 1989

Nucleotide sequence of *Suncus murinus* immunoglobulin μ gene and comparison with mouse and human μ genes

Hiroshi Ishiguro, Yoshikazu Ichihara, Takao Namikawa*, Toshiharu Nagatsu+ and Yoshikazu Kurosawa

Institute for Comprehensive Medical Science, Fujita-Gakuen Health University, Toyoake, Aichi 470-11, **Faculty of Agriculture, Nagoya University, Chikusa-ku, Nagoya 464* and *+Department of Biochemistry, Nagoya University School of Medicine, Tsurumai, Showa-ku, Nagoya 466, Japan*

Received 18 February 1989

Both primates and rodents apparently originated from insectivores and then evolved separately. We isolated the immunoglobulin μ gene from DNA of the insectivore *Suncus murinus* and determined its nucleotide sequence. The gene organization was CH1 exon (318 bp)-intron (89 bp)-CH2 exon (345 bp)-intron (80 bp)-CH3 exon (318 bp)-intron (85 bp)-CH4 exon (392 bp)-coding sequence and 3′-untranslated region. Comparison of nucleotide sequences of μ genes between mouse, human and *Suncus murinus* indicated that the evolutionary distance between human and mouse is equal to that between *Suncus murinus* and human, and that mouse is deviated further from *Suncus murinus* than the two other combinations. This conclusion was further supported by sequence comparison of non-coding regions.

Immunoglobulin μ gene; Reverse genetics; (*Suncus murinus*)

1. INTRODUCTION

Immunoglobulin (Ig) heavy (H) chain gene loci have been extensively characterized in mouse and human (for review see [1]). Several functional sequences have been identified in other than coding regions: signal oligomer sequences for $V_HD_HJ_H$ joinings [2], octamer sequences in promoter regions [3], enhancer sequence for transcription [4], and characteristic repetitive sequences for class switch (so-called S sequences) [5]. However, many problems with other genetic phenomena observed in IgH gene loci have not yet been solved in molecular terms. When nucleotide sequences of different species, such as human and mouse, are compared, homologous sequences are sometimes seen outside of coding regions. These kinds of homologous sequences have functional significance. In fact, many of the above functional segments were originally identified as conserved sequences by sequence comparison. However, available sequences are, in many cases, restricted to coding regions and their surrounding regions of only a few species. We previously reported homologous sequences in human and mouse [6], but it is difficult to discriminate, by comparison of only two species, between biologically meaningful sequence conservation and an accidental coincidence. For such comparisons to be informative, a choice of animals is crucial. Differences of nucleotide sequences in non-functional regions are large enough, and those of functional regions are convincingly small. Both primates and rodents apparently originated from insectivores and evolved separately [7]. In this study, we isolated the Ig μ gene from DNA of the insectivore *Suncus murinus*, and determined the nucleotide sequence. We also

Correspondence address: Y. Kurosawa, Institute for Comprehensive Medical Science, Fujita-Gakuen Health University, Toyoake, Aichi 470-11, Japan

Abbreviations: Ig, immunoglobulin; H, heavy

The nucleotide sequence presented here has been submitted to the EMBL/Gen Bank database under the accession no. Y07500

Published by Elsevier Science Publishers B.V. (Biomedical Division)

discuss the evolutionary relationships between mouse, human and *Suncus murinus*.

2. MATERIALS AND METHODS

Suncus murinus is maintained at Nagoya University animal center. High molecular mass DNA was extracted from a liver of *Suncus murinus* according to the published procedure in [8]. 12 kb *Eco*RI DNAs were separated by agarose gel electrophoresis, ligated with EMBL4 phage arms, packaged into λ phages, and screened with human C_μ probe by the Benton-Davis' method [9]. A mixture of two *Eco*RI fragments (1.2 kb and 0.95 kb) containing human C_μ genes was prepared from clone 6 described in a previous paper [6], and used as C_μ probe. The nucleotide sequence of the C_μ-coding region was determined by the dideoxy method [10].

3. RESULTS AND DISCUSSION

3.1. *Cloning and sequencing of the C_μ gene of* Suncus murinus *DNA*

Human C_μ gene-containing fragments were used as probe (C_μ probe) to identify the C_μ gene of *Suncus murinus*. Southern hybridization of *Eco*RI-digested *Suncus murinus* DNA with the C_μ probe gave a single band at 12 kb. We cloned this band as described in section 2 and named it λSUNM-1. Fig.1 shows the restriction map of clone λSUNM-1. The nucleotide sequence of the C_μ-coding region was determined. Fig.2 shows the total nucleotide sequence of the C_μ gene and the surrounding region. Nucleotide sequences of mouse and human μ-coding regions are given for comparison [11-13]. The gene organization is CH1 exon (318 bp) - intron (89 bp) - CH2 exon (345bp) - intron (80 bp) - CH3 exon (318 bp) - intron (85 bp) - CH4 exon (392 bp)-coding sequence and 3′-untranslated region.

3.2. *Comparison of μ-coding regions in mouse, human and* Suncus murinus

Only parts of the nucleotide sequence of the human C_μ gene have been published [12,13], so we compared the amino acid sequences of μ chains of *Suncus murinus*, mouse and human [14]. Table 1 summarizes the amino acid residues common to two animals: 54, 62 and 62% of amino acid residues in combinations of *Suncus murinus* and mouse, *Suncus murinus* and human, and mouse and human, respectively. 47% of amino acid residues are common to all three animals. These data suggest that the evolutionary distance between human and mouse is equal to that between *Suncus murinus* and human, and that mouse is deviated further from *Suncus murinus* than the two other combinations. This relationship among these three aminals is supported by available data on nucleotide sequences of μ genes; 65, 72 and 74% of available 912 bp-coding sequences are identical in the above respective combinations (table 1).

Amino acid sequences of α- and β-globins of *Suncus murinus* have been published [15]. Although sequences of globin genes are conserved much better than those of C_μ genes, evolutionary relationships between mouse, human and *Suncus murinus* are similar to the above observations. In

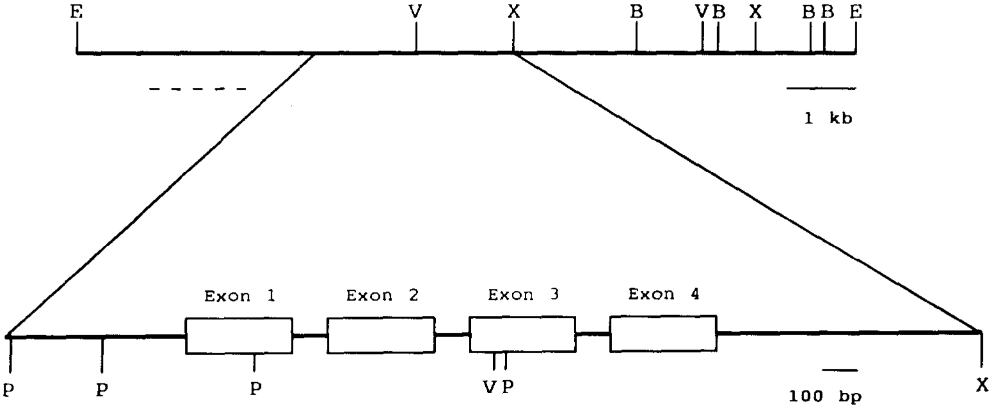

Fig.1. Restriction map of λSUNM-1. *Abbreviations:* E, *Eco*RI; V, *Eco*RV; X, *Xba*I; B, *Bam*HI; P, *Pst*I. Putative S_μ region is indicated by a broken line.

```
Suncus
Mouse
Human
                                                                                                                                120
GGGGAAAACC TCTGCAGAAG AAAATGGATG TTTTATTTTG AGATGATTTT CTCTGCACTG CACAATCTTT AACTAATTTA AACCTGCTGC CACTCTCCTA GAGTGCAAGG ACCTGTTCTT
                                                                                                                                240
GAGCCTTCTG TGTAGGTGGA GGGACTGTCT CCTGACTCCC TGATGCACAG TATGGGGAAG GGGTTTGAAC TGGACAGCTG TGTTGGGGGC TGGGGCCCTC TCAGTCCTAT GTGACCAGGT
                                                                                                                                360
TTAAGTGACC ATTGATTGCT GCGGCTGGAG CCTGAAGAGC CCCTTCAGTC CTGCAGTGCT GGGAGTGGGC CAGGCCAACC AGTGGGGAGC TCCAGGGTGG TAATTGGTGA TGCTGAAAGG

 S  S  S    A  P  L    L  F  P   L  V  S  C    D  S  S    L  P  D   E  T  Q  V    T  L  G    C  L  A   R  D  F  L    P  R  P    V  T  F
AGTCCAGCTC AGCTCCACTC CTCTTCCCTC TTGTCTCCTG TGATTCATCA CTGCCGGATG AGACACAAGT GACCCTGGGC TGTCTGGCCC GTGACTTTTT GCCCCGGCCT GTCACCTTCT
--AGTCAG-- CTTC---AAT G-------C- -C-------- C--GAGCC-C ---T-T---A ---AT-TG-- -G--A----- --C------- -G-----CC- ----A-CA-C A-TT-----A
G-AGTGCA-- C--C---AC- --T-----C- -C-------- ---GAAT--C
                                                                                                                                720
S  W  K  F    K  N  S    S  S  I   S  S  Q  N    I  Y  N    F  P  E   V  F  T  G    G  K  Y    M  A  T   S  Q  V  L    L  P  S    T  A  I
CCTGGAAATT CAAGAACAGC TCCAGCATCA GCAGCCAGAA CATCTACAAT TTCCCTGAGG TCTTCACTGG TGGCAAGTAC ATGGCCACAT CCCAGGTGCT GCTGCCATCC ACGGCCATCC
-------C-A -C------A- A-TGAAG--* **-T----GG T---AGA-CC -----AACAC -GAGG--A-- G--------- C-A-----C- -G-----CT- ----T-TC-- -A-AG-----

L  Q  S  T    D  D  Y    I  T  C   H  T  K  H    T  T  G    E  K  E   K  K  V  E    L  Q                                       840
TGCAGAGCAC CGATGATTAC ATAACGTGCC ATACCAAGCA CACCACAGGG GAAAAAGAAA AGAAAGTAGA ACTACAAGGT AAGCCCGAGA GCAGGCACCA AGTTCCTGGG TCCTCCCTGG
-TG-AG-TT- A-----A--- C-TGTA---A -A-T-C-CT- -GGAGGCAAA A-C-G---TC T-C-T--GCC CA-T-C--

                                                                                                                                960
                                                V  T  P  E  L    P  P  N    V  S  I   F  V  P  P    R  N  S    F  S  G   N  H  P  R
TTGCCTGAGG GTCCCCCCAT CAAGTTGACT CCATGTTCTC ACCACAGTAA CCCCAGAGCT GCCTCCGAAT GTGAGCATCT TCGTTCCGCC CCGAAACTCC TTCTCCGGCA ACCACCCACG
                                                   CTG T-G-----A- -AA---C--- --A-ATG-G- ----C--A-- A--GG-TGG- -----T---C CTGCA-----
                                                   -G- TTG-T----- ------C--A ------G--- ----C--A-- ---CG--GG- ----T----- --***--C--
                                                                                                                               1080
 T  S  Q    L  I  C   Q  A  S  G    F  S  P    R  T  I   V  M  S  W    L  Q  R    G  E  P   V  Q  P  S    L  V  S    T  S  A   V  E  A  E
CACGTCCCAG CTTATCTGTC AGGCCTCTGG CTTCAGCCCC CGCACCATAG TCATGTCCTG GTTGCAGCGG GGTGAACCAG TGCAGCCCAG CCTAGTGTCA ACCAGTGCTG TGGAGGCTGA
--A---TA-A --C-----CG -----A-GAA -----CT--A AAAC-G--CA CAG-A----- -C-AA--GAT --GA-G-TC- --G-A***TC TGGCT-CA-C --AGA-C-G- --ACCATC--
--A----A-- --C-----C- -----A-G-- T-----T--- --GCAG--TC AGG------- -C---GCGA- --GA-G--C- --GG-***TC TGGC--CA-C --GGACCAG- --C-------
                                                                                                                               1200
 P  K  G    S  G  P   T  T  F  R    V  I  S    R  L  T   I  T  E  N    E  W  L    S  Q  R   E  F  T  C    Q  A  L    H  K  G   L  T  F  Q
GCCCAAGGGC TCGGGGCCCA CTACCTTCAG GGTCATCAGC AGACTGACCA TCACCGAGAA TGAATGGCTC AGCCAGAGAG AGTTCACCTG CCAAGCACTC CACAAGGGTC TCACCTTCCA
-AA---A--A --CACA---C AA----A--A ------A--- -C---T---- --T-T--A-T C--C-----G -A--T--AT- T--A------ --GT-TGGAT ----G----- --------TT
-G----A-AG --T------- -G----A--A ---G-C---- -C-------- ---AA---                                           GAT ----G---C- -G--------
                                                                                                                               1320
 K  N  V    S  S  V   C  M  G  D                                                                                            D  T  S
GAAGAATGTT TCATCTGTGT GCATGGGCGA TGGTGAGTCC CACTGTCCCC TGGGAGGATG GGTGGGGGTC ATGTGGGATG GCCATTGAGT GCATCCTCTC TCCTGCCTCC AGACACTTCC
------C--G --C--CACA- -TGCT-C-** *A                                                                                        GTC-C---
-C------CG --C--CA--- -TG-CCC-** *-                                                                                        -TCAAGA-
                                                                                                                               1440
 T  G  I   S  V  F  L    L  P  P    T  F  A   N  I  F  L    T  Q  S    A  Q  L   T  C  L  V    T  G  L    A  T  Y   D  S  L  D    I  S  W
ACCGGAATCA GTGTGTTCCT GCTGCCCCCC ACCTTTGCCA ACATCTTCCT CACCCAGTCG GCCCAGCTGA CCTGCCTGGT CACGGGGCTG GCCACCTACG ACAGCCTGGA TATCTCATGG
--A-AC---C TAACC---AC CA-C------ T--------G ---------- --G-A----C --TA-C---- ----T----- -T-AAAC--- --A-----T- -A-C----A- ------C---
--A-CC---C -G--C---GC CA-C-----A T--------- G--------- G---A----C A--A--T--- ---------- ---A-AC--- A-------T- -----G--AC C-----C---
                                                                                                                               1560
 S  R  Q   N  G  E  A    L  Q  T    H  V  N   I  S  E  S    H  P  N    S  T  F   T  A  K  G    H  A  S    V  C  R   E  E  W  E    S  G  E
AGCCGCCAGA ATGGGGAGGC CCTGCAGACA CACGTCAACA TCTCCGAGAG TCACCCCAAC TCAACCTTCA CCGCCAAGGG CCACGCCAGC GTGTGCCGCG AGGAATGGGA ATCAGGGGAG
GCTTCT--A- G---T--AC- A---G-A--C A-AA-T--A- --ATG--A-- C--T-----T GGC------- GT--T----- TGTG--T--T --T--TGTG- -A--C---A- TAACA--A-G
                                                                                                                     -A- T--G------

 K  F  T   C  T  V  Q    H  S  D    L  P  S   P  L  K  Q    S  L  S    R  P  K                                                 1680
AAGTTCACCT GCACTGTGCA GCACTCCGAC CTGCCCTCAC CACTGAAACA GAGCCTCTCC CGGCCCAAGG GTATGTACCT GCCTCACCCT GCCTGTCCTG CCCTGTGCAT CTCCCCAACC
G-A--TGTG- -T------AC T---AGG--T -----T--G- ---A---GA- ATT-A----A AAA-----T-
-G------G- ----C---AC C---A-A--- --------G- -------G-- --C-A----- ----------
                                                                                                                               1800
                                      D  V    A  N  D   P  P  S  V    F  V  L    P  P  A   Q  E  Q  L    K  L  R    E  S  A   S  I  T  C
CAGCCCATGA CACCCACCCT GCCTTGCCTC CACAGATGTG GCGAACGACC CGCCTTCTGT GTTCGTGCTG CCACCTGCCC AGGAGCAACT GAAGTTGCGT GAGTCTGCCT CTATCACCTG
                                      -G--- CAC--AC-T- -A---G---- -A-C------ -----A--T- GT-------- ---CC-AA-A -----A---A -AG-------
                                                                                          -----G-- ---CC----G -----G---A -C-----G--
                                                                                                                               1920
 L  V  K    D  F  S   P  P  D  V    F  V  Q    W  Q  H   H  G  Q  P    V  D  P    K  H  Y   V  T  S  N    P  T  P    E  P  Q   N  P  G  L
CCTGGTGAAG GACTTTTCGC CACCGGATGT TTTCGTCCAG TGGCAGCATC ACGGGCAACC TGTGGACCCC AAGCATTATG TGACCAGCAA CCCTACACCT GAGCCCCAGA ACCCTGGCCT
---------- -G---C--T- -TG-A--CA- CAGT--G--- ----TT--GA GA-------T CT--CC--AA G--A-G---- -------TGC ---G-TG--A -----TGG-G C---A---T-
--------C- -G---C--T- -CG----C-- C-----G--- ---AT---GA GG-----G-- CT--TC---G G--A-G---- --------GC ---A-TG--- ---------G C---A----G
                                                                                                                               2040
 Y  F  V    H  S  I   L  T  V  S    E  K  D    W  S  S   G  E  S  F    S  C  V    V  G  H   E  A  L  P    L  S  V    T  E  K   A  V  D  K
CTACTTCGTC CACAGCATCC TGACGGTGAG CGAGAAGGAC TGGAGCTCTG GAGAGTCCTT CTCCTGCGTC GTGGGCCACG AGGCCTTGCC CCTGTCAGTG ACCGAGAAGG CTGTGGACAA
------TAC- ---------- ----T----C A---G----A ----A---C- -----A---A TA-------T --A--G---- -----C---- A-ACCT---- -------G-A -C--------
G-------C- ---------- ----C---TC ----G----A ----A-A-G- -G---A---A -A-----*** ----C---T- -C---C---- -AACAGG--C -------G-A -C--------

 T  S  G    K  P  T   L  Y  N  V    S  L  V    L  S  D   T  A  S  T    C  Y                                                    2160
GACATCTGGT AAACCCACCC TGTACAACGT GTCTCTCGTC CTGTCTGACA CCGCCAGCAC CTGCTACTGA CCCCACCACC CTGTCTCTGT GCATGCAAGT GAGCCAGCCC TCTGAGCTTG
-T-CA----- --------A- -------T-- C--C--GA-- A--------- -A-G-G---- ------T
-T-CA-C--- ---------- ---------- ---C--G--- A----C---- -A--TG---- -------


                                                                                                                               2280
GCCACACAGG CTGTGAGCCA CGGAATCAGA AATAAAAAGA TCCGTTGAAA AGATGCTGAG TTAGCAGTGC TCACACCTGT GCTGTCCCCT GCCCTACTCC AGGCACCCAG CTTCTGGCCT
                                                                                                                               2400
CTCTTGGGAC CCCCTGCTGC TTACAGAGTG CTTGCAGCTC ACAGGGCCCA CTAGCTGCTC TCTCCTCTCA CTGCTTTTAT TGGGGACCTG GGGCCTGCCA CACAAGGCGC ACGAGCATGC
                                                                                                                               2520
ACAAACACAC TCAGGCATGT GTATTGCATG TGTATACAAG CTAAACACAA TCACACACTC AGAGGTGCAT AGAGATTACA CATTTGAACA TAGACATACT TAGAGATATA CAGGTATACT
                                                                                                                               2640
TATAGAGTCA TACAAGCATA TGGAACAGGT TGTAGAGTAT ACAGACATGC TCACATATGC ACAGCGATGT ACTTATCTGC CCCTCACATA CGTGAGCATT CAGGTATGTG GAGCTTGTGT
                                                                                                                               2760
AAACAGAGTC ATGTTCATAG GCACATGAAA ATTACATATT TGATACAGAT ATATACAGAT ATGCACAGAC ATATGCCACT CACACTGAAC GCACAAACAT ACAACTTACA CAGTAGCGCA

GATGCAGACC CATTTGTGTG TGCACATGTG TGCACTGTAA TTTACACATG TGGGCACTAT ACAGACACTC TCATATCTAG A
```

Fig.2. The nucleotide sequence of the C_{μ} gene of *Suncus murinus*. Amino acid residues deduced from nucleotide sequences are indicated along the top. The nucleotide sequences of μ-coding regions of mouse [11] and human [12,13] are shown for comparison. Dashes indicate the same nucleotide as that in *Suncus murinus*. Asterisks indicate missing nucleotides.

Table 1

Comparison of amino acid and nucleotide sequences of μ-coding regions in mouse, human and *Suncus murinus*

	Amino acid length			Su/Mo	Su/Hu	Hu/Mo	Su/Mo/Hu	Number of Nucleotides	Su/Mo	Su/Hu	Hu/Mo	Su/Mo/Hu
	Su	Mo	Hu									
CH1	105	105	103	45 (43%)	58 (55%)	48 (46%)	35 (33%)	50	27 (54%)	31 (62%)	36 (72%)	25 (50%)
CH2	115	113	112	58 (50%)	69 (60%)	71 (63%)	52 (45%)	306	174 (57%)	205 (67%)	210 (69%)	153 (50%)
CH3	106	106	105	60 (57%)	68 (64%)	57 (54%)	50 (47%)	211	137 (65%)	160 (76%)	152 (72%)	123 (58%)
CH4	131	131	130	83 (63%)	88 (67%)	106 (81%)	78 (60%)	345	251 (73%)	262 (76%)	281 (81%)	232 (67%)
TOTAL	457	455	450	246 (54%)	283 (62%)	282 (62%)	215 (47%)	912	589 (65%)	658 (72%)	679 (74%)	533 (58%)

Left half: comparison of amino acid sequences in three species. Numbers of amino acid residues commonly observed between Su/Mo, Su/Hu and Hu/Mo, and among the three animals are indicated. Right half: comparison of nucleotide sequences in the three animals. Since only the partial sequence of the human C_μ gene has been published, the data are based on the available regions

141 amino acid residues of α-globin chain, 121, 120 and 121 residues are common in the above respective combinations, and in 146 amino acid residues of β-globin chain, 115, 123 and 117 residues are common.

3.3. *Comparison of non-coding regions of C_μ genes in mouse, human and* Suncus murinus

Nucleotide sequences upstream of C_μ genes responsible for the class switch have been published for mouse and human [12,16]. They are rich in tandem repeated sequences. In mouse, the basic unit of the repeats is GGG$\underline{\text{G}}$T(GAGCT)$_3$ [16]. In human, GGG$\underline{\text{C}}$T(GAGCT)$_n$ is common [12]. We determined partially the nucleotide sequence upstream of the C_μ gene in *Suncus murinus*. Switch sequences were distributed in sequences longer than 2 kb. As shown in fig.3a, they are rich in GGG$\underline{\text{C}}$T, instead of GGG$\underline{\text{G}}$T, in addition to GAGCT. The tendency for the mouse sequence to deviate further from the *Suncus murinus* sequence than those of the two other combinations is also observed in other regions. Although the sizes of introns of the *Suncus murinus* C_μ gene are relatively constant and short compared with mouse and human genes [11,12], splicing donor and acceptor sequences in intron 2 of *Suncus murinus* is more similar to human than mouse (fig.3b). Fig.2c compares poly(A) addition signal and its surrounding regions. The order of sequence similarities is Hu/Mo > Hu/Su > Su/Mo.

3.4. *Reverse genetics of non-coding regions*

In this study, we planned to establish a general method, 'reverse genetics of non-coding regions', i.e. search for conserved sequences across species to identify unknown functional regions. For such comparisons to be informative, a choice of species is crucial. It is generally accepted that both primates and rodents originated from insectivores and evolved separately [7]. Although *Suncus murinus* is an insectivore, it must also have changed from the original insectivore. To explain the above mentioned sequence conservations among the three animals, there are three possibilities (fig.4). Case 1: these three animals had a common ancestor and the nucleotide sequence of the human C_μ gene has changed more slowly than in the other two animals. Case 2: rodents deviated from insectivore earlier than primates in the evolutionary process: i.e. the nucleotide sequence of the *Suncus murinus* C_μ gene has changed more rapidly

```
(a)
GAGCT GGACT GAGCT GGGA  GAGCT GAGCT GGGCT GAGCT GAGCT GGGCT  (49)

GAGCT GGGT  GAACT GAGCT GGGCT GAGCT GGGA  GAGCT GAGCT GGGCT  (97)

AGACT GAGCT AGGCT GAGCT AGGT  GAGCT GAGCT GGGCT GAGCT GAGCT (146)

GGGT  GAGCT GAGCT GGGCT GAGCT GGTT  GAGCT GGGCT GAGTT GGGT  (193)

GAGCT GGACT GAGCT GGGT  GAGCT GGGTT GGGTT GAGCT GAGCT GGGA  (241)

GAACT GAGCT GGACT GAGTT GGGTT GAACT GGGCT GAGC  GGGAT GCACT (290)

GGGCT GAGCT GAGCT GGGT  GAGCT GGGGT GAGCT AGGT  GAGCT GAGCT (338)

GGGCT GAGCT GGGTT GAACT GGCCT GAGCG GGGT  GAGCT GAGCT GGGCT (387)

GAGCT GGGT  GATCT GAGCT GGGCC GAGCT ATGCT GGGCT GAGCT GAGCT (436)

GGGT  GAGCT GAGCT GAGCT GAGCT GAATT GGACT AAACT             (475)
(b)
Human   GTGAGTGACCTGTCCCCAGGGGCAGC(188)ACCTCTGACTCCCTTCTCTTGANTCCAG
Mouse   GTGAGTGG-CTGGGATAAGCCCAATG(226)ACCTGACACAGTGTTCTCTTGACTGCAG

Human   GTGAGTGAC-CTGTCCCCAGGGGCAG(189)ACCTCTGACTCCCTTCTCTTGANTCCAG
Suncus  GTGAGTCCCACTGTCCCCTGGGAGGA( 26)TTGAGTGCATCCTCTCTCCTGCCTCCAG

Suncus  GTGAGTCCCACTGTCCCCTGGGAGGA( 26)TTGAGTGCATCCTCTCTCCTGCCTCCAG
Mouse   GTGAGTGGCTGGGATAAGCCCAATGC(225)ACCTGACACAGTGTTCTCTTGACTGCAG
(c)
Human   TCTTATAAAA-TTAGAAATAAAAAGA-TCCATTCAAAAGATACTGGT
Mouse   TTTTATAAAAATTAGAAATAAAAAAAATCCATTCAAACGTCACTGGT

Human   TCTTATAAAATTAGAAATAAAAAGATCCATTCAAAAGATACTGGTC
Suncus  AGCCACGGAATCAGAAATAAAAAGATCCGTTGAAAAGATGCTGAGT

Suncus  AGCCACGGAATCAGAAATAAAAAGA-TCCGTTGAAAAGATGCTGAG
Mouse   TTTATAAAAATTAGAAATAAAAAAAATCCATTCAAACGTCACTGGT
```

Fig.3. Nucleotide sequences of non-coding regions. (a) Putative S_μ sequence from *Suncus murinus* located several kb upstream of the C_μ gene. (b) Comparison of splicing donor and acceptor sequences in intron 2 of all three species. Commonly observed nucleotides are underlined. Numbers indicate nucleotide length between two sequences. Dashes indicate missing nucleotides. (c) Comparison of poly(A) additional sequences and their surrounding regions in all three species.

than in the other two animals. Case 3: rodents and primates were derived from a common ancestor, and *Suncus murinus* was derived from an older insectivore: i.e. the nucleotide sequence of the mouse C_μ gene has changed more rapidly than in the other

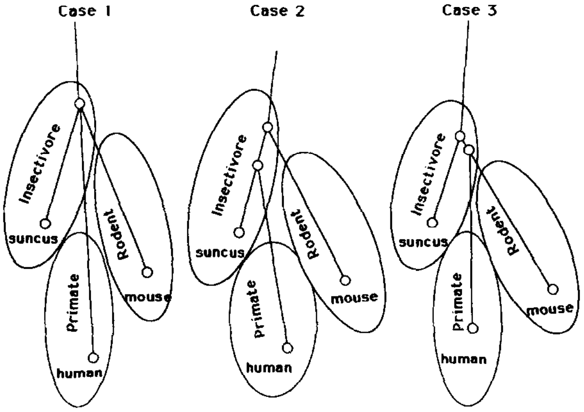

Fig.4. The evolutionary relationships between mouse, human and *Suncus murinus*. Case 1: mouse, human, *Suncus murinus* were derived from a common ancestor. Case 2: mouse was derived from an insectivore older than a common ancestor for *Suncus murinus* and human. Case 3: *Suncus murinus* was derived from an insectivore older than a common ancestor for mouse and human.

two animals. In any case, it is quite unlikely that some functional regions not possessed by insectivores could have evolved independently in both rodents and primates. Therefore, comparison of nucleotide sequences in these three species should be informative. Biologically meaningful regions, if they exist, should be observed in *Suncus murinus*, too.

Acknowledgements: We are grateful to Dr S. Ohno for suggesting this study. We thank Ms A. Nagata for preparation of the manuscript. This work was supported in part by grants from the Ministries of Education, Science and Culture, Health and Welfare, and Agriculture, Forestry and Fisheries in Japan; Fujita-Gakuen Health University; Japan Private School Promotion Foundation; Uehara Science Foundation.

REFERENCES

[1] Honjo, T. (1983) Annu. Rev. Immunol. 1, 499-528.

[2] Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979) Nature 280, 288-294.

[3] Parslow, T.G., Blair, D.L., Murphy, W.J. and Granner, D.K. (1984) Proc. Natl. Acad. Sci. USA 81, 2650-2654.

[4] Gillies, S.D., Morrison, S.L., Oi, V.T. and Tonegawa, S. (1983) Cell 33, 717-728.

[5] Kataoka, T., Miyata, T. and Honjo, T. (1981) Cell 23, 357-368.

[6] Akahori, Y., Handa, H., Imai, K., Abe, M., Kameyama, K., Hibiya, M., Yasui, H., Okamura, K., Naito, M., Matsuoka, H. and Kurosawa, Y. (1988) Nucleic Acids Res. 16, 9479-9511.

[7] Colbert, E.H. (1969) Evolution of the Vertebrates, 2nd edn, Wiley, New York.

[8] Gross-Bellard, M., Oudet, P. and Chambon, P. (1973) Eur. J. Biochem. 36, 32-38.

[9] Benton, W.D. and Davis, R.W. (1977) Science 196, 180-192.

[10] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

[11] Kawakami, T., Takahashi, N. and Honjo, T. (1980) Nucleic Acids Res. 8, 3933-3945.

[12] Rabbitts, T.H., Forster, A. and Milstein, C.P. (1981) Nucleic Acids Res. 9, 4509-4524.

[13] Takahashi, N., Nakai, S. and Honjo, T. (1980) Nucleic Acids Res. 8, 5983-5991.

[14] Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M. and Gottesman, K.S. (1987) Sequences of proteins of immunological interest. U.S. Dept. Health and Human Services, Washington, DC.

[15] Maita, T., Matsuda, G., Takenaka, O. and Takahashi, K. (1981) Hoppe-Seyler's Z. Physiol. Chem. 362, 1465-1474.

[16] Nikaido, T., Nakai, S. and Honjo, T. (1981) Nature 292, 845-848.