

## Short Communication

# Molecular Cloning and Phylogenetic Analysis of Canine $\beta$ -Casein

MANABU WATANABE<sup>ab</sup>, SUMIO SUGANO<sup>b\*</sup>, TAKUSHI TOGASHI<sup>b</sup>, JUN-ICHI IMAI<sup>b</sup>,  
KAZUYUKI UCHIDA<sup>a</sup>, RYOJI YAMAGUCHI<sup>a</sup> and SUSUMU TATEYAMA<sup>a</sup>

<sup>a</sup>Department of Veterinary Pathology, Faculty of Agriculture, Miyazaki University, Miyazaki 889-2155, Japan and <sup>b</sup>Department of Virology, Institute of Medical Science, The University of Tokyo, Tokyo, 108-8639, Japan

(Received March 11, 2000)

A canine  $\beta$ -casein cDNA was isolated from mammary tissue by polymerase chain reaction (PCR) using degenerate primers. It encodes 250 amino acids protein containing the conserved sequence motif of  $\beta$ -casein. It showed the highest homology with snow-leopard (*Uncia uncia*) (55–62 % identity). It also showed 44–53% identity with human, 33–42 % identity with mouse, 29–37 % identity with rat, 43–53 % identity with rabbit, 41–48 % identity with pig, 44–51 % identity with cattle and 44–50 % identity with sheep. A 1.2-kb mRNA was detected in mammary tissue by Northern blot analysis. Phylogenetic analysis revealed that canine  $\beta$ -casein formed a branch with lesser panda and snow leopard, which were grouped into carnivore.

**Keywords:**  $\beta$ -casein, degenerate oligo primer, dog, RT-PCR

**Database Accession No.:** AB035080

Caseins are a member of phosphoproteins which form the major component of milk proteins. They are divided into four kinds, which are  $\alpha_{s1}$ -,  $\alpha_{s2}$ -,  $\beta$ -, and  $\kappa$ -casein (Mercier et al., 1990).  $\beta$ -Casein and another three caseins form large aggregates which termed milk micelles. The

aggregates increase the solubility of calcium phosphate in milk and provide to newborn infants with well-balanced nutrients from mother (Lonnerdal et al, 1985).  $\beta$ -Casein genes have been cloned in various mammals (Lonnerdal et al, 1990, Gatesy et al, 1996), but canine  $\beta$ -casein sequence is not reported.

Here, we report the cloning of canine  $\beta$ -casein cDNA and the phylogenetic analysis of various species of mammals using their amino acids sequences.

We cloned canine  $\beta$ -casein cDNA using degenerate primer-based PCR from canine mammary tissue. The sequence of the primers used PCR reactions was based on conserved nucleotide sequence of  $\beta$ -casein from human, mouse, rat, rabbit, pig, cattle and sheep (5'-ATGAAGITCATICTIGCCTG-3' and 5'-CAAATGACTGGIAGGAAAT-3', 1; inosine). The PCR conditions were: 1 cycle of 95C for 2min, 40 cycles of 95C for 30 sec, 44C for 1min

\* Address for correspondence: Department of Virology, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. Telephone: +81-3-5449-5286. Fax: +81-3-5449-5416. E-mail: ssugano@ims.u-tokyo.ac.jp.

ATCATCCACTTAGCTTCCCCTTCACTTTCTGTACTCCACCTTGGAAAAAAGGACTTGATA	60
ACCATGAAGGTCTTCATCCTGGCCTGCCTTGTGGCTCTTGCTCTTGCAAGAGAGAAGGAA	120
<u>M K V F I L A C L V A L A L A R E K E</u>	19
GAACTCACTCTATCCAATGAGACTGTGGAAAGCCTTCAAGCAGTGAGGAATCTATTACA	180
E L T L S N E T V E S L S S S E E S I T	39
CACATCAACAAGCAGAACTTGAGAATTTTAAACACGAGGAGCAACAGCAAAGAGAGGAT	240
H I N K Q K L E N F K H E E Q Q Q R E D	59
GAAAGCCAGAATAAAAATCCACCCCTTTTCCAGCAACAGCCTCTAGTCTCTCCTTATGCT	300
E R Q N K I H P L F Q Q Q P L V S P Y A	79
GATCCCATCCACTATGCTATCCTTCCACAGAACATCCTGCCTCTGGCTCAGCCTGCTGTG	360
D P I H Y A I L P Q N I L P L A Q P A V	99
GTGGTGCCTTTCCTTCAGCCTGAAATAATGGAAGTCCCTAAAGTTAAGGAAAACATCTTT	420
V V P F L Q P E I M E V P K V K E N I F	119
CCTAGGCACAAAGTAATGCCCTTCTTAAATCTCCAGTGACACCCTTCTTGGACAGCCAA	480
P R H K V M P F L K S P V T P F L D S Q	139
ATCCTGAATGTGGCTGATCTTGAAAATGTGCATTTTCCCTCTCCTCTATCTCTGCCTCTG	540
I L N V A D L E N V H F P L P L S L P L	159
CTCCAGCCTCTGATGCACCAGATCCCTCAGCCTCTTCCCTCTGCTCCAGCCTCTGATGCAC	600
L Q P L M H Q I P Q P L P L L Q P L M H	179
CAGATCCCTCAGCCTCTTCCCTCAGACTCCTATGCTTACTCCTCAGTCAGTGCTGTCTATC	660
Q I P Q P L P Q T P M L T P Q S V L S I	199
CCACAGCCCAAAGTCTTGCCTTTTCCCTCAGCAAGTGGTGCCTACCTCCAGAGAGATATG	720
P Q P K V L P F P Q Q V V P Y L Q R D M	219
CCCCTACAAGCCTTCTGCCATAACCAGGAATCTACCCACCAGGCCCAAACCTGTGACTCAA	780
P L Q A F L P Y Q E S T H Q A Q P V T Q	239
CCACTTGCCCCACTTGTCAACTCTGCTCTTGTTTAAAAGAATTCCAAAGTTAATGTCCCC	840
P L A P L V N S A L V *	250
TCCTCACTTTTGAATTGACTGTGACTGGAAACGTGGCATCTTTTCACTTGGCATCATGTT	900
ACCAAAAATTAATAATTTTAAATGAACCTACATGGAAAAAATGAAAATTTGTTTCTTT	960
ATTTATTTTATGTACCATTGGCATTTCATCTTAATCTGAATTTGGCTCATAAACTCTACAT	1020
TTTCCAAACTTTAATTCAATTATACCATAGAAAATTAATTTTGGAGTTGGAAATACACATA	1080
AGCATGGGAAAAATATGTATAGTTTTTGGCAATCATGTTTGTGTTATTTAAATATCTATT	1140
TCCTAACCAAGTCATTTCAATAAATTAATCCTTTAGGCATAAAAAAAAAAAAAAAAAAAAA	1196

FIGURE 1 Nucleotide and deduced amino acid sequence of canine  $\beta$ -casein cDNA. The signal peptide is underlined and the stop codon is indicated by asterisks (\*). A consensus polyadenylation recognition signal AATAAA, is located 16 nucleotides upstream from the poly (A) tail. The cbca cDNA was corresponded to nucleotide 79–1144 of canine  $\beta$ -casein cDNA

and 72C for 2min, and followed by 1 cycle of 72C for 5min. The PCR products about 1-kb were purified from agarose gel and inserted into a pT7-blue vector (Novagen). 42 of 50 clones had identical sequences which was named cbca

cDNA. To isolate the 5' ends and 3' ends of canine  $\beta$ -casein message, the 5' and 3' RACE was performed. 5' RACE was performed using 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0 (Gibco BRL) and gene specific

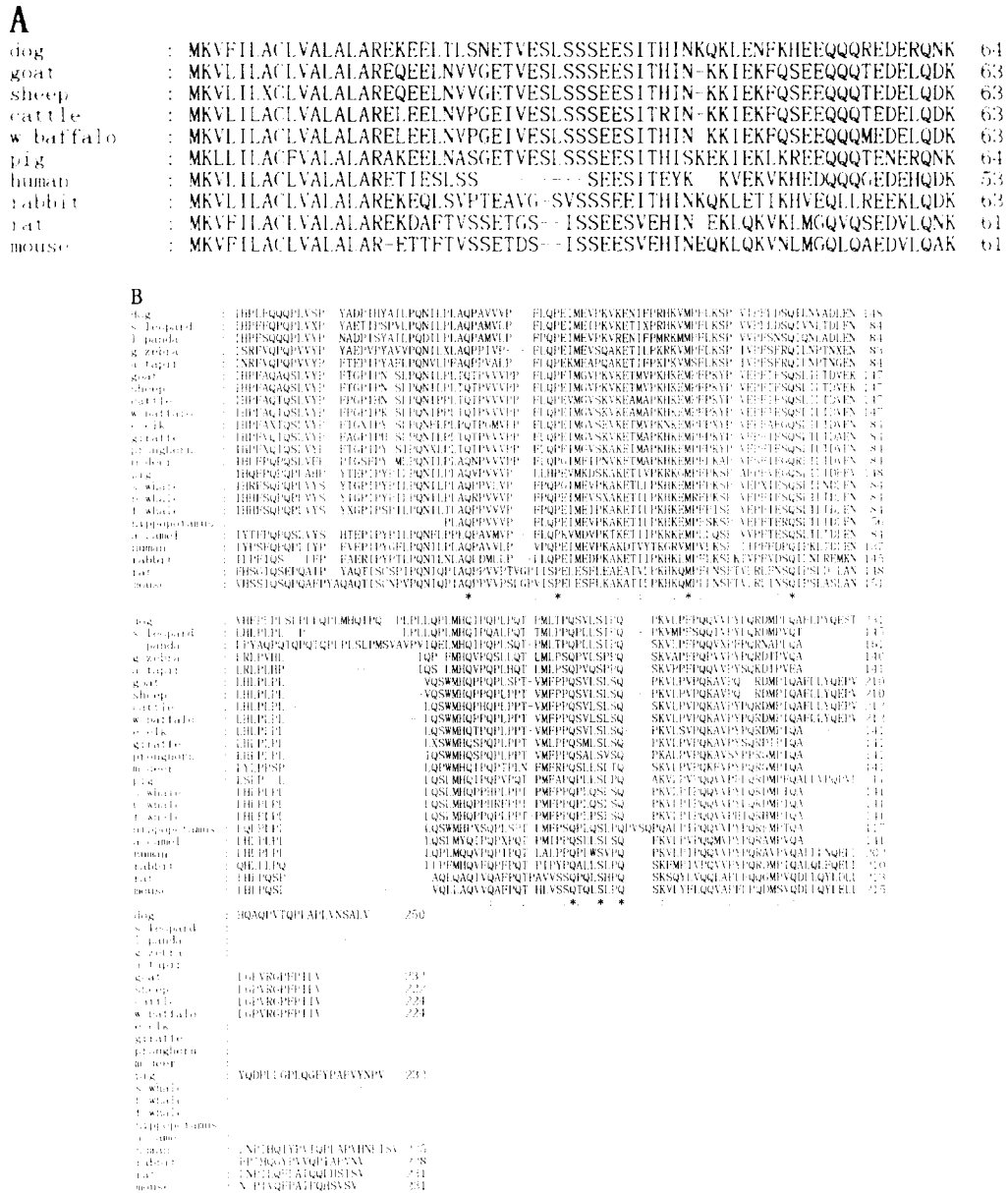


FIGURE 2 Alignment of amino acids sequence of mammalian  $\beta$ -casein genes. The Genbank/EMBL and DDBJ Data Bank accession numbers are: AB035080 (dog; *Canis familiaris*), U53906 (s-leopard; snow-leopard; *Uncia uncia*), U86646 (1-panda; lesser-panda; *Ailurus fulgens*), U53905 (g-zebra; grevyi-zebra; *Equus grevyi*), U53904 (a-tapir; asiatic-tapir; *Tapirus indicus*), AH001195 (goat; *Capra hircus*), X79703 (sheep; *Ovis aries*), M55158 (cattle; *Bos taurus*), AJ005165 (w-buffalo; water-buffalo; *Bubalus arnee bubalis*), U53896 (e-elk; european-elk; *Alces alces*), U53897 (giraffe; *Giraffa camelopardalis*), U53895 (pronghorn; *Antilocapra americana*), U53898 (m-deer; mouse-deer; *Tragulus napu*), X54974 (pig; *Sus scrofa*), U86644 (s-whale; sperm-whale; *Physeter catodon*), U53899 (b-whale; beluga-whale; *Delphinapterus leucas*), U53900 (f-whale; finback-whale; *Balaenoptera physalus*), U53901 (hippopotamus; *Hippopotamus amphibius*), U53902 (a-camel; arabian-camel; *Camelus dromedarius*), X55739 (human; *Homo sapiens*), M10936 (rat; *Rattus norvegicus*), M26940 (mouse; *Mus musculus*). The alignment was made with the CLUSTALW program via the internet (<http://www.ddbj.nig.ac.jp/htmls/E-mail/clustalw-e.html>). . identical residues; : very similar residues; less similar residues

primers, which were based on information from the nucleotide sequence of cbca cDNA (5'-TCTCTTTGCTGTTGCTCCTCGTG-3' and 5'-TGCAAGAGCAAGAGCCACAAGG-3'). We obtained 5' ends of the  $\beta$ -casein consist of 78-bp. To determine 3' ends of canine  $\beta$ -casein message, we made a canine mammary tissue first strand cDNA using dT adapter primer (5'-GCGGCT-GAAGACGGCCTATGT-GGCCITTTTTTTTTTTTTTTT TTT -3'). Using this first strand cDNAs, we obtained 52-bp of 3' ends with the anchor-primer (5'-GCGGCTGAA-GACGGCCTATGT-3') and gene specific primers (5'-AACCTGTGACTCAACCACTTGC-3' and 5'-TTGACTGTGACTGGAAACGTG-3'). The entire canine  $\beta$ -casein cDNA was consist of the 1196 base pair nucleotide sequences. cDNA sequence of  $\beta$ -casein (accession No. AB035080) showed that the start codon is at nucleotide 63-65 and stop codon is at nucleotide 814-816. The cDNA has an open reading frame of 750-bp, and encodes a putative polypeptide of 250 amino acids, including a signal sequence of 15 residues.

A motif search using the MOTIF program (Ogiwara *et al.*, 1996) revealed that the predicted amino acid sequences have casein  $\alpha/\beta$  signature, located at 8 to 15 amino acids. Overall structure of the deduced amino acid sequences exhibited 44% identity and 53% similarity over 225 amino acids to the published human  $\beta$ -casein sequence (Lonnderdal *et al.*, 1990). Thus, we identified the cDNA canine  $\beta$ -casein (Fig. 1).

Comparison of the deduced  $\beta$ -casein amino acid sequence with the database sequence using the BLAST program (Altschul *et al.*, 1990) revealed that it showed the highest homology with snow-leopard (55-62% identity). It also showed similar homology to lesser-panda (54-61% identity). Similarity of predicted canine  $\beta$ -casein polypeptide to another mammalian  $\beta$ -casein amino acid sequence was 33-42% identity with mouse, 29-37% identity with rat, 43-53% identity with rabbit, 41-48% identity with pig, 44-51% identity with cattle and 44-50% identity with sheep.

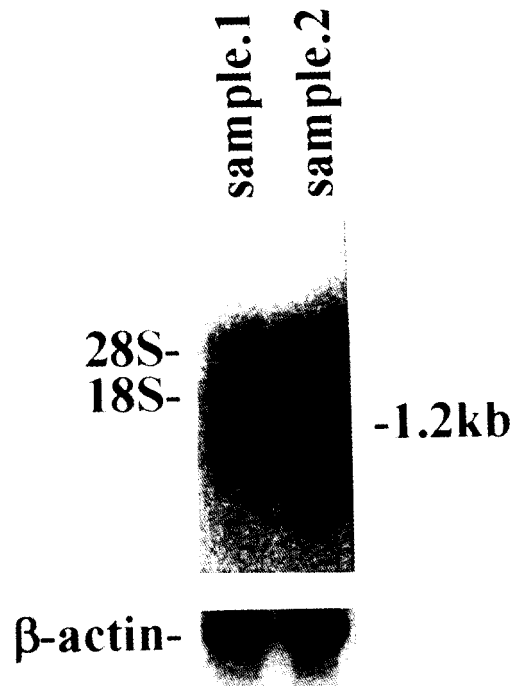


FIGURE 3 Northern blot analysis of canine  $\beta$ -casein mRNA in mammary tissue. Total RNA (10 $\mu$ g) were separated by 1% agarose MOPS-formaldehyde gel electrophoresis and transferred to nylon membrane, and probed with [ $\alpha$ -32] dCTP-labeled cbca cDNA. The RNA was extracted from two mammary tissue samples (Lane 1,2). The positions of ribosomal RNAs (28s and 18s) are shown on the left and size of canine  $\beta$ -casein mRNA (1.2-Kb) is on the right. The same blot was also hybridized with canine  $\beta$ -actin probe (bottom panel)

To compare further the amino acid sequences among mammals, the protein multiple alignments was made using the CLUSTALW program (Tompson *et al.*, 1994). As shown in Fig. 2, alignment of 23 species of  $\beta$ -casein amino acid sequences revealed that overall structure of amino acids sequences were similarly conserved. The 15 residue signal peptide of canine  $\beta$ -casein was identical to those of rat and mouse, and highly similar to those of human, rabbit, goat, sheep, cattle, and water-buffalo (Fig. 2A). In the N-terminal half of the  $\beta$ -casein sequences, glutamine, located at 96 and 138, proline, located

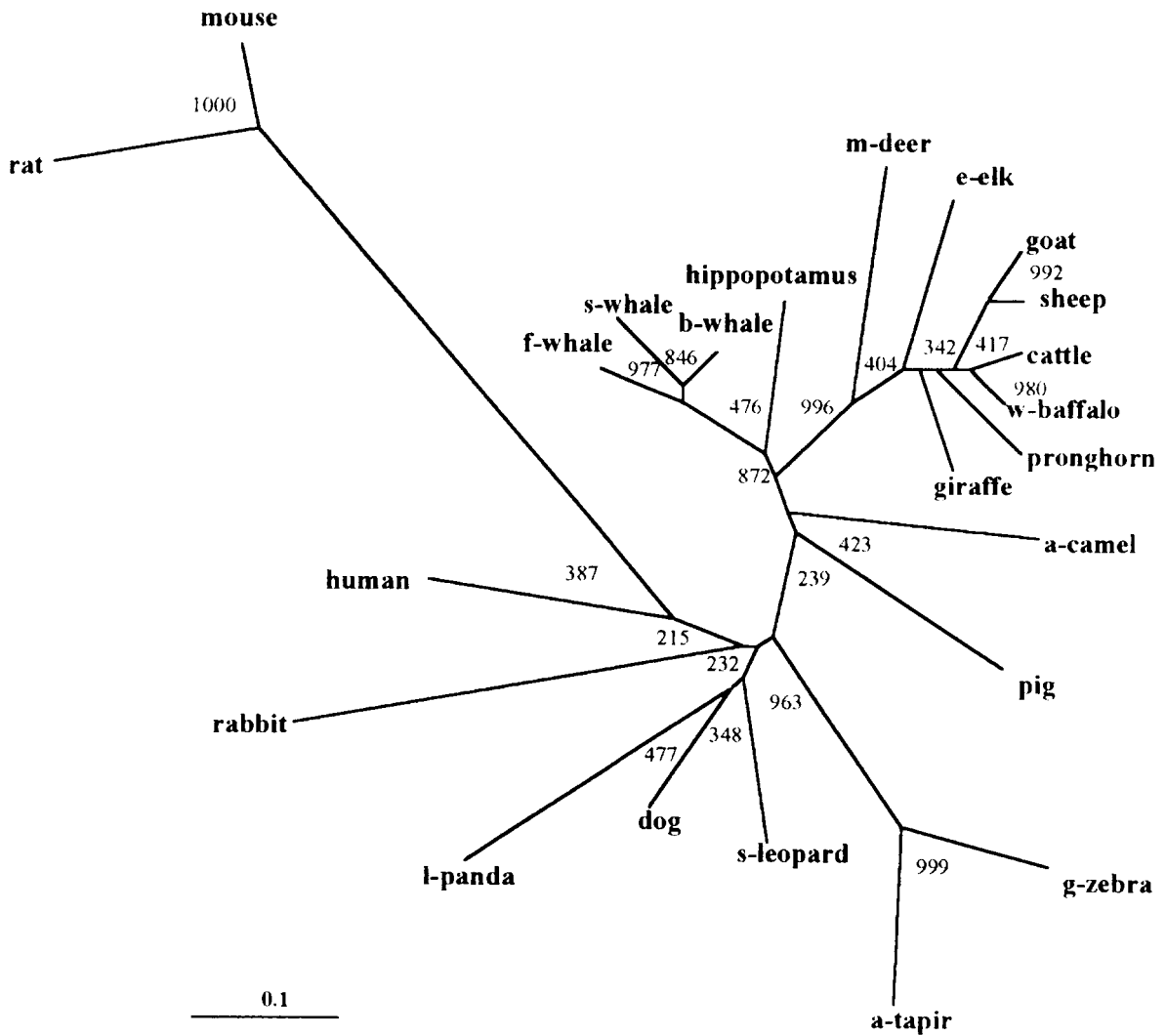


FIGURE 4 Unrooted phylogenetic tree of the amino acid sequences of  $\beta$ -casein among 23 species. The accession numbers of  $\beta$ -casein were described in the figure legend of Figure 2. The amino acids sequences of  $\beta$ -casein were aligned using CLUSTALW and phylogenetic tree was constructed by neighbor-joining method (Saitou *et al*, 1987). The reliability of the tree was measured by bootstrap analysis with 1000 replicates. The length of scale bar is equivalent to 0.1 substitution of amino acid per site

at 106, and methionine, located at 125 were completely conserved. In the C-terminal half of these proteins, there are 8 glutamines between 180 to 222. The spacing between two glutamines is 6 or 7 amino acids except one at 183. Of these 7 glutamine which are regularly spaced, 6 glutamines are conserved for almost all the species. At position 188 which is just next to poor-conserved glutamine, there is a threonine

residue which is conserved for all species, except for mouse-deer.

Canine  $\beta$ -casein has an amino-acids insertion at 156 to 173. An insertion was also found at similar position in lesser-panda and snow-leopard, suggesting that the unique structure may be characteristic of the carnivore order (Fig. 2B). From the gemone structure of  $\beta$ -casein of human (Hansson *et al*, 1994), cattle (Bonsing *et al*, 1998),

goat (Roberts *et al.*, 1992), and mouse (Yoshimura *et al.*, 1990), the insertion may be occurred within one exon in the canine genome. To determine this possibility, we performed PCR using canine genomic DNA as template. A fragment with the predicted size was found using specific primers (5'-CAAATCCTGAATGTGGCTGATC-3' and 5'-TGGTTGAGTCACAGGTTGGGC-3'). Nucleotide sequence of the product was identical to that of canine  $\beta$ -casein cDNA (data not shown). These results suggested that the insertion had occurred in one exon of the canine genome.

To examine canine  $\beta$ -casein mRNA size, Northern blot analysis was performed using canine mammary tissues. As shown in Fig. 3, canine  $\beta$ -casein was expressed in mammary tissue as a major transcripts of 1.2-kb. The size of mRNA was similar to the length of canine  $\beta$ -casein cDNA (1196-bp).

We performed a phylogenetic analysis using amino acids sequences of  $\beta$ -casein. The tree was constructed with aligned sequence by the neighbor-joining methods (Saitou *et al.*, 1987). According to the phylogenetic tree, canine  $\beta$ -casein appeared to form a branch with that of lesser-panda and snow-leopard. This result was consistent with our observation that they have a characteristic amino acid insertion (Fig. 2), and with traditional classification that they are grouped into carnivore order.

In this phylogenetic tree, rodents and perissodactyls formed their own branch. Hippopotamus and cetaceans formed a distinct branch, and close relationship of artiodactyla and cetacea was observed. These results were consistent with the previous report (Gatesy *et al.*, 1996).

Our phylogenetic tree also showed that human was more closely related to rodents than to rabbit, and ruminants shaped a branch among artiodactyla/cetacea cluster. These results were different from Gatesy *et al.* (1996). Two possibilities may be showed as the cause of these discrepancies. First, we constructed the tree using the protein sequences, whereas the tree of the previous study used the regions of exon 7. Second,

while 23 species of  $\beta$ -casein were used in this report, 20 species were used in previous one. Further analysis of  $\beta$ -casein from additional mammals or another milk proteins will contribute to helpful information for mammalian phylogenetics.

### Acknowledgements

We wish to thank Dr. A. Rungsipipat for supplying canine mammary tissue samples. This work was partly supported by Grant-in-aid for Scientific Research from The Ministry of Education, Science, Sports and Culture of Japan.

### References

- Altschul S. F., Gish W., Miller W., Myers E. W., and Lipman D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.
- Bonsing J., Ring J.M., Stewart A.F., Mackinlay A. G (1998). Complete nucleotide sequence of the bovine beta-casein gene. *Aust J Biol Sci* 41, 527-537.
- Gatesy J., Hayashi C., Cronin M.A., Arcander P. (1996). Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. *Mol Biol Evol* 13, 954-963.
- Hansson L., Edlund A., Johansson T., Hernell O., Stromqvist M., Lindquist S., Lonnerdal B., Bergstrom S. (1994). Structure of the human beta-casein encoding gene. *Gene* 139, 193-199.
- Lonnerdal B. (1985). Biochemistry and physiological function of human milk proteins. *Am J Clin Nutr* 42, 1299-1317.
- Lonnerdal B., Bergstrom S., Andersson Y., Hjalmarsson K., Sundqvist A.K., Hernell O. (1990). Cloning and sequencing of a cDNA encoding human milk beta-casein. *FEBS Lett* 269, 153-156.
- Mercier J., Vilotte J., and Provot C. (1990). Structure and function of milk protein genes. In Geldman H. and Ellendorf F. (ed). *Genome analysis in domestic animals*. (Weinheim: New York), pp. 233-258.
- Ogiwara A., Uchiyama I., Takagi T., Kanehisa M. (1996). Construction and analysis of a profile library characterizing groups of structurally known proteins. *Protein Sci* 5, 1991-1999.
- Thompson J. D., Higgins D. G., Gilson T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-4680.
- Roberts B., DiTullio P., Vitale J., Hehir K., Gordon K. (1992). Cloning of the goat beta-casein-encoding gene and expression in transgenic mice. *Gene* 121, 255-262.
- Saitou N., and Nei M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol Biol Evol* 4, 406-425.
- Yoshimura M., Oka T. (1990). Transfection of beta-casein chimeric gene and hormonal induction of its expression in primary murine mammary epithelial cells. *Proc Natl Acad Sci U S A* 87, 3670-3674.