Development and Characterization of CATS Markers for Genetic Linkage Mapping in the House Musk Shrew, Suncus murinus

Samuel ADJEI1), Akira SATO1), Shigekazu TANAKA1), Eiji KOBAYASHI2), Kazuaki TANAKA3), Takao NAMIKAWA1), and Akira ISHIKAWA 1)

1)Laboratory of Animal Genetics, Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Science, Nagoya University, Chikusa-ku, Nagoya 464-8601, 2)Department of Technology, National Livestock Breeding Centre, Nishigo, Fukushima 961-8511, and 3)Department of Animal Science and Biotechnology, School of Veterinary Medicine, Azabu University, Fuchinobe Sagamihara, Kanagawa 229-8501, Japan

Abstract: To serve as an initial step in developing an ideal genetic marker map for the house musk shrew, Suncus murinus, 318 comparative anchor tagged sequence (CATS) primer pairs were assessed for polymorphism ascertainment and linkage mapping. Of the 112 (35.2%) CATS primer pairs that were successfully amplified by PCR in the shrew, 18 (16.1%) showed polymorphism between two mutant strains, BAN-kc, oeb and WZ. Linkage analysis of the polymorphic CATS markers and three visible mutant genes, kc, oeb and wz, genotyped in a 77 F2 mapping panel from a cross of the two mutant strains, assigned wz and five CATS markers into three linkage groups. Sequence analysis revealed that two (ADA and TXN) out of nine CATS amplified sequences had a total of six deletions of varying sizes and 17 single nucleotide polymorphisms (SNPs). BLAST search identified three CATS (ADA, CYP1A2, and TXN) products matching the genes from which they were originally designed, while the remaining six markers could not be identified. Together with the use of the detected SNPs as genetic markers, the five CATS markers linkage mapped in this species will serve as anchors in establishing the first framework map for locating loci affecting all heritable qualitative and quantitative traits in the musk shrew.

Key words: comparative anchor tagged sequence, deletion, linkage mapping, SNP, Suncus murinus

Introduction

Gene mapping in mammalian species has steadily progressed to the stage where precise comparisons can be made using type I (coding) loci, as their DNA sequences can be used to establish homology between divergent species. This is particularly useful for mapping in ‘gene-poor’ species because such loci show high conservation in different species of mammals. One such species deficient in gene maps is the house musk
shrew (*Suncus murinus*, Insectivore). The house musk shrew, widely distributed in Asia, Southwest Pacific islands and East Africa [13], has been domesticated since 1973 for the purpose of providing diversity in animals used in biomedical research [20]. Several spontaneous mutations in coat hair, behavior, morphology, diabetes, etc. have been reported [8, 9, 21, 22] in established laboratory strains. Effects of various emetic and anti-emetic drugs have been extensively studied in the house musk shrew which has shown a potential as a model for emetic research [26], because it is one of the few mammalian species that vomits in response to emetic drugs. It has also been suggested that insectivores, because they lie more directly in the ancestral lineage of mammals than do rodents or lagomorphs, may provide more appropriate models of physiological or pharmacological activity than do the more commonly used laboratory animals such as mice and rats [5, 13]. Recently, the musk shrew has been identified as a model species for studies of nutritional regulation of reproduction [25]. In addition, the peculiar chromosomal characteristics of the shrews [24], said to be one of the groups of species that hold untapped reservoirs of comparative genomes [19], and the absence of a genetic marker map in this species makes the construction of a genetic marker map an invaluable task. Given its ubiquity, its abundance, its early description by Limneaus and its association with human, the house musk shrew is one important insectivore which requires the construction of a gene map.

Lyons and colleagues [15], utilizing the rapidly growing gene mapping databases and the remarkable increase in DNA sequences, designed conserved gene-specific PCR primers based on aligned homologous exon sequences from species of two or more vertebrates. These gene specific PCR primers, named Comparative Anchor Tagged Sequences (CATS), were designed to span introns for polymorphism ascertainment, and to include sufficient exonic sequences (25–400 bp) for gene identification. Markers developed in this manner have been used to generate comparative genome maps in the human, cat, dog, horse and other domesticated species [2, 3, 14, 15]. Expanding gene mapping studies to other eutherian species will add enormously to the knowledge of eutherian genome evolution, especially if the same set of comparative anchor loci are examined, because the relationships of many of the orders of eutherian mammals are difficult to deduce [18].

Taking advantage of the 318 CATS primer pairs [15] offered for integrative mapping of mammalian genomes, we report the development and characterization of CATS markers in the house musk shrew, *Suncus murinus*, and their utility in the construction of a linkage map for identification of mutant genes discovered in the shrew. This will serve as an initial step in elucidating an ideal marker map for this important species.

### Materials and Methods

**Animals:** Animals used included four male parents from the BAN-kc, oeb strain, fixed for two mutations of kinky-coat (*kc*) [8] and open-eyelids at birth (*oeb*) [9], ten female parents from the WZ strain, fixed for the waltzing (*wz*) mutation [21], and 31 F1 and 77 F2 animals of their progeny. The husbandry of all the shrews has been described previously [10].

**PCR optimization:** Genomic DNA samples from the animals described above were obtained by standard phenol-chloroform extraction protocol [23]. The isolated DNA samples from parental BAN-kc, oeb, WZ and their F1 offspring were used in a PCR experiment for optimization and polymorphism ascertainment using the 318 CATS primer pairs synthesized according to the primer sequence described by Lyons *et al.* [15]. Approximately 100 ng of genomic DNA was amplified in a final volume of 14 µl containing 5 pmol of each forward and reverse primer, 1.4 µl 10× Takara PCR buffer (Mg²⁺ free), and 0.07 µl Taq polymerase. To obtain an optimum condition, the basic PCR test amplification conditions were done under different concentrations of MgCl₂ with a cycle of 94°C for 5 min, followed by 30 cycles of 94°C for 1 min, different annealing temperatures for 1 min, and 72°C for 1 min with 10 min as final extension time. This PCR amplification was performed in a Takara thermal cycler MP (Takara Biomedicals, Tokyo). PCR products were separated in 1.5% agarose gel (Seakam GTG, Rockland, ME) for 30 min at 100 V in 1× TBE buffer using a Mupid electrophoresis unit (Cosmo Bio. Co. Ltd., Tokyo), and visualized with ethidium bromide staining. Primers yielding a strong single product were considered as successful.
**Mapping:** Prior to linkage analysis, the segregation ratios of the CATS markers were evaluated for deviations from Mendelian expectations by chi-square test. Linkage mapping was carried out with Map Manager QTxb19 [16] based on the segregation data obtained for the three mutant genes (\(kc, oeb\), and \(wz\)) and the polymorphic CATS markers after genotyping of 77 F\(_2\) individuals. Recombination frequencies (%) were converted to map distances in cM by the Kosambi map function. Linkages with LOD scores of 3.0 or greater were considered significant.

**Cloning and sequencing of CATS amplicons:** DNA fragments produced with CATS primer pairs in a 50 \(\mu\)l PCR containing approximately 350 ng of genomic DNA were cloned into pGEM-T Easy vector (Promega Corporation, Madison, WI) and direct sequences were generated from both strands with BigDye Terminator Cycle Sequencing Ready Reaction Kits version 1.1 (Applied Biosystems, Forster City, CA). The extension reaction in a 20 \(\mu\)l volume was performed and extension products were electrophoresed on ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Forster City, CA). Sequence information was obtained from at least three different cloned PCR products for each DNA sample used. Using the DNASIS software version 3.2 (Hitachi Software Engineering Company Limited, Tokyo) sequences were edited and connected to obtain the fully amplified sequences of both strands.

**Gene identification:** Coding regions of the CATS amplified sequences were predicted with the NCBI Open Reading Frame Finder (ORF Finder), available at http://www.ncbi.nlm.nih.gov/gorf/orf/ orfig.cgi, using standard genetic codes, after which selected regions were manually checked to determine the actual intron region. Sequence homology searches in the non-redundant databases were carried out with the BLAST program [1] available at http://www.ncbi.nlm.nih.gov/BLAST/Blatt.cgi, using the selected coding sequences. Expectation scores < 0.01 were chosen to indicate significant homology and for high stringency because of the shortness of the sequences.

**Results**

**PCR amplification success, polymorphism and inheritance**

Our focus on identifying CATS based on shrew-specific conditions using a final PCR reaction volume of 14 \(\mu\)l resulted in 112 of the 318 (35.2%) CATS primer pairs producing a single PCR product. We investigated the genetic polymorphism of the successful CATS markers in a pedigree comprising four male BAN-\(kc, oeb\) parents, ten female WZ parents and 31 F\(_1\) animals of their progeny using PCR amplification and agarose gel electrophoresis. Of the 33 that were initially observed to be polymorphic between BAN-\(kc, oeb\) and WZ strains, 18 CATS were clearly shown to be polymorphic (Table 1). The remaining 15 CATS were considered ambiguous because of the inability to genotype them in the F\(_2\) animals.

Prior to linkage analysis, the segregation pattern at each of the 18 polymorphic CATS marker loci was checked in the 77 F\(_2\) mapping panel for significant deviations from Mendelian expected ratios. Except for CYP1A2 which deviated significantly from the expected 3:1 ratio at \(P<0.01\) (chi square value of 9.56) for unknown reasons, all the remaining 17 markers were not different at \(P>0.05\): X-linked dominant markers CHGA, FGFR4 and GHR (chi-square values of 0–3.3); autosomal co-dominant markers ADA, GUSB, LMC1, PKM2 and TXN (chi-square values of 0.25–0.66); and autosomal dominant markers ADRBK2, BCL2, CREM, CYP1A2, CYP21, GJB2, MAOA, MET, MYOD1 and NRAS (chi-square values of 0.04–0.97). In addition, segregation patterns of three mutant genes, \(kc, oeb\) and \(wz\), were not significantly different from the expected 3:1 ratios based on autosomal recessive inheritance as previously reported [8, 9, 21]. Genotyping data for this study will be made available upon request.

**Linkage analysis**

Linkage analysis of the three mutant genes (\(kc, oeb\), and \(wz\)) and the 18 polymorphic CATS markers resulted in three linkage groups containing five CATS (ADA, ADRBK2, CREM, MET and TXN) and one mutant gene (\(wz\)), covering a total distance of 42.9 cM with the average distance between loci being 14.3 ± 4.7 cM (Fig. 1). LOD scores ranged from 6.4 to 10.1.
Characterization of CATS and comparative sequence analysis

The five CATS markers that were linkage mapped as described above were analyzed for sequence variants. In addition, one CATS marker, CYP1A2, having shown a possible linkage with the kc mutation (LOD score of 2.6), and three others, CHGA, FGFR4 and GHR, having shown LOD scores of 2.0 and 1.5 respectively for FGFR4-CHGA and CHGA-GHR and a segregation pattern indicating linkage to the shrew chromosome X, were included in the sequence analysis. Six deletions of varying sizes and 17 SNPs (single nucleotide polymorphisms) were detected between the two shrew strains in two co-dominant markers, ADA and TXN (Fig. 2). All SNPs detected were located in the non-coding regions. In contrast, no deletions or SNPs were detected within and/or between strains in sequences obtained for the seven dominant markers, ADRBK2, CHGA, CYP1A2, CREM, FGFR4, GHR and MET.

When compared with the GenBank database by BLAST search, the PCR products of three CATS (ADA, CYP1A2 and TXN) matched the genes from which the CATS primers were designed, with the expectation for best match values (e-values) of $1.0 \times 10^{-4}$, $4.0 \times 10^{-22}$ and $2.0 \times 10^{-11}$ for the human, respectively. For the remaining six CATS markers, even though the sequences matched for all shrew samples used, no proper
match was found in the database. Assessment of the available database sequences of the nine genes, from which the CATS primers were originally designed, showed that primer design errors might have possibly contributed to the inability to identify these markers properly. The sequences for both primers of ADRBK2 and the reverse primer of CHGA could not be located. The 21 bp sequence of the reverse primer of MET was from the last ten nucleotides of exon 19 and the first eleven nucleotides of exon 20, suggesting that it was probably designed from the cDNA sequence of MET. FGFR4 also had four nucleotides at the 3’ end of the reverse primer from an intron region. Except for ADA, CYP1A2 and TXN that matched their original genes, primer pairs for CHGA, CREM, FGFR4 and MET, were not designed from consecutive exons. GHR presented a peculiar case with a very low homology to the human (e-value of 6.9). The best matches were the clones Homo sapiens, X BAC RP11-6471I17 (e-value of 6.0 × 10⁻²²) and Mus musculus, RP23-174C4 (e-value of 2.0 × 10⁻¹⁵) on X chromosome, implying the amplification of a gene located on the shrew X chromosome. This was in agreement with the X-linked segregation pattern obtained for CHGA, FGFR4 and GHR in the F₂ intercross, as described above.

The nucleotide sequence data reported in this paper have been submitted to DDBJ and have been assigned the accession numbers AB167746 to AB167758.

**Genetic relationship**

In analyzing the genetic relationship, exon sequences obtained for ADA, CYP1A2 and TXN in the shrew were aligned with coding sequences for the human, mouse and rat (Fig. 3). For ADA, the shrew sequence of 67 bp showed 82.1%, 73.1% and 77.6% similarity, respectively, for the human, mouse and rat, and the 149 bp CYP1A2 fragment displayed 83.9%, 79.9% and 82.6%. Furthermore, the identity of the 57 bp TXN fragment and the human sequence for TXN was 94.7%, while for both the rat and mouse, it was 79.0%. Taken together, the generated 273 bp sequences displayed a total DNA sequence similarity of 85.7%, 75.5%, 76.9%, 78.0% and 80.6% corresponding to shrew/human, shrew/mouse, shrew/rat, human/mouse and human/rat pairwise identities, respectively. Thus, the shrew consistently showed a closer genetic relationship with the human than with the mouse or rat.
As an initial step to identify genetic markers for the house musk shrew, *Suncus murinus*, we evaluated CATS primer pairs and assessed their usefulness in constructing the first low-resolution genetic map for this species. It is interesting to note that the 35.2% of CATS primer pairs producing a single PCR product in the musk shrew correlates well with the 38.5% obtained for the short-tailed shrew, *Blarina brevicauda* [15]. This figure also compares favorably with that of other eutherian species, with the most efficient being the deer (*Equus davidian*) with a success rate of 52% [15]. It is also significant in the sense that about 40% of all CATS primer pairs are expected to work in any eutherian mammalian species. Failure of the 206 CATS primers to amplify DNA fragments specific to the musk shrew may have been caused by mismatch of the 3’ region of the primer sequence and the bonding site of the musk shrew genomic DNA. Redesigning these primers to anneal to different regions of the targeted genes in the shrew genome would facilitate their amplification and subsequent mapping.

Most importantly, the polymorphisms observed between the two mutant musk shrew strains, BAN-*kc,oeb* and WZ, for 18 (5.7%) CATS primer pairs comprised 5 co-dominant and 13 dominant markers. This is the first report of the detection of length polymorphism in the 5 co-dominant markers in this species in the elucidation of CATS markers. In other works, CATS polymorphism detection was mainly by restriction fragment length polymorphism (RFLP) and/or by single strand conformation polymorphism (SSCP) analysis [3, 14, 15]. Greater polymorphism discrimination would likely have resulted if RFLP screening or SSCP analysis were applied to the musk shrew [15].

A large number of deletions and SNPs between the two strains for the two co-dominant markers, *ADA* and *TXN*, were detected. These polymorphisms must reflect the remarkable genetic difference between the two strains which have different geographical origins. The BAN strain was originally established from wild shrews captured in Bangladesh [11], whereas the WZ strain has been bred from wild shrews captured in Japan [21]. Extreme differences in adult body size and wild-type coat color have been reported between the two strains [10, 12]. Furthermore, a previous sequence analysis of shrew mitochondrial DNA (mtDNA) revealed that the two strains are genetically differentiated from each other to an extent comparable to a mouse intersubspecific
level [27]. The detected SNPs, which happen to be intronic, can add to the marker map of this species as genetic markers for fine-scale mapping of mutant loci, quantitative trait loci and others of interest. Sequence analysis of the 94 CATS that amplified non-polymorphic or ambiguous polymorphic PCR products between the two strains leaves open the possibility that they will become useful SNP markers because of the above noted extreme genetic difference between the strains. Linkage mapping of the two markers, \( \text{ADA} \) and \( \text{TXN} \), and their identification as the correct CATS primer derivative genes, make them ideal for chromosomal localization in the musk shrew. This will assist the effort to make available comparative mapping data for the musk shrew [17], a task we are currently engaged in.

Comparative analysis of the amplified exon sequences of \( \text{ADA} \), \( \text{CYP1A2} \) and \( \text{TXN} \) for the shrew with coding sequences for the human, mouse and rat may provide a certain insight into their genetic relationships. The percentage sequence similarity was highest between the shrew and human in all the three sequences, further confirmed by the mtDNA analysis showing that the insectivore may have more in common with the human than rodents or lagomorphs [5]. This result also agrees well with two works on pancreatic polypeptides [6] and the immunoglobulin \( \mu \) gene [7] in the house musk shrew. Increasing the data set based on nuclear genes for insectivore species will further clarify the position of the house musk shrew in relation to the other shrew species, and will also complement available data based on mtDNA.

The linkage of \( \text{ADA} \) and \( w_z \) in linkage group I prompted the search for a possible candidate for the waltzing (\( w_z \)) mutation using the comparative map information on \( \text{ADA} \) in the human, mouse and rat. Searching for phenotypic mutants on circling behavior and for functional genes related to hindbrain development in the National Center for Bioinformatics (NCBI) database (http://www.ncbi.nlm.nih.gov/mapview) identified the kreisler (\( kr \)) mutation (the present gene symbol, \( \text{Mafb} \) (v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B avian)) at 91.0 cM on mouse chromosome 2 (human and rat ortholog: \( \text{Mafb} \) on chromosome 2q11.2-q13.1 and \( \text{Mafb} \) on chromosome 3q42, respectively) as a possible candidate for the \( w_z \) mutation in the shrew. \( kr \) is the only behavioral mutant around \( \text{ADA} \) in all the three species considered. As a flanking marker on the opposite side of the \( w_z \) mutation is lacking, the construction of a more informative linkage map around \( w_z \) will be needed to specify a candidate for this mutation.

**Acknowledgments**

The authors are grateful to Professor Yoichiro Matsuda (Hokkaido University) for the discussion. S. Adjei is also grateful to Japan International Cooperation Agency (JICA) for sponsoring his two years graduate study in Japan.

**References**