JARQ 28, 52-58 (1994)

Analysis of Swine Genome Organization: Use of Repetitive Sequences as Markers for Genetic Linkage Analysis

Hiroshi TASUE
Department of Animal Breeding and Genetics, National Institute of Animal Industry (Tsukuba, Ibaraki, 305 Japan)

Abstract
In order to analyze the linkage of phenotypic traits in the field of molecular biology, marker sequences which display polymorphism are a prerequisite. Since it is generally accepted that interspersed repetitive sequences are highly variable, we considered that these sequences would be suitable as marker sequences. Therefore, the interspersed repetitive sequences from swine genomic DNA have been cloned and characterized. The cloned repetitive sequences were found to belong on the PRE-1 family. The homology analysis of those sequences indicated that they varied in the range of 68 to 85%, and that an arginine t-RNA gene was their progenitor. The localization of PRE-1 sequences on chromosomes was investigated by *in situ* hybridization which revealed that the PRE-1 sequences were distributed unevenly all over the chromosomes. These findings indicated that the PRE-1 sequences could be used as suitable linkage markers.

Discipline: Biotechnology / Animal industry
Additional key words: chromosomal localization, genomic DNA.

Introduction

Since genetic information was shown to be imprinted in DNA by Avery et al.[1], it is generally recognized that economic traits of farm animals are also transcribed in DNA as the other traits. However, until the advent (in the 1980s) of techniques for gene manipulation, it had not been possible to investigate the relationship between the economic traits and DNA (or genes on DNA). Thus, for many years, economic traits have been studied to clarify their genetic linkage using statistical methods and to introduce useful traits into farm animal breeds.

In the field of medicine, human genome studies focused on genetic diseases have advanced remarkably with the use of gene manipulation techniques. Recently many genes responsible for genetic diseases such as Duchenne muscular dystrophy[16] have been identified. The human genome studies are being currently promoted in taking gene therapy into account.

When we consider the studies on the genome of farm animals compared with human genome studies, it is obvious that genome studies should lead to the identification of genes responsible for economic traits and enable to introduce these genes into farm animal breeds. The genome studies of animals, however, lag behind

those of human beings. In the studies on animal genome it is essential to isolate marker sequences for genetic linkage analysis, and to assign them on the chromosomes. Such marker sequences are required to reflect sequence polymorphism in the species and/or in the breeds.

To achieve this objective, we cloned at the molecular level swine repetitive sequences for use as genetic markers, and characterized them.

Materials and methods

Preparation of genomic DNA: Swine genomic DNA was extracted from spermatozoa of a Large White boar, according to the method of Fujinaga et al.[5].

Labeling of DNA for probe: DNA was labeled with ^{32}P by nick-translation or the kinase method[9] and used for plaque hybridization, dot-blot hybridization, and Southern blot hybridization. For *in situ* hybridization, DNA was labeled with biotin by the method of Forster et al[4].

Molecular cloning of repetitive sequences: Swine genomic DNA was cleaved with *Eco*RI, and the resulting DNA fragments were molecularly-cloned in Charon 4A (lambda phage vector) according to the method of Enquist and Stenberg[3]. The recombinant phages were then subjected to plaque hybridization by using ^{32}P-labeled swine genomic DNA as a probe to select phages containing repetitive sequences[9]. We assumed that signal intensity was proportional to the concentration of the probe DNA in the hybridization mixture when the amount of sample DNA on the membrane was constant. Since the concentration of individual repetitive sequences in genomic DNA probe is proportional to their frequencies, the plaques of phages containing repetitive sequences are likely to give a stronger signal intensity in hybridization than those without repetitive sequences.

Sequence analysis: The nucleotide sequence of the DNA fragment was analyzed by the dideoxy chain termination method[10] using an automatic sequencer (model 370A, Applied Biosystems Co., USA). The sequence homology analysis was performed by using a GENETYX program (Software Development Co., Japan).

Dot-blot hybridization: Appropriate amounts of genomic DNA and DNA fragments for copy controls were denatured in an alkaline solution (0.4 NaOH, and 0.6 M NaCl), then dot-blotted onto a nylon membrane (Boehringer Mannheim Co., FRG). The membrane was processed as described by the manufacturer. The hybridization conditions were the same as those described previously[13].

Southern blot hybridization: DNA was cleaved by appropriate restriction enzymes under the conditions recommended by the manufacturers. The cleaved products were then subjected to electrophoresis in 0.9% agarose gel. Subsequently, the DNA fragments in the gel were denatured *in situ* by alkali, and transferred to a nitrocellulose membrane (BAS 83; S & S Co. FRG), according to the method described previously[13]. The DNA on the membrane was then hybridized with ^{32}P-labeled probe DNA as in the case of dot-blot hybridization.

***In situ* hybridization:** Peripheral blood cells from a Large White boar were cultured in RPMI1640 containing 10% fetal calf serum and 10 μl/m*l* phytohemagglutinin for 43 hr. Colcemid was added to the culture to a final concentration of 0.04 μg/m*l* 30 min before the end of the culture. Then the cultured cells were fixed for chromosome spreads on glass slides[14]. The chromosome spreads thus obtained were subjected to *in situ* hybridization as described previously[15]. The hybridization mixture contained 2 μg/m*l* probe DNA, 250 μg/m*l* *M. lysodeikticus* DNA as a carrier, 6 × SSC, 10 mM EDTA, 5 × Denhald's reagent and 45% formamide. One hundred microliters of the mixture was put on one glass slide and covered with a siliconized coverslip, followed by

incubation at 39°C for 15 hr. After hybridization, the glass slides were washed with 2 × SSC containing 50% formamide at 30°C and with 2 × SSC, followed by dehydration in ethanol. The probe DNA hybridized to chromosomal DNA was visualized by the alkaline phosphatase-avidin/BCIP/NBT system[15].

Results

1) Molecular cloning of repetitive sequences

From the lambda phage library of swine genomic DNA, the phages that may contain repetitive sequences were selected according to the procedures described above. Two primary phage clones containing 6.2 kilobase pairs (kbp) and 9.5 kbp genomic fragments, respectively were selected and the cleavage map of the fragments was constructed using *Bam*HI, *Bgl*II, *Cla*I, *Eco*RI, *Hin*cII, and *Sma*I (Fig. 1). The map was then examined by Southern blot hybridization using genomic DNA as a probe to locate the repetitive regions on the cloned fragments. It was shown that the three segments shaded in Fig. 1 contained repetitive sequences. The segments were then subcloned to pBluescript (Stratagene Co., USA) to be sequenced. The nucleotide sequence of the three cloned fragments was determined by the method described above, and registered in DDBJ with accession numbers D10045, D10046, and D10047, respectively[12].

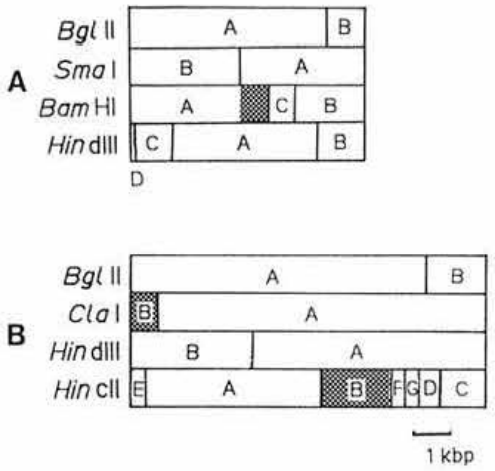

Fig. 1. Restriction endonuclease cleavage maps of (A) *Eco*RI-6.2 kbp and (B) *Eco*RI-9.5 kbp fragments
Segments in each restriction endonuclease cleavage map were designated in alphabetical order in the order of decreasing fragment size. The shaded regions were found to contain sequences with higher repetitive frequencies over the fragments by Southern blot analysis (see text).

2) Homology analysis

The homology analysis between the fragments which was performed using the GENETYX program, revealed that 6 regions in the fragments showed a sequence homology ranging from 68 to 85%. The sequences in the 6 regions were then compared with the sequences registered in GenBank using the program. The comparison showed that either of the 6 sequences showed a homology with the PRE-1 consensus sequence which was reported as a swine SINE (short interspersed repetitive element) by Singer et al.[11]. The extent of the homology was in the range of 70 to 84%[12], indicating that the 6 sequences belonged to the PRE-1 family. We designate our 6 sequences as 'DOS1-6' (domestic swine repeats) hereafter for convenience. When the 6 DOS sequences were aligned, it was revealed that the sequences contained an RNA polymerase III split promoter region and a relatively 'conserved' region (underlined twice in Fig. 2).

To estimate the repetitive frequency in the swine genome, we synthesized the conserved sequence (underlined twice in Fig. 2) and used it as a probe in dot-blot hybridization. The hybridization under the conditions of Tm-20°C revealed that the repetitive frequency was approximately 2×10^6 per diploid genome in each of the five swine breeds (Meishan, Hampshire, Landrace, Large White, and Duroc) (data not

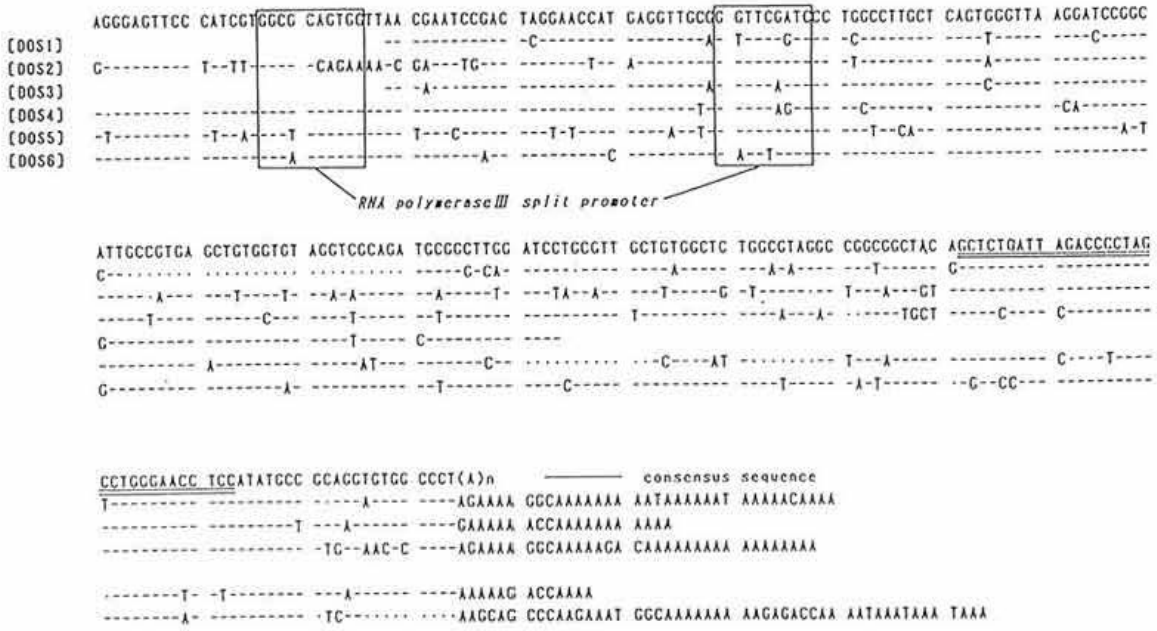

Fig. 2. Comparison of PRE-1 sequences (DOS1 to 6) in *Hin*cII, *Bam*HI/*Sma*I, and *Eco*RI/*Cla*I fragments, and their consensus sequences

The consensus sequence in this figure was constructed based on the 6 PRE-1 sequences (DOS1 to 6) in *Hin*cII, *Bam*HI/*Sma*I, and *Eco*RI/*Cla*I fragments. It was then compared with each of the PRE-1 sequences. Symbols other than A, C, G, and T in the DOS sequences indicate the following:—, the nucleotide is the same as that in the consensus sequence; ●, the nucleotide corresponding to that in the consensus sequence is deleted. The sequence underlined twice in DOS2 was used as a probe DNA in the dot-blot hybridization.

shown). No hybridization signals were detected when bovine, mouse, and human DNAs were dot-blotted[12]. Assuming that the average PRE-1 size is 230bp, the 2×10^6 copies per diploid genome amount to about 7% of the swine genome.

Since mammalian SINEs show a sequence homology with specific genes such as t-RNA genes, it appears that these genes are the progenitors of the SINEs [2,8]. Therefore, the DOS sequences were compared with the sequences registered in GenBank under less stringent conditions than the above using the GENETYX program. The comparison revealed that part of the DOS sequences showed a homology with bovine liver arginine t-RNA gene (72%), *Drosophila melanogaster* arginine t-RNA gene (77%), *Shizosaccharomyces pombe* arginine t-RNA gene (73%)[12]. These findings suggested that the progenitor of the PRE-1 sequences could be an arginine t-RNA gene, indicating for the first time that the progenitor of SINEs is an arginine t-RNA gene. This finding provides an additional evidence for the diversity of progenitor genes of current SINEs in mammalian genomes.

3) *Localization of PRE-1 sequences on chromosomes*

In order to localize the PRE-1 sequences on chromosomes, *in situ* hybridization using biotin-labeled plasmid DNA containing Hinc-fragment (see legend to Fig. 1) as a probe was performed. The Hinc-fragment was shown to

consist of 3 PRE-1 sequences and unique sequences. Almost all the hybridization signals on the chromosomes were therefore expected to be associated with the hybridization of PRE-1 sequences. Plasmid DNA lacking insert DNA was labeled with biotin, and used as a control for the hybridization. The results of *in situ* hybridization are shown in Plate 1A and 1B. While the control probe DNA failed to yield hybridization signals on the chromosomes, as anticipated (Plate 1A), the plasmid DNA containing Hinc-fragment yielded hybridization signals all over the chromosomes (Plate 1B)[15], with different intensities along the chromosomes. These results indicated that the PRE-1 sequences were distributed unevenly all over the chromosomes.

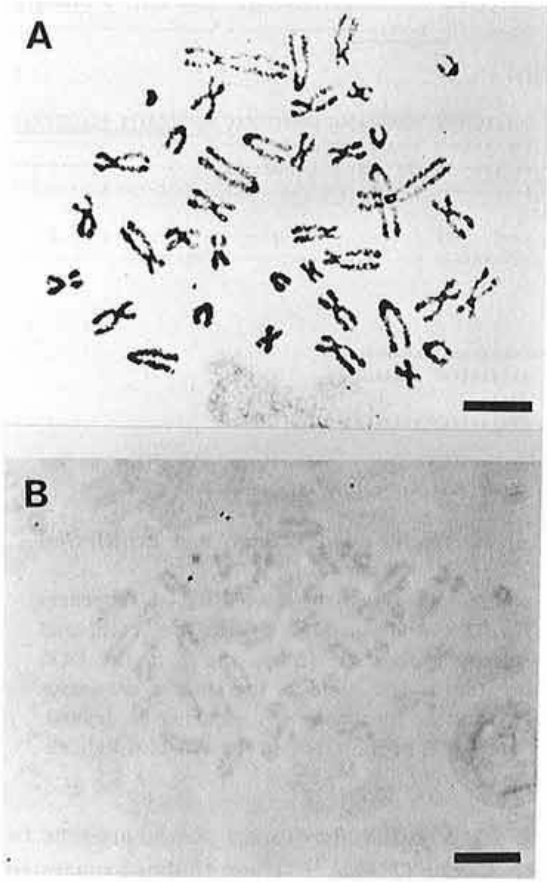

Plate 1. Distribution of "PRE-1" sequences along the chromosomes

The chromosome spreads on glass slides were subjected to *in situ* hybridization by using biotin-labeled pBS-Hinc as described in "materials and methods." As a control, the chromosome spreads were hybridized with biotin-labeled pBS. Panel A shows a representative metaphase cell hybridized with pBS-Hinc; and Panel B, with pBS. For Panel B, the cell was photographed under a longer exposure than that for Panel A, in order to distinguish the chromosomes whithout hybridization signals from the background.
Scale bars: 10 μm.

Discussion

1) Qualification of PRE-1 sequences as genetic linkage markers

Our study on PRE-1 sequences revealed the following aspects: 1) The sequences were present in the swine genome at a frequency of 2×10^6 per genome. 2) The homology of the sequences varied in the range of 68 to 85%. 3) The progenitor of the PRE-1 sequences is considered to be an arginine t-RNA gene. 4) The sequences were distributed unevenly all over the chromosomes.

The prerequisite factors for a genetic linkage marker are as follows: 1) The marker sequences should contain detectable polymorphism among individuals. 2) The marker sequences should be specific to a single locus. As described above, the PRE-1 sequences are highly variable. In addition, our recent experiments revealed that the PRE-1 sequence in one locus was different between two individuals (unpublished data). Therefore, it is possible that the PRE-1 sequence at a given locus varies among individuals. Then, one may raise the question of whether a PRE-1 sequence can be easily assigned to a single locus on the genome. When the flanking sequences of the PRE-1 sequences were compared with each other, no homology was observed between the sequences,

in contrast to the homology between the PRE-1 sequences. We therefore attempted to amplify the PRE-1 sequence of one locus by PCR using the flanking sequences as a set primer. The amplified DNA fragment was examined by agarose gel electrophoresis, then by the method enabling to detect single strand conformation polymorphism (SSCP)[7]. These examinations revealed that the amplified fragment containing a PRE-1 sequence was derived from one locus of the genome (unpublished data).

These findings strongly indicate that PRE-1 sequences could be used as effective genetic linkage markers.

2) Detection of sequence polymorphism of amplified PRE-1 sequences

The restriction fragment length polymorphism (RFLP) has been used as genetic linkage marker. However, as the detection rate of RFLP is not very high in the segments of a given locus, a large number of tests must be performed to detect RFLP at the locus. Recently, an efficient method for the detection of SSCP has been developed by Orita et al.[6]. Thus, SSCP is now considered to be more useful as a genetic marker than RFLP. Therefore, we attempted to detect the SSCP of PCR-amplified PRE-1 fragments. The experiment, though preliminary, revealed the heterozygosity of the locus containing the PRE-1 sequence.

3) Effective strategy for genetic linkage analysis using PRE-1 sequences

As described in the preceding sections, the PRE-1 sequences are considered to be one of the most suitable candidates for genetic linkage markers. However in order to use the genetic linkage marker, we first have to assign marker sequences on chromosomes. To achieve this objective, *in situ* hybridization should be performed on the chromosomes using unique sequences flanking PRE-1 as a probe. The *in situ* hybridization at present empirically requires unique sequences longer than 10 kbp in total to obtain reproducible and ample hybridization signals on the chromosomes. Therefore, we consider that the following procedure is suitable for the PRE-1-based genetic linkage analysis: 1) Genomic segments ranging between 35 kbp and 45 kbp should be molecularly cloned in cosmid. 2) Genomic DNA cosmid clone containing PRE-1 sequences should be selected by colony hybridization. Then the clones thus obtained should be further subjected to selection for PRE-1 sequences showing SSCP in genome. 3) The unique sequence of the cosmid clone should be used for *in situ* hybridization as a probe, in order to assign the sequence to the chromosomes. 4) Based on the assignment, precise linkage between PRE-1 marker sequences should be analyzed by family studies using SSCP, along with studies on the linkage between PRE-1 markers and phenotypic traits.

References

1) Avery, O. T., MacLeod, C. M. & McCarty, M. (1944): Studies on the chemical nature of the substance inducing transformation of *Pneumonococcus* types. *J. Exp. Med.*, **79**, 137-158.
2) Daniels, G. R. & Deininger, P. L. (1985): Repeat sequence families derived from mammalian tRNA genes. *Nature*, **317**, 819-822.
3) Enquist, L. & Stenberg, N. (1979): *In vitro* packaging of λ *Dam* vectors and their use in cloning DNA fragments. *Methods in Enzymology*, **68**, 281-298.
4) Forster, A. C. et al. (1985): Non-radioactive hybridization probes prepared by the chemical labeling of DNA and RNA with a novel reagent, photobiotin. *Nucleic Acids Res.*, **13**, 745-761.
5) Fujinaga, K. et al. (1973): Analysis of multiple viral genome fragments in adenovirus 7-transformed hamster cells. *Cold Spring Harbor Symp.*, **39**, 633-636.
6) Orita, M. et al. (1989): Detection of polymorphism of human DNA by gel electrophoresis as single-strand conformation polymorphism. *Proc. Natl. Acad. Sci. USA*, **86**, 2766-2770.
7) Orita, M. et al. (1989): Rapid and sensitive

detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics*, **5**, 874–879.

8) Sakamoto, K. & Okada, N. (1985): Rodent type 2 *Alu* family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evoluted from tRNA genes. *J. Molec. Evol.*, **22**, 134–140.
9) Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989): Molecular cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
10) Sanger, F., Nicklen, S. & Coulson, A. R. (1977): DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
11) Singer, D. S., Parent, L. J. & Ehrlich, R. (1987): Identification and DNA sequence of an interspersed repetitive DNA element in the genome of the miniature swine. *Nucl. Acids Res.*, **15**, 2780.
12) Takahashi, H., Awata, T. & Yasue, H. (1992): Characterization of swine short interspersed repetitive sequences. *Anim. Genet.*, **23**, 443–448.
13) Yasue, H., Ishibashi, M. & Fujinaga, K. (1980): The oncogenicity of avian adenoviruses. II. The arrangement of the viral DNA in tumors and transformed cells. *Virology*, **106**, 361–373.
14) Yasue, H. & Ishibashi, M. (1982): The oncogenicity of avian adenoviruses. III. *In situ* hybridization of tumor line cells localized a large number of a virocellular sequences in few chromosomes. *Virology*, **116**, 99–115.
15) Yasue, H. et al. (1991): Uneven-distribution of short interspersed repetitive sequence, PRE–1, on swine chromosomes. *Cell Struct. Funct.*, **16**, 475–479.
16) Wapenaar, M. C. et al. (1988): A deletion hot spot in the Duchenne muscular dystrophy gene. *Genomics*, **2**, 101–108.

(Received for publication, Feb. 17, 1993)