The Seoul Journal of Medicine
Vol. 36, No. 2 : 61-68, June 1995

Allele Frequency and Genotype Distribution of STR SE33 locus in Koreans

Jung Bin Lee, and Chan Choi[1]

Department of Forensic Medicine, Seoul National University College of Medicine, and Department of Pathology, Wonkwang University College of Medicine[1]

= Abstract = **DNA was obtained from placental tissue and peripheral blood of 268 unrelated Koreans in order to analyze the applicability of PCR amplification of human beta-actin related pseudogene H-beta-Ac-psi-2(ACTBP-2, SE33) STR locus for forensic medicine in the field of individual identification and paternity tests. After amplification using two primers specific to the SE33 locus, the products were analyzed by polyacrylmide gel electrophoresis, so called Amp-FLP(Amplification Fragment Length Polymorphism) procedure followed by silver staining. To improve the differentiation of the allele an automatic sequencer, Licor 4000-L was used and the result was compared with that of routine acrylamide gel electrophoresis. 36 alleles were deteced, and the distribution was somewhat homogenous, and none of the alleles were above 0.1 in frequency. Allele 12 and allele 14 were most frequent with frequencies of 0.09701 and 0.08022, and alleles of two base repeat such as allele 17, allele 19 were also encountered. 144 kinds of genotypes were found, among which 22 were homozygotes, and the heterozygosity was 91.85%. In 164 gametes of 64 families whose parent-child relationship were confirmed through other studies and information, Mendelian inheritance was well presented although two mutations were observed.**

Key words: *Human beta-actin related, SE33, STR, Amp-FLP, automatic sequencer, Mendelian Inheritance*

INTRODUCTION

With the advent of advanced molecular biological technique, individual identification has achieved a great advance. The use of DNA polymorphism has several advantages over routine blood typing, enzyme typing, serum protein typing or HLA typing(Newall 1981). The first is the quantity of sample needed for the test. The classical methods using antigen-antibody reaction or mobility difference of the protein in electrophoresis need considerable amounts of sample. But in the case of DNA typing, only a small amount, even a tiny blood stain or a hair shaft can afford good results. Second is the quality of the sample. The above methods can usually give reliable results only with fresh samples,

This work was supported by grants partly from the Seoul National University Hospital Research Fund (1989-190) and partly from the JINRO Foundation
서울대학교 의과대학 법의학교실 : 이정빈
원광대학교 의과대학 병리학교실 : 최 찬

after less than 6 months has elapsed after leaving the human body. DNA is more stable than protein, so the DNA test is more useful than protein work in the case of old and/or contaminated samples. Recently several reports about DNA work on an old Egyptian mummy have been published (Lawlor *et al.*, 1991). Furthermore, the DNA work can provide a more powerful discriminating capacity. With the multilocus probe provided by Jeffreys, we can tell one from the other as we can with fingerprints, so the word 'DNA fingerprint' was introduced. The method applied at first was the analysis of restriction fragment length polymorphism (RFLP) via Southern blotting (Guisti *et al.*, 1985, Kanter *et al.*, 1985). Although it was effective, some difficulties appeared in technical aspects. It required a microgram amount of relatively undegraded DNA (Jeffreys *et al.*, 1985) for multilocus analysis, and hundreds of nanogram for single locus analysis (Wong *et al.*, 1987). But usual biological samples cannnot satisfy this. Polymerase chain reaction (PCR) has resolved this problem (Saiki *et al.*, 1985). Only a limited qunatity of DNA is required for the amplification. Furthermore, even with somewhat degraded samples, if the target sequences are intact, it can produce reproducible results. Elimination of the need for isotopes and rapidity are its additional advantages.

PCR can be linked to several methodologies. After amplification, restriction enzyme can be treated and the fragment pattern can be resolved (Saiki *et al.*, 1985) or SSCP (single strand conformation polymorphism) which detects point mutation can be revealed. Direct hybridization procedure can be done to the amplification product as in dot blotting or slot blotting, which is widely applied for HLA typing. But separation of the product according to its size, so called amplification fragment length polymorphism (Amp-FLP) is a simple and rapid method especially in VNTR (variable number of tandem repeats) typing. The alleles are separated by sizes which differ according to the repetition number of core sequences.

The mechanisms producing genetic polymorphism are largely divided into two; sequence polymorphism and length polymorphism. Sequence polymorphism usually derives from point mutation. So the degree of polymorphism is proportional to the frequency and loci of mutation. VNTR loci of DNA are representative of length polymorphism. These loci are usually located in intron and their sequences are comprised of tan dem repetition of certain unique sequences, or core sequences. With individuals, the repetition number is variable producing numerous alleles. Nowadays these VNTR loci are widely used for "DNA fingerprint", individual identification with DNA test. Some researchers have reported genetic polymorphism in D17S5 (Horn *et al.*, 1989), 3' hypervariable region of the apolipoprotein B gene (Boerwinkle *et al.*, 1989, Ludwig *et al.*, 1989)and D1S80 (Budowle *et al.*, 1991, Kasai *et al.*, 1990). Recently numerous repetitive loci with short core sequences usually of two to six nucleotides, STR (Short Tandem Repeat) have been found, and its utility as a genetic marker is increasing. Amp-FLP is a popular method used to reveal these polymorphisms. Genetic polymorphism in STR of Myelin Basic Protein gene (Boylan *et al.*, 1990) and several STR with trimeric and tetrameric STR (Edwards *et al.*, 1991) have reported.

For practical use, allele frequency and genotype distribution must be known in a population, and whether the population obeys the Hardy-Weinberg distribution is an important problem. These population data are known to be different according to race, so we should have our own genetic data in order to use it in Koreans. SE33 is a kind of STR in beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP-2) on chromosome 6 as a tetrameric [AAAG]n repeat. According to Polymeropoulos (1992) 21 alleles were noted. The current report describes the availability of the SE33 in a Korean population and the basic data such as allele frequency and genotype distribution.

Material and Methods

DNA was isolated from 268 unrelated placentas and blood as described elsewhere (Kunkel *et al.*, 1982). After quantitation with a DNA fluorometer (Hoefer Scientific Inc.), the DNA was

used as a template. Amplification was done using two primers, the forward, 5' AAT CTG GGC GAC AAG AGT GA 3' and the reverse, 5' ACA TCT CCC CTA CCG CTA TA 3' with a GeneAmp PCR system 9600 (Perkin Elmer Cetus, USA). 10 nanograms of the above DNA were amplified in 20 μl of a reaction mixture containing 0.3 mM dNTP (each), 50mM KCl, 10mM tris-Hcl pH 8.3, 1.5 mM $MgCl_2$, 0.01% gelatin with 0.8 μM of each primer and 1 unit of Taq polymerase. After denaturation of the DNA at 94℃ for 20 sec, annealing and extension was done at 67℃ for 1 min. The reaction was repeated for 10 cycles and followed by 23 cycles of 94℃ 20 sec, 62℃ 1 min. The product was separated on 7% acrylamide gel and was revealed by silver staining. Briefly described, the gel was treated with 10% ethanol for 5 min, followed by oxidation in nitric acid for 5 min, 0.012 M $AgNO_3$ treatment for 12 min, 10% glacial acetic acid. Thorough washing with distilled water between each procedure was done. After staining the gel was vacuum dried with 3M paper.

For the confirmation of exact allele size, the PCR product was separated using an automatic DNA sequencer Licor-4000L (Li-cor, Lincoln Nebraska, USA). The sequencer used infrared fluorophore. For the detection we modified the forward primer with M13 phage primer tailing (CAC GAC GTT GTA AAA CGA C) to 5' side. The PCR condition was the same except for the primer composition, (+)-strand primer with tail 0.25 μM , (-)-strand primer 0.25 μM, IR-M13 forward primer 0.5 pM. For gel electrophoresis, 1.0 μl of PCR product with appropriate dilution was loaded on 41 cm long, 0.25mm thick, 7% Long Ranger denaturing gel. 0.4×TBE buffer was used and electrophoresis was done with parameters 2000 volts, 35 mA, 70 watts at 50℃.

The family study was done in 164 gametes of 64 families, whose parent-child relationship were confirmed through other studies including HLA-typing and other VNTR typing, as well as history taking.

Table 1. SE33 Allele Frequencies. The allele number was labelled in order of allele size, and did not mean the exact repetition number.

Allele No	Included No	Frequency (%)	Allele No	Included No	Frequency (%)
2	4	0.00746	23	14	0.02612
3	1	0.00187	24	32	0.05970
4	4	0.00746	25	4	0.00746
5	1	0.00187	26	28	0.05224
6	13	0.02425	27	6	0.01119
8	25	0.04664	28	33	0.06156
10	38	0.07090	29	11	0.02052
12	52	0.09701	30	40	0.07462
13	2	0.00373	31	2	0.00373
14	43	0.08022	32	34	0.06343
15	3	0.00560	33	4	0.00746
16	31	0.05784	34	22	0.04104
17	8	0.01493	35	1	0.00187
18	12	0.02239	36	17	0.03172
19	11	0.02052	37	2	0.00373
20	8	0.01493	38	6	0.01119
21	6	0.01119	39	1	0.00187
22	16	0.02985	40	1	0.00187
			36	536	1

RESULTS

After the silver staining the majority of alleles could be seen as a discrete band ranging from about 240 bp to 340 bp. But on acrylamide gel, some alleles of similar size could not be differentiated from each other. So we tried allele separation using sequencing gel with an automatic sequencer. The above alleles separated clearly with two base differences. On sequencing gel 36 alleles were found among which 16 kinds were of two base repeats. Allele 12 was the most frequent with the frequency of 0.09701. Other alleles, allele 14 and allele 30 were also frequently encountered with frequencies of 0.08022 and 0.07462 respectively. None of the alleles were encountered with a frequency over 0.1, and many of the alleles (13 kinds) were encountered less than five in sample number (rare allele), and this made statistical assay difficult. The alleles with their observed frequencies are illustrated in Table 1.

As a combination of the above alleles, 159 kinds of genotypes were noted. Among the 268 DNAs studied, 22 were homozygotes, and the remaining 246 were heterozygotes, the heterozygosity was 91.8 %. The observed genotypes with each frequency are shown in Table 2. Although the No of samples was not small, most genotypes contained less than 5 samples due to the enormous variety of genotypes, so statistical analysis using classical x^2 test could not be done.

In the 64 family study composed of 164 gametes, Mendelian inheritance was confirmed. All the alleles of the children derived from their parents, one from the father, the other from the mother. When one of the parents was absent, the genotype could easily be determined through other family members. Through 164 gametes two cases of mutation were noted. In these two families, one of the child's alleles differed from that of the parent's, and the parent-child relationship in these two families was confirmed through other studies including VNTR loci typing and HLA typing.

DISCUSSION

Use of PCR has many advantages over RFLP. First, we can perform a test with only a small amount of sample, and moreover even with a degraded sample. If the target sequences are intact, reproducible results can be obtained even with a single copy gene as in a sperm (Higuchi *et al.*, 1988). Second, the alleles appear in discrete bands and this reduces measurement error. With RFLP about thirty to fourty base pairs cannot be differentiated due to its low resolution when the band size is over 1000-2000 bases of nucleotide. This results in pseudohomozygotes and quasicontinuous distribution of alleles producing statistical chaos in analyzing the results. Moreover, avoidance of isotopes is another practical advantage. After silver staining the results can be permanently stored. Especially the Amp-FLP procedure, which separate the alleles only with their size, is valuable in VTNR or STR loci typing. As in other VNTR loci the polymorphism in SE33 locus can easily be revealed with Amp-FLP, and this is simple and rapid. The 4 base pair differences can easily be separable in a high percentage of acrylamide gel electrophoresis. But in case of SE33, some of the alleles were difficult to size exactly. To solve this problem we used an automatic sequencer. The sequencing gel can separate alleles with only a base pair difference. Contrary to previous classic isotope labelling, newly introduced automatic sequencers use infrared or ultravisual fluorophore, and this make sequencing more easy and convenient, and this can also be used in STR typing. The Li-cor 4000L automatic sequencer using infrared fluorophore, and the Long Ranger gel we used, the commercial trade name for high performance alternative to acrylamide/bis-acrylamide stock solution, makes the gel condition more profitable. After sizing of the SE33 alleles with an automatic sequencer, many alleles of two base repeats, which were not detected by conventional acrylamide gel electrophoresis, were noted. According to Polymeropoulos (1991) the SE33 locus is a STR with four base repeats, and this is contrary to

Table 2. Genotype distribution of SE33

Genotype	observed		Genotype	observed	
	No	Frequency		No	Frequency
2-18	1	0.00373	10-22	4	0.01493
2-30	1	0.00373	10-23	1	0.00373
2-34	1	0.00373	10-24	2	0.00746
2-36	1	0.00373	10-26	4	0.01483
3-32	1	0.00373	10-27	1	0.00373
4-14	1	0.00373	10-30	2	0.00746
4-27	1	0.00373	10-31	1	0.00373
4-32	1	0.00373	10-32	1	0.00373
4-36	1	0.00373	10-34	1	0.00373
5-26	1	0.00373	10-36	1	0.00373
6-12	1	0.00373	12-14	3	0.01120
6-14	3	0.01120	12-15	1	0.00373
6-16	2	0.00746	12-16	5	0.01866
6-18	2	0.00746	12-17	2	0.00746
6-26	1	0.00373	12-19	2	0.00746
6-30	1	0.00373	12-22	1	0.00373
6-36	1	0.00373	12-23	1	0.00373
8-10	1	0.00373	12-24	6	0.02239
8-13	1	0.00373	12-26	1	0.00373
8-14	1	0.00373	12-28	6	0.02239
8-16	1	0.00373	12-29	2	0.00746
8-17	1	0.00373	12-30	7	0.02612
8-18	4	0.01493	12-32	4	0.01493
8-20	2	0.00746	12-34	2	0.00746
8-22	2	0.00746	12-38	1	0.00373
8-24	1	0.00373	13-27	1	0.00373
8-25	1	0.00373	14-15	1	0.00373
8-26	1	0.00373	14-16	2	0.00746
8-28	3	0.01120	14-19	1	0.00373
8-29	1	0.00373	14-20	1	0.00373
8-32	1	0.00373	14-22	3	0.01120
8-34	1	0.00373	14-23	2	0.00746
8-36	1	0.00373	14-24	6	0.02239
8-38	1	0.00373	14-26	5	0.01866
8-39	1	0.00373	14-28	2	0.00746
10-12	2	0.00746	14-30	1	0.00373
10-14	1	0.00373	14-32	2	0.00746
10-16	2	0.00746	14-36	1	0.00373
10-17	2	0.00746	14-38	1	0.00373
10-18	1	0.00373	15-28	1	0.00373
10-19	2	0.00746	16-20	1	0.00373
10-21	1	0.00373	16-22	1	0.00373

continued

Continuing

Genotype	observed		Genotype	observed	
	No	Frequency		No	Frequency
16-24	2	0.00746	23-24	1	0.00373
16-25	2	0.00746	23-27	1	0.00373
16-26	2	0.00746	23-29	1	0.00373
16-28	3	0.01120	23-30	2	0.00746
16-29	1	0.00373	23-35	1	0.00373
16-30	1	0.00373	24-28	1	0.00373
16-32	1	0.00373	24-30	2	0.00746
16-33	1	0.00373	24-32	3	0.01120
16-34	2	0.00746	24-34	3	0.01120
16-38	1	0.00373	24-36	2	0.00746
17-23	1	0.00373	25-30	1	0.00373
17-32	1	0.00373	26-28	2	0.00746
17-34	1	0.00373	26-30	2	0.00746
18-22	1	0.00373	26-32	1	0.00373
18-26	1	0.00373	26-34	1	0.00373
18-30	1	0.00373	27-29	1	0.00373
18-32	1	0.00373	27-33	1	0.00373
19-26	1	0.00373	28-30	6	0.02239
19-28	2	0.00746	28-32	3	0.01120
19-30	1	0.00373	28-34	1	0.00373
19-36	2	0.00763	28-36	1	0.00373
20-32	3	0.01120	29-31	1	0.00373
20-37	1	0.00373	30-32	4	0.01493
21-29	1	0.00373	30-34	4	0.01493
21-34	1	0.00373	32-34	2	0.00746
21-36	1	0.00373	32-36	2	0.00746
21-37	1	0.00373	32-38	1	0.00373
22-24	1	0.00373	34-36	1	0.00373
22-26	1	0.00373	36-38	1	0.00373
22-30	2	0.00746	36-40	1	0.00373

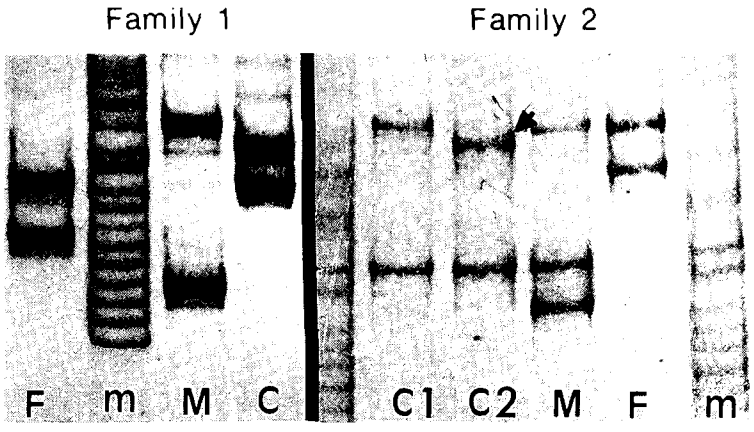

Fig 1. Silver stained polyacrylamide gel after PCR amplification of SE33 in DNA from two families showing mutation. The letter denotes F = father, M = mother, C = child and m = size marker. Bands of similar size in family 2 (arrow) were confirmed using sequencing gel showing two base differences.

our data. As reports on this locus in various races are not available, it is uncertain whether this phenomenon is confined to Koreans or not. Anyhow our report suggests, in case of STR typing with four or six base repeats, co-exitence of shorter core sequence can be possible. So routine acrylamide gel electrophoresis should be carefully used in STR typing. And for this an automatic sequencer is a very useful tool.

In practical application, probability to identification is essential in criminal cases. Thus population data is necessary. Usually the allele distribution is different according to race. So this study can provide valuable population data in Koreans.

In order to apply a genetic locus in personal identification, it is important to confirm that the locus obeys the Hardy-Weinberg distribution. For this the observed frequency of genotypes is compared with the expected one which derives from observed allele frequency. x^2 test is usually used for statistical analysis, which needs at least more than 5 in each sample number. Our population data is characteristic in numerous alleles and genotype numbers compared to others, which is the result of many two base repeats. Although our sample number was not so small compared to routine studies, most of the sample numbers per genotype were less than 5. So routine x^2 cannot be done, and this entails a prudent approach to probability calculation.

According to Jeffreys(1992), the VNTR loci show a relatively high rate of mutation because the core sequence is similar to x sequence, the recombination signal in E. coli. But the mutation rate is quite different according to the genetic loci. About STR loci, mutation rate is rarely available at present. Our results show that mutation could be very rare in STR loci. It is recommended that in case of doubt in regard to paternity tests, we must type another loci to confirm the relation. Sequencing the mutant allele in the above family member may give a clue to the mechanism of the mutation in STR or the mechanism of repetition formation.

REFERENCES

Baird M, Balazs I, Guisty A, Miyazaki L, Wexler K, Kanter E, Glassberg J, Allen F, Rubinstein P, Sussman L. Allele frequency distribution of two highly polymorphic DNA sequences in three ethnic groups and its application to the determination of paternity. Am J Hum Genet 1986; 39:489-501

Boerwinkle E, Xiong W, Fourest E, Chan L. Rapid typing of tandemly repeated hypervariable loci by polymerase chain reaction: application to the apolipoprotein B 3' hypervariable region. Proc Natl Acad Sci USA 1989; 86: 212-216

Boylan KB, Ayres TM, Popko B, Takahashi N, Hood LE, Prusiner SB. Repetitive DNA(TGGA)n 5' to the Human Myelin Basic Protein Gene: A New form of Oligonucleotide Repetitive Sequence Showing Length Polymorphism. Genomics 1990; 6:16-22

Budowle B, Chakraborty R, Guisti AM, Eisenberg AJ, Allen RC. Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am J Hum Genet 1991; 48:137-144

Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and Genetic Mapping with Trimeric and tetrameric Tandem Repeats. Am J Hum Genet. 1991; 49:746-756

Fowler JCS, Scott AC. Examination of the correlation of groupings in blood and semen. J Foren Sci 1985; 30:103-113

Gill P, Jeffreys AJ, Werrett DJ. Forensic application for DNA fingerprints. Nature 1985; 318: 577-579

Giusti A, Baird M, Pasquale S, Balazs I, Glassberg J. Application of DNA polymorphism to the analysis of DNA recovered from sperm. J Foren Sci 1986; 31:409-417

Highichi LI, Gyllensten UB, Cui XB, Saiki RK, Erlich HA, Arnheim N. Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 1988; 335:414-417

Honma M, Ishiyama I. DNA fingerprinting: the

importance in forensic medicine(1) Application to paternity testing by minisatellite DNA probes. Jap J Legal Med 1987; 41:236-241

Honma M, Yoshii T, Ishiyama I, Mitani K, Ko, ima, i R, Muramatsu M. Individual identification from sperm by DNA fingerprint technique. J Foren Sci 1989; 34:222-227

Horn GT, Richard B, Klinger KW. Amplification of a highly polymorphic VNTR segment by polymerase chain reaction. Nucleic Acids Res 1989; 17:2140

Jeffreys AJ, Wilson V, Them SL. Hypervariable minisatellite regions in human DNA. Nature 1985; 314:67-73

Kanter E, Baird M, Shaler R, Balazs I. Analysis of restriction fragment length polymorphisms in DNA recovered from dried bloodstains. J Fores Sci 1986; 31:403-408

Kasai K, Nakamura Y, White R. Amplification of a variable number of tandem repeate(VNTR) locus(pMCT118) by the polymerase chain reaction(PCR) and its application to forensic science. J Foren Sci 1990; 35:1196-1200

Kevin BB, Ayres TM. Repetitive DNA(TGGA)n 5' to the human myelin basic protein gene:A new form of oligonucleotide repetitive sequence showing length polymorphism. Genomics 1990; 6:16-22

Kunkel LM, Tatravahi U, Eisenhard M, Latt SA Regional localization on the human X of DNA segments cloned from flow sorted chromosomes. Nucleic Acids Res 1982; 10:1557-1578

Lawlor DA, Dickel CD, Hauswirth WW, Parham P. Ancient HLA genes from 7500-year-old archaeological remains. Nature 1991; 349: 785-788

Lee JB. Amplification of VNTR locus D1S80 in DNA recovered from vaginal sperm. Seoul J Med 1991a; 32:149-154

Lee JB. Some VNTR loci typing from single hair. Seoul J Med 1991b; 32:209-215

Ludwig EH, Friedl W, McCarthy BJ. High-resolution analysis of a hypervariable region in the human apolipoprotein B gene. Am J Hum Genet 1989; 45:458-464

Polymeropoulos MH, Rath Ds, Xiao H, Merril CR. Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2(ACTBP 2). Nucleic Acids Research 1991; 20:1432

Newall PJ. The identification of HLA-A9 in dried stains of blood and body secretions. Canad Society Foren Sci 1981; 14:104-112

Saiki RK, Scharf SJ, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnhein N. Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 1985; 230:1350-1354

Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 1989; 44:388-396

Wong Z, wilson V, Patel I, Povey S, Jeffreys AJ. Characterization of a panel of highly variable minisatellites cloned from human DNA. Ann Hum Genet 1987: 51:269-288