Int J Legal Med (1997) 110 : 208–212

ORIGINAL ARTICLE

C. Liu · N. Harashima · Y. Katsuyama · M. Ota
A. Arakura · H. Fukushima

ACTBP2 gene frequency distribution and sequencing of the allelic ladder and variants in the Japanese and Chinese populations

Received: 22 November 1996 / Received in revised form: 22 January 1997

Abstract The human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) gene frequency distributions in the Japanese and Chinese Han populations were investigated and compared. Analysis was carried out by applying fluorescently labeled samples and a differently labeled sequenced allelic ladder within the same lanes in denaturing gels, followed by laser detection and automated analysis using Genescan software 672. The discrimination index and the heterozygosity index were calculated to be 0.993 and 0.916 in the Japanese population, and 0.993 and 0.944 in the Chinese Han population, respectively. No deviation from Hardy-Weinberg equilibrium was observed in these two populations. The allelic ladder, which ranged from 233 bp to 319 bp, was constructed from a combination of 23 regularly occurring alleles. The allelic ladder and 12 variants observed in 24 individuals in these two populations were sequenced. The variants could be divided into three types according to their structural variation characteristics. These variants differed from the alleles of the same repeats in the allelic ladder by the presence or absence of hexanucleotides in the central repeat regions, base deletions in the flanking regions, and base insertions in the repeat units.

Key words STR · ACTBP2 · Gene frequency · Variants · Allelic ladder · Direct sequencing · Cloning

Introduction

Short tandem repeat (STR) loci are important polymorphic markers for paternity testing and personal identification. Among the known STR systems, ACTBP2 is one of the most polymorphic and informative (Möller et al. 1995). Compared with other STR systems such as HumFGA, HumFES/FPS, HumvWA, D18S51, and D8S1179, some of which possess much structural variation (Barber et al. 1995,1996; Barber and Parkin 1996), it exhibits higher degrees of length polymorphism as well as sequence polymorphism. The discrimination index (DI) of this system can reach >99% (Wiegand et al. 1993) which equals the total DI of three other STR systems, each with a DI of approximately 80% (Möller et al. 1995).

Because of its high applicability in forensic practice, ACTBP2 is gaining increasing attention worldwide. Some population database studies have been carried out on Caucasians, as well as on some ethnic populations including a western Japanese population and a southwestern minority of the Chinese population (Urquhart et al. 1993; Wiegand et al. 1993; Meyer et al. 1995; Möller et al. 1995; Brinkmann et al. 1996). In recent sequencing studies conducted on Caucasians, Möller and Brinkmann (1994), Möller et al. (1995) and Brinkmann et al. (1996) observed sequence variations such as base deletions, insertions, and transitions in the flanking regions, and sometimes in the repeat regions as well. To date, few reports have been published concerning gene frequency distribution in Japanese population samples obtained from the whole nation and in Chinese Han population samples, as well as on sequencing data of these two populations.

In this study, we investigated and compared the ACTBP2 gene frequency distributions in Japanese population samples collected from the whole nation and in Chinese Han population samples, and sequenced the allelic ladder and 12 variants observed in 24 individuals in these two populations.

Materials and methods

DNA extraction

DNA was extracted from blood samples of unrelated Japanese (n = 179) from the whole nation and Chinese (n = 164) of Han nationality in the northeast part of China, using the proteinase K/SDS lysis and phenol extraction method (Inoko et al. 1986).

C. Liu · N. Harashima · Y. Katsuyama · M. Ota · A. Arakura
H. Fukushima (✉)
Department of Legal Medicine,
Shinshu University School of Medicine, Asahi 3–1–1,
Matsumoto 390, Japan

PCR

PCR was performed in a 25 μl reaction mixture containg 0.5 μM of each primer, 200 μM of each dNTP, 2.5 U Taq polymerase and corresponding buffer (TaKaRa, Japan).

Primer sequences (Polymeropoulos et al. 1992):

I: 5′-AATCTGGGCGACAAGAGTGA-3′

II: 5′-ACATCTCCCCTACCGCTATA-3′.

To determine the size of an allele, one of the primers was 5′ labeled with a fluorescent dye, HEX or TET. Amplification was performed as described by Harashima et al. (1996) in a DNA Thermal Cycler 480 (Perkin Elmer).

Allele typing

Samples were automatically analyzed on an ABI 373A sequencer using Genescan software 672, following electrophoresis on a 6% denaturing PAG (acrylamide/bisacrylamide 19:1, containing 8.3 M urea) in 1 X TBE buffer (89 mM Tris, 89 mM boric acid, 2.2 mM EDTA-Na_2, pH8.3) at 40 mA, 30 W, 40° C for 10–12 h. ROX-labeled GS2500 and an allelic ladder labeled with FAM were used as internal markers.

Allele purification

Samples of known types were run on a 6% native PAG in 1 X TBE buffer. After ethidium bromide staining target alleles were excised from the gel and eluted in TE buffer (Maniatis et al. 1989). The eluate was re-amplified, and purified with Microcon-100 (Amicon).

Direct sequencing

The sequencing reaction was performed using the Dye Terminator Cycle Sequencing Ready Reaction mix (Perkin-Elmer) with the purified alleles as templates, and primer I or II as the primer for sequencing the forward or reverse strand, respectively. Amplification was carried out at 96° C-30 s, 50° C-15 s, 60° C- 4 min for 25 cycles, using a DNA Thermal Cycler 480 (Perkin-Elmer). The reaction products were purified by phenol/chloroform extraction and alcohol precipitation, resuspended and boiled in loading buffer (formamide/50 mM EDTA 5:1) for 3 min before electrophoresis on a 6% denaturing polyacrylamide gel in 1 X TBE buffer on an ABI 373A sequencer. The sequences were automatically analyzed using the Data Collection and SeqEd software (Perkin-Elmer).

Cloning and sequencing

The purified alleles were ligated to the pBlue ScriptII SK+ l plasmid vector through the T-A cloning method (Holton and Graham 1991; Marchuk et al. 1991) and transformed into NovaBlue Competent Cells. Recombinant colonies were randomly selected on LB plates. Plasmids were isolated from the culture by the alkaline extraction method (Birnboim and Doly 1979). The purified plasmids were used as templates for the sequencing reaction, using T3 or T7 as a primer. The reaction conditions and analysis methods were the same as those used in direct sequencing.

Results and discussion

Of the regularly occurring alleles, 23 were selected and combined to construct an evenly spaced allelic ladder ranging from 233 bp to 319 bp. (Fig. 1). The sizes of the allelic ladder were determined by sequencing. By applying the samples and the allelic ladder within one lane of a denaturing gel, the sizes of the alleles were obtained in accordance with the sizes of the matching alleles in the allelic ladder. The measurements by the internal marker GS2500 were usually longer than the sizes determined by sequencing, but they were useful for inferring the sizes of the alleles which did not match the allelic ladder.

Gene frequency distribution in Japanese and Chinese populations

Table 1 shows the gene frequency distributions in the Japanese and Chinese Han populations. In both populations, 22 alleles were found, ranging from alleles 12 to 35.2. In the Japanese population, the high frequency alleles were alleles 27.2, 29.2, 28.2, 19, 20, 21, 25.2, and 26.2 in decreasing order and in the Chinese Han population alleles 26.2, 18, 25.2, 20, and 24.2. Alleles 12 and 13 were found only in the Japanese, while alleles 33.2 and 35.2 were found only in the Chinese. Alleles 18, 26.2, and 29.2 were observed to have a significant distribution difference between these two populations using the χ^2-test.

Moreover, 12 alleles which did not match the allelic ladder and thus were called variants were found in the Japanese and 6 were found in the Chinese. Alleles 17(–2),

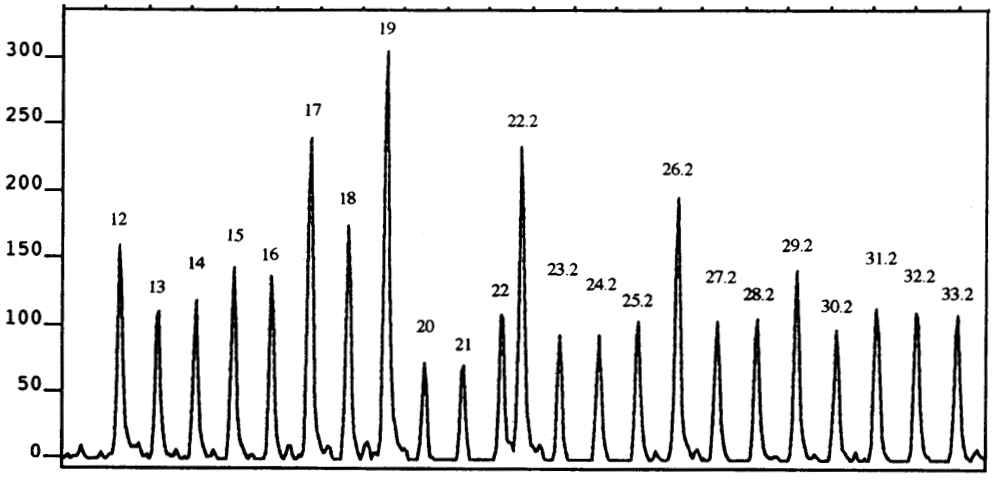

Fig. 1 Electropherogram of the allelic ladder composed of 23 regularly occurring alleles

Table 1 ACTBP2 gene frequency distributions in the Japanese and Chinese Han populations. *: designation showing only the length of the allele

Allele	Japanese		Chinese	
	Number	Frequency	Number	Frequency
12	1	0.0028	0	0
13	1	0.0028	0	0
14	2	0.0056	6	0.0183
14.2	0	0	1	0.003
15	2	0.0056	6	0.0183
16	11	0.0307	9	0.0274
17(–2)	1	0.0028	0	0
17	10	0.0279	15	0.0457
18	18	0.0503	30	0.0915
19	30	0.0838	18	0.0549
19.2*	1	0.0028	0	0
20	27	0.0754	21	0.064
20.2	2	0.0056	4	0.0122
21	26	0.0726	15	0.0457
21.2	3	0.0084	3	0.0091
22	4	0.0112	8	0.0244
22.2	11	0.0307	14	0.0427
23	1	0.0028	2	0.0061
23.2	12	0.0335	11	0.0335
24	1	0.0028	1	0.003
24.2	19	0.0531	20	0.061
25	1	0.0028	0	0
25.2	22	0.0615	29	0.0884
26.2	22	0.0615	36	0.1098
27.2(–2)	1	0.0028	0	0
27.2(–1)	0	0	1	0.003
27.2	34	0.095	17	0.0518
27.3	1	0.0028	0	0
28.2(–2)	1	0.0028	0	0
28.2	31	0.0866	16	0.0488
28.3*	1	0.0028	0	0
29.2	32	0.0894	17	0.0518
30.2	13	0.0363	15	0.0457
31.2	11	0.0307	7	0.0213
32.2(–1)	1	0.0028	0	0
32.2	4	0.0112	4	0.0122
33.2	0	0	1	0.003
35.2	0	0	1	0.003
Total	358	1	328	1
DI	0.993		0.993	
HET	0.916		0.944	
MEC	0.874		0.880	
PIC	0.935		0.938	

19.2*, 25, 27.2(–2), 27.3, 28.2(–2), 28.3*, and 32.2(–1) were found only in the Japanese, while alleles 14.2 and 27.2(–1) were found only in the Chinese. In total, 14 variants were found in 27 individuals in these two populations. Alleles 20.2 and 21.2 were found to be commonly occurring variants.

The values of the discrimination index (DI), the heterozygosity index (HET), the mean exclusion chance (MEC), and the polymorphic information content (PIC) were calculated to be 0.993, 0.916, 0.874, 0.935, and 0.993, 0.944, 0.880, 0.938 in the Japanese and Chinese populations, respectively. No deviations from Hardy-Weinberg equilibrium were observed in these two populations after calculation applying the exact test (Guo and Thompson 1992).

Table 2 Sequences summary of the ACTBP2 allelic ladder. The fourth column shows the base at position 20 in the 3′ flanking region counting from the end of the primer II. FR: flanking region

Allele designation	Length	Repeat region	Base at position 20 3′-FR
12	233	$(AAAG)_{12}$	A
13	237	$(AAAG)_{13}$	A
14	241	$(AAAG)_{14}$	A
15	245	$(AAAG)_{15}$	A
16	249	$(AAAG)_{16}$	A
17	253	$(AAAG)_{17}$	A
18	257	$(AAAG)_{18}$	A
19	261	$(AAAG)_{19}$	A
20	265	$(AAAG)_{20}$	A
21	269	$(AAAG)_{21}$	A
22	273	$(AAAG)_{22}$	A
22.2	275	$(AAAG)_{11}$ AAAAAG $(AAAG)_{10}$	G
23.2	279	$(AAAG)_{10}$ AAAAAG $(AAAG)_{12}$	G
24.2	283	$(AAAG)_{9}$ AAAAAG $(AAAG)_{14}$	G
25.2	287	$(AAAG)_{9}$ AAAAAG $(AAAG)_{15}$	G
26.2	291	$(AAAG)_{14}$ AAAAAG $(AAAG)_{11}$	G
27.2	295	$(AAAG)_{11}$ AAAAAG $(AAAG)_{15}$	G
28.2	299	$(AAAG)_{10}$ AAAAAG $(AAAG)_{17}$	G
29.2	303	$(AAAG)_{11}$ AAAAAG $(AAAG)_{17}$	G
30.2	307	$(AAAG)_{12}$ AAAAAG $(AAAG)_{17}$	G
31.2	311	$(AAAG)_{12}$ AAAAAG $(AAAG)_{18}$	G
32.2	315	$(AAAG)_{14}$ AAAAAG $(AAAG)_{17}$	G
33.2	319	$(AAAG)_{11}$ AAAAAG $(AAAG)_{21}$	G

General sequence of an allele

As sequence polymorphism of the ACTBP2 alleles often occurs in the flanking region, it is necessary to understand the complete sequence of an allele rather than only the central repeat region. However, although sequencing of the forward and reverse strands was performed at the same time, the sequences of about 10 bases at the beginning or end of a flanking region could not be easily determined because the peaks were often overlapped by background “noise”. Thus cloned fragments of these samples were also sequenced in order to obtain the full sequences of the alleles. For each variant at least three clones were tried, and direct sequencing was repeated at least 3 times.

A typical ACTBP2 allele was composed of the primer regions, the 5′ flanking region with 122 bases, central repeat region, and the 3′ flanking region with 23 bases (Fig. 2). The repeat unit was the tetranucleotide AAAG, inter-

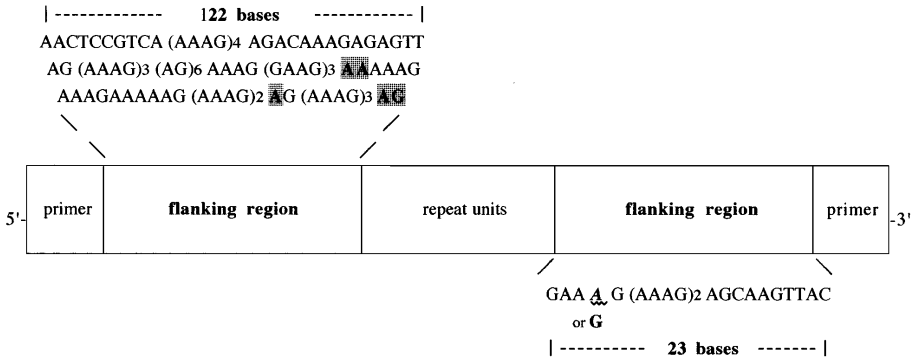

Fig. 2 ACTBP2 allele sequence model of the AAAG strand. In the 5′ flanking region, the shaded bases (AA, A and AG) are those positions where base deletions occurred. In the 3′ flanking region, the A at position 20 (counting from the end of primer II) was transverted into G in some alleles

Table 3 Sequences summary of the ACTBP2 variants. The fourth column shows the base at position 20 in the 3′ flanking region counting from the end of the primer II. FR: flanking region

Type	Allele designation	Repeat region	Base at posion 20, 3′-FR	Deletions in 5′-FR (position)	Number of sequenced alleles
I	14.2	$(AAAG)_4AAAAAG(AAAG)_9$	A		1
II	17(–2)	$(AAAG)_{17}$	A	AG deletion (121, 122)	1
I	20.2	$(AAAG)_8AAAAAG(AAAG)_{11}$	G		2
I	20.2	$(AAAG)_9AAAAAG(AAAG)_{10}$	G		1
I	20.2	$(AAAG)_{10}AAAAAG(AAAG)_9$	G		3
I	21.2	$(AAAG)_9AAAAAG(AAAG)_{11}$	G		1
I	21.2	$(AAAG)_{10}AAAAAG(AAAG)_{10}$	G		4
I	21.2	$(AAAG)_{11}AAAAAG(AAAG)_9$	G		1
I	23	$(AAAG)_{23}$	A		2
I	24	$(AAAG)_{24}$	A		2
I	25	$(AAAG)_{25}$	G		1
II	27.2(–2)	$(AAAG)_8AAAAAG(AAAG)_{18}$	G	AG deletion (121, 122)	1
II	27.2(–1)	$(AAAG)_8AAAAAG(AAAG)_{18}$	G	A deletion (107)	1
III	27.3	$(AAAG)_{10}AAAAAG(AAAG)_9AAAAG(AAAG)_6$	G		1
II	28.2(–2)	$(AAAG)_8AAAAAG(AAAG)_{19}$	G	AA deletion (83, 84)	1
II	32.2(–1)	$(AAAG)_8AAAAAG(AAAG)_{23}$	G	A deletion (107)	1

rupted in some alleles by a hexanucleotide unit AAAAAG at different sites.

Nomenclature and sequences of the allelic ladder

Following the ISFH recommendations (1994), the 23 component alleles in the allelic ladder were designated as alleles 12 to 33.2 (Table 2). Alleles having only AAAG repeats were denominated with full numbers referring to the number of repeats. Alleles showing an AAAAAG hexanucleotide repeat motif were denominated with a number referring to the number of AAAG repeats, followed by a ".2" for the "extra" AA dinucleotide, as the hexanucleotide could be considered as an AAAG repeat inserted with an AA dinucleotide.

From allele 12 to allele 22, in the 3′ flanking region, the base at position 20 counting from the end of primer II was A, while from allele 22.2 to allele 33.2, this A was replaced by G (Table 2, Fig. 2). The remaining parts of the flanking regions of all component alleles were the same as that shown in Fig. 2. No insertion or deletions were found in the flanking regions.

Sequences of the variants

Except for three alleles (19.2*, 28.3*, and one of the three alleles 23) which were difficult to separate, all the sizes and designations of the variants found in these two populations were finally determined by sequencing. The deletions in the flanking regions were shown in brackets as (–) together with the number of deleted base pairs. The sequenced variants in these two populations could be divided into three types according to their sequence characteristics (Table 3).

Type I variants were characterized by the presence or absence of the hexanucleotide unit within the repeat region. Altogether six alleles 20.2 and six alleles 21.2 were detected in these two populations and could be separated into three subtypes each due to the occurrence of the hexanucleotide at different positions within the repeat region (Table 3).

Type II were base deletions in the 5′ flanking region. Base deletions were found easily occurring at position 107 [alleles 27.2(–1), 32.2(–1)], positions 121, 122 [alleles 17(–2), 27.2(–2)], and positions 83, 84 [allele 28.2(–2)], counting from the end of primer I (Table 3, Fig. 2).

Type III was base insertion in the central repeat region (allele 27.3) (Table 3). This kind of structure variation in the repeat units was only observed in the Japanese and has not been reported in any other population.

According to the EDNAP group (Gill et al. 1994), the most reliable typing method in this system involves the application of a denaturing gel system as well as an automated laser fluorescence detection system, with a sequenced allelic ladder as internal marker. However, with this method, the genotypes defined by the allelic ladder reflect only the length polymorphism of the samples, and the actual sequence polymorphism of the samples cannot be determined. Moreover, one designation of an allele based on length may actually be shared by a group of alleles, which are of the same size, but vary in the sequences such as in the number of repeats or in the flanking regions due to base insertions or deletions, while the designation defined through sequencing is unique for the allele. Nevertheless, using this method, the population data from different laboratories can be compared at the level of length polymorphisms.

The variations caused by the hexanucleotide AAAAAG, which may be called a positional polymorphism, is interesting and noteworthy. As our results have shown that alleles 20.2 and 21.2 could be further separated into three subtypes each due to this positional polymorphism, and this polymorphism was also observed in a German population (Möller et al. 1995), there is a great possibility that the positional polymorphism exists in all commonly occurring alleles in all populations. The detection of this positional polymorphism and other sequence polymorphisms may extremely elevate the heterozygosity and the discrimination index. While it is unpractical to sequence all population samples to screen for these polymorphisms, an efficient method is necessary to be developed to make them easily detected and utilized in the forensic practice.

References

Barber MD, Parkin BH (1996) Sequence analysis and allelic designation of the two short tandem repeat loci D18S51 and D8S1179. Int J Legal Med 109:62–65

Barber MD, Piercy RC, Andersen JF, Parkin BH (1995) Structural variation of novel alleles at the Hum vWA and Hum FES/FPS short tandem repeat loci. Int J Legal Med 108:31–35

Barber MD, McKeown BJ, Parkin BH (1996) Structural variation in the alleles of a short tandem repeat system at the human alpha fibrinogen locus. Int J Legal Med 108:180–185

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plamid DNA. Nucleic Acids Res 7:1513–1523

Brinkmann B, Sajantila A, Goedde HW, Matsumoto H, Nishi K, Wiegand P (1996) Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci. Eur J Hum Genet 4:175–182

DNA recommendations – 1994 report concerning further recommendations of the DNA commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. Forensic Sci Int 69:103–104

Gill P, Kimpton C, D'Aloja E, Andersen JF, Bär W, Brinkmann B, Holgersson S, Johnsson V, Kloosterman AD, Lareu MV, Nellemann L, Pfitzinger H, Phillips CP, Schmitter H, Schneider PM, Stenersen M (1994) Report of the European DNA profiling group (EDNAP) – towards standardization of short tandem repeat (STR) loci. Forensic Sci Int 65:51–59

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Harashima N, Liu CY, Katsuyama Y, Ota M, Hama Y, Fukushima H (1996) Automated DNA typing of ACTBP2: a comparison of allelic ladder with GS2500 as internal marker. Nippon Hoigaku Zasshi 50(1):27–29

Holton TA, Graham MW (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res 19:1156

Inoko H, Ando A, Ito M, Tsuji K (1986) Southern hybridization analysis of DNA polymorphism in the HLA-D region. Hum Immunol 16:304–313

Maniatis T, Fritsch EF, Sambrook J (1989) Molecular cloning: a laboratory manual, Cold Spring Harbour Laboratory, Cold Spring Harbour, New York

Marchuk D, Drumm M, Saulino A, Collins FS (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res 19:1154

Meyer E, Wiegand P, Brinkmann B (1995) Phenotype differences of STRs in 7 human populations. Int J Legal Med 107:314–322

Möller A, Brinkmann B (1994) Locus ACTBP2 (SE33): sequencing data reveal considerable polymorphism. Int J Legal Med 106:262–267

Möller A, Schürenkamp M, Brinkmann B (1995) Evaluation of an ACTBP2 ladder composed of 26 sequenced alleles. Int J Legal Med 108:75–78

Polymeropoulos MH, Rath DS, Xiao H, Merril CR (1992) Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2). Nucleic Acids Res 20:1432

Urquhart A, Kimpton CP, Gill P (1993) Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2(ACTBP2) locus. Hum Genet 92:637–638

Wiegand P, Budowle B, Rand S, Brinkmann B (1993) Forensic validation of the STR systems SE33 and TC11. Int J Legal Med 105:315–320