Int J Legal Med (1999) 113 : 15–18

ORIGINAL ARTICLE

Takashi Yoshimoto · Keiji Tamaki · Syun Katsumata
Xiu-Lin Huang · Rieko Uchihi · Miwa Tanaka
Hiroki Uchida · Toshimichi Yamamoto · Song Chen
John A. L. Armour · Yoshinao Katsumata

Sequence analysis of alleles at a microsatellite locus D14S299 (wg1c5) and population genetic comparisons

Received: 28 September 1998 / Received in revised form: 4 January 1999 / Accepted: 8 February 1999

Abstract In order to increase the discriminating power of DNA analysis in personal identification, we evaluated the forensic utility of the microsatellite locus D14S299 (wg1c5) in the Japanese population and also in the Chinese and Caucasian populations. Twelve different alleles were identified in length by gel electrophoresis with silver staining. The major alleles in Japanese were sequenced and designated as the numbers of the variable repeats (GGAT or GGAA). There were five variable regions and extensive homoplasy was found. However, the allele fragment lengths were in 4 bp increments and no "interalleles" were found. The estimated heterozygosity and the polymorphism information content (PIC) were 0.726 and 0.689, respectively in Japanese. Those in Chinese (0.743 and 0.704) were similar to those in Japanese, while those in Caucasians (0.812 and 0.781) were much higher. After adjacent alleles were combined to yield at least five entries, statistical analysis was performed. The power of discrimination (PD) was 0.887 in Japanese, 0.895 in Chinese and 0.935 in Caucasians and no significant deviations from the Hardy-Weinberg equilibrium were found in the three populations. We retyped all apparently homozygous samples using an alternative pair of flanking primers and found them to be true homozygotes. D14S299 appears to be a useful STR locus for forensic practice.

Key words Short tandem repeat · D14S299 · Japanese · Population database · Homoplasy

T. Yoshimoto · K. Tamaki · S. Katsumata · R. Uchihi
M. Tanaka · H. Uchida · T. Yamamoto · Y. Katsumata (✉)
Department of Legal Medicine,
Nagoya University School of Medicine,
65 Tsurumacho, Showa-ku, Nagoya 466–8550, Japan
Fax +81-52-744-2121

X. L. Huang
Department of Legal Medicine,
Tokai University School of Medicine,
Bohseidai Isehara 259–1193, Japan

S. Chen
DNA Analysis Section, Institute of Forensic Science,
Ministry of Public Security, Beijing, 100038 P.R. China

J. A. L. Armour
Department of Genetics, School of Clinical Laboratory Sciences,
University of Nottingham, Queen's Medical Centre,
Nottingham NG7 2UH, UK

Introduction

Microsatellite loci are tandemly repeated stretches of a short DNA motif of 2–5 bases which are variable in length, relatively evenly spaced in eukaryotic genomes and can serve as highly informative polymorphic markers (Hagelberg et al. 1991; Menotti-Raymond et al. 1997). Recently, Armour et al. (1994) discovered 24 polymorphic microsatellite loci in the human genome by a rapid isolation method they had devised.

In a previous paper, we evaluated the forensic utility of one of these loci, D7S809, using 128 Japanese individuals and found it very useful to increase the discriminating power in personal identification (Tamaki et al. 1996). In the present paper, we describe the precise features of another microsatellite D14S299 (wg1c5) which was reported to be highly polymorphic (observed heterozygosity = 0.86) in Caucasians. We found this locus to be highly polymorphic also in Japanese and Chinese although the observed heterozygosities of the two populations were somewhat lower than that of Caucasians. The sequence analysis also revealed that this compound microsatellite locus showed extensive homoplasy where fragments of the same size exhibited different sequence structure.

Materials and methods

DNA was extracted from fresh blood treated by EDTA obtained from healthy unrelated individuals of 208 Japanese (Nagoya area), 113 Chinese (Beijing area) and 112 Caucasians (British) as previously described (Tamaki et al. 1991). The DNA concentration of each sample was quantified fluorometrically using a TKO 100 fluorometer (Hoefer Scientific Instruments) as previously described (Labarca and Paigen 1980). DNA samples were stored at –80 °C before use.

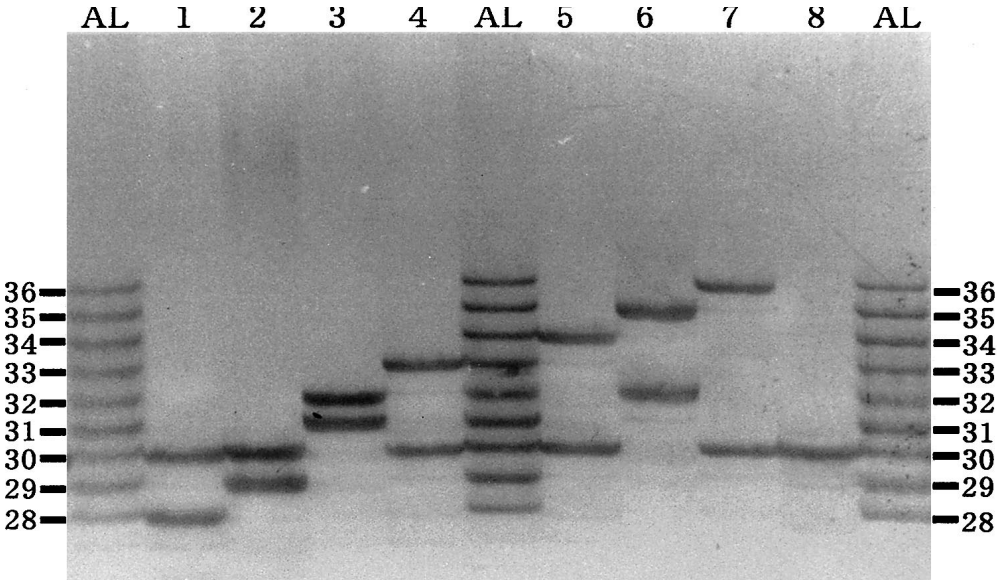

Fig. 1 Examples of genotyping at D14S299(wg1c5) in eight individuals (1–8). Genotyping was carried out using a denaturing polyacrylamide gel (6%) followed by silver staining. AL; allele ladders including allele 28–36

Segments of D14S299 were amplified from genomic DNA (10 ng) using a primer set originally described by Armour et al. (1994) as follows: primer 1 (designated wg1c5a; 5′-GATCTCAATAAACATTGATACTGG-3′) and primer 2 (wg1c5b; 5′-CTGCATGAGCTAAAGCATACTG-3′) were then used with 0.05 U/μl Ampli Taq Gold polymerase (Perkin Elmer) in a 10 μl PCR reaction buffer. Cycling conditions were 95 °C for 11 min for 1 cycle, 94 °C for 1 min, 65 °C for 1 min, and 72 °C for 1 min for 30 cycles, 72 °C for 10 min for 1 cycle in a GeneAmp PCR Systems 9700 (PE Applied Biosystems).

We also designed another pair of primers for retyping the apparent homozygotes using the sequence flanking wg1c5 from the EMBL sequence database in order to confirm whether they were real homozygotes. The primer sequences were: wg1c5a_2:5′-CATTGATACTGGAGGGATGAAAT-3′;wg1c5b_2: 5′-CTCTGCCCATCATCTGTTTGC-3′. Using this alternative primer pair, PCR was performed in the same buffer and the same cycling conditions except that the annealing temperature was changed from 65 °C to 58 °C. The sequence predicts that the amplified product using these alternative primers is 9 bases longer than that amplified using the original primer pair. PCR products were separated by electrophoresis on 6% denaturing polyacrylamide gels (8 M urea), followed by silver staining (Goldman and Merril 1982).

Amplified products were cloned using the Original TA Cloning Kit (Invitrogen). The cloned plasmids were isolated using QIAprep Spin Miniprep Kit (QIAGEN) and sequenced by the dyeprimer methods using 373 DNA sequencer (PE Applied Biosystems) and the Dye primer cycle sequencing FS kit (PE Applied Biosystems).

Results and discussion

Examples of genotyping at D14S299 of eight Japanese individuals are shown in Fig. 1. Twelve different D14S299 alleles were identified in length, at least two each of the seven major alleles in Japanese (> 1.0%) were sequenced, and additionally two rare alleles were analyzed (Griffiths et al. 1998). The alleles differed regularly by multiples of 4 bp, and the allelic ladder was made by mixing the cloned plasmids. An example of the sequence of the most common allele (allele 30) is shown in Fig. 2 which is 307 bp in size and possessed a total of 30 repeats in 5 variable regions named x, m, y, n, z from the 5′ end. We found that an irregular repeat, atatt, in the EMBL database was in fact a regular one, atat, as shown in the fifth row, so the PCR product is one base shorter than that in the database. From these sequence data, the exact sizes of D14S299 alleles, ranging from 299 to 343 bp, were determined.

Using the allelic ladder, 208 Japanese,113 Chinese and 112 Caucasian individuals were typed and the allele frequencies were calculated (Table 1).

The allele frequency distribution was almost unimodal with common alleles in the Japanese population sample, leading to a heterozygosity rate of 0.726 and a discriminating power (Fisher 1951) of 0.887. A similar distribution was observed in the Chinese population sample, while the distribution in the Caucasian population sample was bimodal (Table 1) showing an extremely high heterozygosity rate as reported previously (Armour et al. 1994). Pairwise comparisons clearly showed the similarity of the

				gatctcaataaacattgatactgg		ag
ggat	ggaa	tgaa	ggaa	$(ggat)_2$	agaa	ggct
ataa	ggat	gggt	$(ggat)_2$	aaat	$(GGAT)_{10}$ **x**	agat
$(ggat)_2$	ggaa	aaat	ggat	agat	ggat	gggt
$(GGAT)_2$ **m**	gaat	$(GGAT)_5$ **y**	gaat	atat	gggt	$(ggat)_2$
$(GGGT)_1$ **n**	$(GGAA)_{12}$ **z**	ggat	ggta	gaag	aaag	gtag
tac	cagtatgctttagctcatgcag					

Fig. 2 An example of the sequence of allele 30. Five variable regions named x, m, y, n z from 5′ end are shown in uppercase letters. The corresponding regions for primers wg1c5a(5′ end) and wg1c5b(3′ end) are underlined

Table 1 Allele frequency distribution for D14S299 in three different populations

Allele	Japanese (n = 208)	Chinese (n = 113)	Caucasian (n = 112)
28	0.005	0.004	0
29	0.087	0.071	0.049
30	0.440	0.407	0.246
31	0.231	0.265	0.254
32	0.132	0.119	0.201
33	0.034	0.044	0.063
34	0.053	0.049	0.134
35	0.005	0.009	0.045
36	0.010	0.027	0.009
37	0	0	0
38	0.002	0	0
39	0.002	0	0
40	0	0.004	0
H*	0.726	0.743	0.812

* expected heterozygosity

Table 2 Population comparisons using the χ^2-test

Population comparison	χ^2	P-value
Japanese-Chinese	4.56	0.6013
Japanese-Caucasian	42.96	1.19E-07
Chinese-Caucasian	25.07	3.32E-04

Japanese population to the Chinese one and also the dissimilarity of the Caucasian population to the two Asian populations (Table 2). It is interesting to find that allele frequency distributions and heterozygosity rates differ between the major ethnic groups and may reflect the genetic distance between the major ethnic groups.

No "interalleles" were found in the present study. The deviations from the predicted frequencies assuming Hardy-Weinberg equilibrium were tested by the homozygosity test (Weir 1992), the likelihood ratio test (Chakraborty et al. 1991) and the exact test (Guo and Thompson 1992). No significant deviations were found in all three populations.

The sequence analysis revealed the extensive homoplasy of this compound microsatellite locus in Japanese (Table 3). Overall, homoplasy for size was found in at least more than half (5/7) of the major allelic classes. This sequence variation within the same allele based on size may increase the power of a microsatellite in forensic applications. From these sequence data we could estimate the mutual relationship taking the similarity of allele structure into account.

PCR-based DNA typing sometimes fails to amplify an allele having sequence variation in the primer annealing site (Kurosaki et al. 1996; Gusmao et al. 1996). If such sequence variation occurs frequently, the reliability of genotyping is decreased due to the spurious appearance of homozygotes. We designed an alternative pair of flanking primers ($wg1c5a_2$ and $wg1c5b_2$) and retyped all of the apparently homozygous samples. Each sample showed only one band with the alternative primers, suggesting that there were no spurious homozygotes (data not shown).

The present study reveals that D14S299 is highly polymorphic, with no "interalleles" detected and has many genotypes apparently in Hardy-Weinberg equilibrium. This somewhat long microsatellite locus, may be especially useful in combination with shorter loci for multiplex analysis. Furthermore, its extensive homoplasy might be useful for the analysis of genetic diversity in humans or primates.

Table 3 Structures of the five variable regions in thirty D14S299 alleles in Japanese

Size (bp)	Allele structure					Allele designation	No
	x	m	y	n	z		
299	$(GGAT)_{10}$	$(GGAT)_2$	$(GGAT)_5$	$(GGGT)_1$	$(GGAA)_{10}$	28	2
303	$(GGAT)_{10}$*	$(GGAT)_2$	$(GGAT)_5$	$(GGGT)_1$	$(GGAA)_{11}$	29	3
307	$(GGAT)_{10}$	$(GGAT)_2$	$(GGAT)_5$	$(GGGT)_1$	$(GGAA)_{12}$	30	2
	$(GGAT)_{11}$	$(GGAT)_2$	$(GGAT)_8$		$(GGAA)_9$	30	4
311	$(GGAT)_9$	$(GGAT)_2$	$(GGAT)_9$**		$(GGAA)_{11}$	31	1
	$(GGAT)_{10}$	$(GGAT)_2$	$(GGAT)_5$	$(GGGT)_1$	$(GGAA)_{13}$	31	1
	$(GGAT)_{12}$		$(GGAT)_9$		$(GGAA)_{10}$	31	2
	$(GGAT)_{12}$	$(GGAT)_2$	$(GGAT)_8$		$(GGAA)_9$	31	2
315	$(GGAT)_{10}$	$(GGAT)_2$	$(GGAT)_5$	$(GGGT)_1$	$(GGAA)_{14}$	32	2
319	$(GGAT)_{11}$	$(GGAT)_2$	$(GGAT)_8$		$(GGAA)_{12}$	33	2
323	$(GGAT)_{11}$	$(GGAT)_2$	$(GGAT)_{10}$		$(GGAA)_{11}$	34	2
	$(GGAT)_{11}$	$(GGAT)_2$	$(GGAT)_{11}$		$(GGAA)_{10}$	34	1
	$(GGAT)_{12}$		$(GGAT)_9$		$(GGAA)_{13}$	34	1
327	$(GGAT)_{11}$	$(GGAT)_2$	$(GGAT)_{10}$		$(GGAA)_{12}$	35	2
331	$(GGAT)_{11}$	$(GGAT)_2$	$(GGAT)_{12}$		$(GGAA)_{11}$	36	2
	$(GGAT)_{12}$	$(GGAT)_2$	$(GGAT)_{11}$		$(GGAA)_{11}$	36	1

* Eighth repeat was changed to GGGT in two alleles
** Eighth repeat was changed to GGGT

Acknowledgement This work was supported in part by a Grant-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan.

References

Armour JAL, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. Hum Mol Genet 3:599–605

Chakraborty R, Fornage M, Guegue R, Boerwinkle E (1991) Population genetics of hypervariable loci: analysis of PCR-based VNTR polymorphism within population. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) DNA fingerprinting: approaches and applications. Birkhauser, Berlin, pp 127–143

Fisher RA (1951) Standard calculations for evaluating a blood group system. Heredity 5:95–102

Goldman D, Merril CR (1982) Silver staining of DNA in polyacrylamide gels, linearity and effect of fragment size. Electroporesis 3:24–26

Griffiths RAL, Barber MD, Johnson PE, Gillbard SM, Haywood MD, Smith CD, Arnold J, Burke T, Urquhart AJ, Gill P (1998) New reference allelic ladders to improve allelic designation in a multiplex STR system. Int J Legal Med 111:267–272

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

Gusmao L, Amorim A, Prata MJ, Peaeira L, Lareu MV, Carracedo A (1996) Failed PCR amplifications of MBP-STR alleles due to polymorphism in the primer annealing region. Int J Legal Med 108:313–315

Hagelberg E, Gray IC, Jeffreys AC (1991) Identification of the skeletal remains ofmurder victim by DNA analysis. Nature 352:427–429

Kurosaki K, Oota H, Saitoh H, Kiuchi M, Ueda S (1996) Sequence variation found in the flanking region of a trimeric short tandem repeat at the PLA2 locus: its considerable effect on estimating alleles. Nippon Hoigaku Zasshi 50:1–5

Labarca C, Paigen K (1980) A simple rapid and sensitive DNA assay procedure. Anal Biochem 102:344–352

Menotti-Raymond MA, David VA, O'Brien SJ (1997) Pet cat hair implicates murder suspect. Nature 386:774

Tamaki K, Yamamoto T, Uchihi R, Katsumata Y, Kondo K, Mizuno S, Kimura A, Sasazuki T (1991) Frequency of HLA-DQA1 alleles in the Japanese. Hum Hered 41:209–214

Tamaki K, Huang X-L, Nozawa H, Yamamoto T, Uchihi R, Katumata Y, Armour JAL (1996) Evaluation of tetranucleotide repeat locus D7S809 (wg1g9) in the Japanese population. Forensic Sci Int 81:133–140

Weir BS (1992) Independence of VNTR alleles defined by fixed bins. Genetics 130:873–887