Int J Legal Med
DOI 10.1007/s00414-009-0369-0

SHORT COMMUNICATION

A hypervariable STR polymorphism in the complement factor I (CFI) gene: Asian-specific alleles

Isao Yuasa · Yoshito Irizawa · Hiroaki Nishimukai · Yasuo Fukumori · Kazuo Umetsu · Nori Nakayashiki · Naruya Saitou · Lotte Henke · Jürgen Henke

Received: 29 January 2009 / Accepted: 8 July 2009

Abstract In this study, a short tandem repeat (STR) polymorphism in intron 7 of the human complement factor I (CFI) gene was studied in 637 DNA samples obtained from African, German, Thai, and Japanese populations and German and Japanese families. A total of 41 alleles were observed and classified into two groups, L and H, based on size differences. Group H, which consisted of 16 alleles, was observed only in Thai and Japanese populations at frequencies of 0.162 and 0.116, respectively, and was strongly associated with c.1217A in exon 11 (CFI*Ah). The heterozygosity values ranged from 0.89 in German to 0.93 in Thai populations. This STR would be a useful supplementary marker for forensic individualization.

Keywords Asian-specific · Complement factor I · Microsatellite · Population study · Short tandem repeat

Introduction

Short tandem repeat (STR) markers are capable of possessing a high degree of polymorphism and are the most effective markers for forensic individualization. Population-specific STR alleles or regionally private STR alleles with a frequency above 0.02 have sometimes been detected on autosomal chromosomes. However, no allele with a frequency above 0.13 has been found except for allele 9RA in D9S1120, which is observed at an average frequency of 0.36 in Native Americans [1, 2].

The human complement factor I (CFI) gene (GenBank GeneID: 3426) spans 63 kb on chromosome 4q25 and consists of 13 exons encoding a 583-amino acid polypeptide as an unprocessed precursor. Recently, we described Japanese,

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0369-0) contains supplementary material, which is available to authorized users.

I. Yuasa (✉) · Y. Irizawa
Division of Legal Medicine, Faculty of Medicine,
Tottori University,
Yonago 683-8503, Japan
e-mail: yuasai@med.tottori-u.ac.jp

H. Nishimukai
Division of Legal Medicine,
Ehime University School of Medicine,
Ehime, Japan

Y. Fukumori
Osaka Red Cross Blood Center,
Osaka, Japan

K. Umetsu
Department of Experimental and Forensic Pathology,
Faculty of Medicine, Yamagata University,
Yamagata, Japan

N. Nakayashiki
Department of Legal Medicine,
Iwate Medical University School of Medicine,
Morioka, Japan

N. Saitou
Division of Population Genetics, National Institute of Genetics,
Mishima, Japan

L. Henke · J. Henke
Institut für Blutgruppenforschung,
Cologne, Germany

Published online: 20 August 2009

African, and East Asian-specific nonsynonymous single-nucleotide polymorphisms (SNPs) in exons 4 (c.A603C), 7 (c.G898A), and 11 (c.G1217A) [3]. An STR with 27 repeats of the tetranucleotide polypyrimidine TTTC unit exists in intron 7 of this gene. The nucleotide sequence shows an insertion/deletion polymorphism due to the presence or absence of $(TTTC)_4$, which is registered as refSNP ID rs56875356 (GenBank Entrez SNP) without frequency data. In this study, we scrutinized this CFI STR polymorphism with respect to its repeat structure, allele distribution, and genetic association with the SNPs and present evidence for the hypervariability of the CFI STR with Asian-specific alleles.

Materials and methods

DNA samples

DNA samples were obtained from 68 sub-Saharan African (Nigerian and Ghanaian), 92 German, 114 Thai, and 198 Japanese individuals. Most of these samples were from the same set as those used in a previous study [3]. For family studies, 21 German trios and 33 Japanese families with 36 offsprings were investigated. This study was approved by the Ethical Committee at the Faculty of Medicine, Tottori University. Nonhuman primate DNA samples were also obtained from three chimpanzees (*Pan troglodytes*), a gorilla (*Gorilla gorilla*), and a Japanese monkey (*Macaca fuscata*).

SNP typing

The five SNPs in the CFI gene were typed by a multiplex polymerase chain reaction (PCR) method as described previously [3].

PCR

PCR products containing the STR were amplified with the following primers: forward primer CFI-i7F: 5′-AACGT CAAAGTTCATGATCCTC-3′ and reverse primer CFI-i7R3: 5′-AGCCTGGGCTACAAGACC-3′ (Fig. 1). The forward primer was labeled with VIC at the 5' terminal. PCR was carried out in a volume of 10 μL containing 1 ng genomic DNA, 0.4 μM of each primer, 200 μM of each dNTP, and 0.25 U HotStarTaq Plus polymerase (Qiagen, Hilden, Germany). The cycle conditions were as follows: 95°C for 5 min, followed by 30 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 40 s, and a final extension of 90 min at 72°C.

Typing of STR

Of the PCR products or allelic ladder, 1 μL was added to 9 μL of HiDi-formamide containing GeneScan 500 LIZ size standard (Applied Biosystems, Foster City, CA, USA), and aliquots were separated by electrophoresis on a 3130xl Genetic Analyzer (Applied Biosystems) with a 36 cm array and POP-7 polymer. The allelic ladder was made in-house by mixing PCR products with different genotypes. The sample run data were analyzed together with the allelic ladder and positive and negative controls using GeneMapper Software Version 3.5 (Applied Biosystems). Alleles were designated according to the recommendations from the DNA Commission of the International Society for Forensic Genetics. [4].

Sequencing

The PCR products for sequencing were prepared with sequencing primers under the abovementioned conditions except the reaction volume was 50 μL. The primers used were CFI-i7F5: 5′-GTAAAATGACTCATATCCC CAGTTC-3′ and CFI-i7R2: 5′-GGCAGAGAATTGCT TAAACCC-3′ (Fig. 1). In heterozygotes, each band was excised from a denaturing gel and subjected to reamplification with the same primer set. The products were purified using Qiaquick PCR Purification Spin Columns (Qiagen). Sequence reactions were obtained using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).

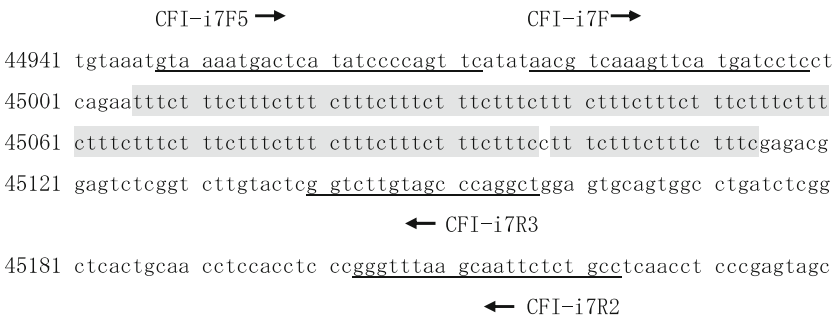

Fig. 1 The GenBank sequence for intron 7 of the CFI gene. The primers are *underlined* and the STR, which consists of 27 repeats of TTTC motif, is *shaded in gray*. The nucleotide numbering was defined according to the genomic sequence (NT_000004.10). The adenine of the translation initiation codon ATG corresponds to nucleotide 15

Statistical analysis

The parameters of forensic interest were calculated using the Powerstats program (http://www.promega.com). Estimation of the allele and haplotype frequencies, testing of the Hardy–Weinberg equilibrium, analysis of molecular variance, and calculation of R_{ST} values were performed using the Arlequin program ver. 3.11 [5]. The linkage disequilibrium coefficients D' and r^2 were calculated between pairs of the STR and SNPs [6].

Results

STR typing

Figure 2 shows electropherograms of the allelic ladder used for the STR typing and a sample with genotype 21–47.3. The allelic sizes fell in a range from 130 to 300 bp. The genotypes were clearly and unambiguously identified. A total of 41 alleles were observed in 637 DNA samples obtained from four populations and 54 families. These alleles were classified into two groups, L and H, based on size differences. Alleles 16–34, which belonged to group L, were characterized by low molecular weights, whereas alleles 40.3–55.3, which belonged to group H, were characterized by high molecular weights. A total of seven microvariant alleles were found in group L. The Control DNA 007 (Applied Biosystems) was identified as genotype 21–25.

Sequencing

To prepare the allelic ladder and to investigate the structure of the repeat motifs, a total of 113 alleles belonging to both groups were sequenced together with primate DNA (Electronic supplementary material Tables S1 and S2). Common formulae for the repeat structures of groups L and H can be given as follows:

$$\text{group L}: (\text{TTTC})_n\text{C}(\text{TTTC})_4 \quad (n = 12 - 30)$$

$$\text{group H}: (\text{TTTC})_p(\text{TTCC})_q\text{TTC}(\text{TTTC})_r\text{C}(\text{TTTC})_3 \quad (p + q + r = 37 - 52)$$

The marked differences between the two groups were due to the insertion or deletion of TTCC units in an interme-

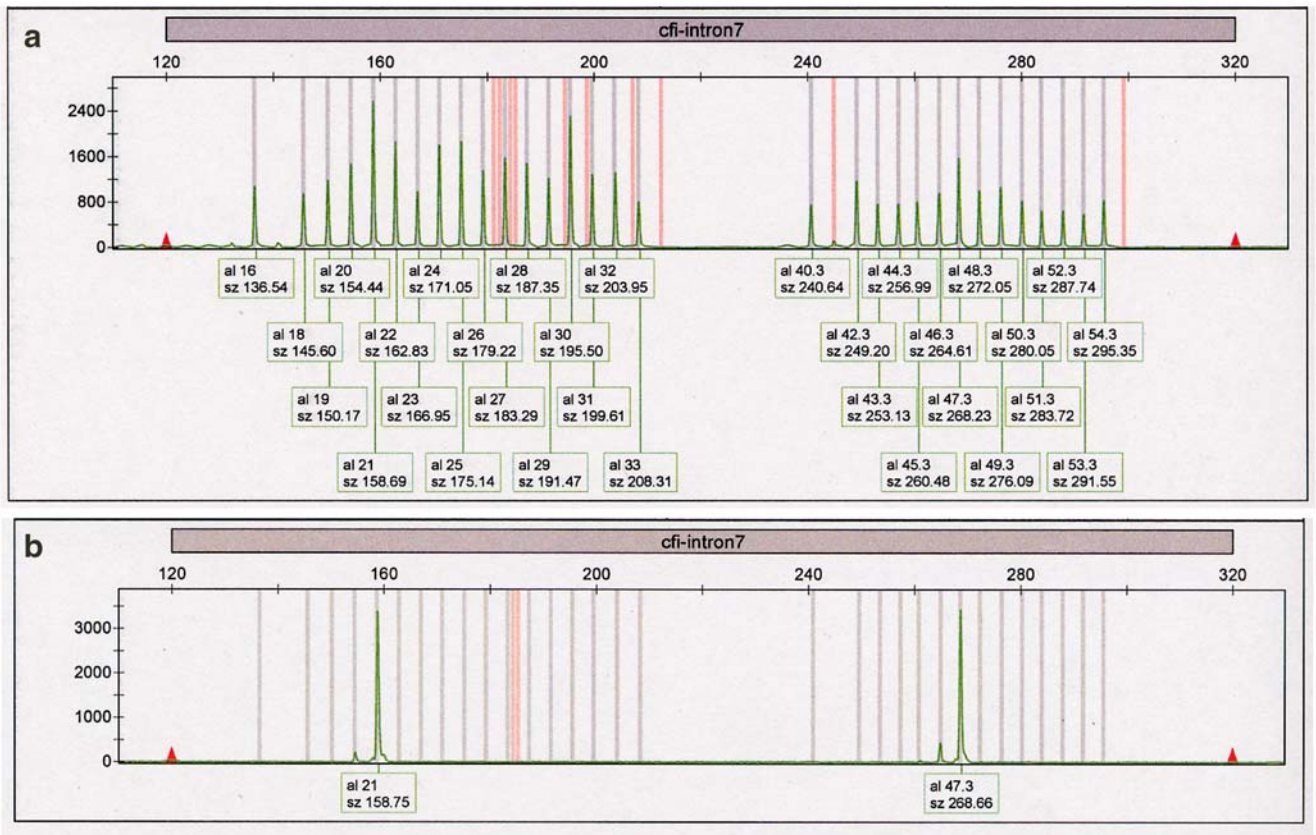

Fig. 2 Automated electropherogram of the CFI STR. (**a**) allelic ladder consisting of 31 bins for allelic products and 10 bins for mutants, and (**b**) genotype 21–47.3 observed in a Japanese

diate region and of a TTTC unit in the 3′ region. The microvariant alleles that contained a 0.3 repeat in group L had the repeat structure of $(TTTC)_mTTC(TTTC)C(TTTC)_4$ (m=21-27). In the microvariant alleles that contained a 0.2 or 0.1 repeat, the tetrameric TTTC repeat block was interrupted by a TT or T insertion, respectively. Some alleles had the same sizes, but differed in their nucleotide and repeat structures. A total of 64 different sequences were observed in the 113 allelic products investigated. No hybrid structure composed of the groups L and H sequences was found, suggesting that the increases and decreases in repeat number have occurred by replication slippage but not by recombination. The PCR products of the nonhuman primates were smaller than those of humans. Amplification of DNA from a Japanese monkey failed with the primers CFI-i7F and CFI-i7R3 due to a deletion of several nucleotides corresponding to the 3' region of the primer CFI-i7R3. PCR products were barely observed using the primers CFI-i7F5 and CFI-i7R2. The sequences in the primates were $(TTTC)_3(T/CTTC)T_{13\text{-}14}$.

Family studies

Family studies including 54 parents and 57 offsprings were performed and alleles were found to be inherited in a regular Mendelian fashion. The results showed that c.1217A (CFI*Ah) was inherited from parent to child together with a group H allele. Similarly, c.603C (CFI*As) was associated with alleles 30, 32, and 34.

Population study

Table S3 of the Electronic supplementary material shows the allele frequencies and parameters of forensic interest for the STR as well as the frequencies of the SNPs c.603C and c.1217A. The frequencies in the German and Japanese populations were estimated using data from population samples and parents. The CFI STR was highly polymorphic in all populations investigated in this study. The Asian populations had more alleles than the African and European populations. The expected heterozygosity ranged from 0.89 in the German to 0.93 in the Thai populations. Interestingly, group H was confined to Asian populations. Although each of the group H alleles was low in frequency, a relatively large number of alleles were observed. The sum of the frequencies of the group H alleles was 0.162 in the Thai and 0.116 in the Japanese populations, and these values were slightly higher than those for c.1217A, which is characteristic of Asian populations [3]. Microvariant alleles containing a 0.3 repeat were observed only in Africans and microvariant alleles containing a 0.2 repeat were only observed in Japanese. Significant R_{ST} values were obtained between the Asians and the other two populations.

Linkage analysis

Our family study has suggested that associations exist between the group H alleles and c.1217A and between the group L alleles which contain a high number of repeats and c.603C. The genotyping results indicated that all of the DNA samples which contained c.1217A also possessed at least one of the group H alleles except for a Thai sample. An association study was carried out between the group H alleles and c.1217A. The group H alleles were combined because the frequency of each allele was low. Table 1 shows haplotype frequencies and linkage disequilibrium coefficients. Significant associations were found in the Thai and Japanese populations. Similarly, c.603C was suggested to be linked to the group L alleles with a high number of repeats (alleles 29–34) and c.898A was suggested to be linked to the group L alleles with an intermediate number of repeats (alleles 23–29).

Discussion

Recent studies have suggested that all modern humans originated from a single location in Africa where populations show the highest genetic diversity. A smooth loss of genetic diversity has occurred through bottleneck events with increasing distance from Africa. Genetic drift is greater in Asians than in Europeans [7, 8]. The CFI STR investigated in this study has a higher diversity in Asians than in Africans and Europeans which comes from the fact that the Asian populations have population-specific alleles that are classified into group H. A comparison of nucleotide sequences between humans and nonhuman primates suggests that an increase in the tetrameric TTTC motif and disappearance of polyT occurred along evolution of the human lineage. The divergence of Asians from Africans and Europeans predated the acquisition of an ancestral allele of the group H alleles in Asians. Subsequently, a c.G1217A

Table 1 Frequencies of haplotypes constructed with the c.G1217A and groups H/L alleles and linkage disequilibrium values

Haplotype	Thai	Japanese
A-H	0.136	0.104
A-L	0.005	–
G-H	0.026	0.011
G-L	0.833	0.884
D'	0.962	1.000
R^2	0.779	0.890
P	<0.001	<0.001

Group L alleles consist of alleles 16–34, and group H alleles, alleles 40.3–55.3

mutation occurred in one of the group H alleles, and then expanded into Asians. The c.A603C mutation must have arised on one of the group L alleles with a high number of repeats.

The group H alleles are confined to Asian populations. Although each of them is low in frequency, the sum of their frequencies was fairly high in the two Asian populations investigated. They provide us with information on the ancestry and population to which an individual belongs. This study has also shown that the group H alleles are associated with c.1217A (CFI*Ah). The frequency of c.1217A is highest in the Han Chinese from Changsha, Southern China and shows a south–north downward geographical cline [3]. For further characterization of the group H alleles, therefore, investigation of their distribution in Chinese and neighboring populations is important. The association of c.1217A with the group H alleles suggests that an STR locus near a population-specific SNP may have population-specific alleles.

The fibrinogen alpha chain (FGA) locus is located on chromosome 4q28, about 45 Mbp distant from the CFI gene. The FGA locus is hypervariable with about 80 alleles, which are also classified into low (alleles 12.1–35.2) and high (alleles 42.2–51.2) molecular weight groups. However, the basic repeat structure is different between alleles≤30 and≥30.2. Some alleles are African-specific, but their frequencies are low [9, 10]. Unlike the FGA locus, groups L and H in the CFI STR are clearly distinguishable by their size and repeat structure.

The gene diversity of the CFI STR ranged from 0.89 in Germans to 0.93 in Thais. This STR shows high diversity not only in Asians but also in Africans and Europeans. These values are comparable to those of hypervariable loci including FGA, D2S1338, and ACTBP2 [11, 12]. The size of allelic products ranged from 130 to 300 bp. It is easy to identify microvariants with differences of 1 bp. In conclusion, the CFI STR is unique in its hypervariability and possession of population-specific alleles. The CFI STR would be a useful marker in forensic and anthropological sciences like other autosomal STR markers which are not included in commercially available kits [13, 14].

Acknowledgements This study was supported in part by a Grant-in-Aid for Scientific Research (20590678 to IY) from the Japan Society for the Promotion of Science.

References

1. Schroeder KB, Schurr TG, Long JC et al (2007) A private allele ubiquitous in the Americas. Biol Lett 3:218–223
2. Phillips C, Rodriguez A, Mosquera-Miguel A et al (2009) D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies. Forens Sci Int Genet 3:7–13
3. Yuasa I, Nakagawa M, Umetsu K et al (2008) Molecular basis of complement factor I (CFI) polymorphism: one of two polymorphic suballeles responsible for CFI A is Japanese-specific. J Hum Genet 53:1016–1021
4. Bär W, Brinkmann B, Budowle B et al (1997) DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. Int J Legal Med 110:175–176
5. Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50
6. Hartl DL, Clark AG (1997) Principles of population genetics, 3rd edn. Sinauer, Sunderland, MA
7. Handley LJL, Manica A, Goudet J, Balloux F (2007) Going the distance: human population genetics in a clinal world. Trends Genet 23:432–439
8. Keinan A, Mullikin JC, Patterson N, Reich D (2007) Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nat Genet 39:1251–1255
9. Griffiths RA, Barber MD, Johnson PE et al (1998) New reference allelic ladders to improve allelic designation in a multiplex STR system. Int J Legal Med 111:267–272
10. Butler JM et al (2005) Forensic DNA typing. Biology, technology, and genetics of STR markers. Elsevier, Amsterdam
11. Montelius K, Karlsson AO, Holmlund G (2008) STR data for the AmpFlSTR Identifiler loci from Swedish population in comparison to European, as well as with non-European population. Forensic Sci Int Genet 2:e49–52
12. Hohoff C, Schürenkamp M, Brinkmann B (2009) Meiosis study in a population sample from Nigeria: allele frequencies and mutation rates of 16 STR loci. Int J Legal Med 123:259–261
13. Grubwieser P, Zimmermann B, Niederstätter H, Pavlic M, Steinlechner M, Parson W (2007) Evaluation of an extended set of 15 candidate STR loci for paternity and kinship analysis in an Austrian population sample. Int J Legal Med 121:85–89
14. Hering S, Augustin C, Edelmann J, Heidel M, Chamaon K, Dressler J, Szibor R (2008) Complex variability of intron 40 of the von Willebrand factor (vWF) gene. Int J Legal Med 122:67–71