**TECHNICAL NOTE**

*Adriano Tagliabracci,[1] M.D., Ph.D.; Loredana Buscemi,[1] M.D., Ph.D.; Francesca Bianchi,[1] Ph.D.; Corrado Sassaroli,[1] Ph.D.; Ugo Ricci,[2] Ph.D.; Tauro M. Neri,[3] M.D., Ph.D.; and Daniele Rodriguez,[1] M.D., Ph.D.*

# Polymorphism and Sequence Variations of the HumCD4 Pentameric Microsatellite in an Italian Population Sample*

**REFERENCE:** Tagliabracci A, Buscemi L, Bianchi F, Sassaroli C, Ricci U, Neri TM, Rodriguez D. Polymorphism and sequence variations of the HumCD4 pentameric microsatellite in an Italian population sample. J Forensic Sci 1998;43(4):841–844.

**ABSTRACT:** A sample of 265 subjects from central Italy was analyzed at the HumCD4 locus by polymerase-chain-reaction (PCR). Phenotypes were identified by comparison with a sequenced ladder, after high-resolution horizontal polyacrylamide gel electrophoresis (PAGE) followed by silver staining. A set of representative alleles was sequenced by Taq-cycle-sequencing with dye terminator labeling and capillary gel electrophoresis strategies. Eight common alleles—5,6,7,8,9,10,11,12—and a rare larger 14, never before described in Caucasians, were found. Allele and genotype frequencies were similar to those described in former studies on Caucasians, with a prevalency of alleles number 5, 6, 10. Sequence analysis showed that the polymorphism is due to a pentameric TTTTC basic motif, tandemly repeated, and that from allele number 10 onwards the fourth repeat presents a T to C translation (CTTTC). Instead, allele number 9 may exist in two forms, because 75% of alleles examined in this study presented the CTTTC motif at the fourth position, while the remaining 25% had the basic repeat structure.

**KEYWORDS:** forensic science, DNA typing, short tandem repeats, HumCD4, population genetics, sequencing analysis, structural variations, Italy

Short tandem repeats (STRs) are an abundant class of microsatellites widely spread throughout the human genome, consisting of tandemly repeated sequences of 2 to 7 bp length monomers (1–3). Related to variations in the number of repeat units displayed, most of these microsatellites have a high degree of length polymorphism, well investigated by the PCR technique (2,4). Although STRs show a lower degree of polymorphism with respect to variable numbers of tandem repeats (VNTR) loci investigated by restriction fragment length polymorphism (RFLP) typing, their very high number in the human genome and their efficiency with minimal amounts of DNA template, even degraded, due to their small size ($<$ 300 to 350 bp) make them very suitable for human identification (5–8).

The HumCD4 system (GenBank M86525) is an STR composed of a pentameric unit repeat—TTTTC (9) located on the short arm of chromosome 12 (12q-12pter) which, in the various ethnic groups studied until now, displayed a polymorphism of 12 alleles in the range 81 to 136 bp (10). Because of its small molecular size, this STR is very useful for forensics, despite its low discriminating power due to the occurrence of the three most common alleles: 5,6,10.

A study on the polymorphism of this system has just been performed on Italians (11), but an excess of homozygotes indicative of population substructure was found, suggesting the need for further studies. The present study analyzed a sample of 265 individuals, to increase the number of observations for better knowledge of polymorphism and to set up a database suitable for paternity testing and personal identification. A set of different alleles was therefore sequenced with the further aim of verifying structural microvariations in the repeat unit previously described (10,12).

## Materials and Methods

DNA was extracted from blood samples of 265 unrelated healthy donors living in Ancona ($n$ = 119) and Florence ($n$ = 146) by the phenol-chloroform method (13) and quantitated by a slot blot procedure using a D17Z1 probe (Gibco-BRL) (14).

PCR amplification was performed in a MiniCycler (MJ Research, Watertown, MA) using the primers proposed by Edwards et al. (9) with 5 ng template in a 25 μL final volume. PCR conditions were: hot start at 94°C for 2 min, followed by 31 cycles at 94°C for 120 s, 62°C for 90 s, 72°C for 60 s, and a final extension step at 72°C for 10 min. Electrophoresis of 3.5 μL of amplified product was carried out on a high-resolution non-denaturing polyacrylamide gel (PAG) using a discontinuous buffer (15). Bands were visualized by silver staining (16) and alleles were identified by side-to-side comparison with a ladder consisting of a mix of sequenced amplified products (Fig. 1).

Statistical analysis—Hardy-Weinberg expectations were verified in the two original subpopulations from Ancona ($n$ = 119)

[1] Istituto di Medicina Legale, Università degli Studi di Ancona, Policlinico Torrette, Ancona, Italy.
[2] Servizio di Genetica Umana, Azienda Meyer, Università di Firenze, Firenze, Italy.
[3] Cattedra di Genetica Medica, Università di Parma, Ospedale Maggiore, Parma, Italy.

* Supported by a grant from the Ministero dell'Universita e della Ricercà Scientifica e Tecnologica.
Received 17 June 1997; and in revised form 22 Oct. 1997 and 3 Dec. 1997; accepted 5 Dec. 1997.

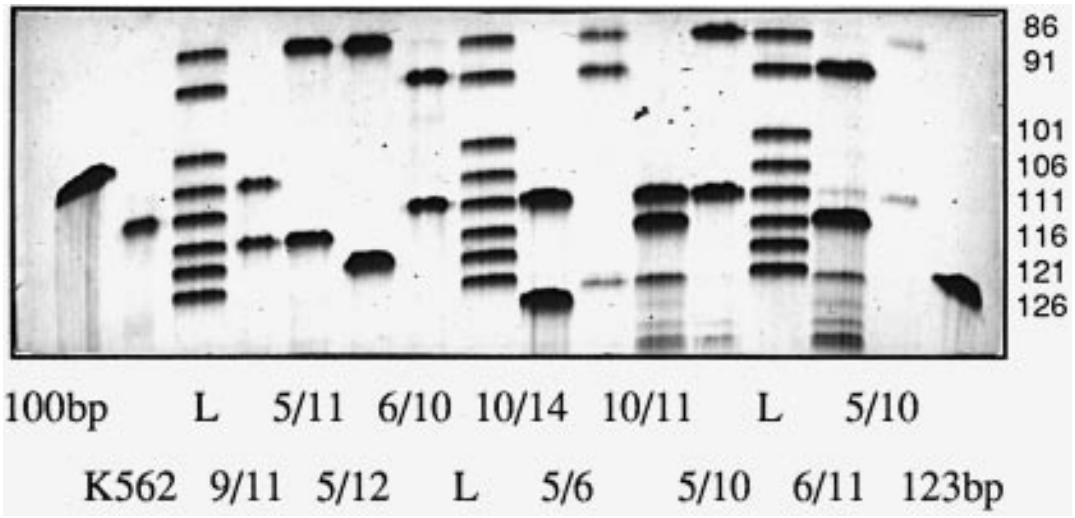


FIG. 1—*PAGE and silver staining of the amplified products. Polyacrylamide gel: 8% T, 3% C, 750 μm thick, piperazine diacrylamide as crosslinker, 28 mM CHES, 80 mM formate (pH 9.0). Electrophoresis in discontinuous buffer with 2% agarose plugs in 2× tris (0.5 M)/borate buffer; separation distance 18 cm; 1000 V, 40 mA, 5 W with ramping every 90 min up to 15 W and stop when the bromophenol blue had reached the anode (anode at the top).* From left to right: *100 bp, 123 bp = molecular weight markers; K562 = positive control; L = sequenced ladder (fragment size bp's in the right of the picture); 9/11, 5/11, 5/12, 6/10, 10/14, 5/6, 10/11, 5/10, 6/11, 5/10 = phenotypes.*

and Florence ($n$ = 146) and in the pooled population ($n$ = 265) in two ways: (1) with the chi-square test, pooling genotypes with an observed frequency below 5. In this case the degree of freedom (d.f.) was calculated as number of classes (observed genotypes) minus 1 and (2) with the Smith chi-square test (17) with small numbers, without pooling genotypes. In this case the d.f. was calculated as the number of observed classes of genotypes minus the number of alleles. Observed and expected heterozygosity frequencies were also compared (18).

The hypothesis of homogeneity in allele frequencies in the Caucasian population samples was tested by pairwise comparisons with the chi-squared and G-statistic tests using an R × C contingency table (19) and a computer program provided by G. Carmody (Carleton University, Ottawa, Canada). The power of discrimination and exclusion chance were also calculated (20,21).

Sequencing analysis was performed on a range of representative alleles. Taq-cycle-sequencing was performed on both strands using the Taq-Dye-Deoxy-Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) in the conditions indicated in the P/N 402078 protocol with some modifications: 15 to 20 ng DNA template and half the amount of dye-labeled terminators were used, taken to 20 μL with deionized water. Analysis was carried out by capillary electrophoresis on an ABI Prism 310 (Applied Biosystems). The data were analyzed with PE/ABD software Sequencing Analysis 3.0.

TABLE 1—*HumCD4 allele frequencies and tests for Hardy-Weinberg equilibrium (HWE) in a sample of 265 Italians (Ancona* n = *119; Florence* n = *146).*

| Allele designation | Ancona | Florence |
|---|---|---|
| 5 | 0.396 | 0.329 |
| 6 | 0.269 | 0.312 |
| 7 | 0.000 | 0.003 |
| 8 | 0.000 | 0.007 |
| 9 | 0.004 | 0.007 |
| 10 | 0.273 | 0.294 |
| 11 | 0.050 | 0.027 |
| 12 | 0.004 | 0.021 |
| 14 | 0.004 | 0.000 |

| HWE | Pooled genotypes | Smith test |
|---|---|---|
| Ancona | $\chi^2$=3.34, p=.8521, d.f.=7 | z=1.03, p=.152, d.f.=21 |
| Florence | $\chi^2$=1.03, p=.9846, d.f.=6 | z=2.85, p=.002, d.f.=36 |
| Pooled popul. | $\chi^2$=7.64, p=.5706, d.f.=9 | z=3.40, p=.001, d.f.=36 |

## Results and Discussion

A study on allele and genotype frequency distribution of the STR HumCD4 was performed on a sample of 265 Italians belonging to two different regions of central Italy (Ancona, $n$ = 119; Florence, $n$ = 146). Electrophoresis in horizontal native high-resolution PAG using discontinuous buffer and silver staining permitted good separation of different bands (alleles). Following the recommendations of the DNA Commission of the International Society of Forensic Haemogenetics (22), allele typing carried out by side-to-side comparison with a ladder composed of sequenced alleles was easy and reliable. The problem of uncommon occurrence of extra bands located in the range of alleles and causing problems with identification was overcome by decreasing the amount of template.

Nine alleles (range 86 to 131 bp) and seventeen genotypes were found in our study (Tables 1 and 2). Among them, a larger allele 14, until now only described in non-Caucasian populations (10), was discovered. Apart from genotypes containing the three common alleles n. 5, 6 and 10, all the remaining ones showed expected frequencies of less than 5. Extremely rare genotypes were also found: two heterozygotes 9,11 and one homozygote 12.

No differences were found in allele frequencies in the two subsamples when tested for homogeneity.

The Hardy-Weinberg equilibrium, verified by pooling genotypes with an observed frequency below 5, showed agreement

between observed and expected genotype both in the whole population and in the two subpopulations. When using the Smith test, the results changed dramatically. Only the smaller subpopulation of Ancona was in Hardy-Weinberg equilibrium, whereas the other populations deviated significantly from Hardy-Weinberg expectations because of the unexpected high frequency of genotypes 9,11 and 12,12 (Table 1). It is worth noting that the degree of polymorphism of HumCD4 locus was lower in the Ancona population because some rare alleles were never found. Instead, the presence of very low frequency genotypes strongly perturbs the equilibrium in the population from Florence and in the pooled population, mainly because of the presence of the two 9,11 heterozygotes, one from Ancona and the other from Florence, above expectations causing highly significant deviation in the statistical test ($\chi^2$ = 31.44).

In our opinion, the deviations from Hardy-Weinberg expectations were caused by sampling error. Because of the presence of low-frequency alleles and genotypes which may be either missed or perturb the equilibrium when found, the HumCD4 system needs a larger population sample to maintain its high degree of polymorphism. Furthermore, the possibility of generation of rare alleles by means of unequal crossovers cannot be ruled out.

The observed heterozygosity frequencies calculated in the whole population were in the expected range (0.705 and 0.704 ± 0.028, respectively). Comparison of allele frequencies with results from previous studies on Caucasians showed good agreement with the data reported for American Caucasians (4), Austrians (23) and Germans (24), while a slight heterogeneity was found with respect to the Italian sample investigated by Casarino et al. (11) (Table 3). It must be noted that these authors also found a much lower number of alleles and an excess of homozygous types, indicating a surplus due to Wahlund's effect or population admixture, as suggested by the same authors.

TABLE 2—*HumCD4 genotype distribution in a sample of 265 Italians (Ancona* n = *119; Florence* n = *146).*

| Genotypes | Observed | | Genotypes | Observed | |
|---|---|---|---|---|---|
| | Ancona | Florence | | Ancona | Florence |
| 5-5 | 19 | 13 | 6-10 | 18 | 30 |
| 5-6 | 21 | 31 | 6-11 | 3 | 3 |
| 5-7 | 0 | 1 | 8-10 | 0 | 1 |
| 5-8 | 0 | 1 | 9-11 | 1 | 1 |
| 5-10 | 27 | 30 | 10-10 | 9 | 12 |
| 5-11 | 7 | 3 | 10-11 | 1 | 1 |
| 5-12 | 1 | 4 | 10-14 | 1 | 0 |
| 6-6 | 11 | 13 | 12-12 | 0 | 1 |
| 6-9 | 0 | 1 | | | |

TABLE 3—*Pairwise comparisons for homogeneity test between Italians (*n = *265, this study); American Caucasians (*n = *191, (4)); Austrian Caucasoids (*n = *300, (23)); Germans (*n = *139, (24)); and Italians (*n = *134, (11)).*

| Populations | $\chi^2$ | P | G-statistic | P |
|---|---|---|---|---|
| Italian-Americans | 7.5079 | 0.5210±0.0158 | 9.2029 | 0.4810±0.0158 |
| Italian-Austrians | 7.0309 | 0.5260±0.0158 | 7.4956 | 0.6020±0.0155 |
| Italian-Germans | 15.3289 | 0.0540±0.0071 | 15.9269 | 0.0930±0.0092 |
| This study-Italians | 14.9423 | 0.0480±0.0068 | 15.4026 | 0.0570±0.0073 |

TABLE 4—*Sequence structure of HumCD4 representative alleles.*

| Allele designation | Fragment length (bp) | Sequence structure | Number of sequenced alleles |
|---|---|---|---|
| 5 | 86 | -(TTTTC)5- | 2 |
| 6 | 91 | -(TTTTC)6- | 2 |
| 7 | 96 | -(TTTTC)7- | 1 |
| 8 | 101 | -(TTTTC)8- | 1 |
| 9 | 106 | -(TTTTC)3 CTTTC (TTTTC)5- | 3* |
| 9' | 106 | -(TTTTC)9- | 1 |
| 10 | 111 | -(TTTTC)3 CTTTC (TTTTC)6- | 3 |
| 11 | 116 | -(TTTTC)3 CTTTC (TTTTC)7- | 2 |
| 12 | 121 | -(TTTTC)3 CTTTC (TTTTC)8- | 2 |
| 14 | 131 | -(TTTTC)3 CTTTC (TTTTC)10- | 1 |

*Two alleles 9 sequenced were out of this study

The discriminating power of the HumCD4 locus was 0.85 and the chance of exclusion 0.45.

Sequence analysis using Taq-cycle-sequencing and dye terminator labeling strategies was performed by means of capillary electrophoresis to confirm the size of the new allele 14, which showed a TTTTC unit motif tandemly repeated fourteen times, and a molecular size of 131 bp.

Structural analysis was also carried out on a set of representative alleles (Table 4). Sequencing studies recently performed on STRs have shown structural differences in repeat units, concerning both length and sequence and affecting several loci (10,12,25,26,27,28). HumCD4 has been described as a simple repeat, consisting of a regular series of TTTTC with a modified CTTTC in the fourth repeat unit from allele 9 onwards (10). Our analysis confirmed the pentameric TTTTC structure as a basic motif, responsible for the length polymorphism and, from allele 10 onwards, the fourth basic motif was CTTTC in all the alleles sequenced. Nevertheless, out of the four alleles 9 sequenced, only three showed the T-to-C transition present in the fourth repeat for the largest alleles, while the last one showed the basic TTTTC motif. The variant allele was indistinguishable from the consensus allele in PAG electrophoresis. Therefore, allele 9 exists in two basic forms, with the unit repeat TTTTC and the T-to-C transition in the fourth repeat, the latter with a frequency of 75% (3 of 4 samples sequenced in this study). Glock et al. (23) reached the same conclusions sequencing the complementary strand.

Although HumCD4 locus displays a low degree of informativeness with respect to other STRs, it is very suitable for individual identification, because positive results can be achieved with minimal amounts of DNA from bloodstains (400 pg in our experience; data not shown). For this reason, the HumCD4 system is a helpful tool in forensic analysis.

## References

1. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 1989;44:388–96.
2. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991;49:746–56.

3. The Utah Marker Development Group. A collection of ordered tetranucleotide-repeat markers from the human genome. Am J Hum Genet 1995;57:619–28.
4. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 Short Tandem Repeat loci for use in personal identification applications. Am J Hum Genet 1994;55:175–89.
5. Jeffreys AJ, Allen MJ, Hagelberg E, Sonnberg A. Identification of the skeletal remains of Josef Mengele by DNA analysis. Forensic Sci Int 1992;56:65–76.
6. Rand S, Puers C, Skowasch K, Wiegand P, Budowle B, et al. Population genetics and forensic efficiency data of 4 AMPFLP's. Int J Leg Med 1992;104:329–33.
7. Oldroyd NJ, Urquart AJ, Kimpton CP, Millican ES, Watson SK, et al. A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification. Electrophoresis 1995;16:334–7.
8. Gill P, Kimpton CP, Urquart A, Oldroyd N, Millican ES, et al. Automated short tandem repeat (STR) analysis in forensic case-Work—a strategy for the future. Electrophoresis 1995;16:1543–52.
9. Edwards MC, Clemens PR, Tristan M, Pizzuti A, Gibbs RA. Pentanucleotide repeat length polymorphism at the human CD4 locus. Nucleic Acids Res 1991;19:4791.
10. Brinkmann B. The STR approach. In: Carracedo A, Brinkmann B, Bar W, editors. Advances in Forensic Haemogenetics 6. Berlin-Heidelberg: Springer-Verlag, 1996;41–51.
11. Casarino L, Mannucci A, Kimpton CP, Presciuttini S, Bruni G, et al. Forensic evaluation of HUMCD4: an Italian database. Int J Leg Med 1996;109:49–51.
12. Urquart A, Kimpton CP, Downes TJ, Gill P. Variation in short tandem repeat sequences—A survey of twelve microsatellite loci for use as forensic identification markers. Int J Leg Med 1994;107: 13–20.
13. Budowle B, Baetchel FS. Modifications to improve the effectiveness of restriction fragment length polymorphism typing. Appl Theor Electrophoresis 1990;1:181–7.
14. Waye JS, Presley L, Budowle B, Shutler GG, Fourney RM. A simple method for quantifying human genomic DNA in forensic specimen extracts. BioTechniques 1989;7:852–5.
15. Allen RC, Graves G, Budowle B. Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver. BioTechniques 1989;7:736–44.
16. Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC. Analysis of the VNTR locus D1S80 by PCR followed by high resolution PAGE. Am J Hum Genet 1991;48:137–44.
17. Smith CAB. Chi-squared tests with small numbers. Ann Hum Genet 1986;50:163–7.
18. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;89:583–90.
19. Roff DA, Bentzen P. The statistical analysis of mitochondrial DNA-polymorphisms and the problem of small samples. Mol Biol Evol 1989;6:539–45.
20. Fisher RA. Standard calculations for evaluating a blood group system. Heredity 1951;5:95–102.
21. Garber RA, Morris JW. General equations for the average power of exclusion for genetic systems of n codominant alleles in one-parent and no-parent cases of disputed parentage. In: Walker RH, editor. Inclusion Probabilities in Parentage Testing. Arlington: AABB, 1983;277–80.
22. DNA Commission of the International Society for Forensic Haemogenetics. DNA recommendations-1994 report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. Int J Leg Med 1994;107:159–60.
23. Glock B, Schwartz DWM, Schwartz-Jungl EM, Mayr WR. Sequence determination of an allelic ladder for the STR polymorphism at the CD4 locus and application of the ladder in testing an Austrian Caucasoid population sample. Forensic Sci Int 1995;78: 125–30.
24. Huckenbeck W, Kuntze K, Scheil HG. German population data on the locus HumCD4. Medicina Legalis Baltica 1996;7:32–5.
25. Urquart A, Kimpton CP, Gill P. Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. Hum Genet 1993;92:637–8.
26. Adams M, Urquart A, Kimpton CP, Gill P. The human D11S554 locus: four distinct families of repeat pattern alleles at one locus. Hum Mol Genet 1993;2:1373–6.
27. Möller A, Meyer E, Brinkmann B. Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. Int J Leg Med 1994;106:319–23.
28. Brinkmann B, Meyer E, Junge A. Complex mutational events at the HumD21S11 locus. Hum Genet 1996;98:60–4.

Additional information and reprint requests:
Dr. Adriano Tagliabracci
Istituto di Medicina Legale
Policlinico Torrette
I-60020 Ancona (Italy)