

CHARACTERISATION OF NEW TETRANUCLEOTIDE POLYMORPHISM OF THE STR D5S1462 LOCUS

Dorota MONIES, Marzanna CIESIELKA, Roman MADRO

Chair and Department of Forensic Medicine, Medical Academy, Lublin

ABSTRACT: The paper characterises a new tetranucleotide D5S1462 system (GDB: 08448). Genotyping of the system was performed in 102 unrelated individuals and in 50 trios (father – mother – child). Seven alleles were detected and sequenced using the automatic sequencer ALFexpress II. The sequence analysis revealed the presence of a complex repetitive sequence $(ATCT)_n(ATCC)_nATCT$. The distribution of genotype frequencies was found to be consistent with the Hardy-Weinberg equilibrium. The calculated values of mean exclusion chance (*MEC*), power of discrimination (*PD*), polymorphism information content (*PIC*) and heterozygosity were 0.521, 0.895, 0.705 and 0.755, respectively.

KEY WORDS: Short tandem repeat; D5S1462; Allelic ladder; Sequencing; Population data.

Problems of Forensic Sciences, vol. LVII, 2004, 81–87

Received 7 April 2004; accepted 20 May 2004

INTRODUCTION

The analysis of human genetic markers is important in the forensic field, where, among other things, it may be used for relationship examinations in immigration cases, paternity testing and identification of biological traces [5, 6, 14, 15].

From the point of view of geneticists performing DNA examinations for forensic purposes the most interesting are polymorphic microsatellite sequences – the human genome contains about 100 000 specific loci in which these sequences occur [2, 7].

However, not all STR markers are useful for forensic purposes. Their inclusion in routine tests is determined by: the degree of heterozygosity, frequency of alleles, low mutation rate, heredity independent of other markers, uncomplicated and reproducible method of allele detection and their stability in biological material [11, 16].

Numerous new polymorphic STR loci which are likely to be a useful tool in genetic expert opinions have not been characterised yet.

The present study characterised a new tetranucleotide D5S1462 system (Gen Bank Accession number G08448) located on chromosome 5. This locus was initially isolated as a sequence tagged site (STS) from the human genome and was labeled human STS CHLC.GATA3H06.P6348 clone GATA3H06 by the CHLC marker development group [13].

MATERIALS AND METHODS

The study involved 102 unrelated inhabitants of South-East Poland of both sexes and 50 blood samples collected from children. DNA was extracted using the DNA Blood Mini Prep Kit (A & A Biotechnology). The primers listed by the CHLC group on the Internet site: <http://gdbwww.gdb.org/> were applied in the PCR:

P1: 5'-CTTTCCCTCTCTTCCCACAT-3;

P2: 5'-CCTGGGGGTAAGAGGAGATA-3.

In the first stage of the study the conditions of amplification were defined. The optimisation of PCR included the following primer concentrations: 0.1, 0.2, 0.3, 0.5, 0.8 μM and their annealing temperature ranged from 52°C to 58°C. Finally, amplification was performed using 2–10 ng human genomic DNA in 15 μl reaction volume. The reaction mixture contained 1 \times Taq buffer, 200 μM dNTPs and 0.6 U Taq polymerase (Promega Corporation, USA), 0.2 μM of each primer. A total of 30 cycles was carried out in a TriBlock (Biometra) with denaturation for 50 s at 94°C, annealing for 70 s at 55°C, extension for 30 s at 72°C and final extension for 7 min at 72°C. Then 30 DNA samples were genotyped. The PCR products were separated in an ALFexpress II sequencer (Amersham Pharmacia Biotech) using the ALFwin Fragment Analyser 1.00.

Selected alleles of various sizes were recovered from the gel according to Lazaruk et al. [10] and sequenced using the AmpliCycle Sequencing Kit (Applied Biosystems) in the above-mentioned sequencer. Sequence analysis was conducted with the ALFwin Sequence Analyser 2.10 programme. After analysing the structure of alleles and classifying them (according to the recommendations of the International Society for Forensic Haemogenetics) on the basis of the number of repetitive sequences [4], the alleles were included in the allelic fragment ladders. For this purpose, the alleles were subjected to re-amplification and mixed in suitable proportions on the basis of signal intensities, and then reamplified, however, this time, in one test-tube.

In the second part of the study using the allele ladder constructed, the remaining DNA samples were genotyped in the ALFexpress II sequencer.

The results were statistically analysed. Consistency with the Hardy-Weinberg equilibrium was evaluated with the exact test according to the

Miller computer programme [12]. Moreover, the usefulness of the marker for forensic purposes was assessed on the basis of mean exclusion chance-MEC [17], polymorphism information content-PIC [1], power of discrimination-PD [8] and heterozygosity of the system [12]. In order to determine whether the segregation of features is consistent with basic rules of heredity, family studies (father – mother – child) were performed.

RESULTS AND DISCUSSION

The optimal concentration of primers for D5S1462 was found to be 0.2 μ M and the optimal temperature of their annealing in PCR (at which most product is obtained without additional artifacts) – 55°C. Under the described conditions, the amplification of each of the 102 examined samples and then their electrophoretic separation were uneventful and the sensitivity of amplification of the D5S1462 locus was high.

Table I presents the sequence structure of 7 identified alleles, which reveals that the locus is characterised by a complex repetitive sequence of the (ATCT)_n (ATCC)_n ATCT type; the subsequent allelic fragments differ in one repetitive unit and the observed alleles are within the range of 14–20 repeats.

TABLE I. DNA SEQUENCES OF ALLELIC FRAGMENTS OF THE D5S1462 LOCUS

Allele	Sequence
14 (182 bp)	S1(20)...20 bp...(ATCT)9 (ATCC)4 ATCT...66 bp...S2(20)
15 (186 bp)	S1(20)...20 bp...(ATCT)10 (ATCC)4 ATCT...66 bp...S2(20)
16 (190 bp)	S1(20)...20 bp...(ATCT)11 (ATCC)4 ATCT...66 bp...S2(20)
17 (194 bp)	S1(20)...20 bp...(ATCT)11 (ATCC)5 ATCT...66 bp...S2(20)
18 (198 bp)	S1(20)...20 bp...(ATCT)12 (ATCC)5 ATCT...66 bp...S2(20)
19 (202 bp)	S1(20)...20 bp...(ATCT)13 (ATCC)5 ATCT...66 bp...S2(20)
20 (206 bp)	S1(20)...20 bp...(ATCT)14 (ATCC)5 ATCT...66 bp...S2(20)

Tables II and III present the frequencies of 7 observed alleles and their phenotypes. The allele size was within the 182–206 bp range. Allele 17 ($f = 0.3431$) and 16 ($f = 0.3137$) occurred most frequently while allele 20 was the least frequent one and occurred with $f = 0.0098$. Eighteen out of 28 phenotypes theoretically possible for this number of alleles were observed and the most frequent phenotype found was 16–17 ($f = 0.1961$).

TABLE II. ALLELE FREQUENCIES OF THE D5S1462 LOCUS IN A GROUP OF 102 UNRELATED INDIVIDUALS OF SOUTH-EAST POLAND

Allele	14	15	16	17	18	19	20
No. of alleles	3	15	64	70	32	18	2
Frequency	0.0147	0.0735	0.3137	0.3432	0.1569	0.0882	0.0098

TABLE III. PHENOTYPE FREQUENCIES OF THE D5S1462 LOCUS IN A GROUP OF 102 UNRELATED INDIVIDUALS OF SOUTH-EAST POLAND

Phenotype	No.	Frequency	Phenotype	No.	Frequency
14–16	2	0.0196	16–18	7	0.0686
14–17	1	0.0098	16–19	6	0.0588
15–15	1	0.0098	17–17	11	0.1078
15–16	7	0.0686	17–18	15	0.1471
15–17	4	0.0392	17–19	7	0.0686
15–18	1	0.0098	17–20	1	0.0098
15–19	1	0.0098	18–18	2	0.0196
16–16	11	0.1078	18–19	4	0.0392
16–17	20	0.1961	18–20	1	0.0098

The locus studied is consistent with the Hardy-Weinberg equilibrium ($P = 0.891$). In 50 trios (father – mother – child) no mutation was observed. The values of *MEC*, *PD*, *PIC* and heterozygosity were: 0.518, 0.895, 0.705 and 0.755, respectively. Thus, with regard to polymorphism, this system is less informative than e.g. Penta E or FGA [9]; however, it is similar to many other systems used in routine studies, e.g. TH01 [3, 18]. These results suggest that D5S1462 is a useful marker for forensic casework and paternity testing.

References:

1. Botstein D., White R. L., Skolnick M. [et al.], Construction of a genetic linkage map in man using restriction fragment length polymorphism, *American Journal of Human Genetics* 1980, vol. 32, pp. 314–331.
2. Edwards A., Civitello A., Hammond H. A. [et al.], DNA typing and genetic mapping with trimeric and tetrameric tandem repeat, *American Journal of Human Genetics* 1991, vol. 49, pp. 746–756.
3. Entrala C., Lorente M., Lorente J. A. [et al.], Fluorescent multiplex analysis of nine STR loci: Spanish population data, *Forensic Science International* 1998, vol. 98, pp. 179–183.

4. Gill P., Brinkmann B., d'Aloja E. [et al.], Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature, *Forensic Science International* 1997, vol. 87, pp. 185–192.
5. Gill P., Sparkes R., Fereday L. [et al.], European Network of Forensic Science Institutes (ENSFI) formulation and testing of principles to evaluate STR multiplexes, *Forensic Science International* 2000, vol. 108, pp. 1–29.
6. Hou Y. P., Tang J. P., Dong J. G. [et al.], Further characterization and population data for the pentanucleotide STR polymorphism D10S2325, *Forensic Science International* 2001, vol. 123, pp. 107–110.
7. Jeffreys A. J., Wilson V., Thein S. L., Hypervariable “minisatellite” regions in human DNA, *Nature* 1985, vol. 314, pp. 67–73.
8. Kloosterman A. D., Daselaar P., Budowle B. [et al.], Population genetic study on the HLA-DQ α and the D1S80 locus in Dutch Caucasians, [in:] Proceedings from the Third International Symposium on Human Identification, Promega Corporation, 1992.
9. Kozioł P., Ciesielka M., Mađro R., Genetic data on nine complex STR systems in the population of South-East Poland and their usefulness in paternity testing, *Problems of Forensic Sciences* 2000, vol. 45, pp. 66–80.
10. Lazaruk K., Wallin J., Holt C. [et al.], Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing, *Forensic Science International* 2001, vol. 119, pp. 1–10.
11. Mayer W. R., DNA markers in forensic medicine, TCB 1995, vol. 4, pp. 325–338.
12. Miller M. P., Tools for Population Genetic Analyses (TFPGA), <http://herb.bio.nau.edu/~miller/tfpga.html>.
13. Murray J., Sheffield V., Weber J. L. [et al.], Cooperative Human Linkage Center, <http://gdbwww.gdb.org/>.
14. Reichenpfader B., Leinzinger E. P., Klintschar M., D1S1171: a new highly variable short tandem repeat polymorphism, *International Journal of Legal Medicine* 2002, vol. 116, pp. 195–197.
15. Reichenpfader B., Zehner R., Klintschar M., Characterization of a highly variable short tandem repeat polymorphism at the D2S1242 locus, *Electrophoresis* 1999, vol. 20, pp. 514–517.
16. Reynolds R., Sensabaugh G., Analysis of genetic markers in forensic samples using the polymerase chain reaction, *Analytical Chemistry* 1991, vol. 63, pp. 2–15.
17. Wiegand P., Budowle B., Rand S. [et al.], Forensic validation of the STR systems SE 33 and TC11, *International Journal of Legal Medicine* 1993, vol. 105, pp. 315–320.
18. Zupanič I., Bala J., Komel R., Analysis of nine short tandem repeat (STR) loci in the Slovenian population, *International Journal of Legal Medicine* 1998, vol. 111, pp. 248–250.

CHARAKTERYSTYKA GENETYCZNA NOWEGO POLIMORFIZMU O CZTEROTETRA nukleotydowej jednostce powtarzalnej typu STR – LOKUS D5S1462

Dorota MONIES, Marzanna CIESIELKA, Roman MADRO

WSTĘP

Badania genetycznych markerów człowieka mają istotne znaczenie dla celów sądowych, gdyż ich wyniki służą do ustalania ojcostwa i pokrewieństwa oraz do identyfikacji śladów biologicznych [5, 6, 14, 15]. W genomie człowieka znajduje się około stu tysięcy specyficznych *loci*, w których występują potencjalnie przydatne do tego celu polimorficzne sekwencje mikrosatelitarne [2, 7]. Jednak nie scharakteryzowano dotychczas wielu polimorficznych *loci* STR. Tymczasem o tym, które z nich mogą być włączone do rutynowych badań, decydują: stopień heterozygotyczności, częstość alleli, niska mutacyjność, dziedziczenie niezależne od innych markerów, nieskomplikowana i powtarzalna metoda fenotypowania oraz trwałość w materiale biologicznym [11, 16].

W niniejszej pracy przedstawiono dane dotyczące czteronukleotydowego układu D5S1462 (GenBank Accession, nr G08448) zlokalizowanego na chromosomie 5, który podczas konstruowania mapy fizycznej genomu człowieka został zidentyfikowany jako STS CHLC.GATA3H06.P6348, klon GATA3H06 [13].

MATERIAŁY I METODY

Próbki krwi uzyskano od 102 niespokrewnionych mieszkańców Polski południowo-wschodniej obojga płci oraz od 50 dzieci par rodzicielskich, które wchodziły w skład tej grupy. DNA wyizolowano za pomocą zestawu DNA Blood Mini Prep Kit firmy A & A Biotechnology. Do reakcji PCR zastosowano startery:

1. 5'-CTTTCCCTCTCTCCACAT-3';
2. 5'-CCTGGGGGTAAGAGGAGATA-3';

podane przez grupę CHLC, zajmującą się badaniami markerów genetycznych, na stronie internetowej: <http://gdbwww.gdb.org/>. Produkty rozdzielano i analizowano za pomocą sekwenatora ALFexpress II (Amersham Pharmacia Biotech).

W pierwszym etapie badań ustalono warunki amplifikacji. Uwzględniono stężenia starterów: 0,1; 0,2; 0,3; 0,5; 0,8 μ M oraz temperaturę ich przyłączenia w zakresie od 52°C do 58°C. Następnie zgenotypowano 30 próbek DNA. Produkty PCR analizowano przy użyciu programu ALFwin Fragment Analyser 1.00. Wybrano następnie allele różniące się wielkością; odzyskano je z żelu wg metody opisanej przez Lazaruk i in. [10] orzaz zsekwencjonowano ze starterami znakowanymi Cy5 przy użyciu zestawu AmpliCycle Sequencing Kit (Applied Biosystems), po czym przeprowadzono analizę sekwencji w oparciu o program ALFwin Sequence Analyser 2.10.

Drabinę allelicznych fragmentów utworzono po przeanalizowaniu struktury alleli i oznaczeniu ich (zgodnie z zaleceniami Międzynarowego Towarzystwa Hemogene-

tyki Sądowej) na podstawie ilości sekwencji repetytywnych [4]. W tym celu allele wycięte z żelu poddano reamplifikacji i zmieszano je w proporcjach odpowiednich do intensywności sygnału, a następnie ponownie namnożono, tym razem w jednej próbce. Ta właśnie drabina posłużyła jako marker do genotypowania pozostałych próbek DNA.

Wyniki fenotypowania poddano analizie statystycznej. Zgodność badanej populacji z regułą Hardy'ego i Weinberga oceniono w oparciu o test „exact” [12] (wg programu Millera). Oceniono również przydatność przebadanego układu STR do ekspertyz sądowych na podstawie: średniej szansy wykluczenia – *MEC* [17], współczynnika informacji o polimorfizmie – *PIC* [1] i współczynnika dyskryminacji – *PD* [8] oraz heterozygotyczności układu [12].

WYNIKI I ICH DYSKUSJA

Wykazano, że dla układu genetycznego D5S1462 optymalne stężenie starterów wynosi 0,2 μ M, a optymalna temperatura (przy której uzyskuje się najwięcej produktu PCR, ale bez dodatkowych artefaktów) to 55°C, bowiem w tych warunkach amplifikacja wszystkich z przebadanych próbek DNA oraz elektroforetyczny rozdział produktów przebiegły bez żadnych problemów, a czułość powielania okazała się wysoka.

Zidentyfikowano 7 alleli o wielkości 182–206 pz i złożonej sekwencji (ATCT)_n (ATCC)_n ATCT obejmującej 14 do 20 powtórzeń, które różnią się między sobą 1 jednostką repetytywną (tabela I). Najczęstszymi allelami okazały się allele 17 i 16, a najrzadszymi 14 i 20 (tabela II). Zidentyfikowano również 18 z 28 możliwych teoretycznie fenotypów (tabela III), spośród których 6 wystąpiło sporadycznie (z częstością 0,0098).

Wykazano, że badany locus znajduje się w równowadze Hardy'ego-Weinberga ($P = 0,891$). W 50 „trójkach” ojciec – matka – dziecko nie zaobserwowano mutacji. Obliczono współczynniki *MEC*, *PD* i *PIC* oraz jego heterozygotyczność (0,518; 0,895; 0,705 i 0,755). Pod względem polimorfizmu układ ten jest zatem mniej informatywny od np. układu Penta E, ewentualnie FGA [9], podobny jest natomiast do wielu innych stosowanych rutynowo [3, 18]. Uzyskane wyniki świadczą więc o tym, że układ D5S1462 jest przydatny w badaniach z zakresu genetyki sądowej.