

## IDENTIFICATION OF THE RARE DYS393\*10 ALLELE IN THE NORTHERN POLISH POPULATION

Anita DETTLAFF-KĄKOL<sup>1</sup>, Ryszard PAWŁOWSKI<sup>1, 2</sup>

<sup>1</sup>*Chair and Department of Forensic Medicine, Medical Academy, Gdańsk*

<sup>2</sup>*Institute of Forensic Research, Cracow*

**ABSTRACT:** The paper describes a rare new allele of the DYS393 locus in a Polish population – DYS393\*10. This allele was found in a population sample of 740 unrelated males originating from north Poland. The presence of the new allele was proved using DNA sequencing analysis. Moreover, the homogeneity of distribution of DYS393 alleles between different populations was analysed. Very significant statistical differences were observed between the north Polish and the whole Polish population and the whole European population (Y-STR DNA base). This observation strongly supports the supposition that haplotype frequencies must be calculated on the basis of one's own database, and not on the basis of the pan-European database.

**KEY WORDS:** DYS393 locus; DYS393\*10; Population genetics; North Poland.

*Z Zagadnień Nauk Sądowych, z. XLVIII, 2001, 89-97*

*Received 20 August 2001; accepted 29 September 2001*

### INTRODUCTION

Locus DYS393 belongs to a group of loci which recently have been introduced to male DNA identification studies in biological traces. Up till now 8 alleles of this locus (from 9 to 16), within the size range from 107–135 bp, have been identified. DYS393 is characterised by the AGAT repetitive sequence [5]. Thanks to the setting up of an Internet web site (<http://ystr.charite.de>) containing a Y chromosome haplotypes database, comparison of frequencies of particular alleles and haplotypes for many different populations has become possible. This data base consists of 5708 haplotypes originating from 46 European populations for 9 Y-STR loci, including almost 600 haplotypes from four different Polish population samples (Bydgoszcz, Gdańsk, Warsaw, Wrocław). Analysis of the contents of the database suggests that 3 alleles of locus DYS393 are very common (DYS393\*12, \*13 and \*14), 2 of them are rare (DYS393\*11 and \*15) and 3 very rare (DYS393\*9, \*10, \*16). The presence of the very rare allele DYS393\*16 in the Polish population has already been described. This allele appeared in the population of 5708 males from Europe only twice [4].

The present work describes the identification of the next very rare DYS393 allele – allele \*10. Because of significant differences in allele frequency distributions among European populations for Y-chromosome loci, (contrarily to frequency distribution of autosomal alleles), an attempt was made to compare allele DYS393 frequencies for Polish populations with frequencies for other populations gained from the above mentioned database – with the aim of assessing it's usefulness for drawing statistical conclusions.

#### MATERIALS AND METHODS

Biological material (blood and buccal swabs) collected from 740 unrelated males living in northern Poland was investigated. This population sample included genetic profiles of 213 men described previously [4]. DNA extraction, measurement of DNA concentration, amplification of alleles of DYS393 locus and detection and sequencing of PCR products on the ABI310 automatic DNA sequencer and appropriate statistical calculations were done according to previously described methods [4]. The DYS393 allele frequencies for other population samples from Poland (Bydgoszcz, Warsaw, Wrocław) were taken from the generally accessible world database of Y-chromosomal loci haplotypes (<http://ystr.charite.de>). This database included, on 23 June 2001, 5708 DNA profiles, with 596 haplotypes from Poland. The allele frequencies ascertained for the population sample from Cracow were taken from the original paper [2]. The allelic ladder for locus DYS393 was kindly supplied by Professor B. Brinkmann from the Institute of Legal Medicine in Münster (Germany).

#### RESULTS AND DISCUSSION

The frequency distribution of 740 alleles ascertained for locus DYS393 in the population sample from north Poland is presented in Table I. During the population study the presence of 6 out of 8 alleles described till now was observed. The allele frequency distribution is very similar to that observed for a population of 213 males (test  $\chi^2$ :  $P = 0.979$ ; G-statistic:  $P = 0.999$ ) in previous studies [4]. It was also found that the coefficient of diversity ( $D = 0.320$ ) calculated for the 740 alleles is very similar to that previously described. The obtained results show that increasing the analysed population sample from 213 to 740 does not significantly change the distribution of allele frequencies – and so the threshold value of 200 alleles, as is generally acknowledged, is sufficient for gaining a reliable distribution. During this study the presence of a 107 bp allele (ABI 310, internal standard GS500), 8 bp shorter than

allele 12, situated in the allelic ladder used, was identified (Figure 1). The new allele was initially classified as allele DYS393\*10. In order to ultimately prove that the observed allele can be classified as allele 10 of the DYS393 locus, sequencing of the new allele was performed. This technique allowed us to establish the exact length of the DNA fragment as 111 bp and its nucleotide sequence as follows: 5' – **GTGGTCTTCTACTTGTGTCAATAC** (AGAT)<sub>10</sub> (ATGT)<sub>2</sub> **CTTTTCTATGAGACATACCTCATT**TTTTTTGGACTTGAGTT – 3', where primer sequences are printed in bold. A fully complementary sequence was obtained by sequencing of the opposite DNA strand of the PCR product (Figure 2). The obtained sequence is in very good concordance with the sequence ascertained for allele DYS393\*11 by Pestoni et al. [5], except, of course, that it has one less AGAT repetitive unit.

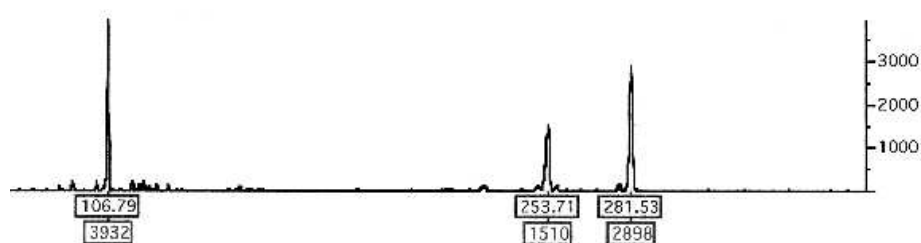


Fig. 1. Identification of the rare allele DYS393\*10 using capillary electrophoresis on ABI310 sequencer. From left to the right, alleles DYS393\*10, DYS392\*14 and DYS391\*11).

TABLE I. COMPARISON OF DYS393 LOCUS ALLELES DISTRIBUTION BETWEEN FIVE POPULATION SAMPLES FROM POLAND AND EUROPE

Allel	Gdańsk		Bydgoszcz		Kraków		Warszawa		Wrocław		E	
	n	[%]	n	[%]	n	[%]	n	[%]	n	[%]	n	[%]
9	0	0.0000	0	0	0	0	0	0	0	0	2	0.0405
10	1	0.1351	0	0	0	0	0	0	0	0	6	0.1216
11	0	0.0000	0	0	0	0	0	0	0	0	39	0.7906
12	52	7.0270	13	8.2802	15	14.0187	18	11.4650	2	1.6529	592	12.0008
13	603	81.4865	139	88.5350	83	77.5701	122	77.7070	103	85.1240	3550	71.9643
14	78	10.5405	15	9.5541	8	7.4766	15	9.5541	13	10.7438	664	13.4604
15	5	0.6756	1	0.6369	1	0.9345	2	1.2739	0	0	78	1.5812
16	1	0.1351	0	0	0	0	0	0	0	0	2	0.04505
	740	100	168	168	107	100	157	100	121	100	4933	100

E – sum of DYS393 allele frequencies for many European populations (except Polish populations); data obtained from Internet page: <http://ystr.charite.de>; state on 5 June 2001.



cumulative frequency of occurrence of alleles of the DYS393 locus in all remaining populations included in the Internet Y-chromosome DNA database with the exception of the data from Poland (4933 alleles). A comparison of the frequencies distribution between the studied population sample from north Poland and other Polish populations revealed statistically significant homogeneity (Table II). The lowest  $P$  value was observed for the population sample from Cracow ( $\sim 0.31$ ), and the highest for the population sample from Bydgoszcz ( $P = 0.94$ ). Comparison of the homogeneity of the DYS393 alleles distribution between the population sample from north Poland and also between the Polish populations and the European population (database) revealed very significant differences in DYS393 allelic distribution. In all cases  $P$  values equalling 0.0000 were noted (Table II), which strongly suggests that all calculations of the Y-STR haplotype frequencies have to be based on frequencies ascertained for local population or joint frequencies for Poland accessible in the Internet Y-STR database. Similar differences were observed previously, when performing comparative analysis of the northern Polish population with western populations based on data for the DYS19 and the DYS390 loci [4], where a tendency towards allele frequencies shift in the direction of those consisting of a greater number of repetitive units was clearly visible. The necessity for the above proposed course of action was also postulated by Pascali et al. [3] and Forster et al. [1].

TABLE II. COMPARISON OF DYS393 ALLELES HOMOGENEITY DISTRIBUTION BETWEEN DIFFERENT POPULATIONS

Compared populations	$\chi^2$	P	G-statistic	P
Gd $\times$ B	0.94	0.9440	1.31	0.9400
Gd $\times$ Wa	4.64	0.4630	4.59	0.4810
Gd $\times$ Ww	2.63	0.6450	3.78	0.5953
Gd $\times$ K	7.20	0.5530	6.59	0.3100
Gd $\times$ E	37.04	0.0000	44.76	0.0000
PL $\times$ E	53.51	0.0000	64.37	0.0000

Gd – population sample from north Poland (740 males); B; Wa; Ww; K – population samples analysed respectively in Institutes of Forensic Medicine in Bydgoszcz, Warsaw, Wrocław, and Institute of Forensic Research in Cracow; E – sum of DYS393 allele frequencies for many European populations (except Polish populations); data obtained from Internet page: <http://ystr.charite.de>; state on 23 June 2001.

## References:

1. Forster P., Kayser M., Meyer E. [et al.], Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans, *Molecular and Biological Evolution* 1998, vol. 15, pp. 1108–1114.
2. <http://ystr.charite.de>.
3. Kupiec T., Branicki W., van Mersbergen C. [et al.], Population genetic study of Y chromosome specific STR-loci in a population sample from Southern Poland, *Progress in Forensic Genetics* 2000, vol. 8, pp. 312–314.
4. Pascali V. L., Dobosz M., Brinkmann B., Coordinating Y-chromosomal STR research for the Courts, *International Journal of Legal Medicine* 1999, vol. 112, p. 1.
5. Pawłowski R., Branicki W., Kupiec T., Y-chromosomal polymorphic loci DYS19, DYS390, DYS393 in a population sample from northern Poland, *Electrophoresis* 1999, vol. 20, pp. 1702–1706.
6. Pestoni C., Cal M. L., Lareu M. V. [et al.], Y chromosome STR haplotypes: genetic and sequencing data of the Galician population (NW Spain), *International Journal of Legal Medicine* 1998, vol. 112, pp. 15–21.

## IDENTYFIKACJA RZADKIEGO ALLELA DYS393\*10 W POPULACJI POLSKI PÓŁNOCNEJ

Anita DETTLAFF-KĄKOL, Ryszard PAWŁOWSKI

### WSTĘP

Lokus DYS393 należy do grupy loci, które ostatnio wprowadzono do badań identyfikacyjnych DNA mężczyzn w śladach biologicznych. Do tej pory w tym lokus zidentyfikowano 8 alleli (od 9 do 16) zawartych w przedziale wielkości między 107–135 pz. DYS393 charakteryzuje się jednostką repetytywną typu AGAT [5]. Dzięki uruchomieniu strony internetowej (<http://ystr.charite.de>) zawierającej bazę danych haplotypów DNA z chromosomu Y, stało się możliwe zestawienie częstości poszczególnych alleli jak i haplotypów dla wielu różnych populacji. Baza ta zawiera 5708 haplotypów z 46 europejskich populacji dla 9 loci typu Y-STR, w tym również dla 4 różnych populacji polskich (Bydgoszcz, Gdańsk, Warszawa, Wrocław) liczących prawie 600 haplotypów. Analiza zawartości bazy pokazuje, iż w zakresie lokus DYS393 występują 3 powszechne allele (DYS393\*12, \*13 i \*14), 2 rzadkie (DYS393\*11 i \*15) oraz trzy bardzo rzadkie (DYS393\*9, \*10, \*16). W populacji polskiej dotychczas opisano obecność bardzo rzadkiego allela DYS393\*16, który w populacji 5708 mężczyzn z całej Europy pojawił się tylko 2 razy [4].

Niniejsza praca opisuje zidentyfikowanie następnego bardzo rzadkiego allela w zakresie lokus DYS393, a mianowicie DYS393\*10. Ze względu na istniejące znaczne różnice w rozkładach częstości alleli w populacjach europejskich dla loci z chromosomu Y – w odróżnieniu od rozkładu częstości alleli obecnych na autosomach – podjęto próbę porównania rozkładu częstości alleli lokus DYS393 w populacjach polskich z innymi populacjami prezentowanymi w bazie danych w celu oceny jej przydatności dla wnioskowania statystycznego.

### MATERIAŁY I METODY

Badaniom poddano próbki materiału biologicznego (krew i wymazy z jamy ustnej) pochodzące od 740 niespokrewnionych mężczyzn z obszaru Polski północnej. Ta próbka populacyjna zawierała w sobie profile DNA 213 mężczyzn opisanych uprzednio [4]. Izolację DNA, pomiar jego stężenia, amplifikację alleli lokus DYS393 oraz detekcję i sekwencjonowanie produktów na automatycznym sekwencjonatorze DNA ABI310, jak również stosowne obliczenia statystyczne, prowadzono jak podano uprzednio [4]. Częstości alleli DYS393 dla innych populacji z obszaru Polski (Bydgoszcz, Warszawa, Wrocław) uzyskano z ogólnie dostępnej światowej bazy haplotypów loci z chromosomu Y (<http://ystr.charite.de>). Baza ta w dniu 23 czerwca 2001 r. zawierała 5708 profili DNA, w tym 596 haplotypów z obszaru Polski. Częstości alleli dla populacji krakowskiej zaczerpnięto z publikacji [2]. Drabinę alleli dla lokus DYS393 uzyskano dzięki uprzejmości Profesora B. Brinkmanna z Instytutu Medycyny Sądowej w Münster (RFN).

## WYNIKI I DYSKUSJA

Rozkład częstości 740 alleli lokus DYS393 zaobserwowany w próbie populacyjnej z obszaru Polski północnej przedstawia tabela I. W badaniach populacyjnych stwierdzono istnienie 6 z 8 do tej pory opisanych alleli. Wspomniany rozkład jest bardzo zbliżony do tego, który zaobserwowano dla populacji 213 mężczyzn w poprzednich badaniach (test  $\chi^2$ :  $P = 0,979$ ; test G-statistic:  $P = 0,999$ ) [4]. Ustalono również, że dla 740 alleli współczynnik różnorodności ( $D = 0,320$ ) jest bardzo zbliżony do poprzedniego. Jak wskazują uzyskane wyniki, powiększenie próbki populacyjnej z 213 do 740 osób nie zmienia znacząco rozkładu alleli, a więc wartość progowa 200 alleli, jak się powszechnie uważa, jest wystarczająca dla uzyskania wiarygodnego rozkładu. Podczas badania zidentyfikowano obecność allele o wielkości ok. 107 pz (ABI310, standard wewnętrzny GS500), krótszego o 8 pz od allele 12 znajdującego się w dostępnej drabinie alleli (rycina 1). Allel ten wstępnie określono jako DYS393\*10. W celu ostatecznego potwierdzenia, iż obserwowany allel jest 10 allelem w lokus DYS393, przeprowadzono jego sekwencjonowanie. Badaniem tym wykazano obecność fragmentu DNA o długości 111 pz i następującej sekwencji: 5' – **GTGGTCTTCTACTTGTGTCAATAC** (AGAT)<sub>10</sub> (ATGT)<sub>2</sub> **CTTTTCTATGAGACATACCTCATT TTTTGGACTTGAGTT** – 3', gdzie tłustym drukiem zaznaczono sekwencje starterów. W pełni komplementarny obraz uzyskano, sekwencjonując drugą nić produktu PCR (rycina 2). Uzyskana sekwencja w pełni odpowiada tej, którą – dla allele DYS393\*11 – obserwowali Pestoni i in. [5], z wyjątkiem, oczywiście, o jeden mniejszej liczbie jednostek repetytywnych typu AGAT.

Allel ten nie był dotychczas obserwowany w przebadanej grupie 596 mężczyzn z populacji polskiej (<http://ystr.charite.de>). Jest on drugim bardzo rzadkim allelem, który, oprócz opisanego już DYS393\*16 [4], ujawniono w populacji Polski północnej. Na obszarze Europy DYS393\*10, obok allele DYS393\*9 i DYS393\*16, należy również do bardzo rzadkich, gdyż pojawił się tylko u 6 mężczyzn; występowanie allele DYS393\*10 stwierdzono bowiem w populacji Austrii, Szwajcarii i Norwegii [6].

W ostatnim czasie pojawia się też coraz więcej doniesień udowadniających bardzo duże różnice w częstości występowania poszczególnych alleli i haplotypów pomiędzy różnymi populacjami [1, 3]. Różnice te mają więc bezpośredni i duży wpływ na określenie częstości haplotypów dla poszczególnych populacji. Rzutuje to oczywiście na właściwe kalkulacje podawane przy ocenie statystycznej częstości profili Y-STR w opiniach z zakresu genetyki sądowej. Rozwiązaniem problemu jest stworzenie odpowiednio dużej własnej bazy zawierającej minimum 200 anonimowych profili DNA z chromosomu Y lub skorzystanie z dostępnej w Internecie bazy danych zawierającej obecnie prawie 600 haplotypów DNA występujących na obszarze Polski.

W celu stwierdzenia ewentualnych różnic pomiędzy badanymi populacjami a innymi populacjami polskimi oraz innymi populacjami prezentowanymi w bazie danych, dokonano analizy rozkładu częstości alleli w zakresie badanego lokus. Tabela I przedstawia rozkład obserwowanych częstości alleli lokus DYS393 dla 5 populacji z obszaru Polski oraz łącznej częstości występowania alleli DYS393 we wszystkich pozostałych populacjach ujętych w internetowej bazie danych z pominięciem danych z obszaru Polski (łącznie 4933 allele). Wzajemne porównanie rozkładu częstości pomiędzy badaną populacją Polski północnej oraz innymi populacjami polskimi wykazało statystycznie znamiennej homogenność rozkładu (tabela II).



Najniższą wartość  $P$  zaobserwowano dla populacji krakowskiej (ok. 0,31), a najwyższą dla populacji bydgoskiej, gdzie wartość  $P$  wynosiła 0,94.

Porównanie homogenności rozkładu alleli DYS393 pomiędzy próbką populacyjną z Polski północnej, jak również pomiędzy populacjami polskimi i populacją Europy (baza danych), wykazało bardzo znamienne różnice w rozkładzie alleli tego lokus. W każdym przypadku zaobserwowano wartości  $P$  równe 0,0000 (tabela II), co zdecydowanie wskazuje na konieczność opierania wszelkich obliczeń częstości haplotypów Y-STR na częstościach zaobserwowanych w lokalnej populacji lub łącznych częstościach dla obszaru Polski podanych w internetowej bazie Y-STR. Podobne różnice obserwowano uprzednio, porównując populację Polski północnej z populacjami zachodnimi w zakresie *loci* DYS19 i DYS390 [4], gdzie wyraźnie widoczna była tendencja przesunięcia częstości alleli w kierunku tych, które posiadały większą ilość jednostek repetytywnych. Na konieczność tego typu postępowania wskazywał również Pascali i in. [3] oraz Forster i in. [1].