

# FREQUENCY OF OCCURRENCE OF SPECIAL COMBINATIONS OF STR ALLELES IN HUMAN Y CHROMOSOMES CLASSIFIED INTO HAPLOGROUP I\*

Marcin WOŹNIAK, Jakub CZARNY, Tomasz GRZYBOWSKI,  
Danuta MIŚCICKA-ŚLIWKA

*Chair and Department of Forensic Medicine, Collegium Medicum in Bydgoszcz,  
Nicolaus Copernicus University, Toruń*

**ABSTRACT:** This paper presents an analysis of correlation between occurrence of certain Y-STR alleles and Y chromosome haplogroup, as defined by biallelic genetic markers. DNA samples from 134 unrelated men from the Pomerania-Kujawy region of Poland have been typed. For each Y chromosome, its STR haplotype for 9 loci (“minimal haplotype”) and haplogroup defined by M9 and M170 markers have been determined. Y chromosomes have been defined as belonging to I\* or non-I\* haplogroup. The results suggest that there could be a strong relationship between some of the Y-STR haplotypes and the chromosome assignment to the defined haplogroup.

**KEY WORDS:** Y chromosome polymorphism; Y-SNP; Y-STR; Population analysis.

*Problems of Forensic Sciences, vol. LIX, 2004, 79–99*  
*Received 28 August 2004; accepted 14 October 2004*

## INTRODUCTION

Microsatellite and biallelic loci of the Y chromosome (Y-STR and Y-SNP) are tools that are frequently utilised in molecular phylogenetics [14], medical genetics [9], genealogy [10] and forensic medicine [11]. Y-SNP loci reveal a lower mutation rate and diversity and therefore are especially useful in comparative studies of population genetic structure [18] as well as phylogenetic research [22]. Highly variable Y-STR markers are particularly useful in personal identification studies [3], but they are also increasingly frequently applied to phylogenetic analyses [6].

In the area of phylogenetic studies, Y-SNP polymorphism is used for definition of haplogroups – i.e. sets of chromosomes belonging to the same paternal line. The criterion of haplogrouping is based on the presence or lack of characteristic mutations in particular loci. Highly polymorphic Y-STR haplotypes (i.e. different combinations of alleles characteristic of a particular set of loci) are mainly used for the purpose of diversification among chromosomes included in a particular haplogroup. The term “haplogroup” is used

with reference to paternal lines defined by Y-SNP polymorphism, whereas the term “haplotype” describes paternal lines defined by Y-STR polymorphism [5].

The present work is an attempt to figure out whether it is possible to identify particular haplotypes or single alleles of the Y-STR loci in the population of the Pomerania-Kujawy region, which reveal correlation with the haplogroup characteristic of the analysed chromosome. To this end Y-chromosome markers were analysed in the set of chromosomes defining haplogroup I\*. The determination of haplogroup I\* is based on the presence of point mutation in the locus M170 and a simultaneously non-mutated M9 locus [20, 21]. Potential correlations between occurrence of particular Y-STR alleles and the haplogroup characteristic of the analysed Y chromosome were studied by analysis of 9 Y-STR loci defining the so-called minimal haplotype (according to <http://www.yhrd.org>) and determination of I\* haplogroup by analysis of the above mentioned I\* haplogroup characteristic diagnostic mutations. The usefulness of the obtained results in forensic genetics is also discussed.

## MATERIALS AND METHODS

The study was conducted on nuclear DNA extracted with the organic method [19] from peripheral blood samples collected from 134 unrelated males living in the Bydgoszcz area (population of Pomerania-Kujawy). DNA concentration was determined with the hybridisation method using QuantiBlot kit (Applied Biosystems). The polymorphism of locus M9 was analysed using the procedure described by Kayser et al. [12]. Analysis of the M170 locus polymorphism was carried out with the unpublished protocol supplied by Prof. M. Kayser (Erasmus MC – University Medical Centre Rotterdam, Department of Forensic Molecular Biology, The Netherlands). Amplification of the Y-STR loci: DYS19, DYS385, DYS389 I and II, DYS390, DYS391, DYS392 and DYS393, consisting of the so-called minimal haplotype, was performed using 7 pairs of primers designed by Kayser et al. [13]. PCR was conducted in two multiplex and one monoplex reaction. The first multiplex (multiplex I) included primers for the loci DYS19 (0.32  $\mu$ M), DYS390 (0.32  $\mu$ M) and DYS389I/II (0.4  $\mu$ M), the second multiplex (multiplex II) included primers for the loci: DYS391 (0.4  $\mu$ M), DYS392 (0.96  $\mu$ M), and DYS393 (0.24  $\mu$ M) and the locus DYS385 was amplified in a monoplex reaction using 0.12  $\mu$ M concentration of primers. Primer concentrations were optimised for the purpose of this study; parameters concerning PCR mixture components and amplification conditions were optimised by other authors [13]. One  $\mu$ l of each PCR product was mixed with solutions contain-

ing 1.5 µl of Blue Dextran Loading Solution (Promega) and 0.5 µl of Fluorescent Ladder 60–400 bp (Promega). Amplification products were resolved in denaturing polyacrylamide gel (5% polyacrylamide:bisacrylamide 19:1, 6 M urea) using ABI377 apparatus with the set of filters “A”. Electrophoresis results were analysed with GeneScan 3.7 and Genotyper 1.2 software (ABI). Correctness of the haplotyping was controlled by simultaneous electrophoresis of samples with known haplotype. Data concerning frequencies of the particular haplotypes characteristic of chromosome Y, subjected to inter-population comparisons, were taken from the Y chromosome Reference Database (YHRD), accessible at the internet address <http://www.ystr.org>. Statistical significance of differences in allele and haplotype frequencies of the STR loci localised on chromosomes including and not including haplogroup I\* was calculated based on a double-sided test of the difference between structural parameters. The program Statistica ver. 5.1 was used for calculations and statistically significant differences were considered when  $p < 0.05$ .

## RESULTS

Twenty out of 134 Y chromosomes subjected to analysis of Y-SNP polymorphism were included in haplogroup I\* (frequency 0.1493). The study of the 9 Y-STR loci (minimal haplotype) revealed 111 different haplotypes in the analysed population sample.

Table I presents frequencies of occurrence of individual alleles of the analysed STR loci according to chromosome haplogroup classification. Particularly significant frequency differences in individual groups of chromosomes characteristic of alleles 11 and 13 in locus DYS385, alleles 12 and 28 in locus DYS389 and alleles 22 and 25 in locus DYS390 are noticeable. Allele 11 of locus DYS385 occurs in every third chromosome outside haplogroup I\*, while it was not observed in chromosomes classified in haplogroup I\*. Allele 12 of locus DYS389 occurs in half of the chromosomes from haplogroup I\* and merely 8% of chromosomes outside this haplogroup. Similarly, allele 28 of this locus was revealed in every third chromosome from haplogroup I\* and only in one out of 20 outside this haplogroup. Moreover, allele 22 of locus DYS390 appears in 30% of samples classified in haplogroup I\*, while in the case of samples classified outside this haplogroup, this allele only occurs at a frequency about 0.02. Allele 25 of locus DYS390 is characteristic for half of the samples from outside haplogroup I\* and it was not detected in any sample classified in this haplogroup.

Analysis of the frequencies characteristic for particular haplotypes of locus DYS385 (Table II) revealed that haplotype 11,14, the most frequent in the studied population is present in almost 50% of chromosomes from out-

TABLE I. FREQUENCIES DETERMINED FOR ALLELES OF PARTICULAR LOCI ACCORDING TO HAPLOGROUP CLASSIFICATION

Allele	A	B	A + B	<i>p</i>
DYS19				
13	0.0000	0.0702	0.0597	0.2239
<b>14</b>	<b>0.3500</b>	<b>0.1140</b>	<b>0.1493</b>	<b>0.0072</b>
15	0.2000	0.2368	0.2313	0.7194
16	0.3500	0.3596	0.3582	0.9343
17	0.1000	0.2193	0.2191	0.2221
DYS385				
9	0.0000	0.0044	0.0037	0.7668
10	0.0000	0.1009	0.0858	0.1397
<b>11</b>	<b>0.0000</b>	<b>0.3289</b>	<b>0.2799</b>	<b>0.0003</b>
12	0.0250	0.0307	0.0299	0.8903
<b>13</b>	<b>0.2250</b>	<b>0.0263</b>	<b>0.0560</b>	<b>0.0005</b>
14	0.4750	0.3465	0.3657	0.2731
15	0.2250	0.0877	0.1082	0.0705
16	0.0500	0.0263	0.0299	0.5665
17	0.0000	0.0219	0.0187	0.5053
18	0.0000	0.0175	0.0149	0.5522
19	0.0000	0.0088	0.0075	0.6744
DYS389I				
<b>12</b>	<b>0.5000</b>	<b>0.0789</b>	<b>0.1418</b>	<b>0.0000</b>
<b>13</b>	<b>0.5000</b>	<b>0.8070</b>	<b>0.7612</b>	<b>0.0035</b>
14	0.0000	0.1053	0.0896	0.1307
15	0.0000	0.0088	0.0075	0.6744
DYS389II				
27	0.0000	0.0088	0.0075	0.6744
<b>28</b>	<b>0.3000</b>	<b>0.0526</b>	<b>0.0896</b>	<b>0.0005</b>
29	0.1500	0.3158	0.2910	0.1346
30	0.3000	0.4474	0.4254	0.2210
31	0.2000	0.1140	0.1269	0.2884
32	0.0500	0.0526	0.0522	0.9616
33	0.0000	0.0088	0.0075	0.6744

Allele	A	B	A + B	<i>p</i>
DYS390				
<b>22</b>	<b>0.3000</b>	<b>0.0263</b>	<b>0.0672</b>	<b>0.0000</b>
23	0.1500	0.1140	0.1194	0.6477
<b>24</b>	<b>0.5500</b>	<b>0.2807</b>	<b>0.3209</b>	<b>0.0188</b>
<b>25</b>	<b>0.0000</b>	<b>0.5088</b>	<b>0.4328</b>	<b>0.0000</b>
26	0.0000	0.0702	0.0597	0.2239
DYS391				
9	0.0000	0.0088	0.0075	0.6744
10	0.4000	0.6228	0.5896	0.0639
11	0.5500	0.3596	0.3881	0.1094
12	0.0500	0.0088	0.0149	0.1637
DYS392				
11	0.9000	0.8070	0.8209	0.3189
12	0.1000	0.0263	0.0373	0.1110
13	0.0000	0.1491	0.1269	0.0668
14	0.0000	0.0175	0.0149	0.5522
DYS393				
12	0.0000	0.0877	0.0746	0.1709
13	0.9500	0.8070	0.8284	0.1201
14	0.0500	0.0965	0.0896	0.5029
15	0.0000	0.0088	0.0075	0.6744

A – frequency of an allele in samples classified into haplogroup I\*; B – frequency of an allele in samples from outside haplogroup I\*; A + B – frequency of an allele in the total population; *p* – statistical significance of difference between frequencies. Alleles characterised by significantly higher frequency in chromosomes from haplogroup I\* are shown in bold type. Alleles characterised by significantly higher frequency in chromosomes from outside haplogroup I\* are shown in bold and italic type.

side haplogroup I\*, while it was not determined among chromosomes classified within this haplogroup. Furthermore, haplotypes of locus *DYS385*, with the shorter allele 13 or 14 (13,14; 13,15; 14,14; 14,15 and 14,16), which are rather rare in the studied population (total frequency of these haplotypes in the analysed population sample equals approximately 17%) are represented in haplogroup I\* with a total frequency of 95%. This means that almost each

chromosome of this haplogroup has one of the above haplotypes of locus DYS385, while total frequency of these haplotypes in chromosomes from outside haplogroup I\* barely exceeds 3%.

TABLE II. FREQUENCY OF OCCURRENCE OF DYS385 LOCUS HAPLOTYPES

Haplotype	A	B	A + B	<i>p</i>
9,14	0.0000	0.0088	0.0075	0.6744
10,10	0.0000	0.0088	0.0075	0.6744
10,11	0.0000	0.0175	0.0149	0.5522
10,14	0.0000	0.1667	0.1418	0.0508
11,11	0.0000	0.0175	0.0149	0.5522
11,13	0.0000	0.0263	0.0224	0.4645
<b><i>11,14</i></b>	<b><i>0.0000</i></b>	<b><i>0.4561</i></b>	<b><i>0.3881</i></b>	<b><i>0.0002</i></b>
11,15	0.0000	0.1228	0.1045	0.1001
<b>12,13</b>	<b>0.0500</b>	<b>0.0000</b>	<b>0.0075</b>	<b>0.0180</b>
12,14	0.0000	0.0439	0.0373	0.3413
12,16	0.0000	0.0088	0.0075	0.6744
12,17	0.0000	0.0088	0.0075	0.6744
13,14	0.2500	0.0000	0.0373	0.0000

Haplotype	A	B	A + B	<i>p</i>
<b>13,15</b>	<b>0.1500</b>	<b>0.0263</b>	<b>0.0448</b>	<b>0.0149</b>
<b>14,14</b>	<b>0.1500</b>	<b>0.0000</b>	<b>0.0224</b>	<b>0.0001</b>
<b>14,15</b>	<b>0.3000</b>	<b>0.0088</b>	<b>0.0522</b>	<b>0.0000</b>
<b>14,16</b>	<b>0.1000</b>	<b>0.0000</b>	<b>0.0149</b>	<b>0.0009</b>
14,17	0.0000	0.0088	0.0075	0.6744
15,17	0.0000	0.0088	0.0075	0.6744
15,18	0.0000	0.0088	0.0075	0.6744
16,17	0.0000	0.0088	0.0075	0.6744
16,18	0.0000	0.0175	0.0149	0.5522
16,19	0.0000	0.0175	0.0149	0.5522
17,18	0.0000	0.0088	0.0075	0.6744
9,14	0.0000	0.0088	0.0075	0.6744

A – frequency of haplotype in samples classified into haplogroup I\*; B – frequency of haplotype in samples from outside haplogroup I\*; A + B – frequency of a haplotype in the total population; *p* – statistical significance of difference between frequencies. Haplotypes characterised by significantly higher frequency in chromosomes from haplogroup I\* are shown in bold type. Haplotypes characterised by significantly higher frequency in chromosomes from outside haplogroup I\* are shown in bold and italic type.

Haplotypes 12-28, 12-29 and 13-31 of locus DYS389 I & II appear in chromosomes classified into haplogroup I\* much more frequently than in chromosomes from outside this haplogroup. An extreme example, describing preferential occurrence of some haplotypes of locus DYS389 I & II associated with classification of the analysed chromosome within haplogroup I\* or outside this haplogroup, consists in the 13-29 allele combination.

In spite of the fact that the frequency of this haplotype in the analysed population is significant and exceeds 0.25, its occurrence was not noted among chromosomes from haplogroup I\* (Table III).

Detailed inspection of frequencies of complete minimal haplotypes (Table IV and Figure 1) revealed that the most frequent haplotype among chromosomes of haplogroup I\* (marked as modal haplotype I\* – HMI\*), was not found among chromosomes from outside this haplogroup. The average frequency of this haplotype in European populations equals approximately 0.0015 (34 times in 22 093 samples included in the YHRD database) and it was noticeable this haplotype was found only in populations from central and eastern Europe. Moreover, it was found that the most frequent haplotype in the general population of Pomerania and Kujawy (marked as

nHMI\*) revealing a frequency of 0.05, was not determined in the analysed population among chromosomes from haplogroup I\* (Figure 1).

TABLE III. FREQUENCY OF OCCURRENCE OF THE DYS389 I & II HAPLOTYPES

Haplotype	A	B	A + B	<i>p</i>
12-27	0.0000	0.0088	0.0075	0.6744
<b>12-28</b>	<b>0.3000</b>	<b>0.0351</b>	<b>0.0746</b>	<b>0.0001</b>
<b>12-29</b>	<b>0.1500</b>	<b>0.0088</b>	<b>0.0299</b>	<b>0.0008</b>
12-30	0.0500	0.0175	0.0224	0.3661
12-31	0.0000	0.0088	0.0075	0.6744
13-28	0.0000	0.0175	0.0149	0.5522
<b><i>13-29</i></b>	<b><i>0.0000</i></b>	<b><i>0.3070</i></b>	<b><i>0.2612</i></b>	<b><i>0.0046</i></b>
13-30	0.2500	0.3947	0.3731	0.2193

Haplotype	A	B	A + B	<i>p</i>
<b>13-31</b>	<b>0.2000</b>	<b>0.0614</b>	<b>0.0821</b>	<b>0.0392</b>
13-32	0.0500	0.0263	0.0299	0.5665
14-30	0.0000	0.0351	0.0299	0.3965
14-31	0.0000	0.0351	0.0299	0.3965
14-32	0.0000	0.0263	0.0224	0.4645
14-33	0.0000	0.0088	0.0075	0.67
15-31	0.0000	0.0088	0.0075	0.67

A – frequency of haplotype in samples classified into haplogroup I\*; B – frequency of haplotype in samples from outside haplogroup I\*; A + B – frequency of a haplotype in the total population; *p* – statistical significance of difference between frequencies. Haplotypes characterised by significantly higher frequency in chromosomes from haplogroup I\* are shown in bold type. Haplotypes characterised by significantly higher frequency in chromosomes from outside haplogroup I\* are shown in bold and italic type.

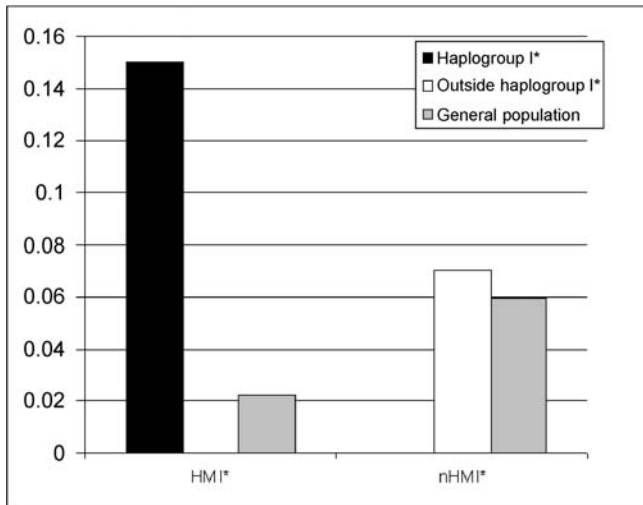


Fig. 1. Frequency of occurrence of haplotypes HMI\* and nHMI\* in sets of chromosomes belonging and not belonging to haplogroup I\* and in the general population.

The performed analyses of frequencies of the minimal haplotypes which most frequently occur in the population of Pomerania and Kujawy in haplogroup I\* (HMI\*) and outside haplogroup I\* (nHMI\*) for other population samples from Poland indicate that haplotype HMI\* appears in all tested

populations relatively rarely. This applies particularly to regions of Białystok and Warsaw (Figure 2). The frequency of occurrence of haplotype HMI\* in the studied population significantly differs from the frequencies characteristic of populations from the above mentioned regions. There are statistically significant differences in the frequencies of haplotype HMI\* between the population of Bydgoszcz and the population of Białystok and Warsaw, between the population of Cracow and the population of Białystok and Gdańsk and Warsaw as well as between the population of Wrocław and the population of Warsaw (Table V).

TABLE IV. FREQUENCY OF OCCURRENCE OF COMPLETE MINIMAL HAPLOTYPES

Haplotype	A	B	A + B	<i>p</i>
<b>16-13-30-24-11-11-13-14,15</b>	<b>0.15</b>	<b>0</b>	<b>0.0224</b>	<b>0.0000</b>
14-12-28-22-10-11-14-14,16	0.05	0	0.0075	0.0180
14-12-28-22-10-12-13-13,14	0.05	0	0.0075	0.0180
14-12-28-22-11-12-13-14,14	0.05	0	0.0075	0.0180
14-12-28-23-10-11-13-13,14	0.05	0	0.0075	0.0180
14-12-28-23-10-11-13-13,15	0.05	0	0.0075	0.0180
14-12-29-22-10-11-13-12,13	0.05	0	0.0075	0.0180
14-12-29-23-10-11-13-14,15	0.05	0	0.0075	0.0180
15-12-28-22-10-11-13-13,14	0.05	0	0.0075	0.0180
15-12-29-22-10-11-13-13,14	0.05	0	0.0075	0.0180
15-13-30-24-11-11-13-13,15	0.05	0	0.0075	0.0180
15-13-31-24-11-11-13-14,16	0.05	0	0.0075	0.0180
16-12-30-24-11-11-13-14,14	0.05	0	0.0075	0.0180
16-13-31-24-11-11-13-13,15	0.05	0	0.0075	0.0180
16-13-31-24-11-11-13-14,14	0.05	0	0.0075	0.0180
16-13-31-24-12-11-13-14,15	0.05	0	0.0075	0.0180
17-13-30-24-11-11-13-13,14	0.05	0	0.0075	0.0180
17-13-32-24-11-11-13-14,15	0.05	0	0.0075	0.0180

A – frequency of haplotype in samples classified into haplogroup I\*; B – frequency of haplotype in samples from outside haplogroup I\*; A + B – frequency of a haplotype in the total population; *p* – statistical significance of difference between frequencies. Haplotype HMI\* is shown in bold type. The allelic order in the presented haplotypes is concordant with the sequence of loci: DYS19-DYS389I-DYS389II-DYS390-DYS391-DYS392-DYS393-DYS385.

Results of performed comparative analyses of the minimal haplotype frequencies of chromosomes from outside haplogroup I\* most frequently occurring in the area of Bydgoszcz (nHMI\*) did not indicate statistically significant differences in the occurrence of this haplotype in various regions of Poland. The exception in this case is the population of Bydgoszcz, where haplotype nHMI\* seems to occur significantly more frequently than in populations of Białystok and Warsaw.

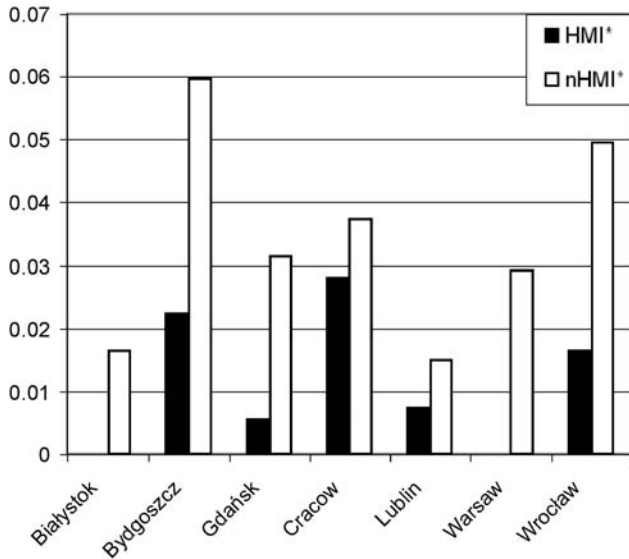


Fig. 2. Frequency of occurrence of haplotypes HMI\* and nHMI\* in populations of Poland.

TABLE V. STATISTICAL SIGNIFICANCE ( $p$ ) OF THE DIFFERENCES BETWEEN FREQUENCIES OF OCCURRENCE OF HAPLOTYPES HMI\* AND NHMI\* IN POPULATIONS FROM POLAND

	16-13-30-24-11-11-13-14,15 (HMI*)						
	Białystok	Bydgoszcz	Gdańsk	Cracow	Lublin	Warsaw	Wrocław
Białystok		0.0433	0.3164	0.0240	0.2428	1.0000	0.0831
Bydgoszcz	0.0392		0.0617	0.6844	0.3804	0.0285	0.8354
Gdańsk	0.4278	0.1189		0.0262	0.7895	0.1204	0.5389
Cracow	0.3741	0.4245	0.7444		0.2171	0.0074	0.4966
Lublin	0.7352	0.0529	0.3346	0.2664		0.1799	0.6487
Warsaw	0.5509	0.0147	0.9516	0.6008	0.3876		0.0286
Wrocław	0.1516	0.7165	0.2795	0.7277	0.0601	0.3184	
	16-13-29-25-10-11-13-11,14 (nHMI*)						

Statistical significance was determined by  $p$  value calculated on the basis of a double-sided test of the difference between structural parameters. Significantly different frequencies ( $p < 0.05$ ) are shown in bold type. The allelic order in the presented haplotypes is concordant with the sequence of loci: DYS19-DYS389I-DYS389II-DYS390-DYS391-DYS392-DYS393-DYS385.

## DISCUSSION

The characteristic feature of the Y chromosome is that out of 60 billion bp making up this chromosome, a section of 50 billion bp is practically free of re-



combination [4]. Mutation alterations occurring in the area of the non-recombining portion of the Y chromosome (abbreviated to NRY – non recombining Y chromosome) reveal common inheritance (like a genetic unit) independently of physical distance between them. Taking this into account, it is possible to consider the hypothesis assuming that individual mutations in SNP and STR loci of the Y chromosome which arose at different times on chromosomes constituting a particular paternal line are inherited together as well as to further evaluate potential implications of this phenomenon in legal medicine.

Classical methods of phylogenetic analysis are often based on the assumption that each mutation which constitutes a diagnostic marker determining individual clades (groups) is a unique event in the evolutionary history of a particular species [15]. This assumption can be true in the case of markers characterised by low polymorphism, such as SNP [17], but in the case of highly variable markers such as STR loci, the probability of multiple mutation events which generate the same particular allele is higher [8]. Therefore, it is possible that Y chromosomes gathered in different haplogroups have identical STR alleles, as a result of parallel mutations.

Unification of nomenclature performed by the Y Chromosome Consortium for Y-SNP loci gave rise to a universal system enabling classification of each studied Y chromosome into one of the 153 haplogroups, based on analysis of just a few markers [21]. Nomenclature of Y-STR loci does not cause so many difficulties because of a commonly accepted system of their identification which is concordant with directions of the International Society for Forensic Genetics (ISFG) [7]. The recently achieved unification of nomenclature for haplogroups and haplotypes of the Y chromosome enables genetic comparative studies of Y chromosomes originating from different populations that are geographically distant and analysed by different laboratories. Earlier, such studies were difficult, mainly because of the non-uniform nomenclature of Y chromosome haplogroups [21]. The literature data indicates, for instance, that in the Croatian population almost all chromosomes of haplogroup I\* have a similar Y-STR haplotype. It is described by the authors as the Dinarsko modal haplotype (DMH) [2]. This haplotype is presented as 16-24-11-11-13 (with respect to loci DYS19-390-391-392-393) and direct descendants of this haplotype (differing in terms of mutations consisting in presence or lack of single repetitive units assembling microsatellites of the Y chromosome) also occur among Y chromosomes of haplogroup I\* of the Pomerania and Kujawy population; however, their association with chromosomes of this haplogroup is not so obvious.

It seems that the results of the present study indicate the existence of a correlation between classification of a Y chromosome into a particular haplogroup, defined by Y-SNP polymorphism, and occurrence on this chro-

mosome of particular alleles and sets of alleles of Y-STR loci. Especially, the presence of haplotypes with a shorter allele 13 or 14 in the locus DYS385 (Table II), the occurrence of sets 12-28 and 12-29 in locus DYS389 I & II (Table III) and the presence of allele 22 in locus DYS390 (Table I) can be considered as an indicator of I\* haplogroup origins of the studied chromosome. Similarly, the presence of the set of alleles 11 and 14 in locus DYS385 and the occurrence of allele 25 in locus DYS390 indicate that the chromosome is from outside haplogroup I\*. The minimal haplotype, which in the studied population is strongly associated with chromosomes of haplogroup I\* was also identified. According to data included in the YHRD database, this haplotype occurs in European populations with a frequency of about 0.0015.

Haplogroup I\*, characterised and described for the first time as Eu7 by Semino et al. [20] is the second most frequent haplogroup determined in the Polish population, just after haplogroup Eu19 (R1a1\*). The quoted authors assessed that its frequency is approximately 0.236, which means that almost 25% of Polish males should belong to haplogroup I\*.

The study conducted on the population sample from Pomerania-Kujawy (Bydgoszcz area) proved that the frequency of this haplogroup equals approximately 0.1493. This value is much lower than the frequency calculated by Semino et al. [20], but the difference appears statistically insignificant ( $p = 0.1554$ ) after taking into consideration the divergence between the number of chromosomes studied by Semino et al. ( $n = 55$ ) and the much more abundant sampling in the present study ( $n = 134$ ). Comparative analysis of the frequency of the haplogroup I\* modal haplotype (HMI\*) characteristic of the population of Pomerania-Kujawy with frequency of its occurrence in other Polish populations, based on data from the YHRD database, shows statistically significant differences between the population of Pomerania-Kujawy and the population of Białystok, Cracow and Warsaw (Table V). Such a result can be explained by variation in frequencies of haplogroup I\* in the analysed populations, but it can also be a result of a lack of direct correlation between classification of a chromosome into haplogroup I\* and occurrence of the HMI\* haplotype. Because of the low frequency of haplotype HMI\* in all the studied populations, the hypothesis assuming that observed differences are artefacts caused by insufficient representation of haplotypes in the analysed populations also seems probable.

Taking into consideration results suggesting homogeneity of the Polish population in terms of frequency of occurrence of individual alleles and haplotypes of STR loci [16] as well as the results of present calculations indicating a strong correlation between individual Y-STR alleles, particularly alleles of loci DYS385, DYS389 and DYS390, and haplogroup classification, it is possible to attempt to estimate (approximately) the frequency of haplogroup I\* in other Polish populations (besides the population of Pome-

rania-Kujawy). Data revealing real frequencies of this haplogroup in the studied populations can be compared with calculated theoretical values, which opens up the possibility of checking the hypothesis assuming a relationship between individual alleles/haplotypes of Y-STR systems and chromosomes of haplogroup I\*. Assuming that the Polish population is homogeneous, frequencies of alleles/haplotypes of Y-STR systems correlated with haplogroup I\* should be constant among chromosomes of this haplogroup. Assuming this, the expected theoretical frequency of chromosomes included in haplogroup I\* for a particular population can be calculated with the formula:

$$phgX = paX \times phgW / paW \quad \{1\}$$

where:  $phgX$  – theoretical frequency of chromosomes of any haplogroup in a population  $X$ ;  $paX$  – frequency of individual allele/haplotype or their set in a particular locus and a population  $X$ ;  $phgW$  – frequency of a haplogroup in a model population;  $paW$  – frequency of allele/haplotype or their set in a particular locus and model population.

Theoretical frequencies of chromosomes from haplogroup I\* calculated on the basis of the above assumptions do not significantly differ between populations of Bydgoszcz and Białystok, Gdansk, Cracow, Lublin, Warsaw and Wrocław (Table VI). Taking into consideration the above quoted data concerning homogeneity of the Polish population in terms of Y-STR allele

TABLE VI. THEORETICAL FREQUENCIES OF CHROMOSOMES OF HAPLOGROUP I\* IN DIFFERENT POPULATIONS FROM POLAND CALCULATED ON THE BASIS OF FREQUENCIES OF DYS385 LOCUS HAPLOTYPES SIGNIFICANTLY ASSOCIATED WITH CHROMOSOMES OF HAPLOGROUP I\*

Population	Haplotypes DYS385						Theoretical frequency of haplogroup I*	$n$	$p$
	13/14	13/15	14/14	14/15	14/16	Total frequency			
Białystok	0.0275	0.0000	0.0220	0.1099	0.0165	0.1758	0.1529	182	0.9210
Gdańsk	0.0295	0.0092	0.0350	0.0847	0.0221	0.1805	0.1570	543	0.9954
Cracow	0.0280	0.0093	0.0374	0.0748	0.0187	0.1682	0.1463	107	0.8215
Lublin	0.0385	0.0075	0.0224	0.0373	0.0149	0.1206	0.1049	104	0.2555
Warsaw	0.0583	0.0125	0.0292	0.0583	0.0125	0.1708	0.1486	240	0.8308
Wrocław	0.0165	0.0083	0.0413	0.0496	0.0165	0.1322	0.1150	121	0.9443

The  $p$  value describes the statistical significance of the difference between frequency calculated for the individual population and the frequency of chromosomes from haplogroup I\* determined for the population of Pomerania and Kujawy. Data concerning frequencies of particular haplotypes characteristic for other populations were retrieved from the YHRD database;  $n$  – number of individuals characteristic for each population from the database.

frequencies, then expectations concerning similar homogeneity in terms of SNP markers defining haplogroup classification should have been assumed. It can be supposed that the obtained values of theoretical frequency of chromosomes from haplogroup I\* in populations of Poland confirm the possibility of determination of frequency of any Y chromosome haplogroup on the basis of frequencies of microsatellite loci of this chromosome. However, there must exist characteristic Y-STR alleles associated with a particular haplogroup. It seems that the success of such a calculation would be dependent on a close genetic relationship between the studied population and the model population on the basis of which correlations between frequencies of particular haplogroups and haplotypes have been determined. The analysis discussed above should be regarded as an introduction to wider studies on relationships between occurrence of particular Y-STR alleles and individual haplogroups. Verification of the assumed association between particular haplotypes or Y-STR alleles and the haplogroup classification of the analysed chromosome requires data on Y chromosome SNP loci for other Polish populations. Bearing in mind that the present results were obtained for a single haplogroup in one single population, we can not exclude the possibility that the observed associations between particular alleles of the Y-STR loci and classification of the chromosome into haplogroup I\*, despite in some cases being statistically significant, can be caused by the stochastic phenomenon related to the small number of analysed chromosomes ( $n = 20$ ) included in haplogroup I\*. Studies concerning haplotyping and haplogrouping of Y chromosomes seem to be useful for legal medicine. Determination of relations between Y-SNP alleles and Y-STR alleles could be used for the development of a quality control system, similar to the efficient quality systems applied to the studies of mtDNA polymorphisms [1]. The hypothetical quality system could be based on independent analysis of Y-STR and Y-SNP loci followed by comparative verification of obtained results based on concordance or discordance between the ascertained haplogroup and haplotype. Application of such an approach could, e.g. in some cases, reveal mutations in positions of annealing of primers used for Y-STR amplification. This system could also be applied to particularly complicated paternity cases, e.g. when available reference DNA material comes from individuals positioned on further branches of the hereditary tree of the putative father. In such cases, haplogrouping in addition to haplotype determination connected with verification of compatibility between results of both examinations could provide an opportunity for a conclusive decision as to whether there is a common paternal line for the determined haplotypes or whether the similarity is coincidental – caused by high mutation frequency of Y-STR loci. Before the described tests are applied in casework, it is necessary to unambiguously establish the associations between haplogroups and haplo-

types of chromosome Y and, moreover, develop a simple, reliable and relatively inexpensive method of determination of the Y chromosome haplogroup, since current methods based on PCR-RFLP technique are too expensive and time consuming.

If the association between occurrence of particular alleles or Y-STR haplotypes and haplogroup classification of chromosome is proved by analysis of a broader range of material originating from genetically different populations, a deeper insight into the evolutionary history of the Y chromosome will become possible. At the same time it should be possible to apply this finding in the applied sciences including forensic genetics.

#### References:

1. Bandelt H. J., Lahermo P., Richards M. [et al.], Detecting errors in mtDNA data by phylogenetic analysis, *International Journal of Legal Medicine* 2001, vol. 115, pp. 64–69.
2. Barać L., Perićić L., Klarić I. M. [et al.], Y chromosomal heritage of Croatian population and its island isolates, *European Journal of Human Genetics* 2003, vol. 11, pp. 535–542.
3. Beleza S., Alves C., Gonzalez-Neira A. [et al.], Extending STR markers in Y chromosome haplotypes, *International Journal of Legal Medicine* 2003, vol. 117, pp. 27–33.
4. Bosch E., Calafell F., Comas D. [et al.], High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between Northwestern Africa and the Iberian Peninsula, *American Journal of Human Genetics* 2001, vol. 68, pp. 1019–1029.
5. de Knijff P., Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome, *American Journal of Human Genetics* 2000, vol. 67, pp. 1055–1061.
6. Di Giacomo F., Luca F., Anagnou N., Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects, *Molecular Phylogenetics and Evolution* 2003, vol. 28, pp. 387–395.
7. Gill P., Brenner C., Brinkmann B. [et al.], DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, *Forensic Science International* 2001, vol. 124, pp. 5–10.
8. Holtkemper U., Rolf B., Hohoff C. [et al.], Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques, *Human Molecular Genetics* 2001, vol. 10, pp. 629–633.
9. Jobling M. A., Tyler-Smith C., New uses for new haplotypes the human Y chromosome, disease and selection, *Trends in Genetics* 2000, vol. 16, pp. 356–362.
10. Jobling M. A., In the name of the father: surnames and genetics, *Trends in Genetics* 2001, vol. 17, pp. 353–357.

11. Jobling M. A., Pandya A., Tyler-Smith C., The Y chromosome in forensic analysis and paternity testing, *International Journal of Legal Medicine* 1997, vol. 110, pp. 118–124.
12. Kayser M., Brauer S., Weiss G. [et al.], Melanesian origin of Polynesian Y chromosomes, *Current Biology* 2000, vol. 10, pp. 1237–1246.
13. Kayser M., Caglia A., Corach D. [et al.], Evaluation of Y-chromosomal STRs: a multicenter study, *International Journal of Legal Medicine* 1997, vol. 110, pp. 125–133.
14. Mitchell R. J., Hammer M. F., Human evolution and the Y chromosome, *Current Opinions in Genetics & Development* 1996, vol. 6, pp. 737–742.
15. Nei M., Kumar S., Population trees from genetic markers, [in:] *Molecular evolution and phylogenetics*, Oxford University Press, London 2000.
16. Płoski R., Woźniak M., Pawłowski R. [et al.], Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis, *Human Genetics* 2002, vol. 110, pp. 592–600.
17. Raitio M., Lindroos K., Laukkanen M. [et al.], Y-chromosomal SNPs in Finno-Ugric-Speaking populations analyzed by minisequencing on microarrays, *Genome Research* 2001, vol. 11, pp. 471–482.
18. Rosser Z. H., Zerjal T., Hurles M. E., Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language, *American Journal of Human Genetics* 2000, vol. 67, pp. 1526–1543.
19. Sambrook J., Fritsh E. F., Maniatis T., Analysis and cloning of eucariotic genomic DNA, *Molecular cloning, a laboratory manual*, Cold Spring Harbor 1989.
20. Semino O., Passarino G., Oefner P. J. [et al.], The genetic legacy of Paleolithic Homo Sapiens in extant Europeans: A Y chromosome perspective, *Science* 2000, vol. 290, pp. 1155–1159.
21. The Y Chromosome Consortium, a nomenclature system for the tree of human Y-chromosomal binary haplogroups, *Genome Research* 2002, vol. 12, pp. 339–348.
22. Underhill P. A., Shen P., Lin A. A. [et al.], Y chromosome sequence variation and the history of human populations, *Nature Genetics* 2000, vol. 26, pp. 358–361.

# WYSTĘPOWANIE SZCZEGÓLNYCH KOMBINACJI ALLELI STR W LUDZKICH CHROMOSOMACH Y NALEŻĄCYCH DO HAPLOGRUPY I\*

Marcin WOŹNIAK, Jakub CZARNY, Tomasz GRZYBOWSKI,  
Danuta MIŚCICKA-ŚLIWKA

## WSTĘP

Mikrosatelitarne oraz bialleliczne *loci* chromosomu Y (Y-STR i Y-SNP) są narzędziami często wykorzystywanymi w analizach z zakresu filogenetyki molekularnej [14], w genetyce medycznej [9], genealogii [10] oraz w medycynie sądowej [11]. Ze względu na niższą częstość mutacji i mniejsze zróżnicowanie, *loci* Y-SNP są przydatne zwłaszcza w porównawczych analizach struktury genetycznej populacji [18] oraz w badaniach filogenetycznych [22]. Wysoce zmienne układy Y-STR znajdują zastosowanie przede wszystkim w badaniach z zakresu identyfikacji osobniczej [3], choć coraz szerzej są stosowane również w analizach filogenetycznych [6].

W zakresie analiz filogenetycznych polimorfizm *loci* Y-SNP wykorzystuje się do wyodrębniania haplogrup – tj. zbiorów chromosomów należących do tej samej linii ojcowskiej. Kryteriami służącymi do wyodrębnienia haplogrup jest obecność lub brak charakterystycznych mutacji w poszczególnych *loci*. Z kolei wysoce polimorficzne haplotypy *loci* Y-STR (tzn. rozmaite kombinacje alleli występujących w danym zestawie *loci*) służą głównie jako narzędzie umożliwiające zróżnicowanie chromosomów przypisanych już do określonej haplogrupy. Termin „haplogrupa” stosowany jest zatem w odniesieniu do linii ojcowskich definiowanych przez polimorfizm Y-SNP, natomiast termin „haplotyp” określa linie ojcowskie definiowane przez polimorfizm Y-STR [5].

Niniejsza praca stanowi próbę odpowiedzi na pytanie, czy możliwe jest zidentyfikowanie w populacji kujawsko-pomorskiej takich haplotypów chromosomu Y, bądź pojedynczych alleli *loci* Y-STR, które występują w korelacji z przynależnością danego chromosomu do określonej haplogrupy. W tym celu podjęto analizę markerów chromosomu Y należących do haplogrupy I\*. Kryterium służącym do wyodrębnienia I\* jako odrębnej haplogrupy chromosomu Y jest obecność mutacji punktowej w lokus M170 przy jednoczesnym zachowaniu lokus M9 w stanie niezmutowanym [20, 21]. W celu zbadania dla potrzeb niniejszej pracy ewentualnych zależności pomiędzy występowaniem określonych alleli Y-STR a przynależnością haplogrupową chromosomu Y, analizowano 9 *loci* Y-STR tworzących tzw. haplotyp minimalny (wg <http://www.yhrd.org>) oraz określano przynależność chromosomów do haplogrupy I\* na podstawie występowania mutacji diagnostycznych dla tej haplogrupy. Podjęto również próbę oceny przydatności otrzymanych wyników dla celów genetyki sądowej.



## MATERIAŁY I METODY

Materiałem do badań był genomowy DNA pozyskany metodą ekstrakcji organicznej [19] z próbek krwi obwodowej 134 niespokrewnionych mężczyzn zamieszkujących okolice Bydgoszczy (populacja kujawsko-pomorska). Stężenie uzyskiwanych preparatów DNA oceniano metodą hybrydyzacyjną z wykorzystaniem zestawu QuantiBlot (Applied Biosystems). Analizę polimorfizmu w locus M9 prowadzono według procedury opisanej przez Kaysera i in. [12]. Do analizy polimorfizmu w locus M170 wykorzystywano nieopublikowany protokół uzyskany od prof. M. Kaysera (Erasmus MC – University Medical Center Rotterdam, Department of Forensic Molecular Biology, Holandia). Do amplifikacji produktów z *loci* Y-STR: DYS19, DYS385, DYS389I i II, DYS390, DYS391, DYS392 i DYS393, których allele stanowią zestaw tzw. haplotypu minimalnego, stosowano 7 par starterów o sekwencjach zaprojektowanych przez Kaysera i in. [13]. PCR przeprowadzano w dwóch reakcjach multipleksowych i jednej reakcji monopleksowej. Pierwsza reakcja multipleksowa (multipleks I) zawierała startery dla *loci* DYS19 (0,32  $\mu$ M), DYS390 (0,32  $\mu$ M) i DYS389I/II (0,4  $\mu$ M). Druga reakcja multipleksowa (multipleks II) zawierała startery dla *loci* DYS391 (0,4  $\mu$ M), DYS392 (0,96  $\mu$ M), DYS393 (0,24  $\mu$ M), natomiast reakcja monopleksowa zawierała startery umożliwiające amplifikację alleli locus DYS385 (0,12  $\mu$ M). Stężenia starterów zoptymalizowano we własnym zakresie; pozostałe parametry dotyczące składu mieszaniny reakcyjnej PCR i warunków amplifikacji zostały opracowane przez innych autorów [13]. Do 1  $\mu$ l produktu każdej reakcji PCR dodawano mieszaniny zawierającej 1,5  $\mu$ l Blue Dextran Loading Solution (Promega) i 0,5  $\mu$ l Fluorescent Ladder 60–400 bp (Promega). Produkty amplifikacji rozdzielano w denaturującym żelu poliakrylamidowym (5% poliakrylamid:bisakrylamid 19:1, 6 M mocznik) przy użyciu aparatu ABI377 z zestawem filtrów A. Wyniki elektroforezy analizowano z wykorzystaniem oprogramowania GeneScan 3.7 i Genotyper 1.2 (ABI). Poprawność haplotypowania kontrolowano, przeprowadzając równoczesną elektroforezę produktów amplifikacji próbek o znanym haplotypie.

Dane odnośnie do częstości poszczególnych haplotypów chromosomu Y, wykorzystane do porównań międzypopulacyjnych, zaczerpnięto z Referencyjnej Bazy Danych Haplotypów Chromosomu Y (YHRD), dostępnej pod internetowym adresem <http://www.ystr.org>.

Statystyczną istotność różnic w częstości alleli i haplotypów loci STR chromosomów należących i nienależących do haplogrupy I\* obliczano na podstawie dwustronnego testu różnicy pomiędzy wskaźnikami struktury. W obliczeniach wykorzystywano oprogramowanie Statistica w wersji 5.1, uznając różnice za istotne statystycznie, gdy  $p < 0,05$ .

## WYNIKI

Wśród 134 chromosomów Y testowanych w zakresie polimorfizmu Y-SNP 20 należało do haplogrupy I\* (częstość haplogrupy = 0,1493). Wyniki badań polimorfizmu Y-STR w 9 *loci* (tzw. haplotyp minimalny) umożliwiły wyodrębnienie w całym badanym materiale 111 różnych haplotypów.

W tabeli I przedstawiono częstość występowania poszczególnych alleli badanych loci STR w zależności od przynależności haplogrupowej chromosomów. Uwagę zwraca-



cją szczególnie znaczące różnice w częstości występowania w poszczególnych grupach chromosomów Y alleli 11 i 13 w lokus DYS385, alleli 12 i 28 w lokus DYS389 i alleli 22 i 25 w lokus DYS390. Allel 11 w lokus DYS385 występuje w co trzecim chromosomie nienależącym do haplogrupy I\*, podczas gdy w chromosomach należących do tej haplogrupy w ogóle nie stwierdzono jego występowania. Allel 12 lokus DYS389 występuje w połowie chromosomów należących do haplogrupy I\* oraz w zaledwie 8% chromosomów nienależących do tej haplogrupy. Podobnie allel 28 tego lokus ujawniono w co trzecim chromosomie z haplogrupy I\* i w co 20 spoza niej. Z kolei allel 22 w lokus DYS390 występuje w 30% próbek przypisanych do haplogrupy I\*, podczas gdy w próbkach spoza I\* allel ten występuje z częstością zaledwie około 0,02. Allel 25 lokus DYS390 występuje w połowie próbek spoza haplogrupy I\* i nie zanotowano jego obecności w żadnej z próbek należących do tej haplogrupy.

Analiza częstości występowania poszczególnych haplotypów w lokus DYS385 (tabela II) wykazała, że haplotyp 11,14, najczęstszy w badanej populacji, obecny jest w niemal 50% chromosomów nienależących do haplogrupy I\*, lecz nie odnotowano go w ogóle w grupie chromosomów należących do tej haplogrupy. Z kolei haplotypy lokus DYS385, w których krótszym allelem jest allel 13 lub 14 (13,14; 13,15; 14,14; 14,15 i 14,16), stosunkowo rzadko występujące w badanej populacji (sumaryczna częstość tych haplotypów w badanym zestawie próbek to ok. 17%), pojawiają się w grupie chromosomów z haplogrupy I\* z sumaryczną częstością 95%, czyli niemal każdy chromosom z tej haplogrupy posiada jeden z wymienionych haplotypów lokus DYS385, podczas gdy sumaryczna częstość występowania tych haplotypów w chromosomach spoza haplogrupy I\* wynosi nieco ponad 3%.

Haplotypy 12-28, 12-29 i 13-31 lokus DYS389I&II występują w chromosomach haplogrupy I\* znacznie częściej niż w chromosomach nienależących do tej haplogrupy. Skrajnym przykładem ilustrującym preferencję w występowaniu niektórych haplotypów w lokus DYS389I&II w związku z przynależnością bądź nieprzynależnością chromosomu do haplogrupy I\* jest kombinacja alleli 13-29. Mimo, że częstość występowania tego haplotypu w badanej populacji jest znaczna i przekracza 0,25, nie zanotowano jakiegokolwiek przypadku jego wystąpienia w chromosomach z haplogrupy I\* (tabela III).

Analiza częstości występowania kompletnych haplotypów minimalnych (tabela IV i rycina 1) wykazała, że haplotyp – najczęstszy w grupie chromosomów należących do haplogrupy I\* (oznaczany jako haplotyp modalny I\*-HMI\*) – w ogóle nie wystąpił w grupie chromosomów spoza tej haplogrupy. W populacjach europejskich haplotyp ten występuje ze średnią częstością ok. 0,0015 (34 razy na 22 093 próbek w bazie YHRD), przy czym zaobserwowano go, jak dotąd, wyłącznie w populacjach Europy centralnej i wschodniej. Ponadto stwierdzono, że najczęstszy haplotyp w populacji ogólnej Pomorza i Kujaw (oznaczony symbolem nHMI\*), a występujący z częstością ok. 0,05, nie wystąpił w ogóle w tej populacji wśród chromosomów należących do haplogrupy I\* (rycina 1).

Analiza częstości występowania w innych populacjach z terenu Polski haplotypów minimalnych najczęściej spotykanych w populacji Pomorza i Kujaw w haplogrupie I\* (HMI\*) i poza haplogrupą I\* (nHMI\*) wykazała, że haplotyp HMI\* jest we wszystkich porównywanych populacjach stosunkowo rzadki. Dotyczy to zwłaszcza regionów Białegostoku i Warszawy (rycina 2). Częstość występowania haplotypu HMI\* w badanej populacji znacznie różni się od częstości rejestrowanej w popula-

cjach ze wspomnianych regionów. Wykazano istnienie statystycznie istotnych różnic w częstości haplotypu HMI\* pomiędzy populacją Bydgoszczy a populacją Białegostoku i Warszawy, między populacją Krakowa a populacjami Białegostoku, Gdańska i Warszawy oraz pomiędzy populacją Wrocławia a populacją Warszawy (tabela V).

Wyniki porównań częstości haplotypu minimalnego chromosomów spoza haplogrupy I\* najczęściej występującego w okolicach Bydgoszczy (nHMI\*), nie wskazują zasadniczo na istnienie statystycznie istotnych różnic w występowaniu tego haplotypu w różnych rejonach Polski. Wyjątkiem w tym względzie jest jednak populacja Bydgoszczy, w której występowanie haplotypu nHMI\* zdaje się być znacząco częstsze niż w populacjach Białegostoku i Warszawy.

#### DYSKUSJA

Charakterystyczny dla chromosomu Y jest fakt, iż na około 60 milionów p.z. wchodzących w jego skład, na obszarze ok. 50 milionów p.z. praktycznie nie występują przypadki rekombinacji [4]. Mutacyjne zmiany w sekwencji zachodzące w obrębie wspomnianej nierekombinującej części chromosomu Y (określanej w skrócie NRY – ang. non recombinig Y chromosome) dziedziczą się zatem wspólnie (tak jakby były jednostką genetyczną), niezależnie od tego, jaki dystans fizyczny dzieli poszczególne pozycje. Biorąc pod uwagę ten fakt, możliwe jest zatem rozpatrzenie hipotezy zakładającej, że wybrane mutacje w *loci* SNP i STR chromosomu Y powstałe w różnym czasie w chromosomach danej linii ojcowskiej, dziedziczą się razem, a także rozważenie potencjalnych zastosowań takiego zjawiska dla potrzeb medycyny sądowej.

Klasyczne metody analizy filogenetycznej opierają się często na założeniu, iż każda mutacja będąca markerem diagnostycznym wyznaczającym podział na poszczególne klady (grupy) jest zdarzeniem unikalnym w ewolucyjnej historii danego gatunku [15]. Założenie takie może być prawdziwe dla markerów o niskim polimorfizmie, takich jak SNP [17], jednak dla markerów wysoce polimorficznych, jakimi są *loci* STR, prawdopodobieństwo wielokrotnego powtórzenia mutacji prowadzącej do powstania danego allele jest wyższe [8]. Jest więc prawdopodobne, że chromosomy Y należące do różnych haplogrup mogą w wyniku mutacji równoległych posiadać identyczne allele w *loci* STR.

Unifikacja nomenklatury *loci* Y-STR dokonana przez Y Chromosome Consortium pozwoliła na stworzenie uniwersalnego systemu umożliwiającego przypisanie każdego badanego chromosomu Y do jednej ze 153 haplogrup na podstawie wyników analizy kilku markerów [21]. Nomenklatura *loci* Y-STR nie nastęrcza tyłu trudności ze względu na powszechnie przyjęty system ich oznaczania, zgodnie z wytycznymi Międzynarodowego Stowarzyszenia Genetyki Sądowej (ISFG) [7]. Istniejąca od niedawna unifikacja nazewnictwa haplogrup i haplotypów chromosomu Y pozwala na porównywanie właściwości genetycznych chromosomów Y pochodzących z różnorodnych populacji oddalonych od siebie geograficznie i badanych przez różne laboratoria, co do niedawna było utrudnione ze względu na – przede wszystkim – zróżnicowane systemy nazewnictwa haplogrup chromosomu Y [21]. Na podstawie danych zawartych w literaturze przedmiotu można np. stwierdzić, iż w populacji Chorwacji niemal wszystkie chromosomy haplogrupy I\* posiadają podobny haplotyp Y-STR. Jest on określany przez autorów jako dynarski haplotyp modalny (DMH) [2].

Haplotyp ten, zapisywany jako 16-24-11-11-13 (wg kolejności *loci* DYS19-390-391-392-393) oraz bezpośrednie pochodne tego haplotypu (różniące się od niego mutacjami w postaci obecności bądź nieobecności pojedynczych powtórzeń jednostek tworzących sekwencje mikrosatelitarne chromosomu Y) występują również wśród chromosomów Y haplogrupy I\* populacji Pomorza i Kujaw, jednak ich związek z przynależnością chromosomu do tej haplogrupy nie jest tak wyraźny.

Wydaje się, że wyniki analiz przeprowadzonych w ramach niniejszej pracy wskazują na istnienie współzależności pomiędzy przynależnością chromosomu Y do określonej haplogrupy, definiowanej przez polimorfizm Y-SNP, a występowaniem na tym chromosomie konkretnych alleli i zestawów alleli *loci* Y-STR. W szczególności obecność w lokus DYS385 haplotypów, których krótszym allelem jest allel 13 lub 14 (tabela II), występowanie w lokus DYS389I&II zestawów 12-28 i 12-29 (tabela III) i występowanie allela 22 w lokus DYS390 (tabela I) można uznać za wskazówkę świadczącą o przynależności danego chromosomu do haplogrupy I\*, podczas gdy obecność w lokus DYS385 zestawu alleli 11,14 i obecność allela 25 w lokus DYS390 wskazuje na przynależność danego chromosomu do haplogrupy innej niż I\*. Zidentyfikowano również haplotyp minimalny, który w obrębie badanej populacji jest silnie związany z chromosomami haplogrupy I\*. Według informacji zawartych w bazie danych YHRD występuje on w populacjach Europy z częstością ok. 0,0015.

Haplogrupa I\*, scharakteryzowana i opisana po raz pierwszy pod nazwą Eu7 przez Semino i in. [20], jest po Eu19 (R1a1\*) drugą pod względem częstości haplogrupą spotykaną w polskiej populacji. Cytowani autorzy szacują jej częstość na ok. 0,236, co oznaczałoby, iż haplogrupa I\* byłaby obecna w polskiej populacji u prawie 25% mężczyzn. Badania przeprowadzone dla populacji bydgoskiej (kujawsko-pomorskiej) wykazały, że częstość tej haplogrupy wynosi ok. 0,1493. Jest to wprawdzie wartość znacznie niższa od skalkulowanej przez Semino i in. [20], jednakże po uwzględnieniu ilości chromosomów przebadanych przez wspomnianych autorów ( $n = 55$ ) i znacznie liczniejszego materiału analizowanego w niniejszej pracy ( $n = 134$ ), obserwowane różnice częstości okazują się nieistotne statystycznie ( $p = 0.1554$ ). Porównania częstości charakterystycznego dla populacji kujawsko-pomorskiej haplotypu modalnego haplogrupy I\* (HMI\*) z częstością jego występowania w innych polskich populacjach, oparte na danych bazy YHRD, pozwalają na wykazanie istotnych statystycznie różnic pomiędzy populacją kujawsko-pomorską a populacją Białegostoku, Krakowa i Warszawy (tabela V). Wynik taki może sugerować zróżnicowanie częstości haplogrupy I\* w badanych populacjach, jednak może on również wynikać z braku ścisłego związku pomiędzy przynależnością chromosomu do haplogrupy I\* a występowaniem haplotypu HMI\*. Ze względu na niską częstość haplotypu HMI\* we wszystkich badanych populacjach prawdopodobna wydaje się również hipoteza, że obserwowane rozbieżności są artefaktami wywołanymi przez niedostateczną reprezentację haplotypów w badanych populacjach.

Biorąc pod uwagę badania wskazujące na homogenność polskiej populacji pod względem częstości występowania poszczególnych alleli i haplotypów *loci* STR [16] oraz sugerowany przez wyniki zaprezentowanych obliczeń silny związek poszczególnych alleli Y-STR, w szczególności alleli *loci* DYS385, DYS389 i DYS390 z przynależnością haplogrupową, możliwe jest podjęcie próby przybliżonego oszacowania częstości haplogrupy I\* w populacjach Polski innych niż kujawsko-pomorska. Dysponując danymi na temat rzeczywistej częstości tej haplogrupy w badanych popula-

cyjach i porównując te dane z wyliczonymi wartościami teoretycznymi, możliwe stałoby się przetestowanie prawdziwości hipotezy zakładającej związek poszczególnych alleli/haplotypów Y-STR z chromosomami haplogrupy I\*. Zakładając homogenność populacji Polski, częstość alleli/haplotypów Y-STR związanych z haplogrupą I\* powinna być stała wśród chromosomów tej haplogrupy. Przyjmując to założenie, można obliczyć oczekiwaną, teoretyczną częstość chromosomów należących do haplogrupy I\* dla danej populacji ze wzoru:

$$phgX = paX \times phgW / paW \quad \{1\}$$

gdzie:  $phgX$  – teoretyczna częstość chromosomów danej haplogrupy w populacji X;  $paX$  – częstość danego allelu/haplotypu lub ich grupy danego lokus w populacji X;  $phgW$  – częstość danej haplogrupy w populacji wzorcowej;  $paW$  – częstość allela/haplotypu lub ich grupy dla danego lokus w populacji wzorcowej.

Obliczone na podstawie powyższych założeń teoretyczne częstości chromosomów haplogrupy I\* w populacjach Polski nie różnią się znacząco pomiędzy populacją Bydgoszczy a populacjami Białegostoku, Gdańska, Krakowa, Lublina, Warszawy i Wrocławia (tabela VI). Biorąc pod uwagę cytowane wcześniej dane na temat homogenności populacji Polski pod względem częstości alleli *loci* Y-STR, należało oczekiwać wystąpienia podobnej homogenności układów SNP definiujących przynależność haplogrupową. Można zatem przypuszczać, że otrzymane wyniki obliczeń teoretycznej częstości chromosomów haplogrupy I\* w populacjach Polski potwierdzają możliwość określenia częstości dowolnej haplogrupy chromosomu Y na podstawie częstości alleli mikrosatelitarnych występujących na tym chromosomie pod warunkiem, że istnieją dla danej haplogrupy charakterystyczne allele Y-STR związane z jej występowaniem. Koniecznym warunkiem powodzenia takiej analizy wydaje się również bliskie pokrewieństwo genetyczne pomiędzy populacją badaną a populacją wzorcową, na podstawie której określano zależności występujące między częstościami poszczególnych haplogrup i haplotypów.

Przedstawiona powyżej analiza ma z konieczności charakter wstępu do dalszych badań nad zależnościami pomiędzy występowaniem na chromosomach Y konkretnych alleli Y-STR w powiązaniu z poszczególnymi haplogrupami. Zweryfikowanie istnienia faktycznego związku określonych haplotypów lub alleli Y-STR z przynależnością danego chromosomu do konkretnej haplogrupy wymaga zatem zgromadzenia danych na temat polimorfizmu *loci* SNP chromosomu Y w innych polskich populacjach. Biorąc pod uwagę fakt, że otrzymane wyniki badań dotyczą jednej haplogrupy i jednej populacji, nie można wykluczyć, że zaobserwowane zależności pomiędzy występowaniem określonych alleli *loci* Y-STR a przynależnością chromosomu do haplogrupy I\*, choć w niektórych przypadkach okazały się statystycznie istotne, mają swe źródło w zjawiskach stochastycznych mogących zaistnieć podczas typowania niewielkiej liczby chromosomów ( $n = 20$ ) zaliczonych do haplogrupy I\*.

Z punktu widzenia medycyny sądowej badania nad przynależnością haplogrupową i haplotypową chromosomów Y wydają się celowe, gdyż zdefiniowanie relacji zachodzących między allelami Y-SNP i Y-STR mogłoby posłużyć do opracowania systemu kontroli jakości wyników badań, podobnego do skutecznych systemów już stosowanych w badaniach polimorfizmu mtDNA [1]. System taki opierałby się na niezależnym typowaniu *loci* Y-STR i Y-SNP, a następnie weryfikacji otrzymanych wyników polegającej na sprawdzaniu zgodności oznaczonej haplogrupy z haplotypem

danej próbki. Zastosowanie takiej analizy mogłoby np. pozwolić w niektórych przypadkach na wykrycie mutacji zachodzących w miejscach wiązania starterów stosowanych do amplifikacji *loci* Y-STR. System taki mógłby również znaleźć zastosowanie w trudnych przypadkach dochodzenia spornego ojcostwa, np. gdy dostępny jest tylko DNA osób z dalszych gałęzi linii rodowej domniemanego ojca. W takich przypadkach analiza przynależności haplogrupowej w połączeniu z określeniem haplotypu i zbadaniem relacji pomiędzy wynikami tych badań mogłaby pozwolić na rozstrzygnięcie wątpliwości, czy obserwowane haplotypy rzeczywiście pochodzą z jednej linii ojcowskiej, czy też ich ewentualne podobieństwo wynika wyłącznie z wysokiej częstości mutacji *loci* Y-STR. Aby opisane wyżej badania zostały wprowadzone do rutynowej praktyki, konieczne jest jednak, oprócz dogłębnego zbadania zależności pomiędzy haplogrupami i haplotypami chromosomu Y, opracowanie prostej, wiarygodnej i stosunkowo taniej metody określania przynależności haplogrupowej chromosomów Y, gdyż stosowane obecnie powszechnie metody oparte na technice PCR-RFLP są zbyt drogie i czasochłonne.

Jeśli istnienie zależności pomiędzy występowaniem określonych alleli lub haplotypów *loci* Y-STR a przynależnością do określonej haplogrupy zostanie potwierdzone na podstawie wyników badań większej liczby próbek pochodzących ze zróżnicowanych genetycznie populacji, możliwy stanie się głębszy wgląd w ewolucyjną historię chromosomu Y, a co za tym idzie, lepsze wykorzystanie zebranej wiedzy również w zastosowaniach aplikacyjnych, w tym w hemogenetyce sądowej.