

POPULATION DIFFERENTIATION AND THE VALUE OF DNA EVIDENCE

Paulina WOLAŃSKA-NOWAK
Institute of Forensic Research, Cracow

ABSTRACT: The strength of any evidence can be assessed using a likelihood ratio (from Bayes' point of view). This is the ratio of the probabilities that the evidence would have been obtained given that the suspect is guilty and innocent, respectively. This, in turn, depends upon the probability that a match will be produced if the suspect is innocent. A key population genetics parameter is the "coancestry coefficient", or θ , or F_{ST} , which is the correlation between two genes sampled from distinct individuals within a subpopulation. In this paper θ coefficients for Polish population were calculated in four loci. The results suggest that values of θ appropriate to forensic applications are quite small in the Polish population, and the values of $\theta = 0.01$ and even more so $\theta = 0.03$ suggested by National Research Council are too conservative for the defendant.

KEY WORDS: DNA profiles; Identification; Statistics; Population genetics; Evidence value; Coancestry coefficient.

Z Zagadnień Nauk Sądowych, z. XXXIX, 1999, 103–121
Received 16 November 1998; accepted 2 March 1999

INTRODUCTION

The main aim of forensic haemogenetics is to establish if there is a connection between the DNA profile of evidence material from the crime scene and the DNA profile of the suspect or victim.

Assessing each matching allele is treated as a statistically independent event (one may compare it with drawing a single ball from an urn and then replacing it), and with such an assumption one can multiply the frequencies of occurrence of each of the alleles in order to calculate the frequency of occurrence of the complete multilocus profile. This method of profile calculation is known as the multiplication rule (product rule), and can be applied only when the events (i.e., the matches at each allele) are actually statistically independent.

Meanwhile, from the mathematical point of view, applying the multiplication rule to calculating the frequency of the DNA profile is analogous to estimating the proportion of blond, blue-eyed, fair-skinned individuals in European population by separately counting the frequencies of people with

blond hair, people with blue eyes, and people of fair skin and calculating their proportions. If a population survey of Europe showed that 1 out of 10 people had blond hair, 1 out of 10 had blue eyes, and 1 out of 10 had fair skin, one would be wrong to multiply these frequencies and conclude that the frequency of people in Europe with all three traits was 1 in 1000. We know that these traits are very popular in Northern Europe, so the chance of meeting an individual matching this description is probably higher than 1 in 1000. In other words, the multiplication rule can produce an underestimate in this case, because the traits are correlated owing to population substructure.

Although the effect of population differentiation on the value of DNA profiling evidence has been the subject of controversy for some years now, there are no answers to the questions that are most relevant to the forensic context. It was proved that because of genetic drift even large, uniform (with regard to race) populations differ in their genetic composition. In the case of using DNA profiles for forensic identification, genetic differentiation is potentially problematic: although a particular DNA profile may be rare overall, it may be more common in a subpopulation which contains both the true perpetrator of the crime and an innocent defendant.

The diagnostic value or strength of DNA evidence is inversely proportional to the frequency of occurrence of the DNA pattern in the general population. Hence, if it is possible that the defendant is innocent but has an ethnic background similar to that of the actual culprit, then forensic assessments that ignore population differentiation may overstate the value of the evidence. How then, bearing in mind the subdivision of population, can we correctly calculate the probability of occurrence of a DNA profile incriminating a perpetrator?

Evett, Balding and Nichols propose introducing the coancestry coefficient F_{ST} , or θ which is the probability that two alleles of the same locus have descended from the same ancestral allele. For all real populations the coancestry coefficient has the value $\theta > 0$; in an ideal population in Hardy-Weinberg equilibrium $\theta = 0$.

Assume that the suspect has the same DNA profile as the material secured at the crime scene. A simple way to estimate the value of such evidence is to assume conditions of the Hardy-Weinberg equilibrium for the ideal population and calculate the frequency of heterozygote occurrence according to the equation $2p_i p_j$, where p_i, p_j are the frequencies of particular alleles in the population. However real human populations do not fulfil these assumptions. To be more precise: let's denote the DNA profile of the offender which is recovered from the crime scene as X^S , and the DNA profile of the suspect as X^P . Then, if the profiles match and X denotes the shared profile, the likelihood ratio representing the strength of evidence in support of the hypothesis that the suspect and offender are the same person is given by

$$LR(X) = \frac{1}{p(X^S = X | X^P = X, P \neq S)}. \quad \{1\}$$

The denominator of the likelihood ratio is referred to as the conditional match probability and represents how likely it is that an unknown true offender exhibits the profile X , given that the suspect does. Thus, the match probability will vary according to how the suspect and the offender are genetically related, when they are not the same people. In this situation one have to consider two cases:

1. the offender belongs to the same subpopulation of the suspect's racial group;
2. the offender belongs to another subpopulation in the same or different racial group as the suspect.

It is assumed that subpopulation is a group of persons loosely related to the suspect who are more likely to carry copies of his/her alleles than other members of the general population.

The match probability of X under the first scenario may be computed using a formula derived for a single-locus genotype by Balding and Nichols [3]:

$$\text{for homozygote } (A_i A_i) \quad \frac{[2\theta + (1-\theta)p_i][3\theta + (1-\theta)p_i]}{(1+\theta)(1+2\theta)}; \quad \{2\}$$

$$\text{for heterozygote } (A_i A_j) \quad \frac{2[\theta + (1-\theta)p_i][\theta + (1-\theta)p_j]}{(1+\theta)(1+2\theta)},$$

where $i \neq j$.

The above equations for the match probability of the profiles are expressed in terms of the subpopulation coancestry coefficient and allele proportions: p_i and p_j exhibited in the suspect's racial group. The conditional match probability for the full profile X is then obtained by multiplying terms such as {2} across all loci.

In the second scenario, where the offender and the suspect are effectively unrelated, the conditional match probability is reduced to the frequency of occurrence of profile X in the racial group in question and may be computed using the product rule; i.e. by effectively multiplying together component allele proportions. This is equivalent to substituting $\theta = 0$ in {2} for all loci.

One issue which may be important when presenting DNA evidence in court relates to the idea of conservativeness; i.e. do the methods of LR calculation tend to yield values which err in favour of the suspect? One can show that the match probability expression in {2} is an increasing function of θ given p_i , resulting in more conservative values for larger θ . Therefore, assuming that the suspect and offender belong to the same subpopulation

(where $\theta > 0$ and {2} is used) is more conservative than assuming they belong to a different subpopulation (where $\theta = 0$ and the product rule is implemented), since the former will yield larger match probabilities and, hence, lower LR values. Thus, excluding special cases, where the suspect and offender are close blood relatives, which will usually be dealt with separately, it is common practice at the European and Australian laboratories to calculate the LR under scenario 1 for the offender's racial group when known; otherwise, the minimum value across all racial groups is reported.

It was suggested in the recent National Research Council report that routine adoption of $\theta = 0.01$ would be appropriate when evaluating STR evidence, although a provisional and more conservative value of 0.03 might be chosen until the completion of more extensive studies on population differentiation.

The recent studies of Balding and Nichols [2] and Foreman [11] indicate that respective values of coefficients of genetic correlations are too large, to be ignored in forensic practice. Budowle [5] and Gill [15] represent the opposite opinion, whilst some, like Foreman, have changed their minds, over time [10].

Hence, it seemed interesting to find answers to the following questions:

1. How does the value of coefficient θ , including "conservative" values suggested by National Research Council, influence the values of DNA evidence?
2. What is the mean value of the coancestry coefficients in the Polish population in systems used in forensic DNA typing?
3. To what extent does applying mean real coancestry coefficients lower the value of DNA evidence?

MATERIAL AND METHODS

To illustrate the influence of coancestry coefficient θ on the change of genotype frequency, and thus on the likelihood ratios, simulation calculations for an abstract genetic system, including three alleles in one locus, were performed. The frequencies of alleles were assumed to be 0.01; 0.09 and 0.9; this gives all possible genotype frequencies – respectively: 0.0001; 0.0018; 0.0081; 0.018; 0.162, and 0.81. A broad range of values permits a test of the influence of the coancestry coefficient on low and high genotype frequencies. The distributions of allele frequencies for locus DQA1 [6, 15, 21, 28, 30], D1S80 [4, 5, 15, 18, 25, 29], HUMvWA and HUMTH01 [4, 12, 16, 17, 19, 20, 23, 24, 26, 27] in different populations were collected from the literature.

Sewall Wright introduced the term "F-statistics" in 1943. According to it θ , or F_{ST} , can be interpreted as a measure of the differentiation between the

suspect's subpopulation and the database population which is being used to estimate the allele frequency in the matching profile. Using Weir's terminology [31], for a set of r populations with sample allele frequencies p_i ($i = 1, 2, \dots, r$) for an allele A , the statistic could be defined as:

$$F_{ST} = \frac{\sum_i (\tilde{p}_i - \bar{p})^2 / (r-1)}{\bar{p}(1-\bar{p})} = \frac{s^2}{\bar{p}(1-\bar{p})}, \quad \{3\}$$

where $\bar{p} = \sum_i \frac{\tilde{p}_i}{r}$ is the average sample frequency of the allele in the samples and s^2 is the sample variance. Equal allele frequencies in all populations will cause θ to be zero and for a pair of populations θ may serve as a measure of genetic distance. Under the pure drift model, $D = -\ln(1-\theta)$, will increase linearly with time because of the divergence of the two populations from the ancestral population.

Calculations of coefficients were performed with the TFGA program available through the Internet (<http://herb.bio.nau.edu/~miller/tfpga.htm>). Evaluations of genotype frequencies according to formulas of Balding and Nichols [2] and likelihood ratios assuming different values of θ , and computations of θ according to the formula published in the National Research Council Report in 1996 were performed using "home-made" simple computer programs.

RESULTS

In the paper simulation calculations were performed on the influence of coancestry coefficients on the likelihood ratios for different genotype frequencies. The results are listed in Table I.

Mean values of θ were calculated on the basis of allele frequencies available from databases cited in references.

In the DQA1 system: $\theta = 0.0044$ for populations living around Warsaw, Kraków, Wrocław, Bydgoszcz and Katowice; $\theta = 0.0016$ for populations around Warsaw, Kraków, Wrocław, Katowice, Switzerland and Hungary.

In the STR TH01 system: $\theta = 0.0003$ for populations living around Gdańsk, Kraków, Lublin; $\theta = 0.0014$ for populations of Gdańsk, Spanish-Galicia, Hungary, South Spain, Denmark, Switzerland, Austria and Germany.

TABLE I. CHANGES IN THE VALUE OF THE LIKELIHOOD RATIO DEPENDING ON THE VALUE OF THE COANCESTRY COEFFICIENT FOR DIFFERENT GENOTYPE FREQUENCIES.

Genotype frequency	Value of likelihood ratio assuming different value of θ						
	0.0	0.0005	0.003	0.005	0.01	0.03	0.05
0.0001	10000	7924	3330	863.5	863.5	66.1	66.1
0.0018	555.5	527	420	261.2	261.2	71.6	71.6
0.0081	123.4	120	107	79.3	79.3	26.4	26.4
0.018	55.5	53	43	28.7	28.7	10.7	10.7
0.162	6.2	6	6	5.7	5.7	4.7	4.7
0.81	1.2	1.2	1.2	1.2	1.2	1.2	1.2
$1.12 \cdot 10^{-34}$ *	$8.9 \cdot 10^{33}$	$1.7 \cdot 10^{33}$	$2.9 \cdot 10^{30}$	$6.3 \cdot 10^{28}$	$6.0 \cdot 10^{25}$	$3.0 \cdot 10^{19}$	$1.6 \cdot 10^{16}$
$0.34 \cdot 10^{-10}$ **	$2.95 \cdot 10^{10}$	$2.7 \cdot 10^{10}$	$2.0 \cdot 10^{10}$	$1.5 \cdot 10^{10}$	$8.3 \cdot 10^9$	$1.0 \cdot 10^9$	$0.16 \cdot 10^9$

*The minimal combined multilocus frequency which can be obtained with the Profiler Plus kit.

**The maximal combined multilocus frequency which can be obtained with the Profiler Plus kit.

TABLE II. RESULTS OF THE ASSESSMENT OF LIKELIHOOD RATIOS FOR CHOSEN GENOTYPES CALCULATED UNDER DIFFERENT ASSUMPTIONS ABOUT THE VALUE OF COANCESTRY COEFFICIENTS θ .

Genotype				Likelihood ratio (LR)					
				Resulting from product rule		Suggested by NRC: $\theta = 0.01$		Assessed on the basis of calculated values of θ	
D1S80	TH01	DQA1	vWA	LR $\theta = 0$	log LR	LR $\theta = 0.01$	log LR	LR $\theta = 0.0021;$ $\theta = 0.0003;$ $\theta = 0.0044;$ $\theta = 0.0015$	log LR
14/40	5/10	1.3/1.1	13/13	$3.1 \cdot 10^{14}$	14.5	$1.6 \cdot 10^{11}$	11.2	$0.4 \cdot 10^{14}$	13.6
20/26	7/8	2/3	14/19	$6.0 \cdot 10^7$	7.8	$2.1 \cdot 10^7$	7.3	$4.7 \cdot 10^7$	7.7
18/24	6/9.3	1.2/4	17/18	$2.5 \cdot 10^3$	3.4	$2.3 \cdot 10^3$	3.4	$2.5 \cdot 10^3$	3.4
14/40	7/8	1.2/4	15/19	$5.2 \cdot 10^8$	8.7	$3.1 \cdot 10^7$	7.5	$2.1 \cdot 10^8$	8.3
18/24	7/8	3/1.1	16/20	$1.7 \cdot 10^6$	6.2	$0.9 \cdot 10^6$	5.9	$1.4 \cdot 10^6$	6.1
14/40	5/10	2/3	13/14	$6.0 \cdot 10^{12}$	12.8	$3.2 \cdot 10^{10}$	10.5	$1.8 \cdot 10^{12}$	12.3
14/40	6/9.3	1.2/4	19/20	$3.7 \cdot 10^8$	8.6	$1.9 \cdot 10^7$	7.3	$1.4 \cdot 10^8$	8.1
20/26	5/10	1.2/4	15/16	$5.8 \cdot 10^8$	8.8	$6.0 \cdot 10^7$	7.8	$4.4 \cdot 10^8$	8.6
14/40	6/9.3	1.3/1.1	18/19	$2.9 \cdot 10^8$	8.5	$2.0 \cdot 10^7$	7.3	$1.3 \cdot 10^8$	8.1
18/24	5/10	2/4	13/20	$8.3 \cdot 10^8$	8.9	$4.5 \cdot 10^7$	7.7	$5.4 \cdot 10^8$	8.7
20/26	5/10	1.2/4	13/13	$7.5 \cdot 10^{11}$	11.9	$2.9 \cdot 10^9$	9.5	$2.0 \cdot 10^{11}$	11.3
14/14	5/5	1.3/1.3	13/13	$2.0 \cdot 10^{16}$	16.3	$1.8 \cdot 10^{11}$	11.2	$7.4 \cdot 10^{14}$	14.9
14/14	7/8	2/3	13/13	$2.6 \cdot 10^{12}$	12.4	$1.3 \cdot 10^9$	9.1	$1.6 \cdot 10^{11}$	11.2
23/19	5/10	1.3/1.1	13/14	$2.3 \cdot 10^{12}$	12.4	$3.3 \cdot 10^{10}$	10.5	$1.0 \cdot 10^{12}$	12.0
14/19	5/5	1.3/1.3	13/14	$2.1 \cdot 10^{14}$	14.3	$1.2 \cdot 10^{11}$	11.1	$4.8 \cdot 10^{13}$	13.7
14/14	5/5	1.3/1.1	13/13	$7.8 \cdot 10^{15}$	15.9	$1.8 \cdot 10^{11}$	11.2	$1.6 \cdot 10^{14}$	14.2
14/40	5/10	2/3	13/13	$1.4 \cdot 10^{14}$	14.1	$8.1 \cdot 10^{10}$	10.9	$8.1 \cdot 10^{13}$	13.2

In D1S80 system: $\theta = 0.0021$ for populations of Kraków and Lublin; $\theta = 0.0050$ for populations of Gdańsk, Kraków, Lublin, Slovenia and Holland.

In the vWA system: $\theta = 0.0015$ for the population of Gdańsk and Lublin; $\theta = 0.0008$ for the populations of Austria, Germany, Holland, Switzerland, Gdańsk, Spain (Galicia) and Hungary.

Making use of calculated values of θ coefficients for systems applied in forensic DNA typing, calculations of frequency of selected genotypes were carried out in the three following variants:

1. allele frequencies were drawn from the South Polish database [29, 30], assuming $\theta = 0$ (equivalent to ignoring genetic correlations);
2. allele frequencies were drawn from the South Polish database [29, 30], assuming $\theta = 0.01$ as recommended by the National Research Council;
3. allele frequencies were drawn from the South Polish database [29, 30] and implemented values of θ calculated for all Polish populations for each genetic system respectively.

Results are summarised in Table II.

DISCUSSION

The results of simulation calculations, which are shown in Figure 1, indicate that, along with an increase in the value of the coancestry coefficient θ due to the increasing extent of relatedness among the members of a popula-

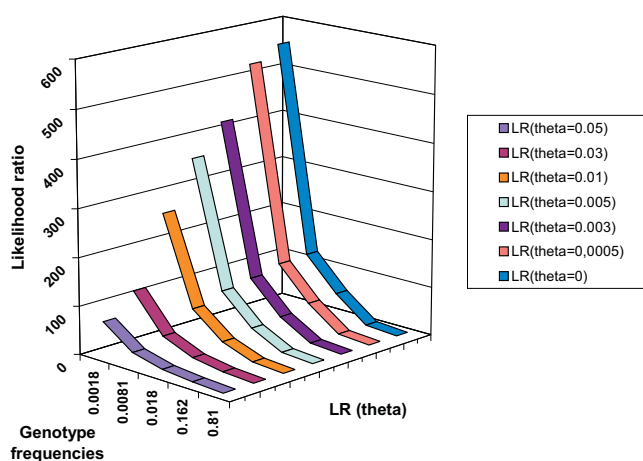


Fig. 1. Plot of the value of likelihood ratio affected by different values of coancestry coefficients and genotype frequencies.

tion, the match probabilities also grow which is equivalent to a decrease in likelihood ratio values. This tendency is expressed more in the case of rare genotypes, which is particularly visible in the case of minimal genotype frequency obtained with Profiler Plus kit (Figure 1b). After implementing $\theta = 0.01$ the likelihood ratio is lowered by about eight orders of magnitude,

and when $\theta = 0.01$ it is lowered by fourteen orders! Analogous calculations performed for the maximum genotype frequency obtainable with the Profiler Plus kit show the slight “sensitivity” of genotype frequencies (of the order 10^{-10}) even at high values of the coefficient θ .

Fig. 1B. Illustration of the sharp decay of evidence value in concordance with growing θ for a minimum combined multilocus genotype frequency obtained with Profiler Plus – $1.12 \cdot 10^{-34}$. Analogical calculations were performed on maximal combined multilocus genotype frequency obtained with Profiler Plus – $3.4 \cdot 10^{-9}$.

Population geneticists have extensively studied the magnitude of genetic differentiation between human populations, because it provides information about human demographic history, but conventionally analysed characters are coded by genes demonstrating a low level of polymorphism and mutation. Hence, when selection is weak, the expected genetic differentiation

is the same at each locus. The loci used in forensic science are highly polymorphic, and this often seems to be associated with higher mutation rates. Furthermore, they are typically found in introns of genes that may possibly be subject to geographically varying selection. High mutation rates and some kind of selection can produce differences in coefficient θ across particular loci, so in this paper the coancestry coefficients were calculated separately for each locus.

Performed calculations of genetic correlations for the Polish population indicate a rather low level of differentiation in the range of studied loci. Coancestry coefficients for D1S80, TH01 and vWA systems have very low values, which is expressed in their weak influence on the values of genotype frequencies, and consequently, their weak influence on the value of DNA evidence. As is shown in Figures 2 and 3, for genotype frequency at the 10^{-10} level one can observe a decline in likelihood ratio of about ten, and for genotype frequency of 10^{-16} – about a hundred.

Fig. 2. Comparison of likelihood ratios for chosen profiles (see Tab. II) for different values of θ used for calculation.

Fig. 3. Illustration of influence of the value $\theta = 0.0042$ calculated for Southern Poland population on likelihood ratios of all possible genotypes in D1S80 system.

In contrast to the above values the relatively high value of the coancestry coefficient in the DQA1 locus in the Polish population ($\theta = 0.0044$) results from a difficult-to-explain peculiarity of the population living in the area around Bydgoszcz [23]. If one takes into consideration the populations of Switzerland and Hungary, whilst omitting those of the Bydgoszcz area, the resulting coancestry coefficient is much lower ($\theta = 0.0016$) and seems to be more rational.

The Table III represents comparison of coancestry coefficients θ calculated in this paper for the Polish population with data for other Caucasian populations. It can be seen that these values do not vary greatly across the loci.

It was suggested in the recent National Research Council report [24] that routine adoption of $\theta = 0.01$ would be appropriate when evaluating STR evidence, although a provisional and more conservative value of $\theta = 0.03$ might be chosen until the completion of more extensive studies on such PCR-based systems. For small, isolated populations, a value of 0.03 would be adequate.

In the USA the values of θ , averaged across all forensically studied loci, calculated on the base of FBI database are markedly lower than 0.01; for Caucasian – 0.002, Black – 0.007 and Hispanic – 0.009 [5, 6].

TABLE III. COMPARISON OF COANCESTRY COEFFICIENTS CALCULATED FOR DIFFERENT CAUCASIAN POPULATIONS.

Source				D1S80
Foreman et al. [11]	0.0125–0.0051	0.0135–0.0053		
Evetts et al. [9]	–0.001–0.01	0.001–0.002		
Budowle [5]			0.0040–0.0089	
Polish population	0.0003	0.0015	0.0044	0.021

From the calculations performed it appears that applying values of coancestry coefficients of $\theta = 0.01$ to genotype frequencies of about 10^{-8} causes a 10-fold decline in the likelihood ratio, and to genotype frequency of 10^{-16} causes a decline of LR of about four orders of magnitude. The real values of coancestry coefficients both for Poland and other Caucasian populations are markedly lower than 0.01. Hence, applying the value of $\theta = 0.01/0.03$ for routine casework by Forensic Science Service is rather strange [9].

Figure 3 illustrates the application of the coefficient $\theta = 0.0042$ (calculated for locus D1S80) in the South Polish population in the context of the general Polish population.

Deviation from relation when $\theta = 0$ is observed only for genotype frequency of about 10^{-3} and lower. Calculations carried out indicate that it is necessary to take into account genetic differentiation in the use of DNA profiles for forensic identification. Furthermore if there is a suspicion that the defendant is innocent but has an ethnic background similar to that of the true perpetrator, then forensic assessments which ignore population differentiation may overstate the strength of the evidence.

One ought to remember that when using, in routine forensic practice, about nine STR systems for individualisation of biological traces (e.g. the Profiler Plus kit commonly used in the Forensic Hemogenetics Section of the Institute of Forensic Research in Cracow) we can obtain mean values of combined multilocus genotype frequencies of about 10^{-12} . These values are just in the range where even low values of θ result in the reduction of the evidential value of matching DNA profiles. Therefore, it seems necessary to perform a precise population study to assess the real distribution of allele frequencies throughout the people living in a small or partially isolated area. The study could detect a subtle genetic differentiation of the Polish popula-

tion. The assumption that $\theta = 0$, which is still commonly applied, is wrong, because it ignores the laws of population genetics and has the effect of exaggerating the strength of the evidence against the defendant. However, implementing arbitrary values of $\theta \geq 0.01$ does not seem to be justified and is too "conservative" with respect to the defendant.

References:

1. Balding D. J., Donnelly P., Inference in forensic identification, *Journal of the Royal Statistical Society A* 1995, vol. 158, part 1, pp. 21–53.
2. Balding D. J., Nichols R. A., Significant genetic correlations among Caucasians at forensic DNA loci, *Heredity* 1997, vol. 78, pp. 583–589.
3. Balding D. J., Greenhalgh M., Nichols R. A., Population genetics of STR loci in Caucasians, *International Journal of Legal Medicine* 1996, vol. 108, pp. 300–305.
4. Balding D. J., Nichols R. A., DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, *Forensic Science International* 1994, vol. 64, pp. 125–140.
5. Budowle B., The effects of inbreeding on DNA profile frequency estimates using PCR-based loci, [in:] Weir B. S. [ed.], *Human identification: the use of DNA markers*, Kluwer Academic Publishers, Netherlands 1995, pp. 21–25.
6. Budowle B., Woller J., Koons B. [et al.], Hungarian population data on seven PCR-based loci, *Journal of Forensic Sciences* 1996, vol. 41, pp. 667–670.
7. Ciesielka M., Koziol P., Krajka A., Allele frequency distributions of D1S80 in the Polish population, *Forensic Science International* 1996, vol. 81, pp. 141–147.
8. Dmochowska G., Częstość alleli i genotypów HLA DQ alfa w badanej próbie populacji polskiej, *Archiwum Medycyny sądowej i Kryminologii* 1994, t. XLIV, s. 405–407.
9. Evett I. W., Gill P. D., Lambert J. A. [et al.], Statistical analysis of data for three British ethnic groups from a new STR multiplex, *International Journal of Legal Medicine* 1997, vol. 110, pp. 5–9.
10. Evett I. W., Lambert J. A., Buckleton J. S. [et al.], Statistical analysis of a large file data from STR profiles of British Caucasians to support forensic casework, *International Journal of Legal Medicine* 1996, vol. 109, pp. 173–177.
11. Foreman L. A., Lambert J. A., Evett I. W., Regional genetic variation in Caucasians, *Forensic Science International* 1998, vol. 95, pp. 27–37.
12. Foreman L. A., Smith A. F. M., Evett I. W., A Bayesian approach to validating STR multiplex databases for use in forensic casework, *International Journal of Legal Medicine* 1997, vol. 110, pp. 244–250.
13. Foreman L.A., Smith A. F. M., Evett I. W., Bayesian analysis of DNA profiling data in forensic identification applications, *Journal of the Royal Statistical Society A* 1997, vol. 160, part 3, pp. 429–469.
14. Furedi S., Woller J., Padar Z., Hungarian population data for the STR systems TH01 and VWA, *International Journal of Legal Medicine* 1995, vol. 108, pp. 48–49.

15. Gallo J. C., Thomas E., Novick G. [et al.], Effects of subpopulation structure on probability calculations of DNA profiles from forensic PCR analysis, *Genetica* 1997, vol. 101, pp. 1–12.
16. Gill P., Evett I., Population genetics of short tandem repeat (STR) loci [in:] Weir B. S. [ed.], *Human identification: the use of DNA markers*, Kluwer Academic Publishers, Amsterdam 1995, pp. 69–87.
17. Hochmeister M., Budowle B., Borer U. [et al.], Swiss population data on the loci HLA-DQA1, LDLR, GYPA, HBG, D7S8, Gc and D1S80, *Forensic Science International* 1994, vol. 67, pp. 175–184.
18. Hochmeister M. N., Jung J. M., Budowle B. [et al.], Swiss population data on three tetrameric short tandem repeat loci – VWA, HUMTH01 and F13A1 – derived using multiplex PCR and laser fluorescence detection, *International Journal of Legal Medicine* 1994, vol. 107, pp. 34–36.
19. Klintschar M., Kubat M., A study of the short tandem repeat systems HUMVWA and HUMTH01 in an Austrian population sample, *International Journal of Legal Medicine* 1995, vol. 107, pp. 329–330.
20. Kloosterman A., Budowle B., Daselaar P., PCR-amplification and detection of the human D1S80 VNTR locus, *International Journal of Legal Medicine* 1993, vol. 105, pp. 257–264.
21. Martin P., Alonso A., Budowle B. [et al.], Spanish population data on & tetrameris short tandem repeat loci, *International Journal of Legal Medicine* 1995, vol. 108, pp. 145–149.
22. Monies D., Ciesielka M., Kozioł P. [i in.], Częstości alleli sześciu układów STR (CSF1PO, TPOX, TH01, F13A01, FESFPS, VWA) w populacji Polski południowo-wschodniej, VI Ogólnopolska Konferencja Naukowa „Błąd medyczny”, Książ k/Wałbrzycha, 22–25 maja 1996, Streszczenia, s. 30–31.
23. Mścicka-Śliwka D., Pleszyńska B., Polimorfizm lokus HLA-DQA1 w populacji regionu pomorsko-kujawskiego, *Postępy Medycyny Sądowej i Kryminologii* 1997, t. 3, s. 315–318.
24. National Research Council, *The evaluation of forensic DNA evidence*, National Academy Press, Washington 1996.
25. Pawłowski R., Allele frequencies of HUMTH01 in the Polish population, *Z Zagadnień Nauk Sądowych* 1996, z. XXXIV, s. 9–18.
26. Pawłowski R., Maciejewska A., Population genetics of the STR-HUMVW system in the population of Northern Poland, *Z Zagadnień Nauk Sądowych* 1996, z. XXXIV, s. 19–29.
27. Pawłowski R., Paszkowska R., Hauser R. [et al.], Population studies of three AMPFLPs systems in a North Polish population, *International Journal of Legal Medicine* 1996, vol. 109, pp. 155–156.
28. Pestoni C., Lareu M., Rodriguez M., The use of the STR's HUMTH01, HUMVWA31/A, HUMF13A1, HUMFES/FPS, HUMLPL in forensic application: validation studies and population data for Galicia (NW Spain), *International Journal of Legal Medicine* 1995, vol. 107, pp. 283–290.
29. Sjerps M., van der Geest N., Pieron C. [et al.], A Dutch population study of the STR loci HUMTH01, HUMFES/FPS, HUMVWA31/1 and HUMF13A1, conducted for forensic purposes, *International Journal of Legal Medicine* 1995, vol. 108, pp. 127–134.

30. Tucholska-Lenart A., Miąskiewicz H., Suszczewski W. [i in.], Badania częstotliwości występowania genotypów HLA locus DQA1 w populacji warszawskiej, *Problemy Kryminalistyki* 1994, nr 206, s. 14–16.
31. Turowska B., Sanak M., D1S80 VNTR locus genotypes in population of South Poland; meta-analysis pointer to genetic disequilibrium of human populations, *Forensic Science International* 1995, vol. 75, pp. 207–216.
32. Turowska B., Sanak M., Nowicka L. [i in.], Polimorfizm lokus HLA DQA1 w populacji Polski południowej, *Archiwum Medycyny Sądowej i Kryminologii* 1996, t. XLVI, s. 261–268.
33. Weir B. S., Genetic data analysis II, Sinauer Associates Inc., Sunderland 1996.

ZRÓŻNICOWANIE POPULACJI A WARTOŚĆ DOWODU Z BADANIA DNA

Paulina WOLAŃSKA-NOWAK

WSTĘP

Zasadniczym zadaniem hemogenetyka sądowego jest ustalenie, czy istnieje związek między profilem DNA materiału dowodowego zabezpieczonego na miejscu zdarzenia a profilem DNA podejrzanego bądź ofiary.

Scharakteryzowanie każdego zgodnego allela traktuje się jako zdarzenie statystycznie niezależne (można je porównać z losowaniem ze zwracaniem kul z urny), a przy takim założeniu wystarczy pomnożyć częstości występowania poszczególnych alleli w celu wyliczenia częstości występowania całego profilu. Wspomniany wyżej sposób postępowania określa się jako regułę mnożenia (ang. multiplication rule, product rule), którą można zastosować wyłącznie pod warunkiem, że zdarzenia, to jest zgodność każdego allela, są niezależne względem siebie.

Tymczasem z matematycznego punktu widzenia zastosowanie reguły mnożenia do obliczania częstości profilu DNA jest analogiczne do oszacowania proporcji osób o blond włosach, błękitnookich i o jasnej cerze w populacji europejskiej poprzez niezależne zliczenie ilości osób o jasnych włosach, błękitnookich i o jasnej cerze. Jeżeli więc weźmiemy pod uwagę ludność całej Europy i uzyskamy średni wynik, na przykład 1 osoba na 10 ma blond włosy, 1 osoba na 10 błękitne oczy i 1 osoba na 10 ma jasną cerę, to błędem byłoby pomnożenie tych częstości i wnioskowanie, że proporcja osobników w Europie posiadających łącznie te trzy cechy wynosi jak 1 do 1000. Wiemy, że te cechy są bardzo popularne wśród ludności zamieszkującej Europę północną i spotkanie tam osoby o powyższych cechach jest na pewno wyższe niż 1 do 1000. Innymi słowy, reguła mnożenia w tym wypadku prowadzi do zaniżenia częstości wspólnego występowania danych cech ze względu na istnienie zróżnicowania populacji.

Pomimo, iż wpływ zróżnicowania populacji na wartość dowodu z profilowania DNA był w ostatnich latach przedmiotem kontrowersji, nie znaleziono jednoznacznej odpowiedzi na najbardziej istotne w kontekście sądowym pytania. Udowodniono, iż w wyniku działania dryftu genetycznego nawet duże jednolite rasowo populacje ludzkie różnią się między sobą pulą genową. W przypadku stosowania profili DNA do identyfikacji sądowej, genetyczne zróżnicowanie populacji jest potencjalnym problemem; pojedynczy profil DNA może być bardzo rzadki, a jednak wśród subpopulacji, do której należy prawdziwy sprawca zbrodni i niewinny podejrzany, może okazać się bardziej powszechny.

Wartość diagnostyczna, inaczej moc dowodu z badania DNA, jest odwrotnie proporcjonalna do częstości występowania profilu DNA w danej populacji. A zatem, jeżeli jest możliwe, iż podejrzany jest niewinny, ale ma pochodzenie etniczne podobne do rzeczywistego sprawcy, to sądowe oszacowanie ignorujące zróżnicowanie populacyjne zawiąza wartość dowodu. Jak więc, uwzględniając problem zróżnicowania populacji, obliczać prawidłowo prawdopodobieństwo częstości występowania profilu DNA inkryminującego sprawcę?

Evett, Balding i Nicholls proponują wprowadzenie współczynnika odziedziczalności F_{ST} , inaczej θ (ang. coancestry coefficient), który jest prawdopodobieństwem, że dwa allele tego samego *locus* u dwóch przypadkowych osób mają to samo pochodzenie, czyli że zostały odziedziczone po wspólnym przodku. Dla każdej realnie istniejącej populacji współczynnik odziedziczalności $\theta > 0$; szczególnym przypadkiem jest populacja idealna spełniająca wymagania prawa Hardy’ego-Weinberga, kiedy $\theta = 0$.

Załóżmy, że podejrzany ma taki sam profil DNA, jak materiał zabezpieczony na miejscu zdarzenia. Prostem oszacowaniem wagi takiego dowodu jest obliczenie prawdopodobieństwa, że przypadkowo wybrana osoba może posiadać taki profil. Najłatwiejszym sposobem jest przyjęcie założeń Hardy’ego-Weinberga dla populacji idealnej i obliczenie częstości wystąpienia heterozygoty według wzoru $2p_i p_j$, gdzie p_i, p_j to częstości poszczególnych alleli występujących w populacji. Jednakże realnie istniejące populacje ludzkie nie spełniają tych założeń. Aby uściślić dalszy tok rozumowania, oznacmy przez X^S profil DNA sprawcy uzyskany ze śladu ujawnionego na miejscu zdarzenia (S), a przez X^P profil podejrzanego. Jeżeli te profile są identyczne i X oznacza wspólny profil, to iloraz wiarygodności (LR, ang. likelihood ratio) wyraża siłę dowodu na korzyść hipotezy, że podejrzany i sprawca są tą samą osobą. Przedstawia go wzór:

$$LR(X) = \frac{1}{p(X^S = X | X^P = X, P \neq S)}. \quad \{1\}$$

Mianownik ilorazu wiarygodności jest warunkowym prawdopodobieństwem zgodności (ang. match probability) opisującym, jak prawdopodobne jest posiadanie przez nieznanego prawdziwego sprawcę profilu X , jeśli wiadomo, iż podejrzany ma ten sam profil. Stąd prawdopodobieństwo zgodności będzie się zmieniać w zależności od tego, jak podejrzany i sprawca są genetycznie powiązani, gdy nie są tą samą osobą. W tej sytuacji należy więc rozważyć dwie możliwości:

1. sprawca należy do tej samej subpopulacji dużej rasowo grupy (populacji generalnej), co podejrzany;
2. sprawca należy do innej niż podejrzany subpopulacji tej samej lub innej dużej rasowo grupy.

Zakłada się, że subpopulacje to grupy osób luźno spokrewnionych między sobą i podejrzanym, pośród których prawdopodobieństwo posiadania wspólnych alleli jest większe niż pośród populacji generalnej.

Prawdopodobieństwo zgodności profilu X według scenariusza pierwszego można wyliczyć używając wzoru wyprowadzonego dla genotypu jednolokusowego przez Baldinga i Nicholsa [3]:

$$\begin{aligned} \text{dla homozygoty } (A_i A_i): & \frac{[2\theta + (1-\theta)p_i][3\theta + (1-\theta)p_i]}{(1+\theta)(1+2\theta)}; \\ & \{2\} \\ \text{dla heterozygoty } (A_i A_j): & \frac{2[\theta + (1-\theta)p_i][\theta + (1-\theta)p_j]}{(1+\theta)(1+2\theta)}, \text{ gdzie } i \neq j. \end{aligned}$$

Powyższe wzory na prawdopodobieństwo zgodności profili zawierają współczynnik θ i częstości alleli w populacji – w obrębie rasy, do której należy podejrzany, wyrażone jako proporcje p_i oraz p_j . A zatem warunkowe prawdopodobieństwo zgodności całego profilu X oblicza się poprzez przemnożenie wartości uzyskanych dla każdego *locus*.

Zakładając scenariusz drugi, kiedy sprawca i podejrzany z założenia nie są spokrewnieni, warunkowe prawdopodobieństwo zgodności redukuje się do częstości występowania profilu X w interesującej nas dużej rasowo populacji generalnej i może być wyliczone według reguły mnożenia, to jest poprzez mnożenie częstości występowania poszczególnych alleli. Jest to równoważne podstawieniu $\theta = 0$ w wyrażeniach {2} dla wszystkich *loci*.

W tym miejscu należy także zwrócić uwagę na problem, który może mieć istotne znaczenie podczas prezentowania dowodu w sądzie, a dotyczący idei „konserwatywności”, to jest stwierdzenia, czy metody wyliczenia ilorazu wiarygodności są korzystne dla podejrzanego. Okazuje się, że wyrażenie określające prawdopodobieństwo przypadkowej zgodności {2} jest wzrastającą funkcją θ w zależności od p_i , co daje w wyniku bardziej „konserwatywne” wartości dla większych wartości θ . Stąd założenie, że podejrzany i sprawca należą do tej samej subpopulacji (gdzie $\theta > 0$ i ma zastosowanie wzór {2}) jest bardziej „konserwatywne” niż założenie, że należą do różnych subpopulacji w obrębie gatunku ludzkiego (gdzie $\theta = 0$ i stosuje się regułę mnożenia). To pierwsze będzie bowiem dawało w wyniku większe prawdopodobieństwo zgodności i niższy iloraz wiarygodności (LR). Wyłączając zatem wyjątkowe przypadki, kiedy podejrzany i sprawca są bliskimi krewnymi (podobne wypadki należy rozważać oddzielnie), częstą praktyką w laboratoriach europejskich i australijskich jest obliczanie ilorazu wiarygodności według scenariusza pierwszego.

Ostatni raport National Research Council [23] sugeruje, aby w trakcie rutynowej pracy do określania wartości dowodu z badania DNA w *loci* typu STR stosować do obliczeń współczynnik $\theta = 0,01$, a nawet $\theta = 0,03$ do czasu dokładniejszego zbadania populacyjnego zróżnicowania gatunku ludzkiego w obrębie tych *loci*. Badania Baldinga i Nicholasa [2] oraz Foremana i in. [11] wskazują, że odpowiednie wartości współczynników korelacji genetycznej są zbyt wysokie, aby je ignorować w opiniowaniu sądowym. Przeciwnego zdania są Budowle [5] i Gill [15], a niektórzy z czasem zmieniają zdanie, jak Foreman [10].

Dlatego więc wydawało się interesujące znaleźć odpowiedź na następujące pytania:

1. Jak wartość współczynnika θ , w tym „konserwatywne” wartości sugerowane przez National Research Council, praktycznie wpływają na wartość dowodu z badania DNA?
2. Jaka średnią wartość mają te współczynniki dla populacji polskiej w układach genetycznych stosowanych sądowym typowaniu DNA?
3. W jakim stopniu zastosowanie uśrednionych realnych współczynników odziedziczalności obniża wartość dowodu?

MATERIAŁ I METODY

Dla zilustrowania wpływu wielkości współczynnika odziedziczalności θ na zmianę częstości genotypów, a tym samym na wartość ilorazu wiarygodności, przeprowa-

dzono obliczenia symulacyjne dla abstrakcyjnego, zawierającego 3 allele w jednym *locus* układu genetycznego, w którym przyjęto częstości występowania alleli jako: 0,01; 0,09 i 0,9 – co daje wszelkie możliwe częstości genotypowe, odpowiednio: 0,0001; 0,0018; 0,0081; 0,0180; 0,162; 0,8100. Duży rozrzut wartości pozwala sprawdzić wpływ współczynnika θ na niskie i wysokie częstości genotypowe.

Rozkłady częstości alleli dla *locus* DQA1 [6, 15, 21, 28, 30], D1S80 [4, 5, 15, 18, 25, 29], HUMVWA i HUMTH01 [4, 12, 16, 17, 19, 20, 23, 24, 26, 27] w różnych populacjach podaje literatura.

Termin „F-statystyka” wprowadził S. Wright w 1943 roku. W jej obrębie θ lub F_{ST} jest miarą wzajemnej zależności między allelami różnych osobników należących do tej samej subpopulacji w odniesieniu do populacji generalnej; używając terminologii Weira [31] dla szeregu r populacji z częstościami populacyjnymi p_i ($i = 1, 2, \dots, r$) dla pewnego allela A statystykę F_{ST} , inaczej θ , definiuje się jako:

$$F_{ST} = \frac{\sum_i (\tilde{p}_i - \bar{p})^2 / (r-1)}{\bar{p}(1-\bar{p})} = \frac{s^2}{\bar{p}(1-\bar{p})} \quad \{3\}$$

gdzie $\bar{p} = \sum_i \frac{\tilde{p}_i}{r}$ jest średnią frekwencją allela w próbach, a s^2 jest wariancją próby.

Co więcej, przyjmując model dryfu genetycznego, F_{ST} jest miarą dystansu genetycznego $D = -\ln(1-\theta)$, który jest proporcjonalny do czasu rozdzielania się porównywanych populacji.

Do obliczenia współczynników odziedziczalności korzystano z ogólnie dostępnego w sieci Internet (pod adresem <http://herb.bio.nau.edu/~miller/tfpga.htm>) programu komputerowego TFPGA. Częstości genotypowe według wzorów Baldinga i Nichollsa {2} i wartości ilorazu wiarygodności dla różnych θ oraz wartości θ według wzoru podanego w Raporcie National Research Council z 1996 roku wyliczono za pomocą programów napisanych we własnym zakresie.

WYNIKI

W pracy przeprowadzono symulacyjne obliczenia wpływu współczynnika odziedziczalności na wartości ilorazu wiarygodności dla różnych częstości genotypowych. Wyniki obliczeń zestawiono w tabeli I.

Na podstawie doniesień literaturowych wyznaczono odpowiednie uśrednione współczynniki θ .

W układzie DQA1: $\theta = 0,0044$ dla populacji warszawskiej, małopolskiej, wrocławskiej, pomorsko-kujawskiej i górnośląskiej; $\theta = 0,0016$ dla populacji: warszawskiej, małopolskiej, wrocławskiej, górnośląskiej, szwajcarskiej i węgierskiej.

W układzie THO1: $\theta = 0,0003$ dla populacji gdańskiej, małopolskiej i lubelskiej; $\theta = 0,0014$ dla populacji gdańskiej, hiszpańskiej (Galicja), węgierskiej, południowo-hiszpańskiej, duńskiej, szwajcarskiej, austriackiej i niemieckiej.

W układzie D1S80: $\theta = 0,0021$ dla populacji gdańskiej, małopolskiej i lubelskiej; $\theta = 0,0050$ dla populacji gdańskiej, małopolskiej, lubelskiej, słoweńskiej i holenderskiej.

W układzie VWA: $\theta = 0,0015$ dla populacji gdańskiej i lubelskiej; $\theta = 0,0008$ dla populacji austriackiej, niemieckiej, holenderskiej, szwajcarskiej, gdańskiej, hiszpańskiej (Galicja) i węgierskiej.

Wykorzystując obliczone wartości współczynników θ dla układów stosowanych w sądowym typowaniu DNA, przeprowadzono obliczenia częstości wybranych genotypów w trzech następujących wariantach:

1. zastosowano częstości alleli zbadane dla populacji małopolskiej [29, 30] i przyjęto $\theta = 0$;
2. zastosowano częstości alleli zbadane dla populacji małopolskiej [29, 30], a częstości genotypowe wyliczono według wzoru {2} i przyjęto konserwatywną wartość $\theta = 0,01$ rekomendowaną dla *loci* STR przez National Research Council [22];
3. zastosowano częstości alleli zbadane dla populacji małopolskiej [29, 30] i wartości θ wyliczone dla obszaru Polski, odpowiednie dla każdego układu.

Wyniki obliczeń przedstawia tabela II.

DYSKUSJA

Obliczenia symulacyjne, których wyniki przedstawiono na rycinie 1, wyraźnie wskazują, iż ze wzrostem współczynnika θ , czyli wraz z podwyższeniem się stopnia pokrewieństwa w obrębie danej subpopulacji, zwiększa się obliczona częstość genotypowa danego profilu, liczbowo równoważna prawdopodobieństwu zgodności. Tendencja ta szczególnie uwidacznia się w przypadku rzadkich genotypów, co wyraźnie widać na przykładzie minimalnej możliwej do uzyskania przy zastosowaniu zestawu Profiler Plus (Perkin Elmer) częstości genotypowej (rycina 1b). Przy przyjęciu wartości $\theta = 0,01$ iloraz wiarygodności obniża się o osiem, a przy $\theta = 0,03$ aż o czternaście rzędów wielkości! Przeprowadzone analogiczne obliczenia dla maksymalnej możliwej do uzyskania częstości genotypowej przy zastosowaniu tego zestawu wykazują niewielką „wrażliwość” częstości genotypowych (rzędu 10^{-10}) nawet na wysokie wartości współczynnika θ .

Wielkość genetycznego zróżnicowania między ludzkimi populacjami była intensywnie badana przez genetyków populacyjnych, ale cechy przez nich analizowane są kodowane przez geny wykazujące niski poziom polimorfizmu i mutacji. Przy niskim poziomie selekcji można zatem oczekiwać, iż zróżnicowanie w każdym *locus* będzie podobne. Natomiast *loci* typu STR stosowane w hemogenetyce sądowej są wysoce polimorficzne, co często wiąże się z ich wysoką mutacyjnością. Ponadto są one zlokalizowane często w intronach genów, które podlegają geograficznie zmieniającej się selekcji. Skutkiem tego może być zróżnicowanie współczynnika θ w poszczególnych *loci*, dlatego więc w niniejszej pracy współczynnik θ policzono oddzielnie dla każdego układu.

Przeprowadzone obliczenia korelacji genetycznej dla populacji Polski wskazują raczej na jej jednolity charakter w zakresie badanych *loci*. Współczynniki odziedziczalności dla układu D1S80, TH01 i VWA mają bardzo niskie wartości, co wyraża się ich niewielkim wpływem na wartości częstości genotypowych, a co za tym idzie, z niewielkim wpływem na wartość dowodu z badania DNA. Jak obrazują ryciny 2 i 3, dopiero dla częstości genotypowych rzędu 10^{-10} obserwuje się spadek ilorazu wiarygodności o około jeden rząd wielkości, a dla częstości rzędu 10^{-16} o dwa rzędy wielkości.

Stosunkowo wysoka wartość θ w układzie DQA1 dla populacji Polski ($\theta = 0,0044$) wynika z trudnej do wyjaśnienia odrębności populacji pomorsko-kujawskiej [23], podczas gdy wynik obliczeń uwzględniających populację szwajcarską i węgierską, przy pominięciu pomorsko-kujawskiej, jest znacznie niższy ($\theta = 0,0016$) i wydaje się być bardziej racjonalny.

Tabela III przedstawia porównanie wartości współczynnika θ otrzymanych dla populacji Polski z danymi wyliczonymi dla innych populacji kaukazydalnych.

Ostatni raport National Research Council z 1996 roku proponuje wprowadzenie współczynnika odziedziczalności θ jako miary zróżnicowania struktury populacji do obliczania częstości genotypowych. Poprawkę tę powinno wyznaczać się empirycznie dla poszczególnych populacji. Raport sugeruje stosowanie konserwatywnej wartości współczynnika $\theta = 0,01$, a dla małych, izolowanych populacji $\theta = 0,03$.

W USA wartości θ , uśrednione przez wszystkie stosowane do identyfikacji *loci*, a wyliczone na podstawie baz danych FBI, są zwykle znacznie niższe niż 0,01; dla populacji białej – 0,002, czarnej – 0,007 i hiszpańskiej – 0,009 [5, 6].

Jak wynika z przeprowadzonych obliczeń, zastosowanie wartości współczynnika $\theta = 0,01$ przy częstościach genotypowych rzędu 10^{-8} powoduje spadek wartości ilorazu wiarygodności o jeden rząd wielkości, a przy częstościach rzędu 10^{-16} już o cztery rzędy wielkości. Rzeczywiste wartości współczynników odziedziczalności, zarówno dla populacji Polski, jak i dla innych populacji kaukazydalnych, są znacznie niższe niż 0,01. Dlatego może dziwić rutynowe przyjmowanie przez Forensic Science Service w sprawach karnych wartości współczynnika $\theta = 0,03$ [9].

Na rycinie 3 zilustrowano zastosowanie współczynnika $\theta = 0,0042$ (wyliczonego dla *locus* D1S80) dla populacji małopolskiej w odniesieniu do populacji generalnej Polski. Wpływ ten zaznacza się tylko dla częstości rzędu 10^{-3} i niższych.

Przeprowadzone obliczenia wskazują, iż uwzględnienie współczynnika odziedziczalności θ w obliczeniach wartości dowodów z badania DNA jest niezbędne, jeżeli zachodzi podejrzenie, iż podejrzany pochodzi z izolowanej geograficznie bądź kulturowo społeczności lub gdy istnieje potencjalna możliwość genetycznego, nawet wyrażonego w niewielkim stopniu, zróżnicowania populacji, do której należy podejrzany i sprawca. Należy bowiem pamiętać, że stosując rutynowo do indywidualizacji śladów zestawu około 9 układów typu STR (jak Profiler Plus używany w Pracowni Hemo-genetyki Sądowej Instytutu Ekspertyz Sądowych) otrzymujemy przeciętne częstości genotypowe osobników rzędu 10^{-12} , a zatem w zakresie, w którym nawet niskie wartości współczynnika θ mają wpływ na obniżenie wartości dowodu. Niezbędne wydaje się zatem przeprowadzenie badania rozkładu częstości genotypowych w małych lub częściowo izolowanych środowiskach celem wyznaczenia realnej wielkości subtelnego zróżnicowania populacji Polski. Wciąż powszechnie stosowane założenie, iż $\theta = 0$, oznacza ignorowanie praw genetyki populacyjnej. Niemniej stosowanie arbitralnie przyjętych wartości dla współczynnika odziedziczalności $\theta \geq 0,01$ wydaje się być nieuzasadnione i zbyt „konserwatywne” wobec podejrzanego.