Determination of combined sibship indices "gray zone" using 15 STR loci for central Bosnian human population

Jasmin Musanovic · Marijana Filipovska-Musanovic · Lejla Kovacevic · Dzenisa Buljugic · Mirela Dzehverovic · Jasna Avdic · Damir Marjanovic

Abstract In our previous population studies of Bosnia and Herzegovina human population, we have used autosomal STR, Y-STR, and X-STR loci, as well as Y-chromosome NRY biallelic markers. All obtained results were included in Bosnian referent database. In order of future development of applied population molecular genetics researches of Bosnia and Herzegovina human population, we have examined the effectiveness of 15 STR loci system in determination of sibship by using 15 STR loci and calculating different cut-off points of combined sibship indices (CSI) and distribution of sharing alleles. From the perspective of its application, it is very difficult and complicated to establish strict CSI cut-off values for determination of the doubtless sibship. High statistically significant difference between the means of CSI values and in distribution of alleles sharing in siblings and non-siblings was noticed ($P < 0.0001$). After constructing the "gray zone", only one false positive result was found in three CSI cut-off levels with the highest percent of determined sibship/non-sibship at the CSI = 0.067, confirming its practical benefit. Concerning the distribution of sharing alleles, it is recommended as an informative estimator for its usage within Bosnia and Herzegovina human population.

Keywords DNA typing · Short tandem repeat loci (STR) · Cumulative/Combined sibship indices (CSI) · Gray zone · Distribution of sharing alleles · Bosnia and Herzegovina

Introduction

During the last two decades, different methods of forensic DNA testing have been widely established and accepted as the standard procedure for determination of parenthood. However it is proved that same molecular genetic technique could be used within determination of other family relationships, such as sibling and half-sibling. The calculated values of likelihood ratio (LR) will not be statistically impressive as the paternity indices (PI) and combined paternity indices (CPI) are, but certainly could point to the potential sibship of two examined persons [1]. Also, the statistical forms are slightly different, but they are still primarily based on the similar bio-statistical concept of observed population genetic heterogeneity. Focal point of this concept is searching for the answers to the following question: how many times the possibility of two examined persons to be siblings is higher than the possibility of two examined persons to be random non-siblings from the local population?

Based upon the degree of sharing of alleles between two DNA profiles, it is possible to determine a combined sibship indices (CSI) [2]. However, determination of universal CSI cut-off point to doubtless distinguish siblings from

J. Musanovic (✉)
Medical Faculty, University of Sarajevo, Sarajevo, Bosnia and Herzegovina
e-mail: jasmin.m.gen@gmail.com

J. Musanovic
Institute for Genetics, Laboratory for Molecular Medicine, Medical Faculty, University of Sarajevo, Sarajevo, Bosnia and Herzegovina

M. Filipovska-Musanovic
Health House with Polyclinic, Kakanj, Bosnia and Herzegovina

L. Kovacevic · D. Buljugic · M. Dzehverovic · J. Avdic · D. Marjanovic
Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina

D. Marjanovic
Genos doo, Zagreb, Croatia

unrelated individuals, especially in relatively small human populations, is a demanding task. Previous studies have described different values of CSI cut-off were used, such as 0.067, 1, 3, 10, 100, and others [3–6]. From those reasons many different authors suggested construction of the "gray zone", the zone of inconclusive sibship. This approach is decreasing the possibility of false sibling determination, which is very important regarding usage of this type of analysis within wide sphere of forensic analysis [4].

Modern Bosnia and Herzegovina is a multi-ethnic and multi-religion country, located on the southern-eastern part of Europe (Fig. 1). Certain archaeological findings indicate that its territory has continuously been populated since the Paleolithic. During that time joint activities of vast number of different factors created fascinating diversity of local human populations [11]. Therefore, applying of any kind of population or forensic statistical methods within this population represent, almost, challenging experience. Previous studies proved that a multiple-locus STR profile could be highly differentiating and informative even among the members of small relatively isolated communities, such as villages of the mountain area of Bosnia and Herzegovina [12]. At this point, we would like to check efficiency of the same genetic markers (15 STR loci) in determination of family relationships, within this relatively genetically interesting human population. Besides scientific importance of the results obtained in this kind of study, there is also important applicative value of obtained information through these activities for the Bosnia and Herzegovina forensic community, considering existence of large ongoing mission of the war victims' identification in Bosnia and Herzegovina.

Fig. 1 Location of Bosnia and Herzegovina

Materials and methods

DNA profiles of 58 known pairs of siblings (descendents of the same parents) and from 58 known pairs of non-siblings as control group was used in this study. Sibling status for most of the examined pairs was proved by parental DNA profiles. All tested individuals were voluntary donors appropriately informed about this research.

Buccal swabs were used as the DNA source. All specimens were air-dried, placed in 1.5 ml tubes, and immediately transported to the laboratory at the Institute for Genetic Engineering and Biotechnology. Samples were stored at $-80°C$ until DNA was extracted. Qiagen DNA-easy™ Tissue Kit was used as DNA extraction procedure [13]. The PowerPlex 16 kit was used to simultaneously amplify by PCR 15 STR loci [14]. Approximately 1 ng/μl of DNA was used in all PCR reactions. The total volume of each reaction was 5 μl. The PCR amplification was carried out in AB Gene Amp PCR System 9700 Thermal Cycler (ABI, Foster City, CA) according to the manufacturer's recommendations. Electrophoresis of the amplification products was performed on an ABI PRISM 310 Genetic Analyzer (ABI, Foster City, CA). The raw data were compiled and analyzed using accessory software, ABI Data Collection Software and GeneMapper™ 3.2.

Likelihood ratio was calculated for each locus inside all siblings and non-siblings pairs, CSI was calculated accordingly by multiplying LR for all observed STR loci. CSI cut-off values for distinguishing siblings from non-siblings were set up at the 0.067, 1, 3, 10, and 10.3, levels, according to the previously published studies [3–7]. All CSI values over the CSI cut-offs were described as positive values which are verifying positive sibling. On the other hand, CSI values below CSI cut-offs were described as negative results, which are verifying negative sibling results. The "gray zone" (area of inconclusive CSI values) was constructed for each CSI cut-off values to identify pairs with inconclusive CSI values and it was delimited by two cut-off points, calculated by existing formulas [4, 8]. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for each of CSI cut-off levels as the indicators of the efficiency of used method for determining sibship. Likelihood ratio for positive test result (LR+) and likelihood ratio for negative test result (LR−) were defined in terms of determined sensitivity and specificity: LR+ = sensitivity/1 − specificity and LR− = 1 − sensitivity/specificity. Two cut-off points, one associated with the minimal desirable value of LR+ and the other maximal desirable value of LR− were identified delimiting the "gray zone" [4]. The second type of sibling estimation that was performed was distribution of sharing alleles within pairs in siblings/non-sibling groups.

Student *t*-test was used to estimate difference between the means of CSI values between the siblings group and non-siblings group and chi-square test was used to estimate difference in distributions of sharing alleles between those groups. All these analysis were performed by home made specifically designed microsoft-excel sample sheet.

Results

The results of CSI for groups of siblings and non-siblings are presented in Fig. 2. The minimum of all CSI values is 0.0000005 (0.000049999%) detected in one non-sibling pair and the maximum is 2969454702.4902 (99.99999 9966%) detected in one sibling pair. Minimal value of CSI in siblings is very low 0.536, lower than CSI in three pairs of non-siblings. The highest value of CSI in non-siblings is 62.3941, higher than CSI values in 12 pairs of siblings. However, the result of Student *t*-test is showing high statistically significant difference between the means of CSI values in siblings and non-siblings ($P < 0.0001$). Sensitivity, specificity, PPV, and NPV for different CSI cut-off values and the cut-off points for the "gray zone" for each level of CSI cut-off are shown in Table 1. For the CSI cut-off value for level 1 [3, 4], the most used in literature, the sibship is determined in 98.276% pairs of siblings. For the same level CSI cut-off, the sibship is excluded in 96.552% of non-sibling pairs. At this level we determine known sibship and non-sibship in 97.414% of all pairs. Also, 2.586% of all pairs are false positive and false negative. For the CSI cut-off for the level 3 [5] the percent of determined true siblings and non-siblings and the percent of false results are same as for the level 1. For the level 0.067 [4, 6] percent of true results is lower (96.552%) and the percent of false results was higher (3.448%). For the level 10 [6] and 10.3 [4, 6], the percent of true results are
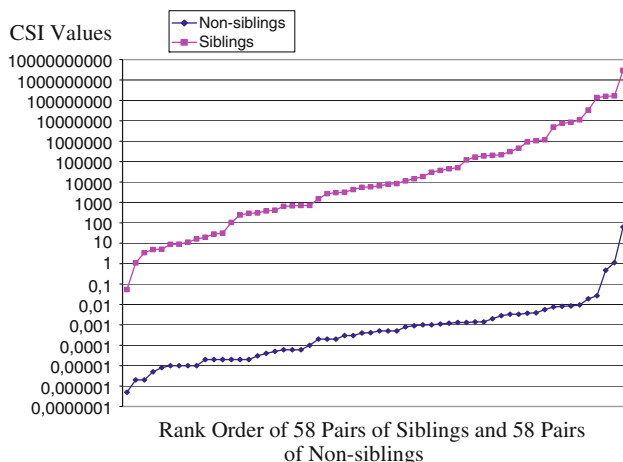
Table 1 Evaluation of the sensitivity, specificity, PPV, NPV, and the minimum and the maximum cut-off points of the "gray zone" for each of CSI cut-off value
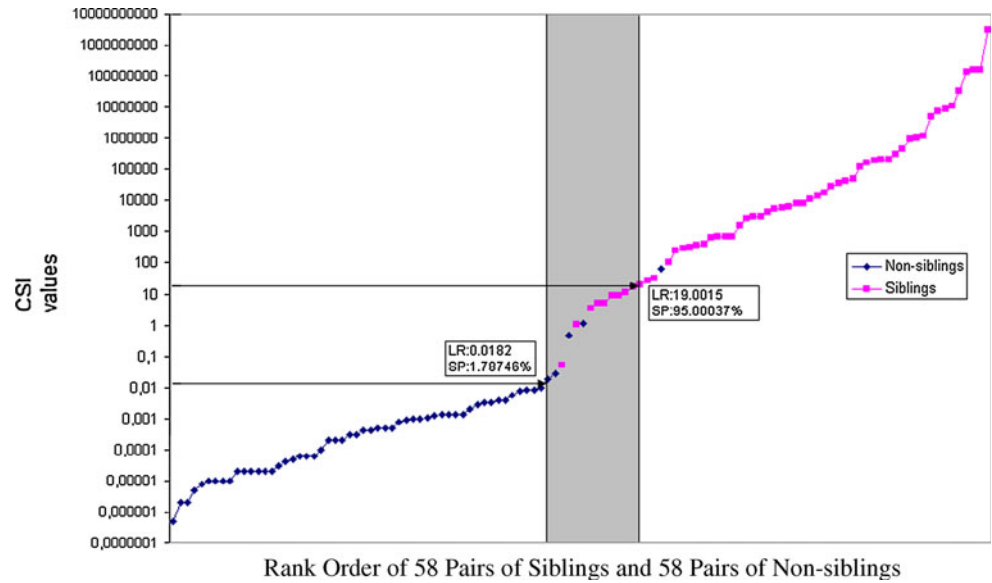
CSI cut-off points	SEN %	SPE %	PPV %	NPV %	MinGZ	MaxGZ
0.067	98.276	94.828	95.000	98.214	0.0182	19.0015
1	98.276	96.552	96.610	98.246	0.0178	28.5023
3	96.552	98.276	98.246	96.610	0.0351	56.0046
10	87.931	98.276	98.077	89.063	0.1228	51.0041
10.3	87.931	98.276	98.077	89.063	0.1228	51.0041
Minimum	87.931	94.828	95.000	89.063	0.0178	19.0015
Maximum	98.276	98.276	98.246	98.246	0.1228	56.0046

SEN % sensitivity, *SPE* % specificity, *PPV* % positive predictive value, *NPV* % negative predictive value, *minGZ* minimum cut-off point of the "gray zone", *maxGZ* maximum cut-off point of the "gray zone"

more decreased (93.103%) and the percent of false results was more increased (6.897%). By constructing the "gray zone" (Fig. 3) decreasing in percent of false results at all CSI cut-off levels is found, but the most at CSI level 0.067 (decreased 2.586%). After constructing the "gray zone", the highest percent of determined known siblings and non-siblings is found at level CSI = 0.067 (87.931%). (Table 2) After constructing the "gray zone", 17 pairs (12 siblings and 5 non-siblings) at all CSI cut-off levels are found as inconclusive or false positive or false negative.

The distribution of alleles sharing for the siblings and the non-siblings is shown in Fig. 4. There were total of 870 observations for each group (15 loci × 58 pairs). The high statistically significant difference in distribution of alleles sharing between the group of siblings and the group of non-siblings ($P < 0.0001$) is noticed. Higher percent of "sharing zero alleles" is found in group of non-siblings and higher percent of "sharing two alleles" is detected in group of siblings. The sibling's pair with the lowest CSI value (0.0536) has "two alleles sharing" at one locus and "zero alleles sharing at three loci". On the other hand, within one sibling pair "two alleles sharing" was detected over nine loci. The highest CSI value (2969454702.4902) with eight "two alleles sharing" and one "zero alleles sharing" was detected. Then again, ten "zero alleles sharing" were detected within two non-siblings pairs. The lowest calculated CSI value was based on nine "zero alleles sharing". Also, the absence of "two alleles sharing" over any loci was detected within nine non-siblings pairs. The highest calculated CSI value from the control non-sibling group was 62.3941 and was based on three "two alleles sharing" and two "zero alleles sharing". Regarding analysis of possible inconclusive and false negative/positive results, within the non-determinative cases in siblings and non-siblings, within three of five non-sibling's pairs with CSI

CSI Values

Fig. 2 Combined sibship indices (CSI) for siblings and non-siblings

Fig. 3 Construction of "gray zone" for CSI cut-off value = 0.067

Rank Order of 58 Pairs of Siblings and 58 Pairs of Non-siblings

Table 2 Sensitivity, specificity, and false results of CSI cut-off method in determining sibship without and with "gray zone"

|	Without constructed "gray zone"						With constructed "gray zone"						
CSI	SEN %	SPE %	(SEN + SPE)/2 %	FP %	FN %	(FP + FN) %	SEN %	SPE %	(SEN + SPE)/2 %	GZ %	FP %	FN %	(FP + FN) %
0.067	98.276	94.828	96.552	2.586	0.862	3.448	84.483	91.379	87.931	11.207	0.862	0	0.862
1	98.276	96.552	97.414	1.724	0.862	2.586	81.034	91.379	86.207	12.931	0.862	0	0.862
3	96.552	98.276	97.414	0.862	1.724	2.586	79.310	94.828	87.069	12.069	0.862	0	0.862
10	87.931	98.276	93.103	0.862	6.035	6.897	79.310	94.828	87.069	11.207	0.862	0.862	1.724

SEN % sensitivity, SPE % specificity, FP % total false positive in %, FN % total false negative in %, GZ % % of cases in "gray zone"

0.0189, 0.0268, and 1.1138, two "two sharing alleles" and five or six "zero alleles sharing" were detected. For two non-sibling's pairs with high CSI (0.4722 and 62.3941) 1 "two alleles sharing" and three "zero alleles sharing" were detected. On the other hand two sibling pairs with the lowest CSI (0.0536 and 1.0747) have one or two "two alleles sharing" and two or three "zero alleles sharing" Within other sibling's pairs with inconclusive sibship by CSI values, the similar distribution of two and zero alleles sharing is detected (over 3–5 loci).

Discussion

There are many disagreements regarding evaluation of the optimal cut-off values for any of existing sibling determination methods [3–6]. From forensic genetic perspective, it is extremely important not to miss "the opinion" [4]. This means that it is very important not to miss any true sibship or any true non-sibship, because these are important forensic proofs that could have great influence on human lives. Because of these reasons it is very uncertain to relay
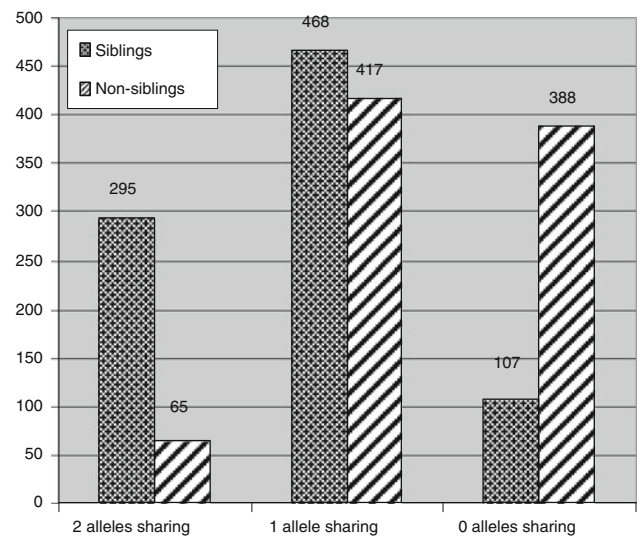
Fig. 4 Distribution of alleles sharing for siblings and non-siblings

on strict cut-offs for determining two siblings or non-siblings with high level of certainty. To overcome this problem, it is recommended in literature to construct a three-

zone partition for sibship test to avoid the constraint of "black or white" decision [4, 8]. This approach allow the most accurate estimation for the results in upper area as positive, in lower area as negative and in the middle area (the "gray zone") as uncertain results that require another genetic test. In our study, for the Bosnia and Herzegovina human population, we found very useful and helpful this method. Before we constructed the "gray zone", we found for the CSI cut-off at the level 1, higher sensitivity and specificity then those found in literature [4]. Also, high sensitivity and specificity has been shown at CSI cut-off level 3, recommended as CSI cut-off values in Tzeng et al. [5]. If we take the first CSI cut-off at level 1, we can found two pairs of known non-siblings as determined siblings but it would be the false positive result in these cases. At this level, one pair of siblings would be false determined as non-siblings. At the level of 3, one pair of non-siblings is false determined as siblings and two pairs of siblings as non-siblings. In the studies where these CSI cut-off values were used, the results derived from larger populations than Bosnian have shown the similar deviation [3, 5]. This could prove how difficult and complicated is establishing strict CSI cut-off values for determining the doubtless sibship. After constructing the "gray zone" for all CSI cut-off levels, we have detected decreasing in true positive and true negative result in favor of constructing the group of pairs with uncertain sibship that require another genetic test such as parentage testing and mtDNA sequencing, performed and recommended in Tzeng et al. [5]. Also, only one single false result, in sense of false positive result, was detected at the CSI cut-off levels 0.067, 1, and 3. At the CSI cut-off level 0.067, that is detected as lower cut-off point of the "gray zone" in Indian population [4], we found after the construction of the "gray zone", the highest percent of true positive and true negative results in sample from central Bosnia and Herzegovinians. This has been shown the practical benefit of constructing the three-zone partition that increases the efficiency of this method for determining the sibship.

Concerning the distribution of sharing alleles, it is postulated that there is no evidence of sibship if there is no two alleles sharing at any loci [6]. But in Pu et al. [6], it has been shown that there is no evidence this method could be recognized as the optimal approach for the sibship determination. More likely, it could be recommended as an informative estimator [4]. In this sample of Bosnians, we found the similar results as in Giroti et al., Pu et al., and Peterson et al. [4, 6, 9]. An expected probability of sharing alleles by two full siblings is 0.25 for two and zero sharing alleles and 0.50 for one sharing alleles. In our study we found statistically significant increase in frequency of two alleles sharing in siblings group (33.908%) and decrease in frequency of zero alleles sharing (12.299%). In comparing with non-siblings group, there is high statistically significant difference in two (7.471%) and zero (44.598%) sharing alleles, but no difference in one allele sharing what match the result found in small Indian sample [4] and Taiwanese population [5]. The significant difference in two and zero alleles sharing was described as phenomenon of polarization and absence polarization in one allele sharing could be due to homogeneity [5]. But, most of the persons that were involved in this study are from central part of Bosnia, so that could be possible reason of the relative homogeneity of this sample as results shows. In order to obtain more realistic picture in our future study, we should consider greater stratified sample included in all sub-populations (regionally and ethnically determined) in Bosnia and Herzegovina.

According to distribution of sharing alleles in group of pairs with inconclusive and false negative and false positive results, we found that in three non-siblings questionable tests (CSI 0.0189, 0.0268, and 1.1138) this could increase the probability of excluding sibship but with recommendation to perform additional genetic test. In two non-siblings questionable tests (CSI 0.4722 and 62.3941) this distribution indicates to uncertain excluding of sibship without additional genetic test. In sibling's group, two pairs with the lowest CSI (0.0536 and 1.0747) have the similar result as in the case of last two non-sibling's pairs what increase the uncertainty of their sibship and require additional test. In other sibling's pairs with inconclusive sibship by CSI values we found profiles that are different from those in determined siblings and confirm the result of inconclusive sibship in these pairs with requirement of additional genetic testing. By combining these two methods, the percent of true positive and true negative results increased for 2.586% and at the level CSI = 0.067 we confirmed sibship/non-sibship in 90.517% pairs what is the greatest percent.

In two sibling's pairs we found two sharing alleles at nine loci, what according to result in Giroti et al. [4] with one sibling pair sharing two alleles at ten loci, indicate to high genetic relatedness. In our sample, we did not find two alleles sharing at zero loci in any of sibling's pairs what is described in the same study [4]. In two non-siblings pairs, we found zero sharing alleles at ten loci and in Giroti et al. [4] described the non-sibling pair with zero sharing alleles at 12 loci. However, this distribution of sharing alleles could provide additional information about siblings and non-siblings and their sibship probability. Nevertheless, additional genetic tests, such as parental testing or (if parental analysis is non-performable) analysis of additional autosomal or lineage (mtDNA/Y-chromosome) markers is required for some cases from the "gray zone". Importance of additional testing for these cases could be emphasized with the fact that every sixth case of paternity testing performed in Bosnia and Herzegovina was negative [10]. Also, regarding the fact that

numerous of more or less isolated local human communities are located all over Bosnia and Herzegovina [12, 15], our future studies will be probably focused on analysis of these parameters within this type of populations.

Conclusion

For the central Bosnian population, the highest percentage of correct findings, and the greatest efficacy in distinguishing siblings from unrelated persons for the observed pattern was found in the marginal value of CSI = 1 and CSI = 3, but without creating a "gray zone", and with the percentage of erroneous findings of 2.586%.

By constructing borders of the "gray zone" and forming three-zone partition of CSI values, the possibility of error is significantly reduced especially for CSI = 0.067 for 2.586%, and after the creation of the "gray zone", this value for the observed pattern gives optimal results. Combining of these two methods, the probability of exact finding increased for 2.586%. This shows the value of creating a "gray zone" and recommends its use in forensic laboratories. However, in some cases it is necessary to implement and additional genetic test, such as parental testing or (if parental analysis is non-performable) analysis of additional autosomal or lineage (mtDNA/Y-chromosome) markers is required for some cases from the "gray zone".

References

1. Butler JM (2005) Forensic DNA typing: biology, technology, and genetics of STR markers, 2nd edn. Elsevier, London
2. Wenk RE, Traver M, Chiafari FA (1996) Determination of sibship in any two persons. Transfusion 36:259–262
3. Reid TM, Wolf CA, Kraemer CM, Lee SC, Baird ML, Lee RF (2004) Specificity of sibship determination using the ABI identifiler multiplex system. J Forensic Sci 49:1262–1264
4. Giroti RI, Verma S, Singh K, Malik R, Talwar I (2007) A gray zone approach for evaluation of 15 short tandem repeat loci in sibship analysis: a pilot study in Indian subjects. J Clin Forensic Med 14:261–265
5. Tzeng CH, Lyou JY, Chen YR, Hu HY, Lin JS, Wang SY et al (2000) Determination of sibship by PCR-amplified short tandem repeat analysis in Taiwan transfusion 40:840–845
6. Pu CE, Linacre A (2008) Systematic evaluation of sensitivity and specificity of sibship determination by using 15 STR loci. J Forensic Leg Med 15:329–334
7. Thomson JA, Ayres KL, Pilotti V, Barrett MN, Walker JIH, Debenham PG (2000) Analysis of disputed single-parent/child and sibling relationships using 16 STR loci. Int J Legal Med 115:128–134
8. Coste J, Pouchot J (2003) A grey zone for quantitative diagnostic and screening tests. Int J Epidemiol 32:304–313
9. Peterson S, Hurst JL (2009) How effective is recognition of siblings on the basis of genotype? J Evol Biol 22:1875–1881
10. Elez N, Hasicic S, Hadzovic E, Halilovic Z, Kamenjas A, Rizvanovic A et al (2009) Ratio of "non-paternity occurrence" in paternity testing in Bosnia and Herzegovina. Proceedings of the 6th ISABS conference on human genome project based application in forensic science, anthropology and individualized medicine 1–5 June. International society for applied biological sciences, Split
11. Marjanovic D, Pojskic N, Kapur L, Haveric S, Durmic-Pasalic A, Bajrovic K et al (2008) Overview of human population–genetic studies in Bosnia and Herzegovina during the last three centuries: history and prospective. Coll Antropol 32:315–321
12. Marjanovic D, Kapur L, Drobnic K, Budowle B, Pojskic N, Hadziselimovic R (2004) Comparative study of genetic variation at 15 STR loci in three isolated populations of the Bosnian mountain area. Hum Biol 76(1):15–31
13. QIAGEN genomic DNA handbook (2003) Qiagen Companies, Hilden
14. Promega (2008) PowerPlex 16 System: technical manual. Promega Corporation, Madison
15. Marjanovic D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, Vidovic S et al (2005) The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. Ann Hum Genet 69:757–763