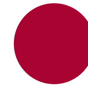
View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Copenhagen University Research Information System

UNIVERSITY OF COPENHAGEN



Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex(®) ESI 17 and AmpFSTR(®) NGM SElect™ in Somalis, Danes and Greenlanders

Tomas Mas, Carmen; Mogensen, H S; Friis, S L; Hallenberg, C; Stene, M C; Morling, N

Published in:
Forensic science international. Genetics

DOI:
10.1016/j.fsigen.2014.04.004

Publication date:
2014

Citation for published version (APA):
Tomas Mas, C., Mogensen, H. S., Friis, S. L., Hallenberg, C., Stene, M. C., & Morling, N. (2014). Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex(®) ESI 17 and AmpFSTR(®) NGM SElect™ in Somalis, Danes and Greenlanders. *Forensic science international. Genetics*, *11*, e18–e21. https://doi.org/10.1016/j.fsigen.2014.04.004

Download date: 08. apr.. 2020

Forensic Science International: Genetics 11 (2014) e18–e21

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Forensic Population Genetics–Short Communication

Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex® ESI 17 and AmpFℓSTR® NGM SElect™ in Somalis, Danes and Greenlanders

C. Tomas *, H.S. Mogensen, S.L. Friis, C. Hallenberg, M.C. Stene, N. Morling

Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Frederik V's Vej 11, DK-2100 Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 12 November 2013
Received in revised form 8 April 2014
Accepted 10 April 2014

Keywords:
PowerPlex® ESI 17
AmpFℓSTR® NGM SElect™
SE33
Somalia
Denmark
Greenland

ABSTRACT

A concordance study of the results of PowerPlex® ESI 17 and AmpFℓSTR® NGM SElect™ kits obtained from 591 individuals from Somalia (*N* = 198), Denmark (*N* = 199) and Greenland (*N* = 194) was performed. Among 9456 STR types, seven discordant results were found with the two kits: one observed in the D19S433 system in an individual from Denmark and six in the SE33 system in six individuals from Somalia. Sequencing of SE33 in the six samples with discordant results showed G > A transition 15 bp downstream of the repeat unit in three of the individuals, and G > A transition 68 bp downstream of the repeat unit in the other three individuals. Population data for 16 autosomal STR systems analyzed in 989 individuals from Somalia, Denmark and Greenland are also presented. The highest mean heterozygosity was observed in Danes (82.5%). With the exception of D8S1179 in Danes, no significant deviations from Hardy–Weinberg expectations were observed. Only one pair of systems (D12S391 and D18S51) showed significant allelic association in Greenlanders (after Holm-Šidák correction). A MDS plot drawn from pairwise F_{ST} values calculated between 21 populations showed a clear displacement of the Greenlandic population versus the other ones included in the analyses. The highest combined chance of exclusion and power of discrimination was observed for Danes reaching values of 99.9999987% and 1 in 1.8×10^{21}, respectively.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

In order to improve the international comparison of DNA profiles and decrease the probability of false matches, the European Standard Set (ESS) loci was extended in 2009 from 7 to 12 autosomal STR loci [1]. Consequently, a number of new STR commercial kits that include the extended ESS loci have been developed by several companies, e.g. PowerPlex® ESI 17 (Promega) and AmpFℓSTR® NGM SElect™ (Life Technologies – LT). Both kits include the same 16 autosomal STR loci and the Amelogenin locus.

A concordance study between the PowerPlex® ESI 17 and AmpFℓSTR® NGM SElect™ kits was performed in our laboratory using DNA from three population samples: Somalis, Danes and Greenlanders. Moreover, a new allele frequency database was built for the 16 autosomal STRs included in the two kits for the three studied populations. Population and forensic statistical parameters are presented.

2. Materials and methods

2.1. Populations and samples

A total of 989 samples from presumably unrelated individuals from Somalia (*N* = 198), Denmark (*N* = 597) and Greenland (*N* = 194) were selected for this study. The samples were collected from paternity and family reunion cases investigated at the Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark.

The work was approved by the Danish ethical committee (H-1-2011-081).

2.2. DNA extraction

DNA from around 60% of the samples was extracted from blood by using the QIAamp DNA blood mini kit (Qiagen GmbH, Hilden,

* Corresponding author. Tel.: +45 35326283; fax: +45 35326120.
E-mail address: carmen.tomas@sund.ku.dk (C. Tomas).

http://dx.doi.org/10.1016/j.fsigen.2014.04.004
1872-4973/© 2014 Elsevier Ireland Ltd. All rights reserved.

Germany). The remaining 40% of the samples were investigated by direct PCR on 1.2 mm disks punched from saliva stains collected on FTA® cards (Whatman Inc., Clifton, NJ).

2.3. STR genotyping and quality control

DNA samples from 591 individuals from Denmark (N = 199), Greenland (N = 194) and Somalia (N = 198) were typed using both AmpFℓSTR® NGM SElect™ (LT) and PowerPlex® ESI 17 (Promega) amplification kits containing the following autosomal STRs: D10S1248, D12S391, D16S539, D18S51, D19S433, D1S1656, D21S11, D22S1045, D2S1338, D2S441, D3S1358, D8S1179, FGA, SE33, TH01, VWA. Furthermore, 398 more samples from Denmark were typed using only the AmpFℓSTR® NGM SElect™ kit (LT). The samples were amplified in a 10 μL reaction volume. A number of 25, 26 or 27 thermo cycles were used for the NGM SElect™ kit and 28 thermo cycles were used for the ESI 17 kit. PCR products were electrophoresed on a 3130*xL* Genetic Analyzer (LT). Alleles were assigned using GeneMapper ID 3.2 or GeneMapperID-X software (LT). All samples were typed in duplicates (using one or two kits) at the laboratory of the Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen that is accredited according to the ISO 17025 standard.

2.4. SE33 system

Six DNA samples were sequenced for the SE33 STR system. The DNA was first amplified by PCR using specific primers for the SE33 system (Table 1).

A library was built from the PCR products following the manufacturer's instructions (Roche, Rapid Library (RL) preparation method manual, May 2010; rev. June 2010) and using six RL Multiplex Identifier (MID) adaptors. Emulsion PCR (emPCR) and pyrosequencing was subsequently performed according to the emPCR amplification method manual – Lib L- (Roche, May 2010; rev. June 2010) and the sequencing method manual (Roche, May 2010; rev. June 2010), respectively. A 454 GS Junior (454 Life Sciences, Branford, CT, USA) was used for the pyrosequencing. Sequencing data was scrutinized using the BioEdit Sequence Alignment Editor [2].

Three of the DNA samples with discordances in the SE33 system were further typed with the PowerPlex® ESI 17 Pro kit (Promega) following the manufacturer's instructions.

2.5. Population and forensic genetic parameters

Population genetic parameters, such as allele frequencies, Hardy–Weinberg equilibrium (HWE), allelic association, AMOVA and pairwise F_{ST} values were calculated using the Arlequin 3.5 software [3]. Allele frequencies of the 16 STRs were used to calculate pairwise F_{ST} values for each STR system between the following 19 populations: Denmark, Somalia, Greenland (this study); CEPH populations from Africa, Middle East, Europe, America, East Asia, South Asia, Oceania [4]; Madagascar, Morocco [5]; Afro Americans and Hispanics [6]; Germany [7]; Poland [8]; Spain [9]; Basques [10]; Brazil [11]. Moreover, F_{ST} distances were also obtained for 15 of the 16 STRs (SE33 was excluded) using allele frequencies from the 19 populations listed above and the POPTREE2 software [12]. The SE33 system was not included in the analyses due to limited published data. The resulting F_{ST} values were represented in a multidimensional scaling (MDS) plot using SPSS Statistics for Windows, ver. 17.0 (Chicago SPSS Inc.). Statistical significance for multiple tests was corrected using the Holm-Šidák procedure [13].

Statistical parameters of forensic interest such as power of discrimination (PD) and chance of exclusion (CE) were calculated using the formulas from Fisher [14] and Ohno [15], respectively. Typical paternity indices (defined as the product of the geometric mean per system [16]) for trios and duos were estimated using DNA VIEW™ version 28.103 [17].

Table 1
PCR primers used to amplify the SE33 STR system for sequencing.

Sample	Primers	Sequence 5′-3′	Product size (bp)	Ref.
1, 2, 3, 5, 6	Forward 1	AAAGAAGGCTGGGCACTGT	676	[20]
	Reverse 1	GCGGTGATAGCATCATCCAT		
4	Forward 2	CAGTGAGCCGAGGTCATG	392	[19]
	Reverse 2	GGCAAGGACAAGGTTCTGTG		

3. Results and discussion

3.1. Concordance between ESI 17 and NGM SElect™ results

Seven discordant results were found when analyzing 591 individuals with both the ESI 17 and the NGM SElect™ kits. One of the seven discordances was observed with the D19S433 system in one of 199 samples from Denmark. The sample was typed as D19S433*13,15 with the ESI 17 kit and D19S433*15 with the NGM SElect™ kit. A mutation, if located on the primer binding site, could explain the discordancy observed for the D19S433 system. A silent allele, probably caused by a G > A transition 32 bp downstream of the D19S433 repeat region, was previously observed in a Japanese population analyzed with the AmpFℓSTR® ® Identifiler® kit (LT) [18]. Six discordant results were observed in SE33 among 198 Somalis. In three of the six samples a heterozygote versus an apparently homozygote discordance was observed (Table 2). In the other three samples, an intermediate X.3 allele was observed with the ESI 17 kit, while an X.2 allele was observed with the NGM SElect™ kit (Table 2). No discordance between the results obtained with NGM SElect™ and ESI 17 Pro kits was observed for the three intermediate alleles in the SE33 system.

3.2. The SE33 system

Sequencing results of the SE33 system are shown in Table 2. In each of the three samples with a heterozygote/apparent homozygote discordance, a G > A transition 15 bp downstream of the repeat unit was observed. This mutation, if located in the primer binding region of the reverse primer of SE33 in the NGM SElect™ kit could explain the observed results. In each of the samples with differences in the intermediate allele X.2 vs. X.3, a G > A transition was observed 68 bp downstream of the repeat unit. This G > A transition has previously been observed in an African American individual, eight individuals with African ancestry and one of Caucasian ancestry [19,20]. Four SNPs within a polymorphic region located between 58 bp and 70 bp downstream of the repeat unit of the SE33 system have been described [19,20]. Such polymorphisms shift the electrophoretic mobility of the amplicons due to changes in the secondary structure of the polymorphic region [19,20]. While the SE33 amplicons obtained using the NGM SElect™ and the ESI 17 Pro kits do not include this polymorphic region, this region is included in the amplicons obtained with the ESI 17 kit. Differences in the sequence of the polymorphic region will result in different electrophoretic mobility of the SE33 alleles amplified with the ESI 17 and the NGM SElect™ kits.

Table 2
Samples sequenced for the SE33 system. STR types obtained with the ESI 17 and NGM SElect[TM] (NGMS) kits, sequence of the repeat units and the mutation found downstream of the repeat unit.

Sample	ESI17	NGMS	Sequence (repeat unit)	Mutation
1	16	–	$(AAAG)_2$-AG-$(AAAG)_3$-AG-$(AAAG)_{16}$-G-$(AAAG)_3$-AG	G > A 15 bp downstream
	25.2	25.2	$(AAAG)_2$-AG-$(AAAG)_3$-AG-$(AAAG)_9$-AAAAAG-$(AAAG)_{15}$-G-AAGG-$(AAAG)_2$-AG	
2	13.2	–	$(AAAG)_2$-AG-$(AAAG)_{17}$-G-$(AAAG)_3$-AG	G > A 15 bp downstream
	30.2	30.2	$(AAAG)_2$-AG-$(AAAG)_3$-AG-$(AAAG)_{10}$-AG-$(AAAG)_{20}$-G-AAGG-$(AAAG)_2$-AG	
3	16	–	$(AAAG)_2$-AG-$(AAAG)_3$-AG-$(AAAG)_{16}$-G-$(AAAG)_3$-AG	G > A 15 bp downstream
	17	17	$(AAAG)_2$-AG-$(AAAG)_3$-AG-$(AAAG)_{17}$-G-$(AAAG)_3$-AG	
4[a]	14.3	14.2	$(AAAG)_2$-AG-$(AAAG)_{18}$-G-$(AAAG)_3$-AG	G > A 68 bp downstream
	17	17	$(AAAG)_2$-AG-$(AAAG)_3$-AG-$(AAAG)_{17}$-G-$(AAAG)_3$-AG	
5[a]	15.3	15.2	$(AAAG)_2$-AG-$(AAAG)_{19}$-G-$(AAAG)_3$-AG	G > A 68 bp downstream
	25.2	25.2	$(AAAG)_2$-AG-$(AAAG)_3$-AG-$(AAAG)_9$-AAAAAG-$(AAAG)_{15}$-G-AAGG-$(AAAG)_2$-AG	
6[a]	16.3	16.2	$(AAAG)_2$-AG-$(AAAG)_{20}$-G-$(AAAG)_3$-AG	G > A 68 bp downstream
	18	18	$(AAAG)_2$-AG-$(AAAG)_3$-AG-$(AAAG)_{18}$-G-$(AAAG)_3$-AG	

[a] Samples 4, 5 and 6 were also typed with the ESI 17 Pro kit. The SE33 types obtained with the ESI 17 Pro kit were identical to those obtained with the NGM SElect[TM] kit.

3.3. Population and forensic parameters

Allele frequencies of the 16 autosomal STRs in Somalis, Danes and Greenlanders are shown in Supplementary material 1. The observed heterozygosity ranged from 0.479 (HUMTH01 in Greenlanders) to 0.970 (SE33 in Danes). The lowest mean heterozygosity was observed in the population in Greenland (0.734) and the highest mean value was observed in the Danish population (0.825). With the exception of the D8S1179 system in Danes, no deviation from Hardy-Weinberg expectation was observed after Holm-Šidák correction (Supplementary material 1). Significant allelic association ($P < 0.05$ after Holm-Šidák correction) was only detected in one out of 120 pairwise comparisons (D12S391-D18S51) in the population in Greenland. The two STR systems located at the same chromosome (D12S391 and vWA) showed significant association ($P < 0.05$) in Greenlanders before the Holm-Šidák correction.

The AMOVA analysis performed for the three studied populations showed that around 4% of the overall genetic variation could be explained by differences among populations ($F_{ST} = 0.039$; $P < 0.00001$). Statistical significant F_{ST} values ($P < 0.00001$) were also observed for the three pairwise comparisons, with values of 0.025 between Danes and Somalis, 0.047 between Danes and Greenlanders and 0.056 between Greenlanders and Somalis.

The STR allele frequencies obtained in this study for Somalis, Danes and Greenlanders were compared with STR allele frequencies from other populations. Pairwise F_{ST} values calculated for each STR system are shown in Supplementary material 2. The two highest F_{ST} values were observed for the TH01 system between Greenland and CEPH-Oceania (0.337) and for the D2S441 system between Morocco and CEPH-Oceania (0.299). Pairwise F_{ST} values were also calculated for 15 of the 16 STR systems (excluding SE33) and presented in an MDS plot (Fig. 1). The populations in Greenland, CEPH-America and CEP-Oceania showed a very distant position from any of the other populations included in the analysis (Fig. 1a). In order to better observe the relative location of the other 16 populations, an MDS plot was drawn excluding the populations of Greenland, CEPH-America and CEPH-Oceania (Fig. 1b). As shown in Fig. 1b, the 16 populations included in the MDS plot tend to group according to their geographic origin. Both Somalis and Danes group together with other Africans and Europeans, respectively. As expected, the genetic heterogeneity among African populations is higher than that observed among Europeans. Populations from Middle East, South America and North Africa present intermediate positions being closer to Europeans than to Asians or Africans.

Statistical parameters of forensic interest calculated for each STR and population are shown in Supplementary material 1. The combined CE was 99.9999971% for Somalis, 99.9999987% for

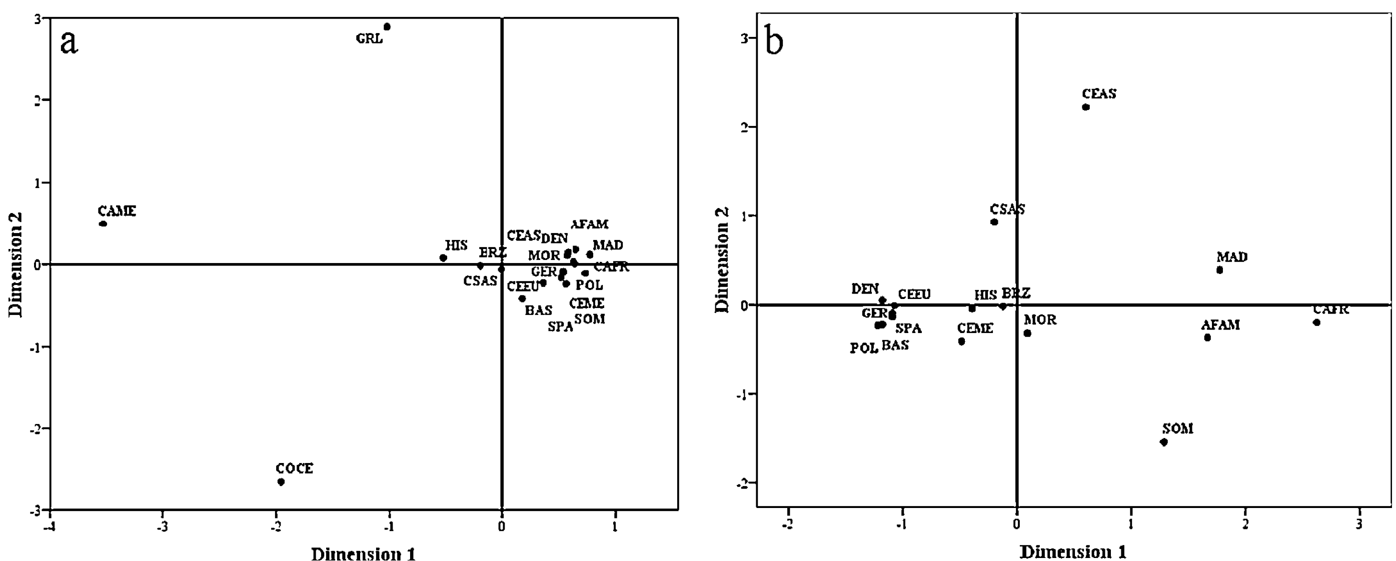


Fig. 1. MDS plot drawn from pairwise F_{ST} values calculated from allele frequencies of 15 autosomal STR loci in 19 (a) and 16 populations (b). Stress = 0.148 (a) and Stress = 0.086 (b). Population abbreviations are as follows: CAFR: CEPH Africa; MAD: Madagascar; SOM: Somalia; AFAM: Afro American; MOR: Morocco; CEME: CEPH Middle East; CEEU: CEPH Europe; DEN: Denmark; GER: Northern Germany; POL: Poland; SPA: Spain; BAS: Basques; CAME; CEPH America; HIS: Hispanics; BRZ: Brazil; CEAS: CEPH East Asia; CSAS: CEPH South Asia; GRL: Greenland; COCE: CEPH Oceania.

Danes and 99.99984% for Greenlanders. The combined PD values ranged from a value of 1 in 6.5×10^{16} in Greenlanders to 1 in 1.8×10^{21} in Danes. The typical paternity index estimated for duos was 1.81×10^{6}, 2.05×10^{6} and 9.48×10^{4} in Somalis, Danes and Greenlanders, respectively. The estimated typical paternity index for trios was 1.43×10^{9} in Somalis, 1.66×10^{9} in Danes and 27.4×10^{6} in Greenlanders.

This short communication follows the guidelines for publication of population data requested by the journal [21].

Acknowledgements

We thank Pernille Jensen for technical assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.04.004.

References

[1] J.M. Butler, C.R. Hill, Biology and genetics of new autosomal STR loci useful for forensic DNA analysis, Forensic Sci. Rev. 24 (2012) 15–26.
[2] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98.
[3] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.
[4] C. Phillips, L. Fernandez-Formoso, M. Garcia-Magarinos, L. Porras, T. Tvedebrink, J. Amigo, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, A. Freire-Aradas, A. Gomez-Carballa, A. Mosquera-Miguel, A. Carracedo, M.V. Lareu, Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel, Forensic Sci. Int. Genet. 5 (2011) 155–169.
[5] M. Poetsch, Z. Ergin, K. Bayer, D. El-Mostaqim, N. Rakotomavo, E.N. Browne, C. Timmann, R.D. Horstmann, T. Schwark, N. von Wurmb-Schwark, The new Powerplex(R) ESX17 and ESI17 kits in paternity and maternity analyses involving people from Africa—including allele frequencies for three African populations, Int. J. Legal Med. 125 (2011) 149–154.
[6] B. Budowle, J. Ge, R. Chakraborty, A.J. Eisenberg, R. Green, J. Mulero, R. Lagace, L. Hennessy, Population genetic analyses of the NGM STR loci, Int. J. Legal Med. 125 (2011) 101–109.
[7] M. Poetsch, K. Bayer, Z. Ergin, M. Milbrath, T. Schwark, N. von Wurmb-Schwark, First experiences using the new Powerplex(R) ESX17 and ESI17 kits in casework analysis and allele frequencies for two different regions in Germany, Int. J. Legal Med. 125 (2011) 733–739.
[8] A. Parys-Proszek, T. Kupiec, P. Wolanska-Nowak, W. Branicki, Genetic variation of 15 autosomal STR loci in a population sample from Poland, Leg. Med. (Tokyo) 12 (2010) 246–248.
[9] O. Garcia, J. Alonso, J.A. Cano, R. Garcia, G.M. Luque, P. Martin, I.M. de Yuso, S. Maulini, D. Parra, I. Yurrebaso, Population genetic data and concordance study for the kits Identifiler, NGM, PowerPlex ESX 17 System and Investigator ESSplex in Spain, Forensic Sci. Int. Genet. 6 (2012) e78–e79.
[10] I. Yurrebaso, J.A. Ajuriagerra, A. Alday, I. Lezama, J.A. Perez, E. Romon, I. Uriarte, O. Garcia, Allele frequencies and concordance study between the Identifiler and the PowerPlex ESX17 systems in the Basque Country population, Forensic Sci. Int. Genet. 5 (2011) e79–e80.
[11] P.E. Raimann, G. Hirschfeld-Campolongo, S. Schumacher, S.E. de Mattos Almeida, C.M. da Silva, R. Rodenbusch, A.S. Bassini, L.F. de Souza, E.C. Nascimento, E.S. de Oliveira, V.R. Santos, E.C. de Oliveira, E.A. Auler-Bittencourt, R.C. da Rocha Hirschfeld-Campolongo, E.S. ML, H.P. de Figueiredo, B.B. de Figueiredo, R. Silva, R.S. Moura-Neto, Population genetic analyses of the AmpFlSTR(R) NGM in Brazil, Int. J. Legal Med. 126 (2012) 337–341.
[12] N. Takezaki, M. Nei, K. Tamura, POPTREE2. Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface, Mol. Biol. Evol. 27 (2010) 747–752.
[13] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (1979) 65–70.
[14] R.A. Fisher, Standard calculations for evaluating a blood-group system, Heredity 5 (1951) 95–102.
[15] Y. Ohno, I.M. Sebetan, S. Akaishi, A simple method for calculating the probability of excluding paternity with any number of codominant alleles, Forensic Sci. Int. 19 (1982) 93–98.
[16] C. Brenner, J. Morris, Paternity index calculation in single locus hypervariable DNA probes: validation and other studies, in: Proceedings of the International Symposium on Human Identification, Memorial Genetics Center, Long Beach, CA, (1989), pp. 21–53.
[17] DNA VIEW™: http://dna-view.com/dnaview.htm.
[18] N. Mizuno, T. Kitayama, K. Fujii, H. Nakahara, K. Yoshida, K. Sekiguchi, N. Yonezawa, M. Nakano, K. Kasai, A D19S433 primer binding site mutation and the frequency in Japanese of the silent allele it causes, J. Forensic Sci. 53 (2008) 1068–1073.
[19] D.Y. Wang, R.L. Green, R.E. Lagace, N.J. Oldroyd, L.K. Hennessy, J.J. Mulero, Identification and secondary structure analysis of a region affecting electrophoretic mobility of the STR locus SE33, Forensic Sci. Int. Genet. 6 (2012) 310–316.
[20] C. Davis, J. Ge, J. King, N. Malik, V. Weirich, A.J. Eisenberg, B. Budowle, Variants observed for STR locus SE33: a concordance study, Forensic Sci. Int. Genet. 6 (2012) 494–497.
[21] A. Carracedo, J.M. Butler, L. Gusmao, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, New guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 7 (2013) 217–220.