International Congress Series 1239 (2003) 53–56

Estimating the ethnic origin (EEO) of individuals using short tandem repeat loci of forensic relevance

M. Klintschar[a,*], S. Füredi[b], B. Egyed[b], B. Reichenpfader[c], M. Kleiber[a]

[a] *Institut für Rechtsmedizin, Martin Luther Universitat Halle-Wittenberg, Franzosenweg 1, D06108 Halle (Saale), Germany*
[b] *Department of Biology, Institute for Forensic Sciences, P.O. Box 314/4, H-1903 Budapest, Hungary*
[c] *Institut für Gerichtsmedizin, Universität Graz, Universitätsplatz 4, A8010 Graz, Austria*

Keywords: Short tandem repeat locus; Estimating the ethnic origin; Forensic DNA profiling; Polymerase chain reaction

1. Introduction

DNA analysis using autosomal short tandem repeat polymorphisms (STRs) is a powerful tool for forensic purposes such as human identification, stain analysis and paternity testing [1,2]. In the last years, national DNA databases comprising up to 13 of these loci were introduced in most countries in Europe and Northern America [3,4]. It has therefore become routine forensic practice to type these loci even without a suspect to compare the profiles. In many instances, however, a database search for a stain that was deposited by an unknown perpetrator at a crime scene is bound to reveal no match. The aim of our study therefore was to investigate whether it is possible to use a large set of autosomal STRs—like the 13 CODIS loci [4]—for estimating the ethnic origin (EEO) of a perpetrator under varying conditions. More precisely the percentage of correct estimations of ethnic origin in three scenarios was tested: for a person who is (1) either Austrian or Egyptian; (2) belongs to one of two Hungarian populations: one from Budapest, the other a Romany minority from Baranya county; and (3) belongs to one of four New York populations: US Caucasians, Afro-Americans, Asians and Hispanics.

---

[*] Corresponding author. Tel.: +49-345-557-1833; fax: +49-345-557-1587.
E-mail address: michael.klintschar@medizin.uni-halle.de (M. Klintschar).

PII: S0531-5131(02)00513-7

2. Materials and methods

STR allelic frequencies for 500 unrelated Austrians and 140 Egyptians (19 STRs), 223 Hungarians from Budapest and 135 Hungarian Romanies from Baranya county (17 STRs), and 125 US Caucasians, 127 Afro-Americans, 126 Asians and 127 Hispanics from New York City (the 13 CODIS core loci) were determined using either singleplex or multiplex reactions with published protocols. The loci applied were D5S818, D13S317, D7S820, D3S1358, VWA, D8S1179, FGA, D21S11, D18S51, TPOX, CSF1PO, TH01, D16S359, CD4, F13B, FES, D12S391, D8S1132, F13A1, LPL, F13A01, D1S1656 and D17S976.

3. Statistical analysis

Several studies to evaluate the feasibility of EEO based on likelihood ratios of the frequencies of DNA profiles in possible original populations [5–8] have been published up to now. The main disadvantage of these approaches is the fact that no information concerning the expected percentage of correct vs. incorrect estimations, that is, the reliability of the procedure, is revealed. We therefore chose a different approach.

In a first step allelic frequencies for 13–19 STRs were established for eight populations in the three scenarios. To simulate the EEO procedure, each individual, one at a time, was removed in a step-by-step procedure from the reference databases and the allelic frequencies were recalculated. For every selected individual, EEO was performed using these newly formed databases by simply choosing that of the two to four possible original populations in which the profile is most common as the most probable ethnic origin. After every step, the selected individual was redistributed to the respective database and the whole procedure was repeated for every single individual in the population samples. The percentage of correct affiliations was used as a measure to test the reliability of the method in forensic practice.

4. Results and discussion

The percentage of correct affiliations when using 13 STRs and a different choice of original populations is given in Table 1. Of both Hungarian populations, slightly less than 90% of all individuals could be identified. This percentage was only around 80% for the

Table 1
Percentage of correct estimations of ethnic origin (EEO) with 13 autosomal STRs

Hungarian populations	Hungarian Romanies	88.8%
	Budapest Caucasians	85.2%
Europeans/Arabs	Austrians	78.9%
	Egyptians	81.1%
US populations	Asians	84.9%
	Afro-Americans	82.7%
	Caucasians	69.1%
	Hispanics	63.8%

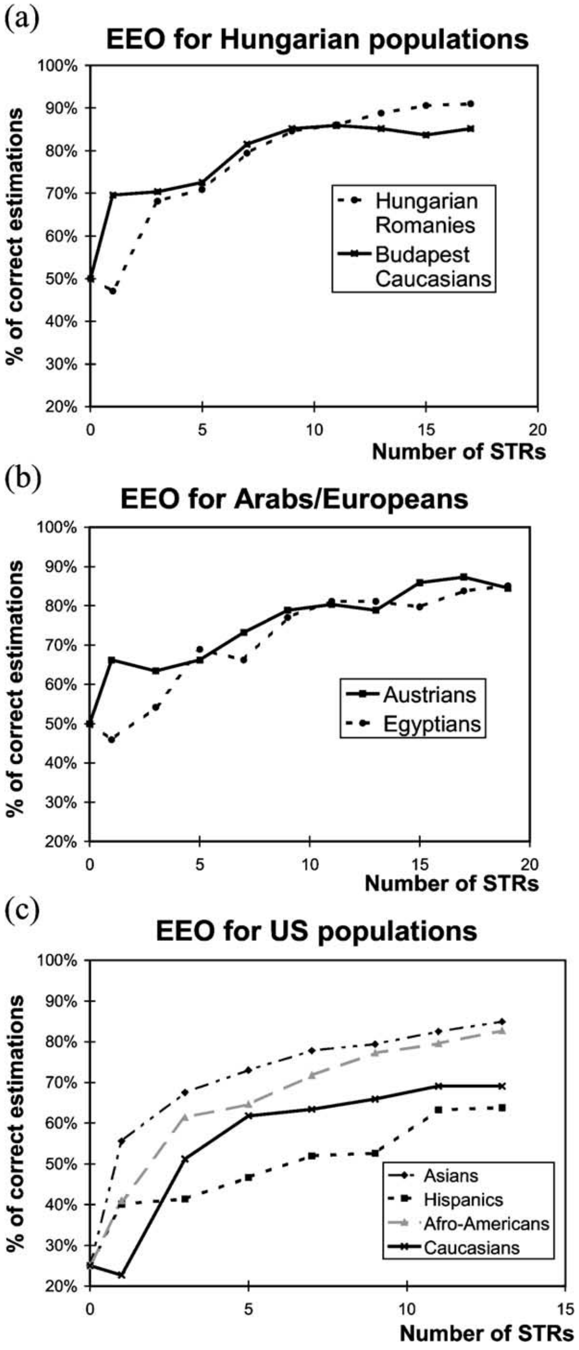

Fig. 1. (a–c) Three graphs showing the percentage of correct estimations of ethnic origin (EEO) with a varying number of loci applied.

Egyptians/Austrians. For the four US populations, the identification of Caucasians and Hispanics were below 70%, but in 93% of all cases, it was correct to state that an individual is either Caucasian or Hispanic.

The percentage of correct vs. incorrect estimations increased when using more loci (Fig. 1a–c). Although even a profile comprising only eight loci could yield valuable information, optimal results should be obtained when using 16–18 STRs. By then the slope of the curve has already flattened out and further increasing the number of loci should yield no additional information concerning the ethnicity of a person in question: When extrapolating the curves in Fig. 1a–c, it can be predicted that it is highly unlikely to correctly identify the ethnic origin in more than 90% even when increasing the number of loci indefinitely. The fact that the "ceiling" of the curve is well below 100% reflects in our eyes the presence of persons of mixed or foreign ancestry in the reference populations: It is obviously difficult to collect ethnically undiluted population samples in modern, globalized societies.

In conclusion, EEO using forensically relevant STRs appears to be a promising approach, especially as no further laboratory procedures are required and DNA database profiles can be employed. Major prerequisites for employing EEO should, however, be that it is (due to the population structure of the area in which the crime has been committed or circumstantial evidence) approximately equally probable that the unknown perpetrator comes from either of two or more genetically distinct population groups and that the investigators are aware of a relatively high rate of wrong estimations.

Acknowledgements

The authors thank M. Prinz, OCME New York, for providing the US-population data.

References

[1] M. Benecke, DNA typing in forensic medicine and in criminal investigations: a current survey, Naturwissenschaften 84 (1997) 181–188.

[2] B. Brinkmann, The use of STRs in stain analysis, Proceedings from the Third International Symposium on Human Identification, Promega, Madison, 1992, pp. 357–373.

[3] P.M. Schneider, DNA databases for offender identification in Europe—the need for technical, legal and political harmonization, Proceedings from the Second European Symposium on Human Identification, Promega, Madison, 1998, pp. 40–44.

[4] B. Budowle, CODIS and PCR-based short tandem repeat loci: law enforcement tools, Proceedings from the Second European Symposium on Human Identification, Promega, Madison, 1998, pp. 73–88.

[5] C.H. Brenner, Probable race of a stain donor, Proceedings from the Sixth International Symposium on Human Identification, Promega, Madison, 1997, pp. 48–52.

[6] I.W. Evett, R. Pinchin, C. Buffery, An investigation of the feasibility of inferring ethnic origin from DNA profiles, J. Forensic Sci. Soc. 32 (1992) 301–306.

[7] P. Smouse, C. Chevillon, Analytical aspects of population-specific DNA fingerprinting for individuals, J. Hered. 89 (1998) 143–150.

[8] E. Meyer, P. Wiegand, B. Brinkmann, Phenotype differences of STRs in 7 human populations, Int. J. Legal Med. 107 (1995) 314–322.