Int J Legal Med (2004) 118 : 111–114
DOI 10.1007/s00414-003-0401-8

SHORT COMMUNICATION

Michael Korostishevsky · Ron Loewenthal
Yelena Slomov · Ephraim Gazit

Erroneous identification in a mixed population: simulation using Israeli STR data

Received: 25 June 2003 / Accepted: 2 September 2003 / Published online: 24 December 2003

Abstract Allele distributions of 10 short tandem repeat (STR) polymorphic DNA loci used in forensic and paternity testing were determined for a cohort comprising 163 individuals representing a mixed Jewish Caucasian population. Typing was carried out by the commercial AmpF/STR SGM Plus kit. The polymorphism and the utility of three of these markers for forensic studies in Israel were established for the first time. Results were compared with data for U.S. Caucasians and African Americans. The probability of identity of two persons of different ethnic origins for identification purposes is discussed. A lemma is presented to show that the chance of erroneous identification of an innocent person who belongs to a population that had not committed a crime will, in most cases, be smaller than for those who belong to a population that had truly committed the crime.

Introduction

Short tandem repeat loci are repetitive sequence elements, 3–7 base pairs in length, which are abundantly distributed throughout the human genome. PCR-based STR analysis is increasingly being used as a means for human identification for forensic and linkage studies (Edwards et al. 1991, 1992; Evett et al. 1996; Hammond et al. 1994). This analysis is based on data concerning the prevalence of alleles at different loci in different ethnic groups. This report presents and compares the allelic frequencies of 10 STR loci in the Israeli Jewish population with that of U.S. Caucasians and African Americans. Furthermore we present simulations concerning the utility of the use of statistical data applied from a data set established for one population to that of a different population when multiple loci are determined. From a practical point of view the question was to determine whether the chance to erroneously identify an innocent individual is increased if STR frequency data determined in a different ethnic group are used. Allelic distribution of 3 of the 10 loci (D16S539, D2S1338 and D19S433) are reported for the first time. This report extends the previous studies (Amar et al. 1999; Motro et al. 2002; Picornell et al. 2002) and accommodates the STR loci data base presently utilized in Israel for forensic studies and paternity determination.

Materials and methods

Whole blood was obtained from 163 unrelated Israeli Jewish individuals and DNA was extracted with the QIAamp DNA mini kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's instructions. DNA was amplified by PCR using the reagents supplied in the AmpF/STR SGM Plus kit (Perkin Elmer Biosystems, Foster City, CA) for the following STR loci: D3S1358, vWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01, and FGA (Cotton et al. 2000). The products were separated on an Applied Biosystems, ABI prism 310 genetic analyzer and analyzed using the software supplied by the manufacturer.

Statistical analysis

Allele frequencies were determined by direct counting of the number of alleles at each locus. Testing for Hardy-Weinberg equilibrium was carried out by the Arlequin software for population genetics data analysis version 2.0 (http://anthropology.unige.ch/arlequin/).

Comparisons for allele frequencies of the different loci in the U.S. Caucasian and African American populations with the Israeli population were determined by the χ^2-test for independent samples as previously described (Siegel 1956; Komlos et al. 1997). Data for the U.S. population was provided by the vendors of the commercial kit (see materials and methods section).

Drs. Korostishevsky and Loewenthal are joint first co-authors.

M. Korostishevsky
Department of Human Genetics, Tel-Aviv University, Israel

R. Loewenthal (✉) · Y. Slomov · E. Gazit
Tissue Typing Laboratory Sheba Medical Center,
52621 Tel-Hashomer, Israel
Tel.: +972-3-5302829, Fax: +972-3-5345964,
e-mail: ronl@sheba.health.gov.il

M. Korostishevsky · R. Loewenthal · E. Gazit
Sackler School of Medicine, Tel-Aviv University, Israel

Definitions

– $I_m(A)$: the intra-population identity is the probability that two individuals randomly selected from the same population (*A*) will have an identical genotype at marker *m*:

$$I_m(A) = 2\sum_i \sum_{j\neq i} (p_{Ai}p_{Aj})^2 + \sum_i (p_{Ai})^4 = 2(1-h_A) - \sum_i (p_{Ai})^4$$

where p_{Ai} is the frequency of allele *i* in population A, h_A is the heterozygosity of the marker.

– $I_m(A,B)$: the inter-population identity is the probability that two individuals randomly selected from two different populations (A and B) will have an identical genotype at marker *m*:

$$I_m(A,B) = 2\sum_i \sum_{j\neq i} (p_{Ai}p_{Aj})(p_{Bi}p_{Bj}) + \sum_i (p_{Ai}p_{Bi})^2 = 2\omega^2 - \sum_i v_i{}^2$$

where $v_i = p_{Ai}p_{Bi}$ and $\omega = \sum_i v_i$.

Note that the definition of intra-population identity is the same as that of the regularly used probability of identity (matching probability).

Table 1 Allele frequencies at each of the 10 loci for an Israeli population sample (*n*=163)

Allele	D3S1358	VWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D19S433	TH01	FGA
5			0.0031						0.0031	
6									0.2791	
7									0.1350	
8			0.0460		0.0061				0.1411	
9			0.1411		0.0031				0.2270	
9.3									0.1902	
10			0.0890		0.0706				0.0245	
11	0.0031		0.3129		0.0767		0.0153	0.0092		
12		0.0031	0.2362		0.1227		0.1564	0.0951		
13	0.0123		0.1411		0.3282		0.1871	0.2301		
13.2							0.0031	0.0337		
14	0.0583	0.0644	0.0245		0.2147		0.1840	0.2699		
14.2							0.0061	0.0491		
15	0.2178	0.1227	0.0031		0.1288		0.1288	0.1472		
15.2								0.0736		
16	0.2485	0.2301		0.0399	0.0460		0.1043	0.0552		0.0031
16.2								0.0215		
17	0.2914	0.3190	0.0031	0.2822	0.0031		0.0767	0.0061		
17.2								0.0061		
18	0.1534	0.1595		0.0675			0.0798			0.0092
18.2								0.0031		
19	0.0153	0.0859		0.1043			0.0276			0.0644
20		0.0123		0.1380			0.0153			0.1135
21		0.0031		0.0491			0.0031			0.1595
22				0.0245			0.0123			0.1902
22.2										0.0031
23				0.0951						0.1748
24				0.1135						0.1564
24.2										0.0031
25				0.0675		0.0031				0.0675
25.2						0.0031				
26				0.0153		0.0031				0.0337
27				0.0031		0.0184				0.0061
28						0.1196				0.0153
29						0.2454				
29.2						0.0031				
30						0.2393				
30.2						0.0215				
31						0.0153				
31.2						0.0982				
32						0.0184				
32.2						0.1380				
33.2						0.0583				
34						0.0031				
34.2						0.0031				
35						0.0092				

Results and discussion

The allele frequencies at each of the loci for the Israeli population are given in Table 1. The allele frequencies of D3S1358, D8S1179, D21S11, D18S51, FGA, vWA and TH01 found in this study, are similar to those previously estimated in Israel (Amar et al. 1999; Motro et al. 2002; Picornell et al. 2002). The allele frequencies of D16S539, D2S1338 and D19S433 are reported for the first time.

The number of alleles, the effective number of alleles, heterozygosity values and the Hardy-Weinberg equilibrium (HWE) test *p*-values are summarized in Table 2 for each of the loci.

The described markers show marked utility for forensic and paternity determination in Israel due to the following findings:

1. These markers are all in HWE i.e. no significant deviation from HWE was found: *p*-values >0.1 for all markers, therefore regular statistical analyses are applicable.
2. All markers are highly polymorphic: the number of alleles ranges from 8 to 17, the effective number of alleles is always greater than four and heterozygosity is not less than 0.76.
3. The power of discrimination is extremely high (not less than 0.916).

In addition, as is shown in Table 4, the combined matching probability is 1.84×10^{-13}. This implies that the error in identification, based on this marker set, is less than 1 divided by the number of humans alive.

Comparisons between allele frequencies in the African American, U.S. Caucasian and Israeli populations, based on the χ^2-test, for each of the different loci are given in Table 3. Results show that the number of loci which are significantly different with regard to allele frequencies is larger when ethnically distant populations are compared. The African American and Israeli comparison shows a *p*-value <0.01 for 9 of the 10 loci. When a comparison is made between the less distant U.S. Caucasian and Israeli populations *p*-values <0.01 are found in 3 of the 10 loci.

A comparison of intra-population and inter-population probability of identities is presented in Table 4. The results show that for each locus the inter-population identity (columns 5 and 6) is always less than the mean intra-population identity (columns 2–4) for a compared pair. Furthermore, in most cases the inter-population identity is even less than each of the intra-population identities (marked in bold). For distant populations (Israeli and African American) this is true for all loci. The mathematical evidence that for any two populations the inter-population identity is not greater than the average intra-population identity is given in the Appendix. The greater the difference in allele frequencies between two populations, the smaller will be the inter-population identity. This means that the chance of erroneous identification of an innocent person who belongs to a population that had not committed a crime will, in most cases, be smaller than for persons who belong to

Table 2 Allelic distribution parameters for the 10 STR loci in the Israeli population sample

Locus	Number of alleles	Effective number of alleles*	Observed heterozygosity	Power of discrimination**	H-W-E *p*-value
D3S1358	8	5.435	0.816	0.916	0.342
VWA	9	6.803	0.853	0.928	0.326
D16S539	10	4.184	0.761	0.930	0.861
D2S1338	12	6.803	0.853	0.965	0.237
D8S1179	10	6.803	0.853	0.935	0.591
D21S11	17	8.130	0.877	0.953	0.970
D18S51	14	7.092	0.859	0.967	0.113
D19S433	13	5.102	0.804	0.951	0.533
TH01	7	4.405	0.773	0.927	0.403
FGA	15	6.536	0.847	0.965	0.503

**K* The effective number of alleles is defined as K=1/(1–h), where h is the heterozygosity.
***Power of discrimination*= 1–probability of identity.

Table 3 Comparisons of allele frequencies for Israeli vs. U.S. Caucasian and African American populations

Locus	U.S. Caucasian			African American		
	χ^2	d.f.[a]	Significance[b]	χ^2	d.f.[a]	Significance[b]
D3S1358	16.90	8	*	49.98	10	***
VWA	18.31	8	*	38.24	11	***
D16S539	24.97	9	**	7.06	9	NS
D2S1338	22.52	13	*	84.37	12	***
D8S1179	9.47	9	NS	42.04	9	***
D21S11	34.94	18	**	75.91	21	***
D18S51	17.29	15	NS	120.69	16	***
D19S433	19.31	12	NS	67.08	16	***
TH01	34.20	6	***	92.72	7	***
FGA	16.21	14	NS	36.11	18	**

[a]*d.f.* degrees of freedom.
[b]*NS* non-significant.
**p*-value<0.05.
***p*-value<0.01.
****p*-value<0.001.

Table 4 Intra and inter-polpuation probability of identities

Locus	Israel	U.S. Caucasian	African American	Israel and US Caucasian [a]	Israel and African American [a]
D3S1358	0.084	0.078	0.102	**0.077**	**0.081**
VWA	0.072	0.065	0.058	**0.063**	**0.050**
D16S539	0.070	0.103	0.066	0.080	**0.066**
D2S1338	0.035	0.024	0.021	0.027	**0.018**
D8S1179	0.065	0.067	0.075	**0.065**	**0.058**
D21S11	0.047	0.045	0.033	**0.043**	**0.032**
D18S51	0.033	0.030	0.028	**0.030**	**0.017**
D19S433	0.049	0.078	0.039	0.059	**0.038**
TH01	0.073	0.094	0.102	**0.071**	**0.052**
FGA	0.035	0.036	0.035	**0.034**	**0.032**
Combined	1.84E-13	2.99E-13	7.91E-14	1.22E-13	1.05E-14

[a]Inter-population probability values that are less than either of the intra-population probability values are marked in bold.

a population that had truly committed the crime. To the extent that if more unrelated markers are used, this assumption is even more likely.

Appendix

Lemma

The inter-population identity is no more than the average of intra-population identities:

$$I_m(A,B) \le \frac{I_m(A)+I_m(B)}{2}$$

Evidence follows from the equations:

$$\begin{aligned}
&2I_m(A,B)-I_m(A)-I_m(B)\\
&= \sum_i \left[p_{Ai}{}^2(p_{Bi}{}^2-p_{Ai}{}^2)+p_{Bi}{}^2(p_{Ai}{}^2-p_{Bi}{}^2)\right]\\
&\quad +\sum_i\sum_{j\neq 1}\left[4p_{Ai}p_{Aj}p_{Bi}p_{Bj}-2p_{Ai}{}^2-p_{Aj}{}^2-2p_{Bi}{}^2-p_{Bj}{}^2)\right]\\
&= -\sum_i (p_{Bi}{}^2-p_{Ai}{}^2)^2-2\sum_i\sum_{j\neq 1}(p_{Bi}p_{Bj}-p_{Ai}p_{Aj})^2\\
&= -\delta_{A,B}\le 0
\end{aligned}$$

The last expression is equal to zero if and only if the same allele frequencies exist in both populations. For any difference in allele frequency between the populations, the value of this expression is less than zero.

A simple consequence is that the inter-population identity for any two populations with different allele frequency sets is less than the average of the intra-population identities.

References

Amar A, Brautbar C, Motro U, Fisher T, Bonne-Tamir B, Israel S (1999) Genetic variation of three tetrameric tandem repeats in four distinct Israeli ethnic groups. J Forensic Sci 44:983–986

Cotton EA, Allsop RF, Guest JL et al. (2000) Validation of the AMPFlSTR SGM plus system for use in forensic casework. Forensic Sci Int 112:151–161

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746–756

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12: 241–253

Evett IW, Gill PD, Scrange JK, Weir BS (1996) Establishing the robustness of short-tandem-repeat statistics for forensic applications. Am J Hum Genet 58:398–407

Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55:175–189

Komlos L, Korostishevsky M, Halbrecht I, Vardimon D, Ben-Rafael Z, Klein T (1997) Possible sex-correlated transmission of maternal class I HLA haplotypes. Eur J Immunogenet 24: 169–177

Motro U, Oz C, Adelman R et al. (2002) Allele frequencies of nine STR loci of Jewish and Arab populations in Israel. Int J Legal Med 116:184–186

Picornell A, Tomas C, Jimenez G, Castro JA, Ramon MM (2002) Jewish population genetic data in 20 polymorphic loci. Forensic Science Int 125:52–58

Siegel S (1956) The χ^2 test for two independent samples. In nonparametric statistics for the behavioral sciences. McGraw Hill, New York, pp 104