Int J Legal Med (2001) 114 : 211–214 

ORIGINAL ARTICLE

**Michael Klintschar · Nabil Al-Hammadi**
**Barbara Reichenpfader**

# Significant differences between Yemenite and Egyptian STR profiles and the influence on frequency estimations in Arabs

Received: 14 November 1999 / Accepted: 5 April 2000

**Abstract** A population genetic study was performed on Yemenites using the set of nine short tandem repeat loci (STRs) D3S1358, VWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820. Analysis of the data revealed that all loci were in Hardy-Weinberg equilibrium and evidence of linkage equilibrium was found for only 1 out of 36 locus pairs. At seven loci the allelic distributions found in the Yemenite sample were significantly different from those found for an Arab population sample from Egypt. Nevertheless, we assume that the Yemenite database can be used for Arabs of unknown or foreign (non-Yemenite) origin in the absence of population-specific databases without exerting a significant bias on the biostatistical interpretation. In an experimental set-up (ethnic profile frequency ratio test), the impact of calculating multi-locus profile frequencies for foreign Arab individuals (Egyptians) using the Yemenite database instead of a region-specific one was negligible.

**Keywords** Forensics · DNA · Ethnic profile · Frequency ratio test · Short tandem repeat · Yemen

## Introduction

Short tandem repeat loci (STRs, microsatellites) are highly variable DNA polymorphisms that have become powerful tools for forensic stain analysis and paternity testing [1, 2, 3, 4].

The STRs D3S1358, VWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 are among those selected for the US Combined DNA Index System (CODIS) [5] and typed in crime labs throughout the world. It would therefore be of benefit for the forensic community to generate data on the genetic variation of these loci in as many populations as possible. However, data for these nine STRs are not available for Arabs from the Arab peninsula, and the only study published for Arabs has been performed on Egyptians [6]. As for other STRs, significant differences between Yemenites and Egyptians have been observed in many other instances [7, 8, 9], and the goal of this study was to investigate the allelic distribution of the nine STRs in a population sample from Yemen and to investigate whether this database might be valid for forensic casework not only involving Yemenites but also persons from other countries of the Arabic-speaking world.

M. Klintschar (✉)
Department of Legal Medicine, University of Halle,
Franzosenweg 1, D-06112 Halle/Saale, Germany
e-mail: michael.klintschar@medizin.uni-halle.de

B. Reichenpfader
Department of Legal Medicine, University of Graz,
Universitätsplatz 4, 8010 Graz, Austria
e-mail: michael.klintschar@email.kfunigraz.ac.at

N. Al-Hammadi
Department of Forensic Medicine, College of Medicine,
Sanaa University, Sanaa, Yemen

## Materials and methods

### Sample preparation

Whole EDTA blood was obtained by venipuncture from 100 unrelated Yemenites from the Sanaa area whereby 39 of these samples had already been used in other studies [7, 8]. For extracting the DNA, a modified alkaline lysis protocol was applied [10]. Aliquots of 1.25 µl of the extracts were used for amplification without prior quantification of the DNA content. Amplification was performed using the AmpF*l*STR Profiler Plus kit (Perkin-Elmer Applied Biosystems, San Jose, Calif.) according to the manufacturer's recommendations. Typing was performed using an ABI Prism 310 Genetic analyzer (Perkin Elmer Applied Biosystems, San Jose, Calif.).

### Statistical analysis

Forensic efficiency parameters were calculated using the PowerStats Microsoft Excel workbook template provided by Promega Corporation (Madison Wis., http://www.promega.com/geneticidtools/). Exact tests for checking the Hardy-Weinberg expectations and linkage equilibrium were performed using the GENEPOP software version 1.2 (M. Raymond and F. Rousset, Montpellier, http://www.cefe.cnrs-mop.fr/). The allele frequencies between different populations were compared using the R × C software provided by G. Carmody (Ottawa, Canada). To investigate whether the Yemenite database from this study could be applicable to forensic casework involving individuals from all Arab countries, the Yemenite database was compared to a database from Egypt [6] using an ethnic profile frequency ratio test. This approach is simi-

lar to that developed by Meyer et al. [11] for estimating the ethnic origin of individuals and yields estimates of the bias introduced when the Yemenite database is used with foreign Arabic (e.g. Egyptian) individuals. Therefore, 56 Egyptian samples were randomly selected and the genotypes pasted into a Microsoft Excel Spreadsheet as multi-locus profiles. For each allele ($x$) at each STR ($y$) an ethnic ratio $ER(x,y)_{e/ym}$ was formed by dividing the allelic frequencies found in Egypt [$f(x_e,y)$] by those found in Yemen [$f(x_{ym},y)$]:

$$ER(x,y)_{e/ym} = f(x_e,y)/f(x_{ym},y) \quad (1)$$

To avoid divisions by zero, the count of alleles not found in one of the populations was set to one. For each multi-locus profile 18 ratios were thus calculated and multiplied to form an individual ethnic ratio of a multi-locus profile in Yemenites and Egyptians (ERPe/ym):

$$ERP_{e/ym} = \prod_{x,y} ER(x,y)e/ym \quad (2)$$

Logarithms ($\log_{10}ERP_{e/ym}$) were calculated to enable graphic display of the dispersed ratios. A positive value indicates that this particular multi-locus profile is more common in Egyptians, a negative value that the profile is more common in Yemenites. Finally, the geometric mean of the individual ratios was used to estimate the average

**Table 1** Allele frequencies and forensic efficiency parameters for the nine STRs in 100 unrelated Yemenites

| Allele | STR locus | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | D5S818 | D13S317 | D7S820 | D3S1358 | VWA | FGA | D8S1179 | D21S11 | D18S51 |
| 7 | – | 0.005 | – | – | – | – | – | – | – |
| 8 | – | 0.125 | 0.160 | – | – | – | 0.005 | – | – |
| 9 | 0.110 | 0.040 | 0.090 | – | – | – | – | – | – |
| 10 | 0.125 | 0.075 | 0.320 | – | – | – | 0.045 | – | 0.015 |
| 11 | 0.235 | 0.255 | 0.240 | – | – | – | 0.115 | – | 0.015 |
| 12 | 0.385 | 0.415 | 0.150 | 0.005 | – | – | 0.215 | – | 0.140 |
| 13 | 0.130 | 0.060 | 0.005 | – | – | – | 0.230 | – | 0.140 |
| 14 | 0.010 | 0.025 | 0.025 | 0.075 | 0.050 | – | 0.160 | – | 0.225 |
| 15 | 0.005 | – | 0.005 | 0.195 | 0.120 | – | 0.165 | – | 0.135 |
| 16 | – | – | – | 0.275 | 0.270 | – | 0.060 | – | 0.060 |
| 17 | – | – | 0.005 | 0.295 | 0.335 | – | 0.005 | – | 0.120 |
| 18 | – | – | – | 0.130 | 0.145 | 0.020 | – | – | 0.090 |
| 18.2 | – | – | – | – | – | 0.005 | – | – | – |
| 19 | – | – | – | 0.025 | 0.065 | 0.015 | – | – | 0.020 |
| 20 | – | – | – | – | 0.010 | 0.090 | – | – | 0.015 |
| 21 | – | – | – | – | 0.005 | 0.140 | – | – | 0.015 |
| 22 | – | – | – | – | – | 0.165 | – | – | 0.005 |
| 23 | – | – | – | – | – | 0.175 | – | – | – |
| 23.2 | – | – | – | – | – | 0.005 | – | – | 0.005 |
| 24 | – | – | – | – | – | 0.235 | – | – | – |
| 25 | – | – | – | – | – | 0.105 | – | – | – |
| 26 | – | – | – | – | – | 0.040 | – | – | – |
| 26.2 | – | – | – | – | – | – | – | 0.005 | – |
| 27 | – | – | – | – | – | – | – | 0.045 | – |
| 28 | – | – | – | – | – | 0.005 | – | 0.145 | – |
| 29 | – | – | – | – | – | – | – | 0.290 | – |
| 30 | – | – | – | – | – | – | – | 0.230 | – |
| 30.2 | – | – | – | – | – | – | – | 0.015 | – |
| 31 | – | – | – | – | – | – | – | 0.040 | – |
| 31.2 | – | – | – | – | – | – | – | 0.060 | – |
| 32 | – | – | – | – | – | – | – | 0.010 | – |
| 32.2 | – | – | – | – | – | – | – | 0.050 | – |
| 33 | – | – | – | – | – | – | – | 0.005 | – |
| 33.2 | – | – | – | – | – | – | – | 0.095 | – |
| 34 | – | – | – | – | – | – | – | 0.010 | – |
| Forensic parameters | | | | | | | | | |
| Matching probability | 0.104 | 0.109 | 0.091 | 0.100 | 0.098 | 0.050 | 0.068 | 0.071 | 0.049 |
| Expressed as 1 in ... | 9.581 | 9.179 | 10.990 | 9.966 | 10.203 | 19.892 | 14.714 | 14.149 | 20.351 |
| Power of discrimination | 0.896 | 0.891 | 0.909 | 0.900 | 0.902 | 0.950 | 0.932 | 0.929 | 0.951 |
| PIC | 0.716 | 0.697 | 0.751 | 0.741 | 0.740 | 0.828 | 0.806 | 0.802 | 0.852 |
| Power of exclusion | 0.455 | 0.407 | 0.397 | 0.518 | 0.506 | 0.666 | 0.723 | 0.747 | 0.737 |
| Typical paternity index | 1.768 | 1.592 | 1.558 | 2.040 | 1.984 | 3.031 | 3.679 | 4.042 | 3.875 |
| Homozygosity rate | 0.283 | 0.314 | 0.321 | 0.245 | 0.252 | 0.165 | 0.136 | 0.124 | 0.129 |
| Heterozygosity rate | 0.717 | 0.686 | 0.679 | 0.755 | 0.748 | 0.835 | 0.864 | 0.876 | 0.871 |

impact on the burden of proof if the Yemenite database is used on casework involving Egyptians and if the Yemenite database might be applicable to casework involving Arabs of unknown origin.

## Results and discussion

The observed allele frequencies for the nine STR loci found in Yemen and the respective forensic efficiency parameters are shown in Table 1. All loci were in Hardy-Weinberg equilibrium ($p > 0.05$). An exact test demonstrated that there is little evidence of correlation between the alleles at any of the pairs of loci tested. There was only one example of departure from expectations (D21S11/D8S1179) out of a total of 36 pairs, which is less than 5% of the comparisons. Thus the degree of departure was not beyond expectations.

Carmody's R × C test and the ethnic profile frequency ratio test both deal with population heterogeneity from a different point of view: the R × C test is a sensitive approach to check the level of significance for population heterogeneity. The ethnic profile frequency ratio test, on the other hand, is aimed at estimating the bias introduced into frequency estimation when using an incorrect reference database. In the R × C test the Yemenite population differs substantially from another Arab sample from Egypt [6] at seven out of nine loci which is the same as for a comparison between Yemenites and Austrians [12] (Table 2). Our study thus suggests distinct genetic heterogeneity within the Arab-speaking world. As no other Arab populations have yet been studied, the forensic community might nevertheless have to apply the Yemenite (or the Egyptian) database to casework involving individuals from other Arabic countries. This practice could lead to overestimating the evidence against the suspect [13]. However, as the potential bias introduced by using the incorrect database in casework involving Arabs is not known it would be of interest to evaluate the influence on Egyptian multi-locus profile frequencies when a Yemenite database is used. Should the impact on multi-locus profile frequencies be insignificant, it might be concluded that the Yemenite database is also applicable to persons of unknown or "foreign" Arabic origin. Therefore, an ethnic profile frequency ratio test was performed on 56 randomly selected individuals from Egypt, the multi-locus profiles for nine STR loci were determined, and the frequency estimated using the (incorrect)Yemenite database and a (correct) Egyptian database [6]. The ethnic ratios of multi-locus profiles ($ERP_{e/ym}$) calculated as described varied between 0.122 and 285.910 ($\log_{10}ERP_{e/ym}$: –0.913 to +2.456) with a geometric mean of 3.068 ($\log_{10}ERP_{e/ym}$: 0.487; Fig. 1). In the other direction (i.e. Yemenites against an Egyptian database), the geometrically expected $ERP_{ym/e}$ value of 3.880 was similar. These values indicate that an Egyptian multi-locus profile comprising the nine STRs can be expected to be on average 3 times more frequent using the Egyptian than using the Yemenite database. Using the Yemenite database for a case involving an Arab from Egypt one might thus expect to introduce a certain bias against the suspect, although in 23% of the cases a bias in favour of the suspect can be expected (Fig. 1). On the other hand, considering an average multi-locus profile frequency of approximately $1 \times 10^{-10}$, one might consider this bias insignificant. This assumption can be illustrated

**Table 2** Comparison between allele distributions in Yemenite (Y, this study), Egyptian (E, [6]) and Austrian populations (A, [12]) (probability values ± standard error, *bold print* indicates statistically significant differences)

| STR | E:Y | SE | E:A | SE | Y:A | SE |
|---|---|---|---|---|---|---|
| D5S818 | **0.007** | 0.0026 | 0.054 | 0.007 | **0.021** | 0.005 |
| D13S317 | **0.001** | 0.001 | 0.346 | 0.015 | 0.444 | 0.016 |
| D7S820 | **0.007** | 0.003 | **0.018** | 0.004 | **0.003** | 0.002 |
| D3S1358 | **0.006** | 0.002 | 0.053 | 0.007 | **0.005** | 0.002 |
| VWA | 0.461 | 0.016 | 0.718 | 0.014 | **0.023** | 0.005 |
| FGA | **<10⁻³** | <10⁻³ | **0.03** | 0.005 | **<10⁻³** | <10⁻³ |
| D8S1179 | **0.01** | 0.003 | **0.004** | 0.002 | **<10⁻³** | <10⁻³ |
| D21S11 | **<10⁻³** | <10⁻³ | **0.001** | 0.001 | **<10⁻³** | <10⁻³ |
| D18S51 | 0.186 | 0.012 | **0.045** | 0.007 | 0.391 | 0.015 |

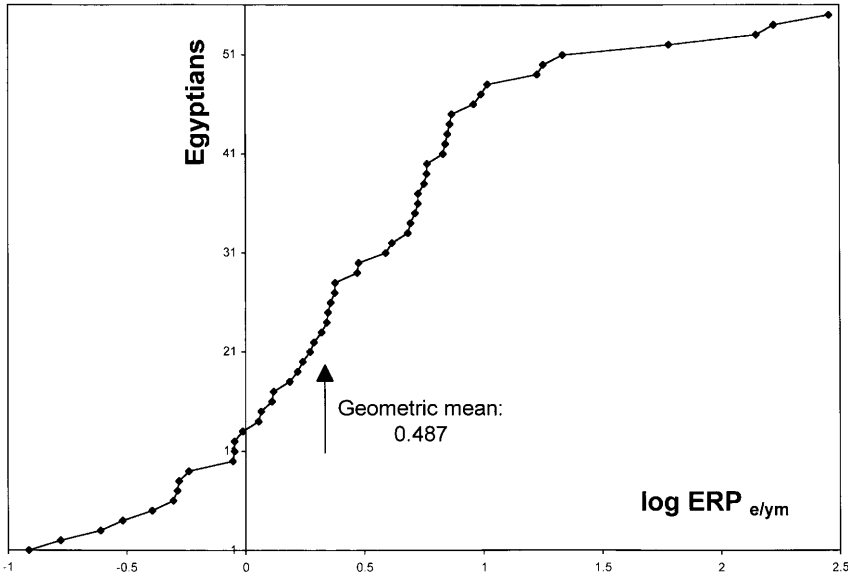


**Fig. 1** Logarithms of the ethnic ratios of a multi-locus profile calculated for 56 randomly selected Egyptian individuals using an Egyptian database [6] and a Yemenite database (this study) (log $ERP_{e/ym}$) (The range of –0.913 to +2.456 corresponds to ratios of 0.122 to 285.910. The geometric mean of the distribution is 0.487 and corresponds to an $ERP_{e/ym}$ – value of 3.068)

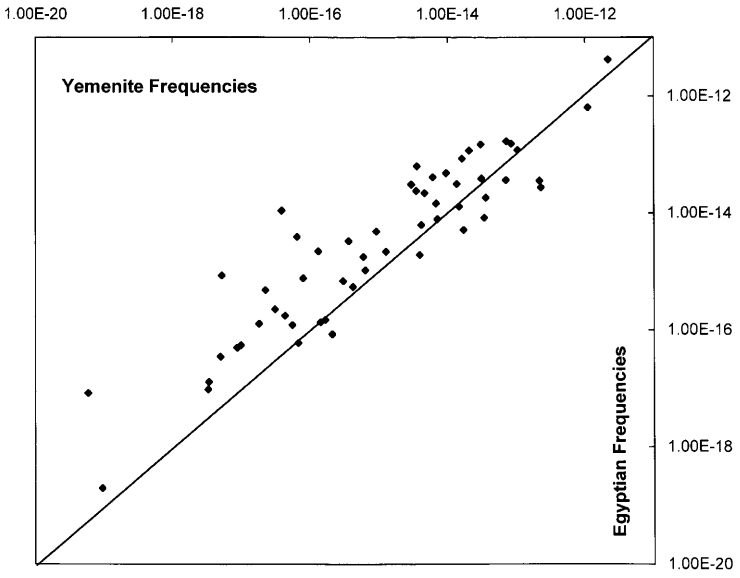


**Fig. 2** Scatterplot of the multi-locus profile frequencies of 56 Egyptian individuals calculated using an Egyptian database (y-axis, [6]) against using the Yemenite database (x-axis, this study)

by a scatterplot of the multi-locus profile frequencies calculated using the Egyptian database against those calculated using the Yemenite database (Fig. 2), which shows that it is of little significance which of these databases is applied. In fact, the largest shift in frequency changed a value of $1.61 \times 10^{-14}$ to $1.97 \times 10^{-17}$, the denominators exceeding the world population by 4–7 orders of magnitude. It seems therefore safe to state that the Yemenite database should allow a sufficiently reliable estimation of the frequency of a multi-locus profile when the suspect originates from a foreign Arab population despite the evident population heterogeneity within the arab-speaking world.

All three populations are caucasoid and from the historical, cultural and geographical point of view one should expect the Yemenites to be more closely related to the Egyptians than to central European Austrians. Surprisingly, however, the geometrically expected ERP values for discriminating Egyptians and Yemenites are higher than those found for Austrians and Egyptians (2.35) and Austrians and Yemenites (3.37; values averaged in both directions, data not shown), thus suggesting a higher diversity within the Arab populations than between Europeans and Arabs. A possible explanation might be the situation of the Arab countries between Africa and Asia. The heterogeneity might thus reflect differences in the amalgamation of subsaharan and caucasoid gene pools to the relatively isolated Yemenite population compared to the more centrally situated Egypt in more recent times or the effect of population migrations during the dispersal of modern humans out of Africa, which is assumed to have taken place through Yemen [14].

In conclusion, the nine STR loci described proved to be suitable and sufficiently polymorphic for most forensic purposes in Arabic populations.

## References

1. 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
2. Benecke M (1997) DNA typing in forensic medicine and in criminal investigations: a current survey. Naturwissenschaften 84: 181–188
3. Alford RL, Hammond HA, Coto I, Caskey CT (1994) Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am J Hum Genet 55: 190–195
4. Brinkmann B (1992) The use of STRs in stain analysis. In: Proceedings of the Third International Symposium on Human Identification. Promega, Madison, Wis., pp 357–373
5. Budowle B (1998) CODIS and PCR-based short tandem repeat loci: Law enforcement tools. In: Proceedings from the Second European Symposium on Human Identification, Promega, Madison, Wis., pp 73–88
6. Klintschar M, Ebner A, Reichenpfader B (1999) Population genetic studies on the tetrameric short tandem repeat loci D3S1358, VWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 in Egypt. Forensic Sci Int 104: 23–31
7. Klintschar M, Al-Hammadi N, Lux T, Reichenpfader B (1998) Genetic variation at the short tandem repeat loci HumVWA, HumFXIIIB, and HumFES/FPS in the Egyptian and Yemenian populations. J Forensic Sci 43: 850–853
8. Klintschar M, Kozma Z, Al Hammadi N, Abdull Fatah M, Nöhammer C (1998) A study on the short tandem repeat systems HumCD4, HumTHO1, and HumFIBRA in a Yemenian and an Egyptian population sample. Int J Legal Med 111: 107–109
9. Klintschar M, Ricci U, Al-Hammadi N, Reichenpfader B, Ebner A, Giovannucci Uzielli ML (1998) Genetic variation at the STR loci D12S391 and CSF1PO in four populations from Austria, Italy, Egypt and Yemen. Forensic Sci Int 97: 37–45
10. Klintschar M, Neuhuber F (2000) Evaluation of an alkaline lysis method for the extraction of DNA from whole blood and forensic stains for STR analysis. J Forensic Sci 44 (in press)
11. Meyer E, Wiegand P, Brinkmann B (1995) Phenotype differences of STRs in 7 human populations. Int J Legal Med 107: 314–322
12. Klintschar M, Ebner A, Reichenpfader B (1999) Population genetic studies on 9 tetrameric short tandem repeat loci using fluorescence dye labeled primers and capillary electrophoresis in the Austrian population. Electrophoresis 20: 1740–1742
13. National Research Council (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington DC
14. Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23: 437–441