Forensic Science International: Genetics 17 (2015) 43–46

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Forensic population genetics – short communication

Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: Persians, Lurs, Kurds and Azeris

L. Poulsen [a,1,*], M. Sharafi Farzad [b], C. Børsting [a], C. Tomas [a], V. Pereira [a], N. Morling [a]

[a] 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
[b] Kawsar Human Genetics Research Center, 41 Majlesi St., Vali Asr St., Tehran, Iran

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form 13 March 2015
Accepted 15 March 2015

Keywords:
Iran
NGM SElect™
DIPplex®
Argus X-12
Forensic genetics

ABSTRACT

A total of 255 individuals (Persians, Lurs, Kurds and Azeris) from Iran were typed for three sets of forensic genetic markers with the NGM SElect™, DIPplex® and Argus X-12 kits. Statistically significant deviations ($P \leq 0.002$) from Hardy–Weinberg expectations were observed for the insertion-deletion markers HLD97 and HLD93 after Holm–Šidák correction. Statistically significant ($P < 0.05$) levels of linkage disequilibrium were observed between markers within two of the four studied X-chromosomal linkage groups. AMOVA analyses of the three sets of markers did not show population structure when the individuals were grouped according to their ethnic group. The Iranian population grouped closely to populations living geographically near to Iran based on pairwise F_{ST} distances. The matching probabilities ranged from 1 in 3.2×10^7 males by using haplotype frequencies of four X-chromosomal haplogroups to 1 in 3.4×10^{21} individuals for the 16 autosomal STRs.

1. Introduction

The Iranian population is ethnically heterogeneous and includes the following ethnic groups: Persians (51%), Azeris (24%), Gilakis and Mazandaranis (8%), Kurds (7%), Arabs (3%), Lurs (2%), Balochis (2%), Turkmen (2%) and others (1%) (proportions estimated in 1999 [1]). Furthermore, the ethnic groups are unevenly distributed in the country with Persians mainly in central Iran, Azeris in the North West, Gilakis and Mazandaranis in the North, Kurds in the West, Arabs and Lurs in the South West, Balochis in the South East and Turkmen in the North East of Iran [2]. Various ethnic groups in Iran have been studied using lineage markers (mitochondrial DNA [3,4], Y-chromosomal short tandem repeats (STRs) [5] and high resolution Y-chromosomal genotyping [6]) and autosomal markers (STRs [7] and single nucleotide polymorphisms (SNPs) [8]).

The purpose of this study was to characterise the diversity of 30 autosomal insertion-deletion polymorphisms (InDels) and 12 X-chromosomal STRs in the Iranian population and to assess their usefulness both in forensic casework and population genetics. Four ethnic groups in Iran were investigated: Kurds, Persians, Lurs and Azeris. Furthermore, data for 16 autosomal STRs are presented in order to update the data in the Iranian population [7].

2. Materials and methods

2.1. Populations samples and DNA extraction

Samples were collected from 255 Iranian individuals as blood on KBC DNA banking cards (DBC™: KBC: Kawsar Biotech Company, Tehran, Iran). The samples were collected at Kawsar Human Genetics Research Center, Tehran, Iran. The individuals were selected using the following criteria: Town of origin, language and the ethnicity of the individuals and their parents. Four ethnic groups were included: Lurs ($N = 76$), Kurds ($N = 70$), Azeris ($N = 58$) and Persians ($N = 51$). The work was approved by the Danish ethical committee (H-1-2011-081).

DNA was extracted from 3.2 mm disks punched from blood stains on DBC™ (KBC) as previously described [8].

2.2. Genotyping and quality control

The 255 samples from Iran were typed for three sets of forensic genetic markers. The number of samples typed for each set of forensic genetic markers was: Autosomal STRs, $N = 214$ (105 males

* Corresponding author. Tel.: +45 3532 6314; fax: +45 3532 6270.
E-mail address: lena.poulsen@sund.ku.dk (L. Poulsen).
[1] URL: http://retsmedicin.ku.dk.

http://dx.doi.org/10.1016/j.fsigen.2015.03.010
1872-4973/

and 109 females), autosomal InDels, $N = 253$ (126 males and 127 females) and X-chromosomal STRs, $N = 251$(127 males and 124 females). All samples were analysed in duplicates.

2.2.1. Autosomal STRs

Sixteen autosomal STR loci and the Amelogenin locus were amplified in one PCR reaction using the AmpFℓSTR® NGMSElect™ kit (AB/LT/TFS: Applied Biosystems™, Life Technologies, Thermo-Fisher Scientific, Waltham, MA USA). The samples were amplified in a 10 μL reaction volume with 25 or 27 cycles.

2.2.2. Autosomal InDels

Thirty autosomal InDels and Amelogenin were amplified in one PCR reaction using the Investigator® DIPplex® kit (Qiagen GmbH, Hilden, Germany). The manufacture's recommendations were followed with the exception that the final PCR amplification volume was 13 μL instead of 25 μL and the total number of PCR cycles was decreased from 30 to 25 or 26 cycles.

2.2.3. X-chromosomal STRs

Twelve X-chromosomal STR markers were amplified using the Investigator® Argus X-12 kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's protocol in a final PCR amplification volume of 12 μL. The total number of PCR cycles varied depending on the quality and quantity of the sample, between 25 and 30 cycles.

The PCR products were electrophoresed on a 3130 xl Genetic Analyzer (AB/LT/TFS). The alleles were assigned using the GeneMapper ID-X v.1.3 software (AB/LT/TFS).

2.2.4. Quality control

The DNA typing was performed at the Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen. The laboratory is accredited according to the ISO 17,025 standard.

2.3. DNA sequencing

Four samples were Sanger-sequenced for the HLD97 InDel marker (rs17238892). Primers for PCR and sequencing were described in [9]. The PCR products were sequenced using the BigDye® Terminator v1.1 cycle sequencing kit (AB/LT/TFS) according to the manufacturer's recommendations. Sequencing products were electrophoresed on a 3130 xl Genetic Analyzer (AB/LT/TFS). The results were analysed with the Sequencing Analysis v.5.2 (AB/LT/TFS) and the Sequencher® 5.1 software [10].

2.4. Population and forensic statistical parameters

Arlequin v. 3.5 software [11] was used to calculate most of the population genetic parameters. Allele and haplotype frequencies as well as parameters of intra-population variability were calculated. A total of 10^6 steps in the Markov chain and 10^6 dememorisation steps were applied to test Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) of data with known gametic phase. The allelic association for data with unknown gametic phase was estimated using 10^5 permutations. The genetic structure of the Iranian population was tested by using two-hierarchical AMOVA tests and pairwise F_{ST} values [12] with 10^3 permutations.

For the X-chromosomal markers, allele and haplotype frequencies were used for the analyses. The 12 analysed X-chromosomal STRs are located in four previously defined linkage groups [13] with three markers per group. Haplotype frequency analyses of the X-chromosomal data were performed considering each of the four linkage groups as a single genetic marker.

Allele/haplotype frequencies estimated for the Iranian population were compared to published frequencies of the populations listed in Supplementary material 2 (Table SM.2.4).

Sample bias corrected F_{ST} distances [14] were calculated from allele/haplotype frequencies using the POPTREE2 software [15]. Due to lack of available data, the SE33 system was not included in the inter-population comparisons. Only 8 out of the 12 studied X-STRs were included in the comparison analyses (DXS10135, DXS8378, DXS7132, DXS10074, HPRTB, DXS10101, DXS10134 and DXS7423).

Multidimensional scaling (MDS) plots were represented from F_{ST} values using IBM SPSS Statistics for Windows, version 22.0 (Released 2013. Armonk, NY: IBM corp.).

Statistical significance for multiple tests was corrected using the Holm–Šidák method [16].

Forensic statistical parameters for autosomal markers were calculated using DNA · VIEW™ version 28.103 [17]. Forensic statistical parameters for the X-chromosomal markers were calculated both from allele and haplotype frequencies for the 4 linkage groups according to the formulas described by Desmarais et al. [18].

3. Results and discussion

3.1. Genotyping: rare events in the Iranian populations

3.1.1. AmpFℓSTR® NGM SElect™ kit (AB/LT/TFS)

One of the 214 samples typed with the NGM SElect™ kit had a tri-allelic type 1 pattern [19] in the D18S51 locus. The allele height ratio of the alleles 13, 15, 16 were 4:1:2. A type 1 pattern indicates that the sum of the peak heights of two of the peaks is equal to that of the third one, which is roughly what was observed. This may indicate that a somatic mutation had taken place in this individual. Tri-allelic patterns for the D18S51 system have previously been observed by others [20].

3.1.2. Investigator® DIPplex® kit (Qiagen GmbH, Hilden, Germany)

Among the 253 Iranians typed with DIPplex®, four individuals (three Kurds and one Persian individual) showed weak signals for the HLD97+ (insertion) allele. All four individuals were homozygous for HLD97+ and no heterozygote with weak signal was detected. Sequencing of the HLD97 region revealed a G to A substitution (rs17245568) 61 bp downstream of the HLD97 InDel marker in all four individuals. This substitution has previously been reported [9,21–23] and is positioned in the PCR primer binding site [9].

3.2. Intra-population genetic variability of the Iranian population

A total of 58 polymorphic markers including 46 autosomal markers and 12 X-chromosomal STRs were typed using three forensic kits. The allele or/and haplotype frequencies, as well as other intra-population variability parameters are shown in Supplementary material 1 (Table SM1.1 to SM1.5). Table 1 shows a summary of the mean intra-population variability values obtained for each of the studied set of markers. The mean expected heterozygosity varied from 49% for the biallelic InDel markers to 83% for autosomal STRs. The mean expected heterozygosity of X-chromosomal markers increased to 99% when haplotype instead of allele frequencies were used as a consequence of the high number of different haplotypes.

No statistical significant deviation from HWE ($P > 0.05$) was observed for the autosomal STRs and X-chromosomal markers studied in females after Holm–Šidák correction ($P > 0.003$ and $P > 0.004$, respectively). However, significant deviations from Hardy–Weinberg expectations ($P \leq 0.002$) were observed for the