academicJournals

Vol. 14(5), pp. 375-399, 4 February, 2015
DOI: 10.5897/AJB2014.14103
Article Number: 7A8AACE50278
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Detection of new variant "Off-ladder" at the D12S391, D19S433 and D1S1656 loci and tri-allelic pattern at the D16S539 locus in a 21 locus autosomal short tandem repeat database of 400 Iraqi Individuals

Mohammed Abdullah Jebor[1], Imad Hadi Hameed[1]* and Muhanned Abdulhasan Kareem[2]

[1]Department of Molecular Biology, Babylon University, Hilla City, Iraq.
[2]Centre of Environmental Research, Babylon University, Hilla City, Iraq.

Received 15 August, 2014; Accepted 26 January, 2015

The objectives of the study included the detection of genetic variation of 21 autosomal short tandem repeats (STRs) loci from random unrelated individuals in the middle and South Arab people of Iraq, the Forensic efficiency parameters of the autosomal 21 genetic loci using power plex21® kit and to evaluate the importance of these loci for forensic genetic purposes and the possibility to use the new kit in routine practical work. FTA® Technology was utilized to extract DNA from blood collected on FTA™ paper. Twenty one (21) STR loci including D3S1358, D13S317, PentaE, D16S539, D18S51, D2S1338, CSF1PO, Penta D, THO1, vWA, D21S11, D7S820, TPOX, D8S1179, FGA, D1S1656, D5S818, D6S1043, D12S391, D19S433 and Amelogenin were amplified by using power plex21® kit. PCR products were detected by genetic analyzer 3130xL then, the data processed and analyzed by PowerStatsV1.2 software. Several statistical parameters of genetic and forensic efficiencies based on allelic frequencies have been estimated. This includes the observed heterozygosity (Ho), expected heterozygosity (He), paternity index (PI), random match probability (RMP), power of discrimination (PD), chance of exclusion (CE), polymorphic information content (PIC) and P-value. The power of discrimination values for autosomal tested loci was from 75 to 96% therefore those loci can be safely used to establish a DNA-based database for Iraq population for identifical purpose. The high PIC values of the selected markers confirm their usefulness for genetic polymorphism studies and linkage mapping programs in human as well. The mean heterozygosity observed, expected and mean PIC values across the 20 loci were 0.79, 0.83 and 0.81, respectively indicating high gene diversity. A total of nine off-ladder alleles and one tri-allelic pattern were detected in this study.

Key words: Autosomal STR, Iraq, off ladder, statistical parameters, tri-allelic pattern.

INTRODUCTION

Microsatellites are a group of molecular markers chosen for a number of purposes including forensics individual identification and relatedness testing (Nakamura et al., 1987; Yamamoto et al., 1999). Microsatellites refer to DNA with varying numbers of short tandem repeats (Allor et al., 2005; Klintschar et al., 2005) between a unique sequence. DNA regions with repeat units that are 2 to 7 bp in length or most generally short tandem repeats (STRs) or simple

sequence repeats (SSRs) are generally known as micro-satellites (Ellegren, 2004). In the core repeated bases, long repeat units may contain several hundred to thousands (Butler and Hill, 2012; Silvia et al., 2009). DNA can be used to study human evolution using human genome analysis regions that are not subjected to selection pressure (Mats et al., 2007; Mohammed and Imad, 2013). Besides, information from DNA typing provides vital information in medico-legal with polymorphisms allowing for more biological studies (Walkinshaw et al., 1996). It has been found that microsatellites are evenly distributed in the genome on all chromosomes and all regions of the chromosome (Ensenberger et al., 2010). They can also be found inside gene coding regions, introns, and in the non-gene sequences. Most microsatellite loci are really small, ranging from a few to a few hundred repeats and this small size of microsatellite loci is important for PCR-facilitated genotyping. Basically, microsatellites containing a higher number of repeats are more polymorphic.

The number of repeats at a given locus can vary

between individuals and there is a statistical probability that a given individual will have a set number of repeats at a particular STR locus (Moxon and Wills, 1999; Butler, 2005; Imad et al., 2014a; Muhanned et al., 2015). To process the results from each analysis, large database known as combined DNA index system (CODIS) has been set up. This database stores profiles from convicted offenders and unsolved casework. Similar databases have been set up in Europe, Japan, and other countries. The information in these databases can be used to detect and apprehend serial offenders by permitting rapid exchange of information between crime laboratories (Budowle et al., 2000; Ruitberg et al., 2001; Imad et al., 2014b).

Autosomal STRs locus information were evaluated and selected at The Cooperative Human Linkage Center. http://www.chlc.org evaluates the genetic markers and the loci are selected from there. Additional STR loci, chromosomal location and repeat sequence for each core STR locus are provided in Tables 1 and 2 (Klintschar et al., 2005). According to the International Society of Forensic Genetics (ISFG) recommendation, the repeat sequence motif was defined so that the first 5'-nucleotides on the Gene Bank forward strand define the repeat motif used; therefore, the repeat motif for each STR marker is listed based on this. A significant fact is that STR allele sizes are measured relative to an internal size standard during electrophoresis. This depends on the DNA strand that is labeled using a dye that may have a different apparent measured size. The PowerPlex® 21 System is compatible with automated PCR instrument and with the ABI PRISM® 3100, 3100-Avant, 3130, 3130xl, 3500 and 3500xL Applied Biosystems Genetic Analyzers. In the United States, Europe and Asia, the PowerPlex® 21 System is used, and it increases the discriminatory power and data-sharing possibilities by incorporating informative loci.

MATERIALS and METHODS

Preparation of blood stain samples

Blood samples were randomly collected from 400 healthy unrelated males and females living in the middle and south of Iraq **and** sent to the forensic genetic laboratories in (DNA Typing of Medico-Legal Institute, Baghdad, Iraq) and the same samples were sent to Department of Biomedical Science, University Putra Malaysia, Selangor, Malaysia to complete other tests.

DNA extraction

DNA was extracted from all dried blood samples on FTA cards following the manufacturer's procedure as described in Whatman FTA Protocol BD01 (Mullen et al., 2009; Imad et al., 2014c). A 1.2 mm diameter disc was punched from each FTA card with a puncher and put directly in PCR tube.

DNA amplification for Autosomal STR

PCR is the process used to amplify a specific region of DNA. It is possible to create multiple copies from small amount of template DNA. Amplifications of 21 STR loci D3S1358, D13S317, PentaE , D16S539, D18S51, D2S1338, CSF1PO, Penta D, THO1, vWA, D21S11, D7S820, TPOX, D8S1179, FGA, D1S1656, D5S818, D6S1043, D12S391, D19S433 and Amelogenin were performed by using the PowerPlex®21 kit (Promega Corporation, Madison, WI 53711-5399 USA, © 2011,2012. All Rights Reserved. www.promega.com). The PCR program is as follows: 96°C for 1 min, then: 94°C for 10 s , 59°C for 1 min, 72°C for 30 s, for 25 cycles, then: 60°C for 20 min 4°C. The amplified samples were kept at -20°C in a light-protected box.

PCR amplicon analysis (capillary electrophoresis)

The major application of capillary electrophoresis (CE) in forensic biology is in the detection and analysis of short tandem repeats (STRs). STR markers are preferred because of the powerful statistical analysis that is possible with these markers and the large databases that exist for convicted offenders' profiles. Other related applications include the analysis of haploid markers in the Y chromosome and in mitochondrial DNA (mtDNA) (Imad et al., 2014). Nonhuman DNA testing can also be performed depending on the circumstances of the case. The techniques involved include genotyping, DNA sequencing and mutation detection. The analysis of STR loci in DNA is the most common method for the between two or more unrelated determination of human identity and can

*Corresponding author. E-mail: imad_dna@yahoo.com. Tel: 009647716150716.

Table 1. Comparison of STR loci present in kits used in the United States.

Locus	Power Plex®18D	Power Plex®21	Power Plex®Fusion	Identifiler	Mini Filer	Extended[a]
TPOX	TPOX	TPOX	TPOX	TPOX	-	-
CSF1PO	CSF1PO	CSF1PO	CSF1PO	CSF1PO	CSF1PO	CSF1PO
D5S818	D5S818	D5S818	D5S818	D5S818	-	D5S818
D7S820	D7S820	D7S820	D7S820	D7S820	D7S820	D7S820
D13S317	D13S317	D13S317	D13S317	D13S317	D13S317	D13S317
FGA	FGA	FGA	FGA	FGA	FGA	FGA
vWA	vWA	vWA	vWA	vWA	-	vWA
D3S1358	D3S1358	D3S1358	D3S1358	D3S1358	-	D3S1358
D8S1179	D8S1179	D8S1179	D8S1179	D8S1179	-	D8S1179
D18S51	D18S51	D18S51	D18S51	D18S51	D18S51	D18S51
D21S11	D21S11	D21S11	D21S11	D21S11	D21S11	D21S11
TH01	TH01	TH01	TH01	TH01	-	TH01
D16S539	D16S539	D16S539	D16S539	D16S539	D16S539	D16S539
-	D2S1338	D2S1338	D2S1338	D2S1338	D2S1338	D2S1338
-	D19S433	D19S433	D19S433	D19S433	-	D19S433
-	-	D12S391	D12S391	-	-	D12S391
-	-	D1S1656	D1S1656	-	-	D1S1656
-	D2S441	-	-	-	-	D2S441
-	-	-	D10S1248	-	-	D10S1248
-	-	D6S1043	-	-	-	-
-	-	-	D22S1045	-	-	-
Penta D	Penta D	Penta D	Penta D	-	-	-
Penta E	Penta E	Penta E	Penta E	-	-	Penta E
-	-	-	DYS391	-	-	DYS391
Amelogenin	Amelogenin	Amelogenin	Amelogenin	Amelogenin	Amelogenin	Amelogenin

[a](Lareu et al., 1996**)** for more information on the proposed U.S. extended core.

indisputably distinguish individuals if sufficient loci can be detected (Budowle et al., 2000).

Using the ABI Prism1 3730xl Genetic Analyzer, 16-capillary array system (Applied Biosystems, Foster City, CA, USA) following manufacturer's protocols, with POP-7™ Polymer and Data Collection Software, GeneMapper® V3.2 software (Applied Biosystems, Foster City, CA, USA) were done. By comparison of the size of a sample's alleles to size of alleles in allelic ladders for the same loci being tested in the sample, the STR genotyping was conducted. By comparison of the size of a sample's alleles to size of alleles in allelic ladders for the same loci being tested in the sample, the STR genotyping was conducted (Weir, 1996; Butler et al., 2004).

Statistical data analysis

Statistical analysis for autosomal STR

The PowerStatsV1.2 (Promega, Madison, USA) was used to calculate the observed heterozygosity (Ho), power of discrimination (PD), probability of exclusion (PE) and polymorphism information content (PIC). Arlequin software program was used to conduct the exact test of population differentiation. In addition, Arlequin software program was used for the expected heterozygosity (He), and Hardy Weinberg Equilibrium (HWE). Where test results with *P*-values less than 0.05 were observed, the Bonferroni correction had to be applied to the data. The Bonferroni procedure (Weir, 1996) adjusted the rejection level for the smallest *P*-value at an overall level of α = 5 % to 0.05/x; where, x is equal to the number of tests conducted on the data. The Ho and He values were calculated by means of the same software program.

Following are the formulas used to compute various parameters for population data analyses:

Expected heterozygosity

Edwards et al. (1991) described the following formula for calculating an unbiased estimate of the expected heterozygosity:

$$H = \frac{n}{(n-1)}\left[1 - \sum_{j=1}^{k}\left(n_j/n\right)^2\right] = \frac{n}{(n-1)}\left[1 - \sum_{j=1}^{k}\left(p_j\right)^2\right]$$

Where, *n1, n2, …, nk* are the allele counts of *K* alleles at a locus in a sample of *n* genes drawn from the population and *pj* is the allele frequency.

Table 2. Information on 21 autosomal STR loci present in The PowerPlex® 21 System kits Adapted from **(**Cotton et al., 2000 ; Wiegand et al., 1993) physical positions are from **(**Schneider et al., 1998**).**

STR Locus[a] [n]	Label	Physical position	Chromosomal Location1	Repeat Sequence (5´ 3´)
Amelogenin	Fluorescein	X and Y	Xp22.1–22.3 and Y	NA
D3S1358	Fluorescein	Chr 3 (45.582 Mb)	3p21.31	TCTA Complex
D1S1656	Fluorescein	Chr 1 (230.905 Mb)	1q42	TAGA Complex
D6S1043	Fluorescein	Chr 6 (92.450 Mb)	6q15	AGAT
D13S317	Fluorescein	Chr 13 (82.692 Mb)	13q31.1	TATC
Penta E	Fluorescein	Chr 15 (97.374 Mb)	15q26.2	AAAGA
D16S539	JOE	Chr.16(86.386 Mb)	16q24.1	GATA
D18S51	JOE	Chr 18 (60.949 Mb)	18q21.33	AGAA
D2S1338	JOE	Chr 2 (218.879 Mb)	2q35	TGCC/TTCC
CSF1PO	JOE	Chr 5 (149.455 Mb)	5q33.1	AGAT
Penta D	JOE	Chr 21 (45.056 Mb)	21q22.3	AAAGA
TH01	TMR-ET	Chr 11 (2.192 Mb)	11p15.5	AATG (19)
vWA	TMR-ET	Chr12(6.093 Mb)	12p13.31	TCTA Complex (19)
D21S11	TMR-ET	Chr 21 (20.554 Mb)	21q21.1	TCTA Complex (19)
D7S820	TMR-ET	Chr 7 (83.789 Mb)	7q21.11	GATA
D5S818	TMR-ET	Chr 5 (123.111 Mb)	5q23.2	AGAT
TPOX	CXR-ET	Chr 2 (1.493 Mb)	2p25.3	AATG
D8S1179	CXR-ET	Chr 8 (125.907 Mb)	8q24.13 (125.976 Mb)	TCTA Complex (19)
D12S391	CXR-ET	Chr 12 (12.450 Mb)	12p12(12.341 Mb)	AGAT/AGAC Complex
D19S433	CXR-ET	Chr19(30.416 Mb)	(35.109Mb)	AAGG Complex
FGA	CXR-ET	Chr 4 (155.509 Mb)	4q28 (155.866Mb)	4q28 (155.866Mb)

*a*database of sequence-tagged sites (STSs) available on the NCBI website: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists.
n The 13 CODIS core loci are highlighted in bold font.

Match probability

The probability of a match at locus *l PMl*, was first described from genotype data (Fisher, 1951):

$$PM_l = \sum_i \tilde{G}_{il}^2$$

Where, $\tilde{G}_i$ is the sample frequency of the *i*th genotype at locus *l*.

$$PM_l = \frac{\sum_{i=1}^{n} \tilde{G}_{il}^2 - 1/N_l}{1 - 1/N_l} \approx \sum_{i=1}^{n} \tilde{G}_{il}^2$$

Where, the first part of this equation is for a sample of size *Nl* at locus *l* (Jones, 1972).

Power of discrimination

Brenner and Morris (1990) described the following formula for calculating the power of discrimination:

$$1 - PM$$

Polymorphism information content

The PIC was also calculated using marker allelic frequencies using the following equation:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

Where, *n* is the number of alleles and *pi* is the allele probability of the *i*th allele (Botstein et al., 1980).

Power of exclusion

$$\text{Power of Exclusion (PE)} = H^2(1-(1-H)H^2)$$

Where, *H* = heterozygosity.

Paternity index

$$\text{Paternity Index (PI)} = \frac{H+h}{2h} = \frac{(1-h)+h}{2h} = \frac{1}{2h} = \frac{1}{2\sum_{i=1}^{n} P_i^2}$$

Where, *Pi* is the frequency of *ith* allele in a population of n samples; *h* = homozygosity and *H*= heterozygosity.

RESULTS AND DISCUSSION

Autosomal STRs

The short tandem repeats (STRs) are rich source of highly polymorphic markers in the human genome. They are relatively small in size, and can be studied with the polymerase chain reaction in a multiplex fashion (Butler, 2005). In this study, we used a commercial available multiplex STRs typing kit to study 21 STRs system (D3S1358, D13S317, D16S539, D18S51, D2S1338, CSF1PO, THO1, vWA, D21S11, D7S820, TPOX, D8S1179, FGA, D5S818, D19S433 and Amelogenin).

These STRs loci include five new nucleotide repeat STR loci: D12S391, PentaE, Penta D, D6S1043 and D1S1656.

Allele frequency of common autosomal genetic loci

After the samples have been collected, DNA extracted and PCR amplified, they were genotyped for the 20 STR loci of interest. The genotyping information was then converted into allele frequencies by counting the number of times each allele was observed. Allele frequencies for each of the 20 STR loci in the Iraq population sample are shown in Tables 3 and 4. Since there are some alleles which were not sampled sufficiently and that an estimate of an allele frequency is uncertain if the allele is so rare that it is represented only once or a few times in a dataset, it is recommended that each allele be observed at least five times to be used in forensic calculations (Butler et al., 2009). The minimum allele frequency is 5/(2n); where, n is the number of individuals sampled and 2n is the number of chromosomes (as autosomes are in pairs due to inheritance of one chromosome from each parent). In the loci D5S818 (allele 13), the highest allele frequencies are found, and the lowest allele frequencies are at allele13 as seen in D3S1358 locus. D21S11 and D18S51 loci illustrate the largest number of different alleles. The following locations, the most common alleles at the 21 loci, were allele 13 for (D5S818, D8S1179 and D18S51), allele 10 for (D7S820 and Penta D), allele 16 for (D2S1338 and D2S1338), allele 11 for (TPOX, D16S539, PentaE, CSF1PO and D6S1043) loci, allele 12 for (D13S317 and D1S1656), allele 30.2 for D21S11 locus, allele 9 for THO1 locus, allele 13.2 for D21S11 locus, allele 6 for THO locus, alleles 17.3 and 16 for VWA locus, allele 30.2 for D19S433 locus, allele 17 for D3S1358 locus, allele 18 for D5S818 locus, allele 18 for D12S391 locus, allele 14.2 for D12S391 locus and allele 25 for FGA locus. The polymorphic nature of microsatellites (STR) makes them the markers of favorite in properties and genetic diversity studies (Ossmani et al., 2009; Chouery et al., 2010; Imad et al., 2014d). The best indicator of the genetic polymorphism within the sample is verified by the number of alleles and the expected heterozygosity found in the Iraq population. Basically the number of alleles is highly associated with the size of the sample. This is due to the presence of unique alleles in populations, which occur in low frequencies. The usefulness of the markers for genetic screening is verified by the number of alleles scored for each marker (Andreini et al., 2007).

The number of alleles and the expected heterozygosities detected in Iraq population are good indicators of the genetic polymorphism within the breed. Generally the number of alleles is highly dependent on the sample size because of the presence of unique alleles in populations, which occur in low frequencies and also because the number of observed alleles tends to increase with increases in population size. The number of alleles scored for each marker is an invaluable indicator of the future usefulness of the marker for genetic screening. Finding the same number of alleles for certain different loci in various populations (for example, Iran, Syrian, Emirates, Qatar and Egyptian populations) may indicate common ancestries (Reyhaneh and Sadeq, 2009; Alshamali et al., 2003; Ana et al., 2003; Clotilde et al., 2007). The frequency and the number of alleles, however, may be an indication for the degree of inbreeding within each population and thus reflects the homogeneity of the population. The analysis of STR polymorphisms by PCR-based method offers certain advantages over RFLP typing. In recent years, short tandem repeat (STR) systems have gained importance in forensic analysis of biological specimens as well as in paternity testing, as an alternative to the use of restriction fragment length polymorphism (RFLP) analysis (Alshamali et al., 2003; Ana et al., 2003; Clotilde et al., 2007; Hammond et al., 1994). The analysis of STR polymorphisms by PCR-based method offers certain advantages over RFLP typing: (1) STR loci can be typed with a high degree of specificity and sensitivity in a short time period, (2) these loci can be successfully amplified from a limited amount of DNA even if it is degraded, and (3) typing of multiple loci can be accomplished in a single multiplex reaction (Hochmeister et al., 1991; Lins et al., 1996).

The amelogenin locus

The Amelogenin locus that occurs on both the X and Y chromosomes and enables sex typing (Sullivan et al., 1993) was also located within the reference human genome sequence. AMELX is located on the X chromosome at 10.676 Mb. AMELY is located on the Y chromosome at 6.441 Mb. Amplification of Amelogenin generates different length products from the X and Y-

Table 3. Allele frequencies and forensic efficiency parameters (D8S1179- D13S317) in Iraq.

Allele	D8S1179	D7S820	D21S11	THO1	vWA	TPOX	D13S317
2.2	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-
6	-	-	-	0.0455	-	-	-
7	-	0.0265	-	0.1713	-	-	-
8	0.127	0.1896	-	0.1326	-	0.0197	0.2318
9	0.0064	0.1003	-	0.2362	-	0.0383	0.0373
9.3	-	-	-	0.1344	-	-	-
10	0.076	0.3304	-	0.18	-	0.1013	0.0546
10.2	-	-	-	-	-	-	-
11	0.1344	0.2054	-	-	-	0.2637	0.2248
12	0.1153	0.1115	-	-	-	0.2005	0.343
12.2	-	-	-	-	-	-	-
13	0.2153	0.0363	-	-	-	0.1985	0.0534
13.2	-	-	-	-	-	-	-
14	0.142	-	-	-	0.0917	0.178	0.0551
14.2	-	-	-	-	-	-	-
15	0.0549	-	-	-	0.0594	-	-
15.2	-	-	-	-	-	-	-
15.3	-	-	-	-	-	-	-
16	0.0181	-	-	-	0.2824	-	-
16.2	-	-	-	-	-	-	-
16.3	-	-	-	-	-	-	-
17	-	-	-	-	0.2806	-	-
18	-	-	-	-	0.0588	-	-
19	-	-	-	-	0.0212	-	-
20	-	-	-	-	0.2059	-	-
21.2	-	-	-	-	-	-	-
27	-	-	-	-	-	-	-
28	-	-	0.0614	-	-	-	-
28.2	-	-	-	-	-	-	-
29	-	-	0.1817	-	-	-	-
29.2	-	-	-	-	-	-	-
30	-	-	0.1393	-	-	-	-
30.2	-	-	0.1615	-	-	-	-
31	-	-	0.0596	-	-	-	-
31.2	-	-	0.0113	-	-	-	-
32	-	-	0.0485	-	-	-	-
32.2	-	-	0.0589	-	-	-	-
33	-	-	0.0589	-	-	-	-
33.2	-	-	0.0983	-	-	-	-
34	-	-	0.0029	-	-	-	-
34.2	-	-	0.1066	-	-	-	-
35	-	-	0.0112	-	-	-	-

chromosomes. Figure 1 shows the possible peaks for both genotypes X/X and X/Y represented as one and two peaks, respectively, in the GeneMapper electropherogram.

New additional autosomal genetic loci

To improve results with challenging DNA samples, a set of 20 autosomal STR loci was characterized in our

Table 4. Allele frequencies and forensic efficiency parameters (D19S433 - CSF1PO) in Iraq.

Allele	D19S433	D2S1338	D18S51	D16S539	FGA	D5S818	D3S1358	CSF1PO
2.2	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-
7	0.0023	-	-	-	-	-	-	-
8	-	-	-	0.1788	-	-	-	0.0158
9	0.0939	-	-	0.1012	-	0.0228	-	0.036
9.3	-	-	-	-	-	-	-	-
10	0.0082	-	0.0086	0.2185	-	0.0124	-	0.2636
10.2	-	-	0.0208		-	-	-	-
11	0.1834	-	0.0454	0.251	-	0.0382	-	0.2858
12	0.157	-	0.1782	0.0987	-	0.0133	-	0.2633
12.2	0.1148	-	-	-	-	-	-	-
13	0.0726	-	0.2459	0.1003	-	0.4375	0.0037	0.0622
13.2	0.2169	-	-	-	-	-	-	
14	0.0745	-	0.17	0.0515	-	0.3588	0.04	0.011
14.2	0.0424	-		-	-	-	-	-
15	0.0109	0.0583	0.1044	-	-	0.0208	0.2267	0.0633
15.2	0.0598	-	-	-	-	-	-	-
15.3	-	-	-	-	-	-	-	-
16	-	0.2015	0.1471	-	-	0.0407	0.2394	-
16.2	-	-	-	-	-	-	0.001	-
16.3	-	-	-	-	-	-	0.001	-
17	-	0.1383	0.0473	-	-	-	0.3471	-
18	-	0.0785	0.0111	-	-	-	0.1231	-
19	-	0.1513	0.0064	-	-	-	0.02	-
20	-	0.0587	0.0112	-	0.1014	-	-	-
20.2	-	-	-	-	-	-	-	-
21	-	0.0388	0.0112	-	0.0379	-	-	-
21.2	-	-	-	-	-	-	-	-
22	-	0.0612	-	-	0.1817	-	-	-
22.2	-	-	-	-	0.1389	-	-	-
23	-	0.0083	-	-	0.0787	-	-	-
23.2	-	-	-	-	-	-	-	-
24	-	0.0717	-	-	0.1373	-	-	-
24.2	-	-	-	-	-	-	-	-
25	-	0.0617	-	-	0.2429	-	-	-
26	-	0.0717	-	-	0.0812	-	-	-
27	-	-	-	-	-	-	-	-
28	-	-	-	-	-	-	-	-
28.2	-	-	-	-	-	-	-	-
30	-	-	-	-	-	-	-	-
30.2	-	-	-	-	-	-	-	-
31	-	-	-	-	-	-	-	-

laboratory at NIST (Hill et al., 2009). The efforts of the Human Genome Project have increased knowledge regarding the human genome, and hence there are many more STR loci available now than there were 10 years ago. In fact, more than 20,000 tetranucleotide STR loci have been characterized in the human genome (Collins

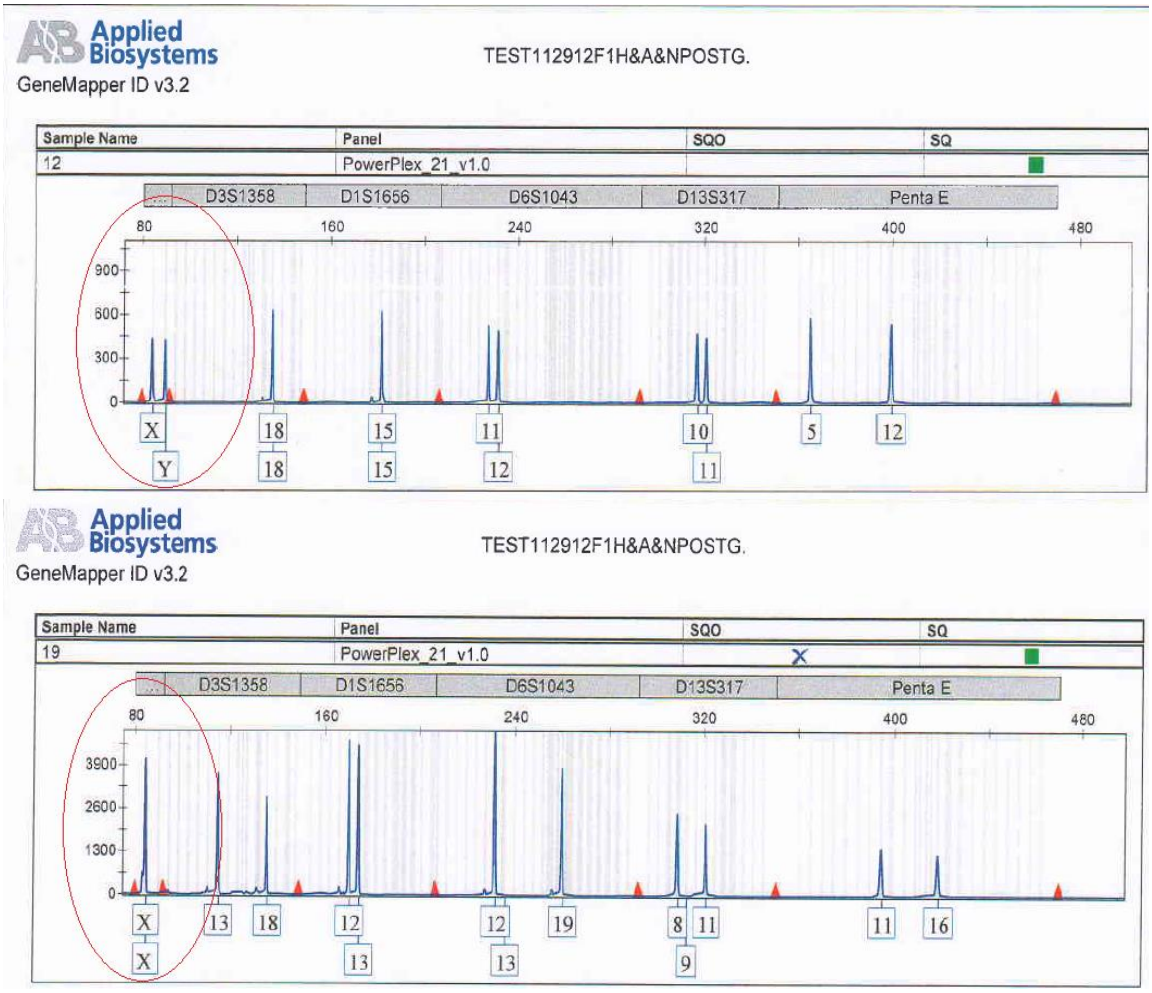

Figure 1. An electropherogram showing Amelogenin locus and its two possible genotypes obtained with PowerPlex®21 amplification and GeneMapper version 4.0.

et al., 2003) and there may be more than a million STR loci present depending on how they are counted (Ellegren, 2004). STR sequences account for approximately 3% of the total human genome (Landerand Linton, 2001).

Allele frequency of D12S391 genetic loci

D12S391 is a highly polymorphic compoundtetranucleotide found on the short arm of chromosome 12 only 6.3 megabases from vWA. It possesses over 50 different alleles ranging from 13 to 27.2 repeats in length. A number of same-size, different-sequence alleles have been identified through sequence analysis. It is part of the extended European Standard Set and is present in NGM™ and NGM SElect™ kits from Applied Biosystems, the PowerPlex® ESI and ESX Systems from Promega, and ESSplex and ESSplex SE kits from Qiagen.

D12S391 is one of new nucleotide repeat STR loci and eight alleles were found in this study. As shown in Figure 2, the allele 18 was most frequent 0.359.

Allele frequency of Penta D genetic loci

Penta D is a pentanucleotide repeat found on chromosome 21 about 25 Mb from D21S11. Alleles ranging from 1.1 to 19 repeats have been observed although some of the shorter alleles are likely due to flanking region deletions (Kline et al., 2011). Penta D is present in the PowerPlex® 16 and PowerPlex® 18D kits. Eleven (11) alleles were found in this study, and as shown in Figure 3, the allele 10 was most frequent 0.398.

Allele frequency of penta E genetic loci

Penta E is a pentanucleotide repeat with very low stutterproduct formation that is located on the long arm of

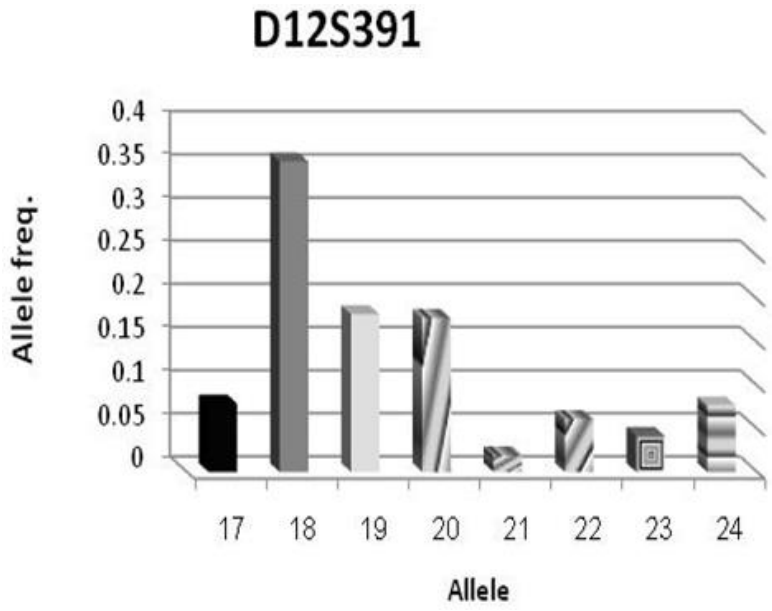

Figure 2. Allele frequency of D12S391 genetic locus.

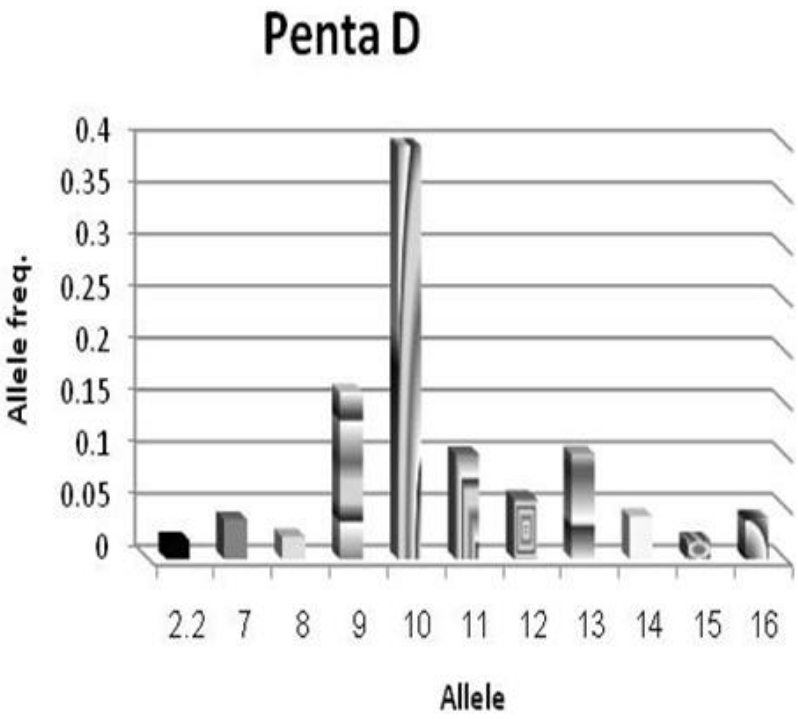

Figure 3. Allele frequency of Penta D genetic locus.

chromosome 15 with alleles ranging from 5 to 32 repeats. Penta E is highly polymorphic and is present in the PowerPlex® 16 and PowerPlex® 18D kits. The Penta E locus on chromosome 15 , one of the new pentanucleotide repeat STR loci examined in this study was highly polymorphism and 10 allele were found in this iraq population. As shown in Figure 4, the allele 11 was most frequent 0.241. The distribution was different from those reported in the Caucasian American and African American populations (Budowle et al., 2001; Levadokou

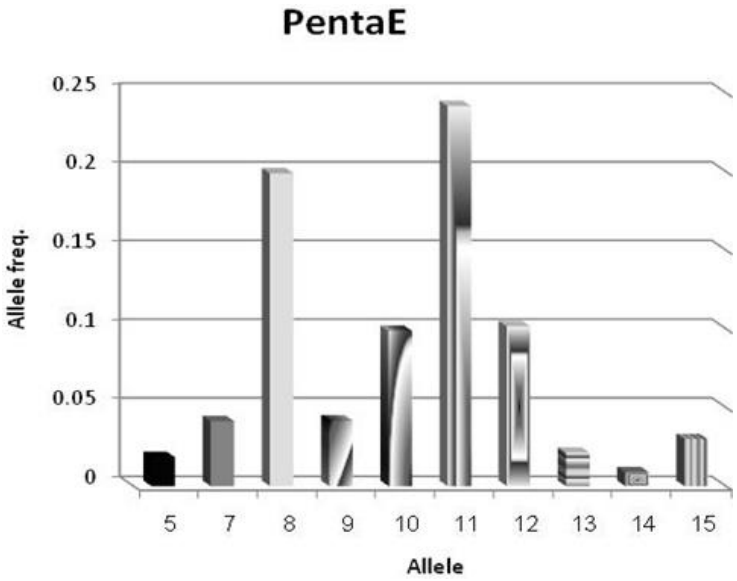

Figure 4. Allele frequency of Penta E genetic locus.

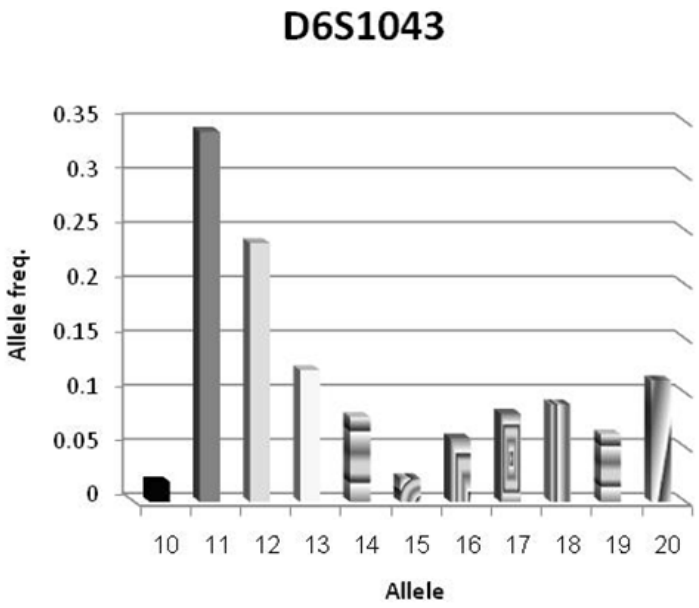

Figure 5. Allele frequency of D6S1043 genetic locus.

et al., 2001). As mentioned above, allele 13 and 14, which had not been detectrd in Caucasian and Africans, were found in the present Iraq population.

Allele frequency of D6S1043 genetic loci

D6S1043 is a compound tetranucleotide repeat with alleles ranging from 8 to 25 AGAT or AGAC repeats. Some x.2 and x.3 alleles have been reported in population studies. D6S1043 is part of the Sinofiler™ kit and has been used to date almost exclusively in Chinese and other Asian population studies. D6S1043 is located less than 4 Mb from SE33 on the long arm of chromosome 6. Eleven (11) alleles were found in this study, and as shown in Figure 5, the allele 11 was most frequent 0.238.

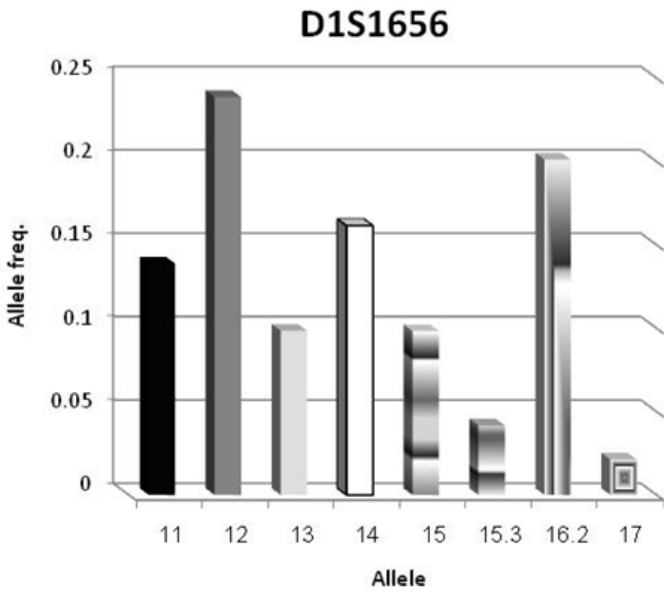

Figure 6. Allele frequency of D1S1656 genetic locus.

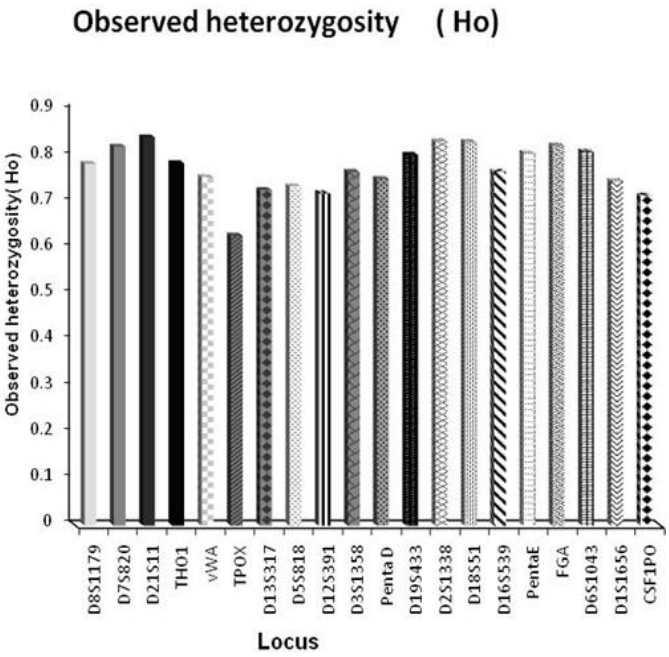

Figure 7. Observed Heterozygosity of twenty autosomal genetic loci.

Allele frequency of D1S1656 genetic loci

D1S1656 is a tetranucleotide repeat found on the long arm of chromosome 1 with alleles ranging from 8 to 20.3 repeats. The x.3 alleles arise from a TGA insertion typically after four full TAGA repeats. It is part of the

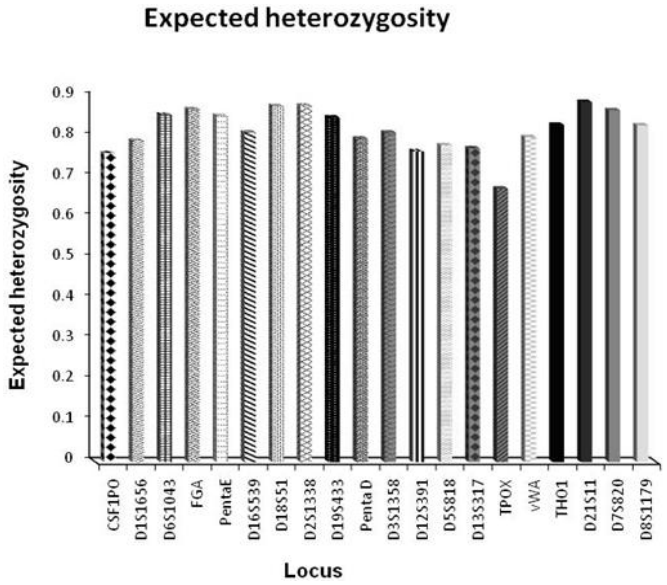

Figure 8. Expected heterozygosity of twenty autosomal genetic loci.

extended European Standard Set and is present in NGM™ and NGM SElect™ kits from Applied Biosystems, the PowerPlex® ESI and ESX Systems from Promega, and ESSplex and ESSplex SE kits from Qiagen. The D1S1656 is one of the new nucleotide repeat STR loci examined in this study which was highly polymorphic and eight alleles were found. As shown in Figure 6, the allele 12 was most frequent (0.238).

Forensic efficiency parameters

Statistical analysis is used to interpret DNA results for genetic identity. In order to determine the significance of a match, it is necessary to support DNA typing results with statistical analysis. These analyses assign a value to the results obtained and enable easier resolution of forensic or paternity cases. Across all loci, the values for the matching probability, power of discrimination, power (chance) of exclusion, polymorphism information content and typical paternity index for the 20 STR loci of the Iraq population were determined and are indicated in Figures 7 to 14.

Observed heterozygosity and expected heterozygosity

The two common ways to report heterozygosity are observed and expected heterozygosities. Observed heterozygosity is calculated by dividing the number of heterozygote individuals at a locus by the total number of all individuals at that locus and describes the heterozygosity actually observed in the members of the sampling group. Expected heterozygosity is calculated as

Table 5. The observed heterozygosity in different populations.

Locus	Iraq	Turkey	Iran	Syria	Emirates	Qatar	Egypt	Gaza
D21S11	0.841	-	-	0.664	-	-	0.682	0.682
TPOX	0.628	0.567	-	-	-	-	-	-
Penta E	-	-	-	0.897	-	-	-	-
FGA	-	-	-	-	-	-	0.890	
D18S51	-	0.790	-	-	0.130	-	-	0.912
D2S1338	-	-	-	-	-	0.839	-	-
vWA	-	-	-	-	-	0.542	-	-
LPL	-	-	0.780	-	-	-	-	-
F13A01	-	-	0.720	-	-	-	-	-
D5S818	-	-	-	-	0.295	-	-	-

1 minus the homozygosity (the sum of squares of all allele frequencies at a locus) and represents the number of heterozygotes that would be expected under HWE based on the observed allele frequencies in the sampling group. A higher heterozygosity means that more allele diversity exists, and therefore there is less chance of a random sample matching. Observed heterozygosity and expected heterozygosity all over the 20 loci are presented in Figures 7 and 8, and the observed heterozygosity oscillated between studied populations as illustrated in Table 5.

The observed heterozygosity in a population relies on the number and the frequency of alleles of each locus. Moreover, the distribution of genotypes in a population sample may deviate from HWE expectation in a number of ways. These include the presence of an excess of homozygotes and a corresponding lack of heterozygotes or an excess (deficiency) of one or more classes of heterozygotes or a combination of those states. There are populations with low heterozygosity, lower than 65% in most tested loci. These populations are small, closed, inbred by cultural or geographical factors; one of those populations is the Qatari population where the levels of observed heterozygosity (*Ho)* oscillated, between 0.339 for D19S433 and 0.839 for D2S1338. Interestingly, *Ho* is lower than expected heterozygosity *(He) in* almost all the analyzed loci, with the unique exception of D5S818 locus. This fact was particularly conspicuous for the loci that deviated from the HWE expectations; the departures from HWE expectations detected in the Qatari population seem to be the result of excess of homozygotes over heterozygotes, which is likely to be the consequence of the high consanguinity rates reported for this population, which is 46% (Ana et al., 2006).

Paternity index

The paternity index (PI) compares the likelihood that a genetic marker (allele) that the alleged father (AF) passed to the child to the probability that a randomly selected unrelated man of similar ethnic background could pass the allele to the child. The potential of a randomly selected man to pass the obligate gene is determined by using a database, which lists the frequency distribution of individual alleles within a given genetic system. Combined paternity index is an odd ratio that indicates how many times more likely it is that the alleged father is the biological father than a randomly selected unrelated man of similar ethnic background. The paternity index was high for all STR analyzed; it ranged from 2.651 (TPOX) to 2.864 (D21S11) (Figure 9).

Random match probability

The match probability is the probability for a random match between two unrelated individuals drawn from the same population. It is the sum of the frequency squared of each genotype its that ranged from 0.011 to 0.168 (Figure 10).

A random match probability, is the probability that a person sampled randomly from the population, would have a particular STR profile. If assumptions can be made that (1) alleles within a locus are inherited independently and (2) that STR loci migrate independently of one another during meiosis, then it is possible to multiply the specific expected genotype frequencies (based on allele frequencies) for all the loci together to obtain the overall random match probability. This combination of all the loci match probabilities by multiplication is termed the product rule. The strength of a genetic match between a questioned sample and a known sample (rarity of a particular DNA profile) relies on the discriminative power of the product rule. Thus the assumptions of independence for alleles within a locus and independence among migrating loci are central to determining the strength of a match (Butler et al., 2005).

Power of discrimination

Power of discrimination is defined as the probability that

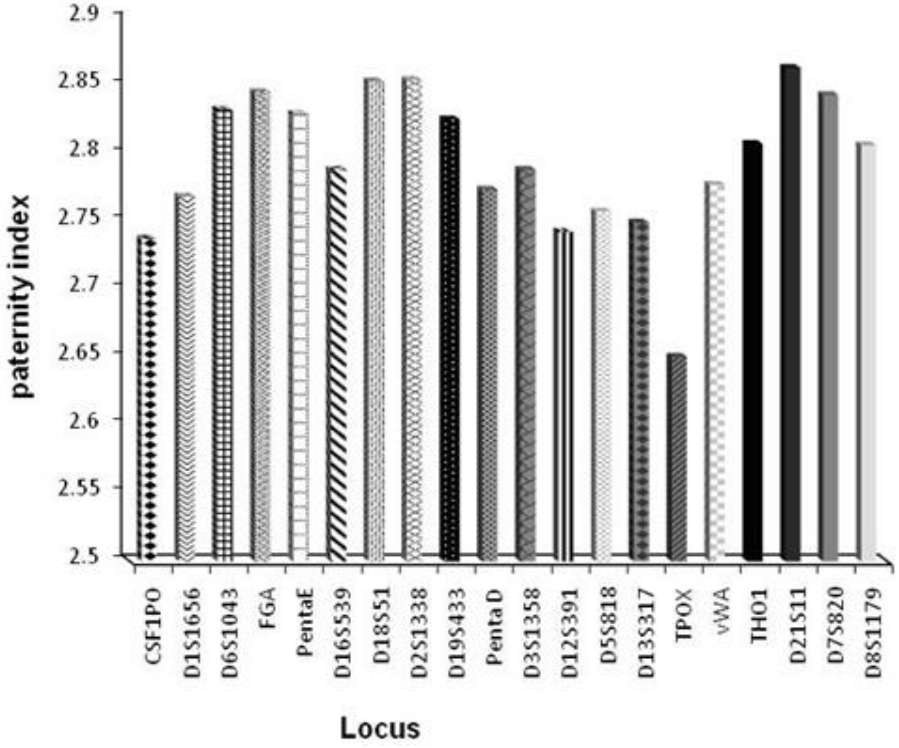

Figure 9. Paternity Index of twenty autosomal genetic loci.

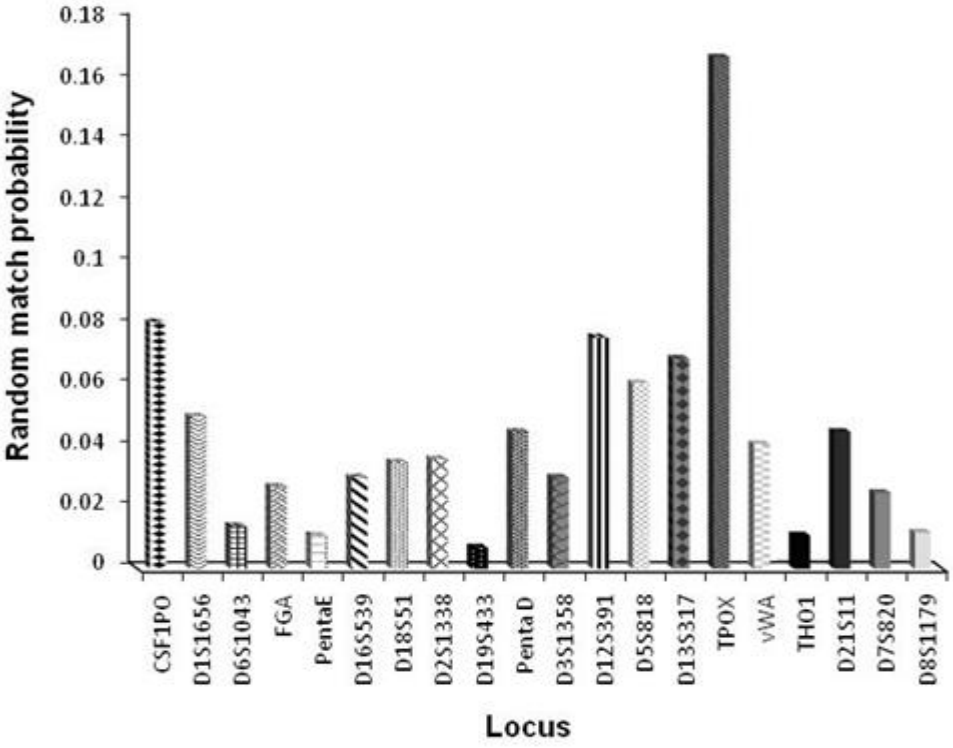

Figure 10. Random match probability of 20 autosomal genetic loci.

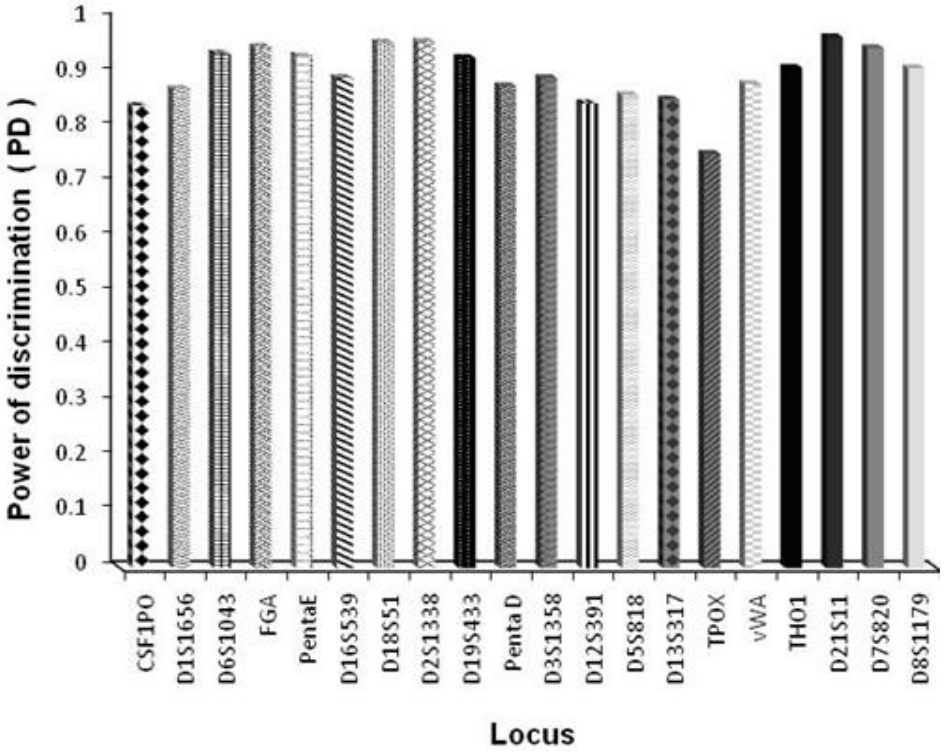

Figure 11. Power of discrimination of 20 autosomal genetic loci.

Table 6. The Power of discrimination in different populations.

Locus	Iraq	Turkey	Iran	Syria	Emirates	Qatar	Egypt	Gaza
D21S11	0.969	-	-	-	-	-	-	-
D2S1338	0.960	-	-	-	-	0.973	-	-
D18S51	0.959	0.953	-	0.963	0.971	-	-	0.962
TPOX	0.759	0.772	-	-	-	0.855	-	-
Penta E	-	-	-	0.974	-	-	-	0.976
Penta D	-	-	-	0.951	-	-	-	0.961
FGA	-	-	-	0.957	-	-	0.973	0.967
vWA	-	-	-	-	-	-	0.937	-
LPL	-	-	0.924	-	-	-	-	-
F13A01	-	-	0.922	-	-	-	-	-
D5S818	-	-	-	-	0.889	-	-	-

two individuals selected at random from the population will not have an identical genotype at the locus. Power of discrimination (PD) for all the 20 loci is presented in Figure 11. Values for all tested loci was 75% for TPOX; ranged from 80 to 89% for the D3S1358, D13S317, D5S818, D12S391, vWA, Penta D, D16S539, D1S1656 and CSF1PO loci, and ranged from 91 to 96% for the rest of the loci. This infers that a DNA-based database for Iraq population can be safely used by using these loci. The possibility to find two persons with the same DNA profile if chosen at random in a population is defined as the matching possibility. It is impossible to find two individuals with the same genotype in the population because most common alleles for account of most common genotypes frequency and matching probability are not present. The highest PD observed in some populations is presented in Table 6.

The Penta E and Penta D loci included in the PowerPlex®21 PCR amplification kits were not typed in the Turkey, Emirates, Iran or Qatari populations because

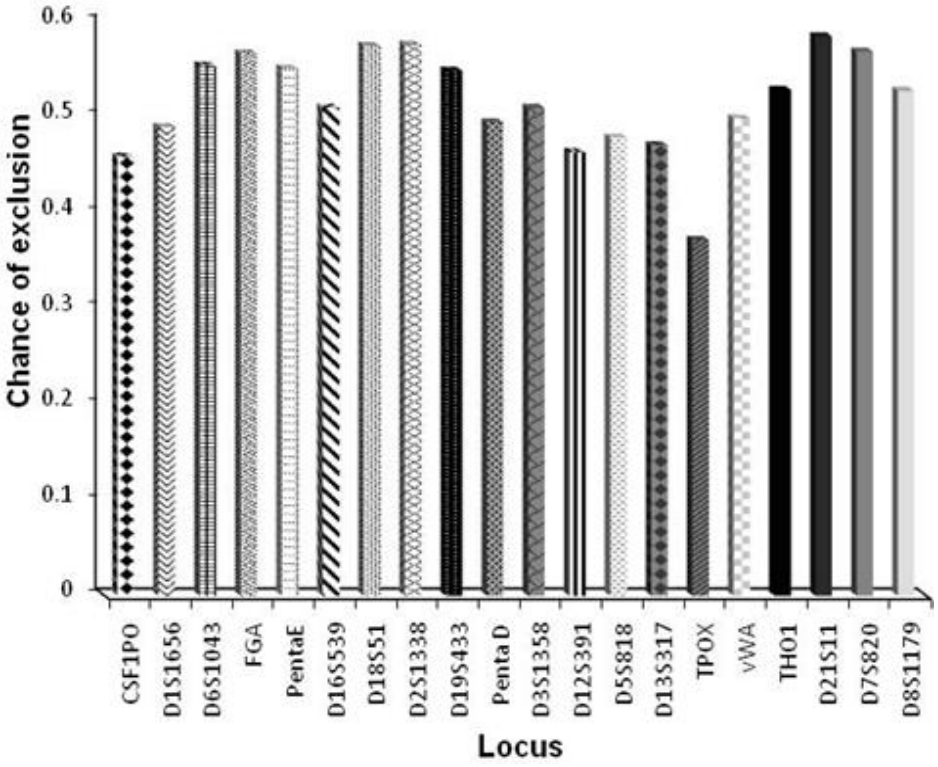

Figure 12. Chance of exclusion of twenty autosomal genetic loci.

Table 7. The power of exclusion in different populations.

Locus	Iraq	Turkey	Iran	Syria	Emirates	Egypt	Gaza
D21S11	0.585	-	-	-	-	-	-
D18S51	-	0.790	-	-	0.762	-	0.820
TPOX	0.372	0.258	-	-	-	0.326	0.364
Penta E	-	-	-	0.788	-	-	-
CSF	-	-	-	0.387	-	-	-
FGA	-	-	-	-	-	0.775	-
LPL	-	-	0.564	-	-	-	-
F13A01	-	-	0.473	-	-	-	-
D5S818	-	-	-	-	0.505	-	-

they used different kits in their genotyping studies. The combined discrimination power (CDP) for the Iraq population of middle and south of Iraq for the corresponding 20 STR loci used, has been calculated as 0.999999972. These results mean that those loci can be safely used to establish a DNA-based database for Iraq population.

Chance of exclusion

Power of exclusion is the probability of excluding a random person as the contributor of an allele to a child at the locus. The power of exclusion (PE) can be calculated to express how rare it would be to find a random man who could not be excluded as the biological father of the child. In the present study and from the genotyping data, PE for every locus was calculated and presented in Figure 12. As expected, the power of exclusion was high for all the microsatellites analyzed; it ranged from 0.372 (TPOX) to 0.820 (D18S51). PE for different populations is given in Table 7. The combined power of exclusion, which is the exclusion probability considering all 20 loci,

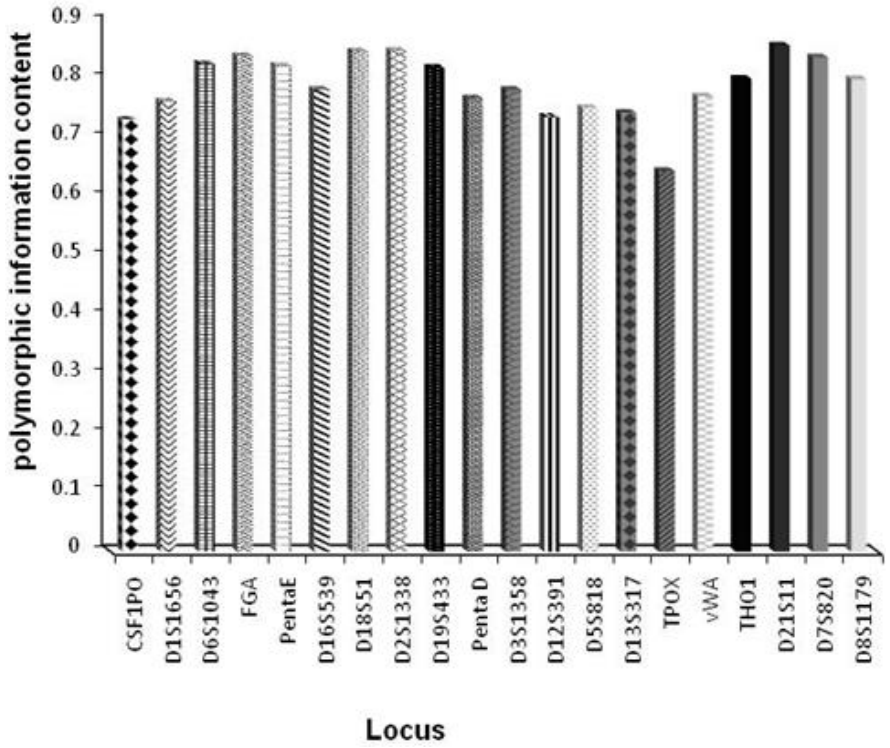

Figure 13. Polymorphic Information Content of twenty autosomal genetic loci.

Table 8. Polymorphic Information Content in different populations.

Locus	Iraq	Turkey	Iran	Syria	Qatar	Egypt	Gaza
D21S11	0.859	-	-	-	-	-	-
D18S51	-	0.840	-	-	-	-	-
D2S1338	-	-	-	-	0.865	-	-
TPOX	0.646	0.540	-	0.640	0.638	0.610	0.650
Penta E	-	-	-	0.890	-	-	0.900
FGA	-	-	-	-	-	0.870	-
LPL	-	-	0.770	-	-	-	-
F13A01	-	-	0.760	-	-	-	-
D5S818	-	-	-	-	-	-	-

was greater than 99.989%, indicating that these loci are appropriate to determine parentage in Iraq population beyond any reasonable doubt.

Polymorphic information content

Polymorphism information content is the probability that a given offspring of a parent possessing a rare allele at a locus will allow inferring of the parental genotype at the locus. The TPOX locus is the least polymorphic marker while D21S11 is the most polymorphic marker based on the degree of polymorphism of every marker, expressed in PIC terms (Figure 13). The usefulness of the findings of genetic polymorphism studies and linkage mapping programs in humans is confirmed by the high PIC values of the selected markers.

Similar degree of polymorphism was found in Syrian population, where the most polymorphic marker was found to be Penta E, and that the least polymorphic one was TPOX (Table 8). On the contrary, the Egyptian population showed that the FGA locus is the most polymorphic marker and that the TPOX locus, as observed in our population, is the least polymorphic marker. Even in other populations like the Iran, Qatar and Turkey, it was found

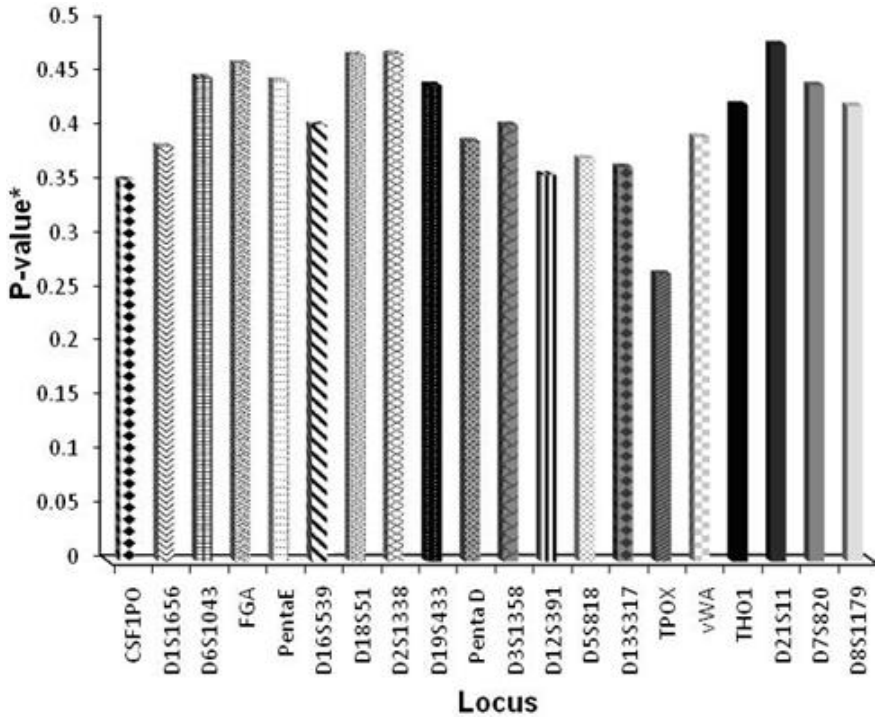

Figure 14. P-value of twenty autosomal genetic loci.

that the previously mentioned loci have a similar degree of polymorphism (Ana et al., 2006; Reyhaneh and Sadeq, 2009). The polymorphic nature of microsatellites makes them the markers of choice in characterization and genetic diversity studies. The high PIC values of the selected markers confirm their usefulness for genetic polymorphism studies and linkage mapping programs in human as well. The mean heterozygosity observed, expected and mean PIC values across the 20 loci were 0.77, 0.81 and 0.78, respectively, indicating high gene diversity.

P-value: probability value of Chi-square test for Hardy Weinberg equilibrium

Checking for HWE is performed by taking the observed allele frequencies and calculating the expected genotype frequencies based on the allele frequencies. If the observed genotype frequencies are close to the expected genotype frequencies calculated from the observed allele frequencies, then the population is in Hardy–Weinberg equilibrium and allele combinations are likely to be independent of one another. The results obtained from the tests for HWE is shown in Figure 14. The null hypothesis states that all the STR loci tested are in HWE and any deviation from HWE expectations is due to sampling error (Hill et al., 2009). The alpha value was set at 0.05. The null hypothesis was rejected if the computed p values were below the alpha value of 0.05. Therefore having a p value above 0.05 to show that the STR alleles do not differ significantly from HWE and does not imply that the samples are in HWE. Butler (2005) states that “if a p-value of < 0.05 is observed with a set of alleles measured at a particular STR locus, it does not mean that a laboratory should avoid using this data because it ‘failed’ a test for Hardy-Weinberg equilibrium”

Most common genotype frequencies

Another measure to reflect the usefulness of a particular set of DNA markers is to examine the frequencies of the most common genotypes, which would therefore be the least powerful in terms of being able to differentiate between two unrelated individuals (Edwards et al., 1991). The theoretically most common type can be calculated by considering a sample type that is heterozygous at all loci possessing the two most common alleles at each locus (Lander and Linton, 2001; Imad et al., 2014e). In Table 9, frequencies from the two most common alleles at each of the twenty loci were used to estimate a theoretical most common STR profile. A number of the newly available STR loci, such as D1S1659 and D6S1043, provide a better probability of identity than widely used loci such as D8S1179 and D2S1338. The “most common genotypes

Table 9. Most common genotype frequencies based on the two most common alleles found in a Iraq population.

Locus	Allele 1	Allele 2	Allele1 Frequency (P)	Allele2 Frequency (q)	2pq	Most common genotype frequency
D8S1179	13	14	0.2153	0.142	2pq	0.0611
D7S820	10	11	0.3304	0.2054	2pq	0.1357
D21S11	29	30.2	0.1817	0.1615	2pq	0.0587
THO1	9	10	0.2362	0.18	2pq	0.0851
vWA	16	20	0.2824	0.2059	2pq	0.1163
TPOX	11	12	0.2637	0.2005	2pq	0.1057
D13S317	8	12	0.2318	0.343	2pq	0.1591
D5S818	13	14	0.4375	0.3588	2pq	0.3139
D12S391	18	19	0.3597	0.1827	2pq	0.1314
D3S1358	16	17	0.2394	0.3471	2pq	0.1662
Penta D	9	10	0.1614	0.3985	2pq	0.1286
D19S433	11	13.2	0.1834	0.2169	2pq	0.0796
D2S1338	16	19	0.2015	0.1513	2pq	0.0609
D18S51	12	13	0.1782	0.2459	2pq	0.0876
D16S539	8	11	0.1788	0.251	2pq	0.0898
PentaE	8	11	0.1988	0.2415	2pq	0.096
FGA	22	25	0.1817	0.2429	2pq	0.0883
D6S1043	11	12	0.3412	0.2358	2pq	0.1609
D1S1656	12	16.2	0.2387	0.2012	2pq	0.2153
CSF1PO	10	11	0.2636	0.2858	2pq	0.0961

frequency" results column can also be a useful metric to locus performance. For example, a comparison of D1S1659 and THO1 is instructive. These two loci both have 44 most common allele frequency, yet D1S1659 has 0.2153 common genotypes frequency while THO1 only has 0.0851. The greater number of genotypes formed with the different combinations of alleles in D1S1659 leads to better probability of identity (0.050 vs. 0.011) values. Furthermore, additional genotype combinations mean that D1S1659 will likely be more useful than THO1 for detecting contributors in DNA mixtures.

New alleles are constantly being discovered that do not size exactly with the ladder alleles. These "off-ladder"

STR typing is typically performed using size comparisons to standardized allelic ladders that possess the most common alleles, which have been sequenced to reveal the true number of repeats. Different STR kit manufacturers may supply allelic ladders with slightly different allele ranges. As more samples are run with STR loci, new alleles are constantly being discovered that do not size exactly with the ladder alleles. These "off-ladder" alleles can be variants with more or less of the core repeat unit than present in the common alleles found in the commercially available allelic ladder. Alternatively, these variant alleles may contain partial repeats or insertions/deletions in the nearby flanking region to the repeat. Insertion/deletion event that creates off-ladder alleles is found in new alleles can be discovered that occur outside the range defined by the commercially available allelic ladder. In many instances, these alleles are simply classified as greater than the largest allele as (or smaller than the smallest allele) in the ladder rather than attempting to extrapolate to a predicted number of repeats. Table 10 contains a list of variant or "off-ladder" alleles that have been reported to the NIST STRBase website as of April 2005. Off-ladder alleles are rare alleles that are not represented in the locus-specific allelic ladders. These off-ladder alleles do not fit within the 0.5 bp range of corresponding alleles in the allelic ladder. Since such alleles cannot be sized by direct comparisons to the reference alleles in the allelic ladders, genotyping software will often designate them as "off-ladder alleles". An off-ladder allele may occur between two alleles in the allelic ladder or it may be smaller (or larger) than the smallest (or largest) allele in the allelic ladder (Butler, 2005). To identify an off-ladder allele, the size (in base pairs) of the off-ladder allele is compared to the sizes of the two closest alleles in the allelic ladder.

A total of nine off-ladder alleles were detected in this study (Table 10). Three of them are located at the D1S1656 locus (Figure 15A, B and C); one is located in D12S391 (Figure 16). One occurred at the D19S433 locus (Figure 17). All of the samples containing off-ladder alleles were analyzed twice on the ABI 310 to confirm the data. Alleles' sizes in base pairs are all generated by the

Table 10. New variant or "off-ladder" discovered in this study and comparison with alleles reported in STRBase.

STR Locus	Number Reported	STRBase	In our study	
		Variant alleles reported as of Apr 2005	STR Locus	Variant alleles
CSF1PO	11	5, 7.3, 8.3, 9.1, 9.3, 10.1, 10.2, 10.3, 11.1, 12.1, 16	-	-
D12S391	0	None reported yet in STRBase	D12S391	18.3
FGA	69	12.2, 13.2, 14, 14.3, 15, 15.3, 16,16.1, 16.2, "<17", 17, 17.2, 18.2,19.1, 19.2, 19.3, 20.1, 20.2, 20.3, 21.1, 21.2,21.3, 22.1, 22.2, 22.3, 23.1, 23.2, 23.3,24.1, 24.2, 24.3, 25.1, 25.2, 25.3, 26.1,26.2, 26.3, 27.3, 29.2, 30.2, 31, 31.2, 32.1,32.2, 33.1, 34.1, 34.2, 35.2, 41.1, 41.2, 42.1, 42.2, 43.1, 43.2, 44, 44.1, 44.2, 44.3, 45.1, 45.2, 46.1, 46.2,47.2, 48.2, 49, 49.1, 49.2, 50.2, 50.3	-	-
TH01	7	4, 7.3, 8.3, 9.1, 10.3, 11, 13.3	-	-
TPOX	7	4, 5, 7.3, 13.1, 14, 15, 16	-	-
VWA	6	16.1, 18.3, 22, 23, 24, 25	-	-
D3S1358	18	8, 8.3, 9, 10, 11, 15.1, 15.2, 15.3, 16.2, 17.1, 17.2, 18.1, 18.2, 18.3, ">19", 20, 20.1, 21.1	-	-
D5S818	5	10.1, 11.1, 12.3, 17, 18	-	-
D7S820	22	5, 5.2, 6.3, 7.1, 7.3, 8.1, 8.2, 8.3, 9.1, 9.2, 9.3, 10.1, 10.3, 11.1, 11.3, 12.1, 12.2, 12.3, 13.1, 14.1, 15, 16	-	-
D8S1179	4	7, 15.3, 18, 20	-	-
D13S317	10	5, 6, 7, 7.1, 8.1, 11.1, 11.3, 13.3, 14.3, 16	-	-
D16S539	10	6, 7, 9.3, 11.3, 12.1, 12.2, 13.1, 13.3, 14.3, 16	-	-
D18S51	30	7, 8, 9, 11.2, 12.2, 12.3, 13.1, 13.3, 14.2, 15.1, 15.2, 16.1, 16.2, 16.3, 17.2,17.3, 18.1, 18.2, 19.2, 20.1, 20.2, 21.2, 22.1, 22.2, 23.2, 24.2, 27, 28.1, 28.3, 40	-	-
D21S11	24	24.3, 25.1, 25.2, 25.3, 26.2, 27.1, 27.2, 28.1, 28.3, 29.1, 29.3, 30.3, 31.1, 31.3, 32.1, 33.1, 34.1, 34.3, 35.1, 36.1, 36.2, 37, 37.2, 39	-	-
Penta D	14	6, 6.4, 7.1, 7.4, 9.4, 10.3, 11.1, 11.2, 12.2, 12.4, 13.2, 13.4, 14.1, 14.4	-	-
Penta E	13	9.4, 11.4, 12.1, 12.2, 13.2, 14.4, 15.2, 15.4, 16.4, 17.4, 18.4, 19.4, 23.4	-	-
D2S1338	3	13, 23.2, 23.3	-	-
D19S433	11	6.2, 7, 8, "<9", 11.1, 12.1, 13.2, 18, 18.2, 19.2, 20	D19S433	16.2
SE33	0	None reported yet in STRBase	-	-
D6S1043	0	None reported yet in STRBase	-	-
D1S1656	0	None reported yet in STRBase	D1S1656	16.3 , 17.3 , 18.3

GeneScan® software. Complicated structures give rise to more variations (mutations) and thus a greater chance of finding microvariants as well as rarely seen full repeating-unit alleles. Microvariants differ from full repeat alleles by containing small sequence variations. The differences can be due to insertions, deletions, or point mutations (Butler, 2005). D7S820, can contain 8, 9, or 10 adjacent T nucleotides starting from 12 nucleotides downstream of the GATA repeat. This flanking region insertion/deletion gives rise to the 9.1, 9.3, 10.1, 10.3, etc. alleles observed in D7S820. An allelic ladder contains only alleles commonly found in the human population. Rare alleles and microvariants are, therefore, not often represented in the allelic ladder. Thus microvariants that fall between two alleles in an allelic ladder or alleles that fall outside of the allelic ladder range are labeled as off-ladder alleles. An FGA microvariant allele 16.1 was observed in sample M-101. A search on the STRBase website as of May 2006 showed that this particular allele was reported ten times. The repeat structure of FGA 16.1 is [TTTC]3TTTTTTCT[CTTT] 5T[CTTT]3CTCC [TTCC]2 [43]. Microvariant allele 24.3 was detected at D21S11of M-68. This allele was seen 9 times on the STRBase website in May 2006. The repeat sequence of D21S11

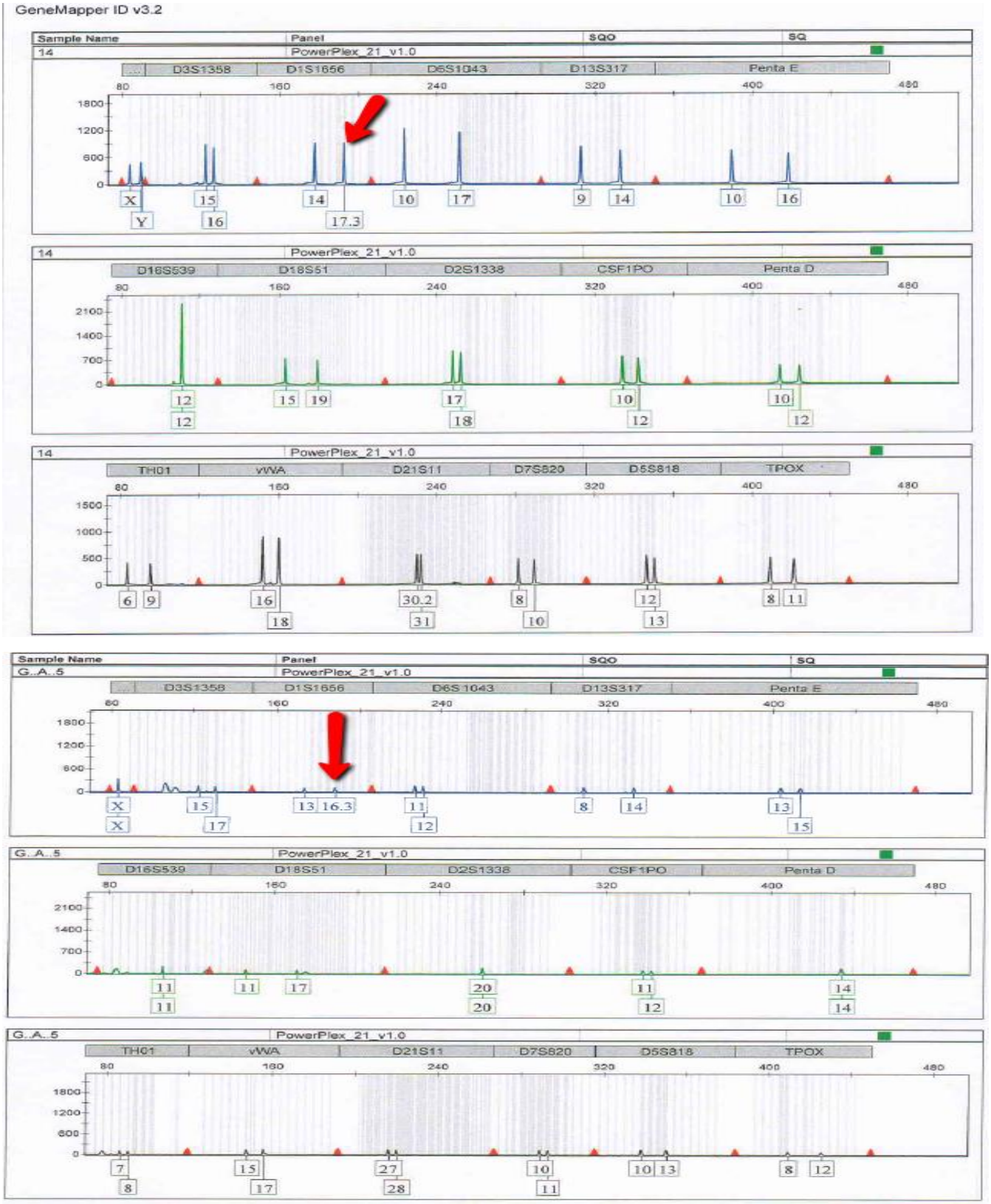

Figure 15. A. Off-ladder allele (17.3) located at the D1S1656 locus **. B.** Off-ladder allele (16.3) located at the D1S1656 locus. **C.** Off-ladder allele (18.3) located at the D1S1656 locus**.**

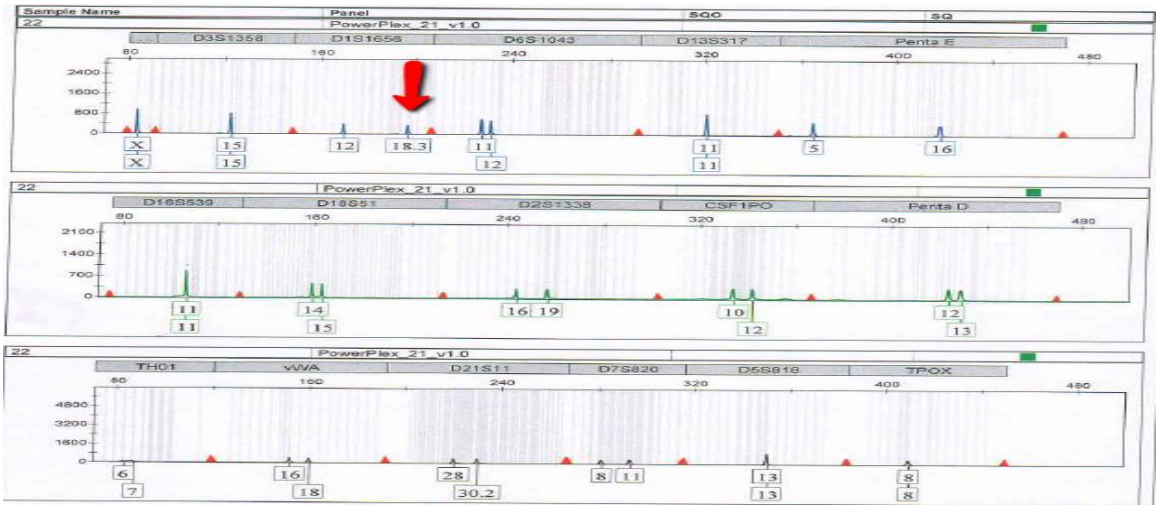

Figure 15. Contd.

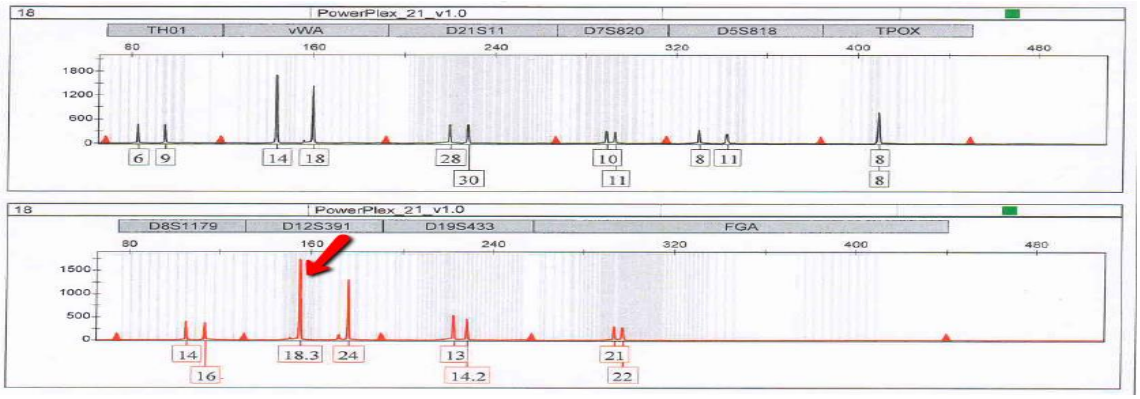

Figure 16. Off-ladder allele located at the D12S391 locus.

allele 24.3 is not published. However, the repeat structure is predicted to be similar to the form [TCTA]n[TCTG]n{[TCTA]3TA [TCTA]3TCA[TCTA]2TCCATA}[TCTA]nTATCTA (Butler, 2005). Off-ladder allele 40 was found at the D21S11 locus of samples M-77 and M-194. A search of the STRBase website (again in May 2006) does not turn up this particular allele and therefore no actual repeat sequence is given. However, the repeat motif is expected to have the similar D21S11 structure previously mentioned. The findings of off-ladder alleles at the FGA and D21S11 loci support the fact that off-ladder alleles are most often encountered at loci that have complex repeat structures.

New Tri-allelic patterns have been observed for many of the core STR loci

Tri-allelic patterns have been observed for many of the core STR loci and recorded on the NIST STRBase website. Clayton et al. (2004) described possible reasons for tri-allelic patterns, which can occur as an imbalance in amounts between the three alleles (type 1) or equal amounts of all three alleles (type 2). A type 1 tri-allelic pattern imbalance is typically a situation where the sum of the peak heights for two of the alleles is approximately equivalent to the third allele (Clayton et al., 2004). It is interesting to note that TPOX, which occurs closest to the tip of a chromosome, has the highest number of observed tri-allelic patterns—most of which are type 2 with equal intensity alleles. Thus, it is possible that this section of chromosome 2 is more likely to be duplicated in some individuals for telomere maintenance to keep the end of the chromosome intact (Chakhparonian and Wellinger, 2003; Louis and Vershinin, 2005). One triallelic pattern D16S539 observed in this study is shown in Table 11 and Figure 18. A number of explanations for these three-banded allele patterns have been suggested: (a) a

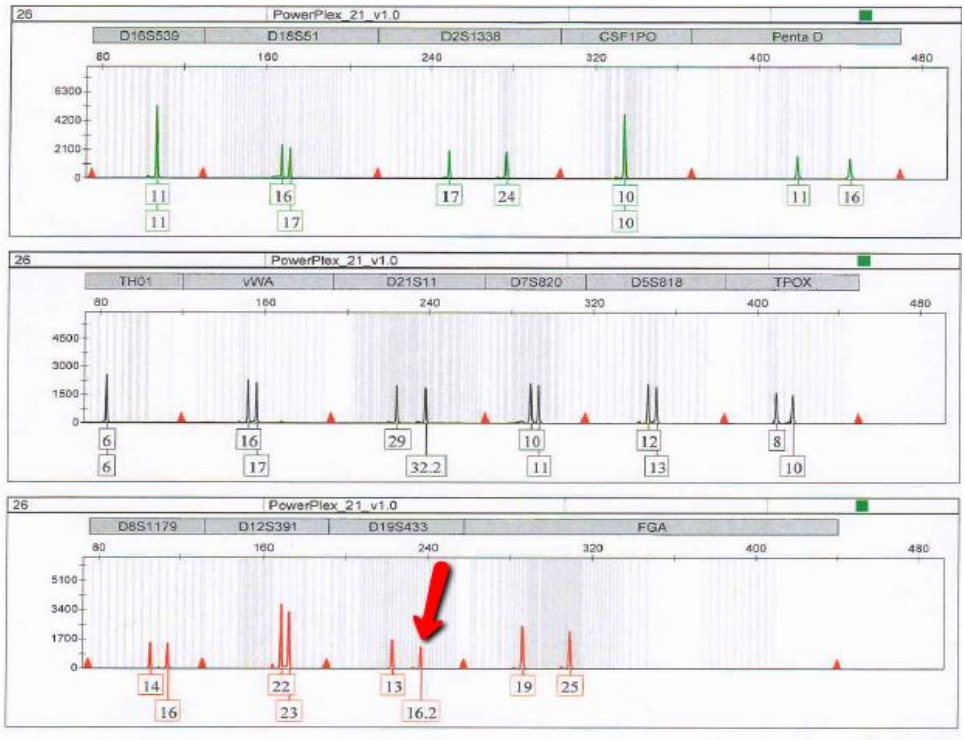

Figure 17. Off-ladder allele located at the D19S433 locus.

Table 11. A total tri-allelic pattern observed in our study and reported on STRBase.

STRBase		**In our study**	
STR Locus	Variant alleles	STR Locus	Variant alleles
CSF1PO	9/11/12; 10/11/12	-	-
FGA	19/20/21; 19/22/23; 19/24/25; 20/21/22; 20/21/24; 20/23/24; 21/22/23; 21/25/26; 22/24/25; 22.2/23/23.2	-	-
TH01	7/8/9	-	-
TPOX	6/8/10; 6/9/10; 6/10/11; 6/10/12; 7/9/10; 7/10/11; 8/9/10; 8/10/11; 8/10/12; 8/11/12; 9/10/11; 9/10/12; 10/11/12	-	-
VWA	11/16/17; 12/18/19; 14/15/17; 14/15/18; 14/16/18; 14/17/18; 15/16/17; 18/19/20	-	-
D3S1358	15/16/17; 15/17/18; 16/17/19; 17/18/19	-	-
D5S818	10/11/12; 11/12/13	-	-
D7S820	8/9/12; 8/10/11	-	-
D8S1179	10/12/13; 10/12/15; 12/13/14; 12/13/15; 13/15/16	-	-
D13S317	8/11/12; 10/11/12; 10/12/13	-	-
D16S539	12/13/14	**D16S539**	**9/12/13**
D18S51	12/13/15; 12/14/15; 12/16/17; 14/15/22; 15/16/20; 16/17/20; 19/22.2/23.2	-	-
D21S11	28/29/30; 28/30.2/31.2; 29/31/32; 30/30.2/31	-	-
Penta D	None reported yet in STRBase	-	-
Penta E	None reported yet in STRBase	-	-
D2S1338	None reported yet in STRBase	-	-
D19S433	None reported yet in STRBase	-	-
SE33	None reported yet in STRBase	-	-
D6S1043	None reported yet in STRBase	-	-
D1S1656	None reported yet in STRBase	-	-

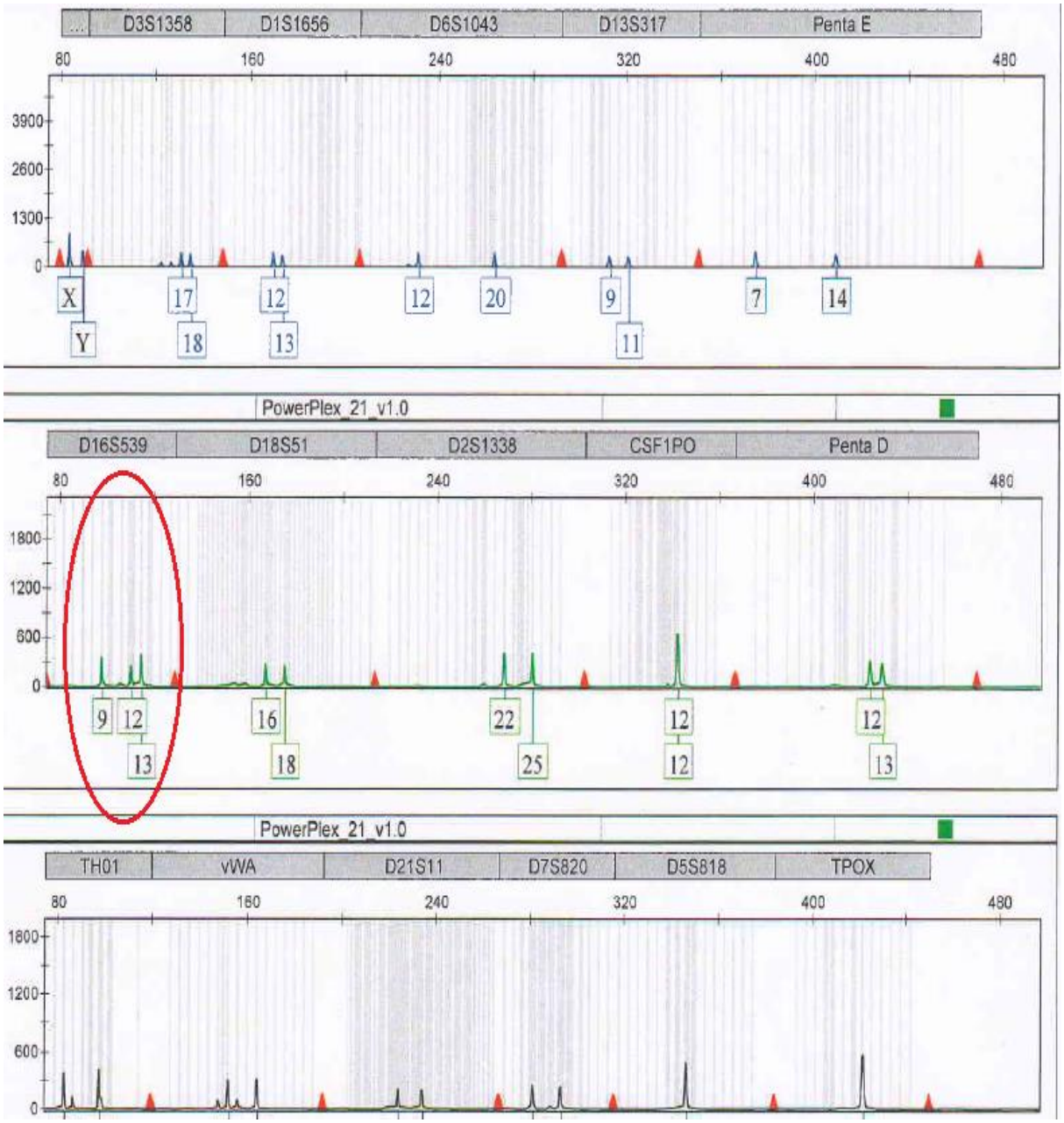

Figure 18. Tri-allelic pattern 9/12/13 observed in D16S539 genetic loci.

genetic duplication of a small chromosomal region containing the STR locus, (b) an improper segregation resulting from chromosomal meiotic or mitotic nondisjunction that leads to either true trisomy or to mosaicism, and (c) chimerism (Rubocki et al., 2001).

Although tri-allelic patterns have been reported by other researchers, it is unclear what their frequency is in the general population. The Ballard Laboratory has observed 10 in a database of individuals across the 15 Identifiler™ loci. However, in an extensive study by Crouse et al. (1999) individuals, only 19 tri-allelic patterns (18 at TPOX and 1 at CSF1PO) were observed (0.19% per locus). In this study, one tri-allelic pattern was observed at the TPOX locus of sample M-200. This sample was genotyped twice on the ABI 310 to verify the result. The TPOX tri-alleles are 8, 10, and 11. This sample was neither included in the STR allele frequencies study nor the HWE database validation. The presence of an extra peak at one locus out of all of the 15 loci tested indicates that this sample is not a mixture. Rather, the extra peak is a real

reproducible artifact of the sample. The sample was re-typed and the same tri-allelic pattern was obtained. The fact that the peak heights of all three alleles are similar suggests that the alleles are probably present in equal copy number (Egyed et al., 2000). Tri-allelic patterns appear to be common at the TPOX locus. A search of the STRBase website (Butler, 2006) as of May 2006 reveals 14 different TPOX tri-allelic patterns.

Conclusion

The conclusions that have been drawn from the present study include: the results of the current study indicate these new autosomal STR Loci useful For DNA typing markers in Iraq can be used for establishment of a DNA database that will be beneficial for the population in terms of resolving social and moral disputes and will contribute to improvements in the justice system. The match probability is the probability for a random match between two unrelated individuals drawn from the same population; it ranged from 0.011 to 0.168. The mean heterozygosity observed, expected and mean PIC values across the 20 loci were 0.77, 0.81 and 0.78, respectively, indicating high gene diversity. A total of nine off-ladder alleles were detected in this study; four of them are located at the D1S1656 locus, three are located in D12S391, one occurred at the D19S433 locus and one discovered outside the range. One triallelic pattern observed in this study was 9/12/13 pattern.

Conflict of interests

The authors did not declare any conflict of interests.

Abbreviations: STRs, Short tandem repeats; Ho**,** observed heterozygosity; **He,** expected heterozygosity; **PI,** paternity index; **RMP**, random match probability; **PD**, power of discrimination; **CE,** chance of exclusion; **PIC,** polymorphic information content.

REFERENCES

Allor, Catherine, David D, Einum, Marco S (2005). Identification and Characterization of Variant Alleles at CODIS STR Loci. J. Forensic Sci. 50(5):1128-1133.

Alshamali FH, Alkhayat AI, Budowle B, Watson ND (2003). Allele frequency distributions and other population genetic parameters for 13 STR loci in a UAE local population from Dubai. Int. Congress Series.1239:249-258.

Ana MP, Miguel AA, José AP , Rene JH (2006). Qatari DNA Variation at a Crossroad of Human Migrations. Hum. Hered. 61:67-79.

Andreini E, Sara F, Elena L, Rosanna T, Nicola D, Francesca P (2007). Allele Frequencies for Nine STR Loci (Des1358, vWA, FGA,

D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820) in the Italian Population. Forensic Science Int. 168(1): e13-e16.

Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. American J. Hum. Gen. 32:324-331.

Brenner C, Morris J (1990). Paternity index calculations in single locus hypervariable DNA probes.Proceedings of the International Symposium on Human Identification: Promega Corporation. Long Beach: California. pp. 21-53.

Budowle B, Chakraborty R, Carmody, G (2000). Source attribution of a forensic DNA profile, Forensic Sci Comm. 2:1.

Budowle B, Sprecher CJ (2001). Concordance study on population database samples using the PowerPlex 16 kit and AmpFISTR profiler Plus kit and AmpFISTR COfiler kit. J. Forensic Sci.46:637-641.

Butler J, Buel E, Crivellente F (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis. 25:1397.

Butler JM, Hill CR (2012). Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Forensic Sci. Rev. 24(1):15-26.

Butler JM, Buel E, Crivellente F, McCord B (2009). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis 25:1397-412.

Butler JM (2005). Forensic DNA Typing. Biology, Technology and Genetic of STR markers. Second Edition. Elsevier Academic Press.

Butler, J (2006). Genetics and genomics of core STR loci used in human identity testing, J. Forensic Sci. 51(2):253-265.

Chakhparonian M, Wellinger RJ (2003). Telomere maintenance and DNA replication: how closely are these two connected?. Trends Genet. 19(8):439-46.

Chouery E, Michael DC, Katharine MS, Jessica LS, Nadine J, Myrna M, Andre M (2010). Population Genetic Data for 17 STR Markers from Lebanon, Legal Medicine. 12(6):324-326.

Clayton TM, Guest JL, Urquhart AJ, Gill PD (2004). A genetic basis for anomalous band patterns encountered during DNA STR profiling. J. Forensic Sci. 49:1207-14.

Clotilde C, Evelyne G, Farha E, Georges L, Jean-Michel D (2007). Allele frequencies of 15 short tandem repeats (STRs) in three Egyptian populations of different ethnic groups. Forensic Science Int. pp. 169260-169265.

Collins JR, Stephens RM, Gold B, Long B, Dean M, Burt SK (2003). An exhaustive DNA micro-satellite map of the human genome using high performance computing. Genomics. 82:10-9.

Edwards A, Civitello A, Hammond HA, Caskey CT (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet 49:746-56.

Egyed B, Furedi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, Padar Z (2000). Analysis of eight STR loci in two Hungarian populations. Forensic Sci. Int. pp. 113:25-27.

Ellegren H. (2004) Microsatellites: simple sequences with complex evolution. Nature Rev. Genet. 5:435-445.

Ensenberger MG, Thompson J, Hill B, Homick K, Kearney V, Mayntz-Press KA (2010). Developmental validation of the PowerPlex 16 HS System: an improved 16-locus fluorescent STR multiplex. Forensic. Sci. Int. Genet. 4(4):257-264.

Fisher RA (1951). Standard calculations for evaluating a blood group system. Heredity. 5:95-102.

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.551:75-89.

Hill CR, Butler JM, Vallone PM (2009). A 26plex Autosomal STR Assay to Aid Human Identity Testing. J. Forensic Sci. 54:1008.

Hochmeister MN, Budowle B, Jung J, Borer UV, Comey CT, Dirnhofer R (1991). PCR-based typing of DNA extracted from cigarette butts. Int. J. Leg. Med. 10:506-513.

Imad H, Ameer I, Mohammed A, Cheah, Y, Aamera J (2014a). Haplotypes and variable position detection in the mitochondrial DNA coding region encompassing nucleotide positions 10,716–11,184. Mitochondrial DNA 1-6.

Imad H, Mohammed A, , Aamera J (2014b). Genetic variation and DNA markers in forensic analysis. Afr. J. Biotechnol. 13(31): 3122-3136.

Imad H, Mohammed A, Aamera J, Ameer I, Cheah Y (2014c) Haplotype data of mitochondrial DNA coding region encompassing nucleotide positions 11,719-12, 184 and evaluate the importance of these

positions for forensic genetic purposes in Iraq. Mitochondrial DNA 1-4.
Imad H, Mohammed A, Cheah Y, Aamera J. (2014d) Genetic variation of twenty autosomal STR loci and evaluate the importance of these loci for forensic genetic purposes. Afr. J. Biotechnl. 13:1-9.
Imad H, Muhanned A, Aamera J, Cheah Y. (2014e) Analysis of eleven Y-chromosomal STR markers in middle and south of Iraq. Afr. J. Biotechnol. 13(38): 3860-3871.
Jones DA (1972). Blood samples: probability of discrimination. J. Forensic Sci. Soc. 12:355-358.
Kline MC, Hill CR, Decker AE, Butler JM (2011). sequence analysis for characterizing normal, variant, and null alleles. Forensic Sci. Int. Genet. 5:329.
Klintschar M, Immel UD, Stiller D, Kleiber M (2005). TH01, a tetrameric short tandem repeat locus in the tyrosine hydroxylase gene: association with myocardial hypertrophy and death from myocardial infarction?. Dis. Markers. 21:9-13.
Lander ES, Linton LM (2001). Initial sequencing and analysis of the human genome. Nat. 409(6822):860-921.
Levadokou EN, Freeman DA, Budzynski MJ, Early BE (2001). Allele frequencies for fourteen STR5 loci of the PowerPlex 1.1 and 2.1 multiplex systems and Penta D locus in Caucasians , African-American ,Hispanic, and other populations of the United states of the America and Brasil. J. Forensic Sci. 124:47-54.
Lins AM, Sprecher CJ, Puers C, Schumm JW (1996). Multiplex sets for amplification of polymorphic short tandem repeat loci- silver stain and fluorescent detection. BioTechniques. 20:882-889.
Louis EJ, Vershinin AV (2005). Chromosome ends: different sequences may provide conserved functions. Bio. Essays. 27:685-97.
Mats O, Mo H, Erik W, Tonia S, Natasha L, Tobias U (2007). Mating system variation and morph fluctuations in a polymorphic lizard. Mol. Ecol. 16:5307-5315.
Mohammed A and Imad H. (2013) Autosomal STR: From locus information to next generation sequencing technology. Res. J. Biotechnol. 8(10):92-105.
Moxon E, Wills C (1999). DNA microsatellites: Agents of evolution? Sci. Am. 280:94.
Muhanned AK, Ameer IA, Imad HH, Mohammed AJ. (2015) A New Polymorphic Positions Discovered in Mitochondrial DNA Hypervariable Region HVIII From Central and North-Central of Iraq. Mitochondrial DNA 1–5.
Mullen MP, Howard DJ, Powell R, Hanrahan JP (2009). A note on the use of FTA™ technology for storage of blood samples for DNA analysis and removal of PCR inhibitors. Irish J. Agric. Food Res. 48:109-113.
Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T (1987). Variable number of repeat (VNTR) markers for human gene mapping. Sci. 235:1616-1622.
Ossmani H, Jalal T, Brahim B, Abdelaziz C (2009). Allele Frequencies of 15 Autosomal STR Loci in the Southern Morocco Population with Phylogenetic Structure Among Worldwide Populations. Legal Med. 11(3):155-158.
Reyhaneh L, Sadeq V (2009). Genetic variation of informative short tandem repeat (STR) loci in an Iranian population. Iranian J. Biotechnol. 7(3):137-141.
Rubocki RJ, McCue BJ, Duffy KJ, Shepard KL, Shepherd SJ, Wisecarver JL (2001). Natural DNA mixtures generated in fraternal twins in utero. J. Forensic Sci. 46(1):120-125.
Ruitberg CM, Reeder DJ, Butler JM (2001). STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Res. 29:320-322.
Silvia B, Campo D, Andrea Z, Dario B, Tatiana G (2009). Integration of Genomic and Gene Expression Data of Childhood ALL Without Known Aberrations Identifies Subgroups with Specific Genetic Hallmarks. Genes Chromosomes & Cancer. 48:22-38.
Sullivan KM, Mannucci A, Kimpton CP, Gill P (1993). A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. BioTechniques. 15(4):636-638.
Walkinshaw M, Strickland L, Hamilton H, Denning K, Gayley T (1996). DNA Profiling in two Alaskan Native Populations Using HLA-DQA1, PM, and D1S80 Loci, J. Forensic Sci. 41:478-484.
Weir B (1996). The second national research council report on forensic DNA evidence, Am. J. Hum. Genet. 59:497.
Yamamoto T, Uchihi R, Nozawa H, Huang XL, Leong YK, Tanaka M (1999). Allele distribution at nine STR loci-D3S1358, vWA, FGA, TH01, TPOX, CSF1PO, D5S818,D13S317 and D7S820-in the Japanese population by multiplex PCR and capillary electrophoresis. J. Forensic Sci. 44(1):167-170.