A Globally used 15 Short Tandem Repeats (STR) Loci in Forensic Human Identification, with their Allele Frequencies and Statistical Values in the Population of Bahrain.

[1]**Aqeela Salman Abuidrees** ; [1]**Malak Jaffar Ishaq** ; [2]**Chang-En Pu**; [1]**Noora Abdullah Alhamad**; [1]**Hasan Ali Alnafea**; and [1]**Ali Mohamed Almehaizea**

[1]Biology and DNA department, Directorate of Forensic Science Evidence, Public prosecution, Manama, Kingdom of Bahrain. [2]Scientific and Technical Research Center, Ministry of Justice Investigation Bureau, 74, Chang-Hua Road, Taipei Country, Taiwan, ROC

ABSTRACT

KEYWORDS

The aim of this study is to estimate the allelic frequencies of Bahraini population using the AmpFlSTR Identifiler® PCR kit which includes 15 universal Short Tandem Repeats (STR) loci. The samples were buccal swabs collected from 302 unrelated individuals from Kingdom of Bahrain. They are selected randomly from Bahraini population in a where to conduct this study. After going through the whole DNA extraction, quantification, Polymerase Chain Reaction (PCR) amplification, Short Tandem Repeats (STR) typing and statistical analysis we got valuable data and results. The most frequent allele types for each locus were: (D8S1179: 13, D21S11: 29, D7S820: 10, CSF1PO: 11, D3S1358: 16, TH01: 7, D13S317: 12, D16S539: 11, D2S1338: 23, D19S433: 13, vWA: 17, TPOX: 8, D18S51: 14, D5S818: 12, FGA: 23). The highest allele frequencies from them were 0.5 for TPOX: 8, 0.3692 for D5S818: 12 and 0.3626 for D16S539: 11. Maximum Homozygosity (H) was 0.3576 shown in TPOX locus. On the other hand, Maximum Heterozygosity (H2) was 0.9106 shown in D2S1338 locus. As expected the most polymorphic loci which had the highest Polymorphic Information Content were the most discriminating ones: D2S1338 (0.9717), D18S51 (0.9668), FGA (0.9640), while the least discriminated locus was TPOX (0.7202), as the discrimination powers (PD) ranged from 0.7202 (TPOX), to 0.9718 (D2S1338). A t-test was applied to the data generated, and therefore it was normally distributed.

<div dir="rtl">

المتكررات الترادفية القصيرة الأكثر إستخداماً في تحديد هوية البصمة الوراثية عالمياً، مع معدلات تكرار أليلاتها وأهم القيم الإحصائية المتعلقة بها لعينة تمثيلية من المجتمع البحريني.

[1]عقيلة سلمان أبوادريس، [1]ملاك جعفر إسحاق، [2]جانغ إن بو، [1]نورة عبدالله الحمد، [1]حسن علي النافع، و[1]علي محمد المهيزع

[1] قسم الأحياء والبصمة الوراثية، إدارة الأدلة المادية، النيابة العامة، مملكة البحرين

[2] مركز البحث العلمي والتقني، مكتب التحقيقات بوزارة العدل، 74، شارع جانغ هو، مدينة تايبيه، تايوان، جمهورية الصين

المُستلخص

الكلمات الدالة

الهدف من هذه الدراسة هو تقدير معدل تكرار الأليلات في البحرين باستخدام نظام ®AmpFlSTR Identifiler للتفاعل البلمري المتسلسل والذي يتضمن 15 موقعاً عالمياً من المتكررات الترادفية القصيرة. كانت العينات عبارة عن مسحات فموية تم جمعها من 302 فرداً غير مرتبطين من مملكة البحرين. هؤلاء الأفراد تم اختيارهم عشوائياً من المجتمع البحريني للقيام بهذه الدراسة. وبعد عمل خطوات استخلاص الحمض النووي الريبوزي منقوص الأوكسجين، وتحديد كميته في العينات، وتكثيره بالتفاعل البلمري المتسلسل، ومن ثم تصنيف المتكررات الترادفية القصيرة، وعمل التحليل الإحصائي اللازم حصلنا على عدد من البيانات والنتائج القيمة. أعلى الأليلات في معدل تكرار حسب كل موقع كانت كالآتي: D21S11 : 29، D7S820 : 10 ، CSF1PO: 11، D3S1358: 16، TH01 :7 ، D13S317: 12، D16S539: 11 ، D19S433: 13 ، vWA: 17 ، TPOX: 8 ، D18S51: 14 ، D5S818: 12 ، FGA: 23 ، D8S1179 : 13 ، D2S1338 : 23. وأعلى معدل تكرار منهم كان على التوالي: 0.5 للأليل TPOX: 8، و0.3692 للأليل D5S818: 12 و 0.3626 للأليل D16S539: 11. أقصى تماثل زيجوتي كان 0.3576 وظهر في موقع TPOX. وفي المقابل، أقصى اختلاف زيجوتي كان 0.9106 وظهر في موقع D2S1338. ومثل ما متوقع كان أكثر المواقع تعدداً شكلياً أي لديها أكبر محتوى معلوماتي للتنوع الشكلي أعلاها تمييزاً وكانت على الترتيب: (0.9668) D18S51، (0.9717) D2S1338، FGA (0.9640)، بينما كانت أقلها قوة تمييزية: (0.7202) TPOX، فكان مدى القوة التمييزية يتراوح من 0.9718 لموقع D2S1338 و 0.7202 لموقع TPOX. كما تم تطبيق اختبار تي (t-test) على البيانات فكانت موزعة بصورة طبيعية البيانات فكانت موزعة بصورة طبيعية (Normally Distributed).

</div>

Introduction

DNA technology was first applied in solving forensic cases in 1986, and since then it was the way leading to personal identification, parentage and tracing the source of any biological samples found in the crime scene (Butler, 2006).

FBI has recommended using certain Short Tandem Repeat (STR) loci known as CODIS 13, and that have been accepted by most of the forensic labs worldwide. Some other labs use 15 loci of STR for more accuracy in personal DNA identification. Therefore those loci are used in Bahrain laboratories too to evaluate the accuracy of DNA expertise and the strength of any association through calculating Paternity Index (PI) which is the numerical summary of the strength of the genetic evidence for a match between an alleged father and a child (Butler,2010), or Matching Probability (MP). And there comes the need to establish and study the allele frequencies of a population and DNA database that are related directly to Bahrain.

Bahrain is an island with a total area of 757.50 km^2, it lies in the middle of the Arabian Gulf, near the East coast of the Kingdom of Saudi Arabia (figure1).

Figure 1: Map of Kingdom of Bahrain Shows its Boarders

Based on 2010 census, Bahrain has 1,234,571 population number with 568,399 Bahrainis and 666,172 non Bahrainis (Central Informatics Organization, 2010). Its location played an important role in shaping its history, politics and social structure. It has been a transit point between Arabia and the Persian and Indian sub-continents in many senses. Socially, this explains its cosmopolitan structure that is formed mainly of Arabs, Persians and Indians (Yarwood, 1988). And when a matching probability (MP) depends on the number of loci in DNA profile, Mutation Rates and population history, the need to calculate such a statistical significance in the Bahraini population became essential for the more understanding of the human diversity at the genetic level.

Materials and Methods

(1) Samples

Buccal swabs were collected from 302 unrelated Bahraini individuals. From 157 males and 145 females, as no age specifications were applied, Donors were Bahraini people involved in forensic cases which have been analyzed in the General Directorate of Forensic Science Evidence. Each donor gave his informed consent before samples collection.

(2) DNA Extraction

DNAs were extracted from epithelial cells present in collected buccal swabs using chelex-100® as chelating agent, or by the use of a robotic instrument which was Qiagen® EZ1 Advanced XL .

The buccal swabs were soaked in 1ml of de-ionized water in 1.5 ml eppendorf tubes for 10 min with frequent shaking of the swabs, centrifugation was done to the samples for 3 min at 13000 rpm, the supernatant was discarded and 170 µl of 5% chelex®-100 resins was added to the cells pellet and incubated for 25-30 min at 56°C, then brought to boil in water bath at 95°C for 8±1 min after vigorous shaking for few seconds using vortex, the samples were then centrifuged and supernatants were taken for further analysis.

For the EZ1 procedure, a cut from the tips of swabs were soaked in diluted 190 µl G2 buffer (1:1, G2: distilled water) with the addition of 10 µl protinase K in 2 ml sample tube and incubated at 56°C for 15 min in thermo mixer, swabs then removed and continued in the instrument with

protocol: DNA purification ("Tip Dance" Protocol) (Qiagen, 2009).

(3) Quantification

The Extracted DNAs were quantified using Quantifiler® Human DNA Quantification kit and applied on 7500 Real Time™ PCR system according to manufacturers' instructions (Applied Biosystem, 2005).

(4) Polymerase Chain Reaction (PCR) Amplification

AmpFlSTR Identifiler® PCR kit was used after applying the correct amount maintained from quantification step with accordance to the manufacturer's instructions (Applied Biosystem, 2001) and PCR was done on GeneAmp® PCR system 9700 in 0.2 ml PCR tubes, using the thermal cycle at these parameters: Hold at 95°C for 11 min followed by 25 cycles of 94°C for 1 min, 59°C for 1 min, 72°C for 1 min, then hold 60°C for 60 min, followed by cooling at 4°C, according to the instructions provided by the kit.

(5) Short Tandem Repeats (STR) Typing

After Polymerase Chain Reaction (PCR) step 1.5 µl of each sample was diluted in 8.5 µl Hi-Di formamide and 1.5 µl Gene Scan™-500LIZ internal size standard, and then analyzed using 3130 XL Genetic Analyzer, with 36 cm capillary, POP-4™ from Applied Biosystem with an allelic ladder in each run. Data collected and analyzed after using GeneMapper® ID software.

(6) Statistical Analysis

All data were collected after analyzing and were entered into Microsoft Office Excel 2003 and statistics then Calculated for the followings: Homozygosity (H), Heterozygosity (H2), Polymorphic Information Content (PIC), Power of Discrimination (PD), Paternity Index (PI).

Results and Discussions

The distribution of the observed allelic frequencies of the 15 STR loci in a representative sample Bahraini population can be seen in (table 1).

The most frequent allele types for each locus were: D8S1179: 13, D21S11: 29, D7S820: 10, CSF1PO: 11, D3S1358: 16, TH01: 7, D13S317: 12, D16S539: 11, D2S1338: 23, D19S433: 13, vWA: 17, TPOX: 8, D18S51: 14, D5S818: 12, FGA: 23. The highest allele frequencies from them were 0.5 for TPOX: 8, 0.3692 for D5S818: 12 and 0.3626 for D16S539: 11 (see, table1).

Maximum Homozygosity (H) was 0.3576 shown in TPOX locus. On the other hand, Maximum Heterozygosity (H) was 0.9106 shown in D2S1338 locus. The heterozygosity (H2) ranged from 0.6424 (TPOX), to 0.9106 (D2S1338), where D2S1338, D18S51, FGA had the highest values. As expected the most polymorphic loci which had the highest Polymorphic Information Content (PIC) were the most discriminating loci: D2S1338 (0.9717), D18S51 (0.9668), FGA (0.9640), while the least discriminated locus is TPOX (0.7202), as the discrimination powers (PD) ranged from 0.7202 (TPOX), to 0.9718 (D2S1338) (see, table2).

A t-test was applied to the data generated, and therefore it was normally distributed (figure2).
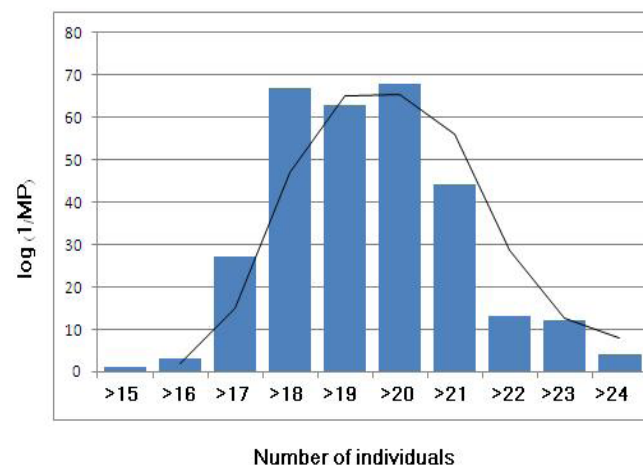
Figure 2: Normal Distribution Curve of the Data of 302 Individuals

As no deviations from Hardy-Weinberg equilibrium should be expected, a little departure from it was observed in the TH01 Locus with p-value of 0.0058 (< .0.05), where a rare allele among the population occurred in that locus lead to that value (see, table 2).

Table1: Distribution of Observed Allele Frequencies on Bahraini Population (N= 302)

Locus	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
4	-	-	-	-	-	0.0083	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	0.2235	-	-	-	-	-	0.0199	-	-	-
7	-	-	0.0298	-	-	0.2864	-	-	-	-	-	0.0083	-	-	-
8	0.0083	-	0.1838	0.0132	-	0.1308	0.1209	0.0331	-	-	-	0.5000	-	0.0166	-
9	0.0083	-	0.0745	0.0331	-	0.2268	0.0430	0.1523	-	-	-	0.1325	-	0.0579	-
9.3	-	-	-	-	-	0.1175	-	-	-	-	-	-	-	-	-
10	0.0877	-	0.2616	0.2914	-	0.0116	0.0480	0.1076	-	-	-	0.0993	0.0083	0.1093	-
10.2	-	-	-	-	-	-	-	-	-	-	-	-	0.0083	-	-
11	0.0861	-	0.2599	0.3096	-	-	0.3030	0.3626	-	0.0149	-	0.2219	0.0215	0.2467	-
12	0.1109	-	0.1606	0.2980	0.0083	-	0.3162	0.2185	-	0.0894	-	0.0199	0.1126	0.3692	-
12.2	-	-	-	-	-	-	-	0.0083	-	0.0116	-	-	0.0083	-	-
13	0.2301	-	0.0248	0.0430	0.0083	-	0.1242	0.1126	-	0.2682	0.0083	0.0083	0.1374	0.1921	-
13.2	-	-	-	-	-	-	-	-	-	0.0546	-	-	0.0083	-	-
14	0.2202	-	0.0083	0.0083	0.0563	-	0.0447	0.0083	0.0083	0.2384	0.1093	0.0083	0.1805	0.0083	-
14.2	-	-	-	-	-	-	-	-	-	0.0530	-	-	0.0083	-	-
15	0.1904	-	-	-	0.2368	-	-	0.0083	0.0083	0.1175	0.1242	-	0.1540	-	-
15.2	-	-	-	-	-	-	-	-	-	0.0646	-	-	-	-	-
16	0.0513	-	-	-	0.2831	-	-	-	0.0232	0.0397	0.2533	-	0.1341	-	-
16.2	-	-	-	-	-	-	-	-	-	0.0248	-	-	-	-	-
17	0.0116	-	-	-	0.2682	-	-	0.0083	0.1391	0.0083	0.2599	-	0.0993	-	0.0083
17.2	-	-	-	-	-	-	-	-	-	0.0199	-	-	-	-	-
18	-	-	-	-	0.1358	-	-	-	0.1325	-	0.1772	-	0.0778	-	0.0083
18.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
19	-	-	-	-	0.0149	-	-	-	0.1341	-	0.0579	-	0.0331	-	0.0629
19.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
20	-	-	-	-	0.0083	-	-	-	0.1308	-	0.0149	-	0.0083	-	0.0894
21	-	-	-	-	-	-	-	-	0.0679	-	0.0083	-	0.0166	-	0.1507
21.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
22	-	-	-	-	-	-	-	-	0.0596	-	-	-	0.0083	-	0.1606
22.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
23	-	-	-	-	-	-	-	-	0.1474	-	-	-	0.0083	-	0.1738

Table1:

Locus	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
23.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
24	-	-	-	-	-	-	-	-	0.0911	-	-	-	0.0083	-	0.1623
24.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
25	-	0.0083	-	-	-	-	-	-	0.0629	-	-	-	-	-	0.0911
26	-	-	-	-	-	-	-	-	0.0083	-	-	-	-	-	0.0546
27	-	0.0397	-	-	-	-	-	-	-	-	-	-	-	-	0.0132
28	-	0.1556	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
29	-	0.2384	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	0.1805	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
30.2	-	0.0166	-	-	-	-	-	-	-	-	-	-	-	-	-
31	-	0.0480	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
31.2	-	0.1060	-	-	-	-	-	-	-	-	-	-	-	-	-
32	-	0.0083	-	-	-	-	-	-	-	-	-	-	-	-	-
32.2	-	0.1440	-	-	-	-	-	-	-	-	-	-	-	-	-
33	-	0.0083	-	-	-	-	-	-	-	-	-	-	-	-	-
33.2	-	0.0480	-	-	-	-	-	-	-	-	-	-	-	-	0.0083
34	-	0.0083	-	-	-	-	-	-	-	-	-	-	-	-	-
34.2	-	0.0083	-	-	-	-	-	-	-	-	-	-	-	-	-
35	-	0.0083	-	-	-	-	-	-	-	-	-	-	-	-	-
36	-	0.0083	-	-	-	-	-	-	-	-	-	-	-	-	-
37	-	0.0083	-	-	-	-	-	-	-	-	-	-	-	-	-

((P-value) (H) *Observed Homozygosity;* (h) *Observed Heterozygosity;* (PIC) *Polymorphic Information Content ; ;* (PD) *Power of Discrimination;* (PI) *Paternity Index)*

Table 2: Observed Probability Value of Chi-square Test for Hardy-Weinberg Equilibirium (p-value), Observed Homozygosity (H), Observed Heterozygosity (h), Polymorphic Information Content (PIC), Power of Discrimination (PD), Paternity Index.

	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	THO1	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
p-value	0.4220	0.9389	0.4525	0.8957	0.9267	0.0058	0.0801	0.9950	0.4140	0.3490	0.6080	0.9220	0.9922	0.1653	1.0000
H	0.2020	0.1523	0.1821	0.2881	0.2517	0.1821	0.2583	0.2351	0.0894	0.2152	0.2417	0.3576	0.1192	0.2318	0.1523
h	0.7980	0.8477	0.8179	0.7119	0.7483	0.8179	0.7417	0.7649	0.9106	0.7848	0.7583	0.6424	0.8808	0.7682	0.8477
PIC	0.8106	0.8302	0.7674	0.6771	0.7322	0.7518	0.7389	0.7407	0.8740	0.8185	0.7783	0.6307	0.8636	0.7118	0.8578
PD	0.9368	0.9483	0.9068	0.8269	0.8791	0.8956	0.8772	0.8758	0.9718	0.9378	0.9146	0.7202	0.9668	0.8515	0.9640
PI	2.4754	3.2826	2.7455	1.7356	1.9868	2.7455	1.9359	2.1268	5.5926	2.3231	2.0685	1.3981	4.1944	2.1571	3.2826

((P-value) (H) *Observed Homozygosity;* (h) *Observed Heterozygosity;* (PIC) *Polymorphic Information Content ; ;* (PD) *Power of Discrimination;* (PI) *Paternity Index)*

During the study, two tri-allelic patterns were observed and confirmed with re-analysis of multiple samples for same individual, but not included in the data. One of them is (9,10,11) in TPOX locus which was already reported in NIST STRBase ,as it is studied that TPOX which occurs close to the tip of the chromosome 2 has the highest number of observed tri-allelic patterns *(Buttler,2006),* another pattern was (15,17,19) in vWA locus which is reported recently in the previous STRBase, and was a valuable finding in our data.

In this study an allele frequency for 15 autosomal STR Loci in a sample of the population of Bahrain was established for further use in Paternity testing and forensic case calculations, as we found some departure from the Hardy-Weinberg equilibrium in some loci among the population due to the rare alleles or new ones which appeared as a result of the mixed marriages or the emigrations to the island.

Acknowledgment

References

Applied Biosystems (2001) Ampflstr® Identifiler™ PCR Amplification Kit user's Manual, Foster City, CA, P/N 4323291, USA.
Available at: http://www.tools.lifetechnologies. co m/content/sfs/manuals/cms_039944.pdf

Applied Biosystems (2005) *Applied Biosystems 7300/7500/7800 Fast Real-Time PCR System, Allelic Discrimination Getting Started Guide, Foster City, CA, P/N 4347822 Rev.2,* USA. pp1-96.

Available at: http://www.fgf.uchicago.edu/ docs/73 00AQ.pdf

Butler JM (2006) Genetic and Genomics of Core Short Tandem Repeat Loci used in Human Identity Testing. *Journal of Forensic Science*, **51** (2): 253-265.
Available at: www.blackwell-synergy.com
Available at: http://www.cstl.nist.gov/strbase/ pub_pres/Butler2006JFS_coreSTRreview.pdf

Butler JM (2010) *Fundamentals of Forensic DNA Typing,* Academic Press of Elsevier, Burlington, USA, pp 1-500

Available at: http://www.booksite.elsevier.com/ samplechapters/9780123749994/.pdf
Available at: http://www.books.google.com. bh/books/about/Fundamentals_of_

Central Informatics Organization (2010) Population. Census Summary Results2010: Population, Housing, Buildings, Establishments and Agriculture Census, , pp 5.
Available at: http://www.cio.gov.bh/CIO_ARA /English/Publications/Census/2011%2009%20 18%20Final%20English%20Census%20201 0%20Summary%20%20Results%20-%20 Review%201.pdf

Lorimer J (1908) *Gazetteer of Persian (Arabian) Gulf.* Calcutta Press, New Delhi, India, pp1306-1308.

Qiagen (2009) *EZ1® DNA Investigator Handbook: For Automated Purification of DNA from Forensic and Biosecurity Samples using EZ1 Instruments,1056888 04/2009, 4th. ed.* QIAGEN Sample and Assay Technologies, Australia, pp1-64.
Available at: http://amanda-knox. wikispaces. com/file/view/.pdf

Yarwood JR (1988) *Al-Muharraq: Architecture, Urbanism and Society in a Historic Arabian. Town.* Faculty of Social Sciences, School of Architecture, University of Sheffield. (PhD Thesis).
Available at: http://www.etheses.whiterose. ac.uk/2994/