جامعة نايف العربية للعلوم الأمنية
NAIF ARAB UNIVERSITY FOR SECURITY SCIENCES
تأسست ١٩٧٨ Est. 1978

Naif Arab University for Security Sciences
Arab Journal of Forensic Sciences & Forensic Medicine
المجلة العربية لعلوم الأدلة الجنائية والطب الشرعي
https://journals.nauss.edu.sa/index.php/AJFSFM

الجمعية العربية لعلوم الأدلة الجنائية والطب الشرعي
Arab Society for Forensic Sciences and Forensic Medicine

Evaluation of Forensic Genetic Parameters of 24 STR Loci and Y indel in a Southern Region Saudi Population Sample Using GlobalFiler™ PCR Amplification Kit

تقييم معايير أدلة الحمض النووي لمواقع التكرار المترادفة القصيرة 24STR والحذف والإضافة على الكروموزوم Y لعينة من سكان جنوب المملكة العربية السعودية باستخدام طقم التكثير نوع GlobalFilter ™

Safia A. Messaoudi[1*], Malak A. Alamri[2], Saranya R. Babu[1], Abrar B. Alsaleh[1], Mohammed H. Albujja[1], Noora AlSnan[3], Mourad Assidi[4]

[1] Department of Forensic Sciences, College of Criminal Justice, Naïf Arab University for Security Sciences, Riyadh, Saudi Arabia.
[2] College of Science, Al-Imam Muhammed Ibn Saud Islamic University, Riyadh, Saudi Arabia.
[3] Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Kingdom of Bahrain.
[4] Center of Excellence in Genomic Medicine Research (CEGMR), Medical Laboratory Technology Department, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

Received 06 June 2021; Accepted 12 July 2021; Available Online 30 Dec. 2021

Abstract

The last three decades have seen rapid advances in the field of short tandem repeats (STRs) genotyping technology. Autosomal STRs have emerged as a powerful tool in forensic identification and paternity investigations. The indigenous population of Saudi Arabia is irregularly distributed and has historically been organized into geographically distinct groups or tribes of patrilineal descent. So far, there has been no detailed investigation of the southern region Saudi population to assist in the interpretation of DNA-based forensic evidence and in the construction of DNA database. The objective of this study is to investigate the genetic structure in 154

Keywords: Forensic Science; Forensic genetics; DNA typing; GlobalFiler™; Saudi population.

المستخلص

شهدت العقود الثلاثة الأخيرة تقدمًا سريعًا في حقل تقنية التنميط الجيني للتكرارات المترادفة القصيرة (STRs). فقد أظهرت التكرارات المترادفة القصيرة (STRs) الجسدية كأداة قوية في التعريف الجنائي والتحقيقات المتعلقة بالأبوة والنسب.

يتوزع السكان الأصليين للمملكة العربية السعودية بشكل غير منتظم، وقد تم تنظيمه تاريخيًا على شكل مجموعات محددة جغرافيًا أو قبائل تتبع لنسب الأبوة. لكن حتى الآن لم يتم إجراء أي تحقيق تفصيلي لسكان جنوب المملكة العربية السعودية يساعد في تفسير الدليل الجنائي الذي يعتمد على الحمض النووي، وكذلك في إعداد قواعد بيانات الحمض النووي.

إن الهدف من هذه الدراسة هو التحقق من التركيب الجيني لمجموعة مؤلفة من 154 شخص سعودي أصحاء البنية لا تربطهم علاقات قرابة

الكلمات المفتاحية: علوم الأدلة الجنائية، علم الوراثة الجنائي، السمات الوراثية للحمض النووي، طقم تكثير الحمض النووي GlobalFiler™، سكان المملكة العربية السعودية.

* Corresponding Author: Safia A. Messaoudi
Email: smassoudi@nauss.edu.sa

doi: 10.26735/UWKA5726

unrelated healthy Saudi subjects within three generations from the southern Saudi regions using a GlobalFiler™ PCR Amplification kit. Intra- and Inter-population genetic diversity as well as the forensic genetics parameters were analyzed. Our results showed that SE33 and TPOX loci were the most and the least polymorphic loci, respectively. The PIC, PE, TPI, Ho and He varied from 0.56116 (TPOX) to 0.94393 (SE33), 0.26638 (TPOX) to 0.83859 (SE33), 1.1875 (TPOX) to 6.33333 (SE33), 0.57894 (TPOX) to 0.92105 (SE33) and 0.6169 (TPOX) to 0.952 (SE33), respectively. The highest PM was observed for D22S1045 (0.223944) and the highest PD for SE33 (0.98935). The combined PD was 99.99999999% and the combined PM was equal to 3.19021E-25. Phylogenetic parameters showed that the southern region Saudi population had the closest genetic relationship with the Saudi, Emirati, Kuwaiti, and Bahraini populations. The study offers some important insights into the southern region Saudi population structure using GlobalFiler™ PCR Amplification kit.

ضمن ثلاثة أجيال جميعهم من المناطق الجنوبية للمملكة العربية السعودية وذلك باستخدام طقم تكثير الحمض النووي GlobalFiler ™ الخاص بتفاعل البلمرة المتسلسل. وتم تحليل معايير التنوع الجيني لهم وفيما بينهم وكذلك معايير علم الوراثة الجنائي. وقد أظهرت النتائج أن أنماط SE33 هي الأكثر انتشارًا وأن أنماط TROX هي الأقل انتشارًا بين التموضعات ذات الأشكال المتعددة. وقد تراوحت كل من (PIC, PE, TPI, HO, HE) بين نسبة 0.56116 لـ(TPOX) و 0.94393 لـ(SE33)، وكذلك 0.26638 لـ(TPOX) و0.83859 لـ(SE33)، و1.1875 لـ(TPOX) و6.33333 لـ(SE33)، و0.57894 لـ(TPOX) و 0.92105 لـ(SE33)، و0.6169 لـ(TPOX) و 0.952 لـ(SE33) على التوالي. وقد لوحظ أن أعلى نسبة PM كانت عند 0.223944 لـ D22S1045 وعلى نسبة PD عند 0.98935 لـ SE33. بينما كانت 99.99999999% لمجموع PD، بينما كان مجموع PM تساوي 3.9021E-25 . وقد أظهرت المعايير الوراثية أن سكان منطقة جنوب المملكة العربية السعودية لهم صلات هي الأقرب مع سكان كل من المملكة العربية السعودية، والإمارات العربية المتحدة والكويت والبحرين. وتقدم الدراسة أفكارًا ذات قيمة علمية عن البنية السكانية لجنوب المملكة العربية السعودية باستخدام طقم التكثير GlobalFiler ™ الخاص بتفاعل البلمرة المتسلسل.

1. Introduction

The Kingdom of Saudi Arabia (KSA) is in the Arabian Peninsula. It borders Bahrain, Qatar, and the United Arab Emirates to the east, Kuwait, Jordan, and Iraq to the north, and Oman and Yemen to the south. With a total area of about 2,149,690 square kilometers, the kingdom covers roughly four-fifths (4/5) of the Arabian Peninsula [1]. KSA is divided into 13 regions as shown in Figure-1 [2]. These regions differ in terms of their surface area and population.

The Saudi population is estimated at 34 million people, based on the General Authority for Statistics [3]. The population of KSA is unevenly distributed, with approximately 50% of the country's population concentrated in two main provinces, Mecca and Riyadh [4]. Native Arabs, who appear to make up approximately 63% of the population, are irregularly distributed and have historically been organized into geographically distinct groups or tribes of patrilineal descent [5], with consanguineous marriages [6]. Furthermore, the genetic structure of the Arab population has been influenced by the nearby African and Asian countries [7, 8].

Comprehensive development in the number of STRs has been demonstrated to be powerful for forensic purposes throughout recent eras [9]. It has increased the biased capacity to diminish the probability of unusual matches and to empower organizations to share data globally as adequately as expected under these circumstances [10]. The GlobalFiler™ Kit was created considering the interest for a multiplex measure that would catch a universally important arrangement of loci. Previous investigations evaluated the robustness, reproducibility, sensitivity and selectivity of the GlobalFiler™ Kit thorough the assessment of the different parameters improved by its structure and convention [11].

According to a recent study that investigated the GlobalFiler™ STR markers among a sample

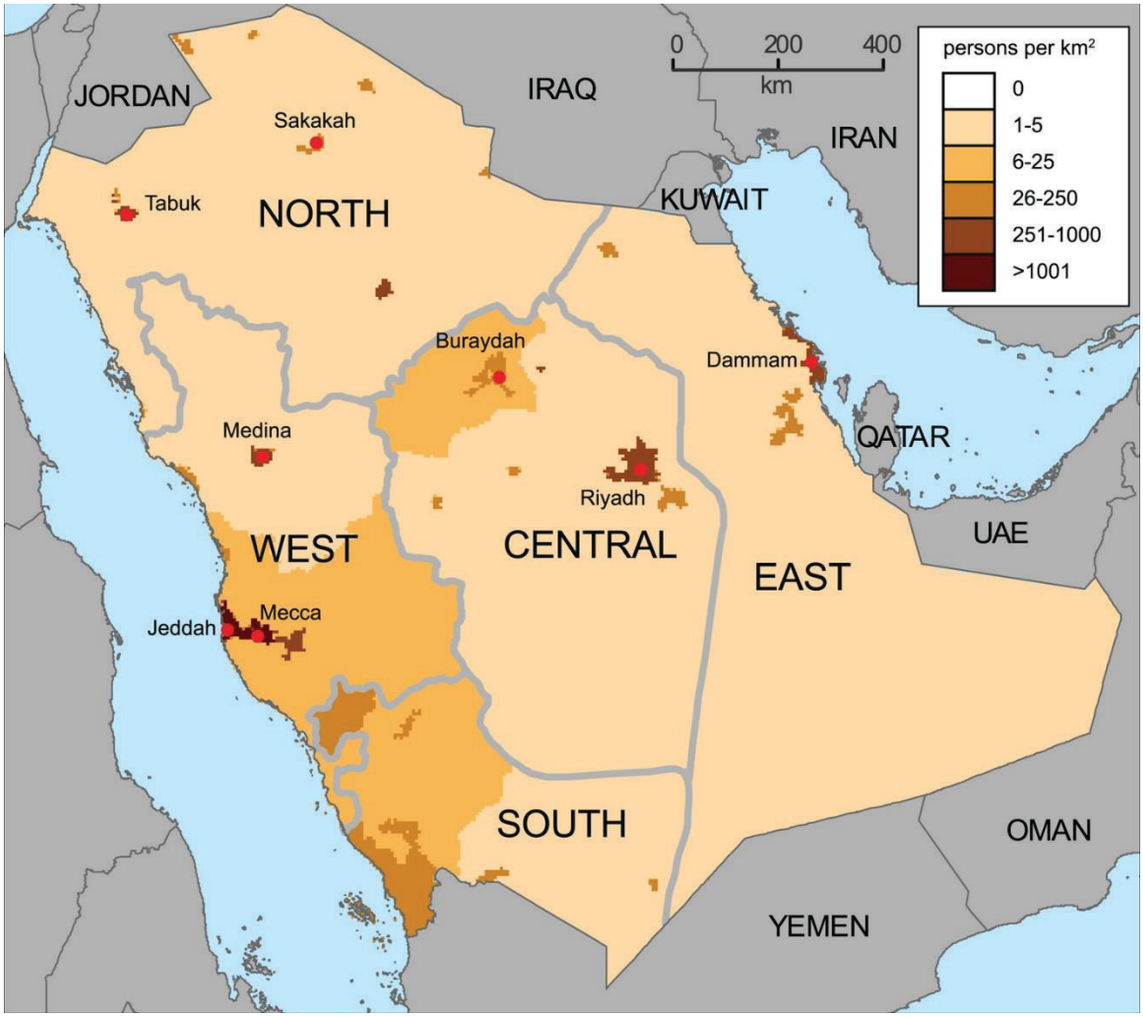

Figure 1- *Map of Saudi Arabia, showing population density and sub-regional divisions (Khubrani et al., 2018). Population density is indicated by shading as shown in the key, top right, and locations of some major cities are shown. Adapted from Global Rural-Urban Mapping Project (sedac.ciesin.columbia.edu/gpw/), under a Creative Commons 3.0 Attribution License*

of 253 indigenous Saudi males recruited from five geographic Saudi regions, the genetic distances between the southern and northern regions of Saudi Arabia were the greatest. In contrast, the central, Western, and Eastern regions showed similar genetic distances due to historical exposures and divergent tribal compositions and to immigration through the strait of Bab-el-Mandeb (the southern route) or the northern route that makes it a possible reservoir of genetic variants carried by these migrants [12].

No previous study has investigated the genetic diversity within the southern region Saudi population. For this reason, the current study was conducted to investigate the genetic structure of the southern region Saudi population to assist in the interpretation of DNA-based forensic evidence and the construction of appropriate DNA database using the 24 loci GlobalFiler™ PCR Amplification kit (Thermo Scientific, USA).

2. Materials and Methods

2.1 Sample collection

The Institutional Research Board of Naif Arab University for Security Sciences (NAUSS), Riyadh (KSA), has approved the study (IRB ref# NAUSS-Rec-21-03), according to the guidelines chartered by the 64th World Medical Association (WMA) General Assembly, Fortaleza, Brazil.

Buccal swabs were collected from selected healthy male (n=100) and female (n=54)) volunteers. A standard questionnaire including the demographic and medical history of the volunteers was prepared to collect the samples. Individuals having a history of bone marrow transplants, genetic disorders, and recent blood transfusions were excluded.

As mentioned in the Helsinki Declaration on Medical Research involving Human Subjects, objectives, methodology, and volunteer rights were explained to the volunteers and a full-informed written consent was obtained from each volynteer.

2.2 DNA extraction and quantification

Genomic DNA was extracted from buccal swabs using a QIAamp DNA Micro Kit following the manufacturer's protocols (Qiagen, Germany). Extracted DNA samples, as well as positive and negative controls, were quantified using the Quantifiler Duo Human DNA Quantification Kit (ThermoScientific, USA) as recommended by the manufacturer in 7500 Real-time PCR System (ThermoScientific, USA) to optimize the input DNA concentration in a multiplex PCR reaction.

2.3 DNA amplification and fragment detection

DNA samples as well as negative and positive controls (DNA Control 007) were amplified for 21 autosomal STR loci and three gender determination loci using a GlobalFiler™ Amplification Kit (ThermoScientific, USA), according to the manufacturer's recommendation. The PCR products were size separated in an ABI 3500 Genetic Analyzer (ThermoScientific, USA), with reference to the LIZ600 size standard v2 (ThermoScientific, USA), and the GeneMapper® ID-X Software v1.4 (ThermoScientific, USA) was used for genotype assignment.

2.4 Confirmation of tri-allelic pattern by next-generation sequencing (NGS)

Sample #154 was analyzed by Next-generation sequencing (NGS), using the Precision ID GlobalFiler™ NGS STR Panel v2 (ThermoScientific, USA) to confirm the existence of tri-allelic pattern. The library and template were prepared using the HID Ion Chef™ Instrument, and the sequencing was conducted on the Ion S5™ System. Torrent Suite™ Software V. 5.10 was used.

2.4 Statistical analysis

Exact test of Hardy–Weinberg equilibrium (HWE), Genetic Diversity (GD), allele frequencies, expected heterozygosity (He), observed heterozygosity (Ho) were estimated using GenAlEx 6.5 software) [13]. Pairwise Linkage disequilibrium (LD) was estimated between pairs of loci using STRAF [14]. Forensic efficiency parameters including Typical Paternity Index (TPI), Power of Discrimination (PD), Polymorphism Information Content (PIC), Probability of Matching (PM), and Power of Exclusion (PE) were estimated, using GenAlEx 6.5 software [13].

Interpopulation pairwise genetic distances between the Southern Saudi population and eleven populations data collected from the literature of Saudi Arabia [15], Bahrain [16], Qatar [17], Emirates [18], Kuwait, Iraq, Iran, Egypt, India, Sri Lanka and Bangladesh [19] were performed based on Fst and calculated utilizing POPTREE2 software [20].

A phylogenetic tree was also constructed using the neighbor-joining (NJ) method [21] with the POPTREE2

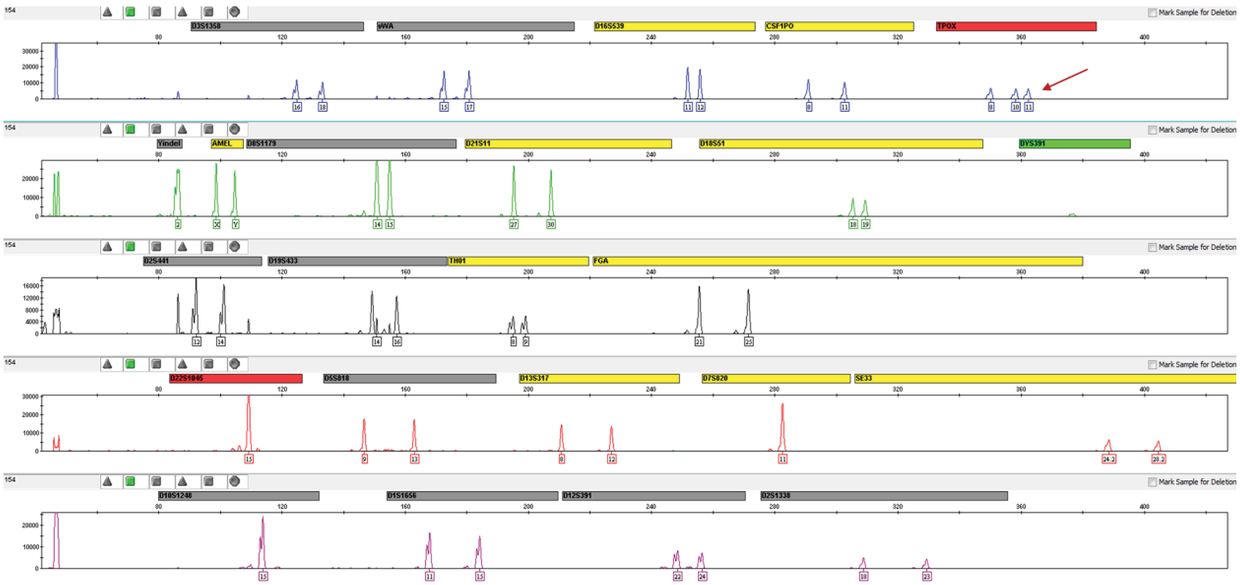

Figure 2- *Tri-allelic variants in TPOX locus observed in sample #154.*

software [20] for all populations in corrected fixation index (Fst) for phylogeny with 1000 permutations [22].

3. Results

3.1 Intrapopulation genetic diversity

In this study, p-value for Hardy-Weinberg equilibrium (HWE), allele frequencies, expected heterozygosity (He), observed heterozygosity (Ho) were performed for 21 autosomal STR markers: TH01D21S11, D7S820, CSF1PO, D3S1358, D13S317, D1S1656D16S539, D2S1338, D8S1179, D18S51, FGA, D12S391, D5S818, D2S441, D19S433, D10S1248, D22S1045, SE33, VWA, and TPOX in the southern region population of Saudi Arabia.

In our study, sample #154 showed tri-allelic variants (8,10,11) in locus TPOX with respective sizes of 350.03bp, 358.23bp, and 362.27bp in CE analysis (Figure-2). As shown in the supplementary file, NGS results showed a tri-allelic profile indicating three alleles (8, 10, 11). Consequently, population genetics and forensic statistics were performed only for 153 samples.

The average allele number per marker was 12, with a total number of discovered alleles equal to 216. SE33 was the most polymorphic with 31 alleles, and TPOX was the least with 6 alleles (Table-1).

All the studied autosomal markers were in HWE (p> 0.05), except for D21S11, D2S1338, TH01, and TPOX loci. However, after Bonferroni's correction was applied to improve the level of significance by dividing the alpha level number (0.05) by the number of autosomal loci (21) (p=0.05/21= 0.00238), no deviation from HWE was observed (Table-1).

The highest allele frequency was found for allele 8 of TPOX with a value of 0.585526, whereas the lowest allele frequency (0.003289) was observed in the following 26 alleles; 6.3, 9, 21.3, 22, 34.2, 35 of SE33, 8, 15 of CSF1PO, 13, 28 of D12S391, 23, 30, 36 of D18S51, 32, 34.2, 35.1 of D21S11, 12, 14, 26 of D2S1338,14.3, 18 of D1S1656, 8 of D22S1045, 8 of D5S818, 28 of FGA, 4 of TH01 and 9 of vWA marker (Table-1). The highest and lowest GD were observed for SE33 and TPOX with values of 0.949648 and 0.603266, respectively (Table-1).

Table 1- *Allele frequencies distribution and statistical parameters of forensic importance for 21 autosomal STRs loci (GlobalFiler™ kit) in 153 unrelated and healthy Saudi Southern region individuals.*

Allele	CSF1PO	D10S1248	D12S391	D13S317	D16S539	D18S51	D19S433	D1S1656	D21S11	D22S1045	D2S1338	D2S441	D3S1358	D5S818	D7S820	D8S1179	FGA	SE33	TH01	TPOX	vWA
4																			0.003289		
6																			0.299342	0.006579	
6.3																		0.003289			
7															0.016447				0.161184		
8	0.003289			0.177632	0.036184					0.003289				0.003289	0.207237			0.019737	0.115132	0.585526	
9	0.019737	0.019737		0.055921	0.167763							0.013158		0.049342	0.095395			0.003289	0.299342	0.177632	0.003289
9.3																			0.115132		
10	0.259868			0.042763	0.082237					0.006579		0.111842		0.128289	0.322368	0.046053			0.006579	0.092105	
11	0.388158	0.019737		0.138158	0.365132	0.032895	0.013158	0.059211		0.088816		0.348684		0.256579	0.190789	0.141447				0.125	
11.3												0.055921									
12	0.276316	0.036184		0.394737	0.200658	0.197368	0.134868	0.148026		0.013158	0.003289	0.108553		0.328947	0.157895	0.161184		0.006579		0.013158	
12.2							0.032895														
13	0.049342	0.197368	0.003289	0.154605	0.134868	0.210526	0.157895	0.111842				0.026316	0.009868	0.217105	0.009868	0.213816		0.036184			
13.2							0.032895														
14		0.309211		0.036184	0.013158	0.157895	0.230263	0.092105		0.029605	0.003289	0.292763	0.032895	0.016447		0.203947		0.095395			0.019737
14.2							0.069079														
14.3								0.003289													
15	0.003289	0.243421	0.016447			0.082237	0.141447	0.190789		0.506579		0.032895	0.180921			0.174342		0.019737			0.111842
15.2							0.075658														
15.3								0.023026													
16		0.125	0.009868			0.085526	0.072368	0.200658		0.322368	0.105263	0.009868	0.299342			0.039474		0.059211			0.223684
16.2							0.039474														

Table 1- Contd..

Allele	CSF1PO	D10S1248	D12S391	D13S317	D16S539	D18S51	D19S433	D1S1656	D21S11	D22S1045	D2S1338	D2S441	D3S1358	D5S818	D7S820	D8S1179	FGA	SE33	TH01	TPOX	vWA
16.3								0.046053													
17		0.049342	0.128289			0.085526		0.0625		0.029605	0.273026		0.325658			0.019737		0.085526			0.296053
17.3								0.042763													
18			0.1875			0.052632		0.003289			0.088816		0.138158				0.006579	0.088816			0.223684
18.3								0.016447													
19			0.144737			0.026316					0.108553		0.013158				0.052632	0.0625			0.115132
20			0.078947			0.046053					0.226974						0.115132	0.055921			0.006579
20.2																		0.009868			
21			0.101974			0.006579					0.016447						0.115132	0.016447			
21.2																		0.023026			
21.3																		0.003289			
22			0.128289			0.006579					0.016447						0.115132	0.003289			
22.2																		0.013158			
23			0.131579			0.003289					0.072368						0.194079				
23.2																		0.026316			
24			0.046053								0.055921						0.213816				
24.2																		0.039474			
25			0.019737								0.026316						0.118421				
25.2																		0.029605			
26											0.003289						0.046053				
26.2																		0.046053			
27									0.009868								0.019737				
27.2																		0.049342			
28			0.003289						0.184211								0.003289				
28.2																		0.055921			
29									0.273026												
29.2																		0.036184			
30						0.003289			0.233553												

Table 1- Contd..

Allele	CSF1PO	D10S1248	D12S391	D13S317	D16S539	D18S51	D19S433	D1S1656	D21S11	D22S1045	D2S1338	D2S441	D3S1358	D5S818	D7S820	D8S1179	FGA	SE33	TH01	TPOX	vWA
30.2									0.023026									0.036184			
31									0.046053												
31.2									0.085526									0.036184			
32									0.003289												
32.2									0.095395									0.023026			
33.2									0.032895									0.009868			
34.2									0.003289									0.003289			
35									0.006579									0.003289			
35.1									0.003289												
36						0.003289															
*MAF	0.357143	0.3125	0.192308	0.357143	0.357143	0.166667	0.227273	0.192308	0.192308	0.3125	0.192308	0.277778	0.357143	0.357143	0.357143	0.3125	0.227273	0.080645	0.357143	0.416667	0.3125
	Forensic parameters																				
* PM	0.140062	0.081631	0.03705	0.088729	0.085093	0.037396	0.038608	0.035925	0.065616	0.223944	0.049429	0.09072	0.103878	0.098078	0.083276	0.051247	0.038695	0.010648	0.095135	0.216672	0.079034
*CPM											3.19021E-25										
* PD	0.859938	0.918369	0.96295	0.911271	0.914907	0.962604	0.961392	0.964075	0.934384	0.776056	0.950571	0.90928	0.896122	0.901922	0.916724	0.948753	0.961305	0.989352	0.904865	0.783328	0.920966
* CPD											99.9999999%										
* PE	0.384915	0.532546	0.731478	0.567689	0.487527	0.692273	0.718334	0.628666	0.567689	0.384915	0.509766	0.532546	0.444732	0.498577	0.591734	0.567689	0.692273	0.838598	0.414156	0.266375	0.455216
* TPI	1.52	2.111111	3.8	2.30303	1.9	3.304348	3.619048	2.714286	2.30303	1.52	2	2.111111	1.727273	1.948718	2.451613	2.30303	3.304348	6.333333	1.617021	1.1875	1.767442
* PIC	0.647259	0.754673	0.86096	0.733608	0.740343	0.849286	0.850039	0.85458	0.792866	0.569101	0.814426	0.729626	0.709965	0.720426	0.749626	0.809919	0.841429	0.943932	0.731617	0.561156	0.753334
* GD	0.704924	0.788627	0.877019	0.765894	0.774383	0.866467	0.867075	0.871113	0.81935	0.631666	0.836308	0.765763	0.753648	0.762159	0.784914	0.834983	0.860322	0.949648	0.770779	0.603266	0.78867
* He	0.7023	0.7891	0.8813	0.7623	0.7755	0.8674	0.8703	0.872	0.8092	0.636	0.8347	0.7668	0.7532	0.7574	0.7843	0.837	0.8579	0.952	0.777	0.6169	0.7884
* Ho	0.671053	0.763158	0.868421	0.782895	0.736842	0.848684	0.861842	0.815789	0.782895	0.671053	0.75	0.763158	0.710526	0.743421	0.796053	0.782895	0.848684	0.921053	0.690789	0.578947	0.717105
* pHW	0.344	0.144	0.168	0.147	0.203	0.786	0.092	0.133	0.033	0.112	0.007	0.559	0.221	0.892	0.946	0.539	0.926	0.581	0.009	0.003	0.312

MAF: Minimum allele frequency based on NRC recommendations; * PM: probability of matching; PD: power of discrimination; PE: power of exclusion; TPI: typical paternity* index; PIC: polymorphism information content; Ho: observed heterozygosity; He: expected heterozygosity; pHW: Hardy–Weinberg equilibrium test (p>0.05);GD: Gene Diversity, CPD: combined power of discrimination, CPM: combined probability of matching

3.2 Forensic efficiency

Based on allele frequency data, we determined the statistical parameters of forensic interest. SE33 was the most informative locus with 31 alleles, whereas TPOX was the least with 6 alleles. In the current study, the TPI, PIC, PE, Ho and He varied from 0.56116 (TPOX) to 0.94393 (SE33), 0.26638 (TPOX) to 0.83859 (SE33), 1.1875 (TPOX) to 6.33333 (SE33), 0.57894 (TPOX) to 0.92105 (SE33) and He 0.6169 (TPOX) to 0.952 (SE33), respectively. The combined PD was 99.99999999% and the combined PM was equal to 3.19021E-25 (Table-1). In addition, the highest PM was observed for D22S1045 (0.223944), while the lowest was observed for SE33 (0.010648). Inversely, for PD, the highest was detected for SE33 (0.98935); however, the lowest was observed for D22S1045 (0.77605) (Table-1).

3.3 Linkage disequilibrium analysis

Linkage disequilibrium (LD) analysis was performed for all pairs of loci, and the results are presented in Table-2. Significant LD ($p \leq 0.05$) was observed for 4 pairs of loci as shown in Figure-3: CSF1PO-D2S1338, SE33-D12S391, D1S1656-D12S391, and vWA-D18S51. Therefore, Bonferroni's correction was performed to increase the significance level ($p \leq 0.05/210$ (0.00024). Statistically significant results were shown in 6 more pairs of loci: D12S391-D21S11, TPOX-SE33, D1S1656-SE33, D10S1248-FGA, D21S11-SE33 and D21S11-TPOX.

3.4 Interpopulation diversity

Based on Fst, interpopulation pairwise genetic distances (Table-3) between the Saudi southern region population and eleven published populations data including Saudi Arabia [15], Bahrain [16], Qatar [17] , Emirates [18], Kuwait, Iraq, Iran, Egypt, India, Sri Lanka and Bangladesh [19] were collected from the literature. The results of the population Fst were substantial, ranging from 0.003 to 0.042. As shown in Table-3, southern Saudi population, Emirati (Fst= 0.003) and Kuwaiti (Fst= 0.007) population are genetically close compared to other populations, followed by Iranian (Fst= 0.009) and Bahraini (Fst= 0.011). On the other hand, SriLankan (Fst= 0.040), Indian (Fst= 0.024), and Bangladeshi (Fst=0.016) revealed relatedness with each other and were genetically distant from the southern Saudi population.

Neighbour-joining (NJ) tree (Figure-4) was graphically generated using Nei's Fst distances based on the Fst data. The NJ tree showed that the indigenous people of the southern region of KSA was closely related to the Middle Eastern populations (rest of Saudi, Emirati, Kuwaiti, and Bahraini).

4. Discussion

In this paper, we utilized the 23 STR markers and one (1) Yindel included in the GlobalFiler™ PCR Amplification kit (Thermo Scientific, USA) to investigate the genetic structure of a Southern Saudi population.

Comparable results were observed with the previous research that studied the nearest population, including Saudis, Bahraini, Emirati, Kuwaiti populations utilizing GlobalFiler TM STR amplification kits [12, 15, 16, 19, 23].

The analysis of allele frequency data showed the evidence of a tri-allelic pattern in locus TPOX (8, 10, 11), which was also reported in the study with the UAE population (6, 8, 10) [23]. Tri-allelic pattern has been previously reported in many other loci. In the study conducted by Al-Snan et al., the tri-allele pattern was observed in both D21S11 (30, 31.2, 32.2) and D2S441 (10, 11, 12) markers [16]. In a Japanese study, a tri-allelic pattern (18, 25, 26) was reported in the D12S391 locus [24]. Alsafiah reported a tri-allelic pattern for D7S820 (9, 12, OL) in a study conducted in a Saudi population [15]. More-

Table 2 - *Pairwise linkage disequilibrium (LD) tested for all pairs of autosomal loci in southern Saudi individuals (n=153).*

	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
D2S1338	3.48E-01	7.03E-01	6.82E-01	1.84E-15	2.28E-01	1.46E-01	8.26E-01	5.57E-01	8.29E-01	1.01E-01	9.70E-01	1.30E-01	4.69E-01	5.37E-01	6.48E-01	3.53E-01	3.65E-01	1.88E-02	7.98E-01	3.85E-01	0
D12S391	1.48E-01	5.89E-01	1.43E-01	2.24E-01	5.58E-03	1.71E-01	1.61E-07	1.00E+00	9.89E-01	8.18E-01	7.04E-01	4.33E-01	9.52E-01	7.34E-01	4.85E-01	6.01E-01	1.97E-10	1.08E-01	4.06E-11	0	
D1S1656	4.83E-02	7.59E-01	4.78E-01	9.41E-01	1.35E-01	3.91E-02	3.35E-01	9.94E-01	2.52E-01	6.01E-01	2.39E-01	1.52E-01	1.76E-01	1.36E-01	2.97E-01	1.05E-01	6.54E-08	1.80E-01	0		
D10S1248	1.99E-01	6.57E-01	6.34E-01	5.60E-01	1.01E-01	2.02E-01	8.53E-01	6.04E-01	8.05E-01	4.71E-02	7.93E-01	5.87E-05	9.51E-01	6.31E-02	2.19E-02	6.31E-01	4.86E-01	0			
SE33	7.45E-01	8.34E-01	3.36E-02	6.58E-01	7.60E-04	3.56E-02	1.18E-04	2.87E-01	6.84E-01	6.70E-01	5.89E-01	7.69E-01	3.42E-01	3.45E-01	1.73E-01	7.18E-01	0				
D7S820	2.91E-02	4.75E-01	4.27E-01	5.13E-01	9.31E-01	5.07E-01	3.56E-01	2.91E-01	3.92E-01	9.83E-01	4.00E-01	4.32E-01	8.99E-01	9.69E-01	8.90E-01	0					
D13S317	8.75E-01	4.54E-01	5.62E-01	6.66E-01	1.05E-01	1.65E-01	1.32E-01	3.92E-01	9.96E-03	8.02E-02	9.48E-01	5.59E-03	6.59E-01	4.73E-01	0						
D5S818	3.42E-01	9.97E-01	3.67E-01	9.47E-01	7.40E-02	1.55E-01	3.38E-02	5.14E-01	5.38E-01	3.00E-01	8.72E-01	5.95E-01	9.49E-01	0							
D22S1045	8.71E-01	2.30E-01	9.69E-01	6.79E-01	8.04E-01	3.38E-01	9.92E-01	5.60E-01	4.44E-01	6.34E-01	9.98E-01	2.18E-01	0								
FGA	2.22E-01	7.65E-01	9.99E-01	7.87E-01	9.82E-01	4.39E-01	9.10E-01	4.81E-01	2.86E-01	3.43E-01	8.99E-01	0									
TH01	5.03E-02	4.77E-01	5.31E-01	3.23E-01	9.60E-01	9.15E-01	7.02E-01	3.43E-01	9.16E-01	2.69E-01	0										
D19S433	8.48E-01	7.52E-01	7.86E-01	4.12E-03	8.59E-01	2.30E-01	4.16E-01	9.48E-01	3.45E-01	0											
D2S441	6.61E-01	8.43E-01	9.83E-02	1.56E-01	9.66E-01	4.29E-01	5.37E-01	1.32E-03	0												
D18S51	2.51E-01	3.94E-09	8.53E-01	1.99E-01	5.81E-01	8.90E-01	9.70E-01	0													
D21S11	7.80E-01	9.69E-01	6.44E-02	4.05E-01	4.09E-05	1.02E-01	0														
D8S1179	3.64E-03	9.70E-01	3.85E-01	1.60E-01	1.84E-01	0															
TPOX	1.92E-01	9.71E-01	1.70E-01	9.60E-01	0																
CSF1PO	6.58E-01	8.10E-01	6.60E-01	0																	
D16S539	8.58E-01	7.13E-01	0																		
vWA	2.12E-01	0																			
D3S1358	0																				

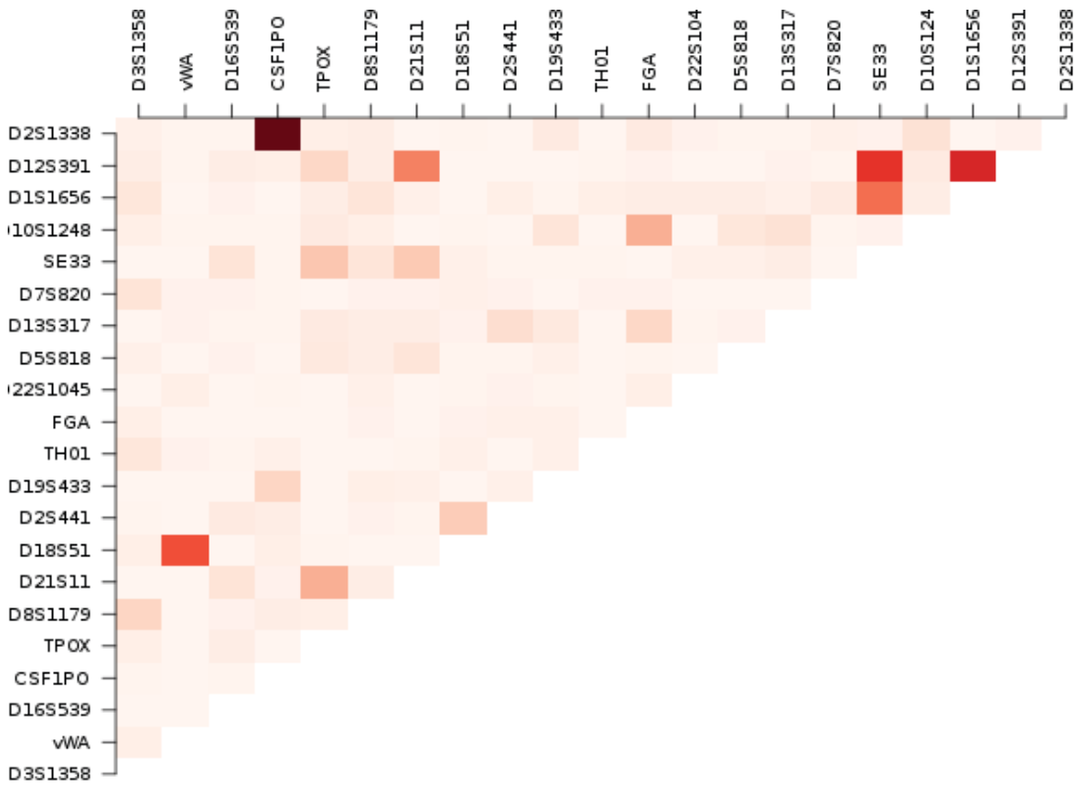

Figure 3- *Plot of Pairwise linkage disequilibrium (LD) of all pairs of autosomal loci.*

over, tri-allelic pattern was also observed in D18S51 (15.2, 16.2, 17.2) D1S1656 (8, 18, 19) in another Saudi population study [12].

We observed that four loci deviated from HWE (D21S11, D2S1338, TH01, and TPOX), which could be explained by the high level of consanguineous marriage rates among Saudi southern region population. This phenomenon is common in communities with a high rate of inbreeding like Saudi Arabia, Bahrain, and UAE, as documented in previous reports [12, 16, 18]. However, after Bonferroni's correction, no deviation from HWE was observed for all markers.

In this study, the most informative marker was SE33, whereas the least informative one was TPOX. These results are consistent with previous studies performed on Saudi population [12, 15] and other populations such as Emirati [23], Japanese [24], and Chinese [25]. Whereas TH01 and D16S539 were the least informative markers in Bahraini and Emirati populations, respectively [16, 18]. In the present study, the highest GD was observed for SE33; however, the lowest was observed for TPOX, matching the previous studies reported for Saudi population [12, 15].

In the present study, the highest PD was observed for locus SE33 (0.989352) which is similar to the literature [12-15]. However, the marker with the least PD value was D22S1045. This result was contrary to the finding from other populations such

Table 3- *Fst genetic distance matrix between Southern Region Saudi population and 11 published populations data (Alsafiah et al., 2017, Alsnan et al., 2019, Pérez et al., 2006, Jones et al., 2017, Al-enizi et al., 2013).*

	Bahraini	Qatari	Kuwaiti	Iraqi	Iranian	Egyptian	Bangladeshi	Srilankan	Indian	Emirati	Saudi	Southern Saudi
Southern Saudi	0.011	0.013	0.007	0.016	0.009	0.014	0.016	0.040	0.024	0.003	0.027	0
Saudi	0.035	0.034	0.033	0.040	0.031	0.039	0.037	0.042	0.042	0.029	0	
Emirati	0.008	0.010	0.006	0.013	0.006	0.011	0.009	0.032	0.016	0		
Indian	0.020	0.018	0.033	0.020	0.015	0.020	0.010	0.026	0			
Srilankan	0.039	0.031	0.042	0.032	0.034	0.026	0.031	0				
Bangla-deshi	0.015	0.021	0.018	0.020	0.014	0.019	0					
Egyptian	0.019	0.011	0.015	0.006	0.016	0						
Iranian	0.010	0.011	0.013	0.014	0							
Iraqi	0.011	0.015	0.017	0								
Kuwaiti	0.015	0.015	0									
Qatari	0.018	0										
Bahraini	0											

as Japanese [24], Chinese [25, 26], Bahraini [16], and Emirati [23], where TPOX was shown to have the lowest PD.

The values of PIC, TPI, and PE for all the investigated STR markers were highly informative, indicating a high level of genetic diversity. The highest values of these forensic efficiency parameters were observed for SE33, while the lowest was for TPOX. These results are consistent with the reports from Bahrain [16] and China [25, 26].

Tests of linkage disequilibrium (LD) were performed for all pairs of loci, and the results were presented in Table 2. Significant LD ($p \leq 0.05$) was observed after Bonferroni corrections was applied ($p \leq 0.05/210 = 0.00024$) for 10 pairs of loci (D2S1338-CSF1PO, D12S391-D1S1656, D12S391-SE33, D18S51-vWA, D12S391-D21S11, TPOX-SE33, D1S1656-SE33, D10S1248-FGA, D21S11-SE33 and D21S11-TPOX) as shown in Figure 3. The expectations of Linkage between loci were based only on physical distances. However, LD could be also a consequence of mutations, selection, founder effects, random genetic drift, and population admixture or stratification [27, 28].

We compared the southern region Saudi population data to 11 selected populations data; using the most accessible loci. The southern region Saudi population and Emirati population are the most genetically related, followed by Kuwaiti and Bahraini populations. The Fst distances of the southern region Saudi population with Emirati, Bahraini, and Kuwaiti were less than the statistics of FST distances < 0.01 [29], supporting the variations in allele frequencies between the studied populations that varied in terms of geography, language, and society. Therefore the present study aids in the classification of the geographically close middle Eastern populations [23].

The phylogenetic (NJ) tree construct, using fixation index FST value (Figure 4) among the 12 select-

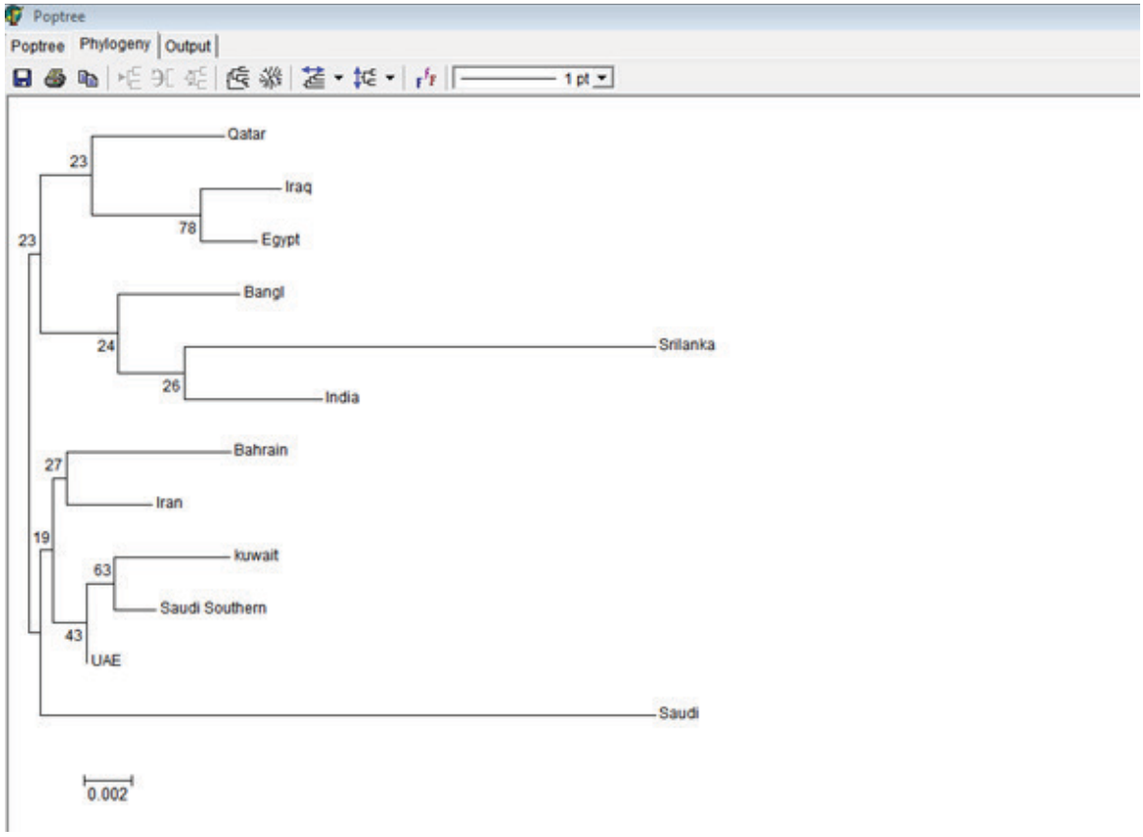

Figure 4- *Phylogenetic tree based on Nei's Fst Distances for the accessible autosomal STR loci estimated for the Southern Saudi population and 11 published populations data.*

ed populations, shows the close relatedness to the Middle Eastern populations (Saudi, Emirati, Kuwaiti and Bahraini). As expected, the genetic distances developed in this study between southern region Saudi and East Asian populations vary widely and might be due to geographic and continental origin distinctions.

5. Conclusion

The present research explores the genetic diversity of the southern region Saudi population utilizing Global Filer ™ PCR Amplification Kit. The results of the present study indicate that Global Filer™ PCR Amplification Kit is reliable for forensic DNA typing to provide higher power of discrimination (PD). SE33 was shown to be highly polymorphic and helpful for paternity testing as well as forensic casework. Overall, further region-specific population genetic studies exploring the northern, eastern, and western regions must be conducted to understand the genetic structure of the Saudi population.

Acknowledgments

This study was supported by Naif Arab University for Security Sciences (NAUSS). We would like to express our gratitude to NAUSS for their valuable help to complete this project and allowing us to utilize the laboratories of Forensic Science Department. We would also like to thank the volunteers for participating in this study.

Financial disclosure

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Authority GS. Statistical yearbook of 2015. 2015.
2. Khubrani YM, Wetton JH, Jobling MA. Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs. Forensic Science International: Genetics. 2018;33:98-105.DOI: https://doi.org/10.1016/j.fsigen.2017.11.015.
3. Statistics H. 1440. General Authority for Statistics Available online: https://wwwstatsgovsa/sites/default/files/haj40enpdf (accessed on 21 May 2020). 2019.
4. Arabia, G.A.F.S.I.S. Demographic Research Bulletin. Available online: https://www.stats.gov.sa/ (accessed on 24 March 2020). 2016.
5. Al-Rasheed M. A history of Saudi Arabia: Cambridge University Press; 2010.
6. Al-Gazali L, Hamamy H, Al-Arrayad S. Genetic disorders in the Arab world. Bmj. 2006;333(7573):831-4. DOI: https://doi.org/10.1136/bmj.38982.704931.AE
7. Abu-Amero KK, González AM, Larruga JM, Bosley TM, Cabrera VM. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC evolutionary biology. 2007;7(1):1-15. DOI: https://doi.org/10.1186/1471-2148-7-32
8. Abu-Amero KK, Hellani A, González AM, Larruga JM, Cabrera VM, Underhill PA. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. BMC genetics. 2009;10(1):1-9. DOI: https://doi.org/10.1186/1471-2156-10-59
9. Jordan D, Mills D. Past, Present, and Future of DNA Typing for Analyzing Human and Non-Human Forensic Samples. Frontiers in Ecology and Evolution. 2021;9:172.DOI: https://doi.org/10.3389/fevo.2021.646130.
10. Arenas M, Pereira F, Oliveira M, Pinto N, Lopes AM, Gomes V, et al. Forensic genetics and genomics: Much more than just a human affair. PLoS genetics. 2017;13(9):e1006960.
11. Ludeman MJ, Zhong C, Mulero JJ, Lagacé RE, Hennessy LK, Short ML, et al. Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. International journal of legal medicine. 2018;132(6):1555-73.DOI: https://doi.org/10.1007/s00414-018-1817-5.
12. Khubrani YM, Wetton JH, Jobling MA. Analysis of 21 autosomal STRs in Saudi Arabia reveals population structure and the influence of consanguinity. Forensic Science International: Genetics. 2019;39:97-102.DOI: https://doi.org/10.1016/j.fsigen.2018.12.006.
13. Peakall R, Smouse P. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics 28: 2537e2539. 2012.DOI: https://doi.org/10.1093/bioinformatics/bts460.
14. Gouy A, Zieger M. STRAF—a convenient online tool for STR data evaluation in forensic genetics. Forensic Science International: Genetics. 2017;30:148-51.DOI: https://doi.org/10.1016/j.fsigen.2017.07.007.
15. Alsafiah HM, Goodwin WH, Hadi S, Alshaikhi MA, Wepeba P-P. Population genetic data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit. Forensic Science International: Genetics. 2017;31:e59-e61.DOI: https://doi.org/10.1016/j.fsigen.2017.09.014.
16. Al-Snan NR, Messaoudi S, R. Babu S, Bakhi-

et M. Population genetic data of the 21 autosomal STRs included in GlobalFiler kit of a population sample from the Kingdom of Bahrain. PloS one. 2019;14(8):e0220620.DOI: https://doi.org/10.1371/journal.pone.0220620

17. Pérez-Miranda AM, Alfonso-Sánchez MA, Peña JA, Herrera RJ. Qatari DNA variation at a crossroad of human migrations. Human heredity. 2006;61(2):67-79.DOI: https://doi.org/10.1159/000092648
18. Jones RJ, Al Tayyare W, Tay GK, Alsafar H, Goodwin WH. Population data for 21 autosomal short tandem repeat markers in the Arabic population of the United Arab Emirates. Forensic Science International: Genetics. 2017;28:e41-e2.DOI: https://doi.org/10.1016/j.fsigen.2017.02.015
19. Al-Enizi M, Ge J, Ismael S, Al-Enezi H, Al-Awadhi A, Al-Duaij W, et al. Population genetic analyses of 15 STR loci from seven forensically-relevant populations residing in the state of Kuwait. Forensic Science International: Genetics. 2013;7(4):e106-e7. DOI: https://doi.org/10.1016/j.fsigen.2013.04.007
20. Takezaki N, Nei M, Tamura K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Molecular biology and evolution. 2010;27(4):747-52.DOI: https://doi.org/10.1093/molbev/msp312
21. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution. 1987;4(4):406-25.DOI: https://doi.org/10.1093/oxfordjournals.molbev.a040454
22. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution. 2018;35(6):1547.DOI: https://doi.org/10.1093/molbev/msy096
23. Ali Alhmoudi O, Jones RJ, Tay GK, Alsafar H, Hadi S. Population genetics data for 21 autosomal STR loci for United Arab Emirates (UAE) population using next generation multiplex STR kit. Forensic Science International: Genetics. 2015;19:190-1. DOI: https://doi.org/10.1016/j.fsigen.2015.07.009
24. Fujii K, Watahiki H, Mita Y, Iwashima Y, Kitayama T, Nakahara H, et al. Allele frequencies for 21 autosomal short tandem repeat loci obtained using GlobalFiler in a sample of 1501 individuals from the Japanese population. Legal Medicine. 2015;17(5):306-8.DOI: https://doi.org/10.1016/j.legalmed.2015.08.007
25. Adnan A, Zhan X, Kasim K, Rakha A, Xin XJ. Population data and phylogenetic structure of Han population from Jiangsu province of China on GlobalFiler STR loci. International journal of legal medicine. 2018;132(5):1301-4.DOI: https://doi.org/10.1007/s00414-018-1815-7
26. Li X, Li L, Wang Q, Zhang J, Ge W, Bai R, et al. Population genetic analysis of the Globalfiler STR loci in 3032 individuals from the Altay Han population of Xinjiang in northwest China. International journal of legal medicine. 2018;132(1):141-3.DOI: https://doi.org/10.1007/s00414-017-1641-3
27. Pereira V, Gusmão L, Valente C, Pereira R, Carneiro J, Gomes I, et al. Refining the genetic portrait of Portuguese Roma through X-chromosomal markers. American journal of physical anthropology. 2012;148(3):389-94.DOI: https://doi.org/10.1002/ajpa.22061
28. Chakravarti A. Population genetics—making sense out of sequence. Nature genetics. 1999;21(1s):56
29. Council NR. National science education standards: National Academies Press; 1996.