



Population Data and Internal Validation of the 21 Short Tandem Repeat Loci in Turkish Population

Emel Hulya Yukselglu¹, Fatma Cavus Yonar^{1*}, Omer Karatas¹, Gulten Rayimoglu¹, Faruk Ascioglu¹, Elif Canpolat¹, Onur Ozturk² and Itir Erkan³

¹Institute of Forensic Sciences and Legal Medicine, Istanbul University-Cerrahpasa, Istanbul, Turkey

²Department of Biophysics, Faculty of Medicine, Inonu University, Malatya, Turkey

³Department of Health Management, Faculty of Health Sciences, Istanbul Yeni Yuzylil University, Istanbul, Turkey

ABSTRACT

The purpose of this study was to identify paternity relations and increase the likelihood success in DNA identification of biological samples found in mass disasters and criminal events by using the 6-dye GlobalFiler™ amplification kit with 21 autosomal STR loci in our laboratory for validation purpose and routine use. In this regard, swab samples from different regions of Turkey (n=350) were collected and studied. The validation of the GlobalFiler™ Amplification Kit was carried out in accordance with the guidelines published by the Scientific Working Group on DNA Analysis Methods (SWGAM). Our study was conducted to show the efficacy and performance of the kit for sensitivity, specificity, limit of detection, dynamic range, limit of quantification, stochastic threshold study, reproducibility and repeatability, mixture and contamination study parameters. Validation parameters and population genetics data were calculated by using Arlequin v 3.5.2.2 version. According to our results, SE33 (PM=0.014) locus showed the greatest power of discrimination and TPOX (PM=0.132) has the least power of discrimination power in Turkish population. Combined power of exclusion and combined power of discrimination were calculated as 99.9999963% and 99.9999999%, respectively. In addition, allelic frequencies from our study were compared to previously published population data of Iraq, Saud Arabia, China (Han), USA (Cauc), Iran, Afghan, Azerbaijan, Romania, and South Africa. The study revealed that there was a significant difference between Turkish and South Africa (0.1198) population and no difference was observed for Azerbaijan (0.0097) population.

Article Information

Received 03 August 2021

Revised 25 September 2021

Accepted 07 October 2021

Available online 08 March 2022

(early access)

Authors' Contribution

FCY, EHY, OK and GR designed the study. FCY, GR and OK executed experimental work. FCY, OK, EC and OO analyzed the data. FCY, OK and IE wrote the paper. EHY and FA provided the laboratory equipment. EHY supervised the research.

Key words

Allele frequency, GlobalFiler™ kit, Internal validation, Population, STR markers

INTRODUCTION

Loci used in early forensic DNA analysis not only have low discrimination power but also they required high amount of DNA with high quality and extended analysis time. Therefore, alternative systems have been searched for better DNA profiling results. One of the efficiently used system is microsatellites, also known as short tandem repeats loci (STRs), composed of repeat units with a length of 1-6 base pairs which have a high power of discrimination (McMahon *et al.*, 2017). The allele size of the STR loci is between 60 to 470 base pairs (Butler, 2012). STRs are ideal genetic markers in forensic sciences for not requiring expensive analysis equipment and possibility of identification of degraded biological samples with automation and multiple analysis (Ezkurdia *et al.*, 2014).

Thirteen core autosomal STR loci (CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11) could be found in any DNA database or the Combined DNA Index Systems (CODIS) that is under the control of the US Federal Bureau of Investigation (FBI) (Connie Milton, 2017; Karantzali *et al.*, 2019). The loci found in the non-coding DNA sequences are used for analysis with commercially available multiplex polymerase chain reaction kits (Gill *et al.*, 2006). Kits used in routine analysis such as Identifiler®, Identifiler Plus®, PowerPlex® comprised of additional loci to enlighten important cases. However, in some cases such as degraded biological samples, paternity-maternity analysis, missing person cases or mutations in DNA, conventionally used STRs might not be efficient enough for good profiling results.

On the other hand, the GlobalFiler™ kit allows for the multiplication of 16 autosomal STR loci as well as additional 8 STRs (D2S441, SE33, DYS391, Y-Indel, D10S1248, D1S1656, D12S391, 22S1045) total of 24

* Corresponding author: fatma.cavus@iuc.edu.tr
0030-9923/2022/0001-0001 \$ 9.00/0
Copyright 2022 Zoological Society of Pakistan

STR loci- in a single tube and with a single polymerase chain reaction with primers marked with five different (6-FAM™, NED™, TAZ™, SID™, VIC®) fluorescent dyes and size standard is marked with sixth dye GeneScan™ 600 LIZ™ Size Standard v2.0 (Thermo Fisher Scientific).

Moreover, the GlobalFiler™ kit which has at least 10 loci contain mini STRs with an amplicon length less than 220 base pairs, which makes it suitable for forensic scenarios including mixed samples. The problem with the mixed samples is that allele loss might be observed in Amelogenin. For instance, deletion in Y-chromosome may bring about false sex determination results. In order to prevent this problem, the GlobalFiler™ kit includes DYS391 and Y-Indel loci specific to gender determination. Furthermore, the kit includes a PCR reaction mixture that offers higher performance and there is an increase in the success rate of the kit even with samples including inhibitors. In addition to that, the GlobalFiler™ kit provides not only good DNA profiling results but also the amplification time of the kit is less than 80 min.

The validation of the GlobalFiler® Kit was performed in accordance with the guidelines published by the Scientific Working Group on DNA Analysis Methods (SWGDM). Studies were performed to demonstrate the effectiveness and performance of the GlobalFiler® Kit for sensitivity, concordance, specificity with limit of detection, dynamic field study, limit of quantification, stochastic threshold study, sensitivity, reproducibility and repeatability, mixture and contamination study parameters. The results validate the multiplex design as well as demonstrate the kit's robustness, reliability, and suitability as an assay for human identification with casework DNA samples.

One of the significant aspects of this study is that Turkey is an important geopolitical country that is the bridge between cross-section of Asia and Europe. Migrations occur inside and out in the several regions of Turkey, which requires mapping the allele frequencies and comparing the data to others published ones. In our country, there are several population studies; however, to the best of our knowledge, there has not been any published study that includes 24 STRs (20 STR loci + 4 gender loci) by using GlobalFiler® amplification kit.

This study aims to expand national population data with frequencies and population statistic parameters for 16 core STR's (CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11) and 8 new STR loci (D2S441, SE33, DYS391, Y- Indel, D10S1248, D1S1656, D12S391, 22S1045). The purpose of the study is conduct the internal validation of GlobalFiler® Amplification Kit with 21 autosomal loci in our laboratory as well as performing a

Turkish population study.

MATERIALS AND METHODS

Oral swab samples were collected and studied for the purpose of population genetic study and internal validation. All the procedures were carried out in accordance with Istanbul University-Cerrahpasa Medical Faculty Clinical Research Ethics Committee after the approval (No. 69885, 06/03/2015). All participants were informed about the study and signed a full informed consent. Oral swab samples were collected from 350 randomly selected individuals (200 men, 150 women) from seven different regions of Turkey (Marmara, Aegean, Mediterranean, Southeastern Anatolia, Eastern Anatolia, Black Sea and Central Anatolia regions) who were aged over 18 years old. Samples were taken from the mouth by using sterile swabs and dried out at room temperature. For method validation, oral swabs were collected from the personnel (3 women, 1 men) of Istanbul University-Cerrahpasa, Institute of Legal Medicine and Forensic Sciences, Forensic Molecular Genetic Laboratory. Given that, our laboratory is accredited according to TSE/IEC 17025 standard; all samples of the laboratory personnel are included in the validation as a standard requirement. All the samples were stored at -20° C prior to analysis. DNA isolation was performed using silica based QIAamp™ DNA Mini Kit (Qiagen, Stanford, CA, USA) using the swab samples collected from volunteers as previously described ([QIAGEN Sample and Assay Technologies, 2016](#)). Quantity of DNA isolates was detected in Qubit™ fluorometer (Applied Biosystems) by using the Quant-iT dsDNA HS (High Sensitive) Assay kit (Invitrogen, Paisley, Renfrewshire, UK) following the steps mentioned (Thermo Fisher Scientific, n.d.). The validation parameters studied by guidelines published by the Scientific Working Group on DNA Analysis Methods (SWGDM). The amount of DNA measured by the fluorimetric method ranged between 1.03 and 12.82 ng/μl.

After isolation and quantification of the samples, GlobalFiler™ PCR Amplification Kit stored at -20°C was used for polymerase chain reaction. Thermal cycling parameters were optimized using an experimental design approach to define the combination of temperatures and retention times that produce the best test performance. Controlling contamination in the polymerase chain reaction stage, negative control sample was used for each replication. In addition, assessing the optimal sensitivity of the method, positive control sample was used ([Thermo Fisher Scientific Inc., 2019](#)). As internal controls oral swab samples from personnel (3 women, 1 men) were used to diminish the individual sample errors. PCR analyses

were carried out using GeneAmp 9700 (Thermo Fisher Scientific Company Carlsbad, USA).

At the electrophoresis stage, capillary electrophoresis device with 4 capillary ABI PRISM® 3130 Genetic Analyzer (Applied Biosystems Foster City, California) was used. Imaging, typing and evaluation of the raw data from the samples after electrophoresis of 24 loci (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391, D2S1338) in ABI 3130 were carried out by using the GeneMapper® ID-X software v1.4 (Applied Biosystems) analysis program (Thermo Fisher Scientific Inc, 2019). For the validation study, each sample was analyzed by each analyst for 3 times.

The validation parameters such as limit of detection (LOD), limit of quantification (LOQ), dynamic range, stochastic threshold, repeatability, reproducibility were calculated by using Microsoft Excel 2010 bundled with Microsoft Office package.

In order to determine the limit of detection of the ABI 3130 genetic analyzer, Relative Fluorescent Unit (RFU) values of the highest peaks observed in the 10 negative control samples which were conducted and analyzed in accordance with the operating and software instructions of the device. The mean and standard deviation of the RFU values of the highest peaks obtained from all samples were calculated. According to the calculation, the cut off value was determined as 50.25 RFU.

Population and genetic data of 21 autosomal STR loci parameters such as matching probability (MP), power of discrimination (PD), polymorphism information content (PIC), power of exclusion (PE) were calculated by using Arlequin v 3.5.2.2. In addition, observed heterozygosity (H_o) and expected heterozygosity (H_e) were calculated as well as the discriminative loci between the inter-population differences with the help of Bonferroni correction ($p < 0.05/21 = 0.00238$) assessed. Moreover, inter-population genetic distances were calculated by using Wright's F-statistics (Fixation Index- F). Allele frequencies and the Z values of allele frequencies from other countries were compared and calculated based on the peak heights of alleles by using Arlequin v 3.5.2.2 (Excoffier and Lischer, 2010).

Ethical approval

This study protocol was approved by ethical committee of the Istanbul University Cerrahpasa Medicine Faculty, Istanbul, Turkey (No. 69885, 06/03/2015) and all the participants consented to use their genetic material and other necessary information.

RESULTS

Positive control samples were prepared for eight different dilution levels (0.005 ng/ μ l, 0.0125 ng/ μ l, 0.025 ng/ μ l, 0.05 ng/ μ l, 0.125 ng/ μ l, 0.25 ng/ μ l, 0.5 ng/ μ l, 1 ng/ μ l) and each concentration level was carried out with 5 PCR and a total of 40 PCRs were performed for dynamic range. The average peak heights were obtained. Then there were 5 amplified samples in each concentration group, and average was calculated for each group and Figure 1 was obtained with the mean peak height versus each concentration. According to our study, dynamic range was determined between 0.05 ng/ μ l to 0.25 ng/ μ l.

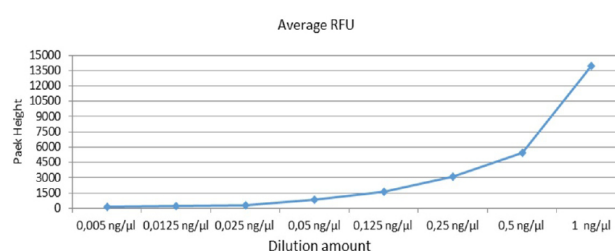


Fig. 1. Dynamic field determination for samples carried out on ABI 3130.

Determining the limit of quantification (LOQ), the applications of 40 PCR products used in the dynamic range study were examined from the lowest to the highest concentration and the sensitivity was determined for the two lowest concentrations obtained in the full profile. For this purpose, the mean and standard deviation of the peak heights of each allele detected in 5 samples in each dilution levels (0.005 ng/ μ l, 0.0125 ng/ μ l, 0.025 ng/ μ l, 0.05 ng/ μ l, 0.125 ng/ μ l, 0.25 ng/ μ l, 0.5 ng/ μ l, 1 ng/ μ l) were calculated. The values obtained for each allele were also averaged and compared with the limit of detection found for the genetic analyzer. The comparison results show that a complete and reliable DNA profile is obtained with a concentration of 0.125 ng/ μ l.

To find the reproducibility and repeatability, oral swab was taken and isolated from 4 people. These samples were run 3 times at different times by 4 people using the same PCR-total of 48 samples and genetic analyzer and their electrophoresis was recorded. For the reproducibility study, the standard deviation of the allele peak size was also calculated. For all loci obtained from the sequencing standard deviation values were averaged. This value is expected to be less than 0.5. For the ABI 3130 device, 7 samples and 21 sequencing results, the average value obtained from the standard deviation of all loci is 0.337.

For repeatability studies, the results of genotypes

were compared with the same sample in our laboratory and the Ministry of Justice Istanbul Forensic Medicine Department of Biology. "Sample No: 18/1 Preparation T: 28.02.2018" coded oral swabs' isolation was profiled in the Forensic Molecular Genetics Laboratory with Manual Silica Membrane Method (Qiagen Mini Kit) and at the Ministry of Justice Istanbul Forensic Medicine Department of Biology with M48 (Qiagen) method. DNA quantification was determined by Quant-it ds DNA HS Qubit 2.0 in Forensic Molecular Genetics Laboratory and by Quantifiler Duo (ABI) ABI7500 RealTime PCR in the Ministry of Justice Istanbul Forensic Medicine Department of Biology. As a PCR Kit, the GlobalFiler™ Kit (24 STR Loci) was used in both laboratories. The PCR products were sequenced in ABI 3130 in Forensic Molecular Genetics Laboratory and analyzed using IDX Gene Mapper program. The PCR products obtained from the Ministry of Justice Istanbul Forensic Medicine Department of Biology were sequenced on the ABI 3500 device and analyzed with the GeneMapper program. The results obtained from the two laboratories were evaluated for 21 autosomal STR loci and results were consistent with each other.

After that, the ratio of allelic peak heights to heterozygous gene regions was found in samples that are profiled and amplified. The stochastic threshold is 0.5 ng/μl (Table I). Then, the values for each allele of each sample and the small peak heights used to find these values are obtained and this is showed in Figure 2. The stochastic threshold can also be expressed as the RFU imbalance points of the target DNA concentration as seen in Figure 3.

Table I. Stochastic threshold determination.

Target DNA concentration	The mean values of allele pairs peak heights ratio	The threshold values for allele pairs peak heights
1 ng/μl	93.25±7.70	70.1559
0.5 ng/μl	90.07±4.42	82.8066
0.25 ng/μl	92.44±3.75	81.1943

The ratio of allele peak heights to heterozygous gene regions was found in the amplified and profiles samples. For this, the small peak was divided by the large peak height and multiplied by 100. The mean and standard deviation of the sister allele peak heights were found. From these averages, one threshold was obtained for the ratio of peak heights by removing three times from standard deviations. As a result of this study, the mean height of sister alleles was found to be 90.004, the standard deviation ± 7.396 and the sister allele peak height were 67.815 as in Figure 4.

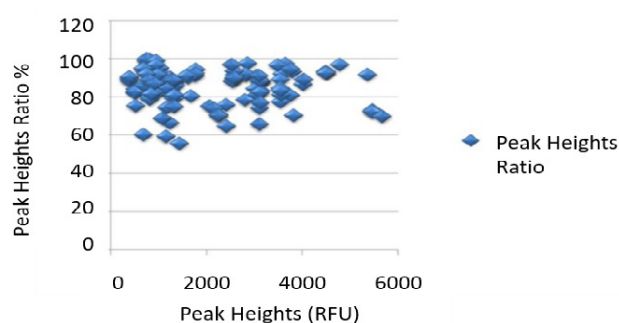


Fig. 2. Peak height versus rate of allele pairs peak heights.

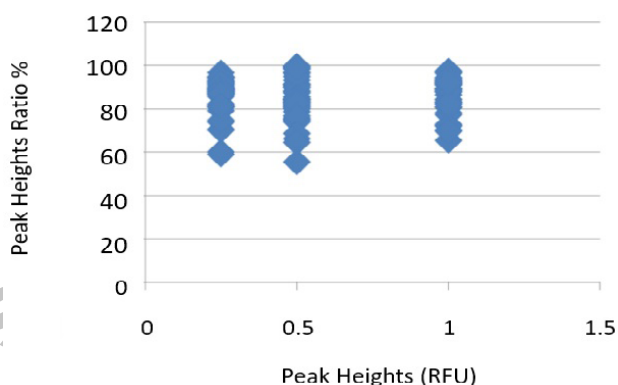


Fig. 3. Sample allele pairs peak heights.

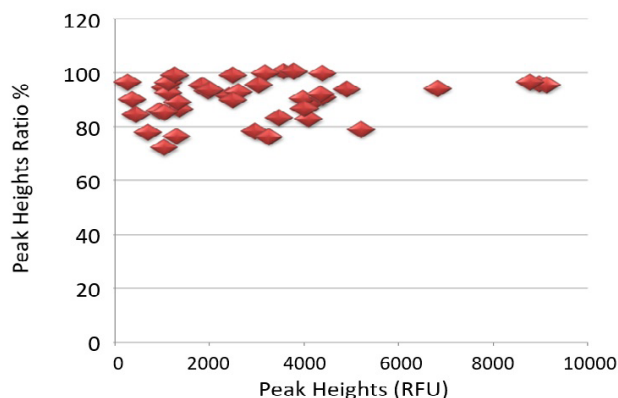


Fig. 4. Peak height versus the rate of allele pairs peak heights.

For mixture studies, two previously known F and G coded samples of the profiles were mixed in a ratio of 1: 1, 1:10, 1:50, 1: 100. As a result of the examination, the observed and expected allele numbers of the profile of both persons were determined (Table II). As shown in Table III, the full profile could not be obtained in the F coded sample

with low dilution, in 1:10, 1:50, 1: 100 dilutions. The full profile could be obtained when the volume of the minor and major component volumes was 1:1.

Table II. Observed and expected allele numbers of the mixture samples.

Mixture ratio (G:F)	Expected allele numbers	Observed allele numbers	Expected allele numbers	Observed allele numbers
Sample name	G	G	F	F
1:1	42	42	44	44
1:10	42	42	44	38
1:50	42	42	44	20
1:100	42	42	44	10

Table III. Sample tray and sample run order.

Sample run order	1	2	3
A	Sample	Allelic ladder	Liz 500
B	Sample	Liz 500	Sample
C	Sample	Sample	Liz 500
D	Allelic ladder	Liz 500	Sample
E	Sample	Sample	Liz 500
F	Sample	Liz 500	Sample
G	Sample	Sample	Liz 500
H	Allelic ladder	Liz 500	Allelic ladder

To determine if there was a cross contamination between the wells during all experimental stages and during the preparation of the sample loading tray in the ABI 3130 genetic analyzer, first, negative control was used during the isolation phase and at the stage of PCR and electrophoresis both negative and positive controls were used. No peak was determined in the electrophoresis obtained from the negative control used in the PCR stage and no different peak was observed except the expected and known peak in the positive control. Thus, no contamination was detected during the experiments. In the second part, Gene Scan 500 Liz Size Standard and duplicated products were included in one of the wells located next to each other and only Gene Scan 500 Liz Size Standard was placed next to them. Hence empty and filled wells were located diagonally. Allelic ladder for each sequencing was also added. The sample tray shown in Table III was loaded on the ABI 3130 genetic analyzer and analyzed. For wells containing samples, we analyzed

whether there was any contamination from the wells and samples containing the Gene Scan 500 Liz Size Standard from the Allelic Ladder. No contamination was detected.

By using GlobalFiler™ PCR Amplification Kit, we have examined 68 different alleles for Turkish population. In addition to that, 44 allelic variants were identified in SE33 locus as shown in Table IV. All the SE33 variants are listed in STRBase (Butler and Redman, 1997).

In our study, significant deviations from Hardy-Weinberg equilibrium expectations ($p < 0.05$) were resulted in vWa ($p = 0.00799$), D19S433 ($p = 0.00018$), TH01 ($p = 0.0039$), D22S1045 ($p = 0.01762$), SE33 ($p = 0.03518$) and D1S1656 ($p = 0.00073$) loci. However, after applying Bonferroni correction ($p < 0.05/21 = 0.00238$), no statistical difference was observed from HWE except for D19S433 ($p = 0.00018$) and D1S1656 ($p = 0.00073$) loci.

According to our results, H_o values ranged from 0.700 (TPOX and D5S818) to 0.920 (SE33) for Turkish population. In addition, H_e values ranged from 0.696 (TPOX) to 0.955 (SE33). These results indicate that TPOX and D5S818 have the lowest discrimination power among 21 autosomal loci since H_e and H_o is used for calculating allele frequencies in population studies. Moreover, H_e and H_o values of SE33 demonstrates the discrimination power of the locus in forensic cases as seen in Table V.

Similar to other population studies (Alsafiah *et al.*, 2017; Park *et al.*, 2016; Zhang *et al.*, 2016), our results indicated that TPOX (PD= 0.868) has the least discriminative power among 21 autosomal loci whereas SE33 (PD= 0.986) has the most discriminative one. Similar results were observed for TPOX (TPI= 1.667) being less useful compared to other loci and SE33 would contribute the most for paternal cases statistically. Thus, according to our data, TPOX is the least polymorphic locus and SE33 was concluded to be the most polymorphic one as seen in Table VI.

According to our results, genetic distance between Turkish population and South African population (0.1198) is the most distinct whereas genetic distance between Turkish population and Azerbaijan (0.0097) is found to be the closest to one another as seen in Table VII.

DISCUSSION

In criminal cases such as murder, sexual assault and theft, detection of the link between crime and criminal, genetic information from biological materials at the scene is used to determine whether there is a connection between the suspect and the incident (Rudin and Inman, 2001).

The rapid development of DNA technology, working with smaller materials, making more accurate and reliable

Table IV. Allele frequencies of 21 autosomal loci profiled with GlobalFiler™ PCR Amplification Kit for Turkish population (N=350).

Allele	D3S 1358	vWa	D16 S539	CSF IPO	TPOX	D8S 1179	D21 S11	D18 S51	D2S 441	D19 S433	TH01	FGA	D22 S1045	D5S 818	D13 S317	D7S 820	SE33	D10 S1248	D1S 1656	D12 S391	D2S 1338		
5					0.007						0.100												
6					0.005	0.007	0.005				0.225						0.009						
6.3																	0.003						
7					0.005	0.005	0.019	0.005			0.165					0.003	0.030	0.005					
7.3						0.007					0.003												
8					0.063	0.010	0.475	0.015			0.145					0.005	0.010	0.190	0.005				
9					0.140	0.070	0.115	0.012		0.005	0.001	0.215			0.051	0.094	0.110	0.005					
9.3											0.215												
10					0.005	0.114	0.325	0.120	0.070		0.012	0.150			0.010	0.062	0.099	0.265	0.015	0.010			
11						0.345	0.370	0.220	0.080		0.023	0.330	0.010	0.019	0.125	0.391	0.300	0.255	0.010	0.010	0.090		
11.2											0.015												
12					0.005	0.220	0.360	0.060	0.140		0.149	0.135	0.100		0.010	0.353	0.305	0.195	0.005	0.030	0.135	0.010	
12.2													0.135										
12.3											0.010												
13					0.005	0.145	0.065	0.005	0.265		0.122				0.005	0.151	0.250	0.053	0.015	0.210	0.090	0.005	0.010
13.2												0.289											
13.3										0.005	0.005	0.341								0.005			
14					0.135	0.100	0.035	0.010	0.205		0.158	0.240	0.035		0.005	0.080	0.003	0.035	0.009	0.010	0.325	0.080	0.005
14.2												0.035											
14.3											0.015									0.030			
15					0.250	0.095		0.130			0.142	0.080	0.158		0.006	0.005	0.050	0.200	0.120	0.045	0.005		
15.2											0.037												
15.3											0.005									0.030			
16					0.235	0.295		0.075			0.136	0.003	0.050		0.001	0.285	0.055	0.165	0.140	0.030	0.035		
16.2											0.020												
16.3												0.050								0.030			
17					0.235	0.275		0.020			0.123	0.007	0.003		0.010	0.080	0.080	0.045	0.065	0.110	0.165		
17.2												0.015											
17.3																				0.080	0.025		
18					0.180	0.205			0.097						0.021		0.060	0.055	0.145	0.100			
18.3																			0.015	0.010			
19					0.040	0.085			0.010	0.041					0.055	0.005	0.010	0.060	0.005	0.135	0.105		
19.3																				0.010	0.010		

Table continue on next page.....

Population Data and Internal Validation of the 21 Short Tandem Repeat Loci

Allele	D3S 1358	vWa	D16 SS39	CSF IPO	TPOX	D8S 1179	D21 S11	D18 S51	D2S 441	D19 S433	TH01 FGA	D22 S1045	D5S 818	D13 S317	D7S 820	SE33 S1248	D10 S1248	D1S 1656	D12 S391	D2S 1338
20	0.015	0.010						0.019			0.139					0.075	0.010	0.110		0.160
20.2											0.001					0.005	0.010		0.005	
20.3		0.010																		
21								0.011			0.170					0.040		0.003	0.115	0.025
21.2											0.005					0.030				
22								0.007			0.175					0.010		0.006	0.085	0.060
22.2											0.013					0.015				
23		0.010									0.180					0.005			0.095	0.130
23.2																0.030				
23.3																				0.005
24		0.005						0.003			0.150					0.01			0.040	0.100
24.2											0.001					0.04				
24.3																				0.005
25								0.005			0.080					0.005			0.025	0.106
25.2								0.003			0.003					0.030				
26								0.001			0.050					0.010			0.005	0.027
26.2								0.001			0.010					0.035				
27											0.010					0.005				0.001
27.2								0.056			0.005					0.040				
28								0.172			0.110					0.015				
28.2																0.055	0.005			
29																				
29.2								0.215			0.320					0.035				
30											0.170									
30.2								0.261	0.005		0.005					0.06				
31								0.024	0.055		0.024					0.003				
31.2								0.081	0.125		0.081					0.04				
32								0.092	0.015		0.092					0.003				
32.2								0.019	0.130		0.019					0.015				
33.2								0.070	0.055		0.070					0.010				
34								0.043			0.043					0.003				
34.2																				
35								0.004			0.004					0.003				
36											0.007					0.003				
36.2											0.003					0.003				
37																0.003				

Table V. Statistical data of 21 autosomal loci profiled with GlobalFiler™ PCR Amplification Kit for Turkish population (N=350).

Loci	N _A	N _E	I	H _O	H _E	U _{H_E}	F	P-value	S.d
D3S1358	9.000	4.797	1.698	0.740	0.792	0.796	0.065	0.28373	0.00031
VWA	12.000	5.485	1.880	0.870	0.818	0.822	-0.064	0.00799	0.00007
D16S539	7.000	4.625	1.699	0.780	0.784	0.788	0.005	0.0588	0.00023
CSF1PO	7.000	3.917	1.486	0.810	0.745	0.748	-0.088	0.06166	0.00022
TPOX	6.000	3.285	1.408	0.700	0.696	0.699	-0.006	0.4369	0.00054
D8S1179	9.000	6.013	1.941	0.890	0.834	0.838	-0.068	0.27765	0.00033
D21S11	11.000	5.484	1.915	0.860	0.818	0.822	-0.052	0.50011	0.00036
D18S51	15.000	8.110	2.262	0.880	0.877	0.881	-0.004	0.03888	0.00015
D2S441	9.000	4.634	1.704	0.830	0.784	0.788	-0.058	0.23349	0.0005
D19S433	15.000	9.166	2.362	0.760	0.891	0.895	0.147	0.00018	0.00001
TH01	7.000	5.219	1.707	0.660	0.808	0.812	0.184	0.0039	0.00006
FGA	15.000	7.369	2.159	0.830	0.864	0.869	0.040	0.79067	0.00026
D22S1045	11.000	3.816	1.587	0.690	0.738	0.742	0.065	0.01762	0.0001
D5S818	8.000	3.941	1.519	0.700	0.746	0.750	0.062	0.3735	0.00045
D13S317	7.000	4.562	1.683	0.810	0.781	0.785	-0.037	0.63718	0.00044
D7S820	9.000	5.068	1.775	0.720	0.803	0.807	0.103	0.02465	0.00014
SE33	36.000	22.247	3.274	0.920	0.955	0.960	0.037	0.03518	0.0001
D10S1248	9.000	4.543	1.676	0.750	0.780	0.784	0.038	0.06409	0.00023
D1S1656	18.000	10.899	2.544	0.830	0.908	0.913	0.086	0.00073	0.00002
D12S391	18.000	10.101	2.474	0.820	0.901	0.906	0.090	0.19723	0.00024
D2S1338	16.000	8.873	2.339	0.810	0.887	0.892	0.087	0.21047	0.00022

Abbreviations: N_A, No. of Different Alleles. N_E, No. of Effective Alleles = $1 / (\sum \pi_i^2)$. I, Shannon's Information Index = $-1 * \sum (\pi_i * \ln(\pi_i))$. H_O, Observed Heterozygosity = No. of Hets / N. H_E, Expected Heterozygosity = $1 - \sum \pi_i^2$. U_{H_E}, Unbiased Expected Heterozygosity = $(2N / (2N-1)) * H_E$. F, Fixation Index = $(H_E - H_O) / H_E = 1 - (H_O / H_E)$. S.d, Standart Deviation. Where pi is the frequency of the ith allele for the population & Sum pi² is the sum of the squared population allele frequencies.

results, the development of automation and the cheapening of technology have provided a great use and success in forensic genetics. In order to provide all these vital expertise services, it is necessary to use reliable and up-to-date methods and systems in accordance with quality standards.

The allelic frequencies of short tandem repeats loci vary from population to population. For this reason, the genetic markers used in forensic sciences should be determined for each population and a database should be established. In the statistical calculations made in the evaluation of DNA analysis results, it is important to use the database of the population (Budowle *et al.*, 2001). Although short tandem repeats databases have been created in many countries around the world, different kits are produced in the criminal research area due to the increasing number of polymorphic loci.-

One of the commercially available GlobalFiler™ Polymerase Chain Reaction kit is used in our study to obtain more rapid and reliable results in forensic cases by

validation of DNA analysis test methods from oral swab. These values correspond to the validity of the polymerase chain reaction kit and our results is other studies. Mentioned in some studies, the capillary electrophoresis device, sample and polymerase chain reaction kits sensitivity and efficiency play a vital role in internal validation studies. This shows that sample collection methods change the efficiency of the study, even if same kit was used (Flores *et al.*, 2014). Flores *et al.* (2014) conducted a study with buccal swab and blood samples by using GlobalFiler Express kit and in ABI 3500xl, according to their results, thermal cycle was found to be 28 cycles, injection time was 12 seconds and minimum detection threshold was indicated as 120 RFU. In addition to that, the study found stochastic threshold as 400 RFU for buccal swab samples and emphasized that increasing the thermal cycle would result in an increase for stochastic threshold value. Moreover, the minimum amount of input DNA to have a full profile is found to be 0.5 g whereas in our study it was 0.125 ng.

Table VI. Analyze results with Arlequin v 3.5.2.2 (N=350).

Loci	PM	Expressed as 1 in	PD	PIC	PE	TPI	Homo zygote	Heterozy- gote	Total Allels
D3S1358	0.080	12.563	0.920	0.759	0.493	1.923	0.260	0.740	700
VWA	0.074	13.441	0.926	0.794	0.735	3.846	0.130	0.870	700
D16S539	0.086	11.628	0.914	0.754	0.562	2.273	0.220	0.780	700
CSF1PO	0.129	7.728	0.871	0.701	0.618	2.632	0.190	0.810	700
TPOX	0.132	7.599	0.868	0.656	0.428	1.667	0.300	0.700	700
D8S1179	0.058	17.182	0.942	0.814	0.775	4.545	0.110	0.890	700
D21S11	0.063	15.974	0.937	0.796	0.715	3.571	0.140	0.860	700
D18S51	0.040	24.876	0.960	0.864	0.755	4.167	0.120	0.880	700
D2S441	0.096	10.417	0.904	0.754	0.656	2.941	0.170	0.830	700
D19S433	0.035	28.736	0.965	0.881	0.527	2.083	0.240	0.760	700
TH01	0.069	14.577	0.931	0.780	0.369	1.471	0.340	0.660	700
FGA	0.037	27.322	0.963	0.849	0.656	2.941	0.170	0.830	700
D22S1045	0.114	8.803	0.886	0.699	0.413	1.613	0.310	0.690	700
D5S818	0.107	9.328	0.893	0.703	0.428	1.667	0.300	0.700	700
D13S317	0.087	11.521	0.913	0.750	0.618	2.632	0.190	0.810	700
D7S820	0.072	13.928	0.928	0.775	0.460	1.786	0.280	0.720	700
SE33	0.014	71.429	0.986	0.953	0.836	6.250	0.080	0.920	700
D10S1248	0.088	11.416	0.912	0.747	0.510	2.000	0.250	0.750	700
D1S1656	0.026	37.879	0.974	0.901	0.656	2.941	0.170	0.830	700
D12S391	0.025	40.650	0.975	0.893	0.637	2.778	0.180	0.820	700
D2S1338	0.029	34.014	0.971	0.877	0.618	2.632	0.190	0.810	700

Abbreviations: PM, match probability; PD, power of discrimination; PIC, polymorphism information content; PE, power of exclusion; TPI, typical paternity index.

Table VII. Genetic distances between populations. Iraq, n = 103 (22); Iran, Farsi, n = 112–126 (23); Saudi Arabia, n = 190 (24); South Africa, n = 634 (25); Afghanistan, n = 348 (26); China, Han, n = 446 (27) and Romania, Wallachia, n = 1910 (28); Azerbaijan, n=765 (29); USA, n=1194 (20).

	Turkey	Iraq	Saud Arabia	China (Han)	USA	Iran	Afghan	Azerbaijan	Romania	South Africa
Turkey	0									
Iraq	0.0272	0								
Saud Arabia	0.0602	0.0572	0							
China (Han)	0.082	0.1042	0.1394	0						
USA	0.0263	0.0491	0.0876	0.1082	0					
Iran	0.0254	0.0482	0.0772	0.0878	0.0389	0				
Afghan	0.0192	0.0374	0.0768	0.0376	0.0378	0.0323	0			
Azerbaijan	0.0097	0.0350	0.0664	0.0748	0.0266	0.0188	0.0175	0		
Romania	0.0178	0.0437	0.0162	0.0175	0.0187	0.0265	0.0288	0.0177	0	
South Africa	0.1198	0.1523	0.1494	0.1897	0.1343	0.1342	0.1371	0.1318	0.1344	0

According to another study Almeida et.al conducted in Brazilian population with 502 participants by using

Globalfiler Express Kit, the cut-off value was found as 50 RFU similar to our results (Almeida *et al.*, 2015).

In one genetic validation study conducted in America population, almost all of the minor components in the 1:9 and 9:1 mix samples were detected and it was reported that a significant proportion of the minor components could be detected in the 1:19 and 19:1 mix samples. In our study, observed alleles of minor components were found as 23% for 1:100 mixture, 45% for 1:50, and 86% for 1:10 mixture samples. However, in 1:1 ratio mixture samples, the alleles of the minor component was observed most efficiently (Ludeman *et al.*, 2018).

Along with the validation parameters that make up the first part of the study; individuals with no consanguineous relationship and taken into consideration of Turkey's population density in seven regions, randomly selected 350 individuals have been studied and allele frequencies of these individuals were calculated in the second part of the study. In our study, it was determined that SE33 locus is the most polymorphic locus Turkish population. In the literature search, many researchers stated that SE33 loci is the most discriminative loci among 21 autosomal loci. This shows that our study is compatible with other studies (Alsafiah *et al.*, 2017; Hennessy *et al.*, 2014; Park *et al.*, 2016; Zhang *et al.*, 2016). For Turkey population, Combined Power of Exclusion is % 99.9999963961 and Combined Power of Discrimination is % 99.99999999999999999999998267. The allele frequencies obtained in our study were compared with the populations of Iraq, Saud Arabia, China (Han), USA (Cauc), Iran, Afghan, Azerbaijan, Romania, South Africa (Table VII). According to the results of the analysis, a significant difference was observed between the Turkish population and the South Africa population (0,1198) and no significant difference was observed for the Azerbaijan (0,0097) population. The differences between populations, the result of the distributions of the alleles on a locus and allele frequencies on a given population should be considered regardless of the geographical distances in between.

CONCLUSIONS

This study was aimed to conduct internal validation of Globalfiler Amplification kit and provide a Turkish population data by using the kit. The results of both validation and population studies were compared with other countries. However, one of the points that should be considered in terms of population study is that the GlobalFiler™ kit is sensitive and has more definite results in terms of some loci than the fact that it has high discriminative power. In many studies some degenerated samples in other kits, Y allele was deficient. Such conditions may occur due to major deletions on the Y chromosome or primers. The absence of the allele can cause problems in

forensic cases or identification. This risk is lower in the GlobalFiler™ kit. The Y-indel, Y-short tandem repeats and DYS391 markers found in the GlobalFiler™ kit are particularly enlightening for the detection of men individuals. As a biological sample we used the oral swabs and GlobalFiler™ kit as a polymerase chain reaction kit to determine the allelic frequencies and population genetic parameters in the Turkish population and validation of the GlobalFiler™ kit was carried out in our laboratory. The GlobalFiler™ polymerase chain reaction kit which was validated in our laboratory, provides faster, more reliable and more informative results in degraded samples for identification, to solve the incidents of sexual assault involving more than one person, in the identification of the victims of the disaster, compared to other kits, because of the inclusion of high discriminative locis.

ACKNOWLEDGMENTS

This study was supported by Istanbul University-Cerrahpasa, Scientific Research Fund (Project No: 56104).

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES

- Älgenäs, C., and Tillmar, A.O., 2014. Population genetics of 29 autosomal STRs and 17 Y-chromosomal STRs in a population sample from Afghanistan. *Int. J. Legal Med.*, **128**: 279-280. <https://doi.org/10.1007/s00414-013-0902-z>
- Almeida, C., Ribeiro, T., Oliveira, A.R., Porto, M.J., Costa-Santos, J., Dias, D., and Dario, P., 2015. Population data of the GlobalFiler® Express loci in South Portuguese population. *Forensic Sci. Int. Genet.*, **19**: 39-41. <https://doi.org/10.1016/j.fsigen.2015.06.001>
- Alsafiah, H.M., Goodwin, W.H., Hadi, S., Alshaikhi, M.A., and Wepeba, P.P., 2017. Population genetic data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit. *Forensic Sci. Int. Genet.*, **9**: 39-41. <https://doi.org/10.1016/j.fsigen.2017.09.014>
- Barni, F., Berti, A., Pianese, A., Boccellino, A., Miller, M.P., Caperna, A., Lago, G., 2007. Allele frequencies of 15 autosomal STR loci in the Iraq population with comparisons to other populations from the middle-eastern region. *Forensic Sci. Int.*, **167**: 87-92. <https://doi.org/10.1016/j.forsciint.2006.03.005>
- Budowle, B., Shea, B., Niezgodna, S., and Chakraborty,

- R., 2001. CODIS STR loci data from 41 sample populations. *J. Forensic Sci.*, **46**: 453-489. <https://doi.org/10.1520/JFS14996J>
- Butler, J.M., 2012. *Advanced topics in forensic DNA typing: Methodology*, First Ed. Academic Press.
- Butler, J.M., and Redman, J.W., 1997. STRBase (SRD-130) [WWW Document]. Natl. Inst. Standards Technol., URL <https://strbase.nist.gov/NIJ/STRBase.htm> (accessed 7.5.21).
- Connie Milton, 2017. Combined DNA Index System (CODIS). In: *Cold case homicides* (ed. R.H. Walton). CRC Press, pp. 365-370-undefined. <https://doi.org/10.1201/9781315228259-18>
- Excoffier, L., and Lischer, H.E.L., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.*, **10**: 564-567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., and Tress, M.L., 2014. Multiple evidence strands suggest that there may be as few as 19000 human protein-coding genes. *Hum. Mol. Genet.*, **23**: 5866-5878. <https://doi.org/10.1093/hmg/ddu309>
- Flores, S., Sun, J., King, J., and Budowle, B., 2014. Internal validation of the GlobalFiler™ Express PCR Amplification Kit for the direct amplification of reference DNA samples on a high-throughput automated workflow. *Forensic Sci. Int. Genet.*, **10**: 33-39. <https://doi.org/10.1016/j.fsigen.2014.01.005>
- Gill, P., Fereday, L., Morling, N., and Schneider, P.M., 2006. The evolution of DNA databases recommendations for new European STR loci. *Forensic Sci. Int.*, **156**: 242-244. <https://doi.org/10.1016/j.forseint.2005.05.036>
- Hedjazi, A., Nikbakht, A., Hosseini, M., Hoseinzadeh, A., and Hosseini, S.M.V., 2013. Allele frequencies for 15 autosomal STR loci in Fars province population, southwest of Iran. *Leg. Med.*, **15**: 226-228. <https://doi.org/10.1016/j.legalmed.2013.01.005>
- Hennessy, L.K., Mehendale, N., Chear, K., Jovanovich, S., Williams, S., Park, C., and Gangano, S., 2014. Developmental validation of the GlobalFiler® express kit, a 24-marker STR assay, on the RapidHIT® System. *Forensic Sci. Int. Genet.*, **13**: 247-258. <https://doi.org/10.1016/j.fsigen.2014.08.011>
- Karantzali, E., Rosmaraki, P., Kotsakis, A., Le Roux-Le Pajolec, M.G., and Fitsialos, G., 2019. The effect of FBI CODIS Core STR loci expansion on familial DNA database searching. *Forensic Sci. Int. Genet.*, **43**: 102129. <https://doi.org/10.1016/j.fsigen.2019.07.008>
- Lucassen, A., Ehlers, K., Grobler, P.J., and Shezi, A.L., 2014. Allele frequency data of 15 autosomal STR loci in four major population groups of South Africa. *Int. J. Legal Med.*, **128**: 275-276. <https://doi.org/10.1007/s00414-013-0898-4>
- Ludeman, M.J., Zhong, C., Mulero, J.J., Lagacé, R.E., Hennessy, L.K., Short, M.L., and Wang, D.Y., 2018. Developmental validation of GlobalFiler™ PCR amplification kit: A 6-dye multiplex assay designed for amplification of casework samples. *Int. J. Legal Med.*, **132**: 1555-1573. <https://doi.org/10.1007/s00414-018-1817-5>
- Mammadov, E.R., Mustafayev, N.S., Mammadov, A.C., Hasanov, A.B., and Huseynova, I.M., 2020. Population-genetic and comparative interpopulation studies of Azerbaijan population based on the 15 autosomal STR markers. *Gene*, **753**: 144804. <https://doi.org/10.1016/j.gene.2020.144804>
- McMahon, K., Paciorkowski, A.R., Walters-Sen, L.C., Milunsky, J.M., Bassuk, A., Darbro, B., Diaz, J., Dobyns, W.B., and Gropman, A., 2017. Neurogenetics in the Genome Era. In: *swaiman's pediatric neurology*. Elsevier. <https://doi.org/10.1016/B978-0-323-37101-8.00034-5>
- Osman, A.E., and Alsafar, H., 2015. Autosomal short tandem repeat (STR) variation based on 15 loci in a population from the central region (Riyadh Province) of Saudi Arabia. *J. Forensic Res.*, **6**: 267-271. <https://doi.org/10.4172/2157-7145.1000267>
- Park, H.C., Kim, K., Nam, Y., Park, J., Lee, J., Lee, H., Kwon, H., Jin, H., Kim, Wook, K., and Won, L.S., 2016. Population genetic study for 24 STR loci and Y indel (GlobalFiler™ PCR Amplification kit and PowerPlex® Fusion system) in 1000 Korean individuals. *Leg. Med.*, **21**: 53-57. <https://doi.org/10.1016/j.legalmed.2016.06.003>
- QIAGEN Sample and Assay Technologies, 2016. QIAamp® DNA Mini and Blood Mini Handbook [WWW Document]. URL <https://www.qiagen.com/ch/resources/download.aspx?id=62a200d6-faf4-469b-b50f-2b59cf738962&lang=en> (accessed 6.20.21).
- Rudin, N., and Inman, K., 2001. An introduction to forensic DNA analysis. In: *Handbook of materials sciences* (ed. C.T. Lynch), vol. 3. CRC Press. <https://doi.org/10.4324/9780367802806>
- Stanciu, F., Stoian, I.M., and Popescu, O.R., 2009. Population data for 15 short tandem repeat loci from Wallachia region, South Romania. *Croat. med. J.*, **50**: 321-325. <https://doi.org/10.3325/cmj.2009.50.321>

- Thermo Fisher Scientific Inc, 2019. GlobalFiler™ and GlobalFiler™ IQC PCR Amplification Kits User Guide [WWW Document]. URL <http://tools.thermofisher.com/content/sfs/manuals/4477604.pdf> (accessed 6.20.21).
- Thermo Fisher Scientific, 2019. User Guide: Qubit® dsDNA HS Assay Kits [WWW Document]. URL https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit_dsDNA_HS_Assay_UG.pdf (accessed 7.6.21).
- Wu, Y.M., Zhang, X.N., Zhou, Y., Chen, Z.Y., and Wang, X.B., 2008. Genetic polymorphisms of 15 STR loci in Chinese Han population living in Xi'an city of Shaanxi Province. *Forensic Sci. Int. Genet.*, **2**: e15-e18. <https://doi.org/10.1016/j.fsigen.2007.11.003>
- Zhang, H., Yang, S., Guo, W., Ren, B., Pu, L., Ma, T., Xia, M., Jin, L., Li, L., and Li, S., 2016. Population genetic analysis of the GlobalFiler STR loci in 748 individuals from the Kazakh population of Xinjiang in northwest China. *Int. J. Legal Med.*, **130**: 1187-1189. <https://doi.org/10.1007/s00414-016-1319-2>

Online First Article