GENETIC CHARACTERISTICS OF 22 a-STR LOCI IN THE VIETNAMESE POPULATION

Nguyen Thi Lan Huong, Vuong Gia Bao*, Mai Quang Truong

Forensic Medicine Center of Ho Chi Minh city, HCMC, Vietnam

SUMMARY

This study reported Short Tandem Repeat (STR) allele data from 4346 Vietnamese individuals across 22 autosomal STR (a-STR) loci: CSF1PO, FGA, TH01, TPOX, vWA, D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, Penta D and Penta E. Population samples consisted of 90% ethnic Kinh and 10% individuals of other ethnic minority or of mixed origin (by questionnaire data). Sample types included buccal swab, blood, hair, nail, tissue, tooth and bone. DNA extraction was performed using QIAamp DNA Mini Kit, QIAamp DNA Investigator Kit and Bone DNA Extraction kit. PCR reactions were carried out with the PowerPlex® Fusion System kit. The PCR products were subjected to capillary electrophoresis on ABI 3500 Genetic Analyzer. Electrophoretic results were analyzed using GeneMapper® ID-X software v1.4. Direct counting method and Excel software were applied to determine allele frequency and forensic genetic parameters. The allele distribution frequencies of 22 a-STR loci were established. The samples were in Hardy–Weinberg equilibrium according to the distribution of genotypes by 22 a-STR loci. The combined matching probability (CMP), combined power of discrimination (CPD), combined power of exclusion (CPE) and combined paternity index (CPI) achieved values of $6.2236x10^{-27}$, 1, 0.999999998952507 and $8.4811x10^{8}$, respectively. The a-STR loci have high discriminatory power and polymorphic information content, which demonstrates that a-STR has great potential for population biodiversity research, human identification, paternity testing, and forensic applications.

Keywords: Allele frequency, forensic genetic parameter, STR, Vietnamese population.

INTRODUCTION

The most widely used genetic markers in forensic DNA profiling (an examination of a DNA sample's sequence and/or length at discrete locations) are Short Tandem Repeats (STRs) or Simple Sequence Repeats (SSRs). STRs (also known as microsatellites) consist of units 2-6 bp long, repeated about 5 to 30 times. Therefore, when multiple loci are analyzed together, the number of possible combinations between STR loci as a group is extremely large. It is the highly variable number of repetitions of units in the STR marker region between individuals that has made STR markers the standard for individual identification. The small size of STR markers makes them much more suitable for forensics because shorter loci are more likely to be found intact (Butler, 2005). Each individual has a DNA profile on the autosomes (a-STR profile) and for each marker there is one paternal allele and one maternal allele. Therefore, the a-STR profile is individual-specific.

When comparing two DNA profiles, a match is always a statistical exercise. To determine the probability that a particular genotype might occur at random in a population as well as to ensure the reliability of DNA analysis based on STR markers, population data must be compiled to make an estimate of the frequency of each possible allele and genotype (Chakraborty, 1992). The number of STR loci and sample size must be large enough to obtain the precise and reliable results. Establishing a DNA database is vital for every forensic laboratory.

This study was carried out to initially create a DNA database, including allele frequencies and forensic genetic parameters for 22 a-STR loci in the Vietnamese population. The DNA database was used to serve the assessment work at the Forensic Medicine Center of Ho Chi Minh City in determining consanguinity relationships and solving complicated cases.

MATERIALS AND METHODS

Materials

Biological samples (buccal cell, blood, hair root, nail, tissue, bone, tooth) were collected for paternity testing and casework at the Forensic Medicine Center of Ho Chi Minh City (Vietnam) from 2016 to 2023. For this study, cases with excluded paternity were selected, including 4346 unrelated individuals (2451 men and 1895 women) of any age, living in Vietnam. Samples were randomly collected from across Vietnam, and all personal information was kept confidential.

Methods

Extraction of genomic DNA: Genomic DNA was extracted from buccal cells, blood and tissues with the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), nails and hairs with QIAamp DNA Investigator Kit (Qiagen, Hilden, Germany), bones and teeth with Bone DNA Extraction Kit (Promega, USA) according to the manufacturer's instructions. The quantities of recovered DNA were determined using Invitrogen™ Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, USA).

Genotyping of autosomal STRs: The 22 a-STR loci, one Y-STR locus and the sex determination locus (Amelogenin gene) were simultaneously amplified using the PowerPlex® Fusion System kit (Promega, USA) on a Veriti™ 96-Well thermal cycler (Applied Biosystems, USA). The PCR products were subjected to capillary electrophoresis on ABI 3500 Genetic Analyzer (Applied Biosystems, USA), following manufacturer's recommendations. Data were collected with Data Collection v1.0 software. Electrophoretic results were analyzed using GeneMapper® ID-X software v1.4 (Applied Biosystems, USA).

Statistical analysis: Allele frequency distribution for each locus and forensic genetic parameters (Matching Probability (MP), Power of Discrimination (PD), Polymorphic Information Content (PIC), Power of Exclusion (PE), Typical Paternity Index (TPI), observed heterozygosity (Ho), and expected heterozygosity (He)) were calculated according to some studies of Butler, 2005; Sensabaugh, 1982; Nei, 1987; Brenner, Morris, 1990.

RESULTS AND DISCUSSION

Allele frequencies and forensic genetic parameters of 22 a-STR loci

In the study, 320 alleles were detected with a mean of 14.55 alleles per locus, ranging from 8 alleles at TPOX to 27 alleles at Penta E. In 4346 individuals, the number of heterozygous individuals in the loci was 3 - 7 times greater than the number of homozygous individuals. The high number of heterozygous individuals showed that there was diversity about alleles and therefore there will be less chance for a random combination (Edwards *et al.*, 1992). The database of allele frequencies of 22 a-STR loci in 4346 individuals was shown in Table 1. The results showed that the allele with the highest frequency was allele 8 of the TPOX marker with a frequency of 0.5802 (Figure 1). Alleles with the lowest frequency were alleles that appeared only once with a frequency of 0.0001. These may be rare alleles and require further investigation with a larger sample size for accurate confirmation.

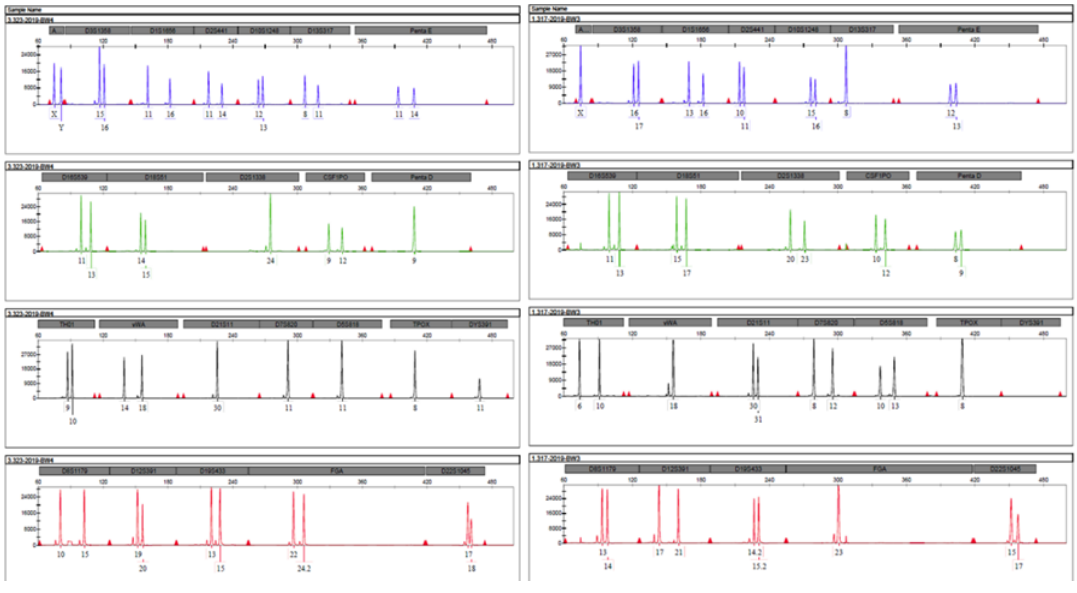

Figure 1. Illustration of DNA profile on autosomes
Left: Male DNA profile (XY), right: Female DNA profile (XX).

Regional variation is one of the factors that contributes to improving the reliability of the matching probability of two DNA profiles. Allele frequencies vary across different regions and areas. For example, surveying the alleles of the TPOX locus, although allele 8 had the highest frequency, following by allele 11, values were achieved in turn of 0.4520 and 0.3571 in Japan (n=1501) (Fujii *et al.*, 2014), 0.5269 and 0.3353 in the Philippines (n=167) (Rodriguez *et al.*, 2015), 0.5746 and 0.2516 in Belarus (n=12225) (Tsybovskiia *et al.*, 2017), and 0.5802 and 0.2640 in the present study (n=4346). Or for allele 22 on the FGA locus, the frequency in Japan was 0.1965 (highest), in the Philippines it was 0.1617 (second highest after allele 23), in Belarus it was 0.2112 (highest) and in this study it was 0.1970 (highest).

Table 1. Allele frequencies and forensic genetic parameters of 22 a-STR loci in 4346 Vietnamese individuals

Allele	D3 S1358	D1 S1656	D2 S441	D10 S1248	D13 S317	Penta E	D16 S539	D18 S51	D2 S1338	CSF1P O	Penta D	TH01	vWA	D21 S11	D7 S820	D5 S818	TPOX	D8 S1179	D12 S391	D19 S433	FGA	D22 S1045
4																				0.0005		
5						0.0413					0.0013	0.0003										
6					0.0001						0.0007	0.1291			0.0001		0.0001					
7					0.0012	0.0028	0.0001	0.0001		0.0068	0.0364	0.3221			0.0116	0.0321	0.0007					
7.2												0.0001										
7.4											0.0005											
8		0.0005	0.0001	0.0003	0.3479	0.0025	0.0024			0.0006	0.0645	0.0602			0.1629	0.0013	0.5802	0.0020				
8.1			0.0001			0.0002									0.0001							
9		0.0001	0.0023		0.1330	0.0114	0.2247	0.0005		0.0354	0.3445	0.3747			0.0637	0.0613	0.0995	0.0007		0.0072		
9.1			0.0145												0.0083							
9.2															0.0002							
9.3												0.0460										
10		0.0018	0.2178	0.0005	0.1245	0.0498	0.1291	0.0012		0.1949	0.1463	0.0656			0.1660	0.2438	0.0358	0.1550		0.0002		
10.1			0.0013												0.0008							
11	0.0002	0.0754	0.2951	0.0025	0.2138	0.2551	0.2816	0.0078		0.2793	0.1275	0.0018			0.3602	0.2693	0.2640	0.1338		0.0036		0.1667
11.1										0.0002					0.0010							
11.2											0.0032				0.0001					0.0005		
11.3			0.1271				0.0002															
12	0.0007	0.0436	0.1691	0.0609	0.1389	0.1205	0.2447	0.0618		0.3894	0.1343		0.0005		0.1910	0.2224	0.0186	0.1268		0.0467		0.0013
12.1																0.0006						
12.2																				0.0037		
12.3			0.0020																			
13	0.0024	0.0972	0.0163	0.3448	0.0321	0.0544	0.0957	0.1414	0.0001	0.0775	0.0958		0.0003		0.0304	0.1539	0.0010	0.1590		0.2531	0.0003	0.0030
13.1								0.0001														
13.2								0.0001												0.0425		
13.3			0.0001																			
14	0.0337	0.0873	0.1439	0.2382	0.0083	0.0893	0.0208	0.2006		0.0132	0.0332		0.2676		0.0033	0.0140		0.1612		0.2593		0.0361
14.1			0.0001																			
14.2																				0.1087		
14.3		0.0001																				
15	0.3102	0.2994	0.0097	0.2231	0.0003	0.0701	0.0006	0.2228		0.0023	0.0074		0.0293		0.0001	0.0013		0.1664	0.0116	0.0699		0.3104
15.2																				0.1550		
15.3		0.0029																				
15.4						0.0001																
16	0.3575	0.1948	0.0005	0.1034		0.0662		0.1585	0.0200	0.0002	0.0037		0.1568					0.0737	0.0048	0.0098	0.0021	0.2208
16.2																				0.0350		
16.3		0.0082																				
17	0.2269	0.0766		0.0240		0.0592		0.0629	0.1096		0.0002		0.2310					0.0171	0.0687	0.0006	0.0007	0.2321
17.2																				0.0033		
17.3		0.0723																	0.0002			
17.4						0.0005																
18	0.0617	0.0161		0.0023		0.0542		0.0385	0.0718		0.0001		0.2102					0.0039	0.1977		0.0115	0.0277
18.2																				0.0005	0.0001	
18.3		0.0185																	0.0005			
18.4						0.0015																
19	0.0064	0.0029				0.0411		0.0368	0.2215		0.0005		0.0866					0.0005	0.1944		0.0801	0.0013
19.1								0.0001														
19.3		0.0023																	0.0001			
20	0.0003	0.0001				0.0301		0.0258	0.1079				0.0162						0.1870		0.0569	0.0007
20.2																					0.0008	
21						0.0183		0.0243	0.0400				0.0013						0.1261		0.1616	
21.2																					0.0148	

Allele	D3 S1358	D1 S1656	D2 S441	D10 S1248	D13 S317	Penta E	D16 S539	D18 S51	D2 S1338	CSF1PO	Penta D	TH01	vWA	D21 S11	D7 S820	D5 S818	TPOX	D8 S1179	D12 S391	D19 S433	FGA	D22 S1045
21.3																						
21.4						0.0001																
22						0.0156		0.0090	0.0438				0.0001						0.0985		0.1970	
22.2						0.0001															0.0118	
23						0.0094		0.0040	0.1711										0.0746		0.1416	
23.2																					0.0133	
24						0.0035		0.0024	0.1498										0.0239		0.1427	
24.2																					0.0136	
25						0.0023		0.0009	0.0579										0.0107		0.0800	
25.2																					0.0078	
25.3																					0.0006	
26						0.0005			0.0058					0.0021					0.0013		0.0435	
26.1									0.0001													
26.2																					0.0024	
27									0.0006					0.0023							0.0151	
27.2																					0.0007	
28								0.0002						0.0535							0.0008	
28.2														0.0015								
29														0.2587								
29.2														0.0005								
29.3														0.0006								
30														0.2598								
30.2														0.0225								
30.3														0.0025								
31														0.0694								
31.2														0.0795								
32														0.0262								
32.1																					0.0001	
32.2														0.1622								
33														0.0035								
33.2														0.0469								
34														0.0013								
34.2														0.0056								
35														0.0001								
35.2														0.0013								
Forensic genetic parameters																						
MP	0.1284	0.0438	0.0693	0.0964	0.0796	0.0214	0.0800	0.0385	0.0335	0.1214	0.0576	0.1176	0.0712	0.0538	0.0822	0.0767	0.2309	0.0378	0.0400	0.0546	0.0277	0.0899
PD	0.8716	0.9562	0.9307	0.9036	0.9204	0.9786	0.9200	0.9615	0.9665	0.8786	0.9424	0.8824	0.9288	0.9462	0.9178	0.9233	0.7691	0.9622	0.9600	0.9454	0.9723	0.9101
PIC	0.6692	0.8188	0.7705	0.7218	0.7488	0.8778	0.7502	0.8364	0.8485	0.6797	0.7889	0.6865	0.7673	0.7998	0.7433	0.7576	0.5258	0.8402	0.8340	0.8005	0.8632	0.7342
PE	0.4499	0.6682	0.6090	0.5400	0.5580	0.7513	0.5514	0.7094	0.7117	0.4624	0.6263	0.4843	0.5893	0.6364	0.5555	0.5704	0.2761	0.7167	0.6935	0.6518	0.7351	0.5493
TPI	1.7468	3.0563	2.5686	2.1494	2.2472	4.1078	2.2106	3.5048	3.5333	1.7959	2.6960	1.8863	2.4361	2.7752	2.2333	2.3191	1.2119	3.5977	3.3176	2.9051	3.8528	2.1994
He	0.7196	0.8360	0.7994	0.7596	0.7797	0.8862	0.7840	0.8529	0.8631	0.7248	0.8098	0.7291	0.7976	0.8218	0.7745	0.7899	0.5821	0.8573	0.8515	0.8226	0.8756	0.7712
Ho	0.7138	0.8364	0.8053	0.7674	0.7775	0.8783	0.7738	0.8573	0.8585	0.7216	0.8145	0.7349	0.7948	0.8198	0.7761	0.7844	0.5874	0.8610	0.8493	0.8279	0.8702	0.7727
No. alleles																						
	10	19	16	10	10	27	10	22	14	11	17	9	11	20	16	10	8	12	15	18	25	10

The matching probability (MP) was highest at the TPOX locus (0.2309), lowest at the Penta E locus (0.0214). The power of Discrimination (PD), Polymorphic Information Content (PIC), Power of Exclusion (PE) and Typical Paternity Index (TPI) were highest at the Penta E locus with values of 0.9786, 0.8778, 0.7513, and 4.1078, respectively. The lowest values for PD, PIC, PE, and TPI were observed at the TPOX locus, with values of 0.7691, 0.5258, 0.2761, and 1.2119, respectively. The lowest and the highest observed heterozygosity (Ho) were obtained in the TPOX and Penta E loci (0.5874 and 0.8783, respectively). The expected heterozygosity (He) values ranged from 0.5821 (TPOX) to 0.8862 (Penta E). The heterozygosity of all loci was greater than 0.7, except for TPOX. These results indicated that these loci have high discriminatory power in the Vietnamese population.

Combined forensic genetic parameters

Combined matching probability (CMP) = $6.2236x10^{-27}$. This means that the chance of any two people having exactly the same DNA profile is 1 in $1.6068x10^{26}$. That is the DNA profile of an individual was unique in the population. Thus, when these 22 STR loci are combined, it is reliable and accurate for individual identification, paternity testing, and research on population genetics.

A Combined Power of Discrimination (CPD) value of 1 indicates that the DNA profiling method can distinguish each individual in the population from all others. Understanding in another way, instead of looking at the probability of matching, the probability of "not matching" was interested in.

Combined Power of Exclusion (CPE) = 0.999999998952507. This index indicated that a person's DNA profile differs from the DNA profile of a randomly selected person in the population by 99.9999998952507%.

Combined Paternity Index (CPI) = $8.4811x10^{8}$. This index is also often used for paternity testing. That means the number of similarities between a child and the biological father is $8.4811x10^{8}$ times.

Almost all countries calculate their allele frequencies and standardize them to improve accuracy in conclusions. In Table 2, the combined matching probability (CMP) decreases when more markers are examined, from $7.29x10^{-16}$ (15 loci) to $1.948x10^{-36}$ (31 loci). This means the possibility of having any two people with identical DNA profiles is increasingly unlikely as the number of markers examined increases. Similarly, the combined indices CPD, CPE and CPI all increase when increasing the number of markers and sample size. This means that the more markers are examined, the greater the chance of accurately identifying each individual and the lower the probability of errors in concluding consanguinity relationships. CPD and CPE values in this study indicate that the discriminatory power of the examined loci is considerably high. The study collected a large number of samples (n=4346) and analyzed with many loci simultaneously (22 loci), which improved the efficiency, reliability, and accuracy of the data.

One of the most important applications of the DNA database created from the study is to determine the accuracy (probability of paternity - POP) in lineal paternity testing. The more loci that are examined, the closer the POP value is to 100% if a match is obtained. For example, when comparing two DNA profiles of a father and a child, if 15 loci are examined, the POP value is 99.999992220147%. If 22 loci are examined, the POP value will be 99.9999999994358%. However, if a new DNA profile does not have the observed allele in the database, the POP calculation cannot be performed. Therefore, the DNA database must be updated annually.

With the advancement of science and technology, a number of new technologies and genetic markers, such as next-generation sequencing and single nucleotide polymorphisms, have been widely used. However, due to the lack of databases for new genetic markers, STR profiling using older methods is still employed in forensic practice. In 1997, the Federal Bureau of Investigation (FBI) laboratory selected 13 STRs as its core loci (CODIS), which were expanded to 20 STRs in 2017. Commercial STR kits are based on these core loci. In recent years, these core loci have been supplemented by an increasing number of new non-CODIS loci to obtain additional genetic information and further improve discrimination. However, the newly approved STRs should be used with caution.

Table 2. Comparison of forensic genetic parameters with other studies (surveyed loci covered 13 CODIS)

Parameter	CMP	CPD	CPE	CPI
This study	$6.2236x10^{-27}$ (n=4346, 22 loci)	1 (n=4346, 22 loci)	0.999999998952507 (n=4346, 22 loci)	$8.4811x10^{8}$ (n=4346, 22 loci)
Shimada *et al.*, 2002		0.9999999999999999998 (n=178, 16 loci)	0.99999994 (n=178, 16 loci)	
Trần Khánh Linh *et al.*, 2010		0.999999999999491 (n=182, 12 loci)	0.9999923 (n=182, 12 loci)	
Liu *et al.*, 2013	$7.29x10^{-16}$ (n=150, 15 loci)	0.99999999999999914651 (n=150, 15 loci)	0.9999916368 (n=150, 15 loci)	
Anghel *et al.*, 2014		0.99999999999999999722 (n=336, 15 loci)	0.999999295 (n=336, 15 loci)	
Rodriguez *et al.*, 2015	$8.21x10^{-28}$ (n=176, 23 loci)		0.999996 (n=176, 23 loci)	
Park *et al.*, 2016	3.601×10^{-26} (n=1000, 22 loci) 2.902×10^{-28} (n=1000, 24 loci)		0.9999999971 (n=1000, 24 loci)	
Tsybovskiia *et al.*, 2017	7.089342×10^{-22} (n=12346, 18 loci)		0.999961 (n=12346, 18 loci)	

Barrio *et al.*, 2019	$2.022x10^{-24}$ (n=496, 20 loci) $1.948x10^{-36}$ (n=496, 31 loci)			$1.473x10^{8}$ (n=496, 20 loci) $1.118x10^{12}$ (n=496, 31 loci)
Vu *et al.*, 2021	$1.8648x10^{-27}$ (n=1000, 22 loci)	1 (n=1000, 22 loci)	0.9999999992 (n=1000, 22 loci)	

CONCLUSION

The study has established the allele frequency database and associated statistical parameters for 22 a-STR markers from a large sample (4346 unrelated individuals distributed among provinces scattered across Vietnam). The a-STR loci have high discriminatory power and polymorphic information content, which demonstrates that a-STR has great potential for population biodiversity research, human identification, paternity testing, and forensic applications.

REFERENCE

Anghel A, Enache A, Seclaman E, Gruin G, Ursoniu S, Alexa A, Antonescu M, Marian C (2014). Genetic polymorphism data on 15 autosomal STR markers in a Western Romanian population sample. *Leg Med (Tokyo)* 16(4): 238-240.

Barrio PA, Martín P, Alonso A, Müller P, Bodner M, Berger B, Parson W, Budowle B & DNASEQEX Consortium (2019). Massively parallel sequence data of 31 autosomal STR loci from 496 Spanish individuals revealed concordance with CE-STR technology and enhanced discrimination power. *Forensic Sci Int Genet* 42: 49-55.

Brenner CH, Morris JW (1990). Paternity Index Calculations in Single Locus Hypervariable DNA Probes: Validation and Other Studies Proceedings for The International Symposium on Human Identification 1989, *Promega Corporation*: 21-53.

Butler JM (2005). Forensic DNA Typing - Biology Technology DNA genetics of STR Markers, 2nd ed., *Acadamic Press*.

Chakraborty R (1992). Sample size requirements for addressing the population genetic issues of forensic use of DNA typing, *Hum Biol* 642: 141-159.

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12(2): 241-53.

Fujii K, Iwashima Y, Kitayama T, Nakahara H, Mizuno N, Sekiguchi K (2014). Allele frequencies for 22 autosomal short tandem repeat loci obtained by PowerPlex Fusion in a sample of 1501 individuals from the Japanese population. *Leg Med* 16(4): 234-237.

Liu J, Li SY, Yin JY, Zhang W, Gao B, Guo L, Qi R (2013). Genetic analysis of 15 STR loci from a Xibe population in Liaoning, Northeast of China. *Ann Hum Biol* 41(3): 244-248.

Nei M (1987). Molecular Evolutionary Genetics. *New York: Columbia University Press*.

Park HC, Kim K, Nam Y, Park J, Lee J, Lee H, Kwon H, Jin H, Kim W, Kim W, Lim 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 (Tokyo)* 21: 53-57.

Rodriguez JJRB, Salvador JM, Calacal GC, Laude RP, Ungria MCA (2015). Allele frequencies of 23 autosomal short tandem repeat loci in the Philippine population. *Leg Med* 17(4): 295-297.

Sensabaugh G (1982). Biochemical markers of individuality. Saferstein R eds. Forensic Science Handbook. *New York: Prentice-Hall, Inc*: 338-415.

Shimada I, Brinkmann B, Nguyen QT, Hohoff C (2002). Allele frequency date for 16 STR loci in the Vietnamese population. *Int J Legal Med* 116(4): 246-248.

Tsybovskiia IS, Veremeichika VM, Kotovaa SA, Kritskayaa SV, Evmenenkob SA, Udinac IG (2017). Developing Forensic Reference Database by 18 Autosomal STR for DNA Identification in Republic of Belarus. *Russ J Genet* 53(2): 275-284.

Thi Thanh Hoan Vu, Thi Thu Mai Do, Thi Hoai Nguyen, Quoc Hai Luyen (2021). Allele frequencies of 23 short tandem repeat loci in the Vietnamese Kinh population. *Forensic Sci Int Rep*. https://doi.org/10.1016/j.fsir.2021.100210.

Trần Khánh Linh, Lê Thúy Quyên, Võ Đức Xuyên An và Phạm Hùng Vân (2010). Bước đầu khảo sát tần suất phân bố kiểu hình của các STR thường được sử dụng trong xét nghiệm dấu vân tay DNA ở người Việt Nam. *Tạp chí Y học Thành phố Hồ Chí Minh*, Tập 13, Phụ bản số 2: 116-120.

ĐẶC ĐIỂM DI TRUYỀN CỦA 22 CHỈ THỊ PHÂN TỬ a-STR TỪ CÁC CÁ THỂ NGƯỜI VIỆT NAM

Nguyễn Thị Lan Hương, Vương Gia Bảo*, Mai Quang Trường

Trung tâm Pháp y Thành phố Hồ Chí Minh, Việt Nam

TÓM TẮT

Nghiên cứu này báo cáo dữ liệu alen STR từ 4346 người Việt Nam được khảo sát trên 22 locus STR nhiễm sắc thể thường (a-STR): CSF1PO, FGA, TH01, TPOX, vWA, D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, Penta D và Penta E. Dân số bao gồm 90% người Kinh và 10% là người dân tộc thiểu số khác hoặc có nguồn gốc hỗn hợp (theo dữ liệu bảng câu hỏi). Các loại mẫu bao gồm niêm mạc miệng, máu, tóc, móng tay, mô, răng, xương. Quá trình tách chiết DNA được thực hiện bằng cách sử dụng các bộ kit QIAamp DNA Mini Kit, QIAamp DNA Investigator Kit và Bone DNA Extraction Kit. Phản ứng PCR được thực hiện với bộ kit PowerPlex® Fusion System. Sản phẩm PCR được điện di mao quản trên máy phân tích di truyền ABI 3500. Kết quả điện di được phân tích bằng phần mềm GeneMapper® ID-X v1.4. Phương pháp đếm trực tiếp và phần mềm excel được áp dụng để xác định tần suất alen và các thông số di truyền pháp y. Bảng tần suất alen của 22 locus a-STR đã được xác lập. Các mẫu ở trạng thái cân bằng Hardy-Weinberg theo sự phân bố kiểu gen của 22 locus STR. Xác suất trùng khớp kết hợp (CMP), khả năng phân biệt kết hợp (CPD), khả năng loại trừ kết hợp (CPE) và chỉ số quan hệ huyết thống kết hợp (CPI) đạt các giá trị lần lượt là 6.2236×10^{-27}, 1, 0.999999998952507 và 8.4811×10^{8}. Các locus a-STR có khả năng phân biệt và thông tin đa hình cao, chứng tỏ rằng a-STR có tiềm năng lớn cho nghiên cứu đa dạng sinh học quần thể, nhận dạng con người, xét nghiệm quan hệ huyết thống và các ứng dụng pháp y.

Từ khóa: Tần suất alen, thông số di truyền pháp y, STR, người Việt Nam.

* Author for correspondence: Tel: 0765230982; Email: giabaovuong10@gmail.com