Population Data of D6S1043, Penta D and Penta E Loci in Calabria (South of Italy)

A. Barbaro*, G. Falcone and A. La Marca

Department of Forensic Genetics, Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy

Abstract: In the present study 200 unrelated healthy natives of Calabria were genetically characterized for D6S1043, Penta D and Penta E STR markers. Allele frequencies distribution and some statistical parameters of forensic interest have been calculated and compared with other published populations data. Results showed Penta E was the most informative locus based on heterozygosity, PIC, PD and TPI, while D6S1043 was the less informative. Our findings confirmed these 3 loci are useful for forensic purposes and paternity tests especially if analyzed in association with other STRs.

Keywords: STRs, population data, Calabria, South Italy.

1. INTRODUCTION

In the last years, new commercial STR multiplexes have been released that include the worldwide routinely used Penta D and Penta E markers [1] together with the relatively new locus D6S1043, commonly analyzed in China [2].

Each autosomal STR marker has unique characteristics in terms of chromosomal position, number of alleles, type and number of repeated sequence, the presence of microvariants.

Penta E is a simple pentanucleotide tandem repeat marker located on the long arm of chromosome 5 (15q26.2), containing an AAAGA repeat motif.

Penta D is a simple pentanucleotide tandem repeat marker located on the long arm of chromosome 21 (21q22.3) containing a AAAGA repeat motif.

D6S1043 is a compound tetranucleotide tandem repeat marker located on the long arm of chromosome 6 (6q15) containing a AGAT or AGAC repeat motif.

The aim of this study is to evaluate alleles distribution for Penta D, Penta E and D6S1043 loci in South of Italy (Calabria), to calculate some forensic statistical parameters and to compare data with those of other populations.

Calabria is a long and narrow peninsula in the South of Italy and because of this it's generally known as the "boot" of Italy. The region has a population of around 2 millions persons and it covers 15,080 km^2. During centuries it was greatly influenced by the Greek and Roman civilizations and less by Arabic and Albanian cultures.

In modern age it was part of the "Kingdom of the Two Sicilies" under Spanish and French domination. Because of its history Calabria still preserves a mixture of different cultures.

Penta D and Penta E are quite diffused and investigated even if no data at date are available for Calabrian population. D6S1043 is less diffused and it has been studied almost exclusively in Asian populations. Because of this, present study is relevant to increase knowledge about these 3 markers and to expand Italian population data.

2. MATERIALS AND METHODS

Blood or saliva samples were collected from 200 unrelated donors belonging to Calabria (South of Italy) since 3 generations. Informed consent was achieved from each donor, according to SIMEF laboratory internal procedures.

DNA was extracted by the InstaGene Matrix (Biorad) [5] and then it was quantified with the Quantifiler™ Human DNA Quantification Kit using a 7300 Real Time PCR System (Applied Biosystems). PCR amplification was performed by the SureID® 21G Human STR (Health Gene Technologies) that amplifies in a single multiplex (using the 5-dye technology) 21 autosomal STRs loci (D3S1358, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, TPOX D8S1179, FGA, D19S433, D12S391, D2S1338, D6S1043, D1S1656, PentaD, Penta E) and the gender marker Amelogenin [3].

PCR products were analyzed by capillary electrophoresis onto an ABI 3130 capillary sequencer.

*Address correspondence to this author at the Department of Forensic Genetics, Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy; Tel: +390965891184; E-mail: simef_dna@tiscali.it

ISSN: 1814-8085 / E-ISSN: 1927-5129/20

Allele assignments was made by comparison with the SureID® 21G Ladder using the internal standard Size-500Plus. Raw data were elaborated by GeneMapper® v.3.2 software with the following standard conditions: analytical threshold 150RFU, stutter peaks ≤ 15%, Heterozygous Peak Height Ratios ≥ 70%.

In order to reduce contamination, each step has been performed in different dedicated areas of the DNA laboratory. Positive and negative controls were used during all steps.

3. STATISTICAL ANALYSIS

Forensic statistical parameters were calculated by STRAF software [4]. Hardy-Weinberg equilibrium and other population data were estimated using Arlequin software v.3.5.2.2 [5].

Allelic frequencies for STRs were compared to previously published population data [6-9].

4. RESULTS

Allele frequencies of analyzed loci are given in Table **1**, while the resulting statistical parameters are reported in Table **2**.

All loci showed a heterozygosity value greater than 0.7 with the highest value in Penta E, that shows a PIC value of 0.8775.

Penta E had the highest discrimination power (PD) equal to 0.9634 and a typical paternity index (TPI) of 50. D6S1043 showed the lowest exclusion power (PE) equal to 0.4738 and consequently the highest matching probability (PM) equal to 0.0612.

Table 1: Allele Frequencies of D6S1043, Penta D and Penta E Loci in Calabria Population

	D6S1043	Penta D	Penta E
Allele			
5			0.1
7			0.10625
8		0.01875	0.01875
9		0.13125	0.03125
10	0.02143	0.08750	0.1
11	0.21429	0.19375	0.18125
12	0.31429	0.15625	0.16875
13	0.02857	0.21875	0.1125
14	0.05000	0.10000	0.05
15	0.00714	0.05625	0.0375
16	0.00714	0.02500	0.0125
17	0.03571	0.01250	0.025
18	0.12143		0.025
19	0.13571		0.0125
20	0.06429		0.00625
21			0.0125

Table 2: Statistical Parameters for D6S1043, Penta D and Penta E Loci

Locus	Nall	GD	PIC	Hobs	He	PM	PD	PE	TPI	P-value
D6S1043	11	0.8187	0.7903	0.7286	0.81871	0.0612	0.9388	0.4738	18.421	0.21298
PENTAD	10	0.8564	0.8336	0.8125	0.85637	0.0478	0.9522	0.6224	26.667	0.56247
PENTAE	16	0.8933	0.8775	0.9000	0.89332	0.0366	0.9634	0.7954	50.000	0.17005

Nall: n° alleles GD: Genetic Distance; PIC: polymorphism information content; Hobs: observed heterozygosity; He: expected heterozygosity; PM: matching probability; PD: power of discrimination; PE: power of exclusion; TPI: typical paternity index, P-value : exact test.

No significant deviations from Hardy–Weinberg expectations were found (p>0.05).

Data from populations comparisons are available in Tables **3A**, **3B**, **3C**; calculated Fst values are reported in Tables **4A**, **4B**, **4C** and corresponding p-values are summarized in Tables **5A**, **5B**, **5C**.

Table 3A: Locus D6S1043 - Comparison with other Populations

	CALABRIA	WEST CHINA	IRAQ
Allele			
7	0	0	0.0009
9	0	0	0.0028
10	0.0214	0.04	0.0211
11	0.2143	0.0975	0.2895
12	0.3143	0.155	0.2670
13	0.0286	0.1388	0.0731
14	0.0500	0.1313	0.0469
15	0.0071	0.0225	0.0064
16	0.0071	0.005	0.0032
17	0.0357	0.0338	0.0225
18	0.1214	0.18	0.0997
19	0.1357	0.1463	0.1149
20	0.0643	0.0425	0.0446
21	0	0.005	0.0064
23	0	0.0013	0.0009
24	0	0.0013	0

Table 3B: Locus Penta D - Comparison with other Populations

	CALABRIA	SICILY	ITALY
Allele			
2.2	0	0	0
3	0	0.003	0.0022
5	0	0	0
7	0	0.007	0.0022
8	0.0187	0.014	0.0170
9	0.1312	0.223	0.1814
10	0.0875	0.164	0.1462
11	0.1937	0.179	0.1575
12	0.1562	0.089	0.1451
13	0.2187	0.190	0.2335
14	0.1000	0.089	0.0714
15	0.0562	0.029	0.0317
16	0.0250	0.003	0.0113
17	0.0125	0.003	0

Table 3C: Locus Penta E - Comparison with other Populations

	CALABRIA	SICILY	ITALY
Allele			
4	0	0	0.0022
5	0.1	0.0078	0.0714
6	0	0	0.0057
7	0.1062	0.175	0.1553
8	0.0187	0.041	0.0215
9	0.0312	0.014	0.0147
10	0.1	0.067	0.076
11	0.1812	0.07	0.1383
12	0.1687	0.175	0.1723
13	0.1125	0.111	0.127
14	0.05	0.041	0.0442
15	0.0375	0.067	0.041
16	0.0125	0.041	0.0238
17	0.025	0.052	0.0567
18	0.025	0.022	0.0249
19	0.0125	0.026	0.0102
20	0.0062	0.003	0.0091
21	0.0125	0.011	0.0045
22	0	0	0.0011

Table 4A: Locus D6S1043 - FST Values

Calabria	0.00000		
W. China	-0.05325	0.00000	
Iraq	-0.08863	-0.03158	0.00000

Table 4B: Locus PENTA D - FST Values

Calabria	0.00000		
Sicily	-0.06721	0.00000	
Italy	-0.08526	-0.08074	0.00000

Table 4C: Locus PENTA E - FST Values

Calabria	0.00000		
Sicily	-0.03546	0.00000	
Italy	-0.05323	-0.05211	0.00000

Table 5A: Locus D61043 - Corresponding P-Values

Calabria	*		
W. China	0.81081+-0.0359	*	
Iraq	0.99099+-0.0030	0.71171+-0.0165	*

Table 5B: Locus PENTA D - Corresponding P-Values

Calabria	*		
Sicily	0.94595+-0.0154	*	
Italy	0.99099+-0.0030	0.98198+-0.0096	*

Table 5C: Locus PENTA E - Corresponding P-Values

Calabria	*		
Sicily	0.91892+-0.0184	*	
Italy	0.98198+-0.0096	0.97297+-0.0184	*

5. DISCUSSION

In the present study we calculated the allele distribution and some statistical parameters for the D6S1043, Penta D and Penta E loci in Calabria (South Italy) using 200 samples from unrelated donors. It's known that a study performed on 100-150 biological samples is generally adequate for calculating STRs population data and related statistical parameters.[10]

Analyzed STR loci were highly polymorphic: all loci showed a heterozygosity value greater than 0.7 Penta E was the most informative locus based on heterozygosity (Het >0.7), Polymorphic Information Content (PIC=0.8775), Discrimination Power (PD = 0.9634) and typical paternity index (TPI=50). On the contrary D6S1043 was the less informative with a PD value = 0.9388.

D6S1043 showed the lowest exclusion power (PE) equal to 0.4738 and consequently the highest matching probability (PM) equal to 0.0612.

Anyway combined RMP using 3 loci was $1.07x10^{-4}$ and therefore combined power of discrimination (PD) was around 0.9999. This means when these 3 loci are used in combination with other ones (such as SE33 or with ESS loci) they can distinguish different individuals with a probability > 99,9999%.

In all loci no significant deviations from Hardy–Weinberg expectations were found (p>0.05).

Comparison for Penta D and Penta E loci was performed with previously published data of Sicily [6] and Italy [7]. Comparison with Sicilian population, showed that Penta E allele frequencies were very similar.

Locus D6S1043 data have been compared with the ones of China [8] and Iraq [9] because no other informations for this locus in closer populations are available as publications.

As expected, comparison showed significant differences.

6. CONCLUSIONS

Data obtained in the present study demonstrated that D6S1043, Penta D and Penta E loci are useful for forensic purposes especially if they are analyzed in combination between them or in association with further loci available in the new commercial STR multiplexes.

Penta D and Penta E are commonly available in commercial multiplexes and they both are quite diffusely investigated. On the contrary, D6S1043 is less diffused and it has been used almost exclusively in Asian population studies.

Because of this allele frequencies and forensic statistical parameters calculated in the present study are useful to increase knowledge about D6S1043 population data.

Furthermore, since no other previous data are available for D6S1043, Penta D and Penta E loci in Calabria, they contribute to expand Italian population data of STR markers used in forensics or paternity tests.

DECLARATIONS OF INTEREST

None.

REFERENCES

[1] Bacher J, Schumm JW. Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter. Profiles DNA 1998; 2: 3-6.

[2] Bright JA, Stevenson KE, Coble MD, Hill CR, Curran JM, Buckleton JS. Characterizing the STR locus D6S1043 and examination of its effect on stutter rates. Forensic Sci Int Genet 2014; 7-8(1): 20-3. https://doi.org/10.1016/j.fsigen.2013.06.012

[3] Arapcic M, Herić A, Dogan S, Čakar J, Pilav A, Džehverović M, Marjanović D. Validation of the SureID® 21G Human STR Identification Kit and concordance study of the new

generation multiplex STR kits. Proceedings of the 10th ISABS Conference on Forensic and Anthropologic Genetics and Mayo Clinic Lectures in Individualized Medicine, Dubrovnik 2017; 19-24.

[4] Gouy A, Zieger M. STRAF - A convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet 2011; 30: 148-151. https://doi.org/10.1016/j.fsigen.2017.07.007

[5] Excoffier L, Lischer HL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 2010; 10: 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

[6] Asmundo A, Sapienza D, Crinò C. Analysis of penta D (15q) and penta E (21q) STR polymorphism in a southern Italian population sample. J Forensic Sci 2004; 49(5): 1128-9. https://doi.org/10.1520/JFS2004159

[7] Brisighelli F, Capelli C, Boschi I, Garagnani P, Lareu MV, Pascali VL, Carracedo A. Allele frequencies of fifteen STRs in a representative sample of the Italian population. Forensic Sci Int Genet 2009; 3(2): e29-30. https://doi.org/10.1016/j.fsigen.2008.05.002

[8] Wang J, Hiao Hue X, Bo Song X. Allele frequencies of nine non-CODIS STR loci in Western Chinese Han population. Forensic Science International: Genetics 2013; 7(3). https://doi.org/10.1016/j.fsigen.2013.01.004

[9] Farhan MM, Hadi S, Iyengar A, Goodwin W. Population genetic data for 20 autosomal STR loci in an Iraqi Arab population: Application to the identification of human remains. Forensic Sci Int Genet 2016; 25:e10-e1. https://doi.org/10.1016/j.fsigen.2016.07.017

[10] Butler JM. Genetics and genomics of core STR loci used in human identity testing. J Forensic Sci 2006; 51: 253-65. https://doi.org/10.1111/j.1556-4029.2006.00046.x

Received on 10-09-2020 Accepted on 07-10-2020 Published on 02-11-2020

https://doi.org/10.29169/1927-5129.2020.16.10