GMR

Allele frequencies in Azuay Population in Ecuador

P.P. Orellana[1], C.F. Andrade[1], C.L. Arciniegas[2] and G.C. Iannacone[3]

[1]Academic Unit of Health and Welfare, Odontology Career,
Laboratory of Molecular Biology and Genetics,
Catholic University of Cuenca, Cuenca, Ecuador
[2]Academic Unit of Social Sciences, Journalism, Information and Law,
Law Career, Catholic University of Cuenca, Cuenca, Ecuador
[3]Laboratory of Molecular Biology and Genetics,
Institute of Legal Medicine, Lima, Peru

Correspondig autor: P.P. Orellana
E-mail: porellana@ucacue.edu.ec

Genet. Mol. Res. 16 (3): gmr16039797
Received August 11, 2017
Accepted September 15, 2017
Published September 27, 2017
DOI http://dx.doi.org/10.4238/gmr16039797

ABSTRACT. One hundred and eighty-two samples of unrelated people who requested the paternity test at the Molecular Biology and Genetics Laboratory of the Catholic University of Cuenca-Ecuador in the province of Azuay were studied, except for the *D1S1656* (180 samples) and *SE33* (89 samples) markers. The STRs *D22S1045*, *D3S1358*, *VWA*, *D16S539*, *D2S1338*, *D8S1179*, *D21S11*, *D18S51*, *D19S433*, *TH01*, *FGA*, *D1S1656*, *D12S391*, *D10S1248*, *D2S441*, and *SE33* were typed from blood samples, amplifying the DNA by polymerase chain reactions and electrophoresis. The allele frequencies were estimated by simple counting and the impartial heterozygosity was also calculated. The Hardy-Weinberg equilibrium theory was studied. In the results obtained with the analyzed markers, the largest number of alleles can be observed in the markers with the highest polymorphic information content (PIC): *D21S11*, *D16S539*, *D2S1338*, *D19S433*, *D18S51*, *FGA*,

D1S1656, and *D12S391*. In addition, *SE33* was analyzed in certain samples, showing as result a high PIC, in fact, the highest one because of its great polymorphisc characteristic. Likewise, these markers are the ones providing the highest probability of discrimination and the lowest probability of coincidence.

Key words: Short tandem repeat; Allele frequencies, Azuay population

INTRODUCTION

Short tandem repeat (STR) markers were first described as effective tools for human identity testing in the early 1990's (Butler, 2006). Over the past decade, the human identity testing community has settled on a set of core STR loci that are widely used for DNA typing applications (Butler, 2006). The usefulness of genetic markers for identity testing and paternity analysis is based on known allele frequencies for the genetic markers analyzed (Cifuentes et al., 2008). STR loci are short, repetitive sequences (3-7 base pairs in length) distributed throughout the human genome (Butler, 2005). A variety of commercial kits enable robust amplification of these core STR loci (Butler, 2006). NGM is a kit of PCR amplification of fifteen STR: *D22S1045*, *D3S1358*, *VWA*, *D16S539*, *D2S1338*, *D8S1179*, *D21S11*, *D18S51*, *D19S433*, *TH01*, *FGA*, *D1S1656*, *D12S391*, *D10S1248*, and *D2S441*, and the gender determination locus, *Amelogenin* (Applied Biosystems, 2015). The NGM SElect kit incorporates an additional STR, the *SE33*, to the mentioned above (Applied Biosystems, 2015).

The Province of Azuay is one of the 24 provinces forming up the Republic of Ecuador. The administrative capital city of Azuay is Cuenca, which is also the largest and most populated city of this province. Ecuadorian individuals originated from a racial miscegenation among ancestral indigenous groups with Spanish Caucasoid settlers and African descendants (Gaviria et al., 2013).

In addition to technical validation, the implementation of STRs requires population studies that include estimating various statistical parameters for forensic studies. At the intra-population level, the frequencies should be estimated as allelic frequencies and should verify the fact that the population in which the genetic analysis system will be used is in Hardy-Weinberg equilibrium, since this allows the use of the binomial squared formula to estimate the frequency of genotypes from allele frequencies. The associations between pairs of loci should be ruled out the linkage disequilibrium (DL), which allows the use of product rule in order to estimate the frequency of genetic profiles. In addition, it is convenient to estimate statistical parameters of forensic interest that indicates the expected utility a priori for each locus and for the genetic system such as: heterozygosity (H_E), power of exclusion (PE), power of discrimination (PD), polymorphic information content (PIC), and index of paternity (IPT). On the other hand, at the inter-population level the validation of STRs usually includes the comparison with other populations to establish their genetic relationships, structure and even the knowledge of their origins (Martínez et al., 2016).

In summary, the forensic parameters and the population validation enable these systems to be used to estimate the frequency of a genetic profile, or to calculate the paternity probability in a criminal law case; when the alleged father and son agree on a paternity test. For this purpose, various population studies have been carried out and many databases have been generated throughout the world using STRs (Martínez et al., 2016).

Yet, there is no information published about gene frequencies of multiallelic loci in the population of Azuay, Ecuador. The present study describes the allele frequencies of fifteen and sixteen STR loci in this population.

MATERIAL AND METHODS

The blood used in the paternity test was obtained from unrelated individuals residing in Azuay, Ecuador, through venipuncture and its collection in FTA classic cards. Later, these cards were perforated with a 1.2-mm micropuncher to obtain the DNA for the next step, the PCR amplification.

The PCR amplification of the first 15 genetic markers was made by using the NGM kit. Then, the NGM Select Kit was employed (including the *SE33*) for the amplification of 16 STRs.

The amplified samples were placed in the ABI3500 genetic analyzer and the Data Collection software helped to obtain the capylar electrophoresis results that were analyzed by the GenneMapper-IDX software.

Statistical analysis

The allele frequencies were determined and adjusted to the genotypic frequencies with EHW for each STR. The statistical parameters of forensic interest were determined using the PowerStats and GDA Softwares.

RESULTS AND DISCUSSION

The allele frequencies of the 16 autosomal STRs were estimated and included in the SElect NGM. Although they are the most basic parameters, the allele frequencies are the most useful data employed by forensic geneticists for biostatistic interpretations of each paternity test and for forensic cases when there is an agreement. Among them, a minimum of allele frequency is essential to interpret cases of null or rare alleles that can be used as a benefit for the accused (Martínez et al., 2016).

In the processes of identifying or analyzing biological ties of kinship for forensic purposes, it is necessary to have the largest number of markers with the highest probability of discrimination to avoid random collation. In this case study, there are 9 markers that are highly polymorphic and among them, there is the *SE33* that shows the highest degree of polymorphic information, although this is a high molecular weight marker that is present in many commercial kits. Therefore, in cases where a small amount of amplifiable DNA is obtained, either because of the intrinsic condition of the samples' type (number of nucleated cells) or quality, which is mainly involved in the degree of degradation of the same DNA and/or the presence of inhibitors, it will partially amplify or not, being the partial amplification the greatest risk, making it difficult to distinguish the homozygous state.

In the results obtained with the analyzed markers, the largest number of alleles can be observed in the markers with the highest PIC: *D21S11*, *D16S539*, *D2S1338*, *D19S433*, *D18S51*, *FGA*, *D1S1656*, and *D12S391*. In addition, SE33 was analyzed in certain samples, showing as a result a high PIC. In fact, the highest one because of its great polymorphism capacity. Likewise, these markers are those providing the highest probability of discrimination and the lowest probability of coincidence (Table 1).

Table 1. Structure of the population at the level of heterozygosity of 15 markers not including SE33.

ALL population	HE	HO	f
Locus			
D8S1179	0.770675	0.711111	0.077487
D21S11	0.839106	0.861111	-0.026300
D3S1358	0.675905	0.677778	-0.002778
TH01	0.676168	0.666667	0.014091
D16S539	0.788889	0.688889	0.127070
D2S1338	0.838440	0.850000	-0.013826
D19S433	0.831492	0.805556	0.031277
vWA	0.702368	0.733333	-0.044216
D18S51	0.855308	0.855556	-0.000290
FGA	0.854968	0.833333	0.025373
D2S441	0.617936	0.594444	0.038118
D22S1045	0.579155	0.516667	0.108165
D10S1248	0.696735	0.683333	0.019287
D1S1656	0.873863	0.861111	0.014632
D12S391	0.816868	0.805556	0.013887
All	0.761192	0.742963	0.024013
SE33*	0.937980	0.876404	0.065995

Hardy-Weimberg equilibrium and linkage imbalance

When analyzing the Hardy-Weimberg equilibrium, it is observed that at a level of 0.05 there is no equilibrium in the markers *D8S1179*, *D16S539*, and *SE 33*, but when the limit is 0.01, the only one that has a highly significant imbalance is the *D16S539* marker (Table 2).

Table 2. Balance Hardy Weimberg.

Population # 1 (Azuay) of 180 individuals	Prob	Locus combination
Runs		
3200*	0.035000*	D8S1179*
3200	0.481875	D21S11
3200	0.946250	D3S1358
3200	0.730625	TH01
3200*	0.001250*	D16S539*
3200	0.757500	D2S1338
3200	0.286250	D19S433
3200	0.350000	vWA
3200	0.941875	D18S51
3200	0.402500	FGA
3200	0.379375	D2S441
3200	0.058125	D22S1045
3200	0.633125	D10S1248
3200	0.545000	D1S1656
3200	0.618750	D12S391
Population # 1 (Azuay) of 89 individuals		
Runs	Prob	Locus combination
3200*	0.033750*	SE33*

*Significative probability to Hardy-Weimberg disequilibrium.

In the case of linkage disequilibrium, there is a very significant imbalance in the *D16S539* marker that can be observed, and could be expected in populations, where historically, there have been large foreign components or in cases of miscegenation (Loh et al., 2013).

Deficit and excess heterozygotes in the study population

When observing the analized makers it can be seen that a total of 5 markers have

excess of heterozygotes while the rest have heterozygotes deficit, being *D16S539* the marker with the greatest value of heterozygotes deficit. (Table 1). Likewise, heterozygosity as a value is the highest in the most polymorphic markers, except for the *D16S539* marker (Frequency Table 3). The general tendency of a slight heterozygote deficit is observed.

Table 3. Allelic frequencies, forensic parameters and of genetic structure in the population of Azuay.

Alleles	D8S1179 1	D21S11 1	D3S1358 1	TH01 1	D16S539 1	D2S1338 1	D19S433 1	vWA 1	D18S51 1	FGA 1	D2S441	D22S1045	D10S1248	D1S1656	D12S391	SE33
6				0.3379												
7				0.4313												
8	0.0027			0.0357	0.0055											
9	0.0027			0.0495	0.2225						0.0027					
9.2							0.0027			0.0027						
9.3				0.1429												
10	0.1154			0.0027	0.2637				0.0165		0.5632					0.0056
11	0.0604				0.1676		0.0165		0.0220		0.2088	0.0275		0.0139		0.0056
11.3											0.0275					
12	0.1538		0.0027		0.2418		0.0302	0.0027	0.0687		0.0165	0.0027	0.0137	0.0639		0.0112
12.2							0.0110									
13	0.3791		0.0027		0.0797		0.2170	0.0082	0.1044		0.0110	0.0027	0.2335	0.1306		
13.2							0.1126									
13.3	0.0027															
14	0.2060		0.0412		0.0165		0.2857	0.0192	0.2720		0.1429	0.0055	0.4203	0.1139		0.0337
14.2							0.0467									
15	0.0604		0.4753		0.0027		0.1401	0.0797	0.1538		0.0275	0.4066	0.2637	0.1361	0.0055	0.0506
15.2							0.0742									
15.3														0.0111		
16	0.0165		0.2720			0.0082	0.0302	0.3846	0.1099			0.5000	0.0632	0.2028	0.0357	0.0674
16.2							0.0302									
16.3														0.0528		
17			0.1538			0.1621		0.3626	0.1236	0.0110		0.0495	0.0055	0.0583	0.0247	0.1236
17.2							0.0027									
17.3														0.1639	0.0082	
18			0.0467			0.0824		0.1016	0.0659	0.0137		0.0055		0.0056	0.1951	0.0506
18.3														0.0361	0.0027	
19			0.0055			0.2280		0.0357	0.0192	0.0495					0.1868	0.1011
19.2																0.0112
19.3														0.0111	0.0275	
20						0.1951		0.0055	0.0110	0.0330					0.3104	0.0449
21						0.0247			0.0165	0.0742					0.1016	0.0169
21.2																0.0056
22						0.0440			0.0082	0.0824					0.0412	
22.2																0.0112
23						0.1868			0.0055	0.1291					0.0357	
23.2																0.0056
24						0.0467			0.0027	0.1978					0.0137	
24.2																0.0169
25						0.0165				0.2280					0.0110	
25.2																0.0337
26						0.0055				0.1538						
26.2																0.0169
27		0.0027								0.0165						
27.2																0.0730
28		0.0742								0.0082						
28.2																0.0955
29		0.1511														
29.2																0.0674
30		0.2610														
30.2		0.0385														0.0787
31		0.0659														
31.2		0.1978														0.0506
32		0.0110														
32.2		0.1346														0.0112
33.2		0.0604														0.0056
34.2																0.0056
35.2		0.0027														
Homozygotes	0.29	0.14	0.32	0.34	0.31	0.15	0.20	0.26	0.14	0.16	0.41	0.48	0.32	0.14	0.19	0.12
Heterozygotes	0.71	0.86	0.68	0.66	0.69	0.85	0.80	0.74	0.86	0.84	0.59	0.52	0.68	0.86	0.81	0.88
Total Alleles	364	364	364	364	364	364	364	364	364	364	364	364	364	360	364	178
Probability of coincidence	0.0852	0.0509	0.1618	0.1539	0.0749	0.0509	0.0533	0.1524	0.0369	0.0404	0.1623	0.2501	0.1472	0.0348	0.0604	0.0213
Power of discrimination	0.9148	0.9491	0.8382	0.8461	0.9251	0.9491	0.9467	0.8476	0.9631	0.9596	0.8377	0.7499	0.8528	0.9652	0.9396	0.9787
Polymorphic Information Content	0.7392	0.8171	0.6224	0.6185	0.7542	0.8164	0.8080	0.6527	0.8385	0.8353	0.5731	0.4960	0.6411	0.8582	0.7919	0.9271
Probability of exclusion	0.4507	0.7199	0.3919	0.3760	0.4165	0.6981	0.6031	0.4866	0.7090	0.6658	0.2766	0.2075	0.4000	0.7169	0.6134	0.7475
Typical paternity index	1.7500	3.6400	1.5424	1.4918	1.6250	3.3704	2.5278	1.8958	3.5000	3.0333	1.2133	1.0460	1.5690	3.6000	2.6000	4.0455
Minimum allele frequency	0.0151	0.0170	0.0147	0.0146	0.0149	0.0168	0.0161	0.0153	0.0169	0.0165	0.0139	0.0134	0.0148	0.0171	0.0162	0.0339
Hardy-Weinberg Equilibrium	0.0350	0.4819	0.9463	0.7306	0.0013	0.7575	0.2863	0.3500	0.9419	0.4025	0.3794	0.0581	0.6331	0.5450	0.6188	0.0338

ACKNOWLEDGMENTS

The authors present our acknowledgments to the Directors of the Universidad Católica de Cuenca (Cuenca-Ecuador) for the help and the economic support provided for the culmination of this investigative work, also to all the individuals who authorized the use of their samples for this study.

REFERENCES

Applied Byosistems (2015). User guide. AmpFlSTR® NGM™ PCR Amplification Kit.

Butler JM (2005). Forensic DNA typing: Biology, Technology, and Genetics of STR Markers. Elsevier, Oxford.

Butler JM (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *J. Forensic Sci.* 51: 253-265. https://doi.org/10.1111/j.1556-4029.2006.00046.x

Cifuentes L, Jorquera H, Acuña M, Ordóñez J, et al. (2008). Allele frequencies for 12 autosomal short tandem repeat loci in two Bolivian populations. *Genet. Mol. Res.* 7: 271-275. https://doi.org/10.4238/vol7-1gmr368

Gaviria A, Zambrano AK, Morejon G, et al. (2013). Twenty two autosomal microsatellite data from Ecuador (Powerplex Fusion). *Forensic Sci. International. Genet. Suppl. Ser.* 4: 330-333. https://doi.org/10.1016/j.fsigss.2013.10.169

Loh PR, Lipson M, Patterson N, Moorjani P, et al. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193: 1233-1254. https://doi.org/10.1534/genetics.112.147330

Martínez VM, Aguilar JA, Inclán A, et al. (2016). Parámetros forenses del sistema Powerplex® 21 (Promega Corp.) en población mestiza del occidente de México. *Revista Española de Medicina Legal.* 42: 10-16. https://doi.org/10.1016/j.reml.2015.03.001