Biology Faculty Publications and Presentations | Department of Biological Sciences

3-1-2018

Allele Frequencies of 15 STR Loci (Identifiler™ Kit) in Basque-Americans

Jason Besecker
Boise State University

Gianluca Peri
Boise State University

Michael Davis
Boise State University

Josu Zubizarreta
Boise State University

Greg Hampikian
Boise State University

Allele Frequencies of 15 STR loci (Identifiler™ kit) in Basque-Americans

Abstract: Individuals with Basque ancestry form a historically and culturally important minority of the population of the western United States. Allele frequencies for the 15 autosomal STRs in the AmpFlSTR® Identifiler® PCR Amplification Kit (Applied Biosystems) from 156 unrelated self-identified Basque individuals born in the United States are presented. Allele frequencies were used to calculate parameters commonly used in genetics and forensics including power of discrimination (PD), power of exclusion (PE), polymorphic information content (PIC), and expected heterozygosity (He). The sample population was also compared with the European Basque population and the major American ethnicities.

Keywords:

Short Tandem Repeats (STR)

Allele Frequency

American Basques

Forensic DNA

Population:

The Basque are a linguistic isolate and ethnic group, inhabiting the Basque Country (Euskal Herria) of northern Spain and southwestern France; they have been the subjects of numerous genetic studies [1-8]. Once considered genetic isolates, they are now known to be more closely related to neighboring European populations than other world populations [1], although they can be differentiated from other Europeans with high-density SNP genotyping [2]. Basque migrants

settled in the region of the Intermountain West in the late 1800's, especially the regions of Southwestern Idaho, Southeastern Oregon, and Northern Nevada. The 2000 census registered a total of 57,793 people of Basque ethnicity residing in the USA [9].

Buccal swab samples (Whatman® Omni Swabs) were collected at the Jaialdi and San Inazio Basque Festivals, in Boise, ID, USA. Samples were taken from 156 unrelated individuals, who identify themselves as Basque and were born in the United States. Informed consent was obtained from each individual as well as information on their ethnic background.

In previous studies it has been shown that genetic diversity in European Basque populations is spatially structured [3-5], and it has been suggested that it is necessary to consider different Basque populations as independent populations [6].

Extraction:

DNA was extracted from buccal swabs using the QIAamp® DNA Mini Kit (QIAGEN) following manufactures protocol for DNA purification from buccal swabs.

PCR:

PCR was performed using AmpFlSTR® Identifiler® PCR Amplification Kit (Applied Biosystems, Thermofisher Scientific) following the manufacturer's protocol.

Typing:

Multiplex PCR samples from each individual and controls were analyzed via capillary gel electrophoresis using the ABI 3130 Genetic Analyzer (Applied Biosystems). The internal lane

standard used was GeneScan™ 500 LIZ™ Size Standard (Applied Biosystems). Genotyping was carried out using GeneMapper ID-X v1.0 software (Applied Biosystems).

Quality control:

The positive control DNA provided in the AmpFlSTR® Identifiler® PCR Amplification Kit, and negative extraction controls were run along with all experimental samples. Individuals analyzing the samples on the ABI 3130 received advanced GeneMapper ID-X training from Applied Biosystems "HID University 300". In a previous study [7], our laboratory underwent a proficiency testing quality control check, in conjunction with submission to the Y-HRD database [10, 11].

Analysis of data:

Arlequin v3.5 software [12] was used to determine observed heterozygosity (Ho), expected heterozygosity (He) [13], p-values (p) for Hardy-Weinberg Equilibrium exact test [14] using a Markov chain, and pair-wise population comparisons of genetic distances (F_{st}) [15] between the sample population and the European Basque population [16], as well as with the main ethnicities of the United States of America [17] (10,000 permutations; significance level 0.05). Powerstats v1.2 software [18] was used to calculate match probability (MP), power of discrimination (PD) [19], power of exclusion (PE), polymorphic information content (PIC) [20] and Typical Paternity Index (TPI) for each locus.

Results:

Statistical parameters and allele frequencies for the 15 STR loci of self-identified Basque individuals born in the United States (N=156) are reported in Table 1. F_{st} values for the pair-wise comparison with other populations of interest are reported in Table 2. Average F_{st} values have been calculated for each comparison [15, 21].

Access of data:

Available upon request to greghampikian@boisestate.edu.

Other remarks:

The results indicate that all loci are in Hardy-Weinberg equilibrium, considering Bonferroni correction (0.05/15 = 0.003) [22]. The most variable loci are D18S51 (PD: 0.963), D21S11 (PD: 0.943), FGA (PD: 0.963) and D2S1338 (PD: 0.962). The average level of expected heterozygosity (He) for the overall population is 0.7815. The combined match probability (MP) and power of exclusion (PE) for all 15 STR loci of American Basques is 3.300 x10^{-17} and 0.999995 respectively.

The average F_{st} index between the sampled American Basques and the European Basques indicated very little difference between these two populations. A close affinity between the American Basque population and the Hispanic and Caucasian ethnicities was noted. On the other hand, the average F_{st} values indicated a significant genetic difference between American Basques and African Americans, and between American Basques and American Asians (see Table 2 for F_{st} values).

Data obtained from this study display that most of the loci have a high power of discrimination and exclusion, confirming their use for individual identification.

Table 1: Allelic frequencies for American Basques population. (N=156)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
6	-	-	-	-	-	0.234	-	-	-	-	-	-	-	-	-
7	-	-	0.026	-	-	0.151	-	-	-	-	-	-	-	-	-
8	0.022	-	0.125	0.006	-	0.096	0.170	0.022	-	-	-	0.519	-	-	-
9	0.010	-	0.090	0.022	-	0.218	0.038	0.071	-	-	-	0.119	-	0.029	-
9.3	-	-	-	-	-	0.288	-	-	-	-	-	-	-	-	-
10	0.071	-	0.337	0.237	-	0.013	0.048	0.029	-	-	-	0.058	0.010	0.093	-
11	0.042	-	0.260	0.343	-	-	0.353	0.256	-	-	-	0.256	0.013	0.365	-
12	0.080	-	0.131	0.343	-	-	0.199	0.365	-	0.119	-	0.048	0.212	0.324	-
13	0.343	-	0.022	0.045	0.010	-	0.099	0.247	-	0.218	0.003	-	0.106	0.183	-
13.2	-	-	-	-	-	-	-	-	-	0.006	-	-	-	-	-
14	0.269	-	0.010	0.003	0.099	-	0.093	0.010	-	0.359	0.103	-	0.112	0.006	-
14.2	-	-	-	-	-	-	-	-	-	0.019	-	-	-	-	-
15	0.151	-	-	-	0.308	-	-	-	-	0.160	0.080	-	0.183	-	-
15.2	-	-	-	-	-	-	-	-	-	0.054	-	-	-	-	-
16	0.010	-	-	-	0.250	-	-	-	0.080	0.032	0.247	-	0.122	-	-
16.2	-	-	-	-	-	-	-	-	-	0.013	-	-	-	-	-
17	0.003	-	-	-	0.147	-	-	-	0.237	0.019	0.269	-	0.138	-	-
18	-	-	-	-	0.167	-	-	-	0.051	-	0.231	-	0.064	-	0.022
19	-	-	-	-	0.019	-	-	-	0.125	-	0.067	-	0.019	-	0.122
20	-	-	-	-	-	-	-	-	0.147	-	-	-	0.003	-	0.099
21	-	-	-	-	-	-	-	-	0.029	-	-	-	0.010	-	0.192
22	-	-	-	-	-	-	-	-	0.019	-	-	-	0.010	-	0.115
22.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.006
23	-	-	-	-	-	-	-	-	0.096	-	-	-	-	-	0.135
24	-	-	-	-	-	-	-	-	0.096	-	-	-	-	-	0.106
24.2	-	0.003	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	-	-	-	-	-	-	-	0.109	-	-	-	-	-	0.128
25.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.003
26	-	0.003	-	-	-	-	-	-	0.010	-	-	-	-	-	0.051
27	-	0.022	-	-	-	-	-	-	-	-	-	-	-	-	0.019
28	-	0.093	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	0.234	-	-	-	-	-	-	-	-	-	-	-	-	-
29.2	-	0.003	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	0.314	-	-	-	-	-	-	-	-	-	-	-	-	-
30.2	-	0.035	-	-	-	-	-	-	-	-	-	-	-	-	-
31	-	0.048	-	-	-	-	-	-	-	-	-	-	-	-	-
31.2	-	0.054	-	-	-	-	-	-	-	-	-	-	-	-	-
32	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
32.2	-	0.138	-	-	-	-	-	-	-	-	-	-	-	-	-
33.2	-	0.032	-	-	-	-	-	-	-	-	-	-	-	-	-
34.2	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
Ho	0.763	0.750	0.782	0.699	0.814	0.795	0.769	0.699	0.872	0.808	0.750	0.673	0.833	0.660	0.846
He	0.776	0.813	0.780	0.708	0.785	0.785	0.788	0.736	0.868	0.781	0.794	0.647	0.862	0.721	0.879
p	0.658	0.164	0.012	0.119	0.548	0.177	0.549	0.246	0.237	0.841	0.077	0.877	0.860	0.032	0.005
PD	0.918	0.943	0.913	0.852	0.911	0.911	0.925	0.883	0.962	0.922	0.920	0.820	0.963	0.865	0.963
PE	0.532	0.510	0.566	0.426	0.625	0.590	0.543	0.426	0.738	0.613	0.510	0.388	0.662	0.369	0.687
PIC	0.740	0.790	0.750	0.650	0.750	0.750	0.760	0.690	0.850	0.750	0.760	0.600	0.840	0.670	0.860
MP	0.082	0.057	0.087	0.148	0.089	0.089	0.075	0.117	0.038	0.078	0.080	0.180	0.037	0.135	0.037
TPI	2.110	2.000	2.290	1.660	2.690	2.440	2.170	1.660	3.900	2.600	2.000	1.530	3.000	1.470	3.250

Ho: Observed Heterozygosity – He: Expected Heterozygosity – p: Exact Test of Hardy-Weinberg – PD: Power of Discrimination – PE: Power of Exclusion– PIC: Polymorphic Information Content – MP: Match Probability – TPI: Typical Paternity Index.

Table 2: F_{st} values calculated for each locus, comparing the differentiation between the sampled America Basque (AB) population and other populations of interest: European Basques (EB), American (all ethnicities) (US), African American (AfA), Caucasian Americans (CA), Hispanic Americans (HA), Asian Americans (AsA).

STR locus	AB vs EB	AB vs US	AB vs AfA	AB vs CA	AB vs HA	AB vs AsA
D8S1179	0.00066^{ns}	0.00620	0.01313	0.01226	0.00295^{ns}	0.02173
D21S11	0.00516	0.00690	0.02993	0.00455	0.00078^{ns}	0.00417^{ns}
D7S820	0.00616	0.00141^{ns}	0.00837	0.00807	-0.00122^{ns}	0.00922
CSF1PO	-0.00195^{ns}	0.00227^{ns}	0.00892	-0.00005^{ns}	0.00112^{ns}	0.01197
D3S1358	0.00599	0.00355	0.01153	0.00132^{ns}	0.00060^{ns}	0.01418
TH01	0.00022^{ns}	0.01839	0.07540	0.00728	0.01717	0.07621
D13S317	0.00573	0.01442	0.04665	0.00556	0.02298	0.01727
D16S539	-0.00071^{ns}	0.02310	0.03811	0.00779	0.02290	0.09458
D2S1338	-0.00179^{ns}	0.00804	0.02417	0.00129^{ns}	0.00744	0.02822
D19S433	-0.00103^{ns}	0.00430	0.02192	0.00084^{ns}	0.00131^{ns}	0.01956
vWA	0.00233^{ns}	0.00197^{ns}	0.01058	0.00098^{ns}	0.00327^{ns}	0.01353
TPOX	-0.00113^{ns}	0.00132^{ns}	0.02473	-0.00211^{ns}	0.00117^{ns}	0.00025^{ns}
D18S51	0.00106^{ns}	0.00859	0.02047	0.00458	0.00559	0.03338
D5S818	-0.00334^{ns}	0.00145^{ns}	0.01353	0.00295^{ns}	0.00393^{ns}	0.02614
FGA	0.00596	0.00689	0.01029	0.00830	0.00214^{ns}	0.01819
Average F_{st}	0.00580	0.01004	0.02385	0.00730	0.01522	0.02955

ns: not significant.

References

[1] H. Laayouni, F. Calafell, J. Bertranpetit. A genome-wide survey does not show the genetic distinctiveness of Basques. Hum Genet. 127 (2010) 455-458.

[2] N. Rodríguez-Ezpeleta, J. Álvarez-Busto, L. Imaz, M. Regueiro, M.N. Azcárate, R. Bilbao, M. Iriondo, A. Gil, A. Estonba, A.M. Aransay. High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations. Hum. Genet. 128 (2010) 113-117.

[3] A.I. Aguirre, A. Vicario, L.I. Mazon, A. Estomba, M.M. de Pancorbo, V. Arrieta-Pico, F. Perez Elortondo, C.M. Lostao. Are the Basques a single and unique population? Am. J. Hum. Genet. 49 (1991) 450-458.

[4] A.M. Pérez-Miranda, M.A. Alfonso-Sánchez, J.A. Peña, R. Calderón. HLADQA1 polymorphism in autochthonous Basques from Navarre (Spain): genetic position within European and Mediterranean scopes. Tissue Antigens 61 (2003) 465-474.

[5] C. Manzano, C. de la Rúa, M. Iriondo, L.I. Mazón, A. Vicario, A.I. Aguirre. Structuring the genetic heterogeneity of the Basque population: a view from classical polymorphisms. Hum. Biol. 74 (2002) 51-74.

[6] A.M. Pérez-Miranda, M.A. Alfonso-Sánchez, A. Kalantar, S. García-Obregón, M.M. de Pancorbo, J.A. Peña, R.J. Herrera. Microsatellite data support subpopulation structuring among the Basques. J. Hum Genet 50 (2005) 403-414.

[7] J. Zubizarreta, M.C. Davis, G. Hampikian. The Y-STR Genetic Diversity of an Idaho Basque Population with Comparison to European Basques and US Caucasians. Hum. Biol. 83 (2011) 685-694.

[8] M. Davis, S.J. Novak, and G. Hampikian "Mitochondrial DNA Analysis of an Immigrant Basque Population: Loss of Diversity Due to Founder Effects". American Journal of Physical Anthropology, 144(4) (2011) 516-525.

[9] American FactFinder. Census 2000 Summary File 3. http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF3_PCT018&prodType=table.

[10] L. Gusmão, J.M. Butler, Á. Carracedo. DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. Forensic Sci. Int. 157 (2006) 187-197.

[11] S. Willuweit, L. Roewer. Y chromosome haplotype reference database (YHRD): update. FSI: Genetics, 1 (2007) 83-87.

[12] L. Excoffier, G. Laval, S. Schneider. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol. Bioinform. 1 (2005) 47-50.

[13] M. Nei. Molecular Evolutionary Genetics. Columbia University Press, New York, 1987.

[14] S.W. Guo, E.A. Thompson. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48 (1992) 361-372.

[15] B.S. Weir, C.C. Cockerham. Estimating F-statistics for the analysis of population structure. Evolution 38 (1984) 1358-1370.

[16] I. Yurrenbaso, J.A. Ajuriagerra, A. Alday, I. Lezama, J.A. Pérez, E. Romón, I. Uriate, O. García. Allele frequencies and concordance study between the Identifiler and the PowerPlex ESX17 systems in the Basque Country population. FSI: Genetics, 5 (2011) e79-e80.

[17] C.R. Hill, D.L. Duewer, M.C. Kline, M.D. Coble, J.M. Butler. U.S. population data for 29 autosomal STR loci. FSI: Genetics, 7 (2013) e82-e83.

[18] Power Stats V12.xls software. Free program distributed by Promega Corporation, USA: Personal Communication.

[19] D.A. Jones. Blood samples: Probabilities of discriminations. J. Forensic Sci. Soc. 12 (1972) 355-359.

[20] D. Botstein, R.I. White, M. Skolnich, R.W. Davis. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32 (1980) 314-331.

[21] J. Reynolds, B.S. Weir, C. Cockerham. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics, 105 (1983) 767-779.

[22] Weir BS. Genetic data analysis II. USA: Sinauer Associates; 1996. p. 134.