

## Announcement of Population Data

Allele frequencies of 15 tetrameric short tandem repeats  
(STRs) in Andalusians from Huelva (Spain)Clotilde Coudray<sup>a</sup>, Rosario Calderon<sup>b</sup>, Evelyne Guitard<sup>a</sup>, Beatriz Ambrosio<sup>b</sup>,  
Antonio González-Martín<sup>b</sup>, Jean-Michel Dugoujon<sup>a,\*</sup><sup>a</sup> Centre d'Anthropologie, FRE 2960 CNRS, 37 allées Jules Guesde, 31073 Toulouse Cedex 4, France<sup>b</sup> Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense, 28040 Madrid, Spain

Received 22 December 2006; received in revised form 15 January 2007; accepted 19 January 2007

Available online 2 March 2007

**Abstract**

The allele frequency distribution of 15 short tandem repeat (STR) loci contained in the AmpF/STR Identifier<sup>TM</sup> PCR Amplification Kit (Applied Biosystems), was determined in 114 individuals from Andalusia (province of Huelva), in the southernmost part of the Iberian Peninsula. After Bonferroni's correction, no deviations from the Hardy–Weinberg equilibrium were observed for all samples at the 15 STR loci. All loci are highly polymorphic. The aim of the study was to obtain accurate allele frequencies relevant for applications in forensics and population genetics. Comparative analyses between our population data and other population samples gathered from the literature are also presented.

© 2007 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** STR; Iberian Peninsula; Andalusia; AmpF/STR Identifier<sup>TM</sup>; Population genetics

**Population:** We studied 114 unrelated and healthy individuals of both sexes coming from the province of Huelva in Andalusia (southern Spain). The sampling localities were El Repilado, Villablanca, El Cerro del Andévalo, Aracena and Valverde del Camino [1] (Fig. 1). Donors were subjected to an interview in order to obtain their informed consent and personal and family information. All the samples correspond to autochthonous people and their parents and all grandparents were born in 37 different municipalities of the province of Huelva. Huelva is the most western province of Andalusia and is divided into four small regions (*Comarcas*) and they are: el Andévalo, el Condado, la Costa, and la Sierra. The existence of human populations in the region goes back to the Upper Paleolithic. Then, many people of different cultures and origins settled there, including the Tartessians and the Phoenicians (11th century B.C.), the Greeks and the Carthaginians (6th century B.C.), the Romans (3rd century B.C.), and the Visigoths (5th century A.D.). In 711, the Arabs and Berbers, crossing the Strait of Gibraltar, established there the center of their western

emirate. Huelva remained under Islamic domination until the reconquest of western Andalusian by the Christians in 1262.

**DNA extraction:** Genomic DNA was isolated from blood samples using a standard proteinase-K digestion followed by phenol–chloroform extraction and ethanol precipitation.

**PCR:** DNA amplification was carried out in a GeneAmp 2700 Thermal cycler (Applied Biosystems) from 1 ng target DNA, following the manufacturer's instructions for the AmpF/STR Identifier<sup>TM</sup> PCR Amplification Kit (Applied Biosystems). The samples were amplified in single reactions for all 15 loci using fluorescently labelled primers.

**Typing:** PCR products were separated by capillary electrophoresis in an ABI PRISM 310 Genetic Analyzer using the separation medium performance optimized polymer (POP) 4 and 47 cm capillaries (Applied Biosystems). The LIZ labelled ladder (GeneScan 500 LIZ) was used for sizing determinations of the amplified fragments in combination with GeneScan software (vs.3.7, Applied Biosystems). Numerical allele designations of the profiles were obtained by comparison to the allelic ladder using Genotyper software (vs.3.7, Applied Biosystems).

**Results:** See Tables 1 and 2.

**Analysis of data:** The frequency of each allele for each locus and the observed heterozygosity (*H<sub>o</sub>*) were calculated

\* Corresponding author. Tel.: +33 5 61 14 59 85; fax: +33 5 61 14 59 79.

E-mail address: [dugoujon@cict.fr](mailto:dugoujon@cict.fr) (J.-M. Dugoujon).



Table 1 (Continued)

Allele	D3S1358	D21S11	D18S51	D5S818	D13S317	D7S820	D16S539	CSF1PO	D8S1179	FGA	TH01	VWA	TPOX	D19S433	D2S1338
31	–	0.101	–	–	–	–	–	–	–	–	–	–	–	–	–
31.2	–	0.079	–	–	–	–	–	–	–	–	–	–	–	–	–
32.2	–	0.083	–	–	–	–	–	–	–	–	–	–	–	–	–
33	–	0.009	–	–	–	–	–	–	–	–	–	–	–	–	–
33.2	–	0.035	–	–	–	–	–	–	–	–	–	–	–	–	–
34.2	–	0.009	–	–	–	–	–	–	–	–	–	–	–	–	–
Ho	0.798	0.851	0.886	0.684	0.789	0.789	0.746	0.763	0.833	0.825	0.737	0.825	0.605	0.702	0.807
He	0.808	0.853	0.867	0.696	0.795	0.797	0.760	0.729	0.823	0.841	0.800	0.800	0.657	0.743	0.858
PD	0.933	0.951	0.960	0.856	0.925	0.921	0.906	0.870	0.935	0.951	0.925	0.910	0.820	0.896	0.961
PIC	0.780	0.837	0.853	0.641	0.767	0.769	0.723	0.680	0.801	0.823	0.769	0.771	0.602	0.704	0.843
PE	0.596	0.697	0.767	0.404	0.580	0.580	0.502	0.533	0.662	0.645	0.488	0.645	0.297	0.431	0.612
<i>p</i>	0.757	0.038	0.625	0.559	0.360	0.054	0.908	0.744	0.150	0.264	0.195	0.059	0.026	0.397	0.271

Ho: observed heterozygosity; He: expected heterozygosity; PIC: polymorphism information content; PD: power of discrimination; PE: power of exclusion; *p*-values of the exact test for Hardy–Weinberg equilibrium (with a 5% significance level). Bonferroni's correction (0.05/15 = 0.0033).

from the number of observed genotypes in the sample (direct gene counting). Expected heterozygosity (He) was estimated as  $1 - \sum_i p_i^2$  where  $p_i$  is the frequency of the *i*th allele in the locus. The Hardy–Weinberg equilibrium exact test (dememorization value 10,000, and 5000 iterations per batch) was assessed by means of GENEPOP (v.s.1.2) statistical package [2]. The Bonferroni correction was applied on loci showing a departure from Hardy–Weinberg equilibrium [3]. Population differentiation was tested for each locus by exact tests [4] as implemented in the ARLEQUIN program [5]. Forensic parameters (polymorphism information content, PIC; power of discrimination, PD; power of exclusion, PE) were calculated using Microsoft Excel-PowerStats program [6] downloaded from Promega corporation website at <http://www.promega.com/geneticid-tools/powerstats/>.

**Access to the data:** Complete data can be obtained from the authors on request to [dugoujon@cict.fr](mailto:dugoujon@cict.fr), or can be found online as supplementary material.

**Other remarks:** Population data for the 15 STRs are shown in Table 1. Based on the results of the exact tests, the D21S11 and TPOX loci showed departures from the Hardy–Weinberg expectations. When using the Bonferroni correction for the number of loci analysed, the two deviations are not significant. So, we can say that all the STR loci analysed in this study met the Hardy–Weinberg equilibrium.

In Andalusians from the province of Huelva, the observed heterozygosity (Ho) ranges from 0.605 (TPOX) to 0.886 (D18S51). The power of discrimination (PD) varies between 0.820 (TPOX) and 0.961 (D2S1338) and the probability of exclusion (PE) varies between 0.297 (TPOX) and 0.767 (D18S51). The most polymorphic genetic marker is D18S51 (PIC value of 0.853) and the least polymorphic is TPOX (PIC value of 0.602). For the 15 loci, the combined power of discrimination and the combined probability of exclusion are  $2.11 \times 10^{-17}$  and 0.9999976, respectively.

Table 2

Population differentiation tests between Andalusians from Huelva and other published population

STR	Huelva vs. Andalusia	Huelva vs. Catalonia	Huelva vs. NE Spain	Huelva vs. Basques from Vizcaya	Huelva vs. Basques from Alava	Huelva vs. Basques from Guipuzcoa	Huelva vs. Basques from Navarre	Huelva vs. Portugal	Huelva vs. Berbers
D3S1358	0.13270	0.52700	<b>0.00235</b>	0.06725	0.03060	<b>0.00000</b>	0.00965	<b>0.00045</b>	<b>0.00000</b>
D21S11	0.07755	<b>0.00085</b>	0.55875	0.01700	<b>0.00030</b>	<b>0.00000</b>	0.01460	0.12620	<b>0.00000</b>
D18S51	0.38410	0.98585	0.36770	0.13600	0.50755	<b>0.00040</b>	0.03590	0.08435	0.60780
D5S818	0.69740	0.08800	0.02525	0.00930	0.03145	<b>0.00000</b>	0.15775	0.36835	<b>0.00030</b>
D13S317	0.23435	0.06115	0.15420	0.25910	0.02580	0.06980	0.21015	0.17615	<b>0.00000</b>
D7S820	0.51515	0.30640	0.18700	0.09540	0.91535	0.54860	0.68025	0.27465	0.23750
D16S539	0.97215	–	0.83860	0.37095	<b>0.00000</b>	0.29420	0.00585	0.92565	<b>0.00170</b>
CSF1PO	0.66540	0.45410	0.31860	0.04055	0.00805	0.04855	0.34695	0.66765	0.24600
D8S1179	0.11605	0.90980	0.49475	0.07525	0.27670	0.31045	0.27330	0.77250	0.25505
FGA	0.17790	0.16645	0.13555	<b>0.00015</b>	<b>0.00145</b>	<b>0.00030</b>	<b>0.00055</b>	0.79320	0.68960
TH01	0.70520	0.07310	0.11820	0.00485	0.00735	<b>0.00045</b>	0.01910	0.80890	0.00710
VWA	0.83670	0.04920	0.24145	0.01805	0.19100	0.33015	0.09670	0.32620	0.38975
TPOX	0.03865	0.58355	0.38555	0.15730	<b>0.00000</b>	0.00305	<b>0.00000</b>	0.12640	<b>0.00025</b>
D19S433	–	–	0.66515	–	–	–	–	0.26285	<b>0.00000</b>
D2S1338	–	–	0.23465	–	–	–	–	0.83395	0.01570

Data for 15 STR loci. In bold: *p*-values that remain significant after Bonferroni's correction. References for the population data used: Andalusians (this study); Andalusia [7]; Catalonia [8]; northeast Spain [9,10]; Basques from Vizcaya [11]; Basques from Alava [12]; Basques from Guipuzcoa [13]; Basques from Navarre [13]; Portugal [14]; Berbers from Bouhria [15].

Present analysis was made in order to complete STR database for Iberian populations. Another work was previously published on an Andalusian population [7] but this is the first study of D19S433 and D2S1338 allele frequency distribution in southern Spain.

We have compared our data with other published data [7–15] on geographically related population samples, using the available sets of loci (Table 2).

After Bonferroni's correction, population differentiation tests showed that, in Spain, the Andalusians from Huelva had no differences with the other surveyed Andalusian population, but significant differences were found with Catalonians in 1 out of 12 loci (D21S11), and with a group from Northeast Spain in 1 out of 15 loci (D3S1358). Considering Iberian Peninsula as a whole, Andalusians from Huelva had significant differences with Portuguese in 1 out of 15 loci (D3S1358), and with Basques from Vizcaya, Alava, Guipuzcoa and Northern Navarre, in 1 (FGA), 4 (D21S11, D16S539, FGA, TPOX), 6 (D3S1358, D21S11, D18S51, D5S818, FGA, TH01) and 2 (FGA, TPOX) out of 13 loci, respectively. As the Berbers settled in Iberian Peninsula during seven centuries, we compared the Andalusians from Huelva with Moroccan Berbers from Bouhria (Northeast Morocco). The population differentiation tests showed that Andalusians from Huelva had significant differences with Moroccan Berbers in 7 out of 15 loci (D3S1358, D21S11, D5S818, D13S317, D16S539, TPOX, D19S433).

This paper follows the guidelines for publication of population data requested by the journal [16].

## Acknowledgments

Blood samples were collected by a group of doctors and nurses from Juan Ramón Jiménez hospital, accompanied by researchers (RC and BA) from the Universidad Complutense, Madrid, Spain. This project was supported by a grant to RC from the Spanish Ministry of Science and Technology (BOS2002-01677). STR's typing of Andalusians and North African population benefited from funding of the action OMLL "Origin of Man, Language, and Languages" (EUROCORES Program, contract no. ERAS-CT-2003-980409).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.forsciint.2007.01.016.

## References

- [1] R. Calderon, B. Ambrosio, E. Guitard, A. Gonzalez-Martin, U. Aresti, J.M. Dugoujon, The genetic position of Andalusians from Huelva in relation to other European and North African populations: a study based on GM and KM allotypes, *Hum. Biol.*, in press.
- [2] M. Raymond, F. Rousset, Genepop version 1.2: a population genetics software for exact tests and ecumenicism, *J. Hered.* 86 (1995) 248–249.
- [3] B.S. Weir, *Multiple Tests in Genetic Data Analysis II*, Sinauer Associates, USA, 1996, p. 134.
- [4] M. Raymond, F. Rousset, An exact test of differentiation, *Evolution* 49 (1995) 1280–1283.
- [5] S. Schneider, D. Roessli, L. Excoffier, Arlequin: A Software for Population Genetics Data Analysis, Ver. 2.0, Genetics and Biometry Laboratory, Department of Anthropology and Ecology, University of Geneva, Geneva, Switzerland, 2000.
- [6] A. Tereba, Tools for analysis of population statistics, *Profile DNA* 2 (1999) 14–16.
- [7] P. Sanz, V. Prieto, I. Flores, Y. Torres, M. Lopez-Soto, M.J. Farfan, Population data of 13 STRs in southern Spain (Andalusia), *Forensic Sci. Int.* 119 (2001) 113–115.
- [8] A. Perez-Lezaun, F. Calafell, J. Clarimon, E. Bosch, E. Mateu, L. Gusmao, A. Amorim, N. Benchemsi, J. Bertranpetit, Allele frequencies of 13 short tandem repeats in population samples from the Iberian Peninsula and northern Africa, *Int. J. Legal Med.* 113 (2000) 208–214.
- [9] M. Paredes, M. Crespillo, J.A. Luque, J.L. Valverde, STR frequencies for the PowerPlex 16 System Kit in a population from Northeast Spain, *Forensic Sci. Int.* 135 (2003) 75–78.
- [10] M. Crespillo, M. Paredes, J.A. Luque, J.L. Valverde, Population data of the STRs D2S1338 and D19S433 from northeast of Spain, *J. Forensic Sci.* 48 (2003) 1183.
- [11] A.M. Perez-Miranda, M.A. Alfonso-Sanchez, A. Kalantar, J.A. Peña, M.M. de Pancorbo, R.J. Herrera, Allelic frequencies of 13 STR loci in autochthonous Basques from the province of Vizcaya (Spain), *Forensic Sci. Int.* 152 (2005) 259–262.
- [12] A.M. Perez-Miranda, M.A. Alfonso-Sanchez, J.A. Peña, M.M. de Pancorbo, R.J. Herrera, Genetic polymorphisms at 13 STR loci in autochthonous Basques from the province of Alava (Spain), *Legal Med.* 7 (2005) 58–61.
- [13] A.M. Perez-Miranda, M.A. Alfonso-Sanchez, A. Kalantar, S. Garcia-Obregon, M.M. de Pancorbo, J.A. Peña, R.J. Herrera, Microsatellite data support subpopulation structuring among Basques, *J. Hum. Genet.* 50 (2005) 403–414.
- [14] D. Abrantes, G. Lima, L. Cain, F. Pinheiro, Allele frequencies of 15 loci using AmpFISTR Identifier Kit in a northern Portuguese population, *J. Forensic Sci.* 49 (2004) 842–843.
- [15] C. Coudray, E. Guitard, C. Keyser-Tracqui, M. Melhaoui, M. Cherkaoui, G. Larrouy, J.M. Dugoujon, Population genetic data of 15 tetrameric short tandem repeats (STRs) in Berbers from Morocco, *Forensic Sci. Int.*, in press.
- [16] P. Lincoln, A. Carracedo, Publication of population data of human polymorphisms, *Forensic Sci. Int.* 110 (2000) 3–5.