

Results of the GEP-ISFG collaborative study on the Y chromosome STRs GATA A10, GATA C4, GATA H4, DYS437, DYS438, DYS439, DYS460 and DYS461: population data

Leonor Gusmão^a, Paula Sánchez-Diz^b, Cíntia Alves^a, Beatriz Quintáns^b, Emilio García-Poveda^c, Helena Geada^d, Eduardo Raimondi^e, Sandra María Silva de la Fuente^f, Maria Conceição Vide^g, Martin R. Whittle^h, María Teresa Zarrabeitiaⁱ, Mónica Carvalho^g, Veranice Negreiros^h, Lourdes Prieto Solla^c, José A. Riancho^{i,j}, Rebeca Campos-Sánchez^f, Cláudia Vieira-Silva^d, Ulises Toscanini^e, António Amorim^{a,k}, Angel Carracedo^{b,*}

GEP-ISFG (The Spanish and Portuguese Working Group of the International Society for Forensic Genetics)¹

^a*IPATIMUP, Institute of Pathology and Immunology of University of Porto, Porto, Portugal*

^b*Faculty of Medicine, Institute of Legal Medicine, c/o San Francisco, s/n University of Santiago de Compostela, E-15782 Santiago de Compostela, A Coruña, Spain*

^c*DNA Laboratory, Scientific Police, Madrid, Spain*

^d*Forensic Genetics Service, Faculty of Medicine, National Institute of Legal Medicine, University of Lisbon, Lisbon, Portugal*

^e*PRICAI—Favaloro Foundation, Buenos Aires, Argentina*

^f*Clinical and Molecular Analysis Laboratory, San José, Costa Rica*

^g*Forensic Genetics Service, Coimbra Department, National Institute of Legal Medicine, Lisbon, Portugal*

^h*Genomic Engenharia Molecular Ltd., São Paulo, Brazil*

ⁱ*Unit of Legal Medicine of Santander, University of Cantabria, Cantabria, Spain*

^j*Internal Medicine Service, University of Cantabria, Cantabria, Spain*

^k*Faculty of Sciences, University of Porto, Porto, Portugal*

Received 31 January 2003; accepted 6 May 2003

Abstract

The Spanish and Portuguese ISFG Working Group (GEP-ISFG) carried out a collaborative exercise in order to assess the performance of two Y chromosome STR tetraplexes, which include the loci DYS461, GATA C4, DYS437 and DYS438 (GEPY I), and DYS460, GATA A10, GATA H4 and DYS439 (GEPY II).

The groups that reported correct results in all the systems were also asked to analyse a population sample in order to evaluate the informative content of these STRs in different populations. A total of 1020 males out of 13 population samples from Argentina, Brazil, Costa Rica, Macao, Mozambique, Portugal and Spain were analysed for all the loci included in the present study.

Haplotype and allele frequencies of these eight Y-STRs were estimated in all samples. The lowest haplotype diversity was found in the Lara (Argentina) population (95.44%) and the highest (99.90%) in Macao (China).

* Corresponding author. Tel.: +34-981-582327; fax: +34-981-580336.

E-mail address: apimlang@usc.es (A. Carracedo).

¹ All authors of this article belong to this working group.

Pairwise haplotype analysis showed the relative homogeneity of the Iberian origin samples, in accordance with what was previously found in the European populations for other Y-STR haplotypes (<http://www.ystr.org>).

As expected, the four non-Caucasian samples, Macao (Chinese), Mozambique (Africans), Costa Rica (Africans) and Argentina (Lara, Amerindians), show highly significant Φ_{st} values in the pairwise comparisons with all the Caucasian samples. © 2003 Elsevier Ireland Ltd. All rights reserved.

Keywords: Y-STR haplotypes; Population genetics; GEP-ISFG

1. Introduction

Out of the Y chromosome polymorphic STRs described until now, DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and YCAII are those from which more data have been accumulated, as they are the most used in forensic genetics. With the help of collaborative efforts made in the construction of large databases (<http://www.ystr.org>; <http://www.ystr.org/usa>; <http://www.ystr.org/asia>), these markers are also the best studied, in what concerns amplification performance and specificity, multiplex amplification strategies, sequence structure and nomenclature, as well as world-wide allele frequency data.

More recently, additional Y-STRs have been described and are already in use in the forensic field: DYS434, DYS435, DYS436, DYS437, DYS438 and DYS439, reported by Ayub et al. [1] and; DYS460 (GATA A7.1), DYS461 (GATA A7.2), GATA A10, GATA C4 and GATA H4, reported by White et al. [2]. Although not in such widespread use in the forensic field as the previous ones, some sequencing, population and multiplex amplification data have started to be accumulated.

In this paper, we report the population data on eight Y-STR loci, DYS437, DYS438, DYS439, DYS460, DYS461, GATA A10, GATA C4 and GATA H4, which were included in the GEP-ISFG collaborative exercise [3]. These data were collected only from the participant laboratories that have reported correct results for all systems in all the distributed samples, which were asked to contribute with population(s) sample(s).

2. Material and methods

A total of 1020 male samples were typed for the loci: DYS437, DYS438, DYS439, DYS460, DYS461, GATA A10, GATA C4 and GATA H4. These 1020 samples included individuals from 13 different population groups from Argentina (50 Caucasians living in Buenos Aires and 27 Lara Amerindian population samples), Brazil (67 samples from Caucasians, residents in São Paulo), Costa Rica (102 Caucasians and 18 Africans), Macao (63 Asians born there), Mozambique (112 Africans residents in Maputo), Portugal (58 residents in Southern, 118 in Central and 208 in Northern regions) and Spain (51 from a general resident population, 54 residents in Cantabria and 92 born in Galicia).

These eight STR loci were amplified in two PCR tetraplex reactions (GEPY I and GEPY II), using the protocol available at the website <http://www.gep-isfg.org>.

Detection of the amplified products was carried out using either monochromatic (ALF-DNA sequencers, Amersham Pharmacia Biotech) or polychromatic platforms (ABI sequencers, Applied Biosystems).

Allele nomenclature was according to Gusmão et al. [4].

2.1. Statistical analysis

Allele frequencies were estimated by gene counting. Gene and haplotype diversities were calculated according to Nei [5]. Analysis of molecular variance (AMOVA) was performed by a Markov test using the Arlequin software version 2.000 [6]. Genetic distance matrices between populations were obtained by using the pairwise difference genetic distance. Unweighted pair group method with arithmetic means (UPGMA) tree was built from the distance matrix using the option Neighbour and Drawtree in the PHYLIP software package [7]. The tree was visualised with the Treeview software [8].

3. Results and discussion

In order to evaluate the informative content of the eight STR loci in different populations, the groups that, in a previous quality control exercise reported correct results in all the systems, were asked to type a small sized population sample. QC approval was essential to ensure the genotyping accuracy and the use of a standard nomenclature.

Results on the haplotype distribution observed in the 13 samples ($N = 1020$) are available through the website <http://www.gep-isfg.org>. A total of 648 different haplotypes could be defined, 496 being unique (found just once in one sample).

3.1. Haplotype sharing

From the 91 haplotypes shared by different populations (Table 1), the most frequent haplotype is shared by 30 males and is present in all the Caucasian population samples (SH72 in Table 1) and in Costa Rica Africans, but it was not found in the Lara, Macao and Mozambique samples. This haplotype is the most frequent in Spanish general population

Table 1

List of shared haplotypes (DYS460, DYS461, GATA A10, GATA C4, GATA H4, DYS437, DYS438 and DYS439)

	CP	SR	CA	CR	BR	SP	BA	LR	SD	NP	MC	MZ	GL	Total
Sample size	118	51	18	102	67	58	50	27	54	208	63	112	92	1021
No. of shared haplotypes	58	32	4	49	36	25	22	8	22	92	3	8	46	405
Shared haplotypes (%)	49.15	62.75	22.22	48.04	53.73	43.10	44.00	29.63	40.74	44.23	4.76	7.14	50.00	39.67
List of shared haplotypes														
SH1	10, 11, 14, 23, 28, 15, 12, 12	1	2							1				4
SH2	10, 11, 14, 25, 26, 16, 10, 11	1								1				2
SH3	10, 11, 15, 21, 28, 15, 10, 11			1	1									2
SH4	10, 11, 15, 23, 28, 15, 12, 11		1										1	2
SH5	10, 11, 15, 23, 28, 15, 12, 12		1		1									2
SH6	10, 11, 16, 21, 27, 14, 9, 11						1	1						2
SH7	10, 12, 13, 22, 27, 14, 11, 12		1									1		2
SH8	10, 12, 14, 21, 26, 14, 10, 13				1									1
SH9	10, 12, 14, 21, 27, 16, 10, 12			1		1				1				3
SH10	10, 12, 14, 22, 27, 14, 10, 12					1					1	1		3
SH11	10, 12, 14, 22, 27, 14, 11, 12		1									3		4
SH12	10, 12, 14, 22, 27, 16, 10, 11					1	1			1				3
SH13	10, 12, 14, 23, 28, 15, 12, 11	1								1				2
SH14	10, 12, 14, 23, 28, 15, 12, 12				1					3				2
SH15	10, 12, 14, 23, 28, 15, 12, 13									1				1
SH16	10, 12, 15, 23, 27, 14, 12, 11	1	1											2
SH17	10, 12, 15, 23, 27, 14, 12, 12		1	1	3				1	2				8
SH18	10, 12, 15, 23, 27, 14, 12, 13	1		1	1									3
SH19	10, 12, 15, 23, 27, 15, 12, 11	1			1					2				2
SH20	10, 12, 15, 23, 27, 15, 12, 12			1	1		1			2				5
SH21	10, 12, 15, 23, 28, 14, 12, 12			2		1				2				5
SH22	10, 12, 15, 23, 28, 14, 12, 13						1		1					2
SH23	10, 12, 15, 23, 28, 15, 10, 12						1			1				2
SH24	10, 12, 15, 23, 28, 15, 12, 11	2		2	1	1				3			1	10
SH25	10, 12, 15, 23, 28, 15, 12, 12	6		1	2	3	1			7				2
SH26	10, 12, 15, 23, 28, 15, 12, 13					2			1					3
SH27	10, 12, 15, 23, 29, 15, 12, 12	2			1					1				1
SH28	10, 12, 15, 24, 27, 14, 12, 12	1							1					2
SH29	10, 12, 15, 24, 28, 15, 12, 11									1				1
SH30	10, 12, 16, 23, 28, 15, 12, 11	1								1				2
SH31	10, 12, 16, 23, 28, 15, 12, 12		1							2				3
SH32	10, 13, 14, 22, 28, 14, 11, 12		1									1		2
SH33	10, 13, 14, 22, 29, 14, 11, 12			1								1		2
SH34	10, 13, 14, 24, 28, 14, 9, 11						1			1				2
SH35	10, 13, 15, 21, 27, 16, 10, 11		1						1					2
SH36	10, 13, 15, 23, 27, 14, 12, 11	1					1							2
SH37	10, 13, 15, 23, 28, 15, 12, 11		1							1				2
SH38	10, 13, 15, 23, 28, 15, 12, 12			1						2				3
SH39	11, 10, 14, 21, 27, 15, 9, 11									1				1
SH40	11, 10, 14, 21, 27, 16, 9, 11						1			1				2
SH41	11, 10, 14, 21, 27, 16, 9, 12	1		2						1				4
SH42	11, 11, 13, 20, 29, 16, 11, 11			1						1				2
SH43	11, 11, 14, 23, 28, 15, 12, 12			1			1							2
SH44	11, 11, 15, 20, 27, 14, 10, 12	4			1									5
SH45	11, 11, 15, 23, 27, 15, 12, 12	1								1				2
SH46	11, 11, 15, 23, 28, 14, 12, 12		1				1							2
SH47	11, 11, 15, 23, 28, 15, 12, 11		1							1				2
SH48	11, 11, 15, 23, 28, 15, 12, 12	2				2	1		1				2	8
SH49	11, 11, 16, 23, 28, 14, 11, 10		1							1		1		3
SH50	11, 12, 14, 21, 27, 14, 10, 11									1	1			2
SH51	11, 12, 14, 22, 28, 14, 10, 10		1							1				2
SH52	11, 12, 14, 23, 27, 14, 12, 11				1		1	1		2				5
SH53	11, 12, 14, 23, 27, 15, 12, 12	1					1			1				3
SH54	11, 12, 14, 23, 28, 14, 12, 11	1								1				2
SH55	11, 12, 14, 23, 28, 14, 12, 12								1	1				2

Table 1 (Continued)

	CP	SR	CA	CR	BR	SP	BA	LR	SD	NP	MC	MZ	GL	Total
SH56	11, 12, 14, 23, 28, 15, 12, 11			1		1			1	1				4
SH57	11, 12, 14, 23, 28, 15, 12, 12	6	1	2	2	1			3				7	23
SH58	11, 12, 14, 23, 28, 15, 12, 13			1		1	1							3
SH59	11, 12, 14, 24, 27, 14, 12, 11		1					5						6
SH60	11, 12, 14, 24, 28, 15, 12, 12		1	1		1				2			1	6
SH61	11, 12, 15, 20, 27, 15, 10, 11	1								1			1	3
SH62	11, 12, 15, 23, 27, 14, 12, 11	1	1							2				4
SH63	11, 12, 15, 23, 27, 14, 12, 12	4	1		1				1	1			2	10
SH64	11, 12, 15, 23, 27, 15, 12, 11	1		3		1			1				1	7
SH65	11, 12, 15, 23, 27, 15, 12, 12	1	1		1					2			2	7
SH66	11, 12, 15, 23, 27, 15, 12, 13		1			1			1	1			2	6
SH67	11, 12, 15, 23, 27, 15, 13, 12		1	1										2
SH68	11, 12, 15, 23, 28, 14, 11, 12			1			1							2
SH69	11, 12, 15, 23, 28, 14, 12, 11	1		1						1				3
SH70	11, 12, 15, 23, 28, 14, 12, 12	1		1						1				3
SH71	11, 12, 15, 23, 28, 15, 12, 11	4	1	6	1	1	1	1	1	3			2	21
SH72	11, 12, 15, 23, 28, 15, 12, 12	1	3	1	2	4	2	4	2	8			3	30
SH73	11, 12, 15, 23, 28, 15, 12, 13		2	5	2				2	2			2	15
SH74	11, 12, 15, 23, 29, 15, 12, 12	1		1	2					2				6
SH75	11, 12, 15, 24, 27, 15, 12, 13		1							1				2
SH76	11, 12, 15, 24, 28, 15, 12, 12									2			1	3
SH77	11, 12, 16, 23, 28, 15, 12, 11			1						1				2
SH78	11, 12, 16, 23, 28, 15, 12, 12	2	1	1		1				2				7
SH79	11, 12, 16, 23, 28, 15, 12, 13				1					1				2
SH80	11, 12, 16, 23, 28, 15, 13, 12	1			1									2
SH81	11, 13, 14, 21, 27, 14, 10, 11				1					1				2
SH82	11, 13, 14, 21, 28, 14, 10, 10	2			3	1	1		1	4			2	14
SH83	11, 13, 14, 21, 29, 14, 10, 10					1				1				2
SH84	11, 13, 14, 22, 28, 14, 10, 10	1											2	3
SH85	11, 13, 14, 23, 28, 15, 12, 12								1				1	2
SH86	11, 13, 15, 23, 28, 15, 12, 11		1										2	3
SH87	11, 13, 15, 23, 28, 15, 12, 12			3			1							4
SH88	11, 13, 15, 23, 29, 15, 12, 12	1								1				2
SH89	11, 14, 14, 21, 28, 14, 10, 10			1						1				2
SH90	12, 12, 15, 24, 28, 15, 12, 11			1	1									2
SH91	9, 11, 15, 21, 27, 14, 10, 11		1								1			2

CP: Central Portugal; SR: Spain, general population; CA: Costa Rica Africans; CR: Costa Rica Caucasians; BR: Brazil, São Paulo; SP: South Portugal; BA: Argentina, Buenos Aires; LR: Argentina, Lara; SD: Spain, Santander; NP: North Portugal; MC: Macao; MZ: Mozambique; GL: Spain, Galicia.

sample, São Paulo, Buenos Aires and North Portugal. Contrarily to what is found for haplotypes defined by the YHRD markers (<http://www.ystr.org>), the most frequent haplotype is not always the same in Iberian rooted samples (in Central Portugal, haplotypes SH25 and SH57; Costa Rica Caucasians, SH71; South Portugal, SH25; Cantabria and Galicia, SH57). However, all these haplotypes can be derived from SH72 by just one repeat unit and, for this reason, they can represent the same ancestral lineage.

Another haplotype (SH82), shared by seven populations, and nine repeats apart from SH72, is the second best represented lineage in Western Peninsula, and SH83, SH84 and SH89 could be derived from SH82 by one repeat unit.

The only haplotype that appears twice in the Costa Rica Africans is not present in any other population sample, as well as the most frequent haplotypes in Macao and Mozam-

bique (data available at <http://www.gep-isfg.org>). From the four haplotypes shared by Costa Rica Africans, two are shared with Mozambique, and the other two with the most frequent Caucasian haplotypes (both absent in Mozambique). Macao shares one haplotype with South Portugal and Mozambique, one with the Spanish general population and another with North Portugal. Mozambique, apart from the haplotypes shared with Costa Rica Africans and Macao, shares one haplotype with Costa Rica Caucasians, one with North Portugal and Spanish general population and another with Spanish general population.

The most frequent haplotype in Lara sample (SH59) is also present in a sample from Spanish general population. Another shared haplotype is the most frequent in Costa Rica Caucasians, sharing also one haplotype with Buenos Aires and another with Buenos Aires, São Paulo and North Portugal.

Table 2
Allele and gene frequencies

	CP	SR	CA	CR	BR	SP	BA	LR	SD	NP	MC	MZ	GL
DYS460													
6				0.0098									
9	0.0169	0.0588		0.0098	0.0448	0.0345	0.0200		0.0370	0.0433	0.3333	0.0357	0.0652
10	0.3729	0.3137	0.2778	0.3627	0.4478	0.5000	0.3800	0.4444	0.2963	0.4327	0.4444	0.4554	0.3152
11	0.5847	0.6275	0.6667	0.5588	0.4328	0.3966	0.5600	0.5185	0.6481	0.4952	0.1905	0.4911	0.5978
12	0.0254		0.0556	0.0588	0.0746	0.0690	0.0400	0.0370	0.0185	0.0288	0.0317	0.0089	0.0109
13												0.0089	0.0109
Gene diversity	52.25	51.45	50.33	55.79	61.37	59.71	55.10	55.27	49.97	56.76	66.46	55.50	54.47
DYS461													
9									0.0185				
10	0.0593	0.0196		0.0196		0.0345	0.0400			0.0337	0.0159	0.0089	0.0217
11	0.1864	0.2745	0.1667	0.1373	0.1194	0.1207	0.2400	0.0741	0.1852	0.2019	0.2857	0.0089	0.2065
12	0.6102	0.5490	0.6111	0.6275	0.7313	0.5862	0.6200	0.7407	0.5926	0.6106	0.5714	0.3304	0.6196
13	0.1356	0.1373	0.2222	0.1863	0.1343	0.2069	0.0800	0.1852	0.1852	0.1298	0.1270	0.5179	0.1522
14	0.0085	0.0196		0.0294	0.0149	0.0345	0.0200		0.0185	0.0240		0.1161	
15						0.0172						0.0179	
Gene diversity	57.58	61.57	58.17	55.70	43.92	60.68	56.08	42.74	59.05	57.06	58.47	61.42	55.59
GATA A10													
12		0.0196						0.0370				0.0179	
13	0.0254		0.0556	0.0588		0.0172	0.0400		0.0370	0.0433	0.0159	0.3393	0.0217
14	0.2712	0.2157	0.5556	0.2451	0.3134	0.3276	0.3400	0.4074	0.4074	0.2885	0.3968	0.4732	0.3804
15	0.5593	0.6667	0.2222	0.6275	0.6119	0.5345	0.4600	0.3333	0.4630	0.5192	0.2857	0.1339	0.4783
16	0.1017	0.0980	0.0556	0.0686	0.0746	0.0862	0.1400	0.2222	0.0926	0.1298	0.2222	0.0357	0.0978
17	0.0424		0.1111				0.0200			0.0144	0.0794		0.0217
18						0.0345				0.0048			
Gene diversity	60.60	50.90	66.01	54.34	52.96	58.05	66.45	69.80	62.12	63.13	71.63	64.72	62.28
GATA C4													
17												0.1339	
19	0.0339		0.0556	0.0098						0.0144	0.0794		0.0109
20	0.0847	0.0196		0.0392	0.0448	0.0172			0.0185	0.0385	0.2698	0.0357	0.0435
21	0.1525	0.2157	0.1111	0.1667	0.2090	0.2069	0.2400	0.1111	0.1481	0.1923	0.3175	0.4911	0.2174
22	0.1102	0.0980	0.2222	0.0980	0.0896	0.1552	0.1400	0.2593	0.0926	0.0577	0.1587	0.2232	0.0978
23	0.5169	0.5294	0.4444	0.5490	0.5970	0.5517	0.5600	0.3704	0.6111	0.5385	0.1111	0.0446	0.5435
24	0.0678	0.1373	0.1667	0.1176	0.0299	0.0690	0.0600	0.2593	0.1296	0.1154	0.0635	0.0536	0.0761
25	0.0339			0.0196	0.0299					0.0337		0.0089	0.0109
26										0.0096		0.0089	
Gene diversity	68.91	65.73	75.16	65.17	59.70	63.46	61.80	74.36	58.98	65.67	79.11	69.10	64.69
GATA H4													
24		0.0196											
25	0.0085			0.0196		0.0172							
26	0.0339	0.0196		0.0196	0.0299	0.0345			0.0556	0.0144	0.0635		0.0870
27	0.3814	0.3726	0.3889	0.3431	0.3284	0.3276	0.4000	0.5556	0.3333	0.3413	0.4603	0.5982	0.2826
28	0.5169	0.5294	0.5556	0.5196	0.5821	0.5345	0.5200	0.2963	0.5185	0.5433	0.4127	0.2946	0.6087
29	0.0593	0.0588	0.0556	0.0980	0.0597	0.0862	0.0800	0.1481	0.0926	0.0913	0.0635	0.1071	0.0217
30										0.0096			
Gene diversity	58.76	58.82	56.86	60.78	55.72	60.86	57.47	60.40	61.98	58.25	61.96	54.87	54.75
DYS437													
13											0.0159		
14	0.3729	0.3726	0.6111	0.3431	0.4328	0.3103	0.3200	0.7037	0.4259	0.3125	0.4127	0.9732	0.3478
15	0.4831	0.5686	0.3333	0.5588	0.4925	0.5862	0.4800	0.1481	0.3889	0.5625	0.3968	0.0268	0.5760
16	0.1441	0.0588	0.0556	0.0980	0.0746	0.1034	0.2000	0.1481	0.1852	0.1250	0.1746		0.0761
Gene diversity	61.21	54.51	54.25	56.59	57.30	55.90	64.00	47.86	64.50	57.31	65.18	5.26	54.73

Table 2 (Continued)

	CP	SR	CA	CR	BR	SP	BA	LR	SD	NP	MC	MZ	GL
DYS438													
7				0.0098									
8		0.0196				0.0345							0.0109
9	0.0847	0.0981		0.0392	0.0597	0.1724	0.0800	0.1481	0.0556	0.1202	0.0159	0.0268	0.1087
10	0.1949	0.2549	0.2222	0.2353	0.3284	0.2414	0.3200	0.0370	0.2778	0.2404	0.7619	0.1964	0.2717
11	0.1102	0.0196	0.5556	0.1275	0.0448	0.0517	0.1400	0.3704	0.1111	0.0385	0.2222	0.7500	0.0435
12	0.5678	0.5490	0.2222	0.5784	0.5373	0.4655	0.4600	0.4444	0.5370	0.5721		0.0268	0.5543
13	0.0424	0.0588		0.0098	0.0299	0.0345			0.0185	0.0288			0.0109
Gene diversity	62.38	63.22	62.75	59.79	60.61	70.24	67.35	66.67	63.03	60.10	37.58	40.11	61.16
DYS439													
9	0.0169					0.0172				0.0096			
10	0.0593	0.0980	0.0556	0.0294	0.0448	0.0517	0.1000		0.1667	0.1010	0.0159	0.0179	0.0761
11	0.2966	0.3137	0.3889	0.3235	0.3284	0.3276	0.2800	0.5926	0.2222	0.3798	0.2540	0.3214	0.3370
12	0.5424	0.3922	0.4444	0.4314	0.4179	0.4310	0.4400	0.1481	0.4630	0.3942	0.5556	0.5268	0.4239
13	0.0678	0.1569	0.1111	0.1863	0.1940	0.1724	0.1600	0.1111	0.1111	0.1106	0.1587	0.1071	0.1413
14	0.0085	0.0392		0.0294	0.0149			0.1111	0.0370	0.0048	0.0159	0.0268	0.0109
15	0.0085						0.0200	0.0370					0.0109
Gene diversity	61.45	72.63	67.32	67.95	68.79	68.60	70.61	62.39	70.79	68.11	61.09	61.21	68.82

CP: Central Portugal; SR: Spain, general population; CA: Costa Rica Africans; CR: Costa Rica Caucasians; BR: Brazil, São Paulo; SP: South Portugal; BA: Argentina, Buenos Aires; LR: Argentina, Lara; SD: Spain, Santander; NP: North Portugal; MC: Macao; MZ: Mozambique; GL: Spain, Galicia.

3.2. Single locus population analysis

The allele frequencies of these eight Y-STRs in all samples are displayed in Table 2, together with the gene diversities.

Three new alleles, not yet previously described in the literature, were found: DYS460 allele 6, in Costa Rica Caucasians; DYS461 allele 9, in Santander and; GATA H4 allele 24, in the Spanish Residents sample.

Except for non-Caucasian samples, the allele frequency distributions and gene diversities do not differ significantly across the different population groups, and gene diversity across loci varies from an average value of 0.5528 in DYS460 to 0.6859 in DYS439. As previously observed for DYS392, DYS438 presents a bimodal allele frequency distribution in European origin populations only [9–12].

In the comparisons with any of the four non-Caucasian samples, significant differences were always found, including those between the two African origin samples from Mozambique and Costa Rica.

3.3. Haplotype population analysis

The haplotype diversity values for each of the 13 population samples are shown in Table 3. The lowest haplotype diversity was found in the Lara (Argentina) population (95.44%) and the highest (99.90%) in Macao (China). However, the lowest average gene diversity (51.51%) was found not in Lara but in the Mozambique sample, also presenting low haplotype diversity (97.97%). This is in accordance with previous findings, showing a lower Y chromosome haplotype diversity in African populations

Table 3
Haplotype and average gene diversity values

	CP	SR	CA	CR	BR	SP	BA	LR	SD	NP	MC	MZ	GL
Sample size	118	51	18	102	67	58	50	27	54	208	63	112	92
No. of haplotypes	92	47	17	82	53	53	47	18	50	166	61	73	72
Average gene diversity	60.39	59.85	61.36	59.51	57.55	62.19	62.37	59.94	61.30	60.80	62.69	51.51	59.56
Haplotype diversity	99.20	99.61	99.35	99.26	99.14	99.64	99.51	95.44	99.65	99.61	99.90	97.97	99.14

CP: Central Portugal; SR: Spain, general population; CA: Costa Rica Africans; CR: Costa Rica Caucasians; BR: Brazil, São Paulo; SP: South Portugal; BA: Argentina, Buenos Aires; LR: Argentina, Lara; SD: Spain, Santander; NP: North Portugal; MC: Macao; MZ: Mozambique; GL: Spain, Galicia.

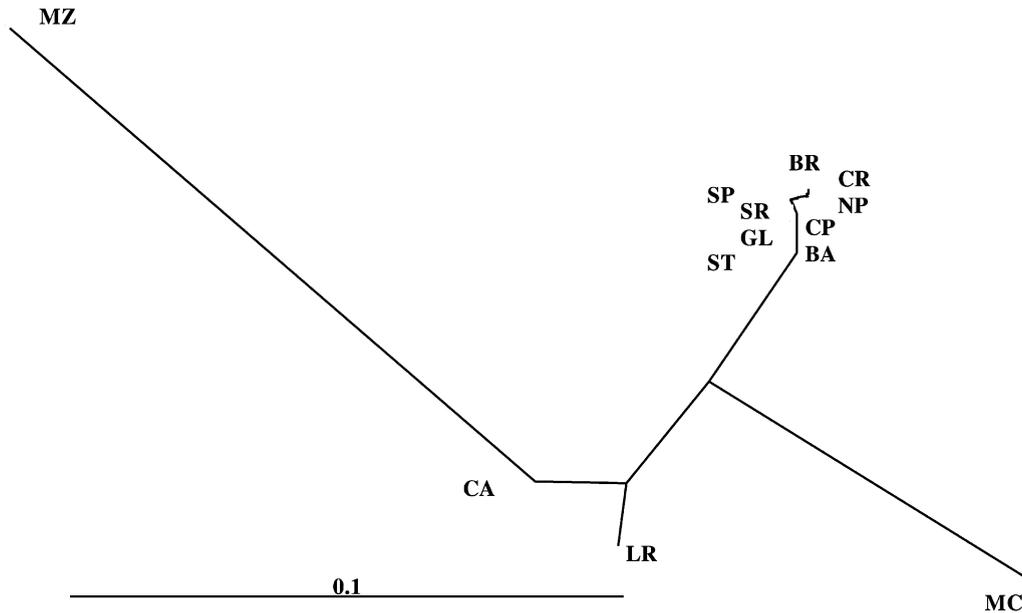


Fig. 1. Unrooted NJ tree based on pairwise Φ_{st} distances.

[12,13], although Kayser et al. [14] report higher single locus diversity.

The high gene and haplotype diversity values observed in Costa Rica Africans, can be explained by a Caucasian admixture, supported by the analysis of the four shared haplotypes, where two of these are the most frequent in the Caucasian samples studied in the present work.

The highest average gene diversity is also observed in the Macao sample, in the same way as it can be found for the average gene diversity of the “minimal haplotype” Y-STRs (<http://www.ystr.org>; <http://www.ystr.org/usa>; <http://www.ystr.org/asia>).

Fig. 1 shows a neighbour joining tree constructed from the 13 populations under study. Pairwise Φ_{st} values between the Iberian rooted samples were all below 1%, with the highest values observed when comparing the Cantabria sample with Brazil and North Portugal (0.6%) and with South Portugal (0.8%). The genetic distance analysis between Caucasian and non-Caucasian samples, shows that Mozambique displays the highest values (between 0.20 and 0.23), followed by Macao (0.08 and 0.12), Lara Amerindians (0.06 and 0.08) and Costa Rica Africans (0.02 and 0.06).

Among the non-Caucasian samples, the lowest genetic distance was found between Lara and Costa Rica Africans (0.02), and the highest between Mozambique and Macao. The pairwise Φ_{st} value between Mozambique and Costa Rica Africans was about one half of those between Mozambique and Caucasian samples.

In summary, pairwise haplotype analysis showed the relative homogeneity of samples of Iberian origin, in agreement with what was previously found in the European populations for other Y-STR haplotypes (<http://www.ystr.org>)

while, as expected, the four non-Caucasian samples, Macao (Chinese), Mozambique (Africans), Costa Rica (Africans) and Argentina (Lara, Amerindians), displayed highly significant Φ_{st} values in the pairwise comparisons with all the Caucasian samples.

References

- [1] Q. Ayub, A. Mohyuddin, R. Qamar, K. Mazhar, T. Zerjal, S. Mehdi, C. Tyler-Smith, Identification and characterization of novel human Y-chromosomal microsatellites from sequence database information, *Nucleic Acids Res.* 28 (2000) e8.
- [2] P. White, O. Tatum, L. Deaven, J. Longmire, New, male-specific microsatellite markers from the human Y chromosome, *Genomics* 57 (1999) 433–437.
- [3] P. Sánchez-Diz, et al., Results of the GEP-ISFG collaborative study on two Y-STRs tetraplexes: GEPY I (DYS461, GATA C4, DYS437 and DYS438 and GEPY II and GEPY II, DYS460, GATA A10, GATA H4, DYS439), *Forensic Sci. Int.*, in press.
- [4] L. Gusmão, A. González-Neira, C. Alves, M. Lareu, S. Costa, A. Amorim, A. Carracedo, Chimpanzee homologous of human Y specific STRs. A comparative study and a proposal for nomenclature, *Forensic Sci. Int.* 126 (2002) 129–136.
- [5] M. Nei, *Molecular Evolutionary Genetics*, Columbia University Press, New York, NY, USA, 1987.
- [6] S. Schneider, D. Roessli, L. Excoffier, Arlequin version 2.000, A Software for Population Genetics Data Analysis, University of Geneva, Geneva, 2000.
- [7] J. Felsenstein, PHYLIP: phylogeny inference package (version 3.2), *Cladistics* 5 (1989) 164–166.
- [8] R.D.M. Page, Treeview: an application to display phylogenetic trees on personal computers, *Comput. Appl. Biosci.* 12 (1996) 357–358.

- [9] M.G. Thomas, T. Parfitt, D.A. Weiss, K. Skorecki, J.F. Wilson, M. le Roux, N. Bradman, D.B. Goldstein, Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba—the “Black Jews of Southern Africa”, *Am. J. Hum. Genet.* 66 (2000) 674–686.
- [10] P. Forster, A. Rohl, P. Lunnemann, C. Brinkmann, T. Zerjal, C. Tyler-Smith, B. Brinkmann, A short tandem repeat-based phylogeny for the human Y chromosome, *Am. J. Hum. Genet.* 67 (2000) 182–196.
- [11] A. Nebel, D. Filon, C. Hohoff, M. Faerman, B. Brinkmann, A. Oppenheim, Haplogroup-specific deviation from the stepwise mutation model at the microsatellite loci DYS388 and DYS392, *Eur. J. Hum. Genet.* 9 (2001) 22–26.
- [12] L. Gusmão, C. Alves, A. Amorim, Molecular characterization of four Y-specific microsatellites (DYS434, DYS437, DYS438, DYS439) for population and forensic studies, *Ann. Hum. Genet.* 65 (2001) 285–291.
- [13] J.K. Pritchard, M.T. Seielstad, A. Perez-Lezaun, M.W. Feldman, Population growth of human Y chromosomes: a study of Y chromosome microsatellites, *Mol. Biol. E* 16 (1999) 1791–1798.
- [14] M. Kayser, M. Krawczak, L. Excoffier, P. Dieltjes, D. Corach, V. Pascali, C. Gehrig, L.F. Bernini, J. Jespersen, E. Bakker, L. Roewer, P. de Knijff, An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations, *Am. J. Hum. Genet.* 68 (2001) 990–1018.