

12-1-2011

# The Y-STR genetic diversity of an Idaho Basque population, with comparison to European Basques and US Caucasians

Josu Zubizarreta

*Department of Biological Sciences, Boise State University, Boise, ID 83725, USA*

Michael C. Davis

*Department of Biological Sciences, Boise State University, Boise, ID 83725, USA*

Greg Hampikian

*Department of Biological Sciences, Boise State University, Boise, ID 83725, USA, [greghampikian@boisestate.edu](mailto:greghampikian@boisestate.edu)*

---

## Recommended Citation

Open access pre-print, subsequently published as Zubizarreta, Josu; Davis, Michael C.; and Hampikian, Greg (2011) "The Y-STR genetic diversity of an Idaho Basque population, with comparison to European Basques and US Caucasians," *Human Biology*: Vol. 83: Iss. 6, Article 7.

Available at: [http://digitalcommons.wayne.edu/humbiol\\_preprints/16](http://digitalcommons.wayne.edu/humbiol_preprints/16)

The Y-STR genetic diversity of an Idaho Basque population, with comparison to European Basques and US Caucasians

Authors: Josu Zubizarreta, Michael C. Davis, and Greg Hampikian

Author's Affiliations: Department of Biological Sciences, Boise State University, Boise, ID 83725, USA.

First Author: Josu Zubizarreta

Corresponding author: Greg Hampikian

Email: [greghampikian@boisestate.edu](mailto:greghampikian@boisestate.edu)

Phone: 208-426-4992 (office), 208-781-0438 (cell)

fax: 208-426-1040

Abbreviated title: Y-STR Diversity in Idaho Basques

Key words: Y-STR, Yfiler, Basques, Idaho, Immigrant

## Abstract

Fifty unrelated Basque males from southwest Idaho were typed for the 17 Y-STR loci in the Yfiler multiplex kit (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, YGATA\_H4.1 and DYS385a/b). A total of 42 haplotypes were identified, with no more than two individuals sharing a single haplotype. The haplotype diversity (HD) was 0.9935, and gene diversity (D) over loci was  $0.457 \pm 0.137$ . The Idaho Basque population was compared to the source population from the Basque autonomous region of Northern Spain and Southern France, as well as a US Caucasian population. The haplotype diversity for the immigrant Basque sample is within 0.4% of the haplotype diversity of the European Basques (0.9903); thus the power of discrimination is similar for each population. The Idaho Basque population has less diversity in 9 out of 16 loci (considering DYS385a/b together) and 3% less diversity across all loci, compared to the European Basque population. A multi-dimensional scaling analysis (MDS) was created using pairwise  $R_{ST}$  values to compare the Idaho Basques to other populations. Based upon  $R_{ST}$  and  $F_{ST}$  measures, no significant differentiation was found between the Idaho and source European Basque population.

The individuals typed for this study are from the community of immigrant Basques, descendants of the Basque migrants who settled in the region of the Intermountain West starting in the late 19<sup>th</sup> and early 20<sup>th</sup> centuries. The Basques of Northern Spain and Southern France

have been characterized as genetically distinct from other European populations, and this is the case for certain markers such as blood types (Mourant 1947). However, studies using the type of DNA analysis common in forensic applications (STRs and mitochondrial DNA) have shown that Basque genetic variation is within the range reported for other European populations, and that the common alleles or haplotypes in Europe are also found in Basque populations (Bertranpetit et al. 1995; Zlojutro et al. 2006; Alfonso-Sánchez et al. 2008; Davis et al. 2011). The Basques can be considered outliers only in the context of the European gene pool, and are more related to neighboring Europeans than to other world populations (Rosenberg et al. 2002; Li et al. 2008; Laayouni et al. 2010). Since the 19<sup>th</sup> century, Idaho has been the destination for a substantial number of Basque immigrants (Douglass and Bilbao 1975). Many young Basque males during the late 1800's and early 1900's came to Idaho in search of work, and found it in the ranching and shepherding industry (Bieter and Bieter 2000). With one of the largest Basque populations outside of the native Basque region, it is estimated that Idaho is home to approximately 20,000 residents claiming Basque heritage (Zubiri 2006). A recent study by Davis et. al (2011) was the first to compare mitochondrial DNA profiles between native European Basques and those of the immigrant Basque population of SW Idaho and SE Oregon. The data presented here compares the Y-STR profiles of an immigrant Basque population with Basque males in the Basque Country, and includes a comparison with a US Caucasian population.

## **Materials and Methods**

**Sampling.** Samples were collected at various Basque cultural events held in Boise, Idaho (Lat: 43.6° N, Lon: 116.3 ° W). Individuals recruited were self-described as having paternal Basque heritage. Questionnaires and cotton buccal swabs were collected from 50 unrelated males. An informed consent approval by the Institutional Review Board of Boise State University was

obtained prior to data collection. The European Basque population data (N = 168) was taken from Garcia et al. 2006 (García et al. 2006), and the US Caucasian data set is from a study published on the NIST STRBase web site of 260 males, typed for Yfiler and additional loci (<http://www.cstl.nist.gov/strbase/NISTpop.htm>). Only data from the Yfiler loci were used for the purposes of this study.

**DNA Extraction.** DNA was extracted from cotton buccal swab samples using the QiaAMP DNA Mini kit (Qiagen). The genomic DNA for individual samples was quantified using the Quantifiler Duo RT-PCR kit (Applied Biosystems), and the DNA levels were adjusted to reach optimal values (0.5-1.0 ng) for microsatellite PCR.

**PCR Amplification.** PCR amplification of 17 Y-STR loci was performed using the AMPFISTR® Yfiler kit (Applied Biosystems) according to the manufacturer's protocol.

**Electrophoresis and typing.** The samples were genotyped using a 3130® Genetic Analyzer (Applied Biosystems). Alleles were typed according to published guidelines for STR analysis (Gusmão et al. 2006; Mulero et al. 2006).

**Data analysis.** Haplotype diversity (HD), Allelic (Gene) diversity (D) were calculated according to the formulas in Nei (Nei 1987), using ARLEQUIN version 3.5 (Excoffier and Lischer 2010). The power of discrimination (PD) was calculated as the number of different haplotypes divided by the sample size (García et al. 2006). An analysis of molecular variance (AMOVA) with 10,000 permutations using ARLEQUIN was performed to generate  $R_{ST}$  and  $F_{ST}$  values and their associated probabilities. From the pair-wise  $R_{ST}$  values, a Multidimensional Scaling (MDS) plot was generated using SAS software, Version 9 for Windows (copyright 2009, SAS Institute Inc.). SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA. The following populations

were used for comparison in the AMOVA and MDS plot: Idaho Basque (this study), European Basque (García et al. 2006), Andalusia (Gaibar et al. 2010), Barcelona (Sánchez et al. 2007), N. Portugal (Pontes et al. 2007), Italy (Turrina et al. 2006), N. Greece (Leda et al. 2009), Serbia (Veselinovic et al. 2008), Romania (Stanciu et al. 2010), and Algeria (Robino et al. 2008).

**Quality control.** A proficiency testing quality control check was performed in conjunction with submission to the Y-hrd.org database (Gusmão et al. 2006; Willuweit and Roewer 2007).

## **Results**

The Idaho Basque population allele frequency and diversity statistics for 17 loci are listed in Table 1. The Idaho Basque samples have less allelic diversity over all loci than the source Basque (0.462 versus 0.477, Table 2). Both Basque populations show less diversity than the US Caucasian samples (0.622). This is consistent with previous reports showing that Basques have lower diversity values than other European populations for both autosomal and Y-chromosome markers (Alonso et al. 2005; Zlojutro et al. 2006). Table 3 summarizes genetic diversity indices for all three populations.

Table 1. Allele frequencies and allelic diversity for 17 Y-STR loci in Idaho Basque, DYS385a/b considered as one locus.

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4	Allele	DYS385
8																11,11	0.08
9					0.04			0.02								11,13	0.02
10					0.26			0.08							0.02	11,14	0.62
11					0.66	0.12		0.02	0.22						0.28	11,15	0.04
12		0.06			0.04		0.08	0.86	0.62						0.58	12,12	0.02
13	0.04	0.66				0.86	0.74	0.02	0.12			0.02			0.12	12,14	0.06
14	0.88	0.28				0.02	0.16		0.04	0.38		0.02	0.04			13,15	0.02
15	0.04						0.02			0.60		0.44	0.12			13,16	0.02
16	0.04									0.02		0.46	0.10			14,14	0.08
17												0.06	0.44			16,18	0.02
18											0.26		0.28			17,18	0.02
19											0.62		0.02				
20											0.04						
21				0.02							0.06			0.04			
22				0.06							0.02			0.08			
23				0.18										0.86			
24				0.64										0.02			
25				0.10													
26																	
27																	
28			0.02														
29			0.70														
30			0.26														
31																	
32																	
33			0.02														
<b>D</b>	0.221	0.482	0.442	0.544	0.494	0.246	0.42	0.253	0.551	0.495	0.542	0.59	0.702	0.252	0.57		0.59

Table 2. Allelic diversity within and across loci for the populations in the present study.

Y-STR	Idaho Basque	Source Basque	US Caucasian
DYS19	0.221	0.307	0.509
DYS389I	0.482	0.617	0.528
DYS389II	0.442	0.674	0.679
DYS390	0.544	0.425	0.703
DYS391	0.494	0.457	0.557
DYS392	0.246	0.253	0.592
DYS393	0.420	0.238	0.366
DYS438	0.253	0.233	0.607
DYS439	0.551	0.586	0.633
DYS437	0.495	0.495	0.579
DYS448	0.542	0.557	0.600
DYS456	0.590	0.616	0.735
DYS458	0.702	0.654	0.753
DYS635	0.252	0.327	0.683
YGATA H4	0.570	0.575	0.574
DYS385a/b	0.590	0.617	0.855
mean	0.462	0.477	0.622

Table 3. Summary statistics for Y-STR diversity in the three populations compared.

Population	N	Haplotypes	HD	DC	D
Idaho Basque	50	42	0.9935	0.840	0.462
Source Basque	168	138	0.9903	0.821	0.477
US Caucasian	260	254	0.9998	0.977	0.622



The haplotype diversity for the Idaho Basque sample was 0.9935. This value is within 0.4% of the haplotype diversity of the European Basques (0.9903). There is no significant difference in average gene diversity over all loci between the Idaho and Source Basques (t test,  $p < 0.05$ ).

In our sample of 50 Idaho Basque males, we found 42 haplotypes. In Table 4 we show the frequency of each of our 42 haplotypes in the YHRD (N = 30,300) and ABI Y-STR (N = 11,393) databases. Of the 42 haplotypes in our study, 22 did not have a match in YHRD, and 33 did not have a match in the ABI Y-STR database. Nine haplotypes were shared between Idaho and Source Basque populations. Five haplotypes were found in both Idaho Basque and US Caucasian populations, and four haplotypes were found in both Source Basque and US Caucasian populations.

Table 4. Hits of Idaho Basque haplotypes in the YHRD and ABI databases.

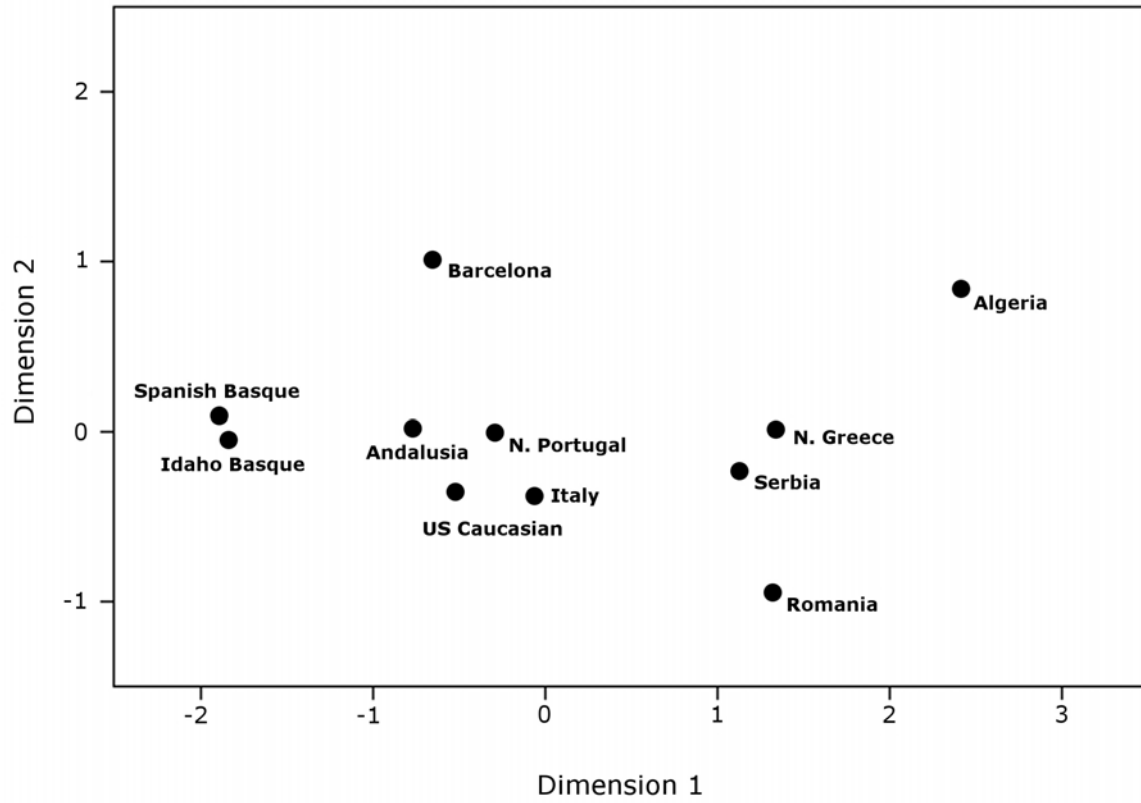
Haplotype ID	YHRD		ABI	
	hits	freq	hits	freq
H1	3	9.90E-05	2	2.00E-04
H2	4	1.32E-04	6	5.00E-04
H3	.	.	.	.
H4	.	.	.	.
H5	.	.	.	.
H6	2	6.60E-05		
H7	.	.	.	.
H8	.	.	.	.
H9	2	6.60E-05	.	.
H10	1	3.30E-05	.	.
H11	.	.	.	.
H12	.	.	.	.
H13	1	3.30E-05	1	1.00E-04
H14	.	.	.	.
H15	2	6.60E-05	.	
H16	.	.	.	.
H17	1	3.30E-05	.	.
H18	.	.	.	.
H19	.	.	.	.
H20	3	9.90E-05	3	3.00E-04
H21	.	.	.	.
H22	.	.	.	.
H23	.	.	.	.
H24	.	.	.	.
H25	9	2.97E-04	4	4.00E-04
H26	3	9.90E-05	.	.
H27	.	.	.	.
H28	.	.	.	.
H29	15	4.95E-05	16	1.40E-03
H30	.	.	.	
H31	5	1.65E-04	1	1.00E-04
H32	.	.	.	.
H33	.	.	1	1.00E-04
H34	.	.	.	.
H35	2	6.60E-05	.	.
H36	29	9.57E-04	11	1.00E-03
H37	1	3.30E-05	.	.
H38	1	3.30E-05	.	.
H39	2	6.60E-05	.	.
H40	.	.	.	.
H41	.	.	.	.
H42	1	3.30E-05	.	.

Considering only the minimal haplotype (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393), our data set of 50 individuals was reduced to 31 haplotypes, and the haplotype diversity is  $0.9445 \pm 0.0234$ , with an average allelic diversity of  $0.4155 \pm 0.2366$ . The most common Idaho Basque haplotype considering just the minimal haplotype loci was 14, 13, 29, 24, 11, 13, 13, 11, 14. That minimal haplotype was shared by 11 of the 50 males in our study. This minimal haplotype is also the most common in YHRD, found in 1637 individuals (n=30,300), or roughly 2% of all individuals in the database.

## **Discussion**

The relative genetic distance between selected populations is shown by a multidimensional scaling plot (MDS) created using  $R_{ST}$  values calculated from Arlequin 3.5 (Figure 1). There was no significant difference between the Idaho and European Basque populations, based on the  $R_{ST}$  genetic distance ( $P < 0.01$ ).  $F_{ST}$  calculations based on haplotype frequencies also showed no significant differentiation (data not shown) between Idaho and European Basque.  $F_{ST}$  values are considered more relevant in forensic analyses since exact haplotype matches are considered for the calculations (Budowle et al. 2009). Both Basque populations were significantly differentiated from all other populations we compared.

Figure 1.



Small sample sizes can result in underestimates of the genetic diversity found within a population, and for haplotypic markers such as Y-STRs or mitochondrial DNA; even sample sizes in the hundreds can underestimate the number of unique haplotypes (Pereira et al. 2004). Our sample size is relatively small, and thus our results may underestimate the diversity found in the immigrant Basque population of Idaho. We detected relatively modest genetic differentiation between immigrant and native populations (i.e., we found no evidence for significant founder effects). Notably, we found no significant difference in average gene diversity over all loci between the Idaho and Source Basques (t test,  $p < 0.05$ ). Our sampling strategy may have helped in this regard, since we purposefully avoided sampling first-degree relatives. In addition, there was no significant differentiation (structure) detected between Idaho and Source Basque populations, based on the  $F_{ST}$  value ( $F_{ST} = 0.004$ ,  $p < 0.05$ ).

In terms of interpopulation comparisons, it was expedient to use a forensic typing kit which contains loci that have been commonly used in previous population studies. Forensic application of Y-STR loci data requires enough polymorphic loci to produce rare or singleton haplotypes. Increasing the number of Y-STR loci would provide more information about both the number of shared lineages, and how closely related the lineages are between populations. If additional loci are to be typed in future studies of Basque populations, loci should be chosen that are highly polymorphic for the population(s) being studied. For example, in a report of Finnish paternal lineages using 16 Y-STR loci, it was discovered that the 10 most polymorphic loci were sufficient to capture most of the diversity (Hedman et al. 2004). A follow-up study (Hedman et al. 2011) revealed that the addition of seven other highly polymorphic loci was able to give unique haplotypes for a set of 53 males who had appeared identical when typed with the 9 locus “minimal haplotype”. The Yfiler results from this study divided these 53 males into 16 haplotypes.

Future studies of the Basques of Idaho (and neighboring states) would be of interest from both an anthropological and population genetics point of view. Additional Y-STR markers should be incorporated (for example, the seven used in Hedman, et al. 2011), as well as phylogenetically informative SNPs, which would allow the immigrant lineages to be placed into the broader phylogeographic history of Y-chromosomes in Europe, and could shed light on both the recent and ancient origins of such lineages.

Many Basque cultural centers (Euskaldunak) exist in the Western United States (Douglass and Bilbao 1975; Bieter and Bieter 2000), and such centers are useful for contacting local communities of descendants of Basque immigrants for genetic studies (Davis et al. 2011). In addition, because many Basque retain their native language and heritage, and because Basque surnames are easily recognizable (because of their unique language), tracing Basque paternal heritage several generations is relatively straightforward.

Studies of immigrant populations like the one presented here are vital to the accurate estimation of haplotype frequencies, which are used for studies of human migration, as well as for determining the power of discrimination and probability of matches used in forensic, missing persons, and paternity cases. The Y-STR haplotypes reported here are available via the Y-HRD website for other researchers.

### **Acknowledgments**

The authors wish to thank: the Basque volunteers of Idaho for participating in this study; the Basque Center Incorporated (Euskaldunak), and the Basque Museum and Cultural Center (Boise, Idaho) for allowing us to collect samples on their sites. This project was supported in part by the US Department of Defense contract W81XWH-07-1-000.

## LITERATURE CITED

- Alfonso-Sánchez, M. A., S. Cardoso, C. Martínez-Bouzas, et al. (2008). "Mitochondrial DNA haplogroup diversity in Basques: A reassessment based on HVI and HVII polymorphisms." American Journal of Human Biology **20**(2): 154-164.
- Alonso, S., C. Flores, V. Cabrera, et al. (2005). "The place of the Basques in the European Y-chromosome diversity landscape." Eur J Hum Genet **13**(12): 1293-1302.
- Bertranpetit, J., J. Sala, F. Calafell, et al. (1995). "Human mitochondrial DNA variation and the origin of Basques." Annals of Human Genetics **59**(1): 63-81.
- Bieter, J. and M. Bieter (2000). An enduring legacy: the story of Basques in Idaho. Reno, NV, University of Nevada Press.
- Budowle, B., J. Ge, X. G. Aranda, et al. (2009). "Texas population substructure and its impact on estimating the rarity of Y STR haplotypes from DNA evidence\*." J Forensic Sci **54**(5): 1016-1021.
- Davis, M. C., S. J. Novak and G. Hampikian (2011). "Mitochondrial DNA analysis of an immigrant basque population: Loss of diversity due to founder effects." Am J Phys Anthropol **144**: 516-525.
- Douglass, W. A. and J. Bilbao (1975). Amerikanuak: Basques in the New World. Reno, Nevada, University of Nevada Press.
- Excoffier, L. and H. E. L. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." Molecular Ecology Resources **10**(3): 564-567.
- Gaibar, M., E. Esteban, P. Moral, et al. (2010). "STR genetic diversity in a Mediterranean population from the south of the Iberian Peninsula." Annals of Human Biology **37**(2): 254-267.
- García, O., I. Yurrebaso, I. Uriarte, et al. (2006). "Distribution of Y-chromosomal haplotypes in the Basque Country autochthonous population using a 17-locus multiplex PCR assay." International Congress Series **1288**: 319-321.
- Gusmão, L., J. M. Butler, A. Carracedo, et al. (2006). "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 Science International **157**(2-3): 187-197.
- Hedman, M., A. M. Neuvonen, A. Sajantila, et al. (2011). "Dissecting the Finnish male uniformity: the value of additional Y-STR loci." Forensic Sci Int Genet **5**(3): 199-201.
- Hedman, M., V. Pimenoff, M. Lukka, et al. (2004). "Analysis of 16 Y STR loci in the Finnish population reveals a local reduction in the diversity of male lineages." Forensic Sci Int **142**(1): 37-43.
- Laayouni, H., F. Calafell and J. Bertranpetit (2010). "A genome-wide survey does not show the genetic distinctiveness of Basques." Human Genetics **127**(4): 455-458.
- Leda, K., L. S. Jessica and A. I. Jodi (2009). "Population genetics of Y-chromosome STRs in a population of Northern Greeks." Forensic science international. Genetics **4**(1): e21-e22.
- Li, J. Z., D. M. Absher, H. Tang, et al. (2008). "Worldwide human relationships inferred from genome-wide patterns of variation." Science **319**(5866): 1100-1104.
- Mourant, A. E. (1947). "The blood groups of the Basques." Nature **160**(4067): 505.
- Mulero, J. J., B. Budowle, J. M. Butler, et al. (2006). "Letter to the Editor—Nomenclature and Allele Repeat Structure Update for the Y-STR Locus GATA H4." Journal of Forensic Sciences **51**(3): 694-694.
- Nei, M. (1987). Molecular Evolutionary Genetics, Columbia University Press.
- Pereira, L., C. Cunha and A. Amorim (2004). "Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database." International Journal of Legal Medicine **118**(3): 132-136.
- Pontes, M. L., L. Cainé, D. Abrantes, et al. (2007). "Allele frequencies and population data for 17 Y-STR loci (AmpFISTR® Y-filer(TM)) in a Northern Portuguese population sample." Forensic Science International **170**(1): 62-67.

- Robino, C., F. Crobu, C. Di Gaetano, et al. (2008). "Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample." International Journal of Legal Medicine **122**(3): 251-255.
- Rosenberg, N. A., J. K. Pritchard, J. L. Weber, et al. (2002). "Genetic structure of human populations." Science **298**(5602): 2381-2385.
- Sánchez, C., C. Barrot, A. Xifró, et al. (2007). "Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer(TM) kit." Forensic Science International **172**(2-3): 211-217.
- Stanciu, F., V. Cutar, S. Pîrlea, et al. (2010). "Population data for Y-chromosome haplotypes defined by 17 STRs in South-East Romania." Legal Medicine **12**(5): 259-264.
- Turrina, S., R. Atzei and D. De Leo (2006). "Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay." International Journal of Legal Medicine **120**(1): 56-59.
- Veselinovic, I. S., D. M. Zgonjanin, M. P. Maletin, et al. (2008). "Allele frequencies and population data for 17 Y-chromosome STR loci in a Serbian population sample from Vojvodina province." Forensic Science International **176**(2-3): e23-e28.
- Willuweit, S. and L. Roewer (2007). "Y chromosome haplotype reference database (YHRD): update." Forensic Sci Int Genet **1**(2): 83-87.
- Zlojutro, M., R. Roy, J. Palikij, et al. (2006). "Autosomal STR variation in a Basque population: Vizcaya Province." Hum Biol **78**(5): 599-618.
- Zubiri, N. (2006). A Travel Guide to Basque America. Reno, Nevada, University of Nevada Press.