Int J Legal Med (2012) 126:293–298
DOI 10.1007/s00414-011-0644-8

ORIGINAL ARTICLE

# Y-STR variation in the Basque diaspora in the Western USA: evolutionary and forensic perspectives

**Laura Valverde · Melania Rosique · Stephan Köhnemann · Sergio Cardoso · Ainara García · Adrián Odriozola · Jose María Aznar · David Celorrio · Marianne Schuerenkamp · Josu Zubizarreta · Michael C. Davis · Greg Hampikian · Heidi Pfeiffer · Marian M. de Pancorbo**

Received: 13 April 2011 / Accepted: 26 October 2011 / Published online: 13 November 2011

**Abstract** Individuals of Basque origin migrated in large numbers to the Western USA in the second half of the nineteenth century, and the flow continued with less intensity during the last century. The European source population, that of the Basque Country, has long been a cultural and geographical isolate. Previous studies have demonstrated that Y-STR frequencies of Basques are different from those of other Spanish and European populations [1]. The Basque diaspora in the Western USA is a recent migration, but the founder effect and the incorporation of new American Y chromosomes into the paternal genetic pool of the Basque diaspora could have influenced its genetic structure and could thus have practical implications for forensic genetics. To check for genetic substructure among the European source and Basque diaspora populations and determine the most suitable population database for the Basque diaspora in the Western USA, we have analysed the haplotype distribution of 17 Y-STRs in both populations. We have found that the Basque diaspora in the Western USA largely conserve the Y chromosome lineage characteristic of the autochthonous European Basque population with no statistically significant differences. This implies that a common 17 Y-STR Basque population database could be used to calculate identification or kinship parameters regardless of whether the Basque individuals are from the European Basque Country or from the Basque diaspora in the Western USA.

**Keywords** Y-STRs · Basque diaspora · Western USA · Basque Country · Migration · Database

## Introduction

The autochthonous Basque population has a long history of migration. The reasons that have led Basques to emigrate are numerous and include the limited economic opportunities available at home, the Basque system of bequeathing property to a single heir, overpopulation in rural areas, the ravages of the First and Third Carlist Wars (fought in 1833–1839 and 1872–1876), the liberalisation of the Spanish economy in 1853 and finally the Spanish Civil War (1936–1939) and the subsequent Franco dictatorship [2].

The Basque community in the Western USA is one of the most representative examples of the Basque diaspora. Emigration to this area peaked in the second half of the nineteenth century, coinciding with the California gold rush. Basques arriving in the Western USA gradually settled in California, Nevada, Idaho, Utah, Oregon and Wyoming, working mainly as sheep herders, livestock keepers or park rangers. The flow of migrants continued up to the early 1970s, albeit with a pause between the 1920s and the 1940s. The US civil authorities recognise the Basques as one of the ethnic groups that comprise the

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00414-011-0644-8) contains supplementary material, which is available to authorized users.

L. Valverde · M. Rosique · S. Cardoso · A. García ·
A. Odriozola · J. M. Aznar · D. Celorrio · M. M. de Pancorbo (✉)
BIOMICs Research Group, CIEA "Lucio Lascaray" Center for Research and Advanced Studies, University of the Basque Country UPV/EHU,
Avda. Miguel de Unamuno 3,
01006 Vitoria-Gasteiz, Spain
e-mail: marian.mdepancorbo@ehu.es

S. Köhnemann · M. Schuerenkamp · H. Pfeiffer
Institute of Legal Medicine, University of Münster,
Münster, Germany

J. Zubizarreta · M. C. Davis · G. Hampikian
Department of Biological Sciences, Boise State University,
Boise, ID, USA

American nation [3] (http://www.lehendakaritza.ejgv.euskadi.net).

Although the Western US Basque diaspora is the result of recent migrations, differences with its source population could arise due to founder effects and the incorporation of new American Y chromosomes into its paternal gene pool. Such differences would influence the genetic structure of this population and, therefore, have implications in forensic practice. In this context, an assessment of the genetic structure of the Basque diaspora in the Western USA is important. To evaluate the influence of this migration on the genetic structure of this population, we conducted a study of 17 Y-STRs (Y Chromosome Short Tandem Repeats). These markers were selected because they are used worldwide in forensic genetics and because there are no data available for the haplotypes of these 17 Y-STR loci in the Basque source population or in its diaspora in the Western USA. One advantage of using markers with uniparental inheritance is that they can be used as lineage markers, enabling inferences to be drawn concerning the recent evolutionary history of populations [4]. They are therefore eminently suitable for assessing the effects of a recent migration such as the Basque diaspora in the Western USA. Furthermore, assessing the genetic structure will help to determine the most suitable Y chromosome forensic database for the Basque diaspora population in the Western USA.

## Materials and methods

### Populations studied

A total of 217 healthy, unrelated male individuals were studied. All participants provided their written informed consent, and the procedures followed were in accordance with the ethical standards of the Helsinki Declaration of 1975, as revised in 2000.

The sample of autochthonous males from the Basque Country (BAS, $n=139$) was selected according to criteria of autochthony, established by requiring individuals to have ancestors with surnames in the Basque language on both sides of their families dating back four generations (Online Resource 1) and all four grandparents born in the Basque Country.

The population sample from the Basque diaspora in the Western USA (BD, $n=78$) includes autochthonous Basque emigrants and their descendents living in Boise (Idaho), Chino (California) and Reno (Nevada).

### Y chromosome analysis

The set of 17 Y-STRs was amplified using the AmpFlSTR ® YfilerTM kit (Applied Biosystems), following the recommendations of the manufacturer. Capillary electrophoresis took place on an ABI Prism 3130 Genetic Analyzer, and sizes were assigned to the different fragments using GeneMapper® v4.0 software. All of the electropherograms were checked by a second researcher. The nomenclature used is that of the latest recommendations by the DNA Commission of the International Society of Forensic Genetics [5], except for locus Y GATA H4, which was named on the basis of the allelic ladder supplied with the AmpFlSTR ® YfilerTM kit.

This paper follows the recently released guidelines for publication of population data [6] and proficiency testing of the GHEP-ISFG working group (http://www.gep-isfg.org).

### Statistical analysis

Y-STR haplotypes with duplicated loci, null alleles or microvariants were excluded from the statistical calculations. Loci DYS385a and DYS385b were treated as a single haplotype in variability calculations and were excluded from the rest of the calculations because it is impossible to assign a specific allele to each of them. The locus DYS389II was treated in all the analyses after subtracting the value of DYS389I (DYS389b). Y-STR haplotype frequencies, diversity, Fst, Rst and pairwise difference tests were calculated with Arlequin v3.1 software [7]. RMP (Random Match Probability) and PD (Power of Discrimination) were also calculated due to their forensic interest. The phylogenetic relationships between the Y-STR haplotypes of the European Basques and the Western US diaspora were estimated by median-joining network using NETWORK v 4.5.1.6 [8]. Greater phylogenetic weight was allocated to the least variable loci, with the variance ($V_L$) of the Y-STRs being calculated by the method described in [9]. The genetic distances based on Fst and Rst between population pairs were plotted with PAST [10] in non-metric multidimensional scaling (MDS) graphs using the Euclidean similarity measure.

In an attempt to detect and quantify the contribution of the American population to the gene pool of the Basque diaspora in the Western USA, we carried out an admixture calculation based on Y-STR haplotypes using ADMIX v.2.0 [11].

## Results

### Y-STR haplotypes in the population of the Basque Country and the Basque diaspora in the Western USA

The haplotypes of 17 Y-STR loci obtained for the 217 male individuals from the autochthonous population of the Basque Country (BAS) and from the Basque diaspora in the Western USA (BD) are available upon request and in the YHRD database (http://www.yhrd.org) [12] under accession numbers YA003672 (Alava), YA003673 (Biscay), YA003674 (Guipuzcoa), YA003675 (California), YA003676 (Idaho) and YA003677 (Nevada). These are the first data from Basque populations genotyped for 17 Y-STR loci.

A null allele was detected in locus DYS391 in one BAS individual and an 18.3 microvariant in locus DYS635 in one BD individual. These samples were confirmed by reamplification.

Haplotype diversity values are shown in Table 1, along with the forensic parameters RMP and PD for each population sample. Pairwise difference tests show no statistically significant differences in haplotype distribution between the Basque diaspora in the Western USA (BD) and the autochthonous population of the Basque Country (BAS) (Online Resource 2). There is also no difference between these Basque populations (BAS and BD) and the previously reported 11-YSTR haplotypes of the European Basque population [1] (Online Resource 3).

The phylogenetic relationship between the Y-STR haplotypes from the European Basques and the diaspora is shown in the median-joining network in Fig. 1. The core of the network shows a star-like structure made up of individuals from both populations studied with no line of independent divergence. This lack of divergence between populations, the lack of significant differences and the maintenance of a high diversity of Y-STR haplotypes seem to indicate that there are no founder haplotypes or significant amounts of different American haplotypes incorporated into the gene pool of the Basque diaspora.

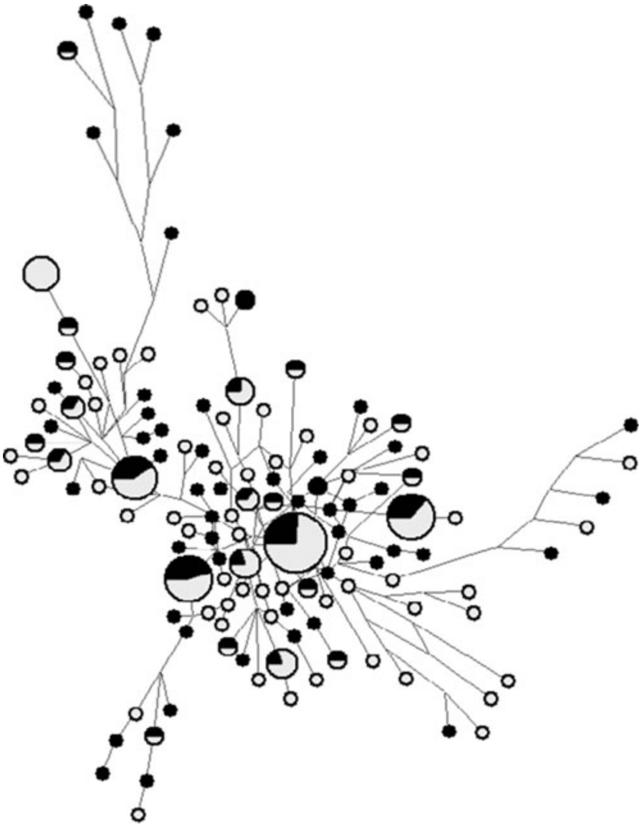

**Fig. 1** Median-joining network of 17 Y-STR haplotypes from the Basque Country (*white nodes*) and the Basque diaspora in the Western USA (*black nodes*)

## Analysis of the contributions to the Y chromosome genetic pool of the Basque diaspora in the Western USA

Basque diaspora individuals can be identified by their Basque surnames (in the Basque language). We determined the frequency of Basque surnames in the Western USA Basque diaspora population sample and found that 86% had a paternal Basque surname, while 14% had a non-Basque paternal ancestry. To check for the origin of haplotypes incorporated into the gene pool of the Basque diaspora in the Western USA, we compared the Y-STR haplotypes of the Basque diaspora with those in the Basque source population and with subpopulations from the USA (Caucasians, Hispanics, Africans and Asians) [13].

**Table 1** Parameters of forensic interest in populations from the Basque Country and the Basque diaspora in the Western USA, calculated with data from 17 Y-STR loci

| Populations | Basques | Basque diaspora |
|---|---|---|
| Population size$^a$ | 138 | 77 |
| Different haplotypes | 116 | 76 |
| Unique haplotypes | 101 | 76 |
| Haplotype diversity ± SD | 0.9966±0.0016 | 0.9997±0.0022 |
| Random match probability (RMP) | 0.0106 | 0.0133 |
| Power of discrimination (PD) | 0.9894 | 0.9867 |

[a] *Individuals with intermediated, duplicated or null alleles were excluded*

First, we looked for shared haplotypes between the populations. As expected, the highest proportion of shared identical haplotypes was observed between the Basque diaspora and its source population (15%), followed by Caucasians (6%) and Hispanic Americans (1%). No shared haplotypes were found between the Basque diaspora and the African and Asian American subpopulations (Online Resource 4).

We carried out admixture calculations based on Y-STR haplotypes, conducted with the Basque source population and various American subpopulations (Caucasian, Hispanic, African and Asian) as possible parental populations for the Basque diaspora. To that end, a molecular distance matrix was constructed by adding up the total number of mutational steps between each pair of different haplotypes. The results show a contribution of 10–20% of Basque

European Y chromosomes (BAS) to the Basque diaspora (BD), according to the data obtained by the analyses of shared haplotypes between BAS and BD populations. The contribution of Y haplotypes from Caucasian and Hispanic Americans was 80–90%, with no contribution from the remaining American subpopulations. These results seem to be an underestimate based on the data from Basque surnames. This is probably due to the fact that the European Basque population is a Caucasoid group, and Hispanic Y chromosome haplotypes are also characterised by having a high component of European Y chromosomes [14]. Perhaps the admixture calculations are reflecting the common origin of 80–90% of the Y chromosome in all the population studied (Basques, Basque diaspora, European and Hispanic Americans), with no capacity to discriminate between the contributions of each one to the genetic pool of the Basque diaspora in the Western USA.

### Population comparisons

Finally, genetic distance calculations based on the Fst statistic were performed to assess the genetic relationships between the Basque founder population, its Western USA diaspora and other populations from North America and Europe [13, 15–23]. Only the populations with the same 17 Y-STR loci data were included in these calculations. Genetic distances between populations were plotted in a multidimensional scaling (MDS) graph (Fig. 2). The Western USA Basque diaspora population sample (BD) is located in quadrant III, very close to the European Basque population (BAS). The next closest populations are Caucasian Americans (CA-US) and Spaniards (NSP, SSP). This agrees with the previous observation that there are few new haplotypes incorporated in the Basque diaspora and that most of them have a more than probable European ancestry. In addition, the Caucasian American population is closest to the Western European populations.

Hispanic Americans (HI-US) are also very close to Caucasian Americans (CA-US), Spaniards (NSP, SSP) and Western European populations, which is consistent with their documented European paternal ancestry [14].

Genetic distances based on the Rst parameter, an analogue of Fst which takes into account mutational steps, did not produce clear results since the genetic distances between some of the populations included in the analysis were incongruent with the known relationship between them (Online Resource 2; see "Discussion").

Therefore, although a few new haplotypes seem to be incorporated into the Y chromosome genetic pool of the diaspora studied here, all of the results indicate a great homogeneity between the Basque diaspora and its source population, which suggests that it is possible to apply the same Y-STR haplotype frequency database to calculate probabilities for forensic purposes.

## Discussion

The distribution and evolution of paternal lineages in the Basque diaspora in the Western USA can be examined by studying their Y-STRs haplotypes. With this aim, we present here the first study of 17 Y-STR loci in autochthonous European Basques and the Basque diaspora in the Western USA.

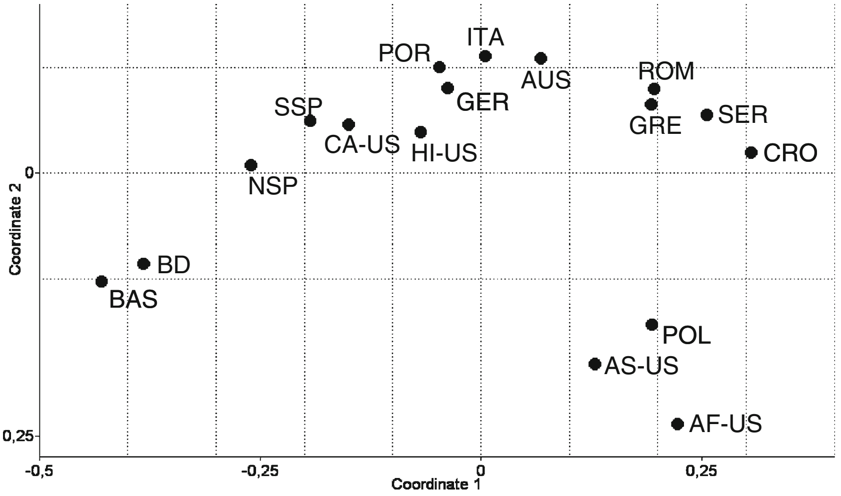

**Fig. 2** Multidimensional scaling analysis based on Fst genetic distances in the populations examined and other populations for which data has been published (stress index 0.05). *BAS* Basque Country, present study; *BD* Basque diaspora in the Western USA, present study; *NSP* Northern Spain, data not published; *SSP* Southern Spain [15]; *CA-US* Caucasian Americans [13]; *HI-US* Hispanic Americans [13]; *AS-US* Asian Americans [13]; *AF-US* African Americans [13]; *POR* Portugal [16]; *ITA* Italy [17]; *AUS* Austria [18]; *GER* Germany [19]; *POL* Poland [19]; *ROM* Romania [20]; *GRE* Greece [21]; *SER* Serbia [22]; *CRO* Croatia [23]

Our results demonstrate a very high-level of conservation of the Y chromosome haplotypes characteristic of the European autochthonous Basque population among individuals of the Basque diaspora in the Western USA. No signs of founder haplotypes have been found, probably because the number of European Basque male individuals who have settled in the Western USA has been large enough for a similar Y-haplotype frequency distribution to be reached between the source and the emigrant populations. Neither geographical separation from their original homeland nor contact with the American population seems to have resulted in significant differentiation.

Basque surnames have proven their advantage in the Basque diaspora for the identification of paternal lineages with Basque ancestry. This is of great interest for identifying individuals with Basque ancestry, for then applying the proper Basque Y chromosome forensic database and even for identifying the most suitable individuals for the construction of specific Y-STR databases for the Basque diaspora.

Additionally, Basque surnames have enabled us to estimate the paternal admixture proportion in the actual Basque diaspora gene pool, revealing that only approximately 14% of the Y chromosomes are not consistent with autochthonous Basque origins. On the other hand, it should be noted that admixture calculations based on Y-STR haplotypes between genetically similar populations may result in an underestimation of admixture [24], as shown in our results, so in the actual case of the Basque diaspora, the identification of paternal Basque ancestry based on the first surname is an extraordinary and valuable peculiarity.

Comparisons between the Basque founder population, its Western USA diaspora and other populations from North America and Europe reveal that the most likely ancestry of the new Y-STR haplotypes incorporated into the genetic pool of the Basque diaspora is Caucasian American, with the genetic structure closest to Western Europe populations. The inclusion in the comparison analyses of populations from both Western and Eastern Europe was necessary since it has been demonstrated that Caucasian Americans from different geographical areas of the USA are substructured due to a non-random form of migration [25, 26].

Another result worth mentioning is the incongruent genetic distances obtained using the statistical parameter Rst, which reveals unexpectedly large distances between populations with known relationships (Online Resource 2). The use of the statistical parameter Rst is recommended to assess genetic differentiation among populations with microsatellites because it takes into account mutational steps; however, it seems to be appropriate only when the mutation rate is higher than the migration rate. In the case of the recent migration of the Basque diaspora, Fst has proved to be more efficient [27–29].

To sum up, the Basque diaspora in the Western USA maintains the characteristics of its European source population in its paternal lineages, so the use of Y-STR haplotype frequencies from the European Basque population is recommended over the use of Caucasian American, Western European or Spanish Y-STR databases to calculate parameters of forensic interest for males of the Basque diaspora. Basque surnames could be used for the identification of individuals with Basque ancestry, which is of special interest for a proper construction and forensic application of Y chromosome databases. Moreover, our findings indicate that the Basque diaspora in the Western USA is beginning to incorporate new haplotypes in a non-random way, with the preferential inclusion of Caucasian American haplotypes whose ancestry is most closely linked to Western Europe.

**Acknowledgements** Funds were provided by the Basque Government (BIOMICs Research Group IT424-07). LV received a grant from the University of the Basque Country UPV/EHU. The technical and human support provided by DNA Bank-General Research Services SGIker (UPV/EHU, Spanish Ministry of Science and Innovation MICINN, Basque Government GV/EJ, European Regional Development Fund ERDF and European Science Foundation ESF) is gratefully acknowledged, especially to M.A. The authors would like to thank Alfonso Sánchez for his statistical support. The authors would also like to thank the Basque Foundation for Health Research and Innovation (BIOEF). We are grateful to G. Eguskiza for his technical support and to E. Sanz and O. Alvarez for their assistance in sample collection. The authors are deeply indebted to the Basque communities of the Western USA, which cooperated generously to the development of this study, especially to P. Miller, J. Navarte and J. Isurza.

**Conflicts of interest** The authors declare that they have no conflicts of interest.

## References

1. García O, Martín P, Gusmão L, Albarrán C, Alonso S, de la Rua C, Flores C, Izagirre N, Peñas R, Pérez JA, Uriarte I, Yurrebaso I, Alonso A (2004) A Basque Country autochthonous population study of 11 Y-chromosome STR loci. Forensic Sci Int 145:65–68
2. Totoricagüena G (2004) Basque diaspora: migration and transnational identity. Basque Textbook Series, Center for Basque Studies, University of Nevada, Reno, Nevada
3. Douglas W, Bilbao J (1975) Amerikanuak: los vascos en el nuevo mundo. University of Nevada Press, Reno, Nevada
4. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612
5. Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM (2006) 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 157:187–197
6. Parson W, Roewer L (2010) Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine. Int J Legal Med 124(5):505–509

7. Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50
8. Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48
9. Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68:990–1018
10. Hammer O, Harper DAT, Ryan PD (2001) PAST: Paleontological Statistics software package for education and data analysis. Palaeontol Electron 4(1):9
11. Dupanloup I, Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol Biol Evol 18(4):672–675
12. Willuweit S, Roewer L (2007) On behalf of the International Forensic Y chromosome User Group, Y Chromosome Haplotype Reference Database (YHRD): update. Forensic Sci Int Genet 1:83–87
13. Decker AE, Kline MC, Redman JW, Reid TM, Butler JM (2008) Analysis of mutations in father–son pairs with 17 Y-STR loci. Forensic Sci Int Genet 2:e31–e35
14. Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H (2010) Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proc Natl Acad Sci USA 107(Suppl 2):8954–8961
15. Gaibar M, Esteban E, Moral P, Gómez-Gallego F, Santiago C, Bandrés F, Luna F, Fernández-Santander A (2010) STR genetic diversity in a Mediterranean population from the south of the Iberian Peninsula. Ann Hum Biol 37(2):254–267
16. Alves C, Gomes V, Prata MJ, Amorim A, Gusmão L (2007) Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFlSTR YFiler) in Portugal. Forensic Sci Int 171:250–255
17. Turrina S, Atzei R, De Leo D (2006) Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. Int J Legal Med 120:56–59
18. Berger B, Lindinger A, Niederstätter H, Grubwieser P, Parson W (2005) Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. Int J Legal Med 119:241–246
19. Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, Lao O, Brauer S, Krüger C, Roewer L, Lessig R, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M (2009) Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR® Yfiler® PCR amplification kit. Int J Legal Med 123:471–482
20. Stanciu F, Cuţăr V, Pîrlea S, Stoian V, Stoian IM, Sevastre O, Popescu OR (2010) Population data for Y-chromosome haplotypes defined by 17 STRs in South-East Romania. Leg Med (Tokyo) 12(5):259–264
21. Kovatsi L, Saunier JL, Irwin JA (2009) Population genetics of Y-chromosome STRs in a population of Northern Greeks. Forensic Sci Int Genet 4:e21–e22
22. Mirabal S, Varljen T, Gayden T, Regueiro M, Vujovic S, Popovic D, Djuric M, Stojkovic O, Herrera RJ (2010) Human Y-chromosome short tandem repeats: a tale of acculturation and migrations as mechanisms for the diffusion of agriculture in the Balkan Peninsula. Am J Phys Anthropol 142:380–390
23. Ljubković J, Stipisić A, Sutlović D, Definis-Gojanović M, Bucan K, Andelinović S (2008) Y-chromosomal short tandem repeat haplotypes in southern Croatian male population defined by 17 loci. Croat Med J 49(2):201–206
24. Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. Mol Biol Evol 15:1298–1311
25. Kayser M, Brauer S, Schadlich H, Prinz M, Batzer MA, Zimmerman PA, Boatin BA, Stoneking M (2003) Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. Genome Res 13:624–634
26. Hammer MF, Chamberlain VF, Kearney VF, Stover D, Zhang G, Karafet T, Walsh B, Redd AJ (2006) Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. Forensic Sci Int 1(164):45–55
27. Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163:1467–1482
28. Balloux F, Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite model. Mol Ecol 11:771–783
29. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462