International Congress Series 1239 (2003) 61–66

Microgeographic patterns of highly informative Y-chromosome haplotypes (using biallelic markers and STRs) in Galicia (NW Spain): forensic and anthropological implications

M. Brión [a], B. Quintáns [a], A. González-Neira [a], M. Zarrabeitia [b], A. Salas [a], J. Riancho [c], M.V. Lareu [a], P. Sánchez-Velasco [d], A. Carracedo [a,*]

[a] *Faculty of Medicine, Institute of Legal Medicine, University of Santiago de Compostela, San Francisco s/n, 15786 Santiago de Compostela, Spain*
[b] *Unit of Legal Medicine, University of Cantabria, Santander, Spain*
[c] *Department of Internal Medicine, Hospital U.M. Valdecilla, Santander, Spain*
[d] *Department of Inmunology, Hospital U.M. Valdecilla, Santander, Spain*

Abstract

The effect of sampling in the estimation of Y-chromosome haplotype frequencies is analyzed in this work in the population of Galicia (NW Spain) and Cantabria, as well as the importance of population substructuring at a local level. Galicia (NW Spain) comprises a population of near 3 million inhabitants with a clear historical matriarchal inheritance. Two population surveys were carried out, one from the general random population (mainly from the most important cities of the area) and another one, excluding the main cities and distributed in seven natural geographical areas. By contrast, Cantabria has a clear patriarchal inheritance, and is distributed in several geographical areas with a certain degree of isolation from each other. Seven Y STRs and 10 binary markers were analyzed in all the samples and significant differences were found between the populations living in towns and those in rural areas. Surprisingly, in Galicia, there are no significant differences among the different natural geographical areas with low Fst values; however, in the population of Cantabria, big differences were found. The possible effect of cultural traits on this unexpected microgeographic distribution of Y-chromosome haplotypes will be discussed in this work.

Keywords: Population substructure; Sampling effects

* Corresponding author. Tel.: +34-981-582327; fax: +34-981-580336.
E-mail address: apimlang@usc.es (A. Carracedo).

0531-5131/03
PII: S0531-5131(02)00394-1

1. Introduction

During the last few years, the interest in Y-chromosome polymorphisms has continuously increased, not only because of the application of Y-chromosome variation in human and evolutionary genetics but also because of its interest in forensics.

The ability to identify male-specific DNA makes highly variable Y-chromosomal STRs an invaluable addition to the standard panel of autosomal loci used in forensic genetics. The male specificity of the Y chromosome and its paternal inheritance makes this chromosome especially useful in cases of mixed stain analysis (male/female cell admixture) and in kinship testing. Still a matter of research but of considerable interest is the potential of Y-chromosome analysis to unravel the ethnic background of a given DNA profile.

One of the problems with the use of Y-chromosome data in forensic cases is how to assess the evidential strength when an inclusion is reported. As with mtDNA, it is invalid to estimate evidential strength by multiplication of frequencies of individual polymorphisms because the Y-chromosome variation from the non-recombinant part of the chromosome is inherited as a haplotype.

There are two ways to report the value of the evidence in this case. Either the counting method may be used where the number of observations may be declared in a relevant database, or, better, a likelihood ratio or Bayesian approaches may be used. The problem is outlined in the DNA Commission guidelines on Y chromosome and mtDNA [1,2].

In any case, the estimated Y-chromosome haplotype frequencies should be interpreted in the light of the data available concerning the distribution of the Y-chromosome haplotypes and the possible subpopulation structures within the relevant populations.

In Europe, Y-chromosomal diversity is clinal and influenced primarily by geography rather than language [3] and ϕ_{ST} values have been found to be <0.05 among many European populations [4]. At a more local level, in the Iberian peninsula, we have demonstrated [5] that despite the peninsula being home to several different populations with diverse languages and diverse histories, there is an important similarity amongst these populations for SNP data and also for STRs (Ref. [5]; Brión et al., submitted for publication). Specifically, the ϕ_{ST} values have been found to be approximately 0.05 among the different populations.

Databases are collected on the basis of unrelatedness; therefore, close relatives are excluded as far as possible. Consequently, common haplotypes, because they come from large pedigrees within a certain population, may be only sampled once and therefore frequencies will be underestimated because of deliberate non-relatedness selection criteria. In a random survey from a large cosmopolitan population with marked mobility, the chance of selecting two people related by patrilineage is low. However, in small isolated rural populations, the effect may be important and must be considered if the frequency estimate from a general population is used to make inferences on the former.

To know the importance of these events, we have analyzed at a microgeographic level the autonomous region of Galicia and we have compared the results obtained with those obtained from another region in Northern Spain (Cantabria).

2. Materials and methods

In Galicia, there are close to 3 million inhabitants, around 1 million living in the main towns (Vigo, A. Coruña, Santiago, etc.) and the other 2 million scattered in small villages in rural areas through the whole region. Galicia can be divided into different districts according to some cultural, geographical and historical traits.

Two different samplings were carried out in Galicia, one including the main towns (104 samples) and the other from small villages from the different districts (292 samples).

To see if the results are the same in other areas, we also analyzed the population of Cantabria. Sampling was carried out with the same requirements in three district areas (159 samples). Although not geographically distant, cultural traditions in Galicia and Cantabria are different and, in addition, the districts in Cantabria remained more isolated from each other due to geographic barriers.

Each of the samples was analyzed for 10 Y-chromosome binary markers (YAP, SRY-8299, SRY-2627, SRY-1532, sY81, M9, 92R7, LLY22g and Tat), and for 7 Y-chromosome STRs (DYS19, DYS389I y II, DYS390, DYS391, DYS392 and DYS393), following the conditions described by Brión et al. ([5], submitted for publication).

Haplogroup frequencies were estimated by single gene counting procedures. STR allele frequencies, haplotype frequencies and diverse Y-chromosome diversity indices were calculated using the Arlequin package version 1.1 [6]. Population differentiation was tested by a Markov test using the Arlequin software.

3. Results and discussion

The analysis of binary polymorphisms in Galician samples has shown that there are significant differences between both sampling methods (Fig. 1).

However, when we analyzed the data from the different districts, we observed with surprise that there were no differences between the different districts. Only a few differences were found between some of the districts from the northeast and southwest using SNPs ($P < 0.02$, 10,000 Markov chain steps), and from the east and west using STRs ($P < 0.03$, 10,000 Markov steps done). Additionally, Fst values among the different districts in Galicia have been found to be < 0.01 (Table 1).

In the case of the Cantabrian region, the isolation could explain the observed, statistically significant, differences among the three districts analysed (Fig. 2). Fst values in this case have been found to be < 0.05 (Table 1).

Results from the population of the Pas Valley were particularly interesting, with a high proportion of haplogroup 21 (28%), which has a very high frequency in North African populations ($\sim$ 70%).

Our initial impression was that the differences between both sampling methods in Galicia could be attributed to relatedness effects in the rural local areas.

However, the homogeneity found among the Galician districts could suggest a higher mobility of males rather than females and this is in agreement with the cultural and legal tradition of Galician rural society. Traditionally, the rural areas in Galicia have a matriarchal structure and from very old times as established in Galician Civil Law, the

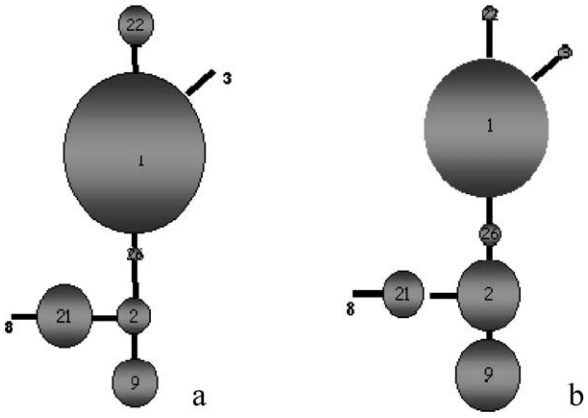

Fig. 1. Unrooted parsimony network trees constructed using 10 biallelic markers showing the differences between both sampling methods in Galicia. (a) Represents the general Galicia population and (b) the population from the rural areas. Circles represent the frequency contribution to each haplogroup and lines represent single mutational steps between them. Circle areas are proportional to the frequency of each haplogroup and arrowheads indicate the derived states.

family house and the main properties are inherited from the parents to the first daughter. This could have led to a higher mobility of males compared with females.

In the light of these results, the differences between the sampling of towns and of rural areas could be better explained by the fact that although Galicia has had high emigration rates for centuries, there is certain level of migration to the main towns from other parts of Spain and recently from South America.

In conclusion we can say, that the sampling method can affect the results and significant differences can be observed if the sampling is made in large cosmopolitan populations with marked mobility but also with more or less important migratory effects compared to rural areas. Additionally, in rural areas, the effect of population substructuring can dramatically change from one region to another due to cultural effects such as the effect

Table 1
Fst values among the different districts belonging to Galicia and Cantabria

	Cantabria			Galicia						
	Liébana	Pas	Santander	Ourense	Noroeste	G. Ártabro	M. Lucense	Santiago	Lugo	R. Baixas
Pas	0.0433									
Santander	0.0213	0.0105								
Ourense	−0.0091	0.0429	0.0119							
Noroeste	0.0043	0.0333	0.0326	−0.0001						
G. Ártabro	−0.0123	0.0400	−0.0013	−0.0172	0.0081					
M. Lucense	−0.0101	0.0515	0.0317	−0.0112	0.0120	−0.0056				
Santiago	−0.0030	0.0425	0.0196	−0.0105	−0.0018	−0.0030	−0.0000			
Lugo	0.0045	0.0346	0.0037	−0.0050	0.0024	−0.0124	0.0062	0.0006		
R. Baixas	−0.0052	0.0613	0.0290	−0.0164	−0.0043	−0.0108	−0.0106	−0.0052	−0.0019	
M. B. Miño	−0.0103	0.0124	0.0057	−0.0136	−0.0101	−0.0126	−0.0117	−0.0040	−0.0009	−0.0067

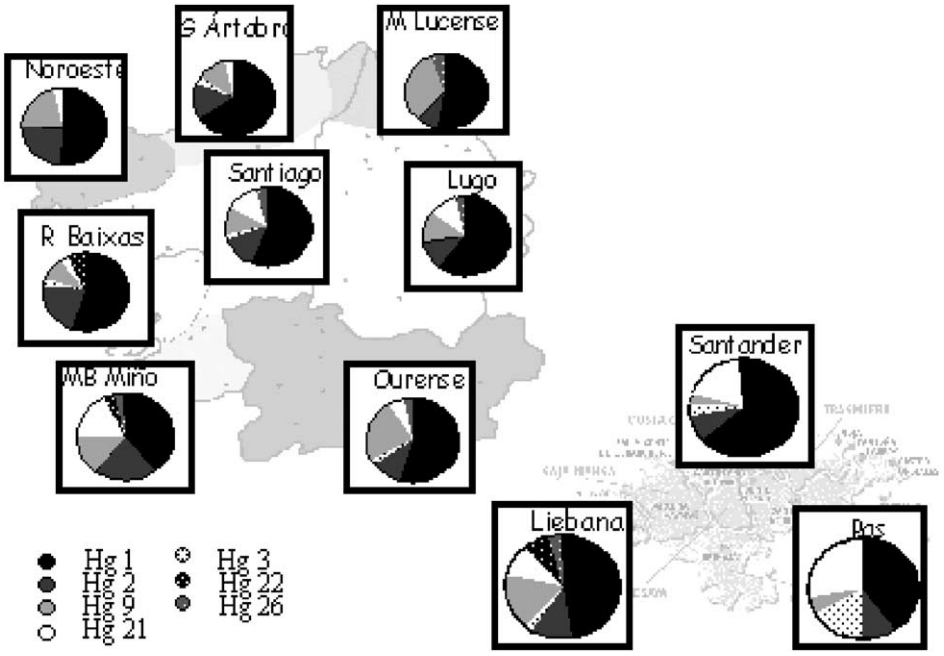

Fig. 2. Frequency distribution of each haplogroup in the different districts belonging to Galicia and Cantabria.

of traditional matriarchal structures in Galicia or the geographical isolation of some areas in Cantabria.

It is clear that there is no universal solution for the effect of population substructuring at a geographic level other than to type samples from all the specific (rural) populations in question wherever possible.

Acknowledgements

We thank Meli Rodríguez for her technical assistance, which is highly appreciated. This work was supported by grants from the ''Xunta de Galicia'' (PGIDT01PXI20806PR) and from the ''Ministerio de Ciencia Tecnología'' (2000/PC092).

References

[1] P. Gill, C. Brenner, B. Brinkman, et al., DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, Int. J. Legal Med. 114 (2001) 305–309.
[2] A. Carracedo, W. Bär, P.J. Lincoln, et al., DNA Commission of the International Society of Forensic Genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.
[3] Z.H. Rosser, T. Zerjal, M.E. Hurles, et al., Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language, Am. J. Hum. Genet. 67 (2000) 1526–1543.
[4] A. Carracedo, A. Beckmann, A. Bengs, et al., Results of a collaborative study of the EDNAP group regarding

the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I y II, DYS390 and DYS393 in a PCR pentaplex format, Forensic Sci. Int. 119 (2001) 28–41.

[5] M. Brión, M.V. Lareu, L. Pereira, et al., Y chromosome lineages: construction of highly informative haplotypes using biallelic markers, STRs and the minisatellite MSY1, in: G.F. Sensabaugh, et al. (Eds.), Progress in Forensic Genetics, vol. 8, Elsevier, Amsterdam, 2000, pp. 263–265.

[6] S. Schneider, J.-M. Kueffer, D. Roessli, L. Excoffier, Arlequin ver 1.1: a software for population genetic data analysis, Genetics and Biometry Laboratory, University of Geneva, Switzerland, 1997.