ANALYSIS OF D1S80 VNTR ALLELE POLYMORPHISM AND ASSOCIATION WITH A NEARBY FLANKING SEQUENCE POLYMORPHISM IN TWO SPANISH POPULATIONS

C. Albarrán[1], O. García[2], R. Deka[3], A. Alonso[1], P. Martín[1], M. Sancho[1], D.N. Stivers[4] and R. Chakraborty[4].

[1] Sección de Biología. Instituto de Toxicología. Mº de Justicia e Interior. Madrid. SPAIN
[2] Laboratorio UTAP. Departamento de Interior. Gobiemo Vasco. Bilbao. SPAIN
[3] Department of Human Genetics, University of Pittsburgh. Pittsburgh. USA
[4] Human Genetics Center, University of Texas, Health Science Center, Houston. Texas. USA

INTRODUCTION

Variable number of tandem repeats (VNTR) loci are highly informative markers for linkage analysis and identity testing. In addition to the variation in the number of repeats, some VNTR loci display variabilty of the repeat sequence so that an additional polymorphism level can be observed analyzing the interspersion pattem of variant repeats along the tandem repeat array (Jeffreys et al. 1991; Neil and Jeffreys 1993). For the majority of VNTR loci, however, the extent of polymorphism of the flanking sequences is not known, despite some early studies showing linkage between VNTR specific-alleles and nearby RFLPs (Higgs et al. 1986; Cox et al. 1988; Renges et al. 1992; Martinson et al. 1994).

We have recently described a polymorphic *Hinf* I restriction site in the flanking region of the VNTR locus D1S80 (Alonso et al. 1995). In this study, we have determined the RFLP/VNTR D1S80 haplotype frequencies in a cosmopolitan population of individuals living in Madrid (Central-Spain) and in a population of autochthonous individuals from the Basque Country (North-Spain) in order to evaluate the magnitude of linkage between VNTR specific-alleles and the biallellic RFLP polymorphism.

MATERIALS AND METHODS

Population Sample and Sample Preparation

EDTA blood samples were collected from two different spanish population samples: a cosmopolitan population sample of 205 unrelated individuals living in Madrid and a population sample of 201 autochthonous individuals from the Basque Country. The DNA was extracted by the standard phenol/chloroform extraction procedure.

PCR, Digestion and Typing

PCR amplification of D1S80 was performed according to the manufacturer's recomendations using the AmpliFLP D1S80 Kit (Perkin-Elmer Corporation, Norwalk, CT). Restriction of the PCR products was performed as previously described (Alonso et al. 1995). Undigested PCR products and *Hinf* I restricted ones were typed by vertical polyacrylamide gel electrophoresis followed by silver stain (Alonso et al. 1995).

Statistical Analysis

The allele frequencies at individual loci were computed by the gene count method (Li CC, 1976). The haplotype frequencies in the two Spanish populations were estimated based on a maximum likelihood approach following an E-M algorithm of haplotype frequency estimation from two-locus genotype data (Long et al. 1995).

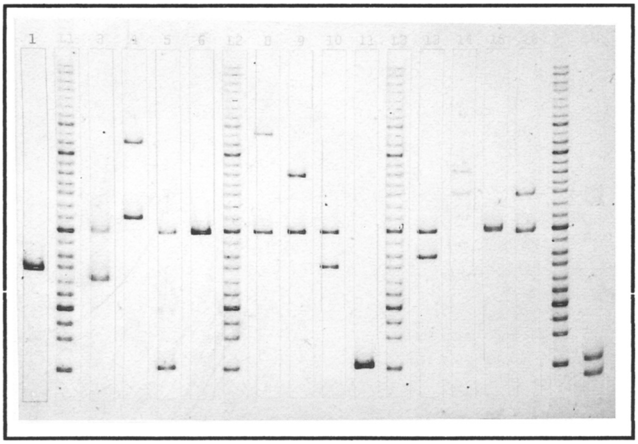

Fig 1. Representative D1S80 profiles as analyzed by restriction of the PCR products with *Hinf* I and subsequent polyacrylamide gel electrophoresis followed by silver stain. All samples were previously typed before *Hinf* I restriction. **Samples** (from left to right): 21-21-; ladder; 24+24-; 29+36+; 18+24-; 24-24-; ladder; 24-37+; 24-29-; 24-25+; 18+18+; ladder; 22-24-; 31+33+; 24-28+; 24-31+; ladder; 17+18+.

Table 1. Haplotype frequencies estimated using (Long 1995)'s haplotype estimation program. Figures in parentheses are expected haplotype frequencies when an absence of linkage disequilibrium is assumed. The G statistic, twice the negative log of likelihood ratios (which is roughly distributed as a χ^2 statistic) was 236.23 for the Madrid sample and 294.34 for the Basque sample; the null hypothesis is no disequilibrium between the two loci. Both are significant ($p < 0.001$), based on 1000 Monte Carlo simulations. Gene frequency totals were estimated directly from the sample, not by adding haplotype frequenies, and hence may differ slightly due to rounding error .

	Basque Country					Madrid				
D1S80	*Hinf* I -		*Hinf* I +		Totals	*Hinf* I -		*Hinf* I +		Totals
17	–	(–)	–	(–)	-	.0024	(.0025)	.0024	(.0024)	.005
18	–	(.1442)	.2637	(.1194)	.264	–	(.1263)	.2488	(.1226)	.25
20	.0368	(.0245)	.008	(.0203)	.045	.0122	(.0062)	–	(.006)	.012
21	.0547	(.0327)	.005	(.027)	.06	.0395	(.0223)	.0045	(.0216)	.044
22	.0327	(.0272)	.017	(.0225)	.05	.025	(.0235)	.0209	(.0228)	.046
23	.0075	(.0041)	–	(.0034)	.007	.0073	(.0037)	–	(.0036)	.007
24	.3279	(.1811)	.0029	(.1498)	.331	.3294	(.1794)	.0243	(.1742)	.354
25	.0373	(.0204)	–	(.0169)	.037	.0272	(.0272)	.0264	(.0264)	.054
26	.0050	(.0027)	–	(.0023)	.005	.0132	(.0074)	.0015	(.0073)	.015
27	–	(.0095)	.0174	(.0079)	.017	–	(.0025)	.0049	(.0024)	.005
28	–	(.0191)	.0348	(.0158)	.035	.0027	(.0223)	.0412	(.0216)	.044
29	.0324	(.0286)	.0198	(.0237)	.052	.033	(.0297)	.0255	(.0288)	.06
30	.0078	(.0109)	.0121	(.009)	.02	–	(.0037)	.0073	(.0036)	.007
31	.0026	(.0381)	.0671	(.0315)	.07	.0025	(.0247)	.0462	(.0240)	.05
32	.0025	(.0027)	.0025	(.0023)	.005	.0076	(.0062)	.0046	(.0060)	.012
33	–	(–)	–	(–)	-	–	(.0025)	.0049	(.0024)	.005
34	–	(–)	–	(–)	-	–	(.0025)	.0049	(.0024)	.005
35	–	(–)	–	(–)	-	–	(.0025)	.0049	(.0024)	.005
36	–	(–)	–	(–)	-	–	(.005)	.0098	(.0048)	.01
37	–	(.0013)	.0025	(.0011)	.002	–	(.005)	.0098	(.0048)	.01
40	–	(–)	–	(-)	-	.0049	(.0025)	–	(.0024)	.005
Total	.547	(.547)	.453	(.453)		.51	(.51)	.49	(.49)	

RESULTS AND DISCUSSION

Fig 1. shows some representative D1S80 profiles as analyzed by restriction of the PCR products with *Hinf* I and subsequent polyacrylamide gel electrophoresis followed by silver stain. A total of 21 VNTR-alleles and 31 RFLP/VNTR haplotypes were observed in the population sample of individuals living in Madrid while a total of 15 VNTR alleles and 23 RFLP/VNTR haplotypes were observed in the population of autochthonous individuals from the Basque Country.

Table 1 shows the estimated haplotype frequencies and allele frequencies in the two Spanish populations. Also shown are the expected haplotype frequencies in the two samples under the assumption of linkage equilibrium between the two sites (i.e., computed by multiplying the respective allele frequencies). As can be seen, the haplotype frequencies determined in both population samples show an extreme association between the *Hinf* I + allele and the VNTR allele of 18 repeats and between the *Hinf* I - allele and VNTR allele of 24 repeats, while the remaining VNTR allele associate more randomly with the two flanking *Hinf* I alleles. The linkage disequilibrium observed between the flanking polymorphism and the two high frequency modal VNTR alleles (18 and 24) that is share by the two different populations analyzed, together with the fact that mutation at minisatellite occurs without exchange of flanking regions (Jeffreys et al. 1994), suggest that the 18 and 24 alleles could be the original VNTR alleles at this locus and the rest of VNTR alleles associated with the *Hinf* I + mutation would arise from the 18 allele, while the VNTR allele associated with the *Hinf* I - mutation would arise from the allele of 24 repeats.

Data gathered in this work also indicates the utility of polymorphisms at these two sites for identity testing and parentage analysis. First, the genotype frequencies in these two samples are in general conformity with their respective Hardy-Weinverg expectations. This is checked by the likelihood ratio test and exact test, as performed in Edwards et al. (1992), Hammond et al. (1994), and Deka et al. (1995), particulary when the rare alleles of frequencies of 0.05 or less at the D1S80 site are merged with their adjacent alleles (data not shown). Second, at the level of the allele frequencies at individual loci, as well as at the level of haplotypes these two Spanish populations are not significantly different (haplotype frequency heterogeneity $\chi^2 = 56.7$ with 41 df, $p \sim 0.06$), suggesting that these two populations are not genetically very dissimilar. Third, with heterozygosity of nerly 50% at the *Hinf* I site and approximately 80% at the D1S80 VNTR loci in these samples, these two polymorphisms offer an average probability of discrimination between genotypes of two unrelated individuals exceeding 94%. In other words, these two polymorphism can constitute important elements of forensic and parentage analysis based on PCR-based DNA typing protocols.

REFERENCES

Alonso A, Martín P, Albarrán C, Sancho M (1995) Int J Legal Med 107: 216-218
Cox NJ, Bell GI, Xiang K-S (1988) Am J Hum Genet 43: 495-501
Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R (1995) Am J Hum Genet 56: 461-474
Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty (1992) Genomics 12: 241-253
Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Am J Hum Genet 55: 175-189
Higgs DR, Wainscoat JS, Flint J, Hill AVS, Thein SL, Nicholls RD, Teal H, et al. (1986) Proc Natl Acad Sci USA 83: 5165-5169
Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Nature 354: 204-209
Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JAL (1994) Nature Genetics 6: 136-145
Li CC (1976) First Course in Population Genetics, Pacific Grove, California, p.16
Long JC, Williams RC, Urbanek M (1995) Am J Hum Genet 56: 799-810
Martinson JJ, Boyce AJ, Clegg JB (1994) Am J Hum Genet 55: 513-525
Neil DL, Jeffreys AJ (1993)Hum Mol Genet 2: 1129-1135
Renges H-H, Peacock R, Dunning AM, Talmud P, Humphries SE (1992) Ann Hum Genet 56: 11-13