GENETIC SUBSTRUCTURE OF THE ITALIAN POPULATION AT THE VNTR LOCI D1S80 AND D17S30: THE TUSCAN REGION

B. Porfirio, T. Ehret, I. Chiarelli, P.L. Mattiuz
Unit of Human Genetics, Department of Clinical Physiopathology, University of Florence, Viale Morgagni 85, 50134 Florence, Italy

Introduction

Hypervariable loci within the human genome are useful tools in several disciplines. Because of their high informational content, they resurrected genetic linkage studies in families (Nakamura et al. 1987) and served extensively in forensic science (Reynolds et al. 1991). It has been estimated that each human genome is heterozygous at $3x10^6$ nucleotide sites. Consequently, it is theoretically possible the unique identification of individuals from their DNA.

However, data concerning the extent of ethnic variation at such highly polymorphic loci as VNTR is still limited. In fact, ethnic groups defined on the basis of geography or national boundaries can still harbor substantial genetic variation even among their subpopulations (Lewontin and Hartl, 1992).

The present study describes the population genetic characteristics of the D1S80 and the D17S30 VNTR loci in a well defined population. We performed PCR-based DNA typing to address whether the genotype distribution at these loci were in accordance with their Hardy-Weinberg equilibrium (HWE) expectations. Several approaches were pursued and compared one each other to evaluate statistically our sample of Tuscan ancestry. Moreover, the allele frequency distributions were compared with published Caucasian databases to assess the presence of population substructure.

Materials and Methods

Subjects. DNA samples from 100 unrelated individuals were tested. They were residents of the Florence area and all their four grand-parents were born in the Tuscan region.

PCR conditions (GeneAmp PCR System 9600).

Locus	D1S80	D17S30
$MgCl_2$	1.5mM	1.2mM
Input DNA	100ng	10ng
Cycles	25	30
Denaturation	20s 94°C	20s 94°C
Annealing	20s 67°C	20s 56°C
Extension	30s 72°C	60s 72°C

Electrophoresis. PCR products of the D1S80 locus were run by vertical PAGE (8% T, 3.3% C, 1mm thick) and stained with ethidium bromide. Alleles at the D17S30 locus were studied after 2% agarose gel electrophoresis and ethidium bromide staining.

Statistics. Tests for HWE from the genotype distributions were carried out by computing the near normally distributed variable z defined as the deviation of $V\bar{x}^2$ from its expected value m divided by its standard error $V\bar{v}$ (Smith, 1986). The same method was also applied to a contingency table to test for heterogeneity between the allele distributions of our Tuscan population and those of published Caucasian databases.

Additional tests for HWE were conducted by using the summary statistics total heterozygosity as well as by studying the congruence between the observed and expected levels of heterozygosity and number of alleles (Chakraborty et al. 1988).

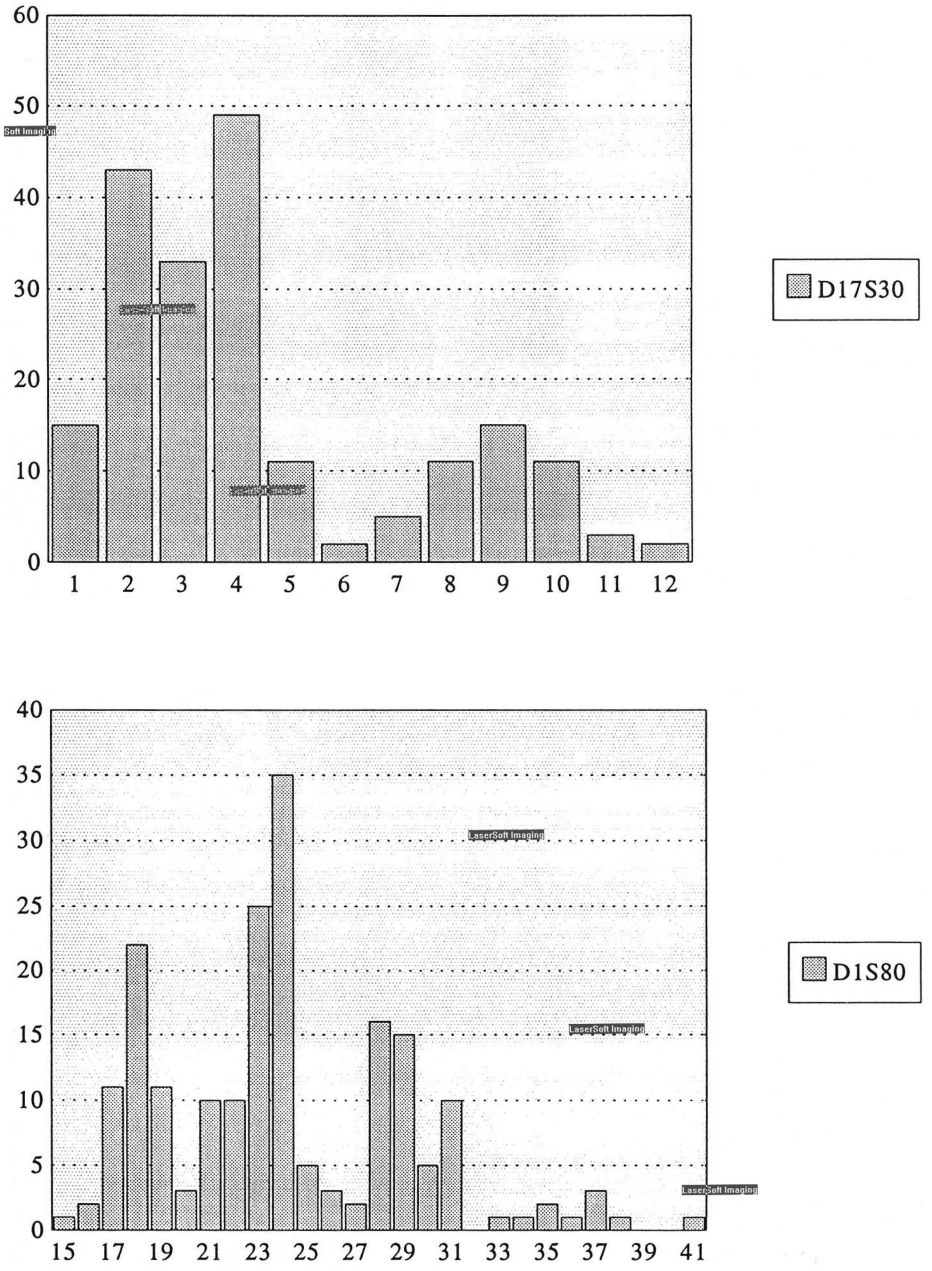

Figure 1. Frequency distribution of D17S30 and D1S80 alleles.

Results

Allele frequency distributions and data analysis are shown in Figure 1 and in Table 1, respectively.

D17S30. Twelve different alleles were observed in 100 unrelated Tuscans. The length of the alleles ranged from 160 to 930 bp. Their designation #1 through #12 reflects the number of repeats.

Out of 78 possible genotypes, given 12 observed alleles, 37 were detected in 100 individuals. Genotypic frequencies were consistent with the assumptions of the HWE. Allelic counts were homogeneous to those found in a Caucasian sample reported by Batanian et al. (1990). The marginal departure from HWE in the heterozygosity test accounts for the slight heterozygote excess found at this locus.

D1S80. Twenty-four different alleles were found in 99 DNA samples (one less than those originally included in the present study, because of an accidental loss before amplification).

The length of the alleles ranged from 387 to 803 bp. Their designation #15 through #41 reflects the number of repeats. The frequency distribution of the D1S80 alleles is multimodal: the most common allele is #24; two additional modes are located approximately 6 repeat units both left- and rightward; a much less frequent mode could also be present 6 repeat units further.

A large excess of single-banded phenotypes is present as judged by the heterozygosity test ($p<10^{-6}$). However, the genotypic distribution was in accordance with HWE expectations. Heterogeneity test was higly significant when comparing the allele frequencies with published Caucasian databases (e.g., Budowle et al. 1991) or with Italian data (Spinella, 1993).

Table 1. Synopsis of the statistical approaches

		D1S80 (n=99)	D17S30 (n=100)
TESTS FOR HWE			
GENOTYPIC DISTRIBUTION			
Standard pooling strategy[a]	X^2	2.54	6.18
	d.f.	3	8
	p	<0.5	<0.75
Chi-squared with small numbers (Smith, 1986)	*z*	1.47	0.38
	p	<0.07	<0.34
HETEROZYGOSITY TEST	obs	71	91
	exp	90.2	83.8
	X^2	45.65	3.86
	p	$<10^{-6}$	<0.05
ALLELIC DISTRIBUTION			
Heterogeneity test[b] (Smith, 1986)	*z*	7.63	0.67
	p	<0.001	<0.25
COMPARISONS BASED ON MUTATION-DRIFT MODELS			
E(*h*) = expected heterozygosity		0.92	0.85
k = observed number of alleles		24	12
E(*k*) (Ewens, 1972)		32.5	20.7

[a] Genotypes with expectations less than 5 were pooled.

[b] Comparisons were made with data from Budowle et al. (1991) for D1S80 and from Batanian et al. (1990) for D17S30.

Discussion

The PCR has provided a simple and rapid method for studying many polymorphisms, including VNTR. Since optimized PCR-based protocols readily detect one or two bands in DNA samples from homozygotes and heterozygotes, the allele frequency estimates were obtained simply by gene counting. Thus, the excess of single-banded phenotypes observed at the locus D1S80 should be regarded as indicating an excess of homozygotes in our sample of Tuscans. As a likely explanation it should be considered that the expected number of heterozygotes in the HWE hypothesis depends on observed allelic frequencies with relatively high sampling errors. On the other hand, the highly significant difference between the allele frequency distribution of D1S80 and that of some Caucasian databases (Budowle et al. 1991; Spinella, 1993) is apparently due to the presence of different "private" alleles in the different samples. However, it could be put forward that comparisons of data gathered through different methodologies may be misleading. For instance, the use of either ultra-thin horizontal or standard vertical PAGE vs. agarose gel analysis could have introduced a serious bias in the allocation of individual alleles at the 16-bp VNTR locus D1S80. It seems to us that previous PCR databases of VNTR loci reflected the binning attitude of the RFLP era. Moreover, the development of allelic ladders has undoubtedly improved the resolution of discrete alleles at the D1S80 locus. These two latter considerations may help to explain why the number of alleles is lower in early works.

The results of the present study provide an estimate of genic diversity in a well defined population. Other regionally founded VNTR databases could be useful in assessing the extent of population substructure.

References

Batanian JR, Ledbetter SA, Wolff RK, Nakamura Y, White R, Dobbins WB, Ledbetter DH (1990) Rapid diagnosis of Miller-Dieker syndrome and isolated lissencephaly sequence by the polymerase chain reaction. Hum Genet 85: 555-559.

Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC (1991) Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am J Hum Genet 48: 137-144.

Chakraborty R, Smouse PE, Neel JV (1988) Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Am J Hum Genet 43: 709-725.

Ewens WJ (1972) The sampling theory of selective neutral alleles. Theor Popul Biol 3:87-112.

Lewontin RC and Hartl DL (1992) Forensic DNA typing. Science 255: 1054-1055.

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235: 1616-1622.

Reynolds R, Sensabaugh G, Blake E (1991) Analysis of genetic markers in forensic DNA samples using the polymerase chain reaction. Anal Chem 63: 1-15.

Smith CAB (1986) Chi-squared tests with small numbers. Ann Hum Genet 50: 163-167.

Spinella A, Italian Scientific Police, Rome (1993) personal communication.