PCR TYPING OF THE COL2A1 SYSTEM: ALLELIC FREQUENCIES IN TWO POPULATION SAMPLES FROM NORTH AND CENTRAL ITALY

*Tagliabracci A., Buscemi L., Cucurachi N.,*Paoli M., Sisti B. and Ferrara S.D.
*Istituto di Medicina Legale e delle Assicurazioni, Università di Ancona, Italy.
Istituto di Medicina Legale e delle Assicurazioni, Università di Parma, Italy.

INTRODUCTION

The COL2A1 system is a VNTR in the 3'- flanking region of the human Type II collagen gene. It was first sequenced by Stoker et al. (1985), who pointed out the polymorphism was due to repetitive sequences of 31 or 34 bp. Mendelian inheritance and polymorphism were investigated at the same time with the polymerase chain reaction (PCR) technique by Wu et al. (1990) and Priestley et al. (1990). Using different primers these authors found four and five alleles respectively.

Further alleles were recently discovered by analysing the amplified products with the more sophisticated high-resolution polyacrylamide gel electrophoresis (PAGE) followed by silver staining. Puers et al. (1992) discovered various alleles whose sizes differed by less than the expected 31 or 34 bp, down to 3 bp. Even more recently Berg and Olaisen (1993) sequenced 19 allelic variants studied by denaturing PAGE, and demonstrated that the polymorphism is due to: 1) a different total number of copies of the core repeat units; or 2) the same total number of copies but containing a different proportion of core repeat units of 34 or 31 bp; or 3) small deletions in some core units.

Population studies to assess the number and the frequency of alleles were published by Priestley et al (1990), Puers et al. (1992), Rand et al. (1992), Berg and Olaisen (1993). Further population studies were performed by Wiegand et al. (pers. comm.) and Hou et al. (pers. comm.), the latter demonstrating allelic variations among different racial groups. The aim of this study was to investigate the polymorphism of this system to verify a) the number of alleles and b) the allelic frequency distribution in a sample of Italians, and c) to compare the results with other Caucasian samples.

MATERIALS AND METHODS

Fresh blood samples were collected from healthy unrelated donors living in Ancona (n = 117) and Parma (n = 99). The two laboratories performed the analysis independently, following the same procedures for DNA extraction (Budowle and Baechtel, 1990), amplification and identification of phenotypes.

Amplification was carried out in a thermal cycler PTC 100-60 and a Minicycler (MJ Research, USA) with primers proposed by Wu et al. (1990): 5'-CCAGGTTAAGGTTGACAGCT-3'; 5'-GTCATGAACTAGCTCTGGTG-3'
for 25 cycles at the following conditions: denaturation at 94°C 1 min; annealing at 64°C 1 min; extension at 70°C 2 min. The amplification mix contained 50 ng DNA template, 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl2, 0.1% Triton X-100, 1 U (0.2 µl) Taq polymerase (Promega Corporation, USA), 0.4 µM each primer, 200 µM each nucleotide, diluted to a total volume of 25 µl and overlayed with 30 µl oil.

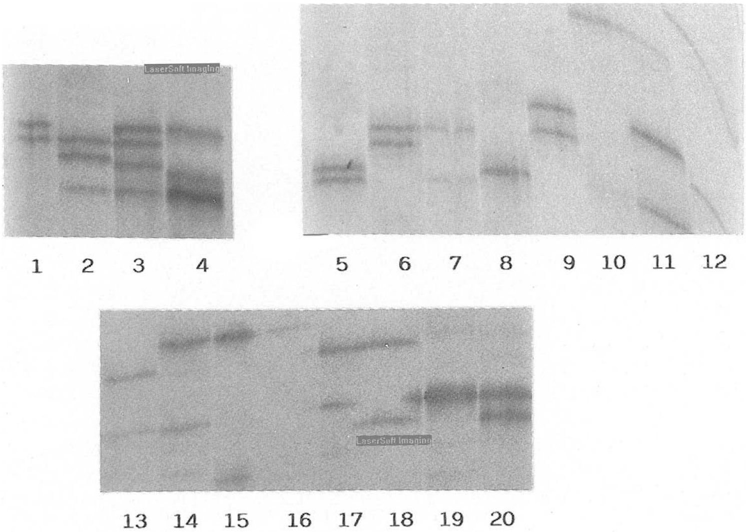
Separation of amplified products was carried out in PAGE (Budowle et al. 1991) at 5% T, 3% C with PDA as cross-linker, thickness 450 µm, using a discontinuous buffer system (Allen et al. 1989). The bands were visualized with silver staining (Budowle et al. 1991), as shown in Fig. 1. Alleles were identified by side-to-side comparison with reference alleles, isolated or mixed in a cocktail obtained from the amplified products.

The results were verified by the chi-square test between observed and expected genotypes under the Hardy-Weinberg law and by comparing the observed and expected heterozygosity frequencies, the latter calculated as allelic diversity ($H = \left[1 - \sum X_{1}^{2}\right]\left[n / n - 1\right]$, where X_i is the allelic frequency and n the sample size (Nei and Roychoudhury, 1974). $\sqrt{\left[h(1-h)/N\right]}$ is the formula to compute the standard error for H, were h is the expected heterozygote frequency and N the number of subjects examined.

The power of discrimination (PD) was calculated using the equation proposed by Fischer (1951): $1-\sum(Pi)^2$, where Pi represents the frequency of each genotype. The exclusion chance was calculated from allele frequencies (Garber and Morris 1983).

The allelic frequencies were statistically compared with previous studies to recognize possible intraethnic variations by a computer program kindly supplied by G. Carmody (Carleton University, Ottawa, Canada).

Fig. 1. COL2A1 amplified products after PAGE and silver staining. Samples: (1) type 5-6; (2) ladder composed of alleles 6,7,10; (3) ladder composed of alleles 5,6,8,10; (4) ladder composed of alleles 5,9,10; (5) type 9-10; (6) type 6-7; (7) type 6-10; (8) type 9-9; (9) type 5-6; (10) type 1A1-9; (11) type 5-10; (12) 123 bp weigth marker BRL; (13) type 1-5; (14) type 1A1-5; (15) type 1A1-9; (16) type 1A1-1A1; (17) type 1-5; (18) type 1-6; (19) type 5-5; (20) type 5-6.



RESULTS AND DISCUSSION

Ten alleles were found in a sample of 216 Italians and named according to the Munster classification and ISFH recommendations. Table 1 shows allelic frequencies and the results of statistical analysis, which demonstrate no deviations from Hardy-Weinberg expectations and from the expected heterozygote rate. The most frequent alleles were COL2A1*10, COL2A1*6, COL2A1*5 and COL2A1*1, in line with the frequencies found in previous studies on Caucasians performed with the same procedures(Wiegand et al., pers. comm.; Prinz et al., pers. comm.). Even if the allelic frequencies follow the same trend, the specific values differ for each allele to the extent that heterogeneity does in fact occur between the Italian and German samples. The same test was not applied to the data of Berg and Olaisen because they used different analytical approaches and nomenclature.

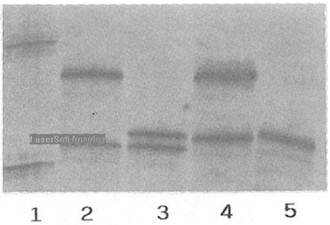
The high PD and chance of exclusion found in this study make this system a powerful tool for forensic identification and paternity testing.

One of the most relevant problems which affect this VNTR is the presence of alleles differing from each other by few bases and not by a sequence repeat unit. In our study we found slight disalignment between some bands, which could be explained only by a difference of a few bp (Fig. 2). Such alleles can only be identified by sequencing (Berg and Olaisen, 1993) since PAGE does not distinguish clearly between similar alleles. Further studies of the analytical methods to improve the resolution of alleles and to make a new nomenclature are desirable.

Table 1. Distribution of COL2A1 alleles in 216 Italians

Alleles	Frequencies	Statistical evaluations
1A1	0.0185	$\chi^2 = 36.57 \quad 0.75 < P < 0.90$ (45 d.f.)
1	0.0787	
5	0.2477	Heterozygosity = 0.810
6	0.2616	Expected heterozygosity = 0.785±0.027
7	0.0301	
8	0.0208	PD = 0.91
9	0.0185	Chance of exclusion = 0.58
10	0.2778	
12	0.0324	Italians/Germans χ^2=57.30 P< 0.0000
15	0.0139	G statistic = 60.2347 P< 0.0000

Fig. 2. Magnification of COL2A1 amplified products after Page and silver staining. Samples: (1) 123 bp weigth marker BRL; (2) type 1-6; (3) type 5-6; (4) type 1-5; (5) type 5-5. Allele 5 in lane (4) is slightly displaced with respect to the same adjacent alleles.



REFERENCES

Allen RC, Graves G, Budowle B (1989) Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver. Bio Tech 7: 736-744

Berg ES, Olaisen B (1993) Characterization of the COL2A1 VNTR polymorphism. Genomics 16: 350-354

Budowle B, Baechtel FS (1990) Modifications to improve effectiveness of restriction fragment length polymorphism typing. Appl Theor Electrophor 1: 181-187

Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC (1991) Analysis of the variable number of tandem repeats locus D1S80 by the polymerase chain reaction followed by high resolution polyacrylamide gel electrophoresis. Am J Hum Genet 48: 137-144

Fisher RA (1951) Standard calculations for evaluating a blood group system. Heredity 5: 95-102

Garber RA, Morris JW (1983) General equations for the average power of exclusion for genetic system of n codominant alleles in one-parent and no-parent cases of disputed parentage. In: Walker RH (ed) Inclusion probabilities in parentage testing. AABB, Arlington, VA, pp 277-280

Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. Genetics 76: 379-390

Priestley L, Kumar D, Sykes B (1990) Amplification of the COL2A1 3' variable region used for segregation analysis in a family with Stickler syndrome. Hum Genet 85: 525-526

Puers C, Wiegand P, Brinkmann B (1992) Investigations to improve allele definition in the "collagen 2A1" system (AMP-FLP). In: Rittner C, Schneider PM (eds) Adv for haemogen 4, Springer Verlag, Berlin Heidelberg, pp 32-34

Rand S, Puers C, Skowasch K, Wiegand P, Budowle B, Brinkmann B (1992) Population genetics and forensic efficiency data of 4 AMPFLP's. Int J Leg Med 104: 329-333

Stoker NG, Cheah HS, Griffin JR, Pope FM, Solomon E (1985) A highly polymorphic region 3'to the human Type II collagen gene. Nucleic Acids Res 13: 4613-4622

Wu S, Seino S, Bell GI (1990) Human collagen, type II, alpha 1, (COL2A1) gene: VNTR polymorphism detected by gene amplification. Nucleic Acids Res 18: 3102