Hum Genet (1996) 97 : 309–313 

ORIGINAL INVESTIGATION

**Fabrício R. Santos · Tudevdagva Gerelsaikhan**
**Bjamba Munkhtuja · Tsendsuren Oyunsuren**
**Jörg T. Epplen · Sérgio D. J. Pena**

# Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19

Received: 6 July 1995

**Abstract** We have studied the allele frequency distribution of the microsatellite locus DYS19 in several populations with different geographical origins worldwide. Three new alleles were found. In addition, remarkable geographic and ethnic differences were observed in the allele frequency profiles and DNA marker (gene) diversity among populations and major ethnic groups. Amerindians showed an overwhelming predominance of the A allele, while in Caucasians the B allele was modal, and in Greater Asians and Africans allele C became predominant. Even within these geographic regions there were significant gradients, as exemplified by the decreasing frequency profile of the B allele from Great Britain over Germany to Slovakia. Thus, DYS19 emerges as a useful tool for studying the structure and dynamics of human populations.

## Introduction

Y-linked loci are haploid and paternally inherited and, with the exception of genes in the pseudo-autosomal region, there is no recombination (Wolf et al. 1992). Thus, Y chromosomal markers are transmitted together as haplotypes. Therefore, each male individual has the same Y chromosome haplotype as his father, brothers, paternal grandfather, paternal uncles, etc., thus establishing a patrilineage. These characteristics should render Y-linked polymorphisms extremely useful as genetic tools to study human evolution. Unfortunately, so far very few DNA polymorphisms have been described in the human Y chromosome and even fewer are amenable to PCR-based assays (Mathias et al. 1994).

Roewer et al. (1992) reported the first human Y-linked polymorphic microsatellite, DYS19, containing a $(GATA)_n$ motif. We studied this polymorphism in the Brazilian population using a simple non-isotopic procedure based on non-denaturing PAGE followed by silver staining and found five different alleles (A–E) with sizes varying from 186 to 202 bp, respectively. The marker (gene) diversity was estimated as 0.66 (Santos et al. 1993a) and no mutations were detected in 100 father-son pairs. DYS19 has been applied in crime investigations (Roewer and Epplen 1992) and paternity testing (Santos et al. 1993b).

To evaluate the usefulness of DYS19 for human evolutionary studies we investigated individuals sampled from populations with a broad geographic distribution. Here, we describe the detection of new alleles and the finding of striking geographical differences in the distribution of allele frequencies of this microsatellite polymorphism.

F. R. Santos · S. D. J. Pena (✉)
Departamento de Bioquímica,
Instituto de Ciências Biológicas - UFMG, Caixa Postal 486,
Belo Horizonte, MG, Brazil 30161-970
Fax (+5531) 227 2792 or 441-5963
e-mail spena@dcc.ufmg.br

T. Gerelsaikhan · B. Munkhtuja · T. Oyunsuren
Institute of Biotechnology, Academy of Sciences,
Ulaanbaatar -51, Mongolia

J. T. Epplen
Molecular Human Genetics, Ruhr-University,
D-44780 Bochum, Germany

S. D. J. Pena
Núcleo de Genética Médica de Minas Gerais (GENE/MG),
Belo Horizonte, Brazil 30130-909

## Materials and methods

### DNA samples

We studied a total of 317 human DNA samples. Ninety-one specimens, including 54 Caucasians, 18 Africans, 10 Asians, 5 Oceanics and 4 individuals of unknown origin, were a gift from C. Tyler-Smith (Oxford University, UK) and have been described in detail elsewhere (Mathias et al. 1994). Nine Amerindian samples were purchased from the Coriell Institute of Medical Research (Camden, N.J., USA; NIGMS Human Diversity Collection). Seventeen samples of African pygmies from Zaire and Central African Republic were a gift from L. L. Cavalli-Sforza (Department of Genetics, Stanford University, Calif., USA). Finally, 48 samples were collected from Mongolian populations and an extended sample of 152 Brazilians was randomly chosen from males undergoing pa-

ternity testing at the Núcleo de Genética Médica de Minas Gerais (GENE/MG).

### PCR and electrophoresis

PCR reactions were performed as described by Santos et al. (1993 a) without multiplexing with D12S67. The products were resolved in 16-cm-long (1-mm-thick) 10% polyacrylamide gels and visualized by silver staining as described previously (Santos et al. 1993 a). We modified the method for silver staining by omitting the rinsing step with water and introducing a change of the developer solution after darkening.

### Cloning and sequencing

PCR products were cloned using either the Sure Clone kit (Pharmacia) or the T-A Cloning Kit-R (Invitrogen). Sequencing was performed either using an automatic fluorescent laser sequencer (Pharmacia A.L.F.) and the AutoRead kit or manually using $\alpha$-[$^{35}$S]dATP (Amersham) and the Sequenase 2.0R kit (US Biochemicals).

### Estimation of genetic diversity

DNA marker (gene) diversity was calculated for each population using the formula $D = 1 - \Sigma\, i^2$, where i represents the allelic frequency (Chakraborty 1985).

## Results

### Novel alleles

In our survey we discovered the existence of three new alleles. One of them, called Z, was 182 bp long and was seen in the only Bushman !Kung studied. It was sequenced and shown to differ from the known alleles for the DYS19 locus (GenBank accession number X77751) only in the number of GATA repeats. In a Brazilian individual, we found three DYS19 PCR products simultaneously, corresponding to alleles B, C, and E. Also, his son carried all three alleles, ruling out the possibility that the apparent locus triplication was a typing artifact (sample mixture) or a karyotypic alteration. After restriction enzyme digestion, hybridization experiments revealed an approximately threefold signal intensity compared to other males using equal amounts of genomic DNA (data not shown), thus supporting the hypothesis of locus triplication. The best explanation is that a region of more than 1 kb, including the microsatellite, was triplicated. Apparently, afterwards, the three copies have diverged.

The third allele was inferred to exist in an Indian individual from whom DYS19 could not be amplified by PCR in several attempts. The presence of the Y chromosome was confirmed by PCR amplification of DYZ1 on the long arm, of alphoid repeats in the centromeric region (Santos et al. 1995), and of a pseudogene adjacent to the *sry* region in the short arm. In addition, several autosomal microsatellites demonstrated normal amplification. On this basis we postulated the existence of a null allele, which could have arisen by a deletion of the DYS19 locus or a mutation in a primer annealing site.

### Allele frequencies in different populations

In order to generate more reliable data on the allele representation of DYS19 in different populations, we pooled our data with those previously described in the literature. To group the populations we followed the nomenclature of Nei and Roychoudhury (1993): four major subdivisions of the phylogenetic tree of human populations are represented by Africans, Caucasians, Greater Asians, and Amerindians (Table 1). Of course, these groups still exhibit considerable internal heterogeneity. We computed allele frequencies and DNA marker diversity values for all groups and subgroups harboring at least 20 individuals. The identification of the geographical origin of the groups and subgroups is shown in Table 1. A large sample of Brazilian Caucasians was separated from other Caucasians because of its heterogeneity, since it apparently contains a significant component of Amerindian, African, Middle East, and Japanese gene flow on a European background. The totalled DYS19 allele frequencies are depicted in Fig. 1 for the four major geographical groups.

In this study we analyzed nine Amerindians from five different groups (Maya, Auca, Quechua, Suruí, and Karitiana), speaking four distinct languages, and with wide geographical origins, ranging from Mexico to Brazil. The allele A (186 bp) was identified in all of them. These results match those obtained on 11 Yanomamis, 10 of which carried the A allele (Roewer et al. 1993). Thus, overall, the A allele had a frequency of 0.95 and the gene diversity of DYS19 in Amerindians was only 0.09. The Amerindian A allele had identical DNA sequence to the European A allele.

Among Caucasians, individuals from Europe, the Middle East, India, and Pakistan were included. The DNA marker diversity for this group was significantly higher than Amerindians, reaching 0.72; the B allele (190 bp) was the mode (frequency 0.40). Interesting trends could be observed within this group. Allele B was frequent in Great Britain (0.80), less abundant in Germany (0.44) and even rarer in Slovakia (0.21). Conversely, allele D grew from a frequency of 0.02 in Great Britain, to 0.19 in Germany and 0.33 in Eastern Europe. Obviously, additional detailed studies are needed for confirmation of these initial results. In particular, the data from Great Britain are remarkable, with a high incidence of the B allele and significantly reduced DNA marker diversity (0.34).

In the Greater Asian group we pooled all data from populations of Northeast, East and Southeast Asia, as well as Australians and Pacific Islanders (although there were very few samples from the latter two populations). Again, DNA marker diversity was high (0.70). Compared to Caucasians, however, the frequencies of the larger alleles were increased, with allele C becoming modal (0.45). Allele A was in very low frequency (0.04), and now a rare allele (F) made its appearance in two Mongolians. This allele had been described recently in a single Japanese individual (Hammer and Horai 1995).

Africans had a relatively high DNA marker diversity (0.77) and, as was the case in Greater Asians, allele C was

**Table 1** Distribution of DYS19 alleles in several populations

| Ethnic group | Population (*n*) | A 186 bp | B 190 bp | C 194 bp | D 198 bp | E 202 bp | Other alleles |
|---|---|---|---|---|---|---|---|
| Amerindians | Suruí, Brazil (3)[a] | 3 | | | | | |
| | Karitiana, Brazil (2)[a] | 2 | | | | | |
| | Maya, Mexico (2)[a] | 2 | | | | | |
| | Quechua, Peru (1)[a] | 1 | | | | | |
| | Auca, Equator (1)[a] | 1 | | | | | |
| | Yanomami, Venezuela (11)[b] | 10 | 1 | | | | |
| | Total 20 | 19 (0.05) | 1 (0.05) | | | | |
| Caucasians | Europeans, Great Britain (41)[a] | 2 (0.05) | 33 (0.80) | 4 (0.10) | 1 (0.02) | 1 (0.02) | |
| | Europeans, Germany (306)[c, d] | 21 (0.07) | 136 (0.44) | 74 (0.24) | 59 (0.19) | 16 (0.05) | |
| | Europeans, Slovakia (81)[c] | 4 (0.05) | 17 (0.21) | 16 (0.20) | 27 (0.33) | 17 (0.21) | |
| | India (39)[a, d] | | 13 (0.33) | 19 (0.49) | 5 (0.13) | 1 (0.03) | Null 1 (0.03) |
| | Pakistan (39)[d] | | 7 (0.18) | 21 (0.54) | 8 (0.21) | 3 (0.08) | |
| | Iraq (1)[a] | | | 1 | | | |
| | Unknown (8)[a] | 2 | 1 | 3 | 2 | | |
| | Total 515 | 29 (0.06) | 207 (0.40) | 138 (0.27) | 102 (0.20) | 38 (0.07) | Null 1 (0) |
| Greater Asians | China (6)[a] | | | 5 | | 1 | |
| | Mongolia (48)[a] | 1 (0.02) | 16 (0.33) | 14 (0.29) | 13 (0.27) | 2 (0.04) | F 2 (0.04) |
| | Japan (136)[a, e] | 10 (0.07) | 4 (0.03) | 66 (0.49) | 33 (0.24) | 22 (0.16) | F 1 (0.01) |
| | Korea (33)[d] | | 8 (0.24) | 15 (0.45) | 7 (0.21) | 3 (0.09) | |
| | Taiwan (14)[e] | 1 | 4 | 4 | 4 | 1 | |
| | Cambodia (1)[a] | | | 1 | | | |
| | Ahom, Burma (21)[d] | 1 (0.05) | 12 (0.57) | 7 (0.33) | 1 (0.05) | | |
| | Kachari, Burma (28)[d] | | 19 (0.68) | 7 (0.25) | 2 (0.07) | | |
| | Thais, North Thailand (42)[d] | | 3 (0.07) | 29 (0.69) | 6 (0.14) | 4 (0.10) | |
| | Malays, South Thailand (12)[d] | | 5 | 6 | 1 | | |
| | Kampuchea (11)[d] | | 2 | 4 | 5 | | |
| | Melanesia (2)[a] | | | 2 | | | |
| | Australia (3)[a] | 1 | 1 | 1 | | | |
| | Total 357 | 14 (0.04) | 74 (0.21) | 161 (0.45) | 72 (0.20) | 33 (0.09) | F 3 (0.01) |
| Africans | Kenya (14)[a] | | 5 | 7 | 2 | | |
| | Pygmies, Zaire (6)[a] | 2 | 2 | | 1 | 1 | |
| | Pygmies, Central African Republic (14)[a]! | | | 4 | 6 | 4 | |
| | !Kung (1)[a] | | | | | | Z 1 |
| | Total 35 | 2 (0.06) | 7 (0.20) | 11 (0.31) | 9 (0.26) | 5 (0.14) | Z 1 (0.03) |
| Others | Brazilians (252)[a, f] | 41 (0.16) | 135 (0.54) | 59 (0.23) | 14 (0.06) | 2 (0.01) | BCE 1 (0) |
| | Unknown (4)[a] | 1 | 2 | | 1 | | |
| Total | 1183 | 106 (0.09) | 426 (0.36) | 369 (0.31) | 198 (0.17) | 78 (0.07) | 6 (0.01) |

[a] This investigation
[b] Roewer et al. (1993)
[c] Muller et al. (1994)
[d] Gomolka et al. (1994)
[e] Hammer and Horai (1995)
[f] Santos et al. (1993 a)

the mode (0.31). A new allele, Z (182 bp), appeared in the single Bushman !Kung tested. Further studies of this latter population might yield interesting results.

Finally, the large group of Brazilian Caucasians presented an allele profile similar to European Caucasians, although with a much higher frequency of the A allele. Among Brazilians, we found the individual with a triplicated locus with the alleles B, C and E.

## Discussion

The worldwide distribution of allele frequencies of the DYS19 tetranucleotide polymorphism exhibited a remarkable heterogeneity. Although allele frequency profiles for the major groups were consistently unimodal, there was variation of the predominant alleles. Amerindians showed an overwhelming predominance of the A allele, while in Caucasians the B allele was modal, and in Greater Asians and Africans allele C became predominant. Even within these geographic regions there were

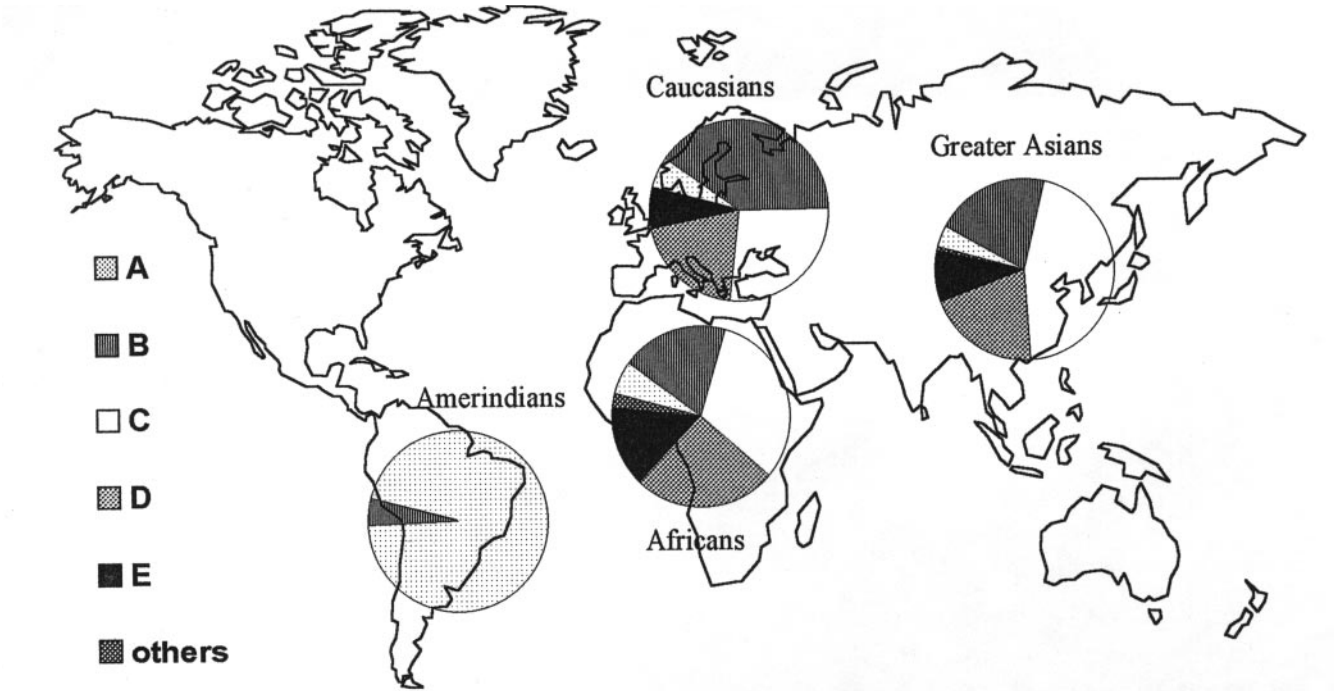


**Fig. 1** Worldwide distribution of DYS19 alleles in the four major subdivisions of man (Nei and Roychoudhury 1993)

significant gradients, as exemplified by the decreasing frequency profile of the B allele from Great Britain over Germany to Slovakia.

These findings were somewhat surprising. Microsatellites are believed to mutate frequently by replication slippage. These mutations would tend to homogenize allele frequency differences between populations. Studies with autosomal microsatellites often show little variation in allele frequencies between populations from diverse geographical origins or between different racial or ethnic groups (Edwards et al. 1992; Gomolka et al. 1994; Hammond et al. 1994). The exceptions are small isolated populations, prone to the effects of genetic drift. Such is hardly the case here, however, because (excepting the Amerindians) the data originate mostly from large outbred populations, subjected to considerable gene flow. Thus, DYS19 appears to be an unusually stable microsatellite. Microsatellites located outside the pseudo-autosomal region of the Y chromosome are regulated by rules different from those governing autosomal ones, since there is haploidy and absence of recombination (Wolf et al. 1992). Therefore, the observed high stability of DYS19 may conceivably reflect a necessity for diploidy in the generation of high microsatellite variability. Previously, we suggested that mechanisms such as gene conversions could play a role in microsatellite mutation, based on the observation of a significant allele size association in haplotypes of two closely positioned microsatellites in intron 40 of the von Willebrandt gene (Pena et al. 1994). Thus, mechanisms that postulate simple replication slippage in the genesis of microsatellite mutations may be too simplistic. Other Y-linked microsatellites should be examined to ascertain if the stability observed on DYS19 is indeed a generalized Y chromosomal phenomenon.

From analysis of our data, DYS19 emerged as a useful tool for studying the structure and dynamics of human populations, despite its simplicity and small number of alleles. The geographic heterogeneity of allele frequencies may give clues to the migratory movements that led to the formation of present-day populations. In particular, the high predominance of a single allele appears relevant, reducing DNA marker diversity in Amerindians and in Caucasians from Great Britain. Although these observations will have to be confirmed with larger samples, in both cases founder effects are possible. Indeed, for Amerindians, a much larger sample was typed simultaneously with DYS19 and the system αh, a new PCR-based Y-linked polymorphism based on sequence variation in alphoid repeats located in the Y centromeric region, typed by heteroduplex analysis (Santos et al. 1995). We could thus identify a major, and perhaps single, ancestral Y chromosomal haplotype in Amerindians, containing the DYS19 A allele (Pena et al. 1995). This fact leads to an interesting paradox. It is believed that the Americas were peopled by Asians, who migrated across a land bridge in Behringia in the Pleistocence (reviewed in Salzano and Callegari-Jacques 1988). However, the data presented in Table 1 indicate that allele A is rare in Greater Asians. In order to reconcile these conflicting data we propose that the Americas were peopled by an Asian subpopulation with a high frequency of the A allele. Extensive population studies using DYS19 in Greater Asians may permit identification of the descendants of this subgroup.

**Acknowledgements** This research was supported by grants-in-aid from Conselho Nacional de Pesquisas (CNPq) and Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), both from Brazil.

## References

Chakraborty R (1985) Paternity testing with genetic markers: are Y-linked genes more efficient than autosomal ones? Am J Med Genet 21 : 297–305

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12 : 241–253

Gomolka M, Hundrieser J, Nurnberg P, Roewer L, Epplen JT, Epplen C (1994) Selected di- and tetranucleotide microsatellites from chromosomes 7, 12, 14 and Y in various Eurasian populations. Hum Genet 93 : 592–596

Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56 : 951–962

Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55 : 175–189

Mathias N, Bayés M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3 : 115–124

Muller S, Gomolka M, Walter H (1994) The Y-specific SSLP of the locus DYS19 in four different European samples. Hum Hered 44 : 298–300

Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. Mol Biol Evol 10 : 927–943

Pena SDJ, Souza KT, Andrade M, Chakraborty R (1994) Allele associations of two polymorphic microsatellites in intron 40 of the human von Willebrandt factor gene. Proc Natl Acad Sci USA 91 : 723–727

Pena SDJ, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B, Oyunsuren T (1995) A major founder Y-chromosome haplotype in Amerindians. Nature Genet 11 : 15–16

Roewer L, Epplen JT (1992) Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. Forensic Sci Int 53 : 163–171

Roewer L, Arnemann J, Spurr NK, Grzeschik K-H, Epplen JT (1991) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 88 : 389–394

Roewer L, Nagy M, Schmidt P, Epplen JT, Herzog-Schroder G (1993) Microsatellite and HLA class II oligonucleotide repeat polymorphism. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) DNA fingerprinting: state of the science. Birkhauser Verlag, Basel, pp 221–230

Salzano FM, Callegari-Jacques SM (1988) South American Indians. Clarendon Press, Oxford

Santos FR, Pena SDJ, Epplen JT (1993 a) Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum Genet 90 : 655–656

Santos FR, Epplen JT, Pena SDJ (1993 b) Testing deficiency paternity cases with a Y-linked tetranucleotide repeat polymorphism. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) state of the science. Birkhauser Verlag, Basel, pp 261–265

Santos FR, Pena SDJ, Tyler-Smith C (1995) PCR haplotypes for the human Y chromosome based on alphoid satellite variants and heteroduplex analysis. Gene 165 : 191–198

Wolf U, Schempp W, Scherer G (1992) Molecular biology of the human Y chromosome. Rev Physiol Biochem Pharmacol 121 : 147–213