Brief Communication

Investigating SNPs Flanking the D1S80 Locus in a Tamil Population from India

KUPPAREDDI BALAMURUGAN,[1] ROBERT POMEROY,[2] GEORGE DUNCAN,[3] AND MARTIN TRACEY[4]

Abstract *D1S80* is a 16-bp variable number of tandem repeats minisatellite. We analyzed single nucleotide polymorphisms (SNPs) flanking this locus in a Tamil population. Alleles ranged from 15 through 41 repeats, with alleles 18 and 24 being predominant with frequencies of 31% and 34.5%, respectively, suggesting a bimodal allelic distribution. All the 18-repeat alleles are associated with *Hin*fI(+) and *Fnu*4HI(–) restriction site polymorphisms at the 5′ and 3′ ends, respectively. Allele 24 is associated with *Hin*fI(–) and *Fnu*4HI(+). Of the alleles tested, 98.5% have a linkage of two specific SNP polymorphisms. If an allele is positive for *Hin*fI, then it is negative for *Fnu*4HI, and if an allele is negative for *Hin*fI, it is then positive for *Fnu*4HI, which demonstrates strong linkage disequilibrium between the two polymorphic SNPs. This suggests that reciprocal crossover is not involved in changes in the number of repeats, as few exchanges are seen in the flanking regions. The repeat allele–SNP association might be involved with the internal structure of the locus micropolymorphisms, possibly a double-strand break hotspot.

Minisatellites are composed of repetitive elements ranging from 9 to 100 base pairs (bp) in length. Although the function of these regions is not well understood, minisatellites have been used in chromosome mapping, linkage analysis, studies of oncology-related deletions (White et al. 2005), and forensic profiling (Balamurugan et al. 2001; Budowle et al. 1991). The minisatellite *D1S80* (1p35–p36) is a variable number of tandem repeats (VNTR) locus with a 16-bp repeat (Kasai et al. 1990). Deka et al. (1994) suggested that this highly polymorphic locus might be useful for evolutionary studies; consequently, we examined a

[1]School of Criminal Justice, 118 College Drive #5127, University of Southern Mississippi, Hattiesburg, MS 39406.
[2]Department of Chemistry, University of California–San Diego, La Jolla, CA 92093.
[3]DNA Laboratory, Broward County Sheriff's Office, 201 SE 6th Street, Fort Lauderdale, FL 33301.
[4]Department of Biology, Florida International University, University Park Campus, Miami, FL 33199.

Human Biology, April 2010, v. 82, no. 2, pp. 221–226.

KEY WORDS: *D1S80* LOCUS, *HIN*FI-*FNU*4HI POLYMORPHISM, MINISATELLITE, SNP, TAMIL POPULATION, LINKAGE DISEQUILIBRIUM.

Tamil population for *D1S80* repeat polymorphisms and for two flanking-region single nucleotide polymorphisms (SNPs) to look for haplotypes that might help us shed some light on mutation and evolution at this locus. Previously, we reported *Hin*fI and *Bso*FI(*Fnu*4HI) restriction site SNPs in the flanking regions of the *D1S80* locus (Duncan et al. 1996).

Allele 18 is strongly associated with a 5′ flanking *Hin*fI(+) site and a 3′ flanking *Fnu*4HI(–) site; allele 24 is associated with *Hin*fI(–) and *Fnu*4HI(+) sites (Alonso et al. 1995; Duncan et al. 1996, 1997). We refer to these polymorphisms as (+, –) and (–, +), respectively. The repeat-SNP association might be involved with the internal structure of the locus micropolymorphisms, possibly by a double-strand break hotspot (Buard et al. 2000; Jeffreys et al. 1994). Instability is conferred on the minisatellite array by flanking DNA elements, which introduce staggered nicks and double-strand breaks in the beginning of an array (Buard and Jeffreys 1997; Jeffreys et al. 1994). Although Jeffreys et al. (1994) reported mutational polarity at the 5′ end of the repeat array, we have observed that for the *D1S80* minisatellite the repeat structure differences are seen within the center of the array while the beginning and end of the array remain stable (Duncan et al. 1997). An important advantage of SNPs is their generally low mutation rates [10^{-8} to 10^{-10} for SNPs compared to 10^{-3} to 10^{-5} for short tandem repeat (STR) polymorphisms] (Butler et al. 2006; Chakraborty et al. 1999).

Materials and Methods

Collection of Samples and DNA Isolation. After obtaining institutional review board approval and informed consent, we collected whole blood samples from 100 unrelated Tamil volunteers residing in Chennai, Tamil Nadu, India. No personal data or other information, such as subgroups or caste affiliation, was collected. DNA was extracted following the method of Miller et al. (1988) and quantitated using a Quantifiler Human DNA Quantitation Kit (Applied Biosystems, Foster City, California). For comparison, Rwandan samples (n = 100) were provided by Rene Herrera (Department of Biology, Florida International University, Miami). All samples were collected in accordance with ethical guidelines of the institutions involved.

Identification of SNPs by Restriction Digestion Analysis. The *D1S80* locus was amplified using the forward and reverse primers described by Kasai et al. (1990). Appropriate restriction enzymes were used for restricting the amplicons, and the digests were run on 5% acrylamide gel to separate the alleles, followed by ethidium bromide staining.

Identification of SNPs Using TaqMan Probe. SNP analysis was performed with real-time PCR using a TaqMan SNP Genotyping Assay kit and custom TaqMan assay probe from Applied Biosystems. The data were collected on a Corbett Rotor-Gene 3000 (Corbett Robotics Inc., San Francisco, California). The

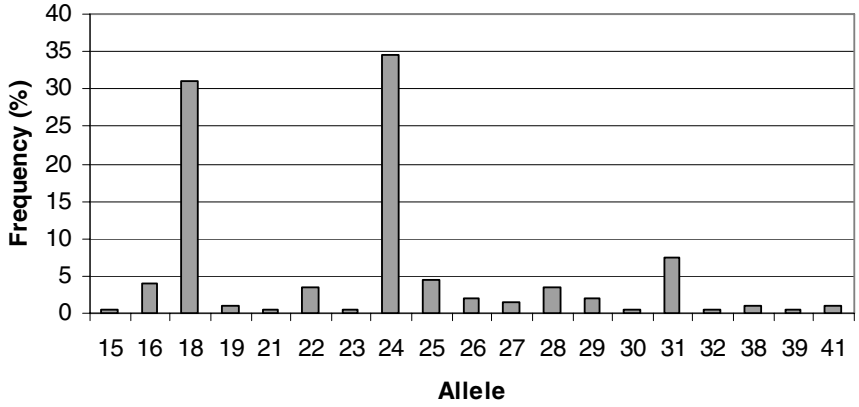

Figure 1. The *D1S80* allele-frequency distribution in a Tamil population. Alleles 18 and 24 are predominant, with frequencies of 31% and 34.5%, respectively, indicating a bimodal distribution.

polymorphisms identified by real-time PCR were confirmed by restriction digest analysis of all alleles.

Results and Discussion

The distribution of the *D1S80* alleles ranged from 15 to 41 repeats, with alleles 18 (31%) and 24 (34.5%) predominating, resulting in a bimodal distribution (Figure 1). All the 18-repeat alleles were found to be positive for the *Hin*fI polymorphism, and 16% of the samples with allele 24 were positive for the *Hin*fI polymorphism and 84% of them were negative for this SNP (Figure 2). Allele 18 is strongly associated with the 5′ flanking *Hin*fI(+) restriction site and the 3′ flanking *Fnu*4HI(–) restriction site. Allele 24 is associated with flanking *Hin*fI(–) and *Fnu*4HI(+) sites. Of the 200 alleles studied, 62.5% were *Hin*fI(+) and 37.5% were *Hin*fI(–). The interesting finding in this study is that 98.5% (197/200) of the SNP haplotypes were either (+, –) or (–, +); only three (–, –) haplotypes were found, and no (+, +) haplotypes were observed (Table 1). When the haplotypes were analyzed for a Rwanda (African) population, all 189 alleles analyzed had either (+, –) or (–, +) haplotypes (Table 1), confirming that these two SNPs are in strong linkage disequilibrium.

These minisatellite allele-flanking SNP associations suggest that crossing over is either nonexistent or rare within the *D1S80* locus. All the primate alleles we have examined (for 11 macaques, 7 orangutans, 1 gorilla, 3 chimpanzees, and 1 bonobo) are *Hin*fI(+)-*Fnu*4HI(–) (unpublished data) and all the 18-repeat alleles in humans share this flanking SNP haplotype. Because all 23 of the primates typed share the allele 18 (+, –) SNP haplotype, we believe that SNP mutations probably occurred in the human lineage.

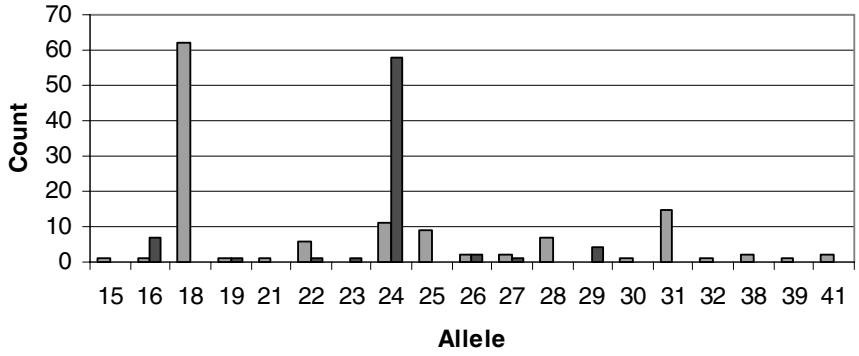

Figure 2. Haplotype frequency distribution of the 5′ flanking region *Hin*fI restriction site polymorphism in a Tamil population. Light bars represent alleles with *Hin*fI(+), and dark bars represent the *Hin*fI(–) polymorphisms.

If we assume that the (+, –) haplotype is ancestral, then we can estimate the age, in generations, of the (–, +) haplotype, which is the second most common haplotype in human populations. Our estimate of the frequencies are 0.625 for the *Hin*fI(+)-*Fnu*4H I(–) haplotype and 0.36 for the *Hin*fI(–)-*Fnu*4HI (+) haplotype. Assuming a rate of change c of 1 in 1,000,000, which includes all mutational mechanisms, we can estimate the age of the (–, +) haplotype as

$$P_{(-,+)} = 0.36 = (1 - c)^g = (0.999999)^g, \quad (1)$$

where g = 1,021,650 generations. Because the SNPs are in total linkage disequilibrium at a frequency of $1/(2N)$ when they occur, this estimate seems reasonable.

For ancestral primates the generation time is estimated to be 10 years (Gibbons 2008). If this is approximately correct, then it suggests that the (+, +) and (–, –) haplotypes occurred far more recently. We cannot resolve these discrepancies in mechanistic terms with our available data, but the haplotype distributions suggest

Table 1. Distribution of *Hin*fI and *Fnu*4HI Polymorphism Haplotypes Flanking the *D1S80* Locus in Tamil (Indian) and Rwandan (African) Populations

	Count	
Haplotype	*Tamil*	*Rwandan*
*Hin*fI(+)-*Fnu*4HI(+)	0	0
*Hin*fI(–)-*Fnu*4HI(–)	3	0
*Hin*fI(+)-*Fnu*4HI(–)	125	101
*Hin*fI(–)-*Fnu*4HI(+)	72	88
Total alleles analyzed	200	189

two things. First, the use of haplotypes provides more genetic variability than the use of single SNPs. The *Hin*fI and *Fnu*4HI SNPs we have described are biallelic; so the maximum heterozygosity is 50%. If, on the other hand, we consider SNP haplotypes such as the *Hin*fI-*Fnu*4HI flanking haplotypes, we have four possible alleles [(+, +), (–, –), (+, –), and (–, +)] with an expected maximum heterozygosity of 75%, a 50% increase. This increases the expected maximum heterozygosity and the value of the locus for identification purposes and for studies of locus evolution. Second, the (+, –) and (–, +) haplotypes are common; they exceed 30% in both the Tamil and Rwandan populations. But the minisatellite allele-frequency distributions are rather different within the SNP haplotypes. For example, allele 18 has a frequency of 31% in the Tamil population and 9% in the Rwandan population when only the (+, –) haplotype is considered. Allele 24 has a frequency of 28% in the Tamil population and 12% in the Rwandan population when only the (–, +) haplotype is considered. We would expect the allele-frequency distribution to reconstitute itself within the SNP haplotypes because of the higher mutation rate for minisatellites. In other words, the SNP mutation occurred on a chromosome bearing a particular *D1S80* allele. Over time we expect the allele-frequency distribution to be restored based on mutation and other possible constraints. It is unlikely, given the age of this locus, that mutational equilibrium is not approximated; so we suggest that the flanking sequences constrain the allele-frequency distribution through an unknown mechanism.

The results of this study indicate that there continues to be a strong association between allele 18 and the *Hin*fI(+) polymorphism and between allele 24 and the *Hin*fI(–) mutation, which predominates in non-African populations but is not represented in African populations (Figure 3). A high level of genetic diversity was observed in African populations, showing that the distribution of haplotypes is multimodal (Figure 3), which was also supported by Albarran et al. (1998). Thus SNP haplotypes may be useful in identifying population of origin for individuals. This study is consistent with a shared and not-too-distant African origin for all the non-African populations studied to date. Obviously more data are needed to validate this proposition.

Received 29 September 2009; revision accepted for publication 8 December 2009.

Literature Cited

Albarran, C., O. Garcia, A. Alonso et al. 1998. Patterns of haplotype variation at the *D1S80* locus and a flanking sequence polymorphism in African and non-African populations. In *Progress in Forensic Genetics 7*, B. Olaisen, B. Brinkmann, and P. J. Lincoln, eds. New York: Elsevier, 401–403.

Alonso, A., P. Martin, C. Albarran et al. 1995. A *Hin*fI polymorphism in the 5´ flanking region of the human VNTR locus *D1S80*. *Int. J. Legal Med.* 107:216–218.

Balamurugan, K., N. Prabakaran, G. Duncan et al. 2001. Allele frequencies of 13 STR loci and the *D1S80* locus in a Tamil population from Madras, India. *J. Forensic Sci.* 46:1515–1517.

Buard, J., and A. J. Jeffreys. 1997. Big, bad minisatellites. *Nat. Genet.* 15:327–328.

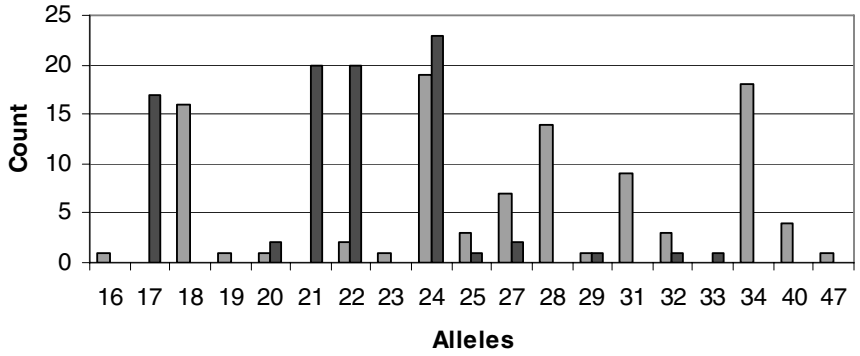

Figure 3. Multimodal allelic distribution and haplotype distribution of the 5´flanking region *Hin*fI restriction site polymorphism in a Rwandan (African) population. Light bars represent alleles with *Hin*fI(+), and dark bars represent the *Hin*fI(–) polymorphisms.

Buard, J., A. C. Shone, and A. J. Jeffreys. 2000. Meiotic recombination and flanking marker exchange at the highly unstable human minisatellite *CEB1* (*D2S90*). *Am. J. Hum. Genet.* 67:333–344.

Budowle, B., R. Chakraborty, A. M. Giusti et al. 1991. Analysis of the VNTR locus *D1S80* by the PCR followed by high-resolution PAGE. *Am. J. Hum. Genet.* 48:137–144.

Butler, J. M., A. E. Decker, P. M. Vallone et al. 2006. Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci. Int.* 156:250–260.

Chakraborty, R., D. N. Stivers, B. Su et al. 1999. The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis* 20:1682–1696.

Deka, R., S. DeCroo, J. Jin et al. 1994. Population genetic characteristics of the *D1S80* locus in seven human populations. *Hum. Genet.* 94:252–258.

Duncan, G. T., K. Balamurugan, B. Budowle et al. 1996. *Hin*fI/*Tsp*509I and *Bso*FI polymorphisms in the flanking regions of the human VNTR locus *D1S80*. *Genet. Anal. Biomol. Eng.* 13:119–121.

Duncan, G. T., K. Balamurugan, B. Budowle et al. 1997. Microvariation at the human *D1S80* locus. *Int. J. Legal Med.* 110:150–154.

Gibbons, A. 2008. Anthropology: The birth of childhood. *Science* 322:1040–1043.

Jeffreys, A. J., K. Tamaki, A. MacLeod et al. 1994. Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* 6:136–145.

Kasai, K., Y. Nakamura, and R. White. 1990. Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. *J. Forensic Sci.* 35:1196–1200.

Miller, S. A., D. D. Dykes, and H. F. Polesky. 1988. A simple salting-out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16:1215.

White, P. S., P. M. Thompson, T. Gotoh et al. 2005. Definition and characterization of a region of 1p36.3 consistently deleted in neuroblastoma. *Oncogene* 24:2684–2694.