Variation in STR Loci of the Human Myelin Basic Protein Gene: North Portugal and São Tomé e Príncipe

LUÍSA PEREIRA,[1,2] LEONOR GUSMÃO,[1] MARIA JOÃO PRATA,[1,2] PAULO MOTA,[1,2] MARIA JESUS TROVOADA,[1,3] AND ANTÓNIO AMORIM[1,2]

Abstract Allele frequencies and a single-base substitution polymorphism for 3 short tandem repeat (STR) loci of the human myelin basic protein (MBP) gene were evaluated in North Portugal and São Tomé e Príncipe. Strong linkage disequilibrium between loci *MBPB* and *MBPC* was found. However, the patterns of nonrandom allelic associations were very different in the 2 populations: levels of haplotypic diversity and heterozygosity were higher in the São Tomé population. Similarly, a difference in the frequency of base substitution G → A at position 124 was found: the frequency reached 4.1% in North Portugal and 0.5% in São Tomé. In both populations it was always found to be associated with haplotypes *B10/C11* and *B12/C9*.

The short tandem repeat (STR) system of the myelin basic protein gene (*MBP*), located 5′ to exon 1 of the gene, is characterized by extensive polymorphism primarily determined by variation of the tetranucleotide motif TGGA (Boylan et al 1990). Using a single pair of primers (Polymeropoulos et al. 1992), 2 loci, *MBPA* and *MBPB*, can be coamplified the last contained within the first. Treatment of the largest amplified fragments with the restriction enzyme *Nla*III (Nelleman et al. 1996), allows individual analysis of the other region that complements *MBPB*, designated as locus *MBPC*. Thus, since *MBPA* is basically the sum of *MBPB* and *MBPC*, it is possible to unequivocally infer compound haplotypes between *MBPB* and *MBPC* alleles (Figure 1). These features of linkage disequilibrium and haplotypic variation have proven to be useful tools in population studies (Tishkoff et al. 1996; Albarrán et al. 1998).

In addition, a single nucleotide polymorphism, specifically a G → A substitution at position 124, has been described (Nelleman et al. 1996; Gus-

[1] Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Roberto Frias s/n, 4200 Porto, Portugal.
[2] Faculdade de Ciências de Universidade do Porto. Pr. Gomes Teixeira, 4050 Porto, Portugal.
[3] Departamento de Antropologia da Universidade de Coimbra, 3049 Coimbra Codex, Portugal.

Human Biology, June 2000, v. 72, no. 3, pp. 481–487.

KEY WORDS: MBP, LINKAGE DISEQUILIBRIUM, SINGLE-BASE SUBSTITUTION, POPULATION GENETICS

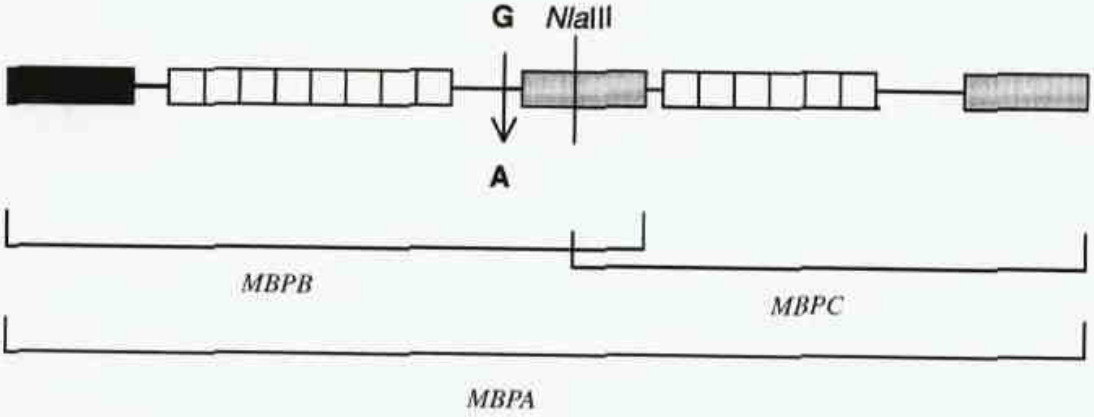

Figure 1. Schematic representation of the MBP region. Black rectangle represents the annealing region of the forward primer; gray rectangles represent the 2 annealing regions of the reverse primer; white squares represent the TGGA repeat motif; vertical arrow represents position 124 with G to A base substitution; vertical line represents the restriction site cut by *Nla*III. Allelic typing of the hypothetical individual in this scheme would be: *MBPA*14, MBPB*8*, and *MBPC*6*; or haplotype *B8/C6*.

mão et al. 1996). Its discrimination obviously increases the information that study of the MBP system can contribute to population genetics.

In this work, we present population data concerning *MBPA*, *MBPB*, and *MBPC* allele frequencies and distribution patterns, as well as the polymorphic substitution at position 124, from North Portugal and São Tomé e Príncipe, a small African archipelago located in the Gulf of Guinea.

Materials and Methods

DNA samples were obtained from 211 and 249 unrelated individuals from North Portugal and São Tomé e Príncipe, respectively.

Polymerase chain reaction (PCR) analysis was performed based on the double amplification method of Gusmão et al. (1996). The amplified samples (corresponding to *MBPA* and *MBPB*) were run on 8% T, 5% C polyacrylamide gels (10 cm × 20 cm × 0.3 cm), using the discontinuous buffer system described by Sajantila et al. (1992).

For the analysis of *MBPC*, 8 μl of PCR-amplified DNA were subjected to restriction enzyme treatment with *Nla*III, as described in Nelleman et al. (1996). The digested samples were analyzed using the electrophoretic conditions described above, except that gels were 12% T, 5% C.

After electrophoresis, DNA fragments were visualized by silver staining following the method of Budowle et al. (1991). The largest fragments (from 209 to 237 base pairs [bp]) represent *MBPA* (according to the original description in Möller et al. [1994], although Nelleman et al. [1996] considered it *MBPC*), the smallest (121 to 145 bp) *MBPB*, and the restricted fragments *MBPC* (the *MBPA* region referred to by Nelleman et al. 1996).

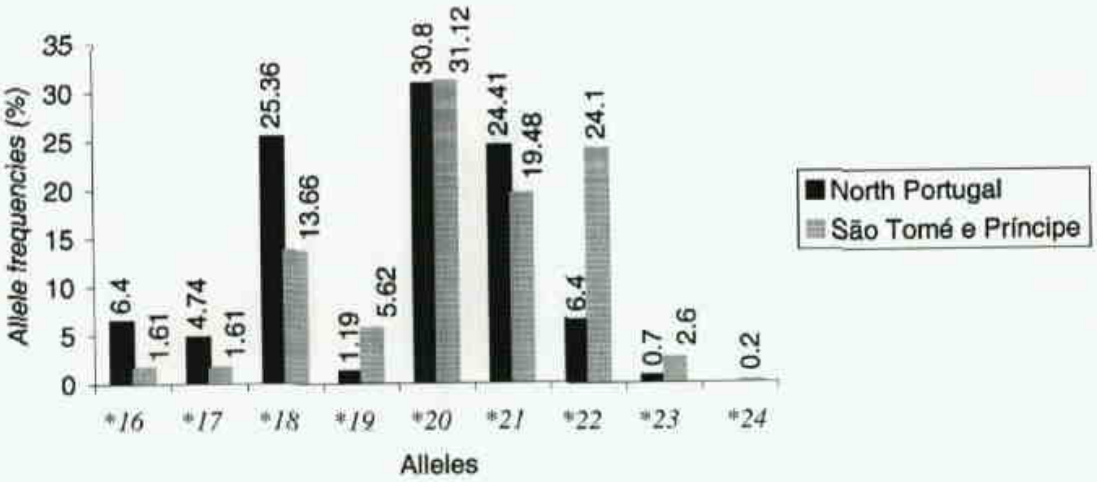

Figure 2. Allele frequencies for *MBPA* in North Portugal and São Tomé e Príncipe.

Statistical analyses for testing Hardy-Weinberg equilibrium and pairwise population comparisons were performed using exact tests (Guo and Thompson 1992), running the statistical software package GENEPOP (Raymond and Rousset 1995). Unbiased heterozygosity was estimated according to Nei (1978). Overall linkage disequilibrium between *MBPB* and *MBPC* was evaluated by p values of exact tests applying χ^2 measures and using Genetic Data Analysis (Lewis and Zaykin 1999) software. Pairwise nonrandom associations between *MBPB* and *MBPC* alleles were measured by D′ values computed with the Arlequin software (Schneider et al. 1997).

Results and Discussion

Variability at the *MBPA*, *MBPB*, and *MBPC* Loci. Both populations shared almost all the alleles at the 3 *MBP* loci. For *MBPA*, 9 alleles were detected, with from 16 to 24 repeats; the largest sized allele was absent from the Portuguese sample. For *MBPB*, 8 alleles, with from 7 to 14 repeats, were found; allele **8* was absent in the Portuguese sample, whereas allele **14* was absent in the São Tomé sample. For *MBPC*, 7 alleles, with from 6 to 12 repeats, were detected; allele **6* was absent in the Portuguese sample.

Allele frequency profiles for *MBPA*, *MBPB*, and *MBPC* in both populations are graphically represented in Figures 2, 3, and 4. The observed genotypic distributions did not deviate significantly from Hardy-Weinberg expectations. For *MBPA* and *MBPB*, highly significant differences were found between the 2 populations analyzed ($p = 0.000$ for both loci), whereas for *MBPC* differences were not statistically significant ($p = 0.229$).

It is interesting to note that despite the registered differences in population allele frequencies, the shapes of the allele distribution patterns are quite

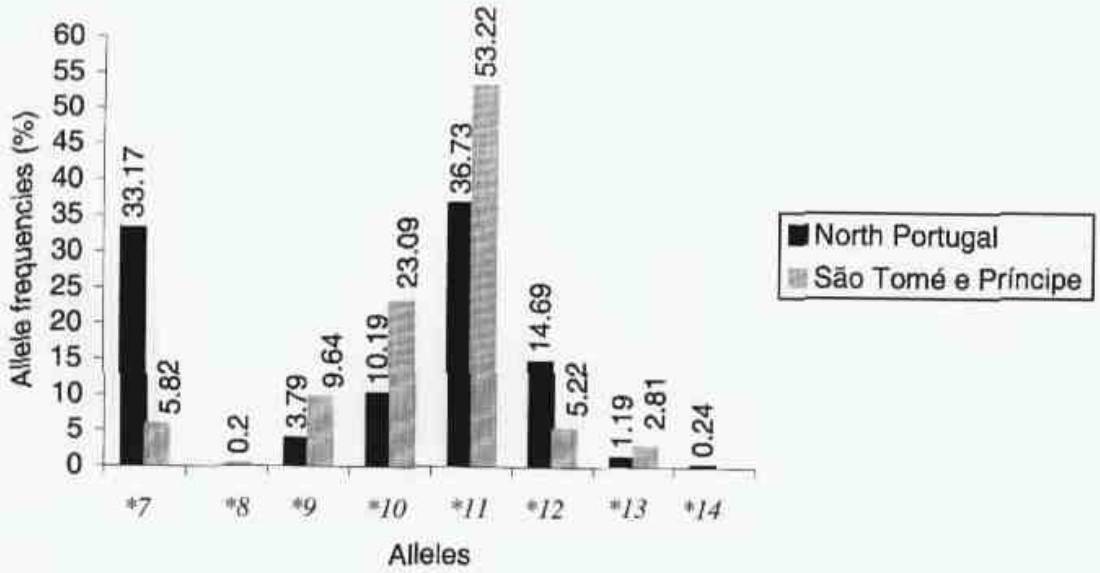

Figure 3. Allele frequencies for *MBPB* in North Portugal and São Tomé e Príncipe.

similar for the 3 loci in both populations. This finding, also described for other STRs in several populations, has been interpreted as evidence for a relatively recent common origin of modern human populations (Pérez-Lezaun et al. 1997).

We conducted further haplotypic analysis between loci *MBPB* and *MBPC*. In the São Tomé e Príncipe population, 26 distinct haplotypes were found, contrasting with 17 in the North Portuguese sample (Fig. 5). Levels of haplotype heterozygosity were slightly higher in the São Tomé (0.8674) sample compared to the Portuguese sample (0.8183).

Highly significant linkage disequilibria were detected between *MBPB* and *MBPC* in both populations. However, the overall extent of nonrandom

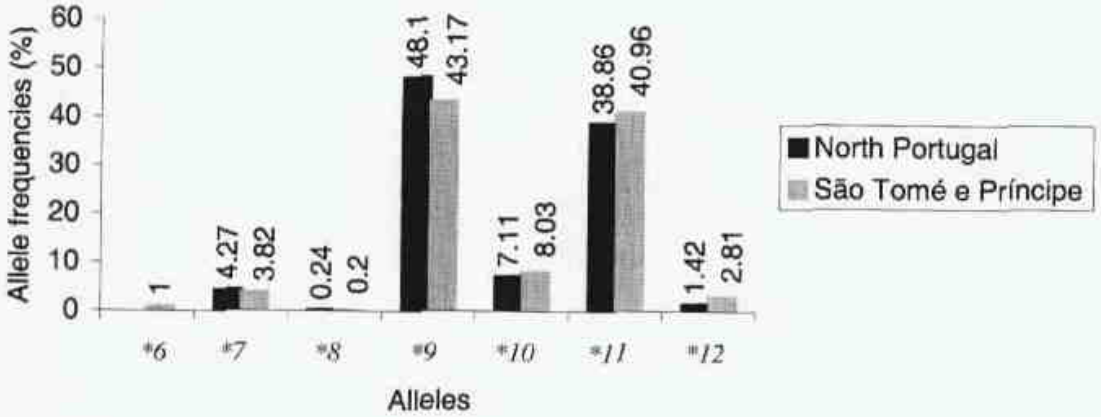

Figure 4. Allele frequencies for *MBPC* in North Portugal and São Tomé e Príncipe.

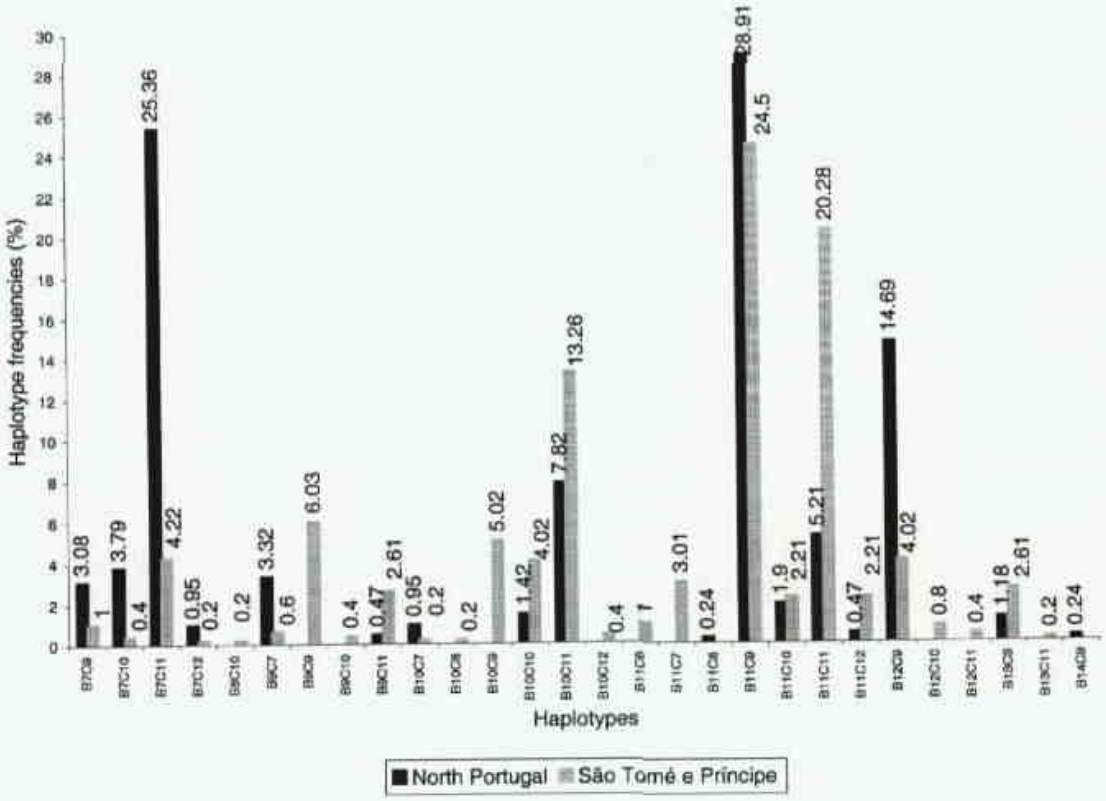

Figure 5. Haplotype frequencies in North Portugal and São Tomé e Príncipe.

allele associations was weaker in the São Tomé population: $p = 0.000$ for North Portugal and $p = 0.009$ for São Tomé e Príncipe (p values of exact test when χ^2 measures were applied to the simultaneous distribution of *MBPB* and *MBPC* loci).

The strongest positive haplotypic associations found between *MBPB* and *MBPC* are displayed in Table 1. These results indicate that the patterns

Table 1. D' Values Registered for some Pairwise Haplotypic Associations between *MBPB* and *MBPC* Alleles in North Portugal and São Tomé e Principe

North Portugal		*São Tomé e Príncipe*	
Haplotypes	*D' Values*	*Haplotypes*	*D' Values*
B7/C10	0.3017	*B7/C11*	0.5327
B7/C11	0.6145	*B9/C9*	0.3401
B7/C12	0.5012	*B10/C10*	0.3499
B9/C7	0.8694	*B11/C7*	0.5500
B10/C11	0.6196	*B11/C12*	0.5420
B11/C9	0.5897	*B12/C9*	0.5939
B12/C9	1.0000	*B13/C9*	0.8743

of pairwise linkage disequilibrium are very different in the African and European populations studied. The strongest positive haplotype associations *B7/C11*, *B9/C7*, *B11/C9*, and *B12/C9* detected in the Portuguese sample were also shared by the Danish and Greenland Eskimo samples studied by Nelleman et al. (1996).

In summary, the São Tomé e Príncipe population studied presented a wider range of haplotypes and a higher expected heterozygosity than the Portuguese sample. Both findings are in agreement with several recent studies that have registered for non-African populations a subset of the genetic diversity found in African populations.

On the other hand, the haplotypic associations shared between European populations (Portuguese, Danish, and Greenland Eskimos) and not shared between Portuguese and São Tomeans, and the globally weaker overall linkage disequilibrium in the African population add further support to the view of recent African origin of modern human populations.

The G → A Substitution at Position 124. When native electrophoretic conditions were applied, a slightly faster mobility of some A21 fragments was observed. It was possible to connect this altered electrophoretic mobility with the presence of the G → A substitution at position 124, previously described by Nelleman et al. (1996). However, as the mobility shift depended on optimal electrophoretic conditions, the discrimination of that base substitution could be reliably achieved only when the double amplification method described by Gusmão et al. (1996) was applied.

The substitution reached a frequency of 4.1% in the North Portugal population and 0.5% in the São Tomé population. In both of these populations, the G → A substitution was present only in alleles *MBPA*21*, *MBPB*10*, and *MBPB*12*. Thus, in these 2 populations the base substitution was found in the background haplotypes *B10/C11* and *B12/C9*, whereas in the Danish and Eskimo samples studied by Nelleman et al. (1996) it was associated only with haplotype *B10/C11*.

In summary, analyzing the extent and organization of genetic variability in the MBP system, as discriminated by the simultaneous analysis of *MBPA*, *MBPB*, and *MBPC* and the polymorphic substitution at position 124, provided a valuable insight into anthropological genetics. It would be interesting to enlarge the study to include a wider range of human populations in order to obtain (1) a more global and informative pattern of haplotypic distribution and linkage disequilibrium, and (2) additional relevant data for inferring the geographical origins of specific *MBP* substitutions.

Acknowledgments This work was partially supported by Programa PRAXIS XXI through project PRAXIS/2/2.1/BIA/196/94, and Junta Nacional de Investigação Científica e Tecnológica through grant BM/ 6723/95.

Received 12 March 1999; revision received 27 July 1999.

Literature Cited

Albarrán, 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. *Progress in Forensic Genetics.* 7:401–403.

Boylan, K.B., T.M. Ayres, B. Popko et al. 1990. Repetitive DNA (TGGA)n 5′ to the human myelin basic protein gene: A new form of oligonucleotide repetitive sequence showing length polymorphism. *Genomics* 6:16–22.

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.

Guo, S.W., and E.A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 48:361–372.

Gusmão, L., A. Amorim, M.J. Prata et al. 1996. Failed PCR amplifications on MBP-STR alleles due to polymorphism in the primer annealing region. *Int. J. Leg. Med.* 108:313–315.

Lewis, P.O., and D. Zaykin. 1999. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0.

Möller, A., P. Wiegand, C. Grüschow et al. 1994. Population data and forensic efficiency values for the STR systems HumVWA, HumMBP and HumFABP. *Int. J. Leg. Med.* 106:183–189.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590.

Nelleman, L.J., J. Frederiksen, and N. Morling. 1996. PCR typing of DNA fragments of two short tandem repeats (STR) system upstream of the human myelin basic protein (MBP) gene in Danes and Greenland Eskimos. *Forensic Sci. Int.* 78:139–155.

Pérez-Lezaun, A., F. Calafell, E. Mateu et al. 1997. Microsatellite variation and the differentiation of modern humans. *Hum. Genet.* 99:1–7.

Polymeropoulos, M.H., H. Xiao, and C.R. Merril. 1992. Tetranucleotide repeat polymorphism at the myelin basic protein gene (MBP). *Hum. Mol. Genet.* 1:658.

Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.* 86:248–249.

Sajantila, A., B. Budowle, M. Ström et al. 1992. PCR amplification of alleles at the *D1S80* locus: Comparison of a Finnish and a North American Caucasian population sample, and forensic casework evaluation. *Am. J. Hum. Genet.* 50:816–825.

Schneider, S., J.M. Kueffer, D. Roessli et al. 1997. Arlequin ver.1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Tishkoff, S.A., E. Dietzsch, W. Speed et al. 1996. Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. *Science* 271:1380–1387.