



## Genetic variability and efficiency of DNA microsatellite markers for paternity testing in horse breeds from the Brazilian Marajó archipelago

Sávio P. Reis, Evonnildo C. Gonçalves, Artur Silva and Maria P.C. Schneider

*Laboratório de Polimorfismo de DNA, Departamento de Genética, Universidade Federal do Pará, Belém, PA, Brazil.*

### Abstract

In this study, 15 microsatellite DNA loci used in comparative tests by the International Society for Animal Genetics were applied to the evaluation of genetic diversity and management, and the efficiency of paternity testing in Marajoara horses and Puruca ponies from the Marajó Archipelago. Based on the genotyping of 93 animals, mean allelic diversity was estimated as 9.14 and 7.00 for the Marajoara and Puruca breeds, respectively. While these values are similar to those recorded in most European breeds, mean levels of heterozygosity were much lower (Marajoara 49%, Puruca 40%), probably as a result of high levels of inbreeding in the Marajó populations. The mean informative polymorphic content of this 15-marker system was over 50% in both breeds, and was slightly higher in the Marajoara horses. The discriminative power and exclusion probabilities derived from this system were over 99% for both populations, emphasizing the efficacy of these markers for paternity testing and genetic management in the two breeds.

*Key words:* Marajoara horses, Puruca ponies, microsatellite markers, genetic diversity, genealogical control.

Received: February 22, 2007; Accepted: October 4, 2007.

The various breeds of horses currently found in Brazil are the result of crosses between European and Arab breeds, which have prospered under specific local conditions (Colbert, 1980; Torres and Jardim, 1985). A prime example from the northern part of the country is the Marajoara breed, derived from Portuguese stock from Cabo Verde, which was introduced into the Marajó archipelago at the beginning of the eighteenth century. These horses have adapted well to the flat, low-lying terrain of the archipelago, acquiring a number of morphological adaptations, such as rounded hooves and modifications of the muscles of the thighs, rump, and back related to trotting in flooded marshland, which is typical of the Marajó region.

A second example is the Puruca breed of pony, derived from nine Shetland ponies imported from France at the end of the nineteenth century. These animals were crossed with Marajoara horses, and proliferated due to their high fecundity and morphological adaptations similar to those of the Marajoaras, related to locomotion in the marshy local conditions. These adaptive changes in the characteristics of the two breeds have had important eco-

nomical implications for ranching activities on the archipelago, where they are used extensively for herding both cattle and buffalo.

The creation of the Marajoara and Puruca breeds in northern Brazil is a classic case of artificial selection, based solely on the preferential crossing of animals which best represented the desired characteristics. However, in the last decades, many animals from both breeds have been deprived of many of their characteristics due to crossings with other breeds introduced in the Marajó Archipelago, leading to a decrease in the number of individuals with complete features. The lack of a conservation program increases this problem, threatening the future of these local breeds. Recent advances in molecular biology have produced considerable advances in the potential for the improvement of domestic animals. This type of intervention nevertheless depends on knowledge of the structure and organization of the genetic variation available in each of the breeds to be improved (Oliveira and Kuabara, 2001).

In the last fifteen years, the development of new analytic tools has brought about major advances in population studies. Of these tools, DNA microsatellite markers have proven to be extremely useful for the elucidation of the population structure of a wide variety of species of plants and animals. They have been used for the evaluation of genetic structure in horses, as well as for paternity testing and

the establishment of phylogenetic relationships among different breeds (Bjørnstad *et al.*, 2000).

In 1993, the Food and Agricultural Organization (FAO) of the United Nations proposed the establishment of a worldwide program for the maintenance of genetic resources, using microsatellite markers for the characterization of various breeds of domestic animals (Barker *et al.*, 1993). Since then, a range of studies have characterized the genetic variability of a number of breeds of horse (Bjørnstad *et al.*, 2000; Cañon *et al.*, 2000; Tosaki *et al.*, 2003; Achmann *et al.*, 2004).

Many of these studies have used the panel of markers recommended by the International Society for Animal Genetics (ISAG) for the genealogical control of domestic species. Given this, the present study applies these markers to the study of the genetic variability of two Brazilian equine breeds - Marajoara and Puruca - and evaluates the applicability of this panel to genealogical control in each breed. These goals are important due to the narrow relationship between the breeds and the importance of these equines to the society of the Marajó archipelago.

A total of 93 individuals representing the two equine breeds were analyzed. Blood samples were collected from 61 Marajoara horses and 32 Puruca ponies on the Camburupy Ranch on Marajó archipelago. Care was taken to sample individuals that were not closely related. Total DNA was extracted from 1 to 3 mL of blood from each animal, following Sambrook *et al.*'s (1989) standard procedure of digestion by K proteinase, separation with phenol-chloroform, and precipitation with ethanol. The samples are part of the DNA bank of the DNA Polymorphism Laboratory of the Department of Genetics at UFPA, Belém, Brazil.

The 15 microsatellite markers (Table 1) were developed specifically for *Equus caballus*, and were chosen here because of their widespread use in the genealogical control of equines by the ISAG, due to their normally high levels of heterozygosity, large numbers of alleles, and relatively consistent amplification. The 15 microsatellites were amplified in multiplex protocols (two or three loci amplified in conjunction) in seven separate polymerase chain reactions. Each reaction had a final volume of 20  $\mu$ L, containing 150 ng of the sample DNA, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.75-1.25 mM MgCl<sub>2</sub>, 375  $\mu$ M dNTP, 1U Taq-DNA Polymerase (Invitrogen, Carlsbad, CA, USA), 0.05-0.3 pmol of Cy5-labeled forward primers and unlabeled reverse primers, and 2.5% of formamide (Sigma-Aldrich, St. Louis, MO, USA). Amplifications were carried out in a PTC-2000 (MJ Research, Reno, NV, USA) thermocycler with the following touchdown cycle program: 3 min at 96 °C for initial denaturation; 10 cycles of 45 s at 94 °C for denaturation, 45 s at temperatures varying between 62 °C and 57 °C (decreasing by 0.5 °C at each cycle) for annealing of the primers and 45 s at 72 °C for extension; 21 to 31 cycles of 45 s at 90 °C for denaturation, 45 s at

**Table 1** - The fifteen ISAG panel markers used in the study of equines of the Marajoara and Puruca breeds. The numbers in brackets indicate the loci analyzed together in multiplex amplification.

| Locus     | Allele size range (bp) | Chromosome | Reference                          |
|-----------|------------------------|------------|------------------------------------|
| HTG4 (1)  | 128-144                | 9          | Ellegren <i>et al.</i> (1992)      |
| HMS7 (1)  | 172-190                | 1          | Guerin <i>et al.</i> (1994)        |
| AHT5 (2)  | 127-147                | 8          | Binns <i>et al.</i> (1995)         |
| HMS6 (2)  | 157-171                | 4          | Guerin <i>et al.</i> (1994)        |
| HTG7 (3)  | 118-130                | 4          | Marklund <i>et al.</i> (1994)      |
| AHT4 (3)  | 142-156                | 24         | Binns <i>et al.</i> (1995)         |
| ASB2 (3)  | 170-188                | 15         | Breen <i>et al.</i> (1997)         |
| HTG10 (4) | 92-114                 | 21         | Marklund <i>et al.</i> (1994)      |
| HMS3 (4)  | 151-171                | 9          | Guerin <i>et al.</i> (1994)        |
| VHL20 (5) | 88-108                 | 30         | Van Haeringen <i>et al.</i> (1994) |
| HMS1 (5)  | 172-182                | 15         | Guerin <i>et al.</i> (1994)        |
| HTG6 (6)  | 82-105                 | 15         | Ellegren <i>et al.</i> (1992)      |
| LEX3 (6)  | 145-167                | X          | Coogle <i>et al.</i> (1996)        |
| ASB17 (7) | 92-124                 | 2          | Breen <i>et al.</i> (1997)         |
| ASB23 (7) | 148-176                | 3          | Irvin <i>et al.</i> (1998)         |

54, 56 or 58 °C for annealing and 45 s at 72 °C for extension, and 5 min at 72 °C for final extension.

The amplified fragments were separated by vertical electrophoresis in 8% polyacrylamide gel and detected using an AlfExpress II automated laser fluorescent DNA sequencer (GE Healthcare, Chalfont St. Giles, UK). The different Cy5-labeled fragments and their respective molecular weights were identified using the Allelelocator program, version 1.03 (GE Healthcare). Fragments of known size by sequencing were used for calibration of the genotyping results.

In order to measure genetic variability in each population, a number of parameters were computed for each locus: number of alleles (A), effective number of alleles (ne), observed heterozygosity (Ho) and expected heterozygosity (He) according to Hardy-Weinberg equilibrium (HWE). Mean values ( $\pm$  standard deviation) were also calculated for the system as a whole, using the Popgene32 program (Yeh *et al.*, 1997). To test the significance of deviations from HWE at each locus, a Markov chain analysis (100 batches of 1000 iterations, with a 1000-step dememorization process) was used to calculate unbiased estimates of exact probabilities (P values). This exact test was run in the Genepop program, version 3.1 (Raymond and Roussett, 1995).

Polymorphic information content (PIC), which permits evaluation of the contribution of each locus to the analysis of segregation (Botstein *et al.*, 1980), and power of discrimination (PD), which is an index of the probability of correctly identifying a randomly selected individual (Jones, 1972; Brenner and Morris, 1990), were calculated using the

PowerStats V1.2 (Promega) spreadsheet in Excel 2003 (Microsoft Corporation). Probability of parentage exclusion, which estimates the chance of excluding an individual selected at random from a population from possible paternity, were calculated using two formulas: (a) when the genotypes of a mother, her offspring, and a putative sire are compared, PE1 (Jamieson, 1994), and (b) when one parental genotype is unavailable, PE2 (Jamieson and Taylor, 1997, based on Garber and Morris, 1983). The locus-specific values of PD, PE1, and PE2 were combined for each population using the formula  $p = 1 - (1 - P_1)(1 - P_2)(1 - P_3)(1 - P_k)$ , for k loci.

The number of alleles recorded per locus varied between 6 and 15 in the Marajoara, and between 3 and 10 in the Puruca. The difference between the two breeds appears to be due to sample size, although the mean number of alleles per locus - 9.1 and 7.0, respectively - were similar to those recorded for other horse breeds. Mean values of approximately seven alleles per locus have been recorded for a number of equine populations, including seven populations of Lipizzaner, which is common in southeastern Europe (Achmann *et al.*, 2004), four horse breeds native to the Basque country (Solis *et al.*, 2005) and four populations of domestic ass (*Equus asinus*), one from Catalonia (Jordana *et al.*, 2001) and three from Croatia (Ivankovic *et al.*, 2002). Values similar to our recorded here for the Marajoara breed were found in four Norwegian (Bjørnstad *et al.*, 2000) six German (Aberle *et al.*, 2004), and seven Celtic Spanish horse breeds (Cañon *et al.*, 2000).

Heterozygosity per locus varied from 11 to 85% in the Marajoara, and from 0 to 61% in the Puruca breeds, with mean values of 49% and 40%, respectively (Table 2). These values are considerably lower than those observed in European horses and asses, for which mean heterozygosity ranges from 60% to 66% (see Ivankovic *et al.*, 2002; Cunningham *et al.*, 2001; Jordana *et al.*, 2001; Aberle *et al.*, 2004; Achmann *et al.*, 2004).

As a parameter, the allele number is more sensitive to the loss of genetic variability than heterozygosity. In this case, the more similar allele numbers recorded here in comparison with European breeds, together with their reduced heterozygosity, suggests that the two Brazilian breeds have not lost many of their original alleles, but have suffered a loss of heterozygosity due to increased inbreeding. This conclusion is supported by the lack of Hardy-Weinberg equilibrium observed in the combined analysis of all loci in both populations, where the majority of loci present a significant deficit of heterozygotes ( $p < 0.05$ ). Inbreeding is common practice in many domestic animals (Oliveira and Kuabara, 2001), but it is especially intense in populations with reduced effective size, as is the case here, with regard to both Marajoara and Puruca populations.

Jordana *et al.* (2001) and Ivankovic *et al.* (2002) also observed Hardy-Weinberg disequilibrium in their study populations. The Catalanian breed analyzed by Jordana *et*

**Table 2** - Levels of genetic variability and percentage values of genealogical parameters in Marajoara horses and Puruca ponies from the Marajó archipelago in comparison with other breeds, based on 15 microsatellite markers from the ISAG panel control (Reis *et al.*, not published).

| Parameters        | Marajoara | Puruca  | Quarto de Milha | Mangalarga Marchador |
|-------------------|-----------|---------|-----------------|----------------------|
| A <sup>1</sup>    | 9.14      | 7.00    | 6.53            | 4.80                 |
| Ne <sup>2</sup>   | 5.01      | 3.98    | 3.96            | 2.97                 |
| Ho <sup>3</sup>   | 0.49      | 0.40    | 0.50            | 0.32                 |
| He <sup>4</sup>   | 0.75      | 0.70    | 0.69            | 0.57                 |
| CPD <sup>5</sup>  | > 99.99   | > 99.99 | > 99.99         | > 99.99              |
| PIC <sup>6</sup>  | 72        | 66      | 64              | 52                   |
| CPE1 <sup>7</sup> | > 99.99   | > 99.99 | > 99.99         | 99.94                |
| CPE2 <sup>8</sup> | 99.96     | 99.74   | 99.79           | 98.39                |

1. Mean number of alleles; 2. Mean effective number of alleles; 3. Mean observed heterozygosity; 4. Mean expected heterozygosity, according to Hardy-Weinberg equilibrium; 5. Power of discrimination for the system; 6. Mean polymorphic information content; 7. Combined probability of parentage exclusion when the genotypes of a mother, her offspring, and a putative sire are compared; 8. Combined probability of parentage exclusion when one parental genotype is unavailable.

*al.* (2001) is known to have a restricted distribution and a population of more or less 100 individuals. Similarly, the Croatian breed studied by Ivankovic *et al.* (2002) is threatened with extinction, and also has a reduced number of individuals. By contrast, the populations studied by Achmann *et al.* (2004), Cunningham *et al.* (2001) and Aberle *et al.* (2004) were all in Hardy-Weinberg equilibrium, although the breeds studied in these cases are all amply distributed and presumably have populations of large effective size, as is the case of the Lipizzaner in particular (Achmann *et al.*, 2004).

The PIC values recorded per locus varied between 24% and 88% for the Marajoara horses, and between 22% and 86% for the Puruca ponies, with means of 72% and 66%, respectively (Table 2). These values reflect the slightly greater genetic variation in the Marajoara breed, which is probably a result of the difference in the effective size of the original populations introduced into the Marajó archipelago. According to Botstein (1980), values below 50% can be considered uninformative, although in the present case, this only applies to locus VHL20, with values of 24% and 22% in the Marajoara and Puruca samples, respectively.

Power of discrimination values for each locus varied between 30% and 96% in the Marajoara, and 23% and 92% in the Puruca breeds. Values of PE1 - the index most used for paternity testing - varied between 14% and 78% in the Marajoara horses, and 12% and 72% in the Puruca ponies. The values of PE2 were slightly lower, varying between 3% and 64% in the Marajoara, and 3% and 55% in the Puruca samples. Once again, the VHL20 locus returned the lowest values for all three parameters. The PD value for the system as a whole was above 99.99% for both breeds.



The combined PE1 value (CPE1) obtained for both populations was also above 99.99%, but was slightly lower in the case of CPE2, in which case the values were 99.96% and 99.74%, respectively, for the Marajoara and Puruca breeds (Table 2). These values for discrimination and exclusion of individuals demonstrate the efficiency of the loci analyzed in this study for genealogical control, and the genetic management of the two breeds. Jordana *et al.* (2001) recorded a mean PIC equal to the value recorded here for the Puruca breed, although a slightly lower value for CPE1 (99.93%) than either recorded here. Ivankovic *et al.* (2002) also recorded slightly lower mean values for PIC (61%) and CPE1 (99.1%) than the present study.

However, CPE values are directly related to the number of loci analyzed, so these differences are consistent with the fact that the study of Jordana *et al.* (2001) included ten loci, and that of Ivankovic *et al.* (2002), only six. Using twelve loci, Solis *et al.* (2005), recorded values of CPE1 and CPE2 much closer to those observed in the present study.

Given the exceptionally low PIC value recorded for the VHL20 locus, the parameters were recalculated omitting this locus, although this only affected the mean values of PIC and CPE2. The PIC values increased from 72% and 66% to 76% and 70%, respectively, in the Marajoara and Puruca breeds, whereas the CPE2 values decreased from 99.96% and 99.74%, respectively, to 99.95% and 99.73%. This indicates that, despite the fact that VHL20 is relatively uninformative in general, it may be relatively useful for paternity testing when parental genotypes are unknown.

When compared with our data from other Brazilian breeds, using the same markers, the values of the Marajo breeds were very similar (Table 2), only the number of alleles from Mangalarga Marchador being lower than others (4.8). Cunningham *et al.* (2001), however, found a mean value of 4.7 alleles for the Thoroughbred breed, which close to this value. The majority of populations of the Brazilian breeds have reduced effective sizes or reduced founder populations. Mangalarga Marchador originated from few individuals of one major founder breed (Alter), while Quarto de Milha originated from a few North American horses introduced only fifty years ago. These data explain the similarity of the values among all breeds.

In conclusion, while the genetic diversity of the Marajoara horses was slightly greater than that recorded for the Puruca ponies, the levels observed in both breeds were similar to those encountered in a variety of European equine breeds. However, this is not the case for heterozygosity, which was inferior to that of European breeds in both Brazilian breeds. We interpret this as being the result of relatively high levels of inbreeding typically practiced in domestic animals, as well as - in the specific case of the Marajoara and Puruca - founder populations of relatively reduced size. The standardized system of molecular markers used in the present study was found to be highly effi-

cient for the genealogical control and genetic and conservation management of Marajoara horses and Puruca ponies, given the discrimination and exclusion values of over 99.99%.

## Acknowledgments

Financial support was provided by SECTAM, CNPq and PRONEX.

## References

- Aberle KS, Hamann H, Drögemüller C and Distl O (2004) Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. *Anim Genet* 35:270-277.
- Achmann R, Curik I, Dovc P, Kavari T, Bodo I, Habe F, Marti E, Sölkner J and Brem G (2004) Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. *Anim Genet* 35:285-292.
- Barker JSF, Bradley DG, Fries R, Hill WG, Nei M and Wayne RK (1993) An Integrated Global Programme to Establish the Genetic Relationships Among the Breeds of Each Domestic Animals Species. Food and Agriculture Organization of the United Nations, Rome, 24 pp.
- Binns MM, Holmes NG, Holliman A and Scott AM (1995) The identification of polymorphic microsatellites loci in the horse and their use in thoroughbred parentage testing. *Br Vet J* 151:9-15.
- Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331.
- Bjørnstad G, Gunby E and Røed KH (2000) Genetic structure of Norwegian horse breeds. *J Anim Breed Genet* 117:307-317.
- Breen M, Lindgren G, Binns MM, Norman J, Irvin Z, Bell K, Sandberg K and Ellegren H (1997) Genetical and physical assignments of equine microsatellites-first integration and anchored markers in horse genome mapping. *Mamm Genome* 8:267-273.
- Brenner C and Morris J (1990) Paternity index calculations in single locus hypervariable DNA probes: Validation and other studies. In: Proceedings for the International Symposium on Human Identification. Promega Corporation, Madison, pp 21-59.
- Cañón J, Checa ML, Carleos C, Vega-Plà JL, Vallejo M and Dunner S (2000) The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Anim Genet* 31:39-48.
- Coogler L, Reid R and Bailey E (1996) Equine dinucleotide repeat polymorphisms at loci LEX002, -003, -004, -005, -006, -007, -008, -009, -010, -011, -012, -013 and -014. *Anim Genet* 27:126-127.
- Colbert EH (1980) Evolution of the Vertebrates. A History of the Back Boned Animals Through Time. 3rd ed. Wiley & Sons, New York, 470 pp.
- Cunningham EP, Dooley JJ, Splan RK and Bradley DG (2001) Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses *Anim Genet* 32:360-364.

- Ellegren H, Johansson M, Sandberg K and Anderson L (1992) Cloning of highly polymorphic microsatellites in the horse. *Anim Genet* 23:133-142.
- Garber RA and Morris JW (1983) General equation for the average power of exclusion for genetic systems of  $n$  codominant alleles in one-parent and in no-parent cases of disputed parentage. In: Walker RH (ed) *Inclusion Probabilities in Parentage Testing*. American Association of Blood Banks, Arlington, pp 277-280.
- Guérin G, Bertaud M and Amigues Y (1994) Characterization of seven new horse microsatellites: HMS1, HMS2, HMS3, HMS5, HMS6, HMS7, HMS8. *Anim Genet* 25:62.
- Irvin Z, Giffard J, Brandon R, Breen M and Bell K (1998) Equine dinucleotide repeat polymorphisms at loci ASB 21, 23, 25 and 37-43. *Anim Genet* 29:67.
- Ivankovic A, Kavar T, Caput P, Mioc B, Pavic V and Dovic P (2002) Genetic diversity of three donkey populations in the Croatian coastal region. *Anim Genet* 33:169-177.
- Jamieson A (1994) The effectiveness of using codominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. *Anim Genet* 25 (suppl 1):37-44.
- Jamieson A and Taylor SC (1997) Comparisons of three probability formulae for parentage exclusion. *Anim Genet* 28:397-400.
- Jones DA (1972) Blood samples: Probability of discrimination. *J Forensic Sci Soc* 12:355-359.
- Jordana J, Folch P and Aranguren JA (2001) Microsatellite analysis of genetic diversity in the Catalonian donkey breed. *J Anim Breed Genet* 118:57-63.
- Marklund S, Ellegren H, Eriksson S, Sandberg K and Anderson L (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Anim Genet* 25:19-23.
- Oliveira DAA and Kuabara MY (2001) A genealogia, sua importância para o melhoramento e para a produção animal e os métodos de confirmação dos *pedigrees*. In: Pereira JCC (ed) *Melhoramento Genético Aplicado à Produção Animal*. FEPMVZ Editora, Belo Horizonte, pp 435-444.
- Raymond M and Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *J Hered*. 86:248-249.
- Sambrook J, Fritsch EF and Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
- Solis A, Jugo BM, Mériaux JC, Iriondo M, Mazón LI, Aguirre AI, Vicario A and Estomba A (2005) Genetic diversity within and among four south European native horse breeds based on microsatellite DNA analysis: Implications for conservation. *J Hered* 96:670-678.
- Torres AP and Jardim WR (1985) *Criação de Cavalo e outros Equinos*. Nobel, São Paulo, 656 pp.
- Tosaki T, Takesaki N, Hasegawa T, Ishida N, Kurosawa M, Tomita M, Saitou N and Mukoyama H (2003) Microsatellite variation in Japanese and Asian horses and their phylogenetic relationship using a European horse outgroup. *J Hered* 94:374-380.
- Van Haeringen H, Bowling AT, Lenstra JA and Zwaagstra KA (1994) A highly polymorphic horse microsatellite locus: VHL20. *Anim Genet* 25:207.
- Yeh FC, Yang RC and Boylet T (1997) POPGENE (version 1.32): Software Microsoft windows-based freeware for population genetics analysis. University of Alberta, Alberta, Canada.

*Associate Editor: Luiz Lehmann Coutinho*