Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Genetic diversity and population structure in Brazilian Mangalarga Marchador horses

J.B. DeAssis[1], D.M. DeLaat[1], M.G.C.D. Peixoto[2], J.A.G. Bergmann[3], C.G. Fonseca[1] and M.R.S. Carvalho[1]

[1]Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[2]EMBRAPA Gado de Leite, Juiz de Fora, MG, Brasil
[3]Departamento de Zootecnia, Escola de Medicina Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

Corresponding author: M.R.S. Carvalho
E-mail: mraquel@icb.ufmg.br

Genet. Mol. Res. 8 (4): 1519-1524 (2009)
Received August 27, 2009
Accepted October 15, 2009
Published December 23, 2009

ABSTRACT. One hundred and fifteen unrelated Mangalarga Marchador horses were sampled from three geographically distinct regions of Minas Gerais State, Brazil (South, Southeast, and Northeast) and tested for 10 microsatellite loci. Genetic diversity and population structure parameters were estimated with ARLEQUIN 3.0, CERVUS 2.0, POPGENE 1.31, GENEPOP on the web, STRUCTURE 2.0, and SPAGEDI 1.2 software packages. Under Hardy-Weinberg assumptions, seven markers were at equilibrium (LEX014, LEX017, LEX019, SGCV23, TKY321, VHL20, and VIASH39), while two (ASB3 and LEX031) presented significant homozygote excess. Seventy-four alleles were identified in these nine markers, with a mean of 8.22 alleles. Mean heterozygosity was 0.637 and polymorphism information content was 0.662. Markers ASB3, LEX019, SGCV23, TKY321, and VHL20 were highly informative (PIC >0.7) and may be useful for eventual expansion of parentage test panels. The F_{ST} value (0.0562) indicated relatively little geographical structure. However, based on a Bayesian-based cluster analysis under a three-cluster model, 94% of the 115 individuals were correctly assigned to the subpopulations from where they were sampled. Mean

pairwise *f* was relatively high (0.11), and in spite of the efforts towards non-consanguineous sampling, 1% of the pairs of individuals shared over 50% of the alleles. These results strongly suggest that the population is genetically structured. Under a conservation genetics approach, two strategies are recommended: avoidance of crosses between highly endogamic individuals and stimulation of crosses between individuals from those regions for which low genetic flow was identified.

Key words: Mangalarga Marchador horse; Genetic diversity; Population genetic structure

INTRODUCTION

Brazil has the third largest equine herd in the world with approximately 5.8 million animals. Mangalarga Marchador (MM) is the most numerous Brazilian horse breed, with over 300,000 registered animals. MM breeding started by the end of the XIX century as a result of crosses between Alter, Andaluz, Arab, Creole, and Quarter Horse animals. Studbooks were closed in 1966 (for males) and in 1984 (for females). The MM breed was established in the South of Minas Gerais State, from which animals dispersed throughout the country. Currently, the most important breeding centers are still located in this region. The MM breed's main features are rusticity and a particularly smooth gait.

This is the first study applying molecular markers to characterize population genetic structure of the MM breed and to provide an initial quantification of its genetic variability.

MATERIAL AND METHODS

Blood samples were collected from 115 unrelated MM horses (50 males; 65 females) born between November 2004 to February 2005 to 18 different studs. Selected studs were located in three regions of Minas Gerais State: Northeast (Governador Valadares Couty, 30 animals), South (south of the state, 63 animals), and Southeast (Cataguases County, 22 animals), representing most of the breed distribution in the state. DNA extraction was performed as described elsewhere (Miller et al., 1988).

The microsatellite loci tested were ASB3 (Breen et al., 1997), ECMPZ001 (Breen et al., 1994), LEX014, LEX017 and LEX019 (Coogle et al., 1996a), LEX031 (Coogle et al., 1996b), SGCV23 (Godard et al., 1997), TKY321 (Tozaki et al., 2001), VHL20 (van Haeringen et al., 1994), and VIASH39 (Ewen and Matthews, 1994).

Tests for allele frequencies, heterozygosity values, and Hardy-Weinberg equilibrium were carried out using the exact tests of GENEPOP (Raymond and Rousset, 1995). Polymorphism information content (PIC), power of exclusion (PE), observed heterozygosity (H_o), and expected heterozygosity (H_e) were computed with CERVUS 2.0 (Slate et al., 2000). The effective number of alleles and Wright F-statistics (F_{ST}, F_{IS}, F_{IT}) were calculated with POPGENE 1.31 (Yeh, 1999). Population substructure was investigated with a Bayesian approach using the prior information model (Markov Chain Monte Carlo) with STRUCTURE 2.0 (Pritchard et al., 2000). The pairwise inbreeding coefficient (*f*) was determined with SPAGEDI 1.2 (Hardy and Vekemans, 2002).

RESULTS

Allele number, effective allele number, observed and expected heterozygosities, PIC, total exclusionary power (PE), and F_{IS} as measure of Hardy-Weinberg equilibrium and their significance (Fisher's method) are shown in Table 1, except for ECMPZ001, which was monomorphic. The total number of alleles was 74. Allele numbers ranged from 4 (LEX017 and LEX031) to 19 (SGCV23). The effective allele number ranged from 1.98 (LEX017) to 5.85 (ASB3). Observed heterozygosity ranged from 0.314 (LEX031) to 0.867 (VHL20). Average PIC was 0.662. PIC values higher than 0.7 were found for ASB3, LEX019, SGCV23, TKY321, and VHL20. Cumulative probabilities of paternity exclusion were 97.5% (PE1) and 99.83% (PE2).

Table 1. Genetic variability parameters of the Mangalarga Marchador breed, F_{IS} (W&C) and P values for Hardy-Weinberg equilibrium test (Fisher's method).

								F_{IS} (W&C[8])		
Locus	K[1]	Ne[2]	Ho[3]	He[4]	PIC[5]	PE1[6]	PE2[7]	Northeast	South	Southeast
ASB3	11	5.85	0.538	0.834	0.810	0.498	0.669	+0.338***	-0.096***	+0.562***
LEX014	8	2.13	0.440	0.541	0.5	0.16	0.328	+0.192n.s.	+0.625n.s.	-
LEX017	4	1.98	0.522	0.506	0.410	0.124	0.227	-0.120n.s.	+0.500n.s.	-
LEX019	7	5.52	0.808	0.763	0.728	0.374	0.556	-0.329**	-0.080***	+0.058*
LEX031	4	2.87	0.314	0.656	0.583	0.217	0.369	+0.163*	+0.547***	+1***
SGCV23	19	4.66	0.750	0.790	0.770	0.447	0.627	+0.072n.s.	+0.033n.s.	+0.048n.s.
TKY321	9	4.98	0.856	0.802	0.770	0.430	0.607	-0.151*	-0.096***	-0.144n.s.
VHL20	6	4.74	0.867	0.816	0.756	0.403	0.581	-0.064n.s.	-	-
VIASH39	6	3.21	0.644	0.692	0.635	0.267	0.434	+0.336*	+0.156***	-0.021n.s.
Mean	8.22	3.99	0.637	0.711	0.662					
Cumulative						0.975	0.998			

[1]Number of alleles; [2]effective allele number; [3]observed heterozygosity; [4]expected heterozygosity; [5]polymorphism information content; [6]total exclusionary power (first parent); [7]total exclusionary power (second parent); [8]Weir and Cockerham (1984); Markers ASB3, LEX019, LEX031, SGCV23, TKY321, and VIASH39 were tested in a total of 115 animals. Markers LEX014, LEX017, ECMPZ001, and VHL20 were tested in a subset of this sample (30 animals). n.s. = non-significant; * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$.

Significant deviations from Hardy-Weinberg equilibrium proportions were observed for ASB3, LEX019, LEX031, TKY321, and VIASH39 when each subpopulation was analyzed separately. When all subpopulations were analyzed together, significant deviations from Hardy-Weinberg equilibrium proportions (heterozygote deficit) were only detected for ASB3 and LEX031. P values close to zero were obtained in a global test for heterozygote deficit, suggesting that the deviations observed were not due to random effects.

Under a Bayesian approach, the most suitable number of clusters for modeling this data set was three. Under a three-cluster model, a conspicuous structure appeared, with 97.1 to 98.5% of the individuals being correctly assigned to their subpopulation of origin (probability of inclusion in the proper cluster >0.9). Six animals were assigned to their proper subpopulations with probabilities ranging from 0.789 to 0.885, and only one animal could be assigned to more than one cluster. The proportion of genetic variability due to differences among subpopulations was 5.62% (F_{ST} = 0.0562) (Table 2), while 94.38% could be ascribed to genetic variability within the subpopulations. Significant differences ($P < 0.05$) in population pairwise F_{ST} were observed between Northeast and South (0.017), Southeast and South (0.053), and Southeast and Northeast (0.175).

Table 2. Wright F-statistics.

Locus	F_{IS}	F_{IT}	F_{ST}
ASB3	0.3584	0.4034	0.0701
LEX019	-0.1164	-0.0369	0.0712
LEX031	0.5735	0.5992	0.0604
SGCV23	0.0353	0.0545	0.0198
TKY321	-0.1340	-0.0433	0.0800
VIASH39	0.0769	0.1072	0.0328
Overall	0.1238	0.1731	0.0562

The mean pairwise inbreeding coefficient (*f*) was estimated to be 0.11, with values equal to or smaller than zero for 53% of the pairs of individuals, suggesting unrelatedness. *f* values in the 0.01-0.2 interval were found for 36% of the pairs and between 0.2 and 0.5 for 10% of them. Approximately 1% of the pairs (42 pairs) showed *f* values higher than 0.5, and in half of them at least one individual came from the South subpopulation. No pairs with *f* values higher than 0.5 were observed between the Northeast and Southeast subpopulations.

DISCUSSION

In an analysis of the genealogical data from the MM breed association (Associação Brasileira dos Criadores do Cavalo Mangalarga Marchador - ABCCMM - http://www.abccmm.org.br) referring to animals born from 1949 to 1999, the effective number of ancestor animals was estimated to be approximately 35 individuals, which would be responsible for up to 50% of the genetic pool in the current population. The current inbreeding coefficient, based on pedigree registers since the studbooks were closed, was estimated to be 1.31% (Costa et al., 2004, 2005). However, since the introduction of artificial insemination and embryo transfer in 1994, the mean number of progeny among top mares increased from one to six per year. Meanwhile, the maximum number of registered offspring per stallion increased from 176 to 1322 (Costa et al., 2004). High numbers of progeny have a direct impact on the effective population size (N_e). Therefore, the inbreeding coefficient (*F*) will probably increase in this population over time.

In the last two decades, studies on genetic diversity and population structure have been conducted for many horse breeds, including some small, endangered feral populations. The MM breed is a large commercial breed that has been developed over a period of less than 200 years. Although some geographical isolation may have prevailed up to the 1950s, some amount of gene flow among these populations occurred most of the time. Since the introduction of assisted reproduction technologies in 1994, there have been increasing forces towards population homogeneity. Artificial selection effects are difficult to foresee, but may strongly impact genetic diversity and create population substructures.

Genetic diversity may be broadly inferred from mean heretozygosity. Low genetic diversity was reported for the Spanish Purebred (0.452), Sorraia (0.486) and Uruguayan Creole (0.503) horse breeds, and higher values were found in Lusitano (0.735) and Garrano (0.788) horses. Intermediate values, similar to those reported here for the MM breed, were also described for Lippizan (0.663), Sable Island (0.626), and Anglo-Arab (0.66) horses (Kelly et al., 2002; Tozaki et al., 2003; Achmann et al., 2004; Luís et al., 2002, 2005; Plante et al., 2007). Recently, a relatively high mean PIC (0.700) was reported for the Pantaneiro horse, an endangered Brazilian feral horse breed (Giacomoni et al., 2008).

Evidence for genetic substructuring of the MM breed was found in multilocus F_{ST} values (5.62%; $P < 0.05$) and on cluster analysis, where under a three-cluster model 93.92% of the 115 unrelated individuals sampled were correctly assigned to the subpopulation from which they originated. Additionally, mean pairwise *f* was relatively high (0.11), and in spite of the efforts toward a non-consanguineous sampling, 1% of the pairs of individuals shared over 50% of alleles.

Taken together, these results suggest that the MM breed is divided into geographic subpopulations, and that in spite of its relatively large population size (300,000 registered animals), the estimated effective population size is proportionally low (N_e = 9,174.24; Costa et al., 2004), as well as the mean observed heterozygosity (0.637; see Table 1).

Artificial selection reduces effective population size and, therefore, may reduce genetic diversity. Local management creates genetic subdivisions within breeds, leading to some degree of reproductive isolation. Besides, artificial reproduction techniques can rapidly decrease genetic diversity, but can also lead to some gene flow into small subsets of a breed. Molecular markers provide valuable information for monitoring artificial selection. They may be used to avoid loss of genetic diversity during selection, helping in the identification of animals or sets of animals that should be kept in the breeding process just to prevent loss of genetic diversity. Estimation of genetic relatedness between individuals may also be helpful in avoiding crosses between highly endogamic animals.

ACKNOWLEDGMENTS

We thank Associação Brasileira dos Criadores do Cavalo Mangalarga Marchador for collecting blood samples, particularly, veterinarians Martinho Bretas, Mitzen Bretas and, especially, Cornélio Maciel for incentive and technical and logistic support. Research supported by Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG). M.G.C.D. Peixoto was the recipient of a FAPEMIG fellowship.

REFERENCES

Achmann R, Curik I, Dovc P, Kavar T, et al. (2004). Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. *Anim. Genet.* 35: 285-292.

Breen M, Downs P, Irvin Z and Bell K (1994). An equine tetranucleotide repeat: microsatellite MPZ001. *Anim. Genet.* 25: 123.

Breen M, Lindgren G, Binns MM, Norman J, et al. (1997). Genetical and physical assignments of equine microsatellites - first integration of anchored markers in horse genome mapping. *Mamm. Genome* 8: 267-273.

Coogle L, Reid R and Bailey E (1996a). Equine dinucleotide repeat loci LEX015-LEX024. *Anim. Genet.* 27: 217-218.

Coogle L, Reid R and Bailey E (1996b). Equine dinucleotide repeat loci from LEX025 to LEX033. *Anim. Genet.* 27: 289-290.

Costa MD, Bergmann JAG, Resende ASC, Martins GA, et al. (2004). Demographic characterization of Mangalarga Marchador breed. [Caracterização demográfica da raça Mangalarga Marchador]. *Arq. Bras. Med. Vet. Zootec.* 56: 687-690.

Costa MD, Bergmann JAG, Resende ASC, Fonseca CG, et al. (2005). Study on genetic subdivision of the Mangalarga Marchador horse breed. [Estudo da subdivisão genética da raça Mangalarga Marchador]. *Arq. Bras. Med. Vet. Zootec.* 57: 272-280.

Ewen KR and Matthews ME (1994). VIAS-H39, an equine tetranucleotide microsatellite repeat polymorphism. *Anim. Genet.* 25: 433.

Giacomoni EH, Fernandez-Stolz GP and Freitas TR (2008). Genetic diversity in the Pantaneiro horse breed assessed using microsatellite DNA markers. *Genet. Mol. Res.* 7: 261-270.

Godard S, Vaiman D, Oustry A, Nocart M, et al. (1997). Characterization, genetic and physical mapping analysis of 36 horse plasmid and cosmid-derived microsatellites. *Mamm. Genome* 8: 745-750.

Hardy OJ and Vekemans X (2002). SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2: 618-620.

Kelly L, Postiglioni A, De Andres DF, Vega-Pla JL, et al. (2002). Genetic characterisation of the Uruguayan Creole horse and analysis of relationships among horse breeds. *Res. Vet. Sci.* 72: 69-73.

Luís C, Cothran EG and Oom MM (2002). Microsatellites in Portuguese autochthonous horse breeds: usefulness for parentage testing. *Genet. Mol. Biol.* 25: 131-134.

Luís C, Cothran EG, Oom MM and Bailey E (2005). Major histocompatibility complex locus DRA polymorphism in the endangered Sorraia horse and related breeds. *J. Anim. Breed. Genet.* 122: 69-72.

Miller SA, Dykes DD and Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16: 1215.

Plante Y, Vega-Pla JL, Lucas Z, Colling D, et al. (2007). Genetic diversity in a feral horse population from Sable Island, Canada. *J. Hered.* 98: 594-602.

Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

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

Slate J, Marshall T and Pemberton J (2000). A retrospective assessment of the accuracy of the paternity inference program CERVUS. *Mol. Ecol.* 9: 801-808.

Tozaki T, Kakoi H, Mashima S, Hirota K, et al. (2001). Population study and validation of paternity testing for Thoroughbred horses by 15 microsatellite loci. *J. Vet. Med. Sci.* 63: 1191-1197.

Tozaki T, Takezaki N, Hasegawa T, Ishida N, et al. (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, Stott ML, Lenstra JA, et al. (1994). A highly polymorphic horse microsatellite locus: VHL20. *Anim. Genet.* 25: 207.

Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Yeh FC, Yang R-C and Boyle T (1999). POPGENE Version 1.31. Microsoft Window-based Freeware for Population Genetic Analysis. Quick User Guide. University of Albert, Centre for International Forestry Research. Available at [http://www.ualberta.ca/~fyeh/popgene.pdf]. Accessed December 12, 2007.