Statistical tools to aid domestic breeds' management.

J. Morais*, J. Malta-Vacas*, C. Luís* and M. M. Oom*

* Universidade de Lisboa, Faculdade de Ciências, Departamento de Biologia Animal/Centro de Biologia Ambiental, Edifício C2 – 3º Piso, 1749-016 Lisboa, Portugal

Abstract

Domestic breeds play a determining role on local economies. There are still many communities that depend on these breeds for their subsistence. For instance, the use of the horse has changed throughout times but horse breeds still have their importance within the agricultural environment. Genetic variability of domestic breeds is usually under very strict local management systems and suffers from the negative effects of selective pressure and cross breeding. For this reason, special attention must be paid to genetic erosion symptoms. If genetic identity of native breeds must be preserved, a delicate balance must be maintained between gene flow and genetic isolation of local populations. New clustering and assignment methods are important statistical aids to address this problem. Bayesian methods and frequency based methods were tested, in the present work, for their efficacy in allocating individuals to their populations of origin and detecting genetic structure within populations. We have genotyped 50 individuals form each of the three Portuguese native horse breeds as a case study and we compared Structure, GeneClass2 and Whichrun for their efficiency versus computing time. We also compared the performance of clustering methods opposing genetic distance measures to a Bayesian method implemented in Structure. Whichrun was the most effective allocation tool whereas genetic distance methods did not perform as well. We were able to confirm internal genetic structure within two of the tested breeds using Structure. Such genetic structure was not detected with traditional genetic distance methods, which could not differentiate between individuals of these two breeds. We have genotyped 50 individuals from each of the three Portuguese native horse breeds for 23 horse specific microsatellite markers (AHT4, AHT5, ASB2, HMS1, HMS2, HMS3, HMS6, HMS7, HTG10, HTG14, HTG4, HTG6, HTG7, HTG8, LEX20, LEX23, LEX36, LEX41, NVHEQ18, UCDEQ405, UCDEQ425, UCDEQ5 and VHL20) and we have used a Li-Cor 4200 Series automated sequencer with 6% poly-acrylamide gels to run and visualize the PCR products.

Methods

Statistical analysis:

<u>Genetic distances</u>: Proportion of shared alleles (Bowcock et al., 1994) (Dps) and the Reynolds et al.'s F_{ST} (1983) - *Microsat2* (Minch et al., 1998). Departures from Hardy-Weinberg equilibrium *Genepop v. 3.2a* (Raymond & Rousset, 1995).

<u>Clustering and assignment methods</u>: Unrooted neighbour-joining trees based on Dps values and on F_{ST} values between individuals - *Phylip 3.6* (Felsenstein, 1993). See Table 1 for software used: *Structure* is a clustering method and *Geneclass2* is rather an assignment tool. We have also used *Whichrun* as an assignemt tool (Banks & Eichert, 2000). We have followed the procedure by Evanno et al. (2005) to better detect the real number of clusters determined by Structure.

Table 1

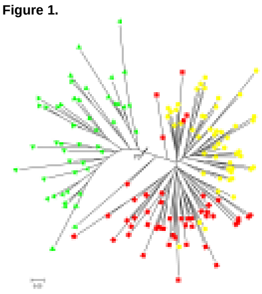
	Bayesian Clustering method - Structure (Pritchard et al., 2000) following the procedure described by Evanno et al., 2005) K=4*	*Bayesian assignment method - GeneClass2* (Piry *et al.*, 2004)	
		Rannala & Mountain (1997)	Paetkau *et al.* (2004)
Number of runs	20	10	10
Run parameters	50000 steps burn-in period + 300000 steps	10000 simulated individuals + significance level of 0.001	10000 simulated individuals + significance level of 0.001
Length of each run	90 min.	25 min.	61 min.

Note: runs performed by a laptop computer with a 1.8 Ghz Pentium 4 (Intel) processor and 512 Mb of RAM memory.
* This procedure was facilitated by a software created by Grosso, A.-R. (2005 - Unpublished) (Faculdade de Ciências da Universidade de Lisboa).

Results and Discussion

- No H-W departures were detected for all populations.

Figure 1.



© Oom, M.M. (Sorraia horses)

Figure 1 represents the N-J tree of individuals between the three Portuguese breeds based on F_{ST} values (Dps performed similarly with lower bootstrap).
○ Garrano horse; 🟥 Lusitano horse; △ Sorraia horse

Figure 2.

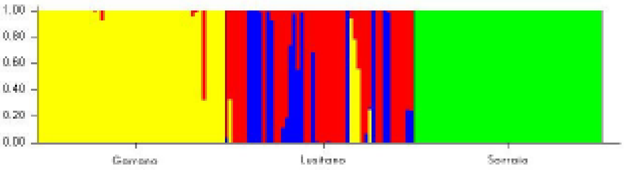


Figure 2 depicts *Structure* data plot (K=4) in terms of proportion of individuals present in the 8 determined clusters**.**

Table 2 indicates the clusters inferred by *Structure* and the assignment of individuals to those clusters. Numbers in bold indicate major clusters.

Inferred clusters K / Sampled populations	K=4			
	1	2	3	4
GAR	**0.016**	0.000	0.000	**0.984**
LUS	**0.646**	0.000	**0.296**	**0.058**
SOR	0.000	**1.000**	0.000	0.000
Est. ln Prob (over 20 runs)	-9677.2			
Ln Likelyhood (over 20 runs)	-9496.8			

Table 3 summarizes the performance of assignment softwares *GeneClass2* and *Whichrun*.

	Whichrun	*GeneClass2* (Rannala & Mountain, 1997)	*GeneClass2* (Paetkau *et al.*, 2004)
GAR	98%	100%	100%
LUS	98%	94%	93%
SOR	100%	96%	96%

1. Distance methods are not as effective as Bayesian methods for clustering individuals.

2. There is internal structure for the Garrano and the Lusitano breeds. This had been demonstrated in previous studies by Morais *et al.* (2003) and Morais *et al.* (2005). This is most likely a direct effect of the local breeding strategies that affect these breeds directly.

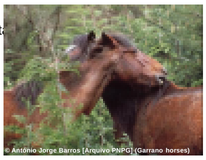
Possible explanations:

- The Sorraia is suffering more severely from the effects of founder effect and genetic drift (closed population since 1937), being more uniform and less variable.
- The Garrano and the Lusitano breeds suffered crossings with other breeds in a recent past, specially the Garrano, for which only recently the studbook was est

3. Assignment methods based in Maximum Likelihood (*Whichrun*) seem more effective than the Bayesian methods applied to the same finality:

- less time consuming
- more robust
- better rate of correctly assigned individuals

www.interagro.com.br
© Bob Langrish (Lusitano stallion)



© António Jorge Barros [Arquivo PNPG] (Garrano horses)

Assignment softwares are powerful tools to detect the origin of unknown individuals and when associated to the Bayesian based clustering tools can be a very useful aid to identify conservation units and establish management plans.

ACKNOWLEDGEMENTS

We thank Dr. João Costa-Ferreira, the director of the Serviço Nacional Coudélico for letting us perform some of the laboratory work at the Molecular Genetics Laboratory of Alter Stud farm (Alentejo, Portugal). Joana Morais and Cristina Luís are supported by PhD fellows from the Portuguese Foundation for Science and Technology (refs. SFRH/BD/10136/2002 and SFRH/BD/3318/2000, respectively). The study was also supported by Programa INTERREG-II/FEOGA) and by SNC/APSL

References

Banks M. A. & Eichert W. (2000) *Journal of Heredity* **91**, 87-89.
BowcockA. M., Linares A. R., Tomforde J., Minch E., Kidd J. R. & Cavalli-Sforza L. L. (1994) *Nature* **398**, 445-457.
Evanno, G., Regnaut, S. & Goudet, J. (2005) *Molecular Ecology* **14**, 2611-2620.
Felsenstein J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
Minch E., Ruiz-Linares A., Goldstein D., Feldman M. & Cavalli-Sforza L. L. (1998) Microsat 2: a computer program for calculating various statistics on microsatellite allele data (http://crick.stanford.edu/microsat).
Morais J., Malta-Vacas J., Luís C. & Oom M. M. (2003). *Comptes Rendus Biologies* **326**, S217-S221.
Morais J., Malta-Vacas J., Luís C. & Oom M. M. (2005) *Biochemical Genetics* **43**, 347-364.
Morais J., Malta-Vacas J., Luís C. & Oom M. M. (2005) 3º Congresso de Ciências Veterinárias da SPCV, Outubro, Fonte Boa, Santarém, Portugal
Paetkau D., Slade R., Burden M. & Estoup A. (2004) *Molecular Ecology* **13**, 55-65.
Piry S., Alapetite A., Cornuet J.-M., Paetkau D., Baudouin L. & Estoup A. (2004) *Journal of Heredity* **95**, 536-539.
Pritchard, J. K., Stephens M. & Donnelly P. J. (2000) *Genetics* **155**, 945-959.
Rannala B. & Mountain J. L. (1997) *Proceedings of the National Academy of Sciences* **94**, 9197-9201.
Raymond M. & Rousset F. (1995) *Journal of Heredity* **86**, 248-249.
Reynolds J., Weir B. S. & Cockerham C. C. (1983) *Genetics* **105**, 767-779.