Abstract no. 424; Session G4.44
joanav@mail.icav.up.pt

EFFECT OF SMALL-HERD CLUSTERING ON THE GENETIC CONNECTEDNESS OF THE PORTUGUESE HOLSTEIN CATTLE POPULATION

J.Vasconcelos*[1,2] and J.Carvalheira[1,2]

[1]'Centro Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Universidade do Porto, Rua Padre Armando Quintas, 4485-661 Vairão, PORTUGAL,
[2]ICBAS-Universidade do Porto

INTRODUCTION

Accuracy of genetic evaluations is directly related with connectedness (Kennedy and Trus, 1993; Hanocq and Boichard, 1999) between management groups (**MG**). Contemporary groups in test-day (**TD**) models may have an unacceptable number of observations (e.g., less then 4) depending on the herd size. Clustering these herds based on, for example, the simultaneous similarity between phenotypic mean and STD may prevent the loss of this information. However, these techniques may induce a false increase in genetic connectedness (**GC**), inflating the accuracy of EBV and promote preferential treatment for some sires. The aim of this study was to evaluate possible effects of clustering herds on the GC of the Portuguese Holstein cattle.

MATERIAL AND METHODS

The Portuguese National Association of Dairy Cattle Breeders (**ANABLE**) provided TD data, produced by 194,980 animals from 1997 to 2003. Ideally, average prediction error variance of differences (**PEVd**) in EBV between animals on different MG should be the basis of measurement of GC (Foulley *et al*, 1992; Kennedy and Trus, 1993). In a dataset of this dimension the estimation of PEVd is computationally infeasible. Two alternative methods were applied to compute the degree of GC:

GLT - Total number of direct genetic links between MG (Roso *et al*., 2004), defined in this work as herd/year:
GLT was calculated by summing the total number of common sires and dams in pairs of MG, forming genetic links (Fries, 1998). For each MG, the overall number of genetic links with all other MG was calculated.

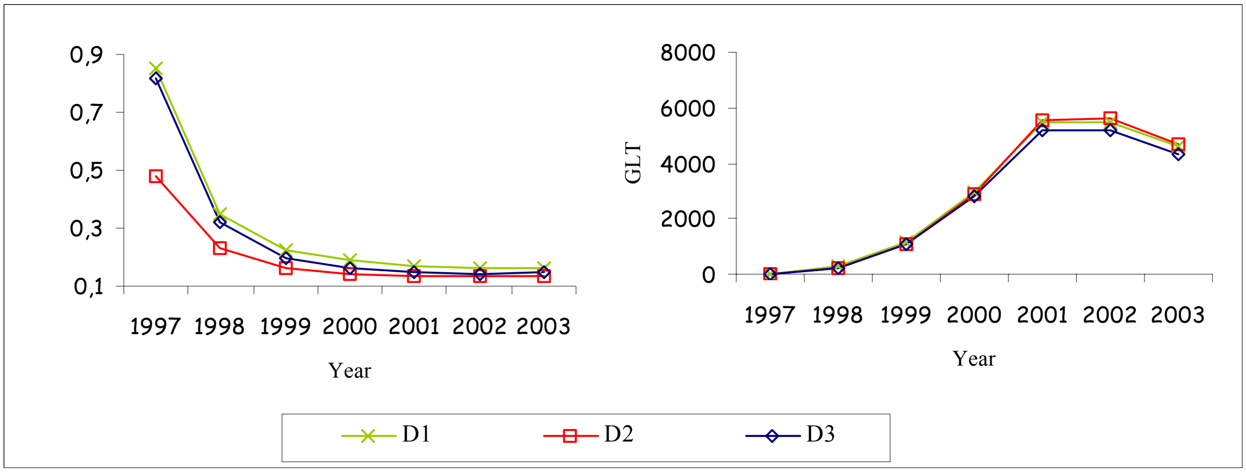
GDV - Genetic drift variance (Kennedy and Trus, 1993):
It was obtained from X'ZAZ'X matrix, that measures the sum of genetic relationship between and within MG. Average relationship between and within MG were obtained by dividing diagonal elements of X'ZAZ'X matrix by the square of the number of records in the MG and the off-diagonal elements by the product of the number of records in each of the MG considered. The resulting matrix, $\bar{\mathbf{A}}$, can be interpreted as the components of the genetic drift (co)variances. The average GDV for each MG with all other was calculated as the average of all variance differences with respect to that MG.

Both methods were applied to 3 datasets: **D1**-containing all herds, **D2**-same as D1, but the small herds (with contemporary groups having less then 4 observations) were clustered based on the simultaneous similarity between phenotypic mean and STD (TD average at 1st lactation) and **D3**-same as D1, but small-herds were removed from the analysis.

RESULTS AND DISCUSSION

Large values of GLT and small values of GDV are preferred, indicating a high level of GC. Results from GLT and GDV for all datasets (Fig. 1) indicate an increase in GC from 1997 to 2001, staying approximately constant towards the end of the period, with a slight decrease in 2003. Table 1 shows the correlations for GLT between all datasets. The observed high correlation obtained in this study suggests that clustering had a negligible effect on GC. The GDV method gave similar results (Table 1), confirming that clustering may be a reliable procedure to join herds with few observations, without significant changes on accuracy of prediction.

Fig. 1. Annual average degree of genetic connectedness between management groups (herd/year) on the basis of total number of direct genetic links (GLT) and genetic drift variance (GDV) [a].



[a]**D1**= contains all herds,
D2= same as D1, but small herds were clustered based on the simultaneous similarity between phenotypic mean and STD,
D3= same as D1, but small herds were removed from the analysis.

The computation of GDV for this study was very costly in CPU-time, especially because of the large number of animals considered in the additive relationship matrix. Kennedy and Trus (1993) obtained a high correlation between GDV and the average PEVd, which, for the purpose of this study, makes the GDV the standard method. Correlations between GDV and GLT were negative, as expected, and ranged from –0.70 to –0.68. This relatively high correlation supports the use of GLT as a viable method to compute GC in large datasets.

Table 1. Correlations between genetic connectedness in the three datasets (D1, D2 and D3) computed by two methods: total number of direct genetic links (GLT - bellow diagonal) and genetic drift variance (GDV – above diagonal).

	D1	D2	D3
D1	-	0,9997	0,9997
D2	0,9998	-	0,9996
D3	0,9999	0,9998	-

CONCLUSIONS

Clustering small herds based on the simultaneous similarity of the phenotypic mean and STD, had a negligible effect on GC of the Portuguese Holstein cattle. These results suggest that changes in accuracy of prediction or the systematic use of sires between clusters, may be minimized, but needs confirmation. Although cluster techniques may be a feasible procedure to join herds with few observations, further studies are necessary to validate their use on genetic evaluations. The next step of this project will evaluate and compare the application of different clustering techniques, their impact on EBV rankings and possible consequences of herds changing clusters between genetic evaluations.

REFERENCES

Fries, L. A. 1998. Proc. 6th World Cong. Genet. Appl. Livest. Prod., Armindale, NSW, Australia. 27:449-500.
Foulley, J. L., E. Hanocq, D. Boichard. 1992. Genet. Sel. Evol. 24:315-330.
Hanocq and Boichard. 1999. Genet. Sel. Evol. 31(2):163-176.
Kennedy, B. W., and D. Trus. 1993. J. Anim. Sci. 71:2341-2352.
Roso, V. M., F. S. Schenkel, S. R. Miller. 2004. Canadian J. Anim. Sci. 84(1):37-47.