Characterisation of linkage disequilibrium and subsequent estimation of effective population size in Thoroughbred horses using single nucleotide polymorphism markers

L J Corbin[1], S C Bishop[1], J E Swinburne[2], M Vaudin[2], S C Blott[2], J A Woolliams[1]
[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom, [2]Animal Health Trust, Newmarket, United Kingdom   *Email: laura.corbin@roslin.ed.ac.uk*

Introduction The extent of linkage disequilibrium (LD) within a population is important when determining the number of markers required for genetic analyses such as association studies and can also impact on the likely accuracy of both marker-assisted and genomic selection. It has been proposed that the average LD over different genomic distances can be used to infer ancestral effective population size (de Roos *et al.* 2008; Tenesa *et al.* 2007). Whilst the extent of genome-wide LD has been investigated in a variety of livestock species, including cattle, pigs and chickens, this paper represents one of the first such studies with respect to the equine.

Materials and methods The data for this study consisted Illumina Equine SNP50 BeadChip genotype data for 817 UK Thoroughbred horses. 28% of the 54,602 available single nucleotide polymorphism (SNP) markers were excluded due to monomorphism, poor genotyping quality (genotyping in <95% of samples), deviation from Hardy-Weinberg equilibrium (p<0.0001) and low minor allele frequencies (<0.05). The LD measure r^2 was calculated for all remaining syntenic markers pairs and expressed as a function of genomic distance for each chromosome. Marker pairs were broadly divided into three distance (Mb) categories: [0,0.1], [0.1,20] and [>20]. Within each category, marker pairs were ordered on the basis of distance and divided into 20 equally sized bins, each containing markers of approximately equal pairwise genomic distance. The effective population size at different stages in the past was estimated from the average r^2 and average marker distance of each of the bins. The equation used for this estimation was $N_T = (1/4c(1/r^2-1))$, where N_T is the effective population T generations ago, and c is the marker distance in Morgans, assuming 1Mb=1cM, and $T = 1/2c$ (Hayes *et al.* 2003; de Roos *et al.* 2008; Tenesa *et al.* 2007).

Results Average r^2 decreased with increasing genomic distance, as expected. The decline in LD appears to be less rapid than that seen in cattle populations, with mean r^2 values at a mean distance between markers approaching 0.1Mb maintained above 0.2 for nearly all chromosomes. By way of example, in the case of Chromosome 1, at mean marker distances of 10kb, 40kb and 100kb, the mean r^2 values (±SE) were calculated as 0.48 (0.022), 0.34 (0.018) and 0.31 (0.017), respectively. The results of most interest in terms of the prediction of past effective population size, were those produced in the distance category 0.1Mb to 20.0Mb; the number of generations ago calculated from marker distances in this group ranged from two to 100 generations in the past and the Thoroughbred population was formed around 300 years ago (representing approximately 25-30 generations). The results for Chromosome 1 can be seen in Figure 1. The pattern of N_T observed is much different to that seen in cattle (de Roos *et al.* 2008).
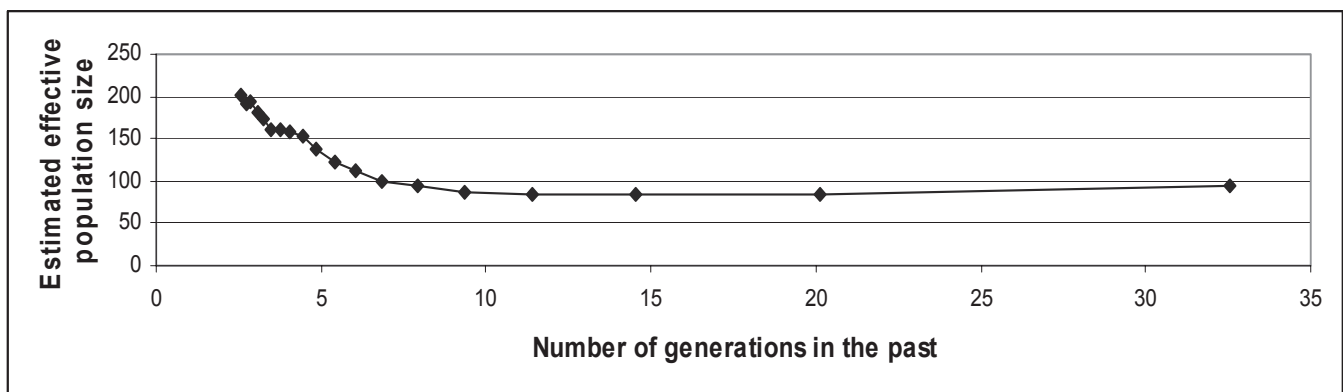

Figure 1 Effective population size along the population history, estimated from the average r^2 at different marker distances of markers on chromosome 1 (the number of generations in the past was truncated at 35).

Conclusion In the Thoroughbred population studied, LD declined with increasing genomic distance but at a less rapid rate than that seen in cattle populations. Estimations of ancestral effective population size suggest that the breed underwent a population expansion, relative to its starting population size, beginning approximately ten generations ago; a pedigree analysis would be useful to validate this hypothesis. The potential effect of population stratification on these analyses has not yet been investigated but is of interest.

References
Hayes, B.J., Visscher, P.M., McPartlan, H.C. and Goddard, M.E. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research 13, 635-643.
De Roos, A.P.W., Hayes, B.J., Spelman, R.J. and Goddard, M.E. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus Cattle. Genetics 179, 1503-1512.
Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E. and Visscher, P.M. 2007. Recent human effective population size estimated from linkage disequilibrium. Genome Research 17, 520-526.