**Application**

The objective is to explore the feasibility of an improved test for differentiating between traditional and modern Hereford cattle populations, preventing admixture with external bloodlines, and utilizing machine learning to classify animals with missing pedigree information.

**Introduction**

The purebred Hereford Original Population (HOP) cattle breed has been present in the United Kingdom since 1627. It originated in Herefordshire, England, and was known for producing high-quality, rich meat despite poor grazing conditions (Willham, 1937). ​The Hereford breed became popular for beef production worldwide, leading to crossbreeding with outside breeds to improve productivity and toughness in local environments (Blott et al., 1998). ​The HOP is the progenitor of the more modern commercial North American Derived (NAD) cattle which has longer legs and larger body size. Previous studies have shown that two groups are genetically distinct due to introgression of genetic material from other breeds into the NAD (Hegarty et al., 2020).

**Materials and Methods**

Breed Assignment and Clustering of Genotypes​

Cattle genotypes from Hereford Herdbook International (HHBI) were combined with previous data from McMahon *et al.* (2015) using the 7K BovineLD assay, to investigate genetic differences between the HOP and NAD populations. GeneSeek data was received in PLINK format and individual files were merged with data from McMahon *et al.* (2015), resulting in a final merged set of 4593 SNPs. Multidimensional Scaling (MDS) analysis was used to identify major sources of variation within the data. ​

Whole genome resequencing​

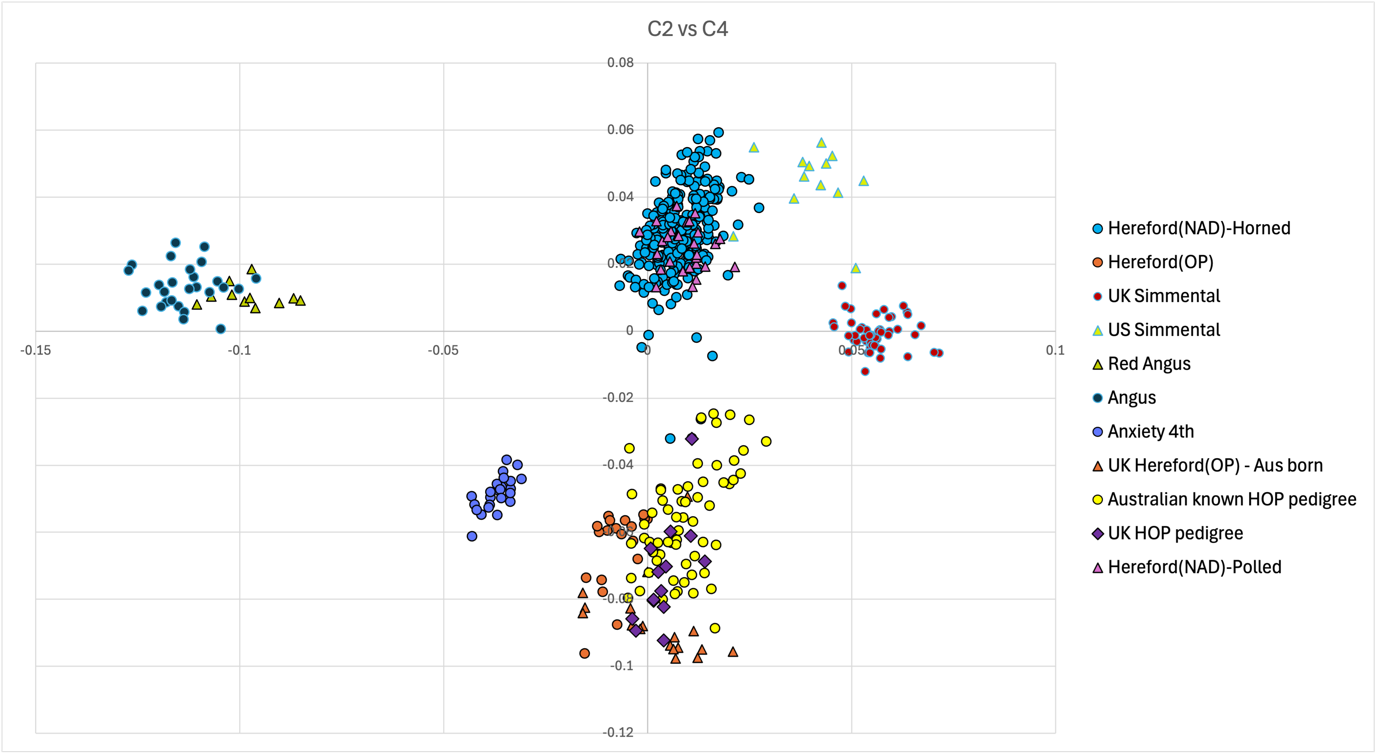
~12-50x genome resequencing data was obtained for 20 animals (mix of HOP, NAD and those of suspect or mixed pedigree). Data for 5 animals/breed for 12 other breeds was downloaded from Sequence Read Archive (SRA). Reads were aligned to the ARS-UCD1.2 bovine reference genome via Bowtie2 and used to call SNPs (~7 million) via SAMtools/bcftools. These were then employed for Principal Component Analysis (PCA) using a custom Python script to identify relevant PC loadings.​

Machine Learning

PCA was employed for dimensionality reduction. We applied non-linear classifier methods; Support Vector Machine (SVM) algorithm and trained the model with K fold (k=5) stratified cross-validation strategies.

**Results**

Multidimensional scaling analysis shows a clear division between Hereford (OP) and Hereford (NAD) (Figure 1). To show the strength of the division between groupings, we also plotted two non-Hereford breed groups - UK and US Simmental profiles (as examples of genetic variation due to environment/breeding practice) and Black/Red Angus (as examples of recently diverged breeds). The clear pattern of division between Hereford (OP) and Hereford (NAD) suggests that the two have been separated by a significant genetic shift, not likely to result from genetic drift or different breeding practices (Figure 1).

**Figure 1.** MDS plot of HOP vs NAD, with reference breeds

For the resequencing analysis, it was found that PC1 split the two breeds. Significant marker loadings were extracted and plotted against the reference genome to identify candidate genes near to these regions, as shown in Figure 2. Candidate genes were then subjected to literature search to identify possible function and any known role in cattle biology.

Some of the candidate genes are related to phenotypic traits involved in meat/milk production, like the MAATS1 gene linked to carcass fat in Nellore cattle, or the ARHGEF26 gene associated with milk cholesterol content in dairy cattle. Other genes are also related to phenotypic traits which indirectly affect economic performance, such as the APP gene that affects cattle temperament, or the DCBLD2 gene linked to hoof disorders and lameness in cattle.

A screen shot of a graph

Description automatically generated

**Figure 2.** Results of PCA loading analysis for chromosome 1, based on WGS resequencing

**Conclusions**

This study identified clear genetic distinctions between the Hereford Original Population (HOP) and the North American Derived (NAD) Hereford cattle. The significant genetic shifts detected are unlikely to have occurred exclusively due to genetic drift or different breeding methods. Machine learning techniques have improved the ability to categorize breeds and manage complex datasets. Overall, the findings indicate the feasibility of developing an improved genetic test for distinguishing traditional and modern Hereford cattle populations, which is essential for maintaining the genetic integrity of the HOP and preventing undesirable admixture with external bloodlines.

**Acknowledgements**

The authors express gratitude to Neogen® for their assistance during the study and to Hereford Herdbook International (HHBI) for their contributions to the Ph.D. bursary.

**References**

Blott, S. C., Williams, J. L., & Haley, C. S. (1998). Genetic variation within the Hereford breed of cattle. Animal Genetics, 29(3), 202–211.​

Hegarty, M., Broadbent, Z. E., & McMahon, R. (2020). Towards the development of a genetic test for distinguishing the Hereford Original Population (Hereford (OP)) from the Hereford North American Derived (Hereford (NAD)) cattle in the UK, Australia and North America, Aberystwyth University.​

McMahon, R., Avila, V., Debbonaire, A., Patten, E., & Hegarty, M. (2015) Verification of breed, individual identity and farm of origin of beef carcasses using genetic assays. Report to Waitrose Supermarket, Aberystwyth University.​

Willham, O. S. (1937). A genetic history of Hereford cattle in the United States. Journal of Heredity, 28(8), 283–294.