Sub-populations within the US Holstein breed can be revealed and used to manage genetic diversity

[1]Yvette Steyn, [2]Thomas J. Lawlor, [3]Andres Legarra, [1]Daniela A.L. Lourenco, [1]Ignacy Misztal
[1]University of Georgia, Athens, GA 30602, USA [2]Holstein Association USA Inc, Brattleboro. Vermont, 05302, USA [3]GenPhySE INRA, Université de Toulouse, Castamet-Tolosan, 3120, France

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Introduction

The existence of sub-populations within a breed can help maintain trait variation and genetic diversity. This is important for long-term genetic gain, as well as the ability to adapt to changes in the environment and consumer demands. The extensive use of reproductive biotechnology in the dairy industry has led to higher inbreeding levels and concern about the loss of diversity. Genetic redundancy is a phenomenon where an excess of beneficial variants exist, which allows multiple genetic pathways to achieve the same phenotype (Goldstein & Holsinger, 1992). Quantitative traits are affected by many different genes, each with a small contribution to the trait expression, which enables genetic redundancy. This can allow sub-populations to change differently over time, which in turn, maintains underlying genetic diversity. The objectives of this study were to identify sub-populations within the US Holstein breed, determine whether these population-specific SNP effects could change the ranking of animals, and observe genetic redundancy (non-parallel changes) over time.

Materials and Methods

Pedigree, performance recording, and genotypes up to 2014 were obtained for the US Holstein population from the USDA. Sires of animals born after 2010 and females with type traits measured in 2012 and 2013 were chosen as selected candidates – animals that would have been considered as breeding animals in 2014. This totaled 20,990 animals. K-means clustering was performed on the genomic relationship matrix to identify clusters (C1, C2, C3, C4, C5). The blupf90 family software were used for further analyses (Misztal et al. 2014). A GWAS for stature was performed within each cluster or with all combined. Indirect genomic predictions (IGP) were obtained for male animals of each cluster when using SNP effects obtained using the females of each cluster. Animals were ranked based on these different IGP. The genetic correlations between the clusters were calculated as the correlation between the five different IGP of the males in each cluster adjusted using the method by Calo et al. (1973). This is an adjusted version of the method by Duenk et al. (2020). Hypothetical matings were performed to determine the expected inbreeding of offspring if males from a specific cluster were mated to females from each cluster. Higher inbreeding was expected within cluster. The fixation index was calculated for each SNP marker using the formula $F_{st} = \frac{s_{\bar{p}}^2}{\bar{p}(1-\bar{p})}$, where p is a vector containing the mean allele frequency across the five clusters and the numerator is the sampling variance of the allele frequency of each SNP across clusters (Wright, 1943). Five families were formed from the respective clusters (F1,F2 F3, F4, F5) by tracing parents back for ten generations (G0, G1,...,G10). The allele frequency (AF) of each SNP was calculated for each generation within a family. To see changes over time, specific SNP markers were identified to plot. Figures show the changes in allele frequency over time for the selected SNP and the 20 nearest SNP markers. This included *DGAT1* due to its known importance in the dairy population, the SNP with the greatest variance of change across families, and the SNP with the greatest range between the family with the least change and the family with the greatest change. The replicate frequency spectrum was used to evaluate whether SNP changed differently within families (Barghi et al. 2019). The 100 SNP markers with the greatest absolute change were identified within each family. Of these sets of 100 SNP markers, those that have changed by more than 0.30 were identified in each separate family.

Genetic redundancy

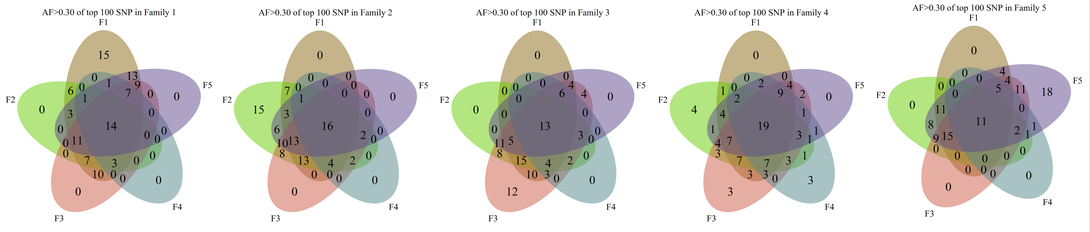

Figure 4: The Replicate Frequency Spectrum approach identifies markers that have changed similarly across families. Each diagram contains the 100 SNP markers that have changed the most in a particular family (here Families 1, 2, and 4). The diagrams identify the number of these markers that have changed allele frequency (AF) by more than 0.30 in the different families. In all scenarios other than Family 4, there are more than 12 markers that have changed by more than 0.30 in only the family the criteria was based on. All scenarios have fewer than 20 markers that have changed at the same rate in all families.

Results and Discussion

The average within-cluster inbreeding is generally higher (from 0.10 to 0.22) than across-cluster (from 0.10 to 0.12). Cluster 4 consistently produced low inbreeding within and across clusters. This is the largest cluster, containing those that remain after more distinct clusters have formed. It contains enough diversity within itself to allow low inbreeding. The average F_{st} of all SNP markers across the five clusters was 0.03, which is approximately half of what was found across three French dairy breeds (Flori et al. 2009). Figure 2 shows the significance level of the F_{st} of each marker. Although none reached significance, a few came close.

Each pair of clusters has two values for genetic correlations – one when SNP effects from cluster X is used to predict cluster Y, and one when cluster Y is used to predict cluster X. The average genetic correlations between clusters was 0.75, but varied from 0.41 to 0.99.

Performing GWAS in separate clusters delivered different results. However, among the 20 SNP with the greatest effect in each cluster, only chromosome 5 was common to all. Table 1 shows that considerable reranking occurred when using different SNP effects. Based on such reranking, it may not be appropriate to evaluate the whole US Holstein population together. Potential GxE interactions and different regional breeding objectives should be accommodated for.

Between 22 to 59 SNP markers reached fixation within a family, but only three did so in the overall population. This shows that, although separate lines can lose genes that might be important in the future, they could still be retained in the overall population and reincorporated at a later stage if necessary. Non-parallel changes in allele frequency over time was observed, both visually using selected SNP (Fig 3) and using the RFS (Fig 4). Visual inspection reveals changes surrounding the selected SNP that reflect hitchhiking, potential epistatic effects, and different selection pressure

Reranking based on breeding values

Table 1: The ranking of specific bulls based on indirect genomic predictions of stature when SNP effects are based on different groups

Bull cluster	Cluster SNP effects were based on						
	Name	All	C1	C2	C3	C4	C5
Cluster 1	Doorman	43	158	825	871	152	777
Cluster 2	Airlift	1	101	3	64	44	276
Cluster 3	Edison	21	94	15	1	233	50
Cluster 4	Chuckie	12	892	173	26	1	234
Cluster 5	Monreal	2	46	97	140	30	2

Conclusions

Genetic redundancy is a powerful phenomenon that can allow long-term selection while still maintaining diversity within the population. This can allow companies to develop their own genetic lines for the same breeding objectives without leading to detrimental effects in the overall US Holstein population. These different lines can be crossed in future if genetic variation decreases in a single line.

References

Barghi et al. 2019. Plos Biol 17:e3000128
Calo et al. 1973. J Anim. Sci 37(3):676-682
Duenk et al. 2020. G310(2):783-795
Flori et al. 2009. Plos One: 4(8):e6595
Goldstein & Holsinger 1992. Evolution 46(2):412-429
Misztal et al. 2014 Manual for BLUPF90 family programs
Wright 1943. Genetics 28(2):114-138

Results

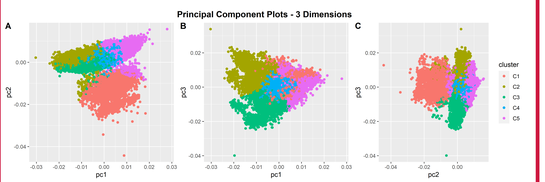

Figure 1: The first three principal components of the genomic relationship matrix

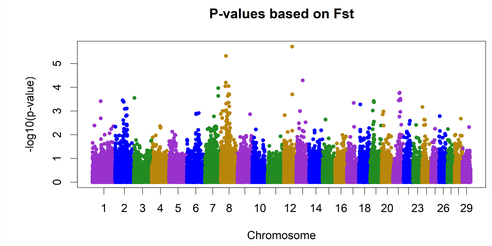

Figure 2: The significance of the Fixation Index (F_{st}) for each SNP across clusters

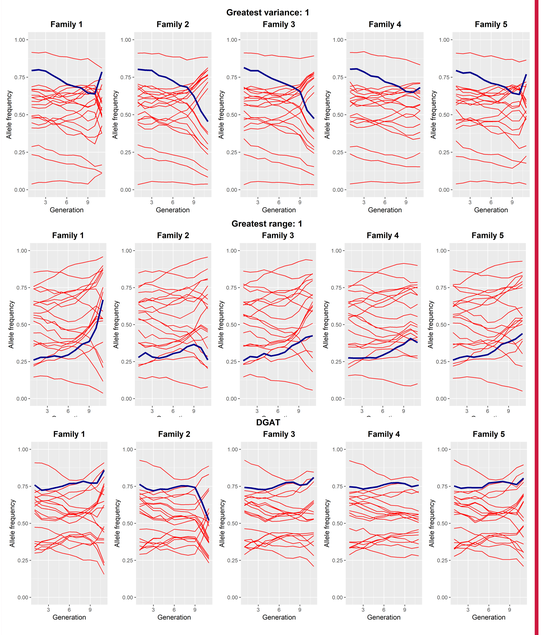

Figure 3: The changes in allele frequencies over time for selected SNP markers (in blue) and the 20 surrounding SNP markers (in red) over 10 generations within each family