211. Population structure of U.S. Holsteins allows for a snapshot of allele frequency changes and family specific SNPs

T.J. Lawlor[1*], Y. Steyn[2], S. Tsuruta[2], Y. Masuda[2], D.A.L. Lourenco[2] and I. Misztal[2]

[1]Holstein Association USA Inc., 1 Holstein Pl, Brattleboro, VT 05301, USA; [2]University of Georgia, 425 River Road, Athens, GA 30605, USA; tlawlor@holstein.com

Abstract

Genetic change occurs in the U.S. Holstein population through the heavy use of specific bulls for a relatively short period of time. By focusing on a specific time-period, we were able to group a high percentage of the descendants of several prominent bulls into five different clusters. The average F_{st} across clusters was 0.03. Comparison between clusters revealed a heterogeneous mixture of allele frequency changes with varying degrees of magnitude and direction. Non-parallel responses between families suggests alternative goals and/or non-additive gene action. SNP effects for the trait stature were estimated independently for the five clusters and used to predict additive breeding values (BV). Correlations of within-cluster BVs with BVs based upon all animals combined varied from 0.70 to 0.88. By stratifying a population into subpopulations, family specific SNPs can be identified and used to increase or maintain genetic diversity.

Introduction

A top dairy bull can have tens of thousands of daughters and an even larger number of granddaughters. These family members, with similar degrees of relationship, appear at different time periods. For example, the first group of descendants of a young prominent bull will be his daughters. They have an additive genetic relationship of 0.5 with their sire. After another year or two, the sons of this prominent bull will start producing granddaughters, and there will be a mixture of both daughters and granddaughters in our population. And after another couple of years, the additive genetic relationship will drop to 0.25 as almost all the living descendants will be granddaughters. As we move through time, the average relationship of this prominent bull with our active breeding population will decrease as his descendants become more distant.

K-mean clustering results in animals with similar relationships being grouped together. By focusing on a specific time-period, the descendants of certain bulls are clustered together. That is the progeny of the most prominent young sires are clustered together, descendants of the slightly older sires with breeding age sons are grouped into a separate cluster, and so on down to the final cluster of animals with the lowest average relationship amongst the current breeding population. When looking at GWAS results of prominent Holstein sires, Ma *et al.* (2012) stated:

> Artificial insemination and genetic selection are major factors contributing to population stratification in dairy cattle. Artificial insemination has been widely used and this has increased the likelihood of the presence of related individuals in randomly selected samples and the presence of large half-sib families; both of which contribute to population stratification.

Materials & methods

A total of 20,099 genotyped Holstein animals were used as target population of breeding animals born between 2010 and 2012. They included 3,902 males with at least 25 progeny and 16,197 females with a classification score. K-means clustering with a built-in R package (kmeans) was applied on the genomic relationship matrix (G) to separate the population into five separate clusters. The pedigrees of the target animals were traced back for 10 generations. The frequency of each SNP was calculated for each generation. Changes in allele frequency (AF) is defined as the difference in frequency of each SNP between

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_211, © T.J. Lawlor *et al.* 2022

two successive generations. The F_{st} across clusters, a measure of the variation in allele frequencies among subpopulations, was calculated as in Bonhomme *et al.* (2010).

Expected inbreeding within or across cluster was calculated based on pedigree information assuming non-zero inbreeding for unknown parents (Aguilar and Misztal, 2008) with the INBUPGf90 package within the BLUPF90 software suite (Misztal *et al.*, 2014). The genetic correlations across clusters were estimated using stature as trait and the method by Duenk *et al.* (2020). This correlated the additive breeding value of a cluster using the SNP effects of the same cluster, or when using the SNP effects of another cluster. Female animals of each cluster were used to estimate SNP effects with the POSTGSF90 package (Misztal *et al.*, 2014) and applied on the males of each cluster. Unlike the method by Duenk *et al.* (2020), the correlations were adjusted using the method by Calo *et al.* (1973). Assuming a heritability of 0.45 in all clusters, this adjustment factor is 1/0.45.

Results & discussion

Successful population stratification was confirmed in several ways. PCA analysis validated the clustering of target animals into five different but overlapping subpopulations. Inbreeding analysis showed higher within cluster relationships and lower relationships between clusters. Average inbreeding by cluster was 0.22, 0.20, 0.18, 0.17 and 0.10; with average between cluster inbreeding of 0.11. A multi-modal distribution of inbreeding values within the first three clusters indicated the inclusion of daughters and granddaughters of the prominent bulls. Average F_{st} value across all cluster was 0.03. Indicating allele frequency differences between the clusters. The age of the prominent bull for each cluster was associated with its average inbreeding Table 1.

All clusters combined. Alternative procedures for identifying the top SNPs associated with selection were investigated. Grouping all target animals together increases the likelihood of identifying those SNPs with a similar and consistent allelic change across all clusters. The method of Rowan *et al.*, 2020, which looks for the SNPs that are highly associated with changes over time was investigated. Most of the top 100 SNPs identified by this method could be described as changing in a parallel way across all clusters. Interpretation of these allele frequency changes are consistent with a polygenic shift.

Separate clusters. Clustering similarly related descendants of a unique bull into individual groups provides a snapshot of the genotypic makeup of different families at a specific point in time. With alternative sources of genetic drift and linkage disequilibrium, each cluster may exhibit its own unique trajectory of AF changes. Three procedures were utilized to determine the most significant family specific SNPs. Two procedures involved comparing within-cluster AF changes and then looking for SNPs with the largest between cluster contrast. The 100 SNPs exhibiting the greatest range or the greatest variance between clusters were selected. The third method involved selecting the 100 SNPs with the highest $F_{st.}$

Table 1. Association of the average inbreeding value of the animals within the cluster with the age of the most prominent ancestor for that cluster.

Cluster description[1]	Most prominent ancestor[2]	Birthdate of bull	Average inbreeding of the cluster
A. Highest	Planet	March 2003	0.22
B.	Goldwyn	January 2000	0.20
C.	Shottle	July 1999	0.18
D.	O Man	March 1998	0.17
E. Lowest	Many	- - -	0.10

[1] Clusters are ranked by the average additive relationship of its members.

[2] Most prominent ancestor was identified as the bull with the largest numbers of descendants within a cluster.

The starting allele frequency for each cluster was similar with all animals tracing back to a common group of ancestors. The SNPs with the highest $F_{st.}$ would usually have one cluster exhibit a sharp change in its trajectory in the final generation. This is an indicator of genetic drift and a strong association between a cluster and a specific family. The genotype of the prominent sire causes a dramatic shift in the allele frequency.

Two procedures aimed at contrasting AF changes between clusters identified rapid directional changes, dramatic reversals in direction, and a mixture of parallel and non-parallel responses across clusters. Replicate Frequency Spectrum (Barghi *et al.* 2019) indicated that individual SNPs that showed a large directional change in AF for a specific cluster were replicated in the other clusters approximately 60% of the time. Heterogeneous changes in AF across clusters indicates genotypic redundancy whereby a variety of different genotype combinations provide different genetic solutions to the selection goals in a specific time-period.

Although we can only speculate on the reasons that may be involved in these non-parallel changes across clusters, two possible explanations would be differences in selection goals in different families and/or non-additive gene actions. The most prominent sires of families 1 and 2 are Planet and Goldwyn who are known to differ dramatically for fat and protein yield. The trajectory of the DGAT gene which has a significant genetic effect on fat and protein changes accordingly Figure 1.

Several known candidate genes identified by Ma *et al.* 2019 as having large AF in a very similar population were investigated. We also observed AF changes of a similar magnitude for the same identical alleles. However, their trajectories are not always the same for all families. For example, *ERBB4* (Chromosome 2) does show a shared and parallel change in AF across all families. Whereas the AF for *SPATA6* (chromosome 3) and *USP13* (chromosome 1) changed in a non-parallel way in the different families.

Population stratification caused by prominent sires allows for the investigation of alleles with strong LD (Barton, 2017). The fate of individual alleles is highly contingent on the allelic makeup at other loci. SNPs

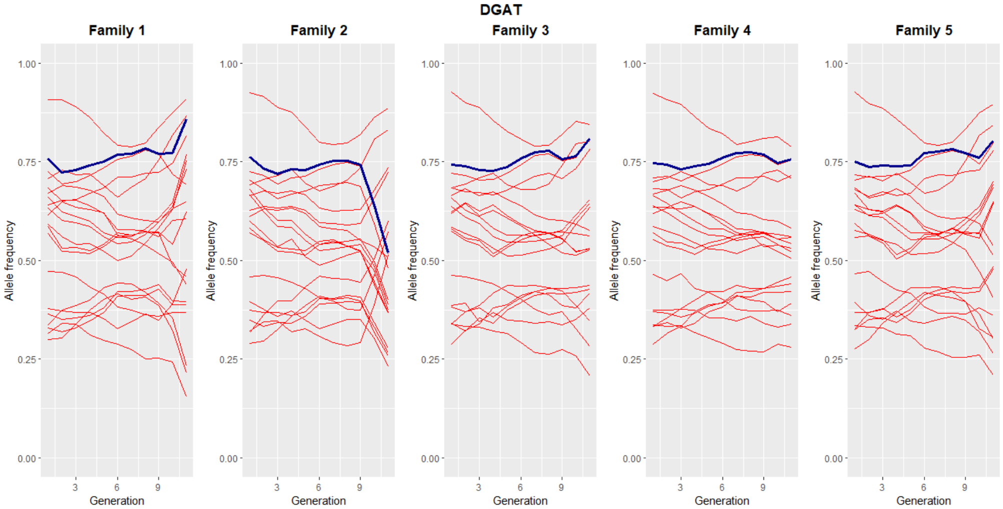

Figure 1. The allele frequency of the SNP that showed the *DGAT* gene (blue) and the surrounding 20 SNP markers (red) per generation within each family.

with an AF trajectory that includes a reversal in direction (or flip) are good candidates for the investigation of epistasis. Using the criteria of an overall reversal in direction of 0.1, we observed 11% or approximately 6,300 SNPs changing direction. In comparison, fixation of alleles was infrequent across the whole population (3 alleles). Number of alleles that became fixed within each family were 38, 22, 22, 59, and 40, respectively.

Conclusions

This study identified different subpopulations within the Holstein breed for a specific time-period. These clusters are associated with unique and identifiable families. AF changes between clusters involved rapid directional changes, dramatic reversals in direction, and a mixture of parallel and non-parallel responses. Heterogeneity in AF changes across families is an indicator of genotypic redundancy and a source of genetic variation.

References

Aguilar, I., and Misztal, I. (2008) J Dairy Sci 91(4):1669-1672. https://doi.org/10.3168/jds.2007-0575

Barghi, N, Tobler, R., Nolte, V., Jaksic, A.M., Mallard, F., *et al.* (2019) PLoS Biol. 17(2):e3000128. https://doi.org/10.1371/journal.pbio.3000128

Barton,N.H. (2017) Heredity (118):96–109. https://doi.org/10.1038/hdy.2016.109

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J. *et al.* (2010) Genetics 186(1):241-262 https://doi.org/10.1534/genetics.110.117275

Calo, L.L., McDowell, R.E., VanVleck, L.D., and Miller, P.D. (1973) J Anim Sci. 37(3):676-682. https://doi.org/10.2527/jas1973.373676x

Duenk, P., Bijma, P., Calus, M.P.L, Wientjes, Y.C.J., and Van der Werf, J.H.J (2020) G3 10(2):783-795 https://doi.org/10.1534/g3.119.400663

Ma, L., Wiggans, G.R., Wang, S., Sonstegard, T.S., Yang, J., *et al.* (2012) BMC Genom. 20(1):128 https://doi.org/10.1186/1471-2164-13-536

Ma, L., Sonstegard, T.S., Cole, J.B, Sonstegard, T.S., VanTassell, C.P., Wiggans, G.R., *et al.* (2019) BMC Genom. 20(1):128 https://doi.org/10.1186/s12864-019-5459-x

Misztal, I., Tsuruta, S., Lourenco, D.A.L., Aguilar, I., Legarra, A. *et al.* (2014) Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf