236. Investigating genetic redundancy as a source of genetic diversity and adaptability in the U.S. Holstein breed

Y. Steyn[1*], T.J. Lawlor[2], S. Tsuruta[1], Y. Masuda[1], A. Legarra[3], D.A.L. Lourenco[1] and I. Misztal[1]

[1]University of Georgia, 425 River Road, Athens, GA 30605, USA; [2]Holstein Association USA Inc., 1 Holstein PI, Brattleboro, VT 05301, USA; [3]INRA, UMR1388 GenPhySE, 31326 Castanet-Tolosam, France; yvette.steyn@uga.edu

Abstract

The continuous increase in inbreeding in the U.S. Holstein is a threat to genetic diversity. Loss of genetic diversity can hinder adaptation to changing environments and consumer demands. K-means clustering on the genomic relationship matrix was used to stratify 20,099 selected candidates into five clusters. These clusters were shown to be different through expected inbreeding within- and across cluster, the fixation index (F_{st}), and genetic correlations for stature between clusters. Expected inbreeding within cluster was mostly higher than across-cluster. The average F_{st} across cluster was 0.03. Genetic correlations ranged from <0.50 to >0.90. Families were formed by tracing pedigrees of each cluster back for 10 generations. Allele frequency changes over generations were evaluated. Polygenic shifts, selective sweeps, hitch-hiking, and epistasis were observed. The Replicate Frequency Spectrum was used to measure genetic redundancy. Results show that sub-populations within the breed have developed differently over time.

Introduction

The U.S. Holstein is generally viewed as a large, homogenous population known for its extensive use of only a few sires. This high level of relatedness among animals is a concern for genetic improvement and overall health of the population. Lack of variation can hinder the ability of populations to adapt to change (Markert *et al.*, 2010), which is a growing concern in the face of climate change and consumer preferences. While trait variation may reduce with selection, genetic variation can still exist. There are two main patterns in which allele frequency (AF) change occurs in a population, namely selective sweeps and polygenic shifts. Genetic hitchhiking can also increase the frequency of markers surrounding the gene of importance simply due to its proximity to the gene. Most traits of economic importance are highly polygenic. These alleles are more likely to undergo small polygenic shifts. Genetic redundancy is a phenomenon where an excess of beneficial variants exists to achieve the same genetic merit (Goldstein and Holsinger, 1992). Thus, different sub-populations may undergo heterogenous changes to converge to the same genotype or phenotype. The objectives of this study were to identify sub-populations within the breed and observe different AF changes over time.

Materials & methods

A total of 20,099 genotyped Holstein animals were used as selected candidates. This includes 3,902 males that were sires of animals born after 2010, with at least 25 progeny, and 16,197 females with measurements in 2013 or 2014. K-means clustering with a built-in R package was applied on the genomic relationship matrix (G) to separate the population into five separate clusters (C1 to C5). Alleles were coded as 0, 1, and 2.

Differences across cluster. 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 F_{st} across clusters was calculated as in Bonhomme *et al.* (2010). The genetic correlations across clusters were estimated using stature as trait and an adjusted version of 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

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

DOI: 10.3920/978-90-8686-940-4_236, © Y. Steyn *et al.* 2022

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. A simple correlation will underestimate the genetic correlation since the accuracy of breeding values is not considered. Therefore, the correlations were multiplied by an adjustment factor using the method by Calo *et al.* (1973). The heritability was considered as a measure of accuracy. Assuming a heritability of 0.45 in all clusters, this adjustment factor is 1/0.45.

Allele frequency changes. Within each cluster, five families were formed (F1 to F5) by tracing pedigrees back for 10 generations (G10 being most recent and G0 earliest). The AF were calculated for each generation within family. Generations overlap and families are connected by shared ancestors, thus animals may appear in more than one generation within a family, or/and in more than 1 family. In G0, 82% of the animals appear in more than one family and 54% are common to all families. The number of shared ancestors decreases over time. By G6, the number of families in common to all is larger than the number of unique animals per family. The total number of unique animals per family is 4,355 (F1), 3,555 (F2), 3,574 (F3). 7,107 (F4), and 4,870 (F5). Specific SNP were identified for visual observation based on previous literature (Ma *et al.*, 2019), the variance of absolute change from G0 to G10 across families, the range between minimum change in a family and maximum change, and the F_{st}. Genetic redundancy was measured using the Replicate Frequency Spectrum (RFS) (Barghi *et al.* 2019). This looks at the number of times the 100 SNP with the greatest change in one family, change by more than 0.30 in each individual family. This reflects non-parallel changes over time.

Results

The inbreeding within each cluster averaged 0.22 for C1, 0.20 for C2, 0.18 for C3, 0.10 for C4, and 0.17 for C5. The average across-cluster inbreeding was 0.11 for C1, 0.11 for C2, 0.12 for C3, 0.10 for C4, and 0.11 for C5. The average F_{st} for markers across clusters was 0.03. The genetic correlations adjusted for heritability are presented in Table 1. The changes in AF over time within each family for the SNP that showed the greatest variance is presented in Figure 1. Table 2 presents the RFS. When the 100 SNP with the greatest change in F1 are observed in each family, 15 SNP change more than 0.30 only in F1, none only in one other family, and 14 in all families. Among the 100 SNP with the greatest change in F2, 15 change more than 0.30 only in F2, none only in one other family, and 16 in all families. Of the 100 SNP with the greatest change in F3, 12 change by more than 0.30 only in F3, none only in one other family, and 13 in all families. When the 100 SNP with the greatest change in F4 are chosen, 3 change by 0.30 only in F4, 3 only in F3, 4 only in F2, and 19 in all. Among the 100 SNP with the greatest change in F5, 18 change by more than 0.30 only in F5, none only in one other family, and 11 in all families. Change was not restricted to be in the same direction in all families.

Table 1. Genetic correlations across clusters, measured as the adjusted correlation between the additive breeding value of animals of each cluster when expressed in the genetic background of each cluster.

Genetic background	Cluster expressed				
	C1	C2	C3	C4	C5
C1	1.00	0.63	0.99	0.76	0.93
C2	0.77	1.00	0.62	0.59	0.51
C3	0.81	0.41	1.00	0.81	0.92
C4	0.80	0.74	0.93	1.00	0.90
C5	0.89	0.45	0.89	0.73	1.00

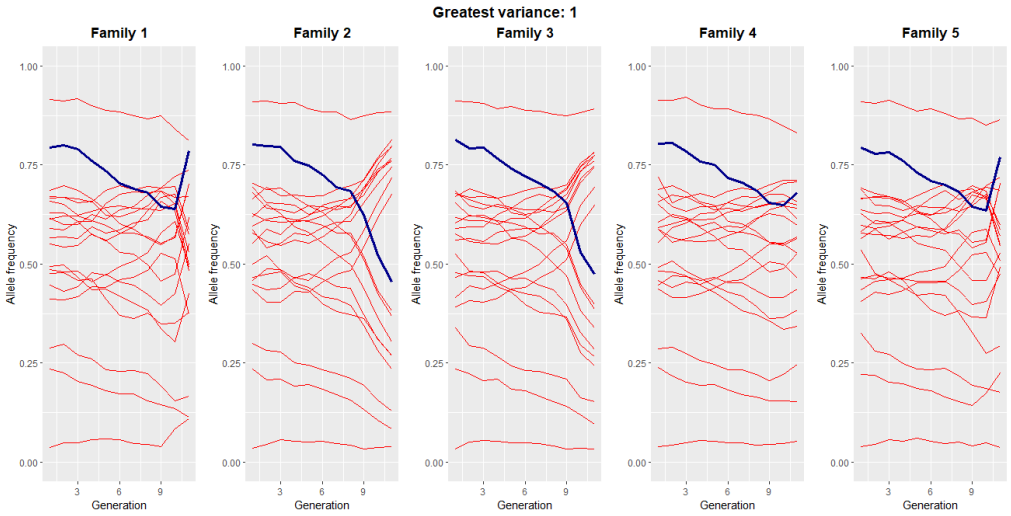

Figure 1. The allele frequency of the SNP that showed the greatest variance in change across families (blue) and the surrounding 20 SNP markers (red) per generation within each family.

Table 2. The Replicate Frequency Spectrum: The number of times the 100 SNP with the greatest change in a specific family (each row), change allele frequency (AF) by more than 0.30 in each family.

Family with top 100 SNP	Number of SNP with AF change >0.30				
	F1	F2	F3	F4	F4
F1	100	45	61	26	59
F2	57	100	68	25	51
F3	60	61	100	31	46
F4	68	67	78	55	60
F5	50	57	57	20	100

Discussion

The generally higher within-cluster inbreeding compared to across-cluster inbreeding, is an indication of genetic differences across clusters. Consistently lower inbreeding occurs for C4, regardless of whether mating was across- or within- cluster. This indicates that, although the clusters are separate from each other, C4 still contains enough variation to allow lower inbreeding. The average F_{st} of 0.03 is higher than what would be expected if no differences exist, and about half of that found between three different dairy breeds in a French study (Flori *et al.*, 2009).

There are two genetic correlations between families – one where the genotype of population 1 is expressed in the genetic background of population 2, and one where population 2 is expressed in the genetic background of population 1. These differ due to the AF in different populations. Although the genetic correlations between some clusters were above 0.85, correlations as low as 0.41 were observed, indicating that clusters show some genetic differences. In general, C2 has the lowest correlations with other clusters.

Some observed SNP showed non-parallel changes across families based on our criteria for selection, indicating genetic redundancy. Figure 1 shows a decrease in AF of the selected SNP (in blue) in F2 and F3

with no change in direction. Relatively small changes in AF are observed in F4, while directional changes in AF occur in F1 and F5. Observed SNP surrounding the selected markers (in red) show signs of potential partial selective sweeps and hitchhiking. Hitchhiking in opposite directions can be achieved when markers are in different phases. Directional changes could be due to a change in selection pressure, or epistasis with a different gene under selection.

The presence of genetic redundancy is further shown with the RFS. Even though many SNP change similarly in all families, there are family-specific SNP that change by more than 0.30 in only one family. While this study shows genetic redundancy, it does not determine whether specific changes were due to selection or drift. Gene-dropping computer simulations could be conducted to identify markers that have changed more than what could be expected under neutral inheritance. Such methods could be applied to identify genes under selection.

Conclusions

The U.S. Holstein breed is a complex mixture of family subgroups with different allelic frequencies and genotypic combinations. The study reveals genetic redundancy that has allowed sub-populations to change differently over time. Genetic redundancy has the ability to continue genetic improvement in the breed without drastically decreasing underlying genetic diversity. The different polygenic shifts in sub-groups avoids the fixation of alleles and the loss of genes that may allow the breed to adapt to climate change and consumer demand.

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

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

Goldstein, D.B., and Holsinger, K.E. (1992) Evolution 46(2):412-429 https://doi.org/10.1111/j.1558-5646.1992.tb02048.x

Flori, L., Fritz, S., Jaffrezic, F., Boussaha, M., Gut, I. *et al.* (2009) PloS One 4(8):e6595 https://doi.org/10.1371/journal.pone.0006595

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

Markert, J.A., Champlin, D.M., Gutjahr-Gobell, R., Grear, J.S., Kuhn, A. *et al.* (2010) BMC Evol. Biol. 10(1):205 https://doi.org/10.1186/1471-2148-10-205

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