Heritable variation in gene expression is the key to maximizing genetic gain and preserving genetic diversity with a properly designed breeding program.

***T. Lawlor*[1]**

[1] *Holstein Association USA, Inc., Brattleboro, VT, USA*

Corresponding author: tlawlor@holstein.com

Abstract

A hierarchical organization of molecular phenotypes provides a biological system of genes and pathways which can lead to different genotypes (redundancy) being selected in different subpopulations for the same phenotype. Heritable variation in transcription and translation is the key driver of genetic change. Redundancy in the regulatory code allows for genetic diversity amongst subpopulations. Gene expression is regulated by transcription factors (TF) ensuring that the right genes are active in the right tissue at the right time. A large amount of standing genetic variation is available from the many potential TF-TF interactions and TF interactions with other regulatory elements. Further diversity is possible in that each TF can target hundreds to thousands of different genes; and many of these genes through exon splicing, can produce functionally diverse transcripts and protein isoforms. These complex interactions form gene regulatory networks controlling specialized metabolic pathways. Which pathway is enriched is dependent upon the epistasis created by different founders, i.e., the ancestral makeup of the subpopulation. Selection on these epistatic effects leads to gametic disequilibrium between replicate populations causing them to differentiate. Different genetic architecture results in varied allele frequencies between subpopulations and intermediary allele frequencies in the global population. Genetic diversity within the Holstein breed can be preserved or increased with a proper population structure. This includes having multiple lines of Holsteins; utilizing multiple reference and target populations; genomic predictions for an overall global population and each separate subpopulation; avoidance of pooling SNPs; and analysis of transcriptomes for possible grouping of animals.

Key words: gene expression, redundancy, subpopulations, genetic variation

Introduction

Increasing rates of inbreeding is a concern amongst all the major Holstein breeding countries. Surveillance of undesirable monogenetic conditions along with genetic testing and selective purging of carriers has minimized the negative impact of an increase in homozygosity at undesirable individual loci. However, over the long-term, a reduction in genetic variation could be a bigger problem.

Rapid genetic change, population differentiation and maintenance of genetic variation has fascinated researchers for a long time. Charles Darwin reported that different species of finches lived on the different Galapagos islands. He speculated that population differentiation was a fundamental component of evolution. Sewall Wright spoke frequently and forcibly about the importance of population structure in maintaining genetic variation. In his 1950 paper on “The Genetic Structure of Populations”, he wrote “the subdivided population maintains more alleles at each locus and more at moderately high frequencies”. That is, selecting for different alleles and different genes in different subpopulations helps maintain genetic diversity across the entire population.

Those scientists would have been greatly aided in their understanding of the changes in the genetics of subpopulations by today's advancements in molecular biology. Much of the genetic variability observed in different polygenic traits originates from differences in gene expression. A meta-study in cattle estimated that 69% of the heritability of polygenic traits was due to variants associated with gene expression (Xiang et al., 2023).

Genes are regulated by transcription factors (TF) ensuring that the right genes are active in the right tissue at the right time. Given that there are thousands of TF, within a population a large amount of genetic variation is created during transcription from specific TF having the ability to interact with many other TF, TF interacting with other regulatory elements, and by each TF having the ability to target hundreds to thousands of different genes. Further variation is created during translation where many genes, through exon splicing, can produce functionally diverse transcripts and protein isoforms. The amount of standing genetic variation within a population can be very large and there can be many different combinations of the genetic variants controlling gene expression that can lead to the same phenotypic change. The phenomenon of multiple genetic solutions leading to the same phenotypic change is known as genetic redundancy (Barghi et al. 2019). That is, there are more variants segregating in the whole population than are needed to achieve a specific phenotypic change.

Genomic testing of populations that have been divided and selected for the same phenotypic goal frequently show non-parallel changes in allele frequencies (Barghi et al. 2019) along with different transcriptomic changes, different gene networks being formed, and different biological pathways being emphasizes (Lai et al., 2023). While the subpopulations differed in which genetic variants were favored or disfavored, the resulting change in metabolites were similar leading to the same overall phenotypic change. This hierarchical organization of molecular phenotypes provides a biological system of genes and pathways which can lead to different genotypes (redundancy) being selected in different subpopulations for the same phenotype.

The biological system of different genes and pathways are known as a gene regulatory network (GRN). Often described as being modular, different GRNs can be used in a similar way. Having redundancy of different GRNs has several benefits. For an individual, the most obvious benefit is that one GRN can compensate for mistakes in other pathways. For a population, different genetic changes in a GRN can lead to an improved function or an evolutionary change.

The ancestral makeup, i.e., the original founders, of the different subpopulations is important for multiple reasons. Given the large number of possible combinations of genetic variants involved in gene expression, different subsets of founders will possess different genotypes, different genes will be enriched, leading to different GRNs and pathways. Another important component is epistasis. Whereby a certain gene has a positive effect in one subpopulation and the opposite effect in another. The value of an epistatic gene differs across subpopulations because its value depends upon what other genes are in that subpopulation. With epistasis, different subpopulations are selecting for different gene combinations.

Integrating new molecular biology information along with quantitative genetic theory provides us with a more accurate prediction of how divided populations change over time. With selection, subpopulations should diverge and become more differentiated over time as different genotypes, i.e., different gene-gene interactions, are favored in different subpopulations. This process is known as gametic disequilibrium (Tomoko, 1982). Rapid changes in genetic architecture in divided populations, when selecting for the same phenotypic goal has been observed in both

plants and animals. The variants changing the most tend to be associated with gene expression. Comparison between subpopulations indicate changes occurring in both shared and unique pathways. In dairy cattle, a decline in predictability of future performance is observed as the time between the animals in the reference and target populations increases.

Materials and Methods

The population structure of U.S. Holsteins was investigated for two different time periods, 2014 and 2022, using the CDCB's National Cooperator Database. The 2014 data set has been discussed by Steyn et al 2023. K-means clustering on the genomic relationships of animals born between 2010 and 2014 identified five subpopulations. Four of the clusters were composed primarily of the descendants of four prominent sires that had been used extensively during that time period. The genetic contribution of the prominent sire for each cluster; Planet, Goldwyn, Shottle and O Man were 28.1%, 18.8%, 19.8% and 21.6%, respectively. The fifth and largest cluster was composed primarily of the offspring of many different sires, with no individual bull having a genetic contribution exceeding 4.3%. Trajectory of allele frequency change for 58,990 SNP markers was calculated across 10 generations for each of the subpopulation.

The combination of genomic selection with sexed semen, advanced reproductive technology and restricted access to young genetics has led individual breeding organizations and countries to genetically diverge from one other. By 2022, population structure was no longer determined by the heavy usage of individual bulls but more by the breeding program of large organizations. Young bulls with a minimum TPI value of 3000 were selected from the December 2021 official genomic evaluations of CDCB. Almost all the 713 young bulls were sired by a bull controlled by the same breeding organization. Four breeding companies controlled 91% of these bulls. Between 83% and 94% of the mothers of these top young bulls were also controlled by the same breeding organization that had control over the sire and his sons. Each of these breeding organizations has created their own subpopulation. To measure genetic differentiation Wright's Fixation Index (F_{st}) was calculated as follows:

$$F_{ST} = (F_{IT} - F_{IS}) / (1- F_{IT})$$

where F_{IT} is the inbreeding within total population and F_{ST} is the inbreeding within subpopulation.

An important component of genomic architecture is the size of the SNP effects or substitution effects which includes the effects due to additive and dominance gene action, inter-locus interactions.

$$\alpha_i = a_i + (1 - 2p_i)d_i + \sum \alpha_{ij}^i$$

Our primary interest in this paper was a researcher's ability to identify inter-locus interactions when data from all subpopulations are pooled together.

Results & Discussion

New information from molecular biology provides valuable insight on the vast amount of redundancy available with respect to alternative genetic solutions to achieve a common phenotypic change. Extensive use of individual bulls or control of access to top genetics are two ways that subpopulations have been created within the U.S. Holstein population. Existence of subpopulations is beneficial in that it helps preserve genetic diversity.

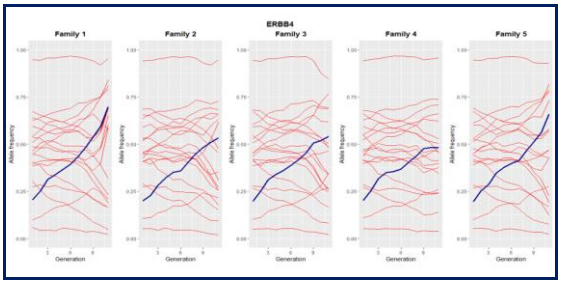
Steyn et al. 2023 reported that different sets of SNPs were changing over time in different subpopulations of U.S. Holsteins. Heterogeneity in SNP frequency changes across subpopulations indicates that different SNPs are being targeted in different subpopulations. While as many as 59 SNPs went to fixation in one of the subpopulations,

fixation of SNPs was infrequent across the whole population (3 alleles).

Much of the genetic variability observed in different polygenic traits originates from differences in gene expression. These heterogenous genomic changes in different subpopulations lead to differences in transcriptomic response, development of distinct GRNs and enrichment of different biological pathways. So why are gene interactions so frequently ignored by quantitative geneticists?

The pooling of all data together into a single national or global evaluation causes us to miss important gene-gene interactions that are important for maintaining genetic variation. Figure 1, adapted from Steyn et al, 2023, provides an illustration of this point. By pooling all data together, genetic interactions which have a heterogeneous effect in different subpopulations are ignored.

A. SNP marker identifying the gene ERBB4



B. SNP marker with no known function

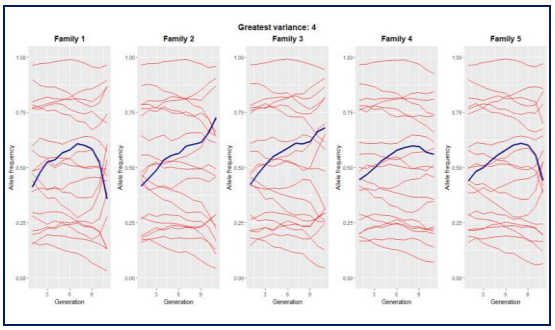


Figure 1. SNP A has a consistent effect in all families resulting it a high SNP effect. SNP B has an inconsistent effect across families, its SNP effect would be low due to being averaged across all families.

This means that our current genomic selection programs are selecting for certain type of gene actions and ignoring other genetic options. For a SNP to have a consistent or large effect across all subpopulations the SNP's action must be direct and largely independent of other genes, e.g., a protein coding gene or using terminology from the omnigenic model a "core" variant involved in gene expression (Mathieson, 2021). SNPs with an inconsistent effect across subpopulations are referred to as "peripheral" genes, affecting the phenotype through a network of interactions with other peripheral genes and core genes.

The important message for our Interbull community is that pooling data sets together for the sake of obtaining higher accuracy of prediction does so by focusing on a limited type of genes, i.e., core genes while ignoring the more numerous peripheral genes. The solution would be to recognized multiple subpopulations within the Holstein breed with separate reference populations and separate genomic evaluations. This allows for the selection of more peripheral gene action, enriching different GRNs in different subpopulations, and preserving genetic diversity across the entire breed.

Combining genomic selection with sexed semen, embryo transfer and restricted access to young genetics has led individual breeding organizations and countries to genetically diverge from one other. Figure 2 presents a measure of genetic differentiation (Fst) or population structure of the U.S. Holstein population in 2022. Each of the different breeding organizations are focusing on a slightly different group of animals. Current genetic differences between breeding organizations approach one quarter to one half of the genetic differences found between dairy breeds.

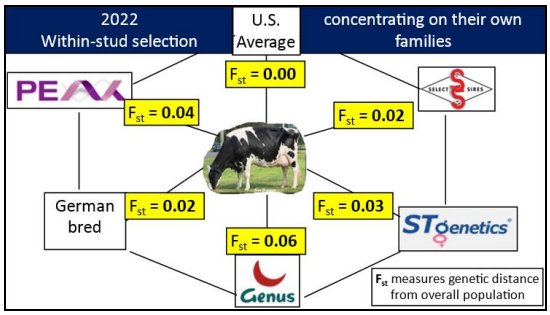


Figure 2. Wright's Fixation Index of major breeding organizations in U.S. Holsteins in 2022.

Within-stud selection has led to the assembly of breeding units made up of slightly different families. This is the start of our breed having different lines of Holsteins to choose from. AI breeding companies can further expand this concept by having multiple lines available within each organization. Farmers could then rotate between lines and continue to make rapid genetic gain while maintaining low inbreeding within their own herd and high genetic diversity across our breed.

Having multiple lines of Holstein does not mean that we all go off in different directions. Quite the contrary. It means that we must use our genetic resources more wisely. Breeding organizations will need to be committed to the program. National genetic evaluation centers will need to provide multiple genetic evaluations, which includes a national overall ranking as well as separate evaluations, with its own genomic reference population, for each domestic line. Our international organizations, such as Interbull, will need to develop genetic tools that routinely monitor the genetic distances between lines and the overall change in inbreeding in our global population. And we will all need to be heavily involved in the educational process of the benefits of this new breeding design and how to properly use multiple lines within a herd.

Conclusions

In our current genetic evaluations, all animals are pooled together causing the unique gene-gene interactions from the different subpopulations to cancel one another out. Rather than selecting for unique and/or epistatic combinations of genes, we select for those genes that have a similar or additive effect across the breed. The highest genetic merit animals are those with the highest total of high average effect SNPs. Genetic diversity within the Holstein breed can be preserved or increased with a proper population structure. This includes having multiple lines of Holsteins; utilizing multiple reference and target populations; and providing genomic predictions for each separate subpopulation as well as the overall global population.

Acknowledgments

The author thanks CDCB staff and industry cooperators for data.

References

Barghi N, Tobler R, Nolte V, Jakšić AM, Mallard F, Otte KA, Dolezal M, Taus T, Kofler R, Schlötterer C. 2019. Genetic redundancy fuels polygenic adaptation in Drosophila. *PLoS Biol. 2019 Feb 4;17(2). https://doi.org/10.1371/journal.pbio.3000128*

Lai WY, Otte KA, Schlötterer C. 2023, Evolution of metabolome and transcriptome supports a hierarchical organization of adaptive traits. *Genome Biol Evol. 2023;15 https://doi.org/10.1093/gbe/evad098*

Mathieson I. 2021, The omnigenic model and polygenic prediction of complex traits, *Am. J. Hum.Genet,108(9):1558-1563. https://doi.org/10.1016/j.ajhg.2021.07.003*

Tomoko O. 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations**.** *Proc. Natl Acad. Sci. USA 79:1940-1944.* https://doi.org/10.1073/pnas.79.6.1940

Steyn, Y. et al., 2023. Nonparallel genome changes within subpopulations over time contributed to genetic diversity within the

US Holstein population. *J Dairy Sci.106(4):2551 – 2572. https://doi.org/10.3168/jds.2022-21914*

Xiang et al., 2023. Gene expression and RNA splicing explain large proportions of the heritability for complex traits in cattle, *Cell Genomics* 3(10). https://doi.org/10.1016/j.xgen.2023.100385

Wright S. 1949. The Genetical Structure of Populations. *Annals of Eugenics.15:*323-354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x