Haplotypes and runs of homozygosity associated with high lifetime milk production among elite Holstein cows

R. D. Shanks[1]

[1] *Holstein Association USA, Inc., 1 Holstein Place, P.O. Box 808, 05302-0808, Brattleboro, Vermont, United States of America*
rshanks@holstein.com (Corresponding Author)

Summary

A profitable dairy cow is a goal in the dairy industry. Fortunately, several strategies can lead to a profitable cow (high lactation production, low costs, elite lifetime production, outstanding unique genetics, great efficiency of production, or disease resistance). The focus of this project was to identify unique genetics of Holstein cows with elite lifetime production (greater than 68,039kg of lifetime milk production and classified very good or excellent). The final sample comprised 814 elite cows and 803 control cows with genomes tested on 50K chips. A haplotype trend regression was conducted on the differences in the two groups and five sets of haplotypes were identified, one set each on chromosomes 6, 13, 23, 24 and 26. Each set of haplotypes contained from three to five overlapping 5-SNP haplotypes. In addition, clusters of runs of homozygosity (ROH) were evaluated and one cluster of approximately 28 SNP from chromosome 22 was found among 83 elite cows but no control cows. The first marker of this unique cluster was ARS-BFGL-NGS-39645. Evaluation of the pedigrees of the elite and control cows revealed more sires and maternal grandsires represented in the pedigrees of the elite than control cows. This pedigree diversity continued among the 83 elite cows, with the unique cluster of ROH, because they were daughters of 61 different sires. Breeding a more profitable cow may be a single goal, but the paths to success are varied, partly as a result of the diversity among elite cows.

Keywords: lifetime milk production, haplotypes, runs of homozygosity, dairy cattle, Holstein

Introduction

This research project was prompted by a letter from a dairy producer, Robert Miller of Mil-r-Mor Farm, to CEO John Meyer of Holstein Association USA. The essence of the letter was a request to use genomic information to identify the value in long-lived, high-producing dairy cows. The premise of the request was the belief that older, high-producing dairy cows had value because they had avoided major diseases. When the letter was written in early 2012, fewer than 10,000 Holstein cows had been genome evaluated with a 50K or greater chip. One objective of the research project was to encourage adoption of genome technology by dairy producers. Another objective of the project was to sample a unique population of older cows. By September 2017, over 38,000 Holstein cows and over 36,000 Holstein bulls had been genome evaluated with a 40K or greater chip. Including all ages, chip densities and imputations, over 1,708,000 Holsteins have been genotyped. (CDCB website). To state the obvious, genome technology has been adopted by the dairy industry to assist with genetic evaluations.

A profitable dairy cow is a goal in the dairy industry. Fortunately, several strategies

can lead to a profitable cow (high lactation production, low costs, elite lifetime production, outstanding unique genetics, or great efficiency of production). Disease resistance or the problem-free cow contributes to success in several of these strategies. High lifetime milk production is the ultimate indicator of disease resistance. Heathy cows are expected to be more productive in their lifetimes than unhealthy cows. The focus of this project was to identify unique genetics of Holstein cows with elite lifetime production.

The research emphasis was to focus on older cows that have been productive. How are these high-producing older cows unique? How are their genetics different from a sample of highly selected contemporaries? The elite cow project was initiated with the research objective to find haplotypes of elite Holstein females that are associated with their valuable and unique phenotypes.

Material and methods

Elite and control cows

Cows with elite lifetime milk production were defined as those cows that had produced more than 68,039 kg (150,000 lb) of milk production during their lifetime. A secondary requirement was that each cow be classified very good (85 to 89 points) or excellent (90 to 97 points). The classification criteria for elite cows reflected the common practice of requiring bull dams to be classified as very good or excellent. More details on linear classification can be found on website of Holstein Association USA. (HAUSA website). Because samples from these cows needed to be collected for DNA analysis, the cow also needed to be alive during the summer or fall of 2013. Elite cows were born between 1998 and 2007. Control females were highly selected and elite in their own right because they had been genome tested through the Holstein Association USA prior to fall 2013. The owners of these cows could be described as early-adopters, progressive and anticipating great genetics among their cows. Control females were born in the decade prior to 2008 and therefore had the opportunity to achieve elite status but had not because either they had not produced 68,039 kg of lifetime milk, or they were not classified very good or excellent. After edits, sample size was 814 elite cows and 803 control cows that were born before 2007. The two populations of cows were compared for differences in lifetime milk production, classification scores and genome.

Elite cows had high lifetime milk production and very good or better classifications. If lifetime milk production and classification scores were plotted, elite cows would be found in one quadrant. Control cows would be found in three quadrants with the majority also being classified very good or excellent. Phenotypic means and standard deviations of final classification score and lifetime milk production for elite and control cows are in Table 1. All elite and most control cows were classified very good or excellent. The major phenotypic difference between the elite and control cows was greater lifetime milk production of the elite cows. Lifetime milk per day in milk was numerically larger for elite than control cows but means were within a single standard deviation. To interpret the means in Table 1, elite cows gave a little more milk per day and did it for many more days. Elite cows survived longer to have greater lifetime milk production than the control cows.

Table 1. Phenotypic means and standard deviations for elite and control cows.

	Final classification score	Lifetime milk (kg)	Lifetime milk per day in milk (kg/day)
Elite (814 cows)	88.4 ± 2.7	89,060 ± 12,231	42.1 ± 5.2
Control (803 cows)	87.9 ± 3.4	41,516 ± 15,675	37.7 ± 5.8

In addition to phenotypic differences, the pedigrees of the elite and control cows were compared for differences in number of sires and maternal grandsires as well as the overlap of sires and maternal grandsires found in pedigrees of both groups.

Genome information

Cows were sampled with Illumina BovineSNP50 v1 BeadChip or BovineSNP50 v2 BeadChip. The 50k v1 chip had 54,001 SNP analyzed for 152 elite cows and 683 control cows. The 50k v2 chip had 54,609 SNP analyzed for 662 elite cows and 120 control cows. In total, 835 cows were analyzed with the 50k v1 chip and 782 cows were analyzed with the 50k v2 chip. Between the two chips, 60,412 SNP were available for analyses. Two alleles were found for 55,309 SNP and 5,103 SNP had only one allele in the data.

Haplotype blocks

A logistic haplotype trend regression was completed as a first approach to detect differences between populations of elite and control cows using SNP & Variation Suite v8.4 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). This approach allowed inclusion of covariates to adjust for differences in year of birth and chip of analysis. Haplotype blocks were defined as a moving window with a fixed width of five markers. The estimation method was Efficient Mixed-Modal Association eXpedited (EMMAX) (Scherer, 2014). A maximum of 50 EM iterations were allowed and the EM convergence tolerance was .001. A total of 49,882 haplotype blocks were analyzed. Regression was successful for 45,816 blocks. Sixteen blocks had too few haplotypes, and 4,050 blocks had failed regressions. The process was successful for 91.8% of the blocks. Year of birth and chip of analysis were covariates in the regression to partially account for genetic trend and call rate tendencies by chip. For the haplotype trend regression, a conservative estimate of significance (-Log10 P > 7.0) was used. Significant blocks were screened to have the cell count of the first marker of the block be five or greater. The intent was to identify major differences rather than all possible differences. For confirmation of detected blocks, a genome wide association study with an additive model and one with a mixed model using genome relationship matrix were also conducted. Details of those analyses will not be presented here.

Runs of homozygosity (ROH)

Rather than limit the analysis to just haplotype differences, runs of homozygosity were also explored to detect differences between the elite and control cows. If one or a few sires were responsible for differences between elite and control cows, several runs of homozygosity were expected. Runs of homozygosity were detected and clusters identified following process of Zhang et al. 2013 as programmed in SNP & Variation Suite v8.4 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). Runs of homozygosity were defined as 20 or more homozygous SNP. One heterozygote and five missing SNP were allowed in the screening for runs of homozygosity. A cluster of run of homozygosity was defined for runs of

homozygosity found in 20 or more cows. Clusters of runs of homozygosity were evaluated for differences between elite and control cows. Analysis of clusters was to focus on the most prevalent runs of homozygosity.

Results and discussion

In addition to phenotypic differences, the pedigrees of the elite and control cows were compared. As expected, the pedigrees of the elite and control cows had many of the same sires and maternal grandsires. Table 2 contains more information on pedigree diversity for elite and control cows.

Table 2. Pedigree diversity of sires or maternal grand sires for elite and control cows.

	Elite cows	Control cows	Both	Total
Total number of cows	814	803		1617
Total number of sires	270	144	71	343
Only sire of elite cow	199			
Only sire of control cow		73		
Total maternal grand sires	332	204	100	436
Only maternal grand sires of elite cow	232			
Only maternal grandsires of control cow		104		

Although the number of control and elite cows were almost the same, many more sires had daughters among the elite than control cows. The same trend was true for maternal grand sires, many more different maternal grand sires were found in the pedigrees of elite than control cows. The 71 sires in pedigrees of both groups represented twenty-six percent of the sires of elite cows and 49% of the sires of control cows. The 100 maternal grand sires in pedigrees of both groups represented thirty percent of the maternal grand sires of elite cows and 49% of the maternal grand sires of the control cows. A major observation from evaluating pedigrees was that elite cows had more pedigree diversity than control cows. Maintaining a broad genetic base is a constant challenge in light of the tremendous genetic progress occurring in the Holstein breed. More offspring of elite cows may help to increase genetic diversity.

Haplotypes

Haplotype trend regression found 33 haplotypes with -Log10 P > 7.0 and 27 of them had the cell count of the first marker greater than five. Therefore six haplotypes were eliminated due to low allele frequency. The 27 haplotypes were found in ten different locations. The most significant of these were grouped together on five different chromosomes with four of them on chromosome 6, five on chromosome 13, three on chromosome 23, four on chromosome 24, and four on chromosome 26. These five areas of significance were also identified by genome wide association study with an additive model and mixed model analyses with genome relationship matrix. To repeat, each of these five areas identified by 20 different first markers had a significantly different allele frequency between elite and control cows. Location of these haplotypes, possible candidate gene and function are listed in Table 3. Because the main difference between the elite and control cows was lifetime milk production, gene functions related to protein coding and calcium-dependent adhesion were plausible. These functions may form a foundation and be the basis for a problem-free cow that has the potential for high lifetime milk production. More detail on the genes in the

vicinity of the key haplotypes is worthy of further pursuit in an attempt to better understand the biology and then have an opportunity to be more successful in a selection program to capture the biological advantage. To state in more general terms, success in disease resistance may be built on the cow having the capacity to do the normal things well. Based solely on the gene functions of the five key haplotypes, redundancy may be an important tool to facilitate disease resistance and survival.

Table 3. Candidate genes for five key haplotypes associated with lifetime milk production.

BTA	Position (Mb)	Candidate Gene	Function
6	109.4 to 109.7	F1MQ42_BOVIN	protein coding
13	59.6 to 59.8	ENSBTAG00000037492	miRNA
23	30.9 to 31.0	ZN184_BOVIN	protein coding
24	10.7 to 11.1	ENSBTAG00000043797	rRNA
		CDH7 (cadherin 7)	calcium-dependent adhesion
26	5.3 to 5.4	PCDH15 (proto cadherin 15)	calcium-dependent adhesion

Runs of homozygosity

Among 1617 cows, 55,557 runs of homozygosity were detected. Runs ranged from a minimum of 20 homozygous SNP for two cows to another cow with a run of 186 homozygous SNP. Of the 55,557 runs, 19,946 ROH had no heterozygotes and the other 35,611 ROH had only one heterozygote within the run. Although up to five SNP were allowed to be missing, the majority of the runs had two or less SNP missing. One ROH comprising 159 SNP occurred 23 times. Another long ROH of 116 SNP occurred 31 times. Many ROH existed in the populations. Rather than interpreting each, clusters were identified of the most prevalent ROH.

From the 55,557 ROH detected, 410 clusters of runs of homozygosity were identified. Clusters of a ROH occurred in 20 or more cows with run length greater than 500Kb comprising a minimum of 25 homozygous SNP and not allowing gaps between SNP greater than 100Kb. After adjusting for chip and year of birth, two significant differences were detected in clusters of ROH between elite and control cows. These clusters were found in 94 cows. The significant cluster from chromosome 17 was found in 11 cows (six control and five elite). The other cluster from chromosome 22 was found in 83 cows. The uniqueness of the cluster from chromosome 22 was that all 83 cows were elite cows. None were control cows. Almost all of the ROH of the cluster from chromosome 22 were 28 consecutive homozygous SNP. These 83 elite cows represented a little more than 10% of the 814 elite cows of the analysis and were daughters of 61 different sires. Twelve sires had two or more elite cows among the 83 and Durham and Blitz each sired six of the 83 cows. One interpretation of this information is that elite cows can become elite based on having diverse genetics. The path to elite cows is not singular.

The 83 elite cows with the unique ROH were born between 2000 and 2006, classified between 85 and 94, with lifetime milk production of 70,302 kg to 129,310 kg. Their lifetime milk production per day in milk ranged from 33 to 54 kg/day. All ROH for these 83 cows started at position 41,417,926 and almost all ended at position 42,616,545. The 28 homozygous SNP represented length of 1,198,619bp.

Wiggans et al (2016) described single genes associated with lactation production but those genes were not found among either the five haplotypes or the unique run of

homozygosity associated with lifetime milk production.

Conclusions

Initially with this project, dairy producers were requesting what makes elite cows different. The hope was that elite cows would have many alleles in common and that selection could be practiced to increase the frequency of those elite alleles in the population. When the pedigrees of elite and control cows were evaluated, the opposite was found. Elite cows had more pedigree diversity than the control cows.

Five haplotypes were identified that had significantly different frequencies between elite and control cows. These haplotypes were not only identified by haplotype trend regression, but also through a genome wide association study with an additive model and a mixed model using genome relationship matrix.

In addition to the five haplotypes that were identified, two clusters of runs of homozygosity were significantly different between elite and control cows with one ROH being of particular interest because it was found only in the data from 83 elite cows sired by 61 different bulls. Breeding a more profitable cow may be a single goal, but the paths to success are varied, partly as a result of the diversity among elite cows.

Acknowledgement

I thank Robert Miller for believing that long-lived, high-producing, disease resistance cows have value.

List of References

CDCB website, www.uscdcb.com/Genotype/cur_density.html, accessed September 25, 2017.

HAUSA website, www.holsteinusa.com/holstein_breed/holstein101.html, accessed January 18, 2018.

Scherer, A., 2014. GWAS. Golden Helix, Inc, Bozeman, MT. 27 pages.SNP & Variation Suite ™ (Version 8.4) [Software]. Bozeman, MT: Golden Helix, Inc. Available from http://www.goldenhelix.com.

Wiggans, G.R., T.A. Cooper, P.M. VanRaden, C.P. Van Tassell, D.M. Bickhart, & T.S. Sonstegard, 2016. Increasing the number of single nucleotide polymorphisms used in genomic evaluation of dairy cattle. Journal of Dairy Science, Volume 99: 4504 – 4511.

Zhang L, M.S. Orloff, S. Reber, S. Li, Y. Zhao, C. Eng, 2013. cgaTOH: Extended Approach for Identifying Tracts of Homozygosity. PLoS ONE 8(3): e57772.