

# Forty Years of Artificial Selection In U.S. Holstein Cattle Had Genome-wide Signatures

T.S. Sonstegard<sup>1</sup>, L. Ma<sup>2</sup>, C.P. Van Tassell<sup>1,3</sup>, E-S. Kim<sup>1</sup>, J.B. Cole<sup>3</sup>, G.R. Wiggins<sup>3</sup>, B.A. Crooker<sup>2</sup>, B.D. Mariani<sup>4</sup>, L.K. Matukumalli<sup>1</sup>, J.R. Garbe<sup>2</sup>, S.C. Fahrenkrug<sup>2</sup>, G. Liu<sup>1</sup>, Y. Da<sup>2</sup>



<sup>1</sup>Bovine Functional Genomics Laboratory, 10300 Baltimore Avenue, Beltsville, MD 20705-2350, USA  
<sup>2</sup>Department of Animal Science, University of Minnesota, Saint Paul, MN 55108, USA  
<sup>3</sup>Animal Improvement Programs Laboratory, 10300 Baltimore Avenue, Beltsville, MD 20705-2350, USA  
<sup>4</sup>Genetics and In Vitro Fertilization Institute, 3015 Williams Drive, Fairfax, VA 22031, USA



Leipzig, Germany  
August 1-6, 2010

9<sup>th</sup> World Congress on Genetics Applied to Livestock Production

**ABSTRACT** Genome signatures of 40 years of artificial selection in U.S. Holstein cattle were identified by comparing allele frequencies, linkage disequilibrium (LD) patterns and extended haplotype homozygosity (EHH) in three Holstein populations representing three selection periods, a contemporary group subjected to 40 years of selection, a group subjected to 20 years of selection since 1964 (born in 1975-1985), and a group unselected since 1964. Results showed that 40 years of artificial selection caused genome-wide changes as characterized by allele frequency differences between selected and unselected groups, by elevated LD signals and by elevated EHH distances in the two selected groups. Many of the selection signatures were associated with SNP effects on 31 dairy traits of contemporary U.S. Holstein cattle or with phenotypic changes over the past 40 years.

## Introduction

Genetic selection of U.S. Holstein cattle since 1964 has resulted in tremendous phenotypic changes. By 2004, average annual yields of milk, protein and fat per cow increased 3740 kg, 167 kg, and 193 kg, respectively, but dairy fertility measured by daughter pregnancy rate (DPR) experienced serious declines (Figure 1). During the same period, selection for body appearance traits achieved many changes towards show-animal standards and better functionality of the cow's anatomy. In this study, we report that the forty years of artificial selection had genome-wide selection signatures associated with the phenotypic changes.

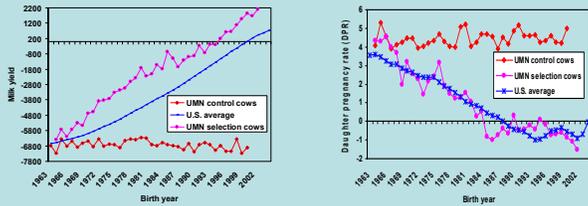


Figure 1: Genetic merit of milk yield (left) increased but DPR (right) decreased steadily for the U.S. and University of Minnesota (UMN) contemporary Holstein cows, while milk yield and DPR were unchanged for the UMN control cows unselected since 1964.

## Material and methods

A random population representing three different selection stages was analyzed: Group I remained unselected since 1964 (no selection for the past 40 years, n = 301), Group II subjected to 20 years of selection between 1964-85 (n = 215, bulls born 1975-1985), and Group III of contemporary cow subjected to 40 years of selection (n = 1656). A total of 45,878 SNP markers from the Illumina Bovine SNP50™ chip that were polymorphic or had allele frequency differences (AFD) of 2% or greater between Groups I and III were genotyped. Selection signature was identified by AFD between selected and unselected groups, by linkage disequilibrium (LD) patterns and by extended haplotype homozygosity (EHH) analysis (Sabeti et al., 2007).

## Results and discussion

Large AFD between different selection stages were observed across all 29 bovine autosomes and the X chromosome, indicating that 40 years of artificial selection caused genome-wide changes. Groups I and III representing approximately 40 years of selection had the largest AFD (red dots in Figure 2a). The first 20 years (II vs. I, blue dots in Figure 2a) and the second 20 years (III vs. II, yellow dots in Figure 2a) of selection also had large AFD, but the first 20 years had considerably more and larger AFD than the second 20 years.

Random AFD were estimated as AFD between a random sample of 250 cows and the remaining 1406 cows of Group III (black dots in Figure 2a). The largest random AFD was 0.103, indicating that AFD = 0.10 could be used as a threshold value for declaring AFD due to selection. However, the comparison of founders and descendants of the UMN control line showed that 1000 markers had AFD of 0.09-0.15 due to genetic drift. Using AFD = 0.15 as a conservative threshold, 22.58% of the Holstein genome would have been affected by 40 years of artificial selection.

The contemporary population (Group III) with 40 years of selection had the largest EHH distances and LD while the unselected population (Group I) had the smallest EHH and LD (Figure 2b, 2c). Many SNP effects were near core regions with elevated EHH in selection populations (Groups II and III, Figure 2d).

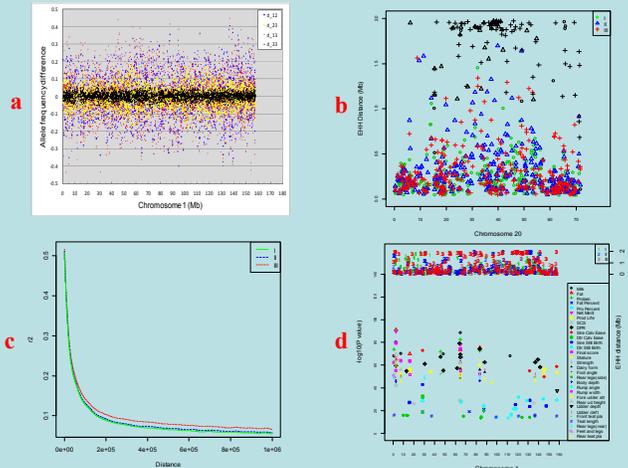


Figure 2: Examples of AFD (a), EHH distances for all core regions (b), average LD of autosomal markers for various distances (in Kb) (c), and QTL effects of SNP markers in Group III along with maximum EHH values > 1.5Mb in at least one group (d). Nodes in black in (b) were minimal maximum EHH distances recorded from the core regions.

The approximate time period of selection was identified by comparing EHH and AFD in the three groups, i.e., whether selection occurred in the first 20 years, second 20 years, the entire 40 years, more than 40 years, or 40 years ago without further selection in the recent 40 years (Table 1). Many QTL effects were placed in or next to core regions of Group II or III with selection signature characterized by long-range EHH and AFD = 0.10, based on our previous reports of QTL maps for 31 dairy traits in U.S. contemporary Holstein population (Cole et al. 2010; Wiggins et al., 2010) and based on our current QTL analysis for the combined population of UMN control cows and the contemporary cows. QTL effects in the combined population (trait names with subscript '2' in table 1) were associated with phenotypic changes due to 40 years of selection, and some of these effects were also highly significant in the contemporary population (effects indicated by trait names without subscripts '1' and '2' in table 1). Selection signatures associated with QTL effects in the combined population were stronger than those associated with QTL effects in the contemporary population only (data not shown). Two selection signature with EHH and LD evidence are shown in Figure 3 and Figure 4.

Table 1: Examples of chromosome regions with selection signature.

Chr	Region (UMD 3.0)	Gene	Selection period <sup>a</sup>	Trait <sup>b</sup>
1	64864266-65025357	<i>KTELC1, CD80, PLA1A, COX17</i>	40 yrs	MY <sub>2</sub> , FY <sub>2</sub> , PY <sub>2</sub> , NM <sub>2</sub> , DPR <sub>2</sub> , FS <sub>2</sub> , DF <sub>2</sub> , BD <sub>2</sub> , RW <sub>2</sub> , UC <sub>2</sub> , FTP <sub>2</sub> , RTP <sub>2</sub>
1	135444745-135491887	<i>EPHBI</i>	1 <sup>st</sup> 20 yrs	SCS, DPR <sub>1</sub>
2	349184373-35043749	<i>TBR1</i>	40 yrs	#1 for RUH <sub>1</sub> ; STA <sub>1</sub> , DF <sub>1</sub> , BD <sub>1</sub> , RW <sub>1</sub> , FUA <sub>1</sub> , UC <sub>1</sub> , RLR <sub>1</sub> , RTP <sub>1</sub>
7	47984663-48218360	<i>CATSPER3</i>	40+ yrs	STR <sub>2</sub> , SCE <sub>2</sub>
10	53560658-53790650	<i>ZNF280D</i>	40 yrs	RW, FUA; STA <sub>2</sub> , BD <sub>2</sub> , RA <sub>1</sub> , UD <sub>2</sub> , UC <sub>2</sub> , RTP <sub>2</sub>
13	58043371-58099969	<i>GNAS</i>	2 <sup>nd</sup> 20 yrs	#1 for MY <sub>1</sub> , FY <sub>1</sub> , PY <sub>1</sub> , NM <sub>1</sub> , STA <sub>1</sub> , STR <sub>1</sub> , RLS <sub>1</sub> , SCE <sub>1</sub> , DCE <sub>2</sub>
14	1801116-2002873	<i>DGATI</i>	1 <sup>st</sup> 20 yrs	#1 for FPC <sub>1</sub> ; FPC
14	4240287-4336714	<i>NBP</i>	1 <sup>st</sup> 20 yrs	#2 for FPC <sub>1</sub> ; FPC, PPC; FY <sub>1</sub> , PY <sub>1</sub>
20	37939597-38252896	<i>CAPSL</i>	1 <sup>st</sup> 20 yrs	MY <sub>2</sub> , FY <sub>2</sub> , PY <sub>2</sub>
20	45887870-45966446	-	40 yrs	*RLS, RTP, MY <sub>2</sub> , FY <sub>2</sub> , PY <sub>2</sub> , SCS <sub>2</sub> , DPR <sub>2</sub> , STA <sub>2</sub> , STR <sub>2</sub> , DF <sub>2</sub> , BD <sub>2</sub> , RW <sub>2</sub> , FUA <sub>2</sub> , RUH <sub>2</sub> , UC <sub>2</sub> , FTP <sub>2</sub> , RLR <sub>2</sub>
20	46209952-46296840	<i>CDH9</i>	2 <sup>nd</sup> 20 yrs	STA <sub>2</sub> , STR <sub>2</sub> , FA <sub>2</sub> , BD <sub>2</sub> , RW <sub>2</sub> , RLR <sub>2</sub>
20	46555323-46900120	<i>CDH9</i>	40 yrs	SCS <sub>2</sub> , STA <sub>2</sub> , STR <sub>2</sub> , FA <sub>2</sub> , BD <sub>2</sub> , RW <sub>2</sub> , RLR <sub>2</sub>
26	20290497-20444634	<i>HPSE2</i>	1 <sup>st</sup> 20 yrs	PPC
26	21870752-22118554	<i>BTRC</i>	40 yrs ago	Unknown

<sup>a</sup>1<sup>st</sup> 20 yrs: AFD and EHH were different between groups I and II but were similar between II and III. <sup>2</sup>2<sup>nd</sup> 20 yrs: AFD and EHH were different between II and III but were similar between I and II. <sup>40</sup> 40 yrs: AFD and EHH were different between I and II and between II and III. <sup>40+</sup> 40+ yrs: AFD and EHH were different between I and II and between II and III, and Group I also had strong EHH patterns. <sup>40</sup> 40 yrs ago: Groups I-III had similar strong EHH patterns but had minor AFD. \*MY = milk yield, FY = fat yield, PY = protein yield, FPC = fat percent, PPC = protein percent, SCS = somatic cell score, DPR = daughter pregnancy rate, PL = productive life, SCE = sire calving ease, STA = stature, STR = strength, BD = body depth, RW = rumen width, DF = dairy form, UD = udder depth, FUA = fore udder attachment, RUH = rear udder height, FIP = front test placement, RTP = rear test placement, RA = ramp angle, UC = udder cleft, RLS = rear legs side-view, RLR = rear legs rear-view. Trait with subscript '1' was significant in contemporary cows, trait with '2' was significant in both contemporary and control cows, and trait without '1' and '2' was significant with or without control cows. \*This SNP effect was upstream and next to this core region.

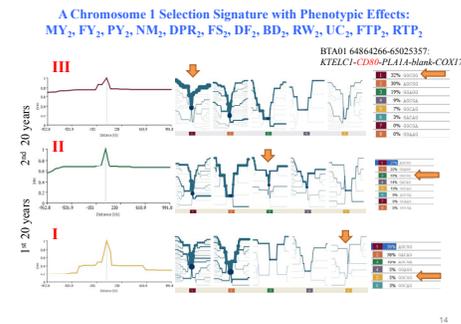


Figure 3: A selection signature with SNP effects.

- 1) GCGGG haplotype was subjected to positive selection for the past 40 years, with frequency increasing from 5% in Group I to 19% in Group II and to 32% in Group III;
- 2) The two selected populations (Groups II and III) had long-range EHH (> 1.8Mb), with Group III having the highest EHH values;
- 3) CD80 had the largest allele frequency changes in this block of five loci, with 'G' allele increasing 23% in 1<sup>st</sup> 20 years and 11% in 2<sup>nd</sup> 20 years;
- 4) The blank locus had the 2<sup>nd</sup> largest allele frequency changes, with 'G' allele increasing 16% in 1<sup>st</sup> 20 years and 14% in 2<sup>nd</sup> 20 years;
- 5) COX17 had no change in allele frequencies in the 40 years.

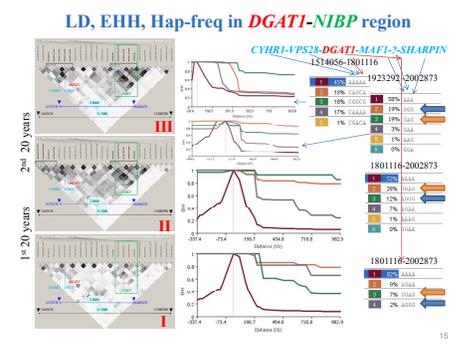


Figure 4: A selection signature in the DGATI-NIBP regions.

- 1) CYHR1-VPS28-DGATI and NIBP regions had strong LD signals in the selected populations (Groups II and III);
- 2) Selection in the 1<sup>st</sup> 20 years (from Group I to Group II) caused the frequencies of AGGG and GGAG to increase by 0.31;
- 3) Selection in the 2<sup>nd</sup> 20 years (from Group II to Group III) was backward selection in CYHR1-VPS28-DGATI region (AFD = 0.09-0.11 decrease);
- 4) 'G' alleles of SHARPIN and MAF1 had largest frequency increases (0.29-0.30) in 1<sup>st</sup> 20 years and slight decreases (0.02-0.05) in 2<sup>nd</sup> 20 years;
- 5) Selection in the NIBP region also occurred during the 1<sup>st</sup> 20 years, with little change in the 2<sup>nd</sup> 20 years (EHH figures not shown).

## References

- [1] Sabeti, P. C. et al. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832-837 (2002).
- [2] Ma, L. et al. Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies. BMC Bioinformatics 9:315 (2008).

## ACKNOWLEDGEMENT

This research was supported by National Research Initiative Grant no. 2008-35205-18846 from the USDA Cooperative State Research, Education, and Extension Service, Holstein Association USA, project MN-16-043 of the Agricultural Experiment Station at the University of Minnesota; and the Minnesota Supercomputer Institute. The assistance of T. Lawlor (Holstein Association USA), M. Cowan (Genetic Visions), R. Wilson (Genex Cooperative), C. Dechow (Pennsylvania State University), B. Harvey (USDA/ARS), L. Hansen (University of Minnesota), D. Sparlock (Iowa State University), A. de Vries (University of Florida), and B. Cassell (Virginia Polytechnic Institute and State University) in obtaining DNA samples is appreciated.