Changes in the genome architecture of two groups of dairy bulls with marked differences in their direct genomic breeding values for production traits in the UK

R. Mrode[1,3], D. Pitt[2], M. Winters[2] and Mike Coffey[1]
[1]Scotland Rural College, Edinburgh, UK
[2]Agriculture and Horticulture Development Board, Coventry, UK
[3]International Livestock Research Institute, Nairobi, Kenya

Abstract

Genomic selection has resulted in a rapid rate of genetic progress in dairy cattle in the last decade which could partly be attributed to a marked reduction in generation interval. For instance, in the UK the average age of a bull when their 100th daughter was born has decreased from over 6 years prior to the introduction of genomics to under 4 years in 2022. It is however not clearly understood if this rapid rate of genetic progress has also been accompanied by changes in the genome architecture in terms of degrees of heterozygosity, allele frequencies and linkage disequilibrium (LD) structure. This study examines changes in these parameters in 9,202 bulls born between 2009 and 2014 in the reference population for production traits and 94,204 young bulls with no daughters born between 2019 and 2023. The mean difference in direct genomic breeding values (DGV) between these reference and young bulls (YG) were 357.7, 24.5 and 16.1 kg for milk, fat and protein yields respectively. The SNP panel used in the UK evaluations consists of 79,051 SNPs and the proportion of SNPs with 0 to 1, 2 to 5, 6 to 10, 11 to 15 and greater than 15 percent change in their allele frequencies between both the reference and young bull groups were 15, 32, 29, 15 and 9% respectively. The SNPs with at least 15% change in allele frequencies accounted for 34, 43 and 39 percent of the mean difference in DGV between reference and young bulls for kg milk, fat and protein respectively, while the corresponding values for SNPs with 0 to 1 % change in allele frequencies were less than one percent for all three traits. In absolute terms the correlation between differences in mean DGV between reference and young bulls and changes in allele frequencies at the chromosome level was about 0.65 for the three traits. Thus, the rapid rate of genetic progress due to genomic selection is significantly changing allele frequencies. The rate of LD decay was similar for both groups of bulls, but tended to be higher for YG, suggesting stronger selective pressures and/or lower effective population size. The increased rate of inbreeding in shorter ROH but slower increase in longer ROH in YG seem to imply that recent inbreeding is being better controlled than very ancient inbreeding.

Keywords: genomic selection, allele frequency changes, runs of homozygosity, linkage disequilibrium

Introduction

The application of genomic selection has resulted in rapid rates of genetic gain, especially in dairy cattle, due to reduced generation intervals, higher selection intensity and higher accuracies for young genomic bulls for production and fitness traits. For an example, the annual rates of genetic gain for production traits in the USA has doubled with a Net Merit$ of $40.33 per year from 2005 to 2010 to $79.20 per year from 2016 to 2020. Consequently, in 2021 young genomic sires with no milking progeny accounted for 71% of A.I. breeding in U.S. dairy herds (CDCB 2023). The equivalent UK statistics are even higher in magnitude, where the Herdbook registered Holstein females saw an average yearly gain in £PLI (Profitable Lifetime Index) of £17.9 between 2005 to 2010, which tripled to £58.2 between 2016 and 2020. Similar to the US, over 72% of matings in 2022 were by genomic young sires. The genetic gains are bolstered by the high usage of sexed dairy semen, with 82% of all Holstein semen sold during 2022 sexed. It is however not clearly understood if this rapid rate of genetic progress has also been accompanied by changes in the genome architecture in terms of allele frequencies, linkage disequilibrium (LD) structure and inbreeding rates. This study examines changes in allele frequencies, LD structure and inbreeding in two groups of bulls with large differences in their mean direct genomic breeding values which were in the reference and selection populations.

Materials and methods

Two groups of bulls used in this study: 9,202 bulls born between 2009 and 2014 which were in the reference (REF) population of the UK genomic prediction model for April 2023 for production traits and 94,204 young (YG) bulls with no daughters born between 2019 and 2023, which were selection candidates. The means (and standard deviations) for the direct genomic breeding values (DGV) for 9,202 in the reference population were 192.5kg (81.3),7.9kg (3.80) and 7.0kg (3.56) for milk, fat and protein yield respectively. Corresponding means and SDs for the young bulls were 550.2kg (68.08), 32.4kg (5.78) and 23.1kg (3.23) respectively. This implies that the mean difference in DGV between these reference and young bulls were 357.7, 24.5 and 16.1 kg for milk, fat and protein yields respectively.

The SNP panel used in the UK evaluations consists of 79,051 SNPs and allele frequencies were computed within each of the two groups of bulls. Using the SNP effects from the UK genomic prediction in April 2023 and the file of genotypes, the relative contribution by each SNP to the DGV, with polygenic contributions ignored, were computed and used to examine the relationship between changes in allele frequencies and contribution of each SNP to the difference between the DGVs for REF and YG bulls at chromosome and SNP levels.

Linkage disequilibrium

Linkage disequilibrium was estimated using PLINK 1.9 (Chang et al. 2015). The pairwise R-squared values were calculated between SNPs along the same chromosome and were categorised into 1 kb bins based on physical distance.

Runs of homozygosity

Runs of homozygosity (ROH) were detected with PLINK 1.9 (Chang et al. 2015). Note that for the ROH analysis, 186 REF and 1,109 YG bulls with substantial Friesian breed contribution were removed from the dataset. For this analysis, a scanning window of 10 SNPs was used, with a maximum of one heterozygote call per window. ROH were further restricted, using the following parameters: a minimum SNP count of 10; at least one SNP per 100 kb; a maximum of one heterozygous SNP per ROH; and a minimum physical length of 1 Mb. The remaining parameters were left as default.

ROH were split into six classes using physical length: 1-2 Mb; >2≤4 Mb; >4≤8 Mb; >8≤16 Mb; >16<32 Mb; and >32 Mb. Longer ROH are expected to represent regions of autozygosity from more recent inbreeding, contrastingly, shorter runs are expected to represent more ancient inbreeding due to the cumulative chance that recombination events have occurred within the ROH. The expected generational source of different ROH length classes was estimated as in Doekes et al. (2019). Briefly, common ancestors that occurred G generations ago gives rise to ROH of various length that follows an exponential distribution with mean 1/2G Morgan. A uniform recombination rate was assumed at 1 Morgan per 100 Mb.

Inbreeding coefficients (F) were calculated from ROH (FROH) using the percentage of the autosome covered by ROH in each length class for a given individual. Density distributions for each length class were compared between the REF and YG bull groups. The goodness of fit between each pair of distributions were tested using the Kolmogorov-Smirnov test. The modes and 95% confidence intervals, calculated from the 0.025th and 0.975th quantiles, were also qualitatively compared.

The data were further split by year of birth and linear regressions were calculated within and across the REF and YG bull groups to estimate the rate of increase in inbreeding for each ROH length class. For the trends calculated by year of birth, bulls belonging to the 2015 and 2023 cohorts were removed due to small sample size (Table 1); bulls from 2023 were retained for visualization only.

Table 1. Distribution of bulls by year of birth for reference (REF) and young (YG) bulls

Reference bulls		Yong bulls	
Year	Count	Year	Count
2009	1341	2019	23319
2010	1436	2020	20341
2011	1466	2021	25244
2012	1667	2022	24001
2013	1532	2023	190
2014	1562		
2015	2		

Results and Discussion

The distribution by chromosome of SNPs with more than 5% changes in their allele frequencies between REF and YG bulls is presented in Figure 1. Of the total 75,091 SNPS, the percentage of those with 0-1, 2-5, 6-10, 11-15 and > 15 percent changes in their allele frequencies between REF and YG were 15, 32, 29, 15 and 9% respectively. The top 5 chromosomes with the highest number of SNPs with changes in their allele frequencies were BTA1, BTA3, BTA9, BTA14 and BTA30. Figure 2 presents the distribution by chromosome of SNPs with more than 15% changes in their allele frequencies between REF and YG bulls. Again, the chromosomes with the highest number of SNPs showing these changes were similar to those in Figure 1 and these were BTA1, BTA9, BTA13, BTA14 and BTA30. No previous studies were found that examined changes in allele frequencies but with about 10% out of 75,091 SNPs with more than 15% changes in allele frequencies, it would appear, substantial changes in allele frequencies have been observed in this study.

The relative contribution of individual SNPs to the overall DGV, indicates that the SNPs with more than 15% changes in their allele frequencies accounted for 34, 43 and 39 percent of the mean difference in DGV between reference and young bulls for kg milk, fat and protein respectively. The corresponding values for SNPs with 0 to 1 % change in allele frequencies were less than one percent for all three traits. The percentage contribution to the difference in DGVs between REF and YG bulls by chromosome is shown in Figure 3. The results show that chromosome BTA14 accounted for about 15% of the positive difference between REF and YG bulls for fat yield with a corresponding similar negative percentage difference between both groups of bulls for milk. This could be attributed to the presence DGAT1 on this chromosome and similar results have been reported by Sun et al, (2009). The chromosome BTA1 contributed the largest positive difference between both REF and YG bulls for milk and protein yields. In absolute terms the correlation between differences in mean DGV between reference and young bulls and changes in allele frequencies at the chromosome level was about 0.65 for the three traits. Thus, the rapid rate of genetic progress due to genomic selection has significantly changed allele frequencies.

Table 2. Summary statistics of FROH distributions across ROH length classes.

	Reference bulls					Young bulls				
Bin	Mean	SD	Mode	Q2.5	Q97.5	Mean	SD	Mode	Q2.5	Q97.5
1-2 Mb	2.31%	0.38%	2.27%	1.62%	3.09%	2.73%	0.52%	02.66%	1.91%	3.79%
2-4 Mb	2.32%	0.57%	2.23%	1.31%	3.51%	3.31%	0.79%	3.09%	1.98%	5.01%
4-8 Mb	3.18%	0.96%	2.97%	1.50%	5.28%	5.10%	1.26%	4.85%	2.74%	7.70%
8-16 Mb	3.42%	1.39%	2.87%	0.98%	6.42%	5.55%	1.80%	5.42%	2.18%	9.24%
16-37 Mb	2.29%	1.60%	1.72%	0.00%	6.04%	3.54%	2.00%	2.57%	0.00%	8.01%
>32 Mb	0.82%	1.32%	0.00%	0.00%	4.36%	1.18%	1.66%	0.00%	0.00%	5.39%

Table 3. Linear regression slope coefficients of inbreeding (FROH) change in bulls by year of birth, calculated for each ROH length class.

	Birth years	1-2 Mb	2-4 Mb	4-8 Mb	4-8 Mb	16-37 Mb	>32 Mb
All bulls	2009-2014, 2019-2022	0.044%	0.109%	0.211%	0.228%	0.135%	0.038%
REF bulls	2009-2014	0.025%	0.082%	0.159%	0.204%	0.181%	0.096%
YG bulls	2019-2022	0.032%	0.107%	0.204%	0.178%	0.101%	0.011%

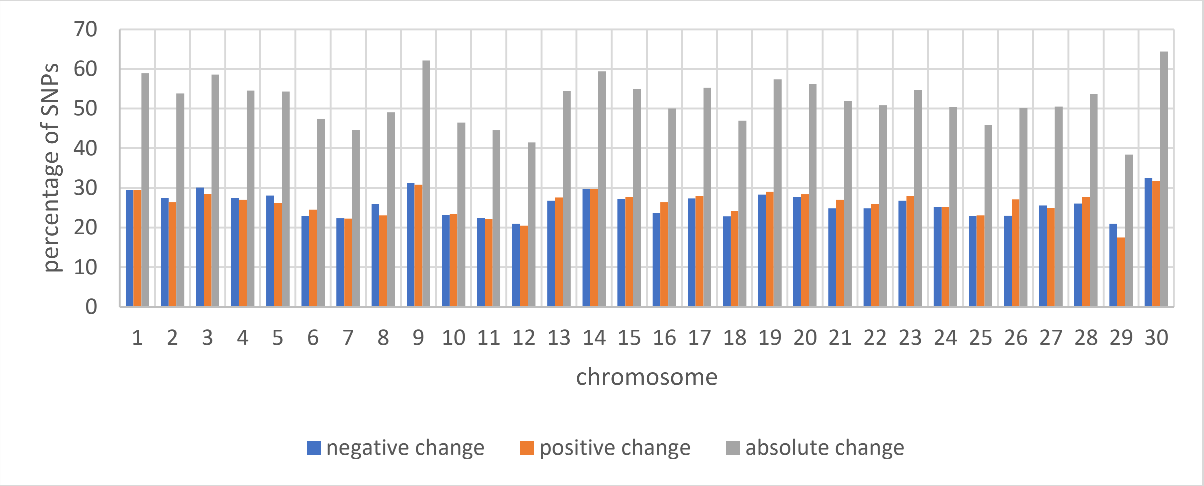

Figure 1: The distribution of SNPs with at least 5% changes in allele frequencies in young bulls relative to the reference bulls by chromosome

Linkage disequilibrium

The decay of LD follows similar curves for both the REF and YG bull groups, with a maximum R^2 at the shortest physical distance which subsequently decays by 48% to a relatively static minimum within approximately 13 kb (Figure 4). While the rate of decay was similar between groups, the magnitude differed. At the shortest physical distance between 0-1 kb, LD was greater in YG bull ($R^2 = 0.87$) than REF bulls ($R^2 = 0.79$), lowering to R^2 of 0.45 and 0.41, respectively at 13 kb. LD decay beyond this distance was at a slower rate, however, binned median R^2 values were consistently greater in YG bulls compared to REF bulls up to the maximum distance of 200 kb analysed.

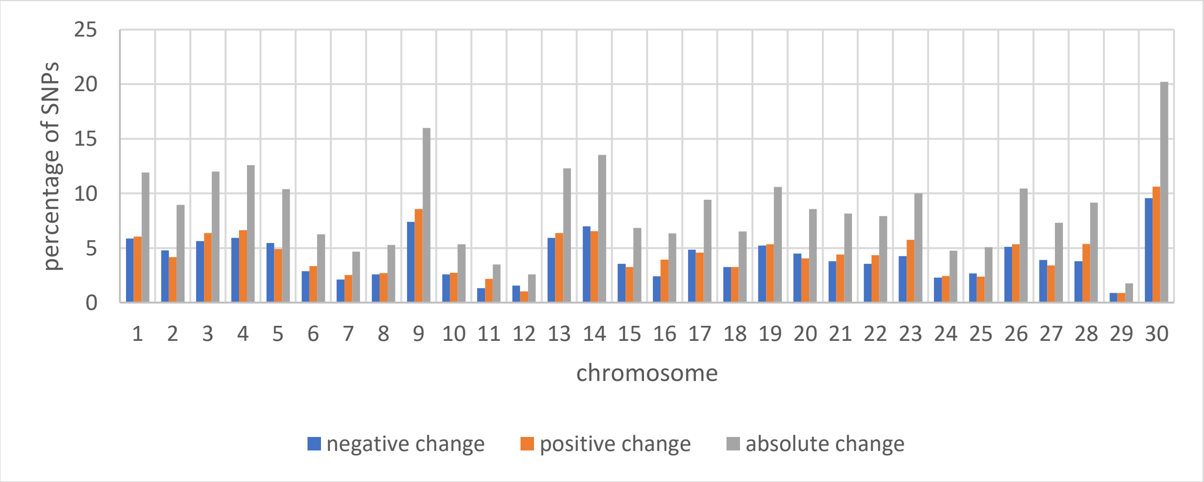

Figure 2: The distribution of SNPs with more than 125% changes in allele frequencies in young bulls relative to the reference bulls by chromosome

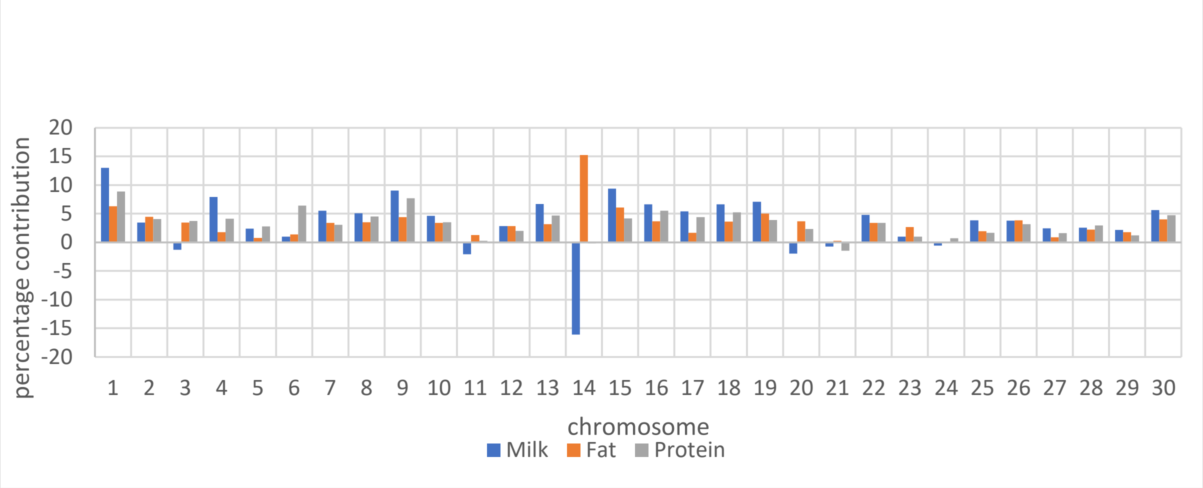

Figure 3. The percentage contribution to the difference in DGVs between reference and young bulls by chromosome

Runs of Homozygosity and Inbreeding

For each ROH size class, the Kolmogorov–Smirnov test was used to test for significant differences in the density distributions between REF and YG bull groups; each comparison was significantly different ($P < 0.001$). For all ROH size classes except >32 Mb, YG had greater inbreeding with a mode at least 17% greater than the reference bulls. The largest proportional difference between modes was for the 5-8 Mb and 9-16 Mb classes, with a 63% and 89% increase, respectively (Table 2). While YG inbreeding has increased across all size classes, the additional increase seen between 4-16 Mb indicates a substantial increase of matings with shared ancestry in the past 10 generations (Figure 5), with 31% of ROH between 9-16 Mb likely originating from the past 2-5 generations. These increases were repeated in the upper and lower bounds of the 95% confidence intervals, with a minimum increase of 17% and the greatest increase observed for 5-8Mb and 9-16Mb (lower confidence intervals for both groups for 17-32Mb and >32Mb remained at zero, therefore no increase was observed). This further validates the overall shift of increased inbreeding between the two groups. Due to

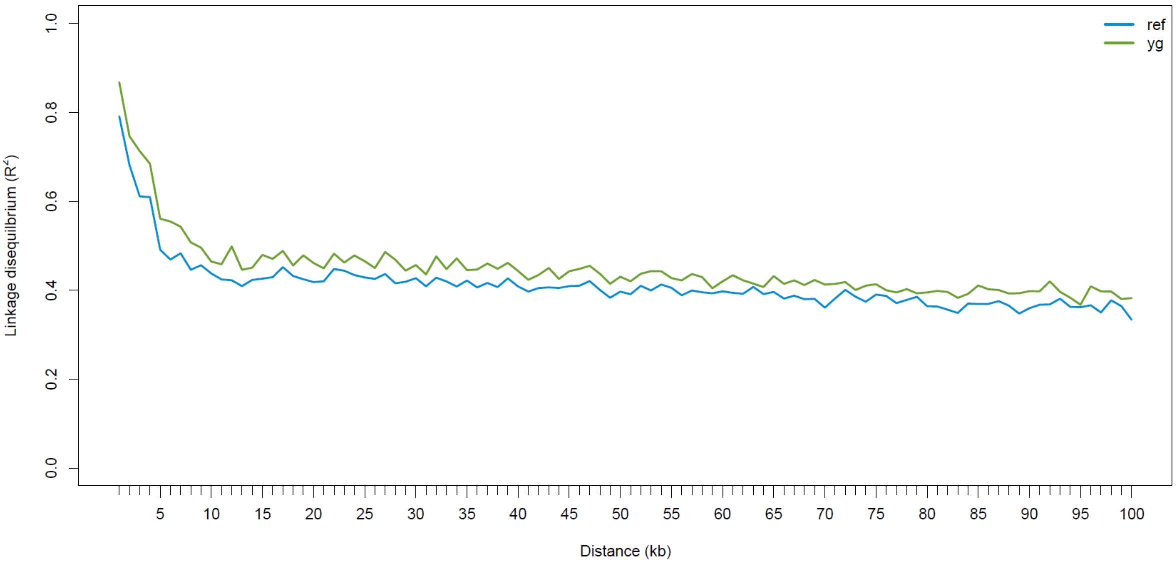

Figure 4. Linkage disequilibrium decay over physical distance in REF and YG bulls

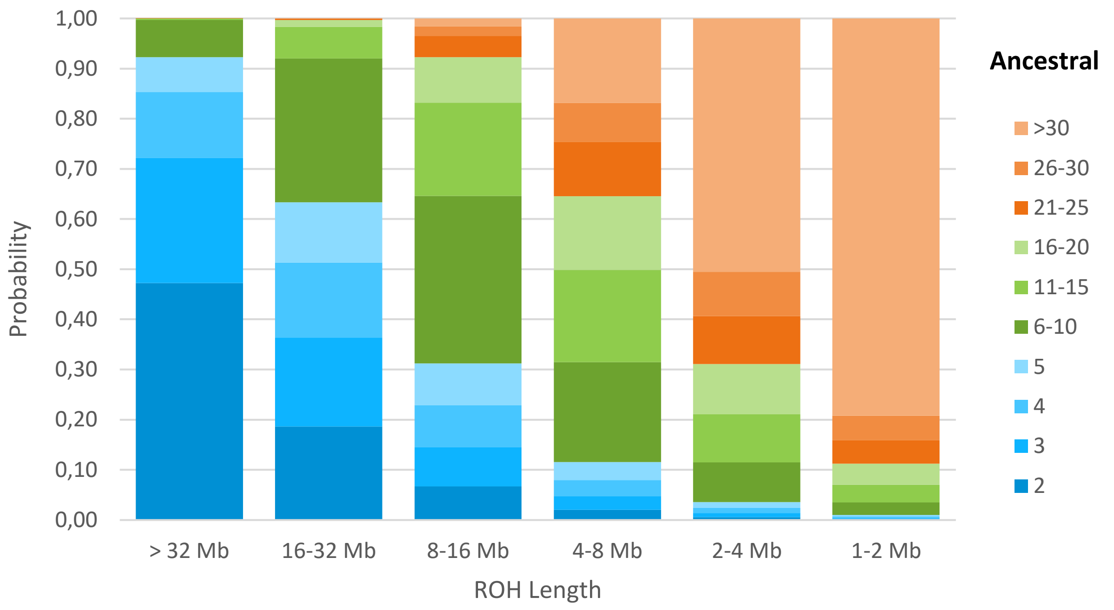

Figure 5. Expected age of ROH by length classes, based on exponential distribution with mean 1/2G Morgan and a uniform recombination rate of 1 Morgan per 100 Mb. Adapted from Doekes et al. (2019).

most individuals having zero ROH exceeding 32 Mb, the density distributions for >32 Mb formed a multimodal curve with a major mode of 0, therefore, differences between the populations were difficult to compare for this size class. The average inbreeding for >32 Mb was 50% greater in YG compared to REF populations which may indicate further inbreeding in the past 2-5 generations for YG, however, mean calculations for this size class are particularly susceptible to skewing from rare outliers due to no upper limit on maximum ROH length.

ROH class bins were further divided by year of birth to characterise trends over time – due to the small sample sizes, 2015 (n = 2) and 2023 (n = 190) cohorts were removed from trend calculations. A linear regression was calculated between year of birth and median inbreeding values, the slope coefficient therefore representing the change in genomic inbreeding per year. Inbreeding across all ROH classes is increasing over time, except for >32 Mb, where all years have a median of zero. Linear regressions were also calculated for data within the both reference and YG groups; for the shorter ROH length classes (1-2, 3-4, and 5-8 Mb) the rate of inbreeding is accelerating, with YG experiencing ~27-30% increase in the rate of increased inbreeding per year, compared to the reference group. Contrastingly, for longer ROH length classes (9-16, 17-32, and >32 Mb), more likely to arise from recent shared common ancestry (Figure 5), the yearly rate of increase is less for YG compared to that observed in the reference group (Table 3). This is potentially capturing the paradigm shift between the breeding schemes of the two groups; with the selection of more recent bulls, there is greater selection pressure on improving GEBVs while less consideration is given for ancient inbreeding allowing the relative rate of inbreeding of shorter ROH to increase.

The linear regression slope coefficient, for 17-32 Mb and >32 Mb, calculated across all bulls (all birth years) falls within the magnitude of the coefficients calculated for the REF and YG groups. This suggests that during the period that is unrepresented in our data (2015-2018, inclusive) the rate of inbreeding increase is approximately in line with the average that is observed across the two groups, and that a gradual change in reducing more recent inbreeding has occurred over time. In contrast, ROH length classes representing relatively more ancient inbreeding (1-2, 3-4, and 5-8 Mb) the slope coefficient exceeded that of either of the two groups (Table 3). It seems that breeding decisions between 2015 and 2018, resulted in a high rate of increase between these years.

Conclusion

The marked increase in DGV of young bulls due to genomic selection has been accompanied with substantial changes in allele frequency. The SNPs with at least 15% changes in allele frequencies accounted for 34 to 43 percent of the mean difference in DGV between reference and young bulls for production traits. The percentage contribution at the chromosome level reveals the high impact of DGAT in chromosome BTA14 for fat and Milk yield. The rate of LD decay was similar for both groups of bulls, but LD tended to be higher for YG, suggesting stronger selective pressures and/or lower effective population size. Increased rate of inbreeding in shorter ROH but slower increase in longer ROH in YG seem to imply that recent inbreeding is being better controlled than very ancient inbreeding. It should be noted however that this study only considered DGV for production, but that the overall breeding goal (PLI) includes many more traits (34% of current PLI is production), which will contribute to the changes observed.

References

CDCB 2023. Impact of genomics. https://uscdcb.com/genomic-impact/). Accessed 18 September, 2023

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), s13742-015. https://doi.org/10.1186/s13742-015-0047-8

Doekes, H. P., Veerkamp, R. F., Bijma, P., de Jong, G., Hiemstra, S. J., & Windig, J. J. 2019. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. *Gen. Selec. Evol.*, 51, 1-16. https://doi.org/10.1186/s12711-019-0497-z

Sun, D., Jia, J., Ma Y., Zhang, Y., Wang, Y., Yu, Y., and Y, Zhang. 2009. Effects of DGAT1 and GHR on milk yield and milk composition in the Chinese dairy population. *Anim Genet*,. 40:997-1000. https://doi.org/10.1111/j.1365-2052.2009.01945.x