

## DOES SELECTING FOR THE A2 B-CASEIN ALLELE INCREASE INBREEDING?

B.A. Scott<sup>1,2</sup>, M. Haile-Mariam<sup>1</sup>, I.M. MacLeod<sup>1</sup> and J.E. Pryce<sup>1,2</sup>

<sup>1</sup>Agriculture Victoria Research, AgriBio, 5 Ring Rd, Bundoora, Victoria 3083, Australia

<sup>2</sup>School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

### SUMMARY

Milk products from cows that are homozygous for the A2  $\beta$ -casein allele are marketed in several countries by the A2 Milk Company. The alleles present at the  $\beta$ -casein locus of genotyped sires is published by DataGene and available to farmers when making selection decisions, alongside the estimated breeding values and selection indices. We hypothesised that intense selection for the A2 allele may have resulted in increased inbreeding. In this study we compared differences in genome wide and regional homozygosity between the two homozygotes of the  $\beta$ -casein alleles (A1 and A2) using medium density genotypes (50K SNP chips) of Holsteins cows. The A2 mutation was imputed into study animals, having first validated this approach in a group of bulls with known or certified genotypes for the  $\beta$ -casein locus. This study shows that the frequency of the A2 homozygote has increased by 20% since 2000 in Holstein cows. Our results suggest that selection for the  $\beta$ -casein A2 allele has increased inbreeding both across the genome and on chromosome 6 in Holstein cows that are homozygous for the A2 allele. Animals that had two A2 alleles were twice more likely to have a run of homozygosity of at least 35 SNP or 1000 kb long across the  $\beta$ -casein locus compared to animals that were homozygous for A1.

### INTRODUCTION

Seventy-five percent of milk protein content and composition can be linked to four casein genes CSN1S1, CSN2, CSN1S2, and CSN3, encoding the casein proteins alpha S1 ( $\alpha$ S1), beta ( $\beta$ ), alpha S2 ( $\alpha$ S2), and kappa ( $\kappa$ ), respectively (Ferretti *et al.* 1990, Threadgill and Womack 1990). This casein cluster spans ~250kb on BTA 6 (Boettcher *et al.* 2004). Selection for the A2  $\beta$ -casein allele, has increased rapidly due to the commercialisation of milk products by the a2 Corporation in 2000. Milk products from the A2 Corporation are produced from cows that are homozygous for the A2  $\beta$ -casein protein where herds need to have cows that are exclusively A2 homozygous. The alleles present at the  $\beta$ -casein locus of genotyped sires is published by DataGene and available to farmers when making selection decisions alongside the estimated breeding values and selection indices. Female genotyping at the  $\beta$ -casein locus can also be pursued by farmers who wish to build an A2 homozygous herd.

An implication of this is that intense selection for homozygosity at a given locus (A2) may result in increased inbreeding. Inbreeding can result in a loss of genetic diversity, decreased response to selection, reduced animal performance and ultimately, decreased farm profitability. Traditional pedigree methods to calculate inbreeding often underestimate the level of inbreeding due to incomplete pedigree and errors. With the availability of genotype information, we can now calculate inbreeding coefficients more accurately and are able to distinguish between recent and ancient inbreeding using runs of homozygosity. With the introduction of genomic selection the rate of inbreeding per year has increased (e.g. Doekes *et al.* 2018, Doublet *et al.* 2019, CDCB 2020), irrespective of the selection for specific alleles.

In this study we determined the frequency of homozygotes of the  $\beta$ -casein alleles (A1 and A2) over an 18-year period (2000-2017) in Australian Holstein cows and compared differences in genome wide and regional homozygosity between the two homozygotes of the  $\beta$ -casein allele using 50K SNP chip genotypes.

## MATERIALS AND METHODS

**Data.** A total of 139,898 genotyped individuals were available for Holsteins, Jerseys and their crosses from DataGene. The genotyping was carried out by various commercial providers. DataGene imputed the genotypes to a standard set of 45,685 SNP genotypes for routine evaluations (Nieuwhof *et al.* 2010). The breed of genotyped cows (Holstein) was validated using the ADMIXTURE program (Alexander *et al.* 2009). After correcting the breed information we had 114,567 Holsteins cows for subsequent analysis. Of these 73,003 cows born between 2000 and 2017 were used.

We imputed the A2 alleles while imputing animals to whole genome sequence. Genotypes were imputed in a stepwise fashion by imputing any low-density genotypes to 50k, then to high density and finally full sequence. The sequenced reference population used for imputation was Run 7 of the 1000 Bull Genomes project that includes 3090 *Bos taurus* animals after QC (Hayes and Daetwyler 2019). Only homozygote animals for the A1 and A2 alleles were selected for analysis. This imputation approach for the A1 and A2 alleles was validated in a group of 443 bulls with known or certified genotypes for the  $\beta$ -casein locus and showed 98.4% concordance.

**Inbreeding coefficients.** Genomic inbreeding coefficients were calculated from runs of homozygosity (ROH), identified across autosomes. A ROH was defined as a homozygous segment of at least 35 SNPs or 1000 kb long, with at least one SNP per 75 kb. Two consecutive SNP could not be included if they were more than 300 kb apart. ROH were identified using the PLINK “*homozyg*” function (Purcell *et al.*, 2007) (command link: `plink --cow --bfile genotyping_data_filename --homozyg --homozyg-kb 1000 --homozyg-snp 35 --homozyg-window-snp 50 --homozyg-window-density 75 --homozyg-gap 300 --out output_filename`). ROH-based inbreeding estimates,  $F_{ROH,i}$ , were computed as the proportion of the genome included in the ROH as follows:

$$F_{ROH,i} = \frac{\sum L_{ROH,i}}{L_{auto}},$$

where  $\sum L_{ROH,i}$  is the total length of ROH for individual  $i$ , and  $L_{auto}$  the length of the autosome genome covered by SNPs after withholding gaps longer than 300 kb between two SNPs, corresponding to the length of the autosomal genome on which ROH can be detected. This parameter allowed for the detection of ROH on 92.2% of the autosomal genome.

For each individual, we also calculated the mean ROH length which is defined as:

$$L_{ROH,mean,i} = \frac{\sum L_{ROH,i}}{N_{ROH,i}},$$

where  $\sum L_{ROH,i}$  is the total length of ROH for individual  $i$  in kb, and  $N_{ROH,i}$  the total number of ROH for individual  $i$ .

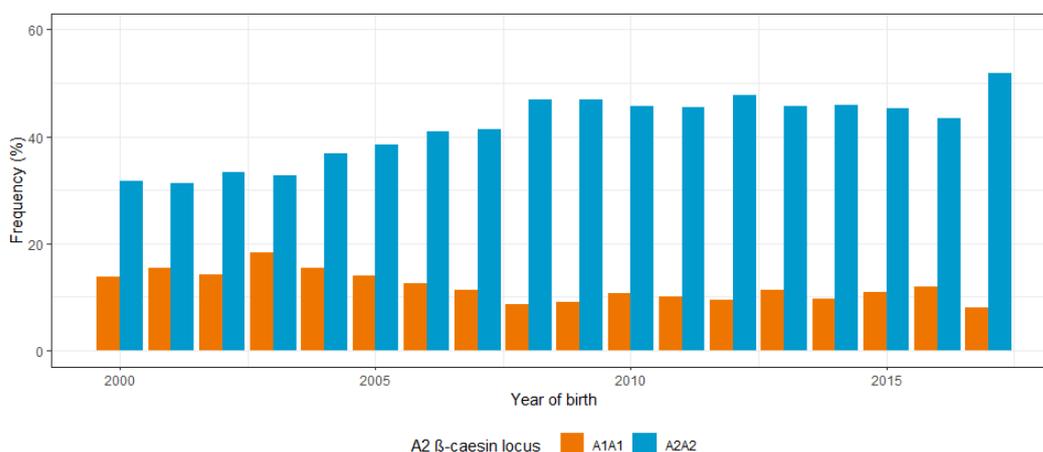
Additionally, we also compared genomic inbreeding (ROH) for Holstein cows that were homozygous for A2 and A1 alleles specifically on chromosome 6 (Chr6:87181619).

Wilcoxon-Mann Whitney tests were used to determine if there were significant differences between the two homozygote groups (A1/A1 versus A2/A2) in inbreeding levels for: all animals (born between 2000-2017) as well as between the same groups but only for young animals (born after 2013). A Chi-squared tested was used to determine if there was a difference between the observed versus expected number of animals with ROH over the A2 position. A total of 39,157 cows had both a known homozygosity over the  $\beta$ -casein locus and inbreeding coefficient and were used for the analysis.

## RESULTS AND DISCUSSION

**Frequency.** There were more than four times the number of A2/A2 cows (31,814) compared to A1/A1 (7,335) in the dataset. Figure 1 demonstrates how the frequency of the  $\beta$ -casein allele has

changed in Holstein females since 2000. The A2/A2 frequency has increased from 32% in 2000 to 52% in 2017, suggesting growth of interest in  $\beta$ -casein possibly sparked by the a2 Corporation.



**Figure 1** Frequency of the A1/A1 and A2/A2 homozygous Holstein cows born in 2000 to 2017

**Inbreeding.** The median  $F_{ROH}$  was consistently higher in A2 than A1 homozygotes (Table 1). Over the whole genome, larger differences were observed when all animals were included (difference of 0.43%) than only young animals (animals born after 2013; difference 0.15,  $p = 0.002$ ).

We observed significantly more regional inbreeding on Chromosome 6 for A2/A2 animals than for A1/A1, with median  $F_{ROH}$  of 9.44% and 8.07%, respectively.

**Table 1** The median genomic inbreeding values ( $F_{ROH}^*$ ) for all animals (born in 2000-2017) or young animals (born in 2014-2017) homozygous for the  $\beta$ -casein A1 or A2 allele and the Wilcoxon-Mann Whitney significance test p-values

	No. A1/A1	No. A2/A2	Median $F_{ROH}$ A1/A1	Median $F_{ROH}$ A2/A2	p-value
<i>Whole Genome</i>					
All	7,335	31,814	7.89	8.32	< 2.2e-16
Young Animals	3,907	18,893	8.63	8.78	0.002
<i>Chromosome 6</i>					
All	5407	25301	8.07	9.44	< 2.2e-16
Young Animals	3055	14977	8.89	9.60	8.211e-05

\*ROH - Runs of homozygosity

**Inbreeding over the  $\beta$ -casein locus.** We found that while A2/A2 animals were more likely to have a ROH over the  $\beta$ -casein locus (chi-square statistic 262,  $p < 0.00001$ ), the length of the ROH was longer for the A1/A1 animals (median ROH length 6,136 kb vs. 3,706 kb). When comparing this subset of animals to the median inbreeding observed over the whole genome, we found that animals with an ROH over the  $\beta$ -casein locus had higher overall level of inbreeding (median  $F_{ROH}$  10.0% subset vs. 7.89% population and 9.2% subset vs. 8.2% population for A1/A1 and A2/A2 animals, respectively).

In this study, we did not determine if the genotyped cows were representative of the entire population. It is possible that farmers who breed strictly for A2/A2 individuals are more likely to genotype their cows than those not breeding for A2/A2, resulting in an overestimation in this study of the frequency of A2/A2 homozygous cows across the Australian dairy industry. Additionally, we did not determine if the differences in inbreeding between A1 and A2 homozygotes was due to selection for A2 rather than simultaneous selection for protein yield or other economic traits. Further work using imputed sequence genotypes across the region encompassing the casein gene cluster as well as the heterozygote individuals may allow us to determine these differences. Understanding these mechanisms could have a wider implication for assessing the benefits and shortcomings of narrow selection strategies. Perhaps farmers that are interested in selecting for particular alleles should pay attention to monitoring inbreeding and its through the use of appropriate mate selection methods. This may have implications for breeders that may consider selection for specific alleles that are currently at low frequency, such as, the polled region.

## CONCLUSIONS

This study shows that the frequency of the A2 homozygote has increased by 20% since 2000 in Holstein cows. Our results suggest that A2/A2 animals were more inbred over the whole genome as well as on chromosome 6 and were more likely to have a ROH over the  $\beta$ -casein locus.

## REFERENCES

- Alexander D.H., Novembre J. and Lange K. (2009) *Genome Res* **19**(9): 1655.
- Boettcher P.J., Caroli A., Stella A., Chessa S., Budelli E., Canavesi F., Ghiroldi S. and Pagnacco G. (2004) *J Dairy Sci* **87**(12): 4311.
- CDCB. 2020. Retrieved June, 2020, from <https://queries.uscdcb.com/eval/summary/inbrd.cfm>.
- Doekes H.P., Veerkamp R. F., Bijma P., Hiemstra S. J. and Windig J. J. (2018) *GSE* **50**(1): 15.
- Doublet A.-C., Croiseau P., Fritz S., Michenet A., Hozé C., Danchin-Burge C., Laloë D. and Restoux G. (2019) *GSE* **51**(1): 52.
- Ferretti L., Leone P. and Sgaramella V. (1990) *Nucleic acids research* **18**(23): 6829.
- Hayes B. J. and Daetwyler H. D. (2019) *Annu Rev Anim Biosci* **7**: 89.
- Nieuwhof G., Beard K., Konstantinov K., Bowman P. and Hayes B. (2010) *Interbull Bulletin* **42**: 35.
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D., Maller J., Sklar P., de Bakker P. I, Daly M. J. and Sham P. C. (2007). *Am J Hum Genet* **81**(3): 559.
- Threadgill, D. W. and Womack J. E. (1990) *Nucleic acids research* **18**(23): 6935.