

EXTENSIVE SEQUENCING OF A TROPICALLY ADAPTED BREED – THE BRAHMAN SEQUENCING PROJECT

L. Koufariotis¹, B.J. Hayes¹, M. Kelly², B. Burns³, R. Lyons⁴ and S. Moore¹

¹ Centre of Animal Science, Queensland Alliance for Agriculture and Food Innovation, The

²University of Queensland, Brisbane, 4072, QLD, Australia

²Australian Agricultural Company (AACo), Brisbane, QLD, Australia

³Queensland Department of Agriculture and Fisheries, Rockhampton, 4702, Queensland, Australia

⁴School of Veterinary Science, University of Queensland, Gatton, Queensland, 4343, Australia

SUMMARY

Brahman cattle are well adapted to tropical environments and are extensively used for beef production in Northern Australia. Identifying mutations in Brahman genomes associated with adaptation, fertility, meat quality and growth rates would facilitate genome selection and therefore accelerate genetic gain for these traits, in both Brahman cattle and composite cattle with Brahman ancestry. In this paper, 36 million high quality variants (SNP and Indels) were discovered from 46 whole genome sequenced Brahman bulls that represent key ancestors of the breed in Australia. As some infusion of *Bos taurus* into Brahman cattle has occurred during breed formation, we investigate regions of the Brahman genome that have high *Bos taurus* introgression. We identified multiple genome regions in Brahmans that were *Bos taurus* in origin, and investigated the roles, pathways and trait associations the of genes found in these regions.

INTRODUCTION

Brahman cattle are a breed of *Bos indicus* cattle, developed in the southern United States that are well adapted to tropical environments. They are a cross breed between four different types of Zebu cattle; Gir, Gujarat, Ongole and Krishna Valley. In northern Australia, Brahman cattle have a major impact on the Australian beef industry and are widely used in beef production due to their suitability to these harsh environments. To increase genetic diversity, Brahman cattle have been “graded up” from existing *Bos taurus* breeds in both Australia and United States.

Identifying mutations in Brahman genomes associated with adaptation, fertility, meat quality and growth rates would facilitate genome selection and therefore accelerate genetic gain for these traits, in both Brahman cattle and composite cattle with Brahman ancestry. With this ultimate aim, 46 Brahman cattle that were key ancestors of the breed were whole genome sequenced in this study. We first identified all the variants in these genomes such as single nucleotides (SNPs) and insertion/deletions (Indels), then annotated the variants into functional classes based on their locations on the genome. Finally, we used the sequence information to identify, for each bull, whether chromosome segments were indicine or taurine in origin.

MATERIALS AND METHODS

Bulls for sequencing were selected using an algorithm that identified 46 bulls that captured the highest proportion of genetic variation in the breed, based on an analysis of an extensive Brahman pedigree and a stepwise regression procedure to avoid double counting of ancestral genomes, and took into account whether DNA, extracted from semen straws or Ampules, was available for a bull or not (Daetwyler 2014; Druet 2014). The selected bulls were sequenced on an Illumina HiSeq sequencer, at an average of 12.5 times genome coverage, and a range of 10 times genome coverage to 30 times genome coverage.

Reads were mapped to the bovine genome (UMD3.1) with BWA, and variants were detected in the sequence with a GATK pipeline (McKenna *et al.* 2010). The variants included single nucleotide

Beef I

polymorphisms (SNP) and small insertion deletions (indels).

SNP variants were filtered based on modified filtering thresholds recommended by the GATK Best-Practices guideline (DePristo *et al.* 2011) to remove SNP that had poor quality scores.

Bos taurus and Gir variants were obtained from the 1000 bulls genomes project (Daetwyler *et al.* 2014) and from the study by (Liao *et al.* 2013), respectively. All SNP were obtained from each study and the allele frequencies for each SNP was calculated using in-house scripts and common SNP found in Brahman cattle were selected for. The population structure between Brahman and *Bos taurus* and between Brahman and Gir cattle was determining by calculating the Fixation value (FST) for each SNP in both taurine and Gir with the SNP in each individual Brahman animal, based on the methodology by the study (Bolormaa *et al.* 2011), using the following formula:

$$FST = \frac{Ht - Hs}{Ht}$$

Where:

$$Hs = PBT(1 - PBT) + Pbrai(1 - Pbrai)$$

And

$$Ht = \frac{2 * (PBT + Pbrai)}{2} \times 1 - \left(\frac{PBT + Pbrai}{2}\right)$$

PBT is the SNP allele frequency of the alternative allele in either *Bos taurus* or Gir and *Pbrai* is the zygosity information of the Brahman SNP for that animal. *Hs* is to calculate heterozygosity in the subpopulation and *Ht* is to calculate the overall heterozygosity. This resulted in two datasets, one is the FST between each individual Brahman animal and the Gir SNPs and the other is between individual Brahman animals with *Bos taurus* SNPs. Following this, we grouped SNPs into fixed windows of 250 kilobases (kb) and calculated the average FST for all SNP within each fixed window across the genome. This was done by simply by adding the FST of all SNP found in the 250 kb fixed windows and dividing by the total number of SNP, as shown in the following formula:

$$FST_{avg} = \frac{\sum(FST_{fw})}{n}$$

Where *FST_{fw}* is the FST for the SNP found within a fixed window, and *n* is the total number of SNP found in that fixed window. Animals were then sorted based on date of birth (from oldest to youngest), and the average FST in each fixed window across all animals was calculated.

SNP Annotation: Brahman SNPs were annotated into intergenic, intragenic, introns, exons, CDS, UTR (both 5' & 3'), 5 kb upstream of TSS, 5 kb downstream of genes, missense, synonymous and splice site classes by querying the Ensembl variant database version 87 (Yates *et al.* 2016). CpG Isles annotations were from the study by (Su *et al.* 2014).

RESULTS AND DISCUSSION

Initial analysis of the sequence data revealed the Brahman genomes had a much higher rate of polymorphism than that observed in *Bos taurus* breeds. This is likely a reflection of a larger ancestral population size for *Bos indicus* cattle than *Bos taurus* cattle (pre-domestication), and the fact there was some infusion of *Bos taurus* breeds into Brahmans during breed formation. Additionally, as Brahman is the synthesis of 4 different indicus breed types, the impact on the breed formation is reduced.

Following SNP annotation, we find that, as expected, the majority of the variants are found within the intergenic regions of the genome, as shown in Table 1, with close to 74 % of all SNP found in this class.

Table 1. Total number of Brahman SNP found in each annotation class

Annotation	No. of SNP	Percent of Genome
Intergenic	26,505,585	73.39 %
Intragenic	9,608,525	26.61 %
Intron	9,088,608	25.17 %
Exon	433,758	1.20 %
CDS	349,093	0.97 %
3' UTR	69,733	0.19 %
5' UTR	16,426	0.05 %
5kb Downstream of TTS	692,447	1.92 %
5kb Upstream of TSS	699,424	1.94 %
Synonymous	128,397	0.36 %
Missense	83,069	0.23 %
CpG Meth	995,319	2.76 %
Splice Sites	21,789	0.06 %
Total	36,114,110	

Just over 25% of Brahman SNPs are found within the intragenic class (Table 1), the majority of which (95%) are actually intronic variants. The percent of variants that are found within coding genes (Exon and CDS class) is 1.2%. Similar results have been observed in *Bos Taurus* (Koufariotis *et al.* 2014).

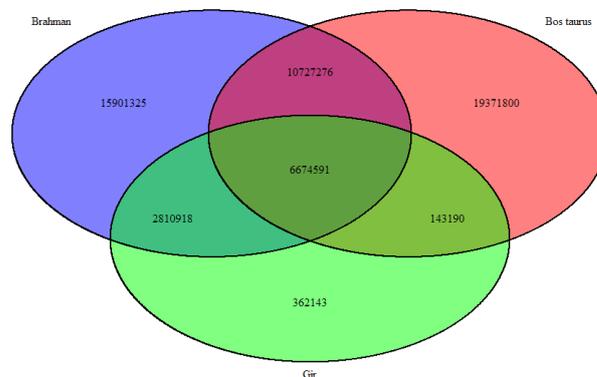


Figure 1. Number of SNP that are common or unique between the breeds

Figure 1 shows a Venn diagram of the overlap of SNP between the breeds. We see that there are more SNP in common between Brahman and Gir (95%) as opposed to Brahman and *Bos taurus* (48%). This is expected as Gir, which are a indicus breed, is one of the 4 ancestors to Brahman. Further 6,674,591 SNP in Brahman are found in both Gir and *Bos taurus*, which is a total of 18.5% of all Brahman SNP.

We examined the population structure between the Brahman cattle and taurine/Gir cattle to determine how much of the genome in Brahman is influenced, or infused with taurine and Gir. Overall, we find that Brahman and Gir cattle are the most similar (average F_{ST} over the whole genome is 0.19) compared with *Bos taurus* (average F_{ST} is 0.26) across the whole genome, as the average F_{ST} remains relatively low (is closer to 0).

However, we do observe some interesting findings on some chromosomes in where *Bos taurus*

Beef I

is more similar to Brahman than Gir, based on the F_{ST} window results. One example of this is on Chromosome 23 (Figure 2), in which we see a large region that is very similar to Brahman. This region is most likely the bovine lymphocyte antigen (BoLA) region. The BoLA regions is part of the MHC complex that display foreign peptides within a cell to cytotoxic T cells, triggering an immune response.

This raises the question, why is Brahman acquiring *Bos taurus* alleles in this region? One suggestion is that this could be due to MHC diversity in the alleles and their heterozygosities, which has been observed in many other vertebrate species (Salmier *et al.* 2016). It must be noted though, that this could also be due to a miss-assembly of that region in the UMD3.1, leading to these results.

Finally, we identified a region of *Bos taurus* introgression around the *PLAG1* gene that was previously described by (Fortes *et al.*, 2013). We find that the *Bos taurus* introgression in this region increases in frequency in younger animals, which could reflect a selection pressure on age at puberty.

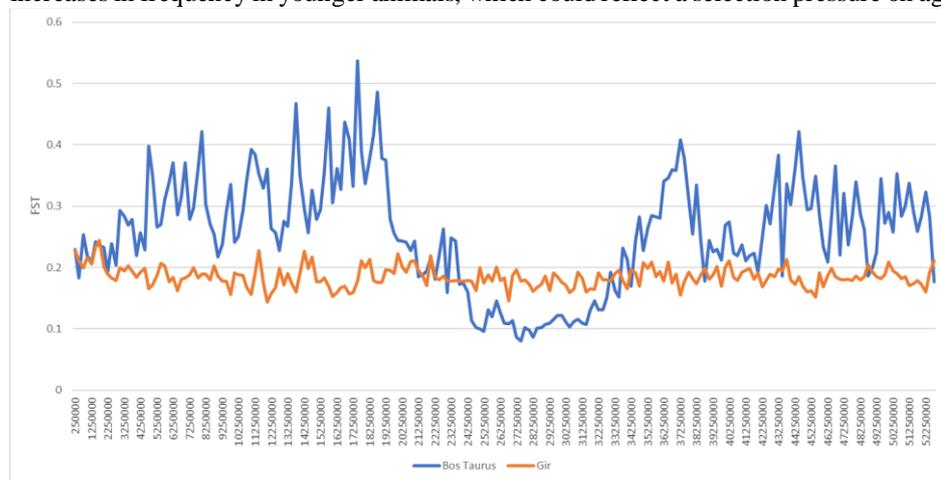


Figure 2: The average F_{ST} in each fixed window on chromosome 23 between Gir cattle (Orange line), and *Bos taurus* (blue line).

The next step in this project is to link genome variation amongst the bulls to variation in key traits such as fertility and meat quality and to examine if variants in certain genomic regions (such as coding regions, non-coding regions, regulatory regions) are more likely to influence complex beef trait variation.

REFERENCES

- Bolormaa, S. *et al.* 2011. *Journal of animal science* **89**: 2050-2060.
Daetwyler, H. D. *et al.* 2014. *Nature Genetics* **46**: 858-865.
DePristo, M. A. *et al.* 2011. *Nature Genetics* **43**: 491-498.
Druet, T., I. M. Macleod, and B. J. Hayes. 2014. *Heredity* **112**: 39-47.
Fortes, M. R. *et al.* 2013. *Animal Genetics* **44**: 636-647.
Koufariotis, L., Y. P. Chen, S. Bolormaa, and B. J. Hayes. 2014. *BMC Genomics* **15**: 436.
Liao, X. *et al.* 2013. *Genome* **56**: 592-598.
McKenna, A. *et al.* 2010. *Genome research* **20**: 1297-1303.
Salmier, A., B. de Thoisy, B. Crouau-Roy, V. Lacoste, and A. Lavergne. 2016. *BMC Evolutionary Biology* **16**: 229.
Su, J., Y. Wang, X. Xing, J. Liu, and Y. Zhang. 2014. *BMC genomics* **15**: 12.
Yates, A. *et al.* 2016. *Nucleic Acids Research* **44**: D710-716.