

# SCIENTIFIC REPORTS



OPEN

## Sequencing the mosaic genome of Brahman cattle identifies historic and recent introgression including polled

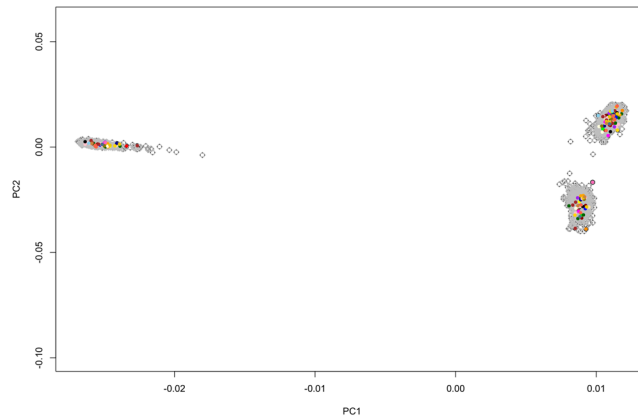
L. Koufariotis<sup>1</sup>, B. J. Hayes<sup>1</sup>, M. Kelly<sup>2</sup>, B. M. Burns<sup>3</sup>, R. Lyons<sup>4</sup>, P. Stothard<sup>5</sup>, A. J. Chamberlain<sup>6</sup> & S. Moore<sup>1</sup>

Brahman cattle have a *Bos indicus* and *Bos taurus* mosaic genome, as a result of the process used to create the breed (repeat backcrossing of *Bos taurus* females to *Bos indicus* bulls). With the aim of identifying *Bos taurus* segments in the Brahman genome at sequence level resolution, we sequenced the genomes of 46 influential Brahman bulls. Using 36 million variants identified in the sequences, we searched for regions close to fixation for *Bos indicus* or *Bos taurus* segments that were longer than expected by chance (from simulation of the breed formation history of Brahman cattle). Regions close to fixation for *Bos indicus* content were enriched for protein synthesis genes, while regions of higher *Bos taurus* content included genes of the G-protein coupled receptor family (including genes implicated in puberty, such as *THRS*). The region with the most extreme *Bos taurus* enrichment was on chromosome 14 surrounding *PLAG1*. The introgressed *Bos taurus* allele at *PLAG1* increases stature and the high frequency of the allele likely reflects strong selection for the trait. Finally, we provide evidence that the polled mutation in Brahmans, a desirable trait under very strong recent selection, is of Celtic origin and is introgressed from *Bos taurus*.

Brahman cattle are an interesting model for studying the impact of natural and artificial selection on the genome. Between 1854 and 1926 four *Bos indicus* cattle breeds (Ongole, Krishna, Gir and Gujarat) from India, that were highly adapted to tropical conditions, were imported into the USA. The Brahman breed was created by repeated crossbreeding of these animals with local *Bos taurus* cattle to generate large numbers and “grade up” to *Bos indicus*<sup>1</sup>. In 1933, Brahman bulls were first imported into Northern Australia, and the Brahman cattle were again “graded up” using local *Bos taurus* breeds. Brahman cattle therefore have a mosaic *Bos indicus* and *Bos taurus* genome, with around 10% of the genome of modern Australian Brahman cattle of *Bos taurus* origin<sup>2</sup>. Brahman cattle today are grazed for beef production in harsh conditions in Northern Australia, so are under strong selection for adaptation to these environments. *Bos indicus* and *Bos taurus* regions of the genome can be readily identified, so we can ask the question are there segments of the genome that are *Bos taurus* in origin and longer than expected by chance (given the history of the breed). This would be evidence for selection in these regions and could lead to identification of causative loci.

Whole genome sequencing should be a powerful method to identify such regions if they exist, provided enough individuals of the mosaic population are sequenced and sequence data is available on the contributing populations. This approach was recently used to provide evidence that Mongolian yak (*Bos grunnius*) has introgressions from *Bos taurus*, finding that Mongolian yaks have inherited 1.3% (on average) of their genome from bovine and the introgressed regions are enriched for genes associated with nervous system development and

<sup>1</sup>Centre of Animal Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Queensland, 4072, Australia. <sup>2</sup>Australian Agricultural Company (AACo), Brisbane, Queensland, 4006, Australia. <sup>3</sup>Department of Agriculture and Fisheries, Rockhampton, Queensland, 4702, Australia. <sup>4</sup>School of Veterinary Science, The University of Queensland, Gatton, Queensland, 4343, Australia. <sup>5</sup>Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, T6G 2C8, Canada. <sup>6</sup>Agriculture Victoria, Agribio, Centre for Agribiosciences, 5 Ring Road, Bundoora, Victoria, 3086, Australia. Correspondence and requests for materials should be addressed to L.K. (email: [r.koufariotis@uq.edu.au](mailto:r.koufariotis@uq.edu.au))



**Figure 1.** Principle component analysis of the genomic relationship between studs. Forty-six bulls selected for sequencing (in colour) relative to the diversity of the Brahman population represented by a large sample of genotyped animals with 24 K SNP (1021 animals in grey). PC1 and PC2 are the first and second principle components of genomic relationship matrix among the genotyped animals.

function (related to aggression)<sup>3</sup>. The authors also demonstrated that a polled mutation segregating in Mongolian yaks was *Bos taurus* in origin.

Our aim here was to identify *Bos taurus* segments in the Brahman genome at sequence level with whole genome sequence data and then to hypothesize if there are regions of the genome that have been under artificial or natural selection, and further some of these can be identified because they result in long segments that are *Bos taurus* in origin. While *Bos indicus* cattle are well adapted to harsh environments, they can have reduced production performance compared to *Bos taurus* cattle, including slower growth and reduced fertility, particularly later puberty and stronger expression of post-partum anoestrus<sup>4</sup>. This suggests an additional hypothesis that we test; that the Brahman genomes could be enriched for *Bos taurus* segments in regions affecting these traits, reflecting some selection during the grading up process.

To look for regions of complete or nearly complete *Bos indicus* or *Bos taurus* fixation in the Brahman population, we sequenced the genomes of 46 Brahman bulls used in Australia between 1953 and the present day. These bulls were selected because they were key ancestors of the modern Brahman population. *Bos indicus* and *Bos taurus* content was assessed in small fixed-size windows using a *bosind\_250* calculation across the genome using allele frequencies of the sequence derived SNP.

In this study, we found that alleles at genes in the G-coupled receptor family, immunity and a range of beef productivity traits are likely of *Bos taurus* origin<sup>5</sup>. As a demonstration of the power of the method, we also investigated the origin of polledness (lack of horns) in Brahmans, a trait which has recently become very valuable for welfare and safety reasons<sup>6</sup>. Using whole genome sequence from three homozygous polled Brahman cattle we found evidence that at least in some Brahman cattle the polled mutation is of *Bos taurus* origin, specifically the Celtic mutation<sup>7,8</sup>.

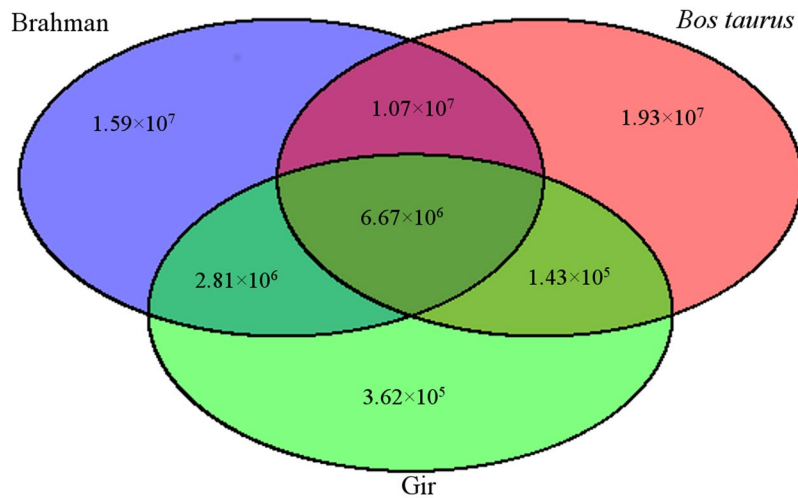
## Results and Discussion

Bulls were selected for sequencing using an algorithm to identify 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<sup>9</sup>. This selection method avoided double counting of ancestral genomes and considered whether DNA, extracted from semen straws or Ampules, was available for a bull or not<sup>10</sup> (Fig. 1). The 46 bulls selected captured 17% of the genetic variation represented in the pedigree. The selected bulls were sequenced on an Illumina HiSeq sequencer<sup>11</sup>, at an average of 12.5 times genome coverage, with a range of 10 to 30 times genome coverage.

The genome sequences were aligned to the UMD3.1 reference genome and variants were called using GATK<sup>12</sup>. After filtering (see Methods), a total of 36,114,110 SNP and 4,779,538 Indels remained. This represents a much higher rate of polymorphism than is observed in *Bos taurus* breeds<sup>10</sup> for the same number of animals sequenced. The higher rate of polymorphism in Brahmans is likely a reflection of a larger ancestral population size for *Bos indicus* than *Bos taurus* cattle (pre-domestication)<sup>13</sup>, the fact that four *Bos indicus* breeds were used to generate the breed, as well as the infusion of *Bos taurus* into Brahmans during breed formation.

Annotation of the variants with the NGS-SNP pipeline<sup>14</sup> indicated that the majority the variants detected were intergenic. Of the intragenic variants 0.35% were missense mutations and only a small number of stop gained mutations were annotated (see Supplementary Table S1).

**Shared variants between the cattle sub-species.** Of the Brahman foundation indicine breeds, whole genome sequence was available for Gir that came from fourteen Gir animals (11 were used to build pooled samples and 3 individual animals were sequenced individually)<sup>15</sup>. The Gir sequences available were taken as a representative of pure *Bos indicus*. However, we must acknowledge that this is a limitation in this analysis as we do not have other indicine breeds to build a more complete *Bos indicus* reference set of SNP. For *Bos taurus*, we used SNP allele frequencies from the 1000 bull genomes project, with 234 bull sequences of *Bos taurus* origin, that includes 129 Holstein-Friesians, 43 Fleckvieh and 15 Jersey breed cattle<sup>10</sup>.



**Figure 2.** Common variants between the cattle breeds. The blue circle represents the Brahman variants, red represents *Bos taurus* and green represents all Gir variants. We find that overall, most of the Gir variants are shared with Brahman breeds and only a small proportion of variants are common between all three groups.

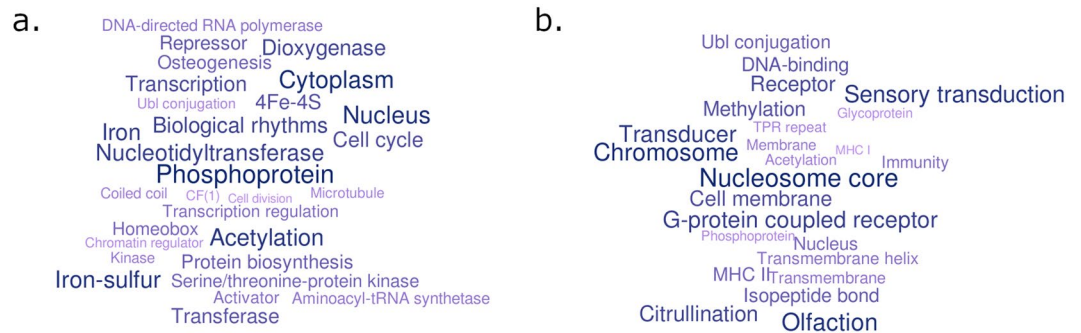
The number of Brahman variants that were in common with variants from *Bos taurus* and Gir were examined. Of the 36.1 million Brahman variants, 15.9 million variants were uniquely found in Brahman when compared with the other 2 datasets (Fig. 2). Most of these variants are likely to be from the other three indicine foundation breeds (Ongole, Gujarat and Krishna) and due to the limitation of using only one sequenced reference indicine breed. 10.7 million variants were shared exclusively between Brahman and *Bos taurus* (and not with Gir), meaning 48% of *Bos taurus* variants were found only in the Brahman genome, and not the Gir genome. There were only nine million SNP in the Gir dataset, reflecting the small number of Gir animals sequenced, however, 95% of variants from Gir were in common with the Brahman variants (expected as Gir was a founding breed). Overall a total of 6.7 million variants (18.5% of Brahman variants) were in common between all three datasets.

**Brahman bull genome regions high in indicine content.** To identify regions of the genome in each animal that were indicine or taurus in origin, we calculated *bosind\_250* values (Methods) using a formula that is based on the *b* values from Bolormaa *et al.*<sup>2</sup> *bosind\_250* statistic is an estimate of comparing two populations based on their sequence, allele frequencies and SNP calls to determine how differentiated the two populations are. *bosind\_250* values for each sequenced Brahman bull with both the Gir genomes and the *Bos taurus* genomes were calculated for SNP in 250 kb fixed-size windows (adjacent and non-overlapping). We used 250 kb windows as this meant enough SNP were in each window to allow for a robust estimate of the *bosind\_250* value. A low *bosind\_250* value in a 250 kb fixed-size window between Brahman and *Bos taurus* indicates that segment is of *Bos taurus* origin. A low *bosind\_250* value between Brahman and Gir indicates that the segment is of Gir (*Bos indicus*) origin.

The global *bosind\_250* value (across all SNP) for the Brahmans was substantially lower for *Bos indicus* (Gir, 0.188) than for *Bos taurus* (0.264), reflecting the high indicine content of Brahman genomes. There was substantial variation between the Brahman animals regarding the proportion of the genome that was indicine derived, indicating both global *bosind\_250* and the proportion of fixed-size windows that had a lower *bosind\_250* values with *Bos indicus* than *Bos taurus*. For example, sire 13 had an average *bosind\_250* value of 0.26 with *Bos taurus*, and 0.24 for *Bos indicus*, with the fixed-size window analysis related to almost 50/50 *indicus/taurus* composition. In contrast, sire 1 had a *bosind\_250* value of 0.27 for *Bos taurus* and 0.17 for *Bos indicus*, showing much higher *indicus* content. Averaged over all Brahman animals, 8.94% of the Brahman genome was taurine derived, which is slightly lower than the 10% previous estimate<sup>2</sup>, perhaps because we included several sires imported directly from the USA. Supplementary Fig. S1 shows each sequenced Brahman animal with its taurine/indicine percentage.

The difference in the percent of *bosind\_250* values for each 250 kb window between Brahman/*Bos indicus* and Brahman/*Bos taurus* was calculated and a heatmap was generated for all chromosomes. These heatmaps are presented in a data repository hosted by LabArchives LLC, (see Data availability). Overall, the heatmaps show that much of the Brahman genome is largely of *Bos indicus* origin with only a few segments showing strong *Bos taurus* introgression. Chromosomes 8, 12, 14, 23, 26 and 29 all show regions in the genome that are of strong *Bos taurus* introgression. This will be discussed below.

Fixed-size windows were sorted in ascending order based on the percent difference in the *bosind\_250* values between *Bos indicus* and Brahman and selected the top 5% windows that had the greatest percent difference (this was 368 fixed-size windows). To determine if these windows were enriched for gene pathways or annotations, all genes found in these fixed-size windows were extracted. This resulted in 1,609 unique genes in the regions of nearly fixed *Bos indicus* origin. Supplementary Table S2 provides the full list of the top 5% most significant fixed-size windows including all genes and gene information found within each fixed-size window. These genes were investigated for enrichment of function using the DAVID package<sup>16,17</sup>. Results from that analysis reveals there is significant enrichment for genes involved in intermediate filament, hormone, protein biosynthesis, cytoplasm and phosphoproteins (Fig. 3a). Protein biosynthesis (Fig. 3a) is interesting as there is speculation that



**Figure 3.** Word cloud showing the keywords after functional categorising genes using DAVID. Larger keywords represent those that show the strongest associations for that set of genes (based on the DAVID p-value). (a) Functional categorised genes that are in genomic segments that are homozygous or nearly homozygous in the sequenced bulls and of *Bos indicus* origin (b). Functional categorised genes located in genomic segments that show strong *Bos taurus* introgression.

Brahmans are more able to cope with harsh conditions because of lower protein turnover. We found 61 genes to be enriched for protein biosynthesis, including, but not limited to, G elongation factor mitochondrial 1 (*GFM1*), *GUF1* homolog, *GTPase* (*GUF1*) and many eukaryotic translation elongation factors. Eukaryotic translation elongation factors and eukaryotic translation initiation factors are key regulators of protein synthesis<sup>18</sup>, protein bond formation<sup>19</sup>, delivery of tRNAs to translating ribosomes<sup>20</sup>, and proliferation (mainly *eEF2*) during skeletal muscle contraction<sup>21</sup>. To determine if the level of enrichment for the above functions could occur by chance, we performed a permutation test by randomly selecting the same number of genes (1,609 genes) as those found in the top 5% most significant *Bos indicus* regions. We performed this random selection process a total of 100 times and functional annotation of each permutation was performed with DAVID. The keywords for each permutation were recorded and grouped together to create a word cloud plot of the most commonly occurring keywords. Over the 100 randomly selected gene permutations, protein biosynthesis, 4Fe-4S nucleotidyltransferase and biological rhythms were never found in the 100 random gene permutations gene set.

Individual regions that were enriched for *Bos indicus* content include chromosome 21 (see Data Availability), at positions 7.5–10 Mb with a 6-fold difference in *bosind\_250* values between *Bos taurus* and *Bos indicus* (*bosind\_250* value 0.67 with *Bos taurus* vs 0.11 with *Bos indicus*). One gene of interest, melanoma antigen family L2 (*MAGEL2*), is implicated in fertility and evidence indicates the gene may influence testicular size, fertility and growth<sup>22</sup>. *MAGEL2* is also involved in circadian rhythms and adaptation to new environments (in mouse knockout studies)<sup>23</sup>.

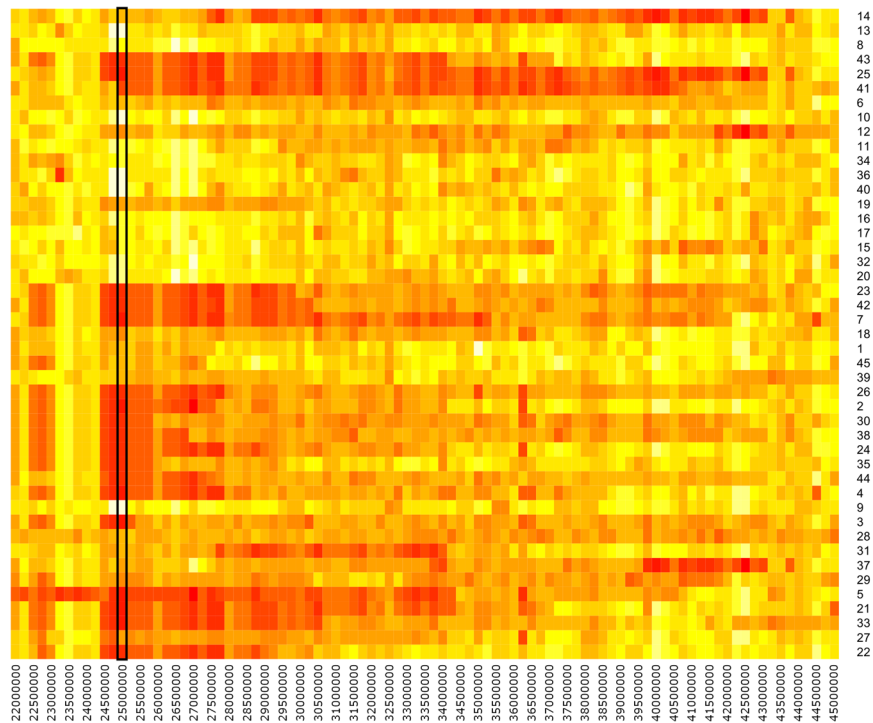
The region between 53–55 Mb on chromosome 8 is of strong *Bos indicus* origin and contains the genes *PRUNE2*, *GNA14*, *GNAQ*, *CEP78* and *PSAT1*. In Simmental cattle, an association analysis with foreshank weight found a region that includes the genes *GNAQ* and *CEP78*<sup>24</sup>.

On chromosome 14 at positions 24,250,000–24,500,000 bp, the XK, Kell blood group complex subunit-related family, member 4 (*XKR4*) gene is found approximately 396 kb upstream of the *PLAG1* gene. The gene is of interest as it has been observed to be associated with subcutaneous rump fat thickness<sup>25</sup>. In addition, polymorphisms in this region have been found to be associated with rump fat thickness, residual feed intake, average daily feed intake and gain in cattle<sup>25</sup>.

**Regions of the genome that are of *Bos taurus* origin.** While *Bos indicus* cattle are well adapted to harsh environments, they can have reduced production performance compared with *Bos taurus* cattle, including slower growth, and reduced fertility (particularly later puberty and stronger expression of *post-partum anoestrus*)<sup>26</sup>. We speculate that the Brahman genome could be enriched for *Bos taurus* segments in regions affecting these traits, reflecting some artificial selection during the grading up process<sup>27,28</sup>. Formation of the Brahman breed involved two different “grading-up” stages, the first was in USA and the second was in Australia. This predominately involved crossbreeding female *Bos taurus* animals with Brahman bulls. This is reflected by the introgression of large segments of *Bos taurus* in the Brahman genome on chromosome X (see Data availability, chromosome X heatmap).

As was done when looking at genes found in regions of significant *Bos indicus* origin, the fixed-size windows were sorted based on the percent difference of the *bosind\_250* values in ascending order between *Bos taurus* and Brahman. The top 5% fixed were selected (253 fixed-size windows) and all genes were extracted from these windows. Supplementary Table S2 provides the full list of the top 5% most significant fixed-size windows including all genes and gene information found within each fixed-size window. This resulted in a total of 892 genes that are found in regions with significant *Bos taurus* introgression and used for functional annotation with the DAVID package<sup>16,17</sup>. The results show enrichment for genes involved with olfactory, major histocompatibility complex (*MHC*) I and II, nucleosome core and the G-coupled protein receptor (*GPR*) family (Fig. 3b). As a control, the same random gene selection permutation test was performed as mentioned previously, in which 892 genes randomly selected over a total of 100 permutations and each permutation was functionally annotated with DAVID. Over the 100 random permutations, we found no enrichment for genes associated with *GPR*, olfactory, immune system and chromatin binding (see Data availability). The *GPR* family is of interest because some genes from *Bos*





**Figure 4.** Heatmap of the Chromosome 14 region with high *Bos taurus* introgression. This includes the *PLAG1* gene found at positions 25,007,291–25,009,296 as indicated by the black box. On the y-axis are all the sequenced animals, ordered by year of birth (with older animals at the top). On the x-axis is the genomic positions, incremented in 250 kb windows. A red colour indicates more *Bos taurus* introgression, a white colour indicates more regions that are of *Bos indicus* origin.

*taurus* introgressed regions that are enriched for this family of proteins, are associated with puberty. Thyrotropin receptor (*TSHR*) is one example, as SNP found in this gene are associated with fertility traits in cattle<sup>5</sup> and *TSHR* was identified as a putative domestication gene in chickens<sup>29</sup>. Other genes of interest that are found to be enriched in the *GPR* family include, but are not limited to, *ADGRL3*, *CELSR2*, *QRFPR*, *RRH*, *VIPR1* and the *TAAR* family.

A Gene Ontology (GO) analysis looking at the biological process (BP) found that of the genes in *Bos taurus* introgressed windows, there is a trend for associations with processes involved in immunity, nucleosome assembly, *GPR* family and sensory perception of smell.

**Chromosome 14, *PLAG1* gene and surrounding region.** The most significant *Bos taurus* introgression observed was an extensive genomic region on chromosome 14 around 2.2–4.2 Mb (Fig. 4). This region contains several widely known and well described genes, including the zinc finger protein Pleomorphic adenoma gene 1 (*PLAG1*) gene. This region is of interest as previous GWAS studies have found significant associations between the *PLAG1* gene, including surrounding genes, to have important associations with stature<sup>30</sup>, fertility<sup>26</sup> and beef production traits, such as carcass weight<sup>26,31–33</sup>. In the study by Fortes *et al.*<sup>26</sup>, there are alleles in this region that have been heavily selected for in Brahmans due to their associations with stature, growth and carcass traits, despite having a negative effect on fertility (later puberty)<sup>26</sup>.

A simulation analysis was performed (Methods) to determine if this large *Bos taurus* introgressed region around the *PLAG1* gene occurred randomly or was the result of selection. The simulation study was performed using the tool *SFS\_CODE*<sup>34</sup> where two ancestral populations of *Bos taurus* and *Bos indicus* animals were simulated. We then simulated a cross-breeding event between the two ancestral breeds to form a simulated Brahman breed. We calculated the *FST* using the Weir & Cockerham calculation from *vcftools* for the simulated and real datasets (Methods). *FST* between the simulated Brahman and *Bos taurus* populations closely matched the *FST* between the real Brahman and *Bos taurus* populations (0.25 and 0.26 respectively). In Methods, the simulation analysis is described in more detail. Supplementary Fig. S2 shows a plot comparing the frequency of the *Bos taurus* introgressed segments between simulated and real Brahman animals. In the simulated Brahman data, we find a higher frequency of single fixed-size windows (<250,000 bp) that are of *Bos taurus* origin with a lower frequency of fixed-size windows of *Bos taurus* origin that are larger. Meanwhile, in the real Brahman data, we observe a higher frequency of *Bos taurus* introgressed segments that extend two or more fixed-size windows (>250,000 bp) (see Supplementary Fig. S2). In the case of the large *Bos taurus* introgressed segment surrounding the *PLAG1* gene, there were no segments of *Bos taurus* introgression as long in the simulated Brahman dataset. This provides some evidence that the large *Bos taurus* introgressed segment on chromosome 14 at the *PLAG1* gene is likely the result of artificial selection and not due to random drift.

Other genes of interest in this region include *LYN*<sup>35,36</sup>, *MOS*<sup>35,36</sup>, *CHCHD7*<sup>37</sup>, *PENK*<sup>35</sup>, and *SDR16C5*<sup>35</sup>, which have been proposed in various studies as associated with milk production, reproduction, muscle formation, carcass traits and body size.

We observed that 26 “older animals” (those born between 1953 to 1989) are less likely to have the *Bos taurus* introgression in this region, while 20 of the “younger” animals (those born between 1990 and 2005) carry the *Bos taurus* introgression in this region (Fig. 4).

The *XKR4* gene as discussed previously is found in this region, just 396 kb upstream of the *PLAG1* gene (Fig. 4) and associated with rump fat thickness, residual feed intake, average daily feed intake and gain in cattle<sup>25</sup>.

**Additional genomic regions with *Bos taurus* introgression.** The second half of chromosome 8 shows several regions with an increased frequency of *Bos taurus* ancestry (see Data availability). The positions around 75,000,000–81,500,000 bp display some of the most substantial *Bos taurus* introgression on the chromosome. The study by Bolormaa *et al.*<sup>2</sup> found that the region around 81.4 Mb on chromosome 8 has a high level of *Bos taurus* ancestry compared to other regions on the chromosome. Isolating the genes found in these positions, a DAVID<sup>16</sup> analysis reveals that the genes are enriched for associations with carbohydrate transmembrane transport, renal water absorption and urea transmembrane transport. One of these genes,  $\beta$ 1,4-galactosyltransferase I (*B4GALT1*), encodes an enzyme that adds glycoprotein glycans to oligosaccharides<sup>38</sup>, and in bovine was found to be associated with milk synthesis and secretion in mammary glands. Further, downregulation of the gene is involved in reduced milking frequency and volume<sup>39</sup>. Possibly better milking ability associated with these alleles means higher weaning rates and therefore increased calf survival.

The genes calpastatin (*CAST*) and heat shock protein beta 1 (*HSPB1*) are both located in regions that exhibit substantial *Bos taurus* introgression on chromosome 7 and 25, respectively; *CAST* is a key gene regulating the rate of protein turnover<sup>40,41</sup> and plays a role in meat quality<sup>42</sup>.

Chromosome 12 and 23 (around the *BoLA* region) show relatively substantial *Bos taurus* introgression at positions 69,000,000–79,000,000 and 22,750,000–34,000,000, respectively. While immune related genes are of great interest in adaptation, this apparent high level of *Bos taurus* introgression could possibly be a cause of assembly errors in the reference genome or due to a high number of copy number variants observed in this region (CNV)<sup>43</sup>.

**Fixed non-reference alleles in the Brahman genome.** If a variant in Brahman cattle and *Bos taurus* is fixed for the alternative allele, it may indicate that this variant has been under strong selection because they confer an advantage for adaptation to harsh environments (although this can happen through drift as well, and this analysis does use Brahman as a proxy for *Bos indicus*). These mutations can occur for at least two reasons. Firstly, a mutation arose in the ancestral animals (before the split between *Bos taurus* and *Bos indicus*) and one allele was fixed in one or more of the four *indicus* foundation breeds while the other allele was fixed in *Bos taurus* due to either genetic drift or natural or (less likely) artificial selection. Secondly, mutations may have occurred spontaneously in the Brahman population following which heavy selection has occurred. There were 20,917 SNPs fixed for alternative alleles in the Brahman and *Bos taurus* (1000 bull genomes<sup>10</sup>) genomes.

Annotation of fixed variants show relatively similar proportions to the annotations of all genomic Brahman variants (67.32% intergenic and 25.35% intronic). However, we find an increased proportion of missense mutations in SNP that are fixed for the alternative allele in Brahmans versus *Bos taurus*. 0.86% of Brahman SNP that are fixed for the alternative allele are annotated as missense, this value varies from the proportion of Brahman genomic SNPs annotated as missense, which is 0.35% (see Table 1 and Supplementary Table S1). This difference in the proportion of missense annotated SNP represents an 85% increase in the percent difference between fixed variants for the alternative allele that are missense, compared to all missense variants in the WGS dataset. A chi squares analysis reveals that this difference is significant, with a chi-square score of 29.629 and a *p*-value of 0.009 making the difference significant at *p* < 0.05.

Some genes with missense mutations fixed in Brahmans for the alternative allele stand out, including *IRGQ*, *KIR3DS1*, *ATG9B*, *ADGRB2*, *PRAME*, *CMPK2*, *CLEC4G* which are all related to immunity. *ATG9B* is one of the main genes involved in autophagy and has been found to be associated with initiating autophagy<sup>44</sup>. *CMPK2* is believed to be associated with macrophage terminal differentiation and was recently found to be a candidate gene for association with lower weight gain and increased viraemia levels in response to porcine reproductive and respiratory syndrome virus infection while showing evidence of allele specific expression in pigs<sup>45</sup>. These findings suggest mutations associated with immunity have become fixed in the Brahman breed, and could be one reason why the Brahman breed is so resilient and resistant to a range of pathogens that affect *Bos taurus* breeds.

The full list of SNP can be found in Supplementary Table S3.

**Brahman Polled mutation.** Four additional Brahmans that were polled (do not grow horns), identified through progeny testing, were sequenced. One animal (47) was heterozygous for the polled allele (*Pp*) and the other three animals (48, 49, 50) were homozygous for the polled allele (*PP*). The three *PP* animals were compared to the 46 Brahmans (as used in this study), in which we refer to as the 46 original animals. The 46 original animals were horned (*pp*) animals (although there is a remote chance that an animal might be *Pp* and was not recorded as such).

The polled allele has been located on chromosome 1 in the positions 1.6–2.2 Mb in cattle<sup>46,47</sup> and also in composite cattle<sup>48</sup>. Three putative mutations have been reported. The first polled mutation is an 80 kb duplication in the region on 1.91–1.99 Mb that is found in Friesian cattle<sup>8</sup>. The second polled mutation is found starting at the position 1.71 Mb and is believed to be of Celtic origin, described as a 212 bp duplication in polled animals that replaces a 10 bp sequence at the positions 1,706,051–1,706,060 bp<sup>7</sup>. This mutation is referred to as the Celtic mutation. The third polled mutation originates from Mongolian Turano cattle and was recently found to be introgressed in Mongolian Yaks<sup>3</sup>. As with the other two polled mutations, this mutation is found on

Annotation	No. of SNP	Percent of total	Percent difference*
3_prime_UTR_variant	56	0.268	34.074
5_prime_UTR_variant	23	0.11	87.260
coding_sequence_variant	10	0.048	196.387
downstream_gene_variant	1,200	5.737	68.453
intergenic_variant	12,066	57.685	-15.416
intron_variant	5,775	27.609	8.543
missense_variant	180	0.861	85.524
non_coding_transcript_variant	9	0.043	39.074
splice_acceptor_variant	3	0.014	157.820
splice_donor_variant	3	0.014	157.125
splice_region_variant	44	0.21	104.743
stop_gained	3	0.014	94.681
synonymous_variant	83	0.397	-41.754
upstream_gene_variant	1,429	6.832	71.557
Unknown	33	0.158	195.742
Total	20,917	100	

**Table 1.** Annotation information of the variants that are fixed in Brahmans for the alternative allele. The percent change with all variants, represents the different (represented as a percentage) between the number of variants found in each functional class that are fixed, and the total number of variants found in each class. \*Percentage difference indicates the difference in the percent between the fixed variants in a class with the total number of variants found in that class (see Supplementary Table S1).

chromosome 1 between positions 1,889,854 bp and 2,010,574 bp. The architecture of this mutation is an 11-bp motif (conserved among Bovidae and duplicated in the Friesian mutation) due to a rearrangement of a 219 bp duplication-insertion, a 7 bp deletion and a 6 bp insertion<sup>3</sup>. This third polled mutation was not observed to be segregating in the Brahman animals, and it is unlikely to be segregating as it has only ever been described in Mongolian Turano cattle and Mongolian yaks.

In Brahmans, a single study has mapped the polled locus in 68 genotyped Brahman cattle to the same region as in *Bos taurus* cattle<sup>49</sup>, yet the origin of the Brahman polled mutation remains unknown.

Examining the fixed-size windows in the polled locus region, shows slight introgression of *Bos taurus* (see data availability, Chromosome 1). Using our sequence data, the total number of reads aligning to the UMD 3.1 reference were counted in 50 bp incrementing windows in and around the polled region and the standard deviation was calculated for each of the 50 bp window (see Methods). We examined the coverage of the reads because if an increase in the coverage occurs, it can indicate a duplication mutation. In this study, the region around 1.65 Mb and 1.90 Mb stands out as an increase in coverage for the three *PP* animals compared to the 46 original Brahman bulls, indicating the presence of structural variants, such as duplications, in the *PP* animals.

In the positions where the Friesian mutation is on chromosome 1, we find no distinct difference or patterns in read coverage across the entire 80 Kb region between the three *PP* animals and the 46 original animals (Fig. 5). If the Friesian 80 Kb duplication was present in Brahman, we would expect to find at least a two-fold increase in the coverage where the mutation exists between the homozygous polled and the 46 original Brahmans. Thus, in this study, we did not find evidence that the Friesian mutation is segregating in the three *PP* polled Brahman animals.

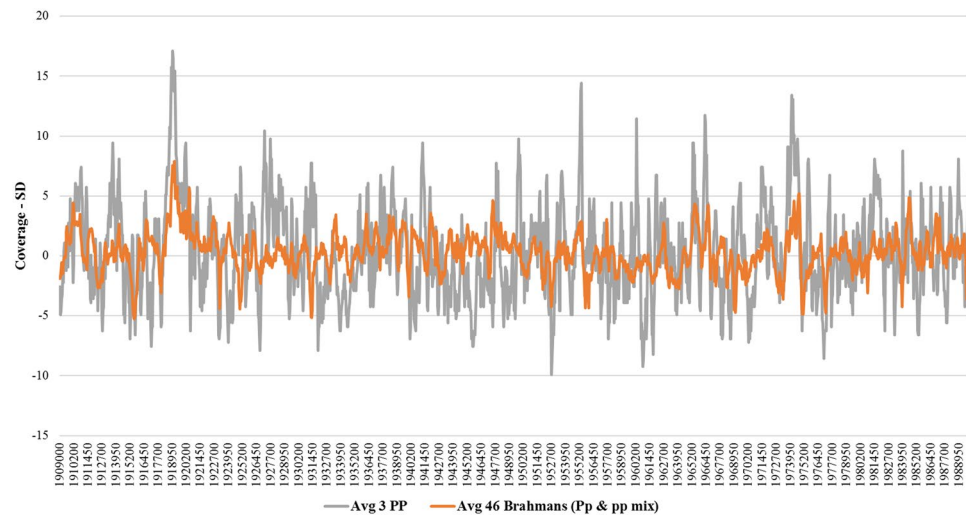
At the positions on chromosome 1 where the Celtic mutation is predicted to be<sup>7</sup>, we find that the 3 *PP* animals show an increase in coverage, when compared to the 46 original Brahman (Fig. 6). In Fig. 6, we see that a duplication is possibly found at this region, around where the actual Celtic mutation is described, as the coverage for the *PP* animals is increased significantly.

Furthermore, the BAM files containing the mapped reads for chromosome 1 were visualized using the Integrative Genomics Viewer (IGV) tool concentrating on the region where the Celtic mutation is found. We find in the three *PP* animals, there is a region between the positions of 1,706,045–1,706,060 bp which shows a 10 bp deletion event, consistent with the Celtic mutation. This deletion can be unambiguously seen in two of the polled animals (48 and 50) (see Supplementary Fig. S3), with minimal reads mapping to this region, especially for these two polled animals. Furthermore, the *PP* animal 49 seems to be carrying some variant of the polled mutation, as few reads map to the 10 bp deletion. This can possibly indicate that while this animal is homozygous for the polled mutation, the mutation affecting polledness might not be at the 10 bp region where the deletion is located. This needs to be followed up with more research and will provide the basis for future work.

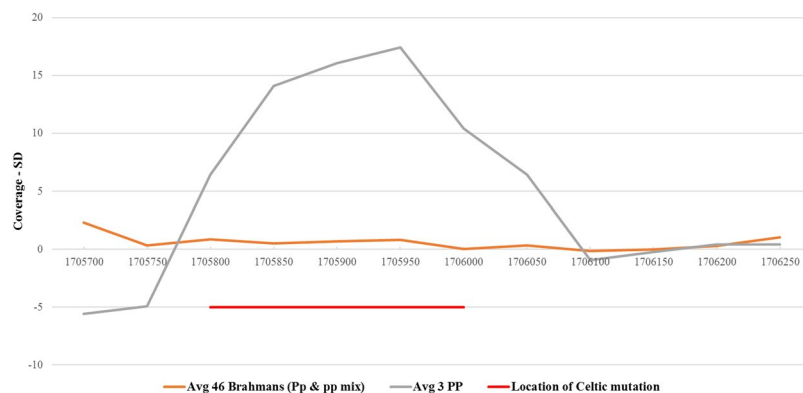
In the 46 original animals that had their genomes sequenced (Data Repository), most do not display this deletion, indicating *pp* (horned) animals with a small chance of *Pp* animals.

## Conclusion

There has been clear introgression of *Bos taurus* alleles into Brahman in several regions, with the largest region on Chromosome 14 in the *PLAG1* region. This introgression could have been driven by selection for increased stature. The regions of the Brahman genomes with near fixation of indicine alleles were enriched for genes associated with protein synthesis, which could reflect selection for reduced protein turnover to cope with harsh



**Figure 5.** Holstein-Friesian 80 Kb polled variation. Figure showing the read coverage (expressed as standard deviations from the mean) between the three homozygous polled Brahms and the 46 original Brahman animals in the region where the 80Kb Holstein-Friesian polled duplication is found.



**Figure 6.** Plot of the number of reads mapping to the reference homed in on where the Celtic mutation has been described. In the 3 PP animals, we see an increase in the coverage, characteristic of a duplication event, while the 46 original animals show consistent coverage. Here the orange line represents the standard deviation in the coverage for the original 46 animals. The grey line represents the standard deviation in the coverage for the 3 PP animals. The red line represents the location of the Celtic mutation as described by Wiedemar *et al.*<sup>7</sup>

environments. However, in general, we found relatively few regions completely fixed for *Bos indicus* content, which suggests adaptation might be a highly polygenic trait (affected by a large number of loci of small effects). It also should be pointed out that because the Brahman genome is largely *Bos indicus* in origin, we had much lower power to differentiate genomes regions enriched for *Bos indicus* content by chance *versus* enriched by selection. Finally, we provide evidence to suggest that the polled mutation in the Brahman cattle of our study, at least in the individuals we had, is a result of the Celtic mutation rather than the Holstein-Friesian mutation.

## Methods

**Animal ethics statement.** DNA was extracted from commercially collected semen straws or Ampoules that have previously been collected for commercial purposes and these were purchased or donated to the project.

**Sequence data.** Sequenced bulls were selected using an algorithm that identified 46 bulls that captured the highest proportion of genetic variation in the breed. The algorithm is based on an analysis of an extensive Brahman pedigree and a stepwise regression procedure to avoid double counting of ancestral genomes. Furthermore, the algorithm considered whether DNA, extracted from commercially collected semen straws or Ampoules, was available for a bull or not<sup>9,10</sup>. Three additional bulls that were known to be homozygous for the polled phenotype were also analysed in this study.

The selected bulls were sequenced on an Illumina HiSeq sequencer, at an average of 12.5 times genome coverage and a range of 10–30 times genome coverage. Data was trimmed and filtered using the tool Trimmomatic<sup>50</sup>, merging singleton reads (reads where one of the matching paired ends passed quality filtering) into a single file. We filtered for reads that had a minimum length of 36 and an average quality of 20.



**Variant calling.** After quality control and trimming, reads (including the singleton reads) were mapped to the bovine genome (UMD3.1) with BWA MEM<sup>51</sup>. All SAM files were converted to BAM format, sorted based on their coordinates and duplicates were marked using the Picard tools Sortsam, and MarkDuplicates<sup>52</sup>. The GATK<sup>12</sup> tool IndelRealigner was used to perform a local realignment of the reads around Indels and the Picard tools MarkDuplicates was additionally used to remove all duplicates in the final BAM files. BAM index files were generated using the Picard tool BuildBamIndex.

Variant discovery was performed using GATK following the GATK Best Practises manual<sup>12</sup>. The GATK<sup>12</sup> tool HaplotypeCaller was used to call all raw variants using the *Bos taurus* reference genome UMD3.1 and all raw variant VCF files were joined via the GenotypeGVCF tool to produce a single VCF file. The variants include single nucleotide polymorphisms (SNP) and small insertion/deletions (Indels). SNP and Indels were filtered using hard filters with thresholds that were based on what is recommended by GATK<sup>53</sup> for removal of variants with poor quality scores. VQSR was not used in this study for filtering of variants.

SNP were filtered with the following annotations and thresholds:

$$QD < 2.0 || FS > 60.0 || MQ < 40.0 || ReadPosRankSum < -8.0 || MQRankSum < -12.5 || SOR > 3.0$$

Descriptions of each filter can be found in the GATK Best Practises manual<sup>12</sup>. In brief, QD is the quality by depth which is the quality score normalized by allele depth. FS is the Fisher Strand test to detect strand bias, lower scores indicate more strand bias. MQ is the root mean square of the mapping quality score. ReadPosRankSum is the ranked sum test for the distance of alleles from the end of the reads, as the closer the variant is to the end of the read, the more error prone it is. MQRankSum looks at the mapping qualities of reads that support the reference allele with those that support the alternative allele and SOR measures strand bias. InbreedingCoeff measures the level of inbreeding in a group of samples by estimating the population allele frequency from the sample genotypes.

The threshold to hard filter Indels as used in this study is:

$$QD < 4.0 || FS > 100.0 || InbreedingCoeff < -0.8 || ReadPosRankSum < -10.0 || SOR > 6.0$$

**bosind\_250 calculation.** BTA and Gir variants were obtained from the 1000 bulls genomes project<sup>10</sup> and from the study by Liao *et al.*<sup>15</sup>, respectively. Only SNP data from each study was used and allele frequencies were calculated using in-house scripts. Following this, SNP found in common between *Bos taurus* and Brahman and between Gir and Brahman were determined, selected and placed into separate files. This was done with a python script that selected SNP that shared the same genomic location.

To determine introgression of *Bos taurus* and Gir genomic DNA in Brahman, a similar method was inspired by the calculation as described in Bolormaa *et al.*<sup>2</sup>, with the difference being our formula uses 250 kb windows. Our method involved calculating *bosind\_250* values, where the *bosind\_250* statistic estimates how differentiated two populations are based on sequence, allele frequency and SNP calls. In this analysis *bosind\_250* values were calculated between *Bos taurus* and *Bos indicus* SNP that are found in the Brahman animals using the following formula:

$$bosind_{250} = \frac{Ht - Hs}{Ht} \quad (1)$$

where:

$$Hs = PBT(1 - PBT) + Pbrai(1 - Pbrai) \quad (2)$$

and

$$Ht = \frac{2 * (PBT + Pbrai)}{2} \times 1 - \frac{PBT + Pbrai}{2} \quad (3)$$

*PBT* is the SNP allele frequency of the alternative allele in either *Bos taurus* or *Bos indicus* and *Pbrai* is the SNP call in that Brahman SNP individual. *Pbrai* is 0 if the SNP call is homozygous for the reference, 0.5 if the SNP call is heterozygous and 1 if the SNP call is homozygous for the alternative allele.

The calculation was performed twice for each Brahman animal across all SNP, the first calculation used common SNP between Brahman and *Bos indicus*, the second calculation used common SNP between Brahman and *Bos taurus*.

All SNPs were grouped into windows of 250 kilobases (kb) and the average *bosind\_250* values (*bosind\_250avg*) were calculated for each window. We used 250 kb windows as this provides enough SNP from sequence data in each window to perform this analysis and obtained robust estimates of the *bosind\_250* value. This was done by adding the *bosind\_250* values across all SNP within a 250 kb window and dividing that number by the total number of SNP found in that window as shown in the following formula:

$$bosind_{250avg} = \frac{\sum(bosind_{250fw})}{n} \quad (4)$$

where *bosind\_250fw* are the *bosind\_250* values for all SNP found in a fixed-size window and *n* is the total number of SNP found in that fixed-size window.

**SNP Annotation.** Annotation of the Brahman SNP was carried out using the NGS-SNP tool<sup>14</sup>. CpG Isles annotations were from the study by Su *et al.*<sup>54</sup>. Micro RNA (miRNA) target site annotations were from the Microcosm Database<sup>55</sup>. Long noncoding RNA annotations were from the paper by Koufariotis *et al.*<sup>56</sup>.

**Simulation study.** A simulation study was performed to investigate if large *Bos taurus* introgressed segments in the Brahman genome (such as *Bos taurus* introgression found around the *PLAG1* gene on chromosome 14) could be due to chance, given the history of the breed, or are due to selection (e.g. very unlikely to have occurred by chance).

To perform this analysis, we first simulated an ancestral founder population using the simulation software tool SFS\_CODE<sup>34</sup>. After 100,000 generations we split the ancestral population into two separate populations. The first population was a simulated *Bos taurus* group and the second population was a simulated *Bos indicus* group. After 10,000 generations, from the simulated founder *Bos taurus* and *Bos indicus* populations, we formed a Brahman population by simulating a cross-breeding event between the *Bos taurus* and the *Bos indicus* populations. The cross-breeding event simulated the “grading up process” of Brahman cattle. After 30 generations, the final Brahman population was simulated with an FST that closely matches the FST in the real Brahman animals. No selection was assumed in the formation of the simulated Brahman breed and the simulation of the founder *Bos taurus* and *Bos indicus* breeds.

For the tool SFS CODE, the following parameters were used: A low mutation and recombination rate of 0.01 per each locus of 10,000 kb. The population size for the ancestral *Bos taurus* and *Bos indicus* breeds was set at 10,000. The Brahman population was simulated to occur after 30 generations of crossbreeding the simulated *Bos taurus* and *Bos indicus* populations. This resulted in a total of 1,000 Brahman animals.

We calculated the FST between the simulated Brahman and the simulated *Bos taurus* animals using VCFtools<sup>57</sup> with the following command:

```
vcftools--vcf simulated_brahman. vcf--weir-fst
-pop simulatedBrahamPopulation.txt--weir-fst
-pop simulatedBos TaurusPopulation.txt--out brahman_vs_bosTau_fst
```

The FST between the simulated Brahman and *Bos taurus* animals was 0.25 which closely matches the 0.26 FST between the real Brahman and *Bos taurus* animals. Both FST values were obtained using the Weir and Cockerham calculation from vcftools. The level of differentiation was achieved by tuning the generations of separation between *Bos taurus* and *Bos indicus*.

The genotypes from the vcf files for animals in both the Brahman and *Bos taurus* simulated populations were used to determine the *Bos taurus* introgression in the simulated Brahman breed, as was described earlier in Methods.

Finally, to examine the chance of a long *Bos taurus* introgressed segment (that spans more than one fixed-size window of *Bos taurus* origin), we calculated the length of *Bos taurus* introgressed chromosomal segments in the Brahman genome. This was done by calculating consecutive fixed-size windows with *Bos taurus* introgression across the Brahman genome, both for the real and simulated Brahman data. Next, the frequency of each *Bos taurus* introgressed segments was determined in both the simulated and real Brahman animals. Supplementary Fig. S2 shows a plot comparing the frequency of the *Bos taurus* introgressed segments between the simulated and real Brahman datasets.

**Fixed alleles.** Using in-house scripts, we identified the SNPs in which every Brahman animal was homozygous for the alternative allele. In the case of the VCF files, when all animals had (1/1). The SNP were compared to *Bos taurus* SNP from the 1000 bulls genome project<sup>10</sup>. Only the SNP that were unique to Brahman and not found in the *Bos taurus* dataset were kept. These SNP are said to be fixed in the Brahman and not in *Bos taurus*. As annotation was carried out with the NGS-SNP tool<sup>14</sup>, we identified the SNP that had a missense mutation annotation.

## Data Availability

The datasets generation during analysis for the current study are available at NCBI BioProject database with project number: PRJNA432125. This includes all the BAM files for the 46 sequenced bulls and the three additional bulls that are known homozygous polled. Heatmaps for all 29 autosomal chromosomes that show the *Bos taurus* and *Bos indicus* introgression are hosted by LabArchives, LLC (<http://www.labarchives.com/>) with <https://doi.org/10.6070/H4RJ4H2D>. All files are in pdf format

## References

1. Briggs, H. M. & Briggs, D. M. *Modern breeds of livestock*. (Macmillan, 1980).
2. Bolormaa, S. *et al.* Detection of chromosome segments of zebu and taurine origin and their effect on beef production and growth. *Journal of animal science* **89**, 2050–2060, <https://doi.org/10.2527/jas.2010-3363> (2011).
3. Medugorac, I. *et al.* Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nature genetics* **49**, 470–475, <https://doi.org/10.1038/ng.3775> (2017).
4. Burrow, H. M. Importance of adaptation and genotype x environment interactions in tropical beef breeding systems. *Animal: an international journal of animal bioscience* **6**, 729–740, <https://doi.org/10.1017/s175173111200002x> (2012).
5. Dias, M. M. *et al.* SNP detection using RNA-sequences of candidate genes associated with puberty in cattle. *Genetics and molecular research: GMR* **16**, <https://doi.org/10.4238/gmr16019522> (2017).
6. Hotzel, M. J. & Sneddon, J. N. The role of extensionists in Santa Catarina, Brazil, in the adoption and rejection of providing pain relief to calves for dehorning. *Journal of dairy science* **96**, 1535–1548, <https://doi.org/10.3168/jds.2012-5780> (2013).
7. Wiedemar, N. *et al.* Independent Polled Mutations Leading to Complex Gene Expression Differences in Cattle. *PLoS one* **9**, e93435, <https://doi.org/10.1371/journal.pone.0093435> (2014).
8. Medugorac, I. *et al.* Bovine Polledness – An Autosomal Dominant Trait with Allelic Heterogeneity. *PLoS one* **7**, e39477, <https://doi.org/10.1371/journal.pone.0039477> (2012).

9. Druet, T., Macleod, I. M. & Hayes, B. J. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* **112**, 39–47, <https://doi.org/10.1038/hdy.2013.13> (2014).
10. Daetwyler, H. D. *et al.* Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature genetics* **46**, 858–865, <https://doi.org/10.1038/ng.3034> (2014).
11. Shen, R. *et al.* High-throughput SNP genotyping on universal bead arrays. *Mutation research* **573**, 70–82, <https://doi.org/10.1016/j.mrfmmm.2004.07.022> (2005).
12. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297–1303, <https://doi.org/10.1101/gr.107524.110> (2010).
13. Tellam, R. L. *et al.* Unlocking the bovine genome. *BMC genomics* **10**, 193, <https://doi.org/10.1186/1471-2164-10-193> (2009).
14. Grant, J. R., Arantes, A. S., Liao, X. & Stothard, P. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. *Bioinformatics* **27**, 2300–2301, <https://doi.org/10.1093/bioinformatics/btr372> (2011).
15. Liao, X. *et al.* Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection. *Genome* **56**, 592–598, <https://doi.org/10.1139/gen-2013-0082> (2013).
16. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).
17. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1–13, <https://doi.org/10.1093/nar/gkn923> (2009).
18. Johanns, M. *et al.* Direct and indirect activation of eukaryotic elongation factor 2 kinase by AMP-activated protein kinase. *Cellular Signalling* **36**, 212–221, <https://doi.org/10.1016/j.cellsig.2017.05.010> (2017).
19. Muñoz-Soriano, V. *et al.* Evolutionary conserved role of eukaryotic translation factor eIF5A in the regulation of actin-nucleating formins. *Scientific Reports* **7**, 9580, <https://doi.org/10.1038/s41598-017-10057-y> (2017).
20. Jank, T. *et al.* Protein glutamylation is a yeast-specific posttranslational modification of elongation factor 1A. *The Journal of biological chemistry*, <https://doi.org/10.1074/jbc.M117.801035> (2017).
21. Lomonosova, Y. N., Belova, S. P., Mirzoev, T. M., Kozlovskaya, I. B. & Shenkman, B. S. Eukaryotic elongation factor 2 kinase activation in M. soleus under 14-day hindlimb unloading of rats. *Doklady Biochemistry and Biophysics* **474**, 165–167, <https://doi.org/10.1134/s1607672917030048> (2017).
22. Utsunomiya, Y. T. *et al.* Genome-wide mapping of loci explaining variance in scrotal circumference in Nellore cattle. *PLoS one* **9**, e88561, <https://doi.org/10.1371/journal.pone.0088561> (2014).
23. Fountain, M. D., Tao, H., Chen, C. A., Yin, J. & Schaaf, C. P. Magel2 knockout mice manifest altered social phenotypes and a deficit in preference for social novelty. *Genes, Brain and Behavior* **16**, 592–600, <https://doi.org/10.1111/gbb.12378> (2017).
24. Wu, Y. *et al.* Genome-wide association studies using haplotypes and individual SNPs in Simmental cattle. *PLoS one* **9**, e109330, <https://doi.org/10.1371/journal.pone.0109330> (2014).
25. Bastin, B. C. *et al.* A polymorphism in XKR4 is significantly associated with serum prolactin concentrations in beef cows grazing tall fescue. *Animal genetics* **45**, 439–441, <https://doi.org/10.1111/age.12134> (2014).
26. Fortes, M. R. *et al.* Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. *Animal genetics* **44**, 636–647, <https://doi.org/10.1111/age.12075> (2013).
27. Fortes, M. R., Reverter, A., Hawken, R. J., Bolormaa, S. & Lehnert, S. A. Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. *Biology of reproduction* **87**, 58, <https://doi.org/10.1095/biolreprod.112.101089> (2012).
28. Fortes, M. R. S. *et al.* Finding genes for economically important traits: Brahman cattle puberty. *Animal Production Science* **52**, 143–150, <https://doi.org/10.1071/AN11165> (2012).
29. Rubin, C.-J. *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591, [http://www.nature.com/nature/journal/v464/n7288/supinfo/nature08832\\_S1.html](http://www.nature.com/nature/journal/v464/n7288/supinfo/nature08832_S1.html) (2010).
30. Karim, L. *et al.* Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature genetics* **43**, 405–413, <https://doi.org/10.1038/ng.814> (2011).
31. Fortes, M. R., Reverter, A., Kelly, M., McCulloch, R. & Lehnert, S. A. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. *Andrology* **1**, 644–650, <https://doi.org/10.1111/j.2047-2927.2013.00101.x> (2013).
32. Ji, J., Zhou, L., Guo, Y., Huang, L. & Ma, J. Genome-wide association study identifies 22 new loci for body dimension and body weight traits in a White Duroc x Erhualian F2 intercross population. *Asian-Australasian journal of animal sciences*. <https://doi.org/10.5713/ajas.16.0679> (2017).
33. Song, Y. *et al.* Genome-Wide Association Study Reveals the PLAG1 Gene for Knuckle, Biceps and Shank Weight in Simmental Beef Cattle. *PLoS one* **11**, e0168316, <https://doi.org/10.1371/journal.pone.0168316> (2016).
34. Hernandez, R. D. A flexible forward simulator for populations subject to selection and demography. *Bioinformatics* **24**, 2786–2787, <https://doi.org/10.1093/bioinformatics/btn522> (2008).
35. Utsunomiya, Y. T. *et al.* Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. *BMC genetics* **14**, 52, <https://doi.org/10.1186/1471-2156-14-52> (2013).
36. Zhao, F., McParland, S., Kearney, F., Du, L. & Berry, D. P. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genetics, selection, evolution: GSE* **47**, 49, <https://doi.org/10.1186/s12711-015-0127-3> (2015).
37. Randhawa, I. A., Khatkar, M. S., Thomson, P. C. & Raadsma, H. W. Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. *BMC genetics* **15**, 34, <https://doi.org/10.1186/1471-2156-15-34> (2014).
38. Geisler, C., Mabashi-Asazuma, H., Kuo, C. W., Khoo, K. H. & Jarvis, D. L. Engineering beta1,4-galactosyltransferase I to reduce secretion and enhance N-glycan elongation in insect cells. *Journal of biotechnology* **193**, 52–65, <https://doi.org/10.1016/j.jbiotec.2014.11.013> (2015).
39. Littlejohn, M. D. *et al.* Effects of reduced frequency of milk removal on gene expression in the bovine mammary gland. *Physiological genomics* **41**, 21–32, <https://doi.org/10.1152/physiolgenomics.00108.2009> (2010).
40. Sevane, N., Armstrong, E., Wiener, P., Pong Wong, R. & Dunner, S. Polymorphisms in twelve candidate genes are associated with growth, muscle lipid profile and meat quality traits in eleven European cattle breeds. *Molecular biology reports* **41**, 4721–4731, <https://doi.org/10.1007/s11033-014-3343-y> (2014).
41. Carvalho, M. E. *et al.* Heat shock and structural proteins associated with meat tenderness in Nellore beef cattle, a Bos indicus breed. *Meat Sci* **96**, 1318–1324, <https://doi.org/10.1016/j.meatsci.2013.11.014> (2014).
42. Wang, D. *et al.* Effects of nutritional level of concentrate-based diets on meat quality and expression levels of genes related to meat quality in Hainan black goats. *Animal science journal = Nihon chikusan Gakkaiho* **86**, 166–173, <https://doi.org/10.1111/asj.12251> (2015).
43. Pausch, H. *et al.* Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genetics, selection, evolution: GSE* **49**, 24, <https://doi.org/10.1186/s12711-017-0301-x> (2017).
44. Popovic, D. & Dikic, I. TBC1D5 and the AP2 complex regulate ATG9 trafficking and initiation of autophagy. *EMBO reports* **15**, 392–401, <https://doi.org/10.1002/embr.201337995> (2014).
45. Kommadath, A. *et al.* Genetic architecture of gene expression underlying variation in host response to porcine reproductive and respiratory syndrome virus infection. *Scientific Reports* **7**, 46203, <https://doi.org/10.1038/srep46203> (2017).

46. Georges, M. *et al.* Microsatellite mapping of a gene affecting horn development in *Bos taurus*. *Nature genetics* **4**, 206–210, <https://doi.org/10.1038/ng0693-206> (1993).
47. Seichter, D. *et al.* SNP-based association mapping of the polled gene in divergent cattle breeds. *Animal genetics* **43**, 595–598, <https://doi.org/10.1111/j.1365-2052.2011.02302.x> (2012).
48. Brenneman, R. A. *et al.* The polled locus maps to BTA1 in a *Bos indicus* x *Bos taurus* cross. *The Journal of heredity* **87**, 156–161 (1996).
49. Mariasegaram, M. *et al.* Fine-mapping the POLL locus in Brahman cattle yields the diagnostic marker CSAFG29. *Animal genetics* **43**, 683–688, <https://doi.org/10.1111/j.1365-2052.2012.02336.x> (2012).
50. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
51. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595, <https://doi.org/10.1093/bioinformatics/btp698> (2010).
52. Picard. *Picard*, <http://broadinstitute.github.io/picard/> (.
53. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491–498, <https://doi.org/10.1038/ng.806> (2011).
54. Su, J., Wang, Y., Xing, X., Liu, J. & Zhang, Y. Genome-wide analysis of DNA methylation in bovine placentas. *BMC genomics* **15**, 12, <https://doi.org/10.1186/1471-2164-15-12> (2014).
55. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic acids research* **36**, D154–158, <https://doi.org/10.1093/nar/gkm952> (2008).
56. Koufariotis, L. T., Chen, Y. P., Chamberlain, A., Vander Jagt, C. & Hayes, B. J. A catalogue of novel bovine long noncoding RNA across 18 tissues. *PLoS one* **10**, e0141225, <https://doi.org/10.1371/journal.pone.0141225> (2015).
57. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158, <https://doi.org/10.1093/bioinformatics/btr330> (2011).

## Acknowledgements

We would like to acknowledge the generous support of seedstock members and producers from the Australian Brahman Breeders Association (ABBA) who generously donated bull semen for this project. We would like to extend a thank you to Dr. Abdulqader Jighly of Agriculture Victoria, for assisting with the simulation study.

## Author Contributions

L.K. performed the analysis and wrote the paper. B.J.H. and S.M. both contributed extensively to the write up of the paper. A.J.C. prepared, performed the quality control and alignment of the 46 Brahman genome sequences, B.M.B., R.L., M.K., P.S. and A.J.C. all contributed to the write up of the paper. P.S. provided the NGS-SNP annotations of the Brahman sequence variants. M.K. and B.B. ran the algorithm to select the 46 bulls that captured the highest proportion of genetic variation in the breed.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-35698-5>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018