ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/j.1365-2052.2010.02053.x

The evolution of tropical adaptation: comparing taurine and zebu cattle

E. K. F. Chan[1], S. H. Nagaraj and A. Reverter

Cooperative Research Centre for Beef Genetic Technologies. CSIRO Livestock Industries, Queensland Bioscience Precinct 306 Carmody Rd., St. Lucia, Brisbane, Qld 4067, Australia

Summary

Beef cattle breeds consist of three major genetic subdivisions. The taurine group is adapted to temperate environments, and the zebu and Sanga groups are both adapted to tropical environments. With the advent of genotyping and sequencing technologies in agriculture, genome-wide exploration of the genetic basis for the differences in tropical adaptation has only just become possible. In this study, approximately 9000 single nucleotide polymorphism markers were genotyped on 317 animals of a selection of taurine, zebu, and composite breeds to characterize any systematic differences between these groups. We identified 91 intra-breed-class markers; 78 were polymorphic only within the zebu animals, while 13 were polymorphic only in the taurine animals. There were no fixed differences (fixed for alternate alleles between the two breed types) between zebu and taurine animals. We found 14 regions with significantly different allele frequencies between zebu and taurine animals indicative of variable selection pressure or genetic drift. We also found 12 independent regions of differential extended haplotype homozygosity (EHH), indicative of recent selection or rapid fixation of the alternate allele within a short period of time in one of the two breed classes. A preliminary functional genomics analysis of these regions pointed towards signatures of tropical attributes including keratins, heat-shock proteins and heat resistance genes. We anticipate this investigation to be a stepping-stone for future studies to identify genomic regions specific to the two cattle groups, and to subsequently assist in the discrimination between temperate and tropically adapted cattle.

Keywords Bovine, single nucleotide polymorphism, tropical adaptation.

Introduction

Modern cattle (*Bos taurus* L.) were probably domesticated several times in Southwest Asia from the aurochs, which had already diverged phenotypically into two major geographic land races: temperate and tropical (Fries & Ruvinsky 1999). Originally, this phenotypic difference was thought of as representing a species difference, hence the use of the species name *Bos indicus* for tropically adapted cattle. However, there is heterosis between cattle from the two geographic races, and they are indeed members of one species, but in animal genetics literature the term *Bos indicus* is universally accepted despite the absence of a species difference.

Lenstra & Bradley (1999) and Bradley *et al.* (1996) provide a review of the phylogenetic analyses that have been performed on wild and domestic cattle species. Arguably, there are three generally recognized cattle breed classes: taurine, zebu and Sanga. Taurines represent those descended from European and Southwest-Asian ancestors, and have short ears and no hump. Zebu breeds represent those descended from South Asian ancestors and have long floppy ears and a prominent hump. Zebu animals were introduced to Africa by the Arab traders more than a thousand years ago, so the geographic influence of zebu includes East Africa. The origins of the Sanga breeds are less clear, but they are found in West and South Africa, and appear to have been in Africa longer than the zebu breeds. In East Africa, there has been a long history of crossing between

Address for correspondence

Dr A. Reverter, Cooperative Research Centre for Beef Genetic Technologies CSIRO Livestock Industries, Queensland Bioscience Precinct 306 Carmody Rd., St. Lucia, Brisbane, Qld 4067, Australia.
E-mail: Tony.Reverter-Gomez@csiro.au

[1]Present address: Monsanto Vegetable Seeds Division 37437 State Highway 16 Woodland, CA 95695, USA.

Accepted for publication 14 February 2010

zebu and Sanga breeds, originally through the use of zebu bulls. In general, purebred Sanga cattle do not have a hump.

As a result of the origins and breeding practices, both natural and artificial, that occurred in different regions of the world, cattle of the different types are broadly divided into temperate (taurine) and tropical (zebu and Sanga) based on the common adaptation characteristics that they possess. For simplicity, zebu and Sanga are grouped as a single group in this article. Temperate cattle have thicker coats, several breeds develop a winter coat, some are susceptible to sunburn, and they have stocky bodies. Tropical cattle have lower rectal temperatures in hot weather, carry lower burdens of the cattle tick *Boophilus microplus*, and show greater ability to tolerate poor feed and inconsistent climate, which are characteristics of tropical environments compared with more temperate, consistent environments. Zebu cattle show different foraging behaviour, and they have a different capacity for reproduction (Lunstra & Cundiff 2003 and Chase *et al.* 2004; also see reviews by Turner (1980) and Mukasa-Mugerwa (1989)).

Evidence also exists at the genetic level that demonstrates differences between the two cattle groups. Kieffer & Carwright (1968) showed that the Y chromosome of *Bos taurus* bulls is submetacentric, while in *Bos indicus* bulls the Y chromosome is acrocentric. Differences between Asian zebu and African taurine cattle have been observed at the level of mitochondrial DNA (Loftus *et al.* 1994; Bradley *et al.* 1996). At the level of autosomal DNA, there is also evidence of differentiation between all three cattle groups, as demonstrated using microsatellite markers (MacHugh *et al.* 1997; Ibeagha-Awemu *et al.* 2004). Microsatellites do not generally show fixed differences between groups because of the large number of alleles that they usually possess. There has been little effort so far to identify DNA polymorphisms on a genome-wide scale that would allow identification of all three groups, although there have been a few cases where DNA variants have been described that are polymorphic in one group, say taurine, but monomorphic in another group (Kemenes *et al.* 1999; Nijman *et al.* 2003).

The recent efforts from The Bovine Genome Sequence Analysis Consortium (2009) and The Bovine HapMap Consortium (2009) represent an unprecedented resource to disentangle the genetic architecture of complex traits in cattle. Animal geneticists have quickly exploited this resource to address a number of questions such as the effect of domestication on molecular evolution (MacEachern *et al.* 2009a), including the examination of positive selection and effective population size (MacEachern *et al.* 2009b), as well as the relationship between regions under positive selection and association to traits (Barendse *et al.* 2009). More recently, Flori *et al.* (2009) have used data from dense genotyping platforms to identify the main regions affected by the strong and recent artificial selection in three breeds of dairy cattle. The authors reported the existence of 13 highly significant regions subjected to strong and/or recent positive selection, and the genomic functionality of these regions pointed towards the antagonism between intensive dairy production and reproductive performance. The same group (Gautier *et al.* 2009) performed a whole genome scan for footprints of adaptive selection in nine West African cattle populations and identified 53 genomic regions.

Complementing these studies, the task of identifying a large number of DNA variants that are different between taurine and zebu groups would facilitate the study of tropical adaptation, as well as provide some practical tools in cattle management. Traditionally, the proportion of zebu contribution to an individual animal is crudely scored based on the extent of observable phenotypic differences such as the presence and size of a hump, and ear floppiness. With a better understanding of the genetic differences between breed-types, DNA variants that are fixed in either taurine or zebu animals would allow animals of composite zebu-taurine ancestry to be identified more efficiently. A desirable chromosomal section originating from zebu cattle could be followed over generations and its contribution to zebu-taurine differentiation may be determined. In particular, genomic regions responsible for major differences between taurine and zebu that show little variation within the individual breed-type could be studied using a larger set of these polymorphisms.

Therefore, the objective of this study is to examine the genotype of cattle of a variety of breeds including both taurine and zebu types of cattle using more than 9000 autosomal and X-linked single nucleotide polymorphisms (SNPs). We put particular emphasis on identifying fixed differences between taurine and zebu animals, as well as identifying regions of the bovine genome that show large allele frequency differences between zebu and taurine animals.

Materials and methods

Animals

A subset of unrelated animals from the Australian Cooperative Research Centre for Beef Genetic Technologies (Beef CRC; http://www.beef.crc.org.au) reported previously (Upton *et al.* 2001; Burrow *et al.* 2003; Wolcott *et al.* 2006; Barendse *et al.* 2009) were used (no full- or half-sibs). They consisted of 317 pure breed cattle, where animals of composite zebu-taurine ancestry were considered to be purebred if their parents were also of the same composite ancestry. None of the animals were crossbred in the sense of having parents from different breeds. These animals consisted of 70 zebu animals of the Brahman (BRM) breed, 24 composite zebu-taurine Santa Gertrudis animals (SGT), 30 composite Sanga-taurine Belmont Red animals (BEL), and the rest were members of 10 taurine breeds of beef or dairy ancestry. These consist of the four beef breeds comprising Angus

(ANG; $n = 42$), Hereford (HFD; $n = 34$), Murray Grey (MGY; $n = 14$) and Shorthorn (SHN; $n = 18$); and seven dairy breeds comprising Brown Swiss (BSW; $n = 4$), Guernsey (GNS; $n = 4$), Jersey (JER; n-10), Illawarra Shorthorn (IWSn=8) and Australian Red (AUR; $n = 7$) and Holstein (HOL; $n = 52$).

SNP genotypes

The animals were genotyped using the MegAllele 10k SNP Panel (Hardenbol *et al.* 2005) by ParAllele Inc. and its parent company Affymetrix. This SNP panel consists of 9919 SNPs that are randomly (and roughly uniformly) distributed across the genome with an average spacing of approximately 325 kb per SNP. Further details of the SNPs can be found at the link ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/snp/Btau20050310/. The bulk of the SNPs on the array were obtained by comparing the genome sequence of a Hereford animal to the partial sequence of a Holstein (72.4%), an Angus (15%), an Limousin (3.1%), and a Brahman (2%) animal, with an additional 7.5% cSNPs (coding SNPs) obtained from the Interactive Bovine *in silico* SNP database (Hawken *et al.* 2004). Thus, in this study, the origin of a SNP is designated by the non-Hereford breed used in its discovery, and consequently, all Holstein, Angus and Limousin SNPs are taurine SNPs, while Brahman SNPs are also referred to as zebu SNPs. In summary, the majority of these SNPs are common differences between a taurine beef and dairy animal, with a small percentage of SNPs being polymorphic between a taurine and a zebu beef animal. Of the genotyped SNPs, those with more than 10% of missing data were excluded, leaving a total of 8427 SNPs. Of these, 7956 were mappable to the BTAU4.0 assembly (Liu *et al.* 2009).

Allelic variations

Zebu and taurine fixed differences were determined by comparing the allele distribution in the Brahman breed with the combined purebred taurine animals. A SNP is defined as private in taurine animals if it is polymorphic in each of the ten taurine breeds with a minor allele frequency (MAF) ≥5% and it is monomorphic in the Brahman breed. Similarly, a SNP is private in Brahman if it is polymorphic with MAF ≥0.05 and it is monomorphic in all taurine animals. A rarefaction approach through the ADZE software (Szpiech *et al.* 2008) was used for estimating the number of private alleles per locus while accounting for sample size differences across breeds and breed-types. For each group of SNPs (described in the previous section), the average numbers of alleles per SNP were estimated for each breed-type for an assumed sample size of 2 to 20.

Fixation indices (F_{ST}) were estimated using the method of Weir and Cockerham (Weir & Cockerham 1984) for (1) between zebu breeds, (2) between taurine breeds and (3) between taurine and zebu breeds, where all taurine animals were grouped into a single population. Estimates were similar irrespective of whether all SNPs, only autosomal SNPs or specific SNP types were used (Table S1). Results were used as a symmetrical distance matrix for the unrooted Neighbour-Joining Tree estimation using R/APE (Paradis *et al.* 2004).

Compound diplotype

We used the SNP density to identify chromosomal regions that are shared identical-by-state by searching for long identical diplotypes shared within taurine or zebu animals but differing between these types. Because allelic phase is unknown for our SNP, we define a compound diplotype as one containing at least 20 consecutive SNPs, all of which must have significantly differential allele frequencies between the two breed types. The test for difference in allelic frequencies was performed using the two-proportion Z-test; for each locus,

$$z = \frac{p_{\text{indicus}} - p_{\text{taurus}}}{\text{SE}} \quad \text{and} \quad \text{SE} = \sqrt{p(1-p)\left(\frac{1}{n_{\text{indicus}}} + \frac{1}{n_{\text{taurus}}}\right)}$$

where p is the total allele frequency and n is the sample size. The H_o: $z = 0$ was assessed with P-values obtained from a normal distribution. A compound diplotype is defined if at least 20 consecutive SNPs have point-wise $P < 0.05$, and a representative SNP per compound diplotype is chosen as the one with the largest $|z|$.

Extended haplotype homozygosity

The counting algorithm of Tang *et al.* (2007) was implemented for identifying differential extended haplotype homozygosity (EHH) regions between taurine and zebu. Individually for each breed type and SNP_i, the proportion of individuals that remain homozygous for a genomic interval extending in both directions from SNP_i was calculated and labelled as $EHHS_{i,j}$. The size of this interval is unique for each SNP_i and is based on SNP_j, the SNP closest to SNP_i, such that $EHHS_{i,k} < 0$; this was determined for both $j<i$ and $i<j$. The EHH at SNP_i was summarized as the integral $iES_i = \Sigma(EHHS_{i,j})$; i.e. the sum of $EHHS_{i,j}$ within the previously identified interval for SNP_i. Differential regions of EHH between taurine and zebu were based on the standardized log-ratio of iES_i between the two breed types (Tang *et al.* 2007): $\ln(Rsb_i) = \ln(iES_{i,T}/iES_{i,Z})$ where T = taurine & Z = zebu. To identify significant regions of positive selection, we estimated (1) the null distribution of $\ln(Rsb_i)$, and (2) distribution of noise: $SD(\ln(Rsb_i))/\ln(Rsb_i)$. SNP i is significantly under differential selection pressure between the two breed types when it satisfied two criteria. First, $\ln(Rsb_i)$ must have bootstrap $P \leq 0.01$: i.e. if $\ln(Rsb_i)$ is more extreme than 1% of 200 bootstrap estimates, where each bootstrap estimate was determined from a repeat analysis with individuals re-sampled from the total population (combining

the two breed types). Second, $\ln(Rsb_i)$ has to be within the mid-50 percentile of its noise distribution, where such a distribution was based on 50 bootstrap analyses with individuals re-sampled within their own breed group. Finally, a genomic region was declared as significant if ≥50% of the SNPs within the region were significant.

STRUCTURE

The Bayesian clustering program STRUCTURE (Pritchard *et al.* 2000) was run assuming admixture model and correlated allele frequencies (Evanno *et al.* 2005) with the degree of admixture inferred from the data. From preliminary STRUCTURE runs we determined that 6000 burn-ins followed by 1000 MCMC iterations were sufficient to ensure convergence of parameter estimates (data not shown). For each K (assumed number of ancestral populations), five replicate runs were performed. The ΔK method of Evanno *et al.* (2005) was employed to determine the K that best represented our data from $K = 1$ to $K = 13$; all five replicate runs revealed a clear peak at $K = 2$ (Fig. S6). We used the modified version of Symmetric Similarity Coefficient (Nordborg *et al.* 2005) initially proposed by Rosenberg *et al.* (2002) to quantify the consistency between replicate runs. The average and standard deviations of the estimated proportions of the two ancestral proportions were estimated for each breed: i.e. estimated across all individuals of a breed.

Results

Allelic privacy: first indication of genetic difference between breed types

The first bovine SNP genotyping array platform (MegAllele 10K SNP panel; Hardenbol *et al.* 2005) provided an excellent resource for identifying breed-type specific polymorphisms as a result of the approach adopted for SNP discovery, namely the identification of SNPs between two breeds. We used this SNP panel to study the genetic differences between 10 taurine breeds, a zebu breed (Brahman), a zebu-taurine composite breed (Santa Gertrudis), and a Sanga-taurine composite breed (Belmont Red). Of the 8238 informative SNPs (polymorphic with minor allele frequency (MAF) exceeding 5% in at least one breed), 13 were private in the taurine breed-type (i.e. polymorphic in taurines but not in Brahmans; Table S2). Based on the method of SNP discovery (Hardenbol *et al.* 2005), ten of these were known to be biallelic within a DNA pool of at least two taurine breeds (Hereford vs. Holstein or Angus or Limousin); i.e. the two alleles could be segregating in one or both breeds, or the two breeds could fixed for the alternate alleles. Our data further showed that these markers are also polymorphic within each of the taurine breeds but monomorphic within Brahmans. By contrast, 78 SNPs were private in Brahmans (Table S2). The majority of these (∼70%) were known to be polymorphic between Brahmans and Herefords (Hardenbol *et al.* 2005). Here, we showed that these markers are polymorphic within Brahmans and fixed in all ten taurine breeds for the same allele. Of these total 91 private SNPs, 67 and 56 were polymorphic in the Santa Gertrudis and Belmont Red samples respectively; more than 53% were polymorphic in both and 18% were polymorphic in only one of the two composite breeds (Table S2). We found no polymorphisms with alternate segregating alleles between the two breed types; i.e. no DNA variants were fixed (i.e. monomorphic) in the Brahman for one allele and fixed for the alternate allele in the combined taurine sample or vice versa.

Two major limitations were recognized in this study, namely unbalanced sampling (many of the smaller breed samples had higher proportions of observed monomorphism; Fig. 1) and SNP discovery bias. The latter limitation is illustrated by the consistently lower proportion of monomorphism in a breed for SNPs obtained by comparing the Hereford reference sequence to an animal from that breed (Figs 1 & 2). This is particularly true for the Angus and Holstein-derived SNPs; a lack of similar evidence in the

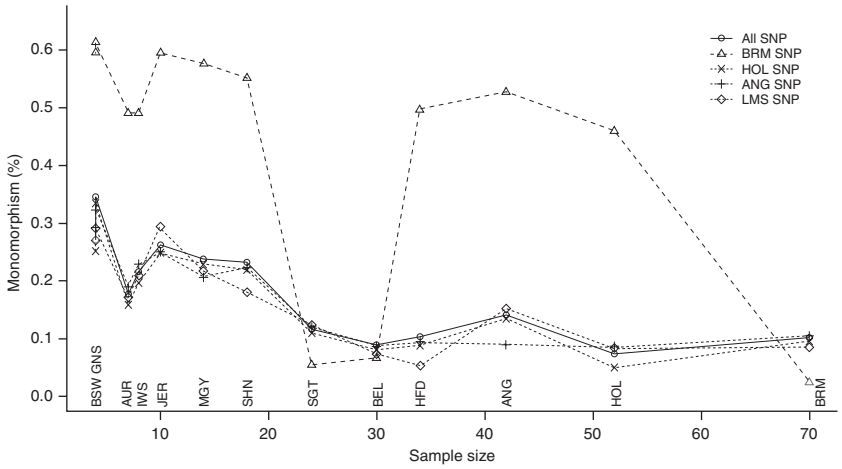

Figure 1 Relationship between proportion of monomorphism and sample size. The proportions of monormophism are shown for all or breed-specific (BRM, HOL, ANG, LMS) SNPs. The breed with the corresponding sample sizes are shown at the bottom of the plot (see Materials and methods for breed code).

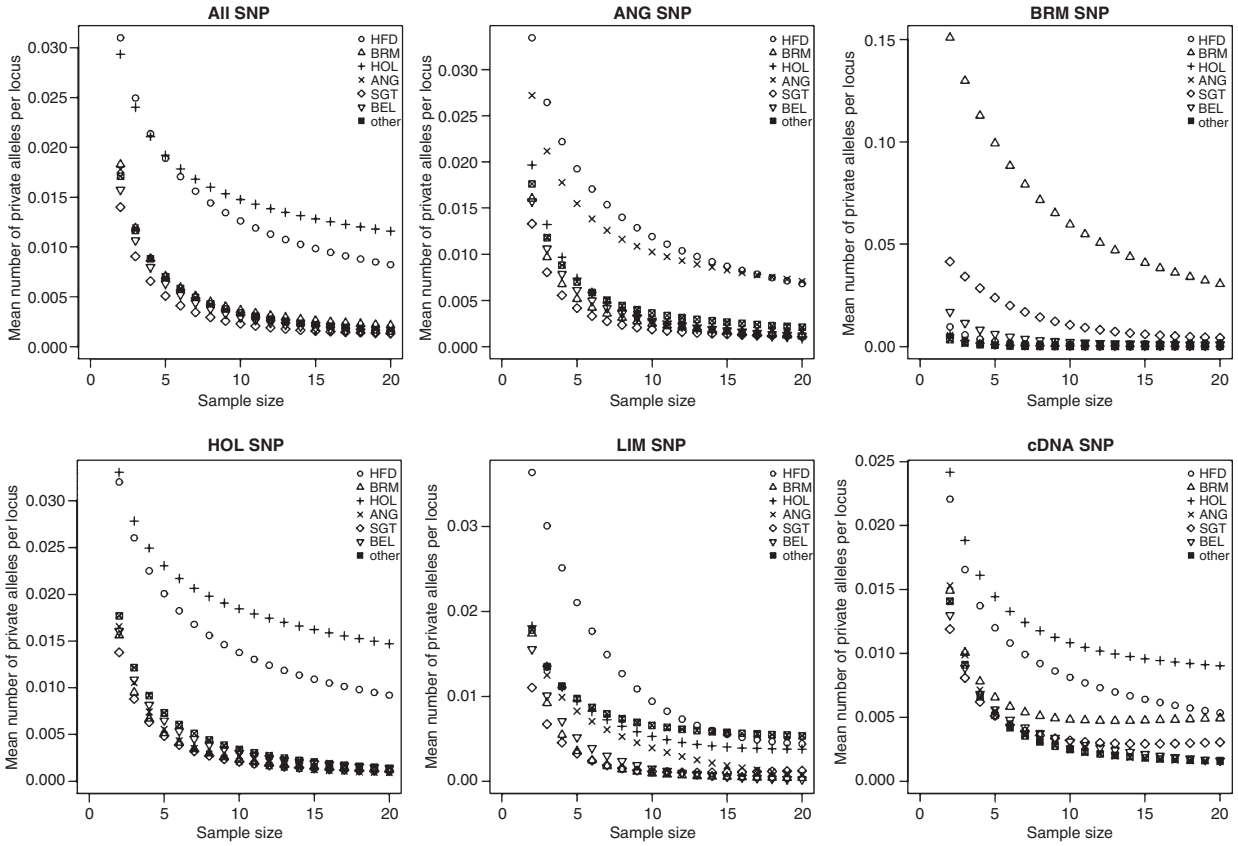

Figure 2 Rarefaction analysis on the number of private alleles per locus. The ADZE software was used to estimate the average number of private alleles per locus for sample sizes of two to twenty.

Limousine-derived SNPs can be explained by the under-representation of Limousine SNPs in the SNP panel. Despite these limitations, our results showed that Brahman is intrinsically more variable than taurine breeds. All taurine breeds showed high levels of monomorphism for DNA variants for Brahman SNPs (Fig. 2). Even the Hereford, in its role as the reference breed for SNP discovery, was monomorphic for 15–30% of the Brahman SNPs. By contrast, while Brahman animals showed similar levels of monomorphism to taurine animals for taurine SNPs not used in SNP discovery, many of the Brahman SNPs had a higher proportion of polymorphisms specific to Brahman.

The composite breeds, Santa Gertrudis and Belmont Red, also showed similar patterns of lower monomorphism for both the taurine and zebu SNPs, but this was unsurprising given their composite origins.

Genetic variations and breed relationships

Although there are few Brahman SNPs in this dataset, they had a disproportionate effect on the estimates of genetic diversity (as per the F_{ST} index) between breeds and breed types due to differences in the extent of polymorphism between breeds. Using taurine-derived SNPs alone, the estimated F_{ST} between taurine breeds was 12.2% and between taurine and zebu breeds was 22.1%. Using the Brahman SNPs alone, the F_{ST} between taurine breeds was 9.5% and between taurine and zebu breeds was 50.6%. Using all the SNPs, F_{ST} between taurine breeds was 12.1% and between taurine and zebu breeds was 22.8%. These inter-taurine breed F_{ST} estimates were consistent with previous reports (Kantanen *et al.* 2000; Wiener *et al.* 2004), thus providing confidence towards the clear difference between inter-taurine F_{ST} and taurine-zebu F_{ST} estimates, despite the notable SNP ascertainment bias (Table S1).

Relationships between breeds were determined by constructing an unrooted Neighbour-Joining tree using breed-pair F_{ST} estimates. These results (Fig. 3) were highly consistent with the known genealogy/history of the breeds. Most notably, Brahman is most distinct from the other 12 breeds; the two composite breeds were clustered together on the same branch as Brahman, all of which were distinct from the taurine breeds. This global picture of breed relationships was also obtained with breed-specific SNPs, the subset of autosomal SNPs, or a subset of equal numbers of Brahman and Holstein SNPs (Figs S4 & S5), suggesting that the differentiation of Brahman from composite breeds from taurine breeds surpasses any inherent SNP discovery biases.

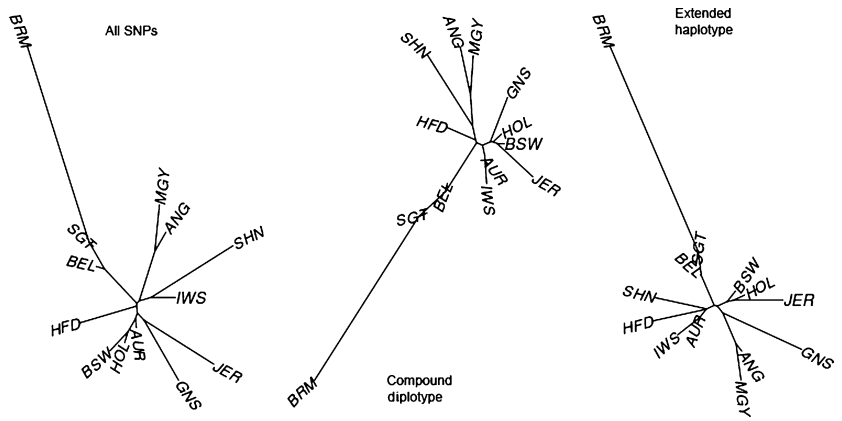

Figure 3 Neighbour-joining tree of all 13 breeds constructed using F_{ST} values estimated for each breed pair. The same analysis was performed using all 8427 SNPs, 326 SNPs within the 14 compound diplotype regions, or 124 SNPs within the 12 extended haplotype homozygosity regions. Breed acronyms are as follows: BRM, Brahman; SGT, Santa Gertrudis; BEL, Belmont Red; HFD, Hereford; BSW, Brown Swiss; HOL, Holstein; AUR, Australian Red; GNS, Guernsey; JER, Jersey; IWS, Illawarra Shorthorn; SHN, Shorthorn; ANG, Angus; and MGY, Murray Grey.

However, despite this clear pattern of breed-type divergence, the current SNP panel does not allow accurate quantification of divergence time between cattle breeds.

Genomic differences between breed types

Because there are only a few (~1% of total SNPs) fixed differences between the zebu and taurine breeds, we examined whether there were regions of the genetic material that showed many SNPs with consistently different allele frequencies. We identified 14 compound diplotypes encompassing 326 SNPs, ranging from 21 to 30 SNPs per compound diplotype (Tables 1 & S3 for full listing of SNPs). We tested the null hypotheses that the 326 SNPs within the 14 compound diplotypes were sampled randomly from the total SNP set without bias for any of the SNP-discovery breeds, using the chi-square test with *P*-values estimated from 5000 permutations. There was evidence that the 326 SNPs were over-represented by Limousin- and Brahman-derived SNPs and under-represented by Holstein-discovered SNPs ($\chi^2 = 23.1$, $P < 0.001$. These 14 compound diplotypes represent genomic regions that have undergone (or are undergoing) independent genetic selection and therefore independent adaptation.

To identify positive selections that have led to complete or near complete fixation, we searched for regions of differential extended haplotype homozygosity (EHH) between the two breed types (Tang *et al.* 2007). A total of 142 SNPs were identified as having significantly differential extended haplotype homozygozity values between the two breed types. Of these, we deduced twelve regions, encompassing a subset of 124 SNPs (6–47 SNPs per region), with significant signals of strong recent positive selection (Table 2). In general, much stronger evidence of selection was observed in taurines when compared with zebus (extent of extended haplotype homozygosity was higher in taurines compared with Brahman; Fig. S6), and this was true for eight of the twelve significant regions (Table S4), thus supporting the common theory that zebus are more ancestral than taurines.

Table 1 Compound diplotypes: genomic regions with significant evidence for differential allele frequencies between taurine and zebu cattle.

Number of SNP	Chromosome: Interval (Mb)	Genes
22	1: 89.7–95.5	
24	3: 98.8–104.6	Solute carrier *SLC1A7* Tick-resistant gene *NDUFA12*
30	4: 47.4–59.9	Solute carrier *SLC26A3* & *SLC26A4* Overlaps QTL for marbling score in cattle from four independent studies
29	5: 64.5–72.9	Solute carriers *SLC17A8*, *SLC25A3* & *SLC5A8*
21	5: 110.8–118.7	Interleukin *IL17RA* CD antigen *CD9* Solute carriers: *SLC16A8*, *SLC6A12*, *SLC6A13* Tick-resistant gene *NDUFA9*
21	6: 34.0–37.6	
30	6: 107.9–116.1	CD antigen *CD38*
28	8: 40.5–47.3	Interleukin *IL33* & *CD274* Solute carrier *SLC1A1*
24	10: 28.3–37.4	Tick-resistant gene *NDUFAF1*
25	13: 24.9–30.3	Heat shock protein *HSPA14*
24	15: 61.6–67.6	CD antigen *CD44* & *CD59* Solute carrier *SLC1A2*
24	16: 35.2–45.2	Solute carriers *SLC25A33*, *SLC2A5*, & *SLC45A1*
24	22: 5.8–10.7	
22	X: 48.7–67.1	Interleukin receptor *IL2RG* Solute carriers *SLC35A2* & *SLC7A3*

Table 2 Extended haplotype homozygosity: genomic regions with significant evidence for relatively recent positive selection between zebu and taurine cattle.

Number of SNP	Chromosome: Interval (Mb)	Genes
6	5: 15.7–18.5	Solute carrier *SLC6A15*
8	10: 9.3–11.5	
2[1]	10: 81.5–82.9	
2	11: 28.1–28.1	No genes found
4	13: 70.8–72.0	
10	18: 14.9–21.1	
5[1]	19: 42.2–44.2	Family of keratin genes Heat shock protein *HSPB9*
13[1]	21: 24.2–30.7	Interleukin *IL16*
4	22:20.8–21.9	No genes found
16[1]	22: 46.1–56.7	Interleukin *IL17RB* Solute carriers *SLC25A20, SLC26A6, SLC38A3, SLC6A1 & SLC6A20*
7	X: 1.2–7.0	
47	X: 39.5–73.5	Interleukin receptor *IL2R2* Solute carriers *SLC35A2* & *SLC7A3* Tick-resistant gene *NDUFV2*

[1]Indicate regions where selection is in the direction of zebu.

Interestingly, these twelve regions did not correspond to our compound diplotypes. In fact, aside from the sex chromosome, the distributions of these two sets of genomic regions appear independent of each other (Fig. 4). These results suggest that the regions of positive selection (EHH regions), likely in taurine breeds, are different to those where both the taurine and zebu are under independent selection (compound diplotypes). Despite the distinction between these two classes of genomic regions, both are able to distinguish and reconstruct the inter-breed relationships as manifested by unrooted Neighbour-Joining trees from using F_{ST} estimates (Fig. 3).

Estimating cattle ancestry

Finally, we used the program STRUCTURE (Pritchard *et al.* 2000) to estimate the proportion of common ancestry between the 13 breeds. Based on 7821 autosomal SNPs, STRUCTURE clearly indicated two ancestral populations corresponding to the 13 breeds (Fig. S7), confirming previously observed results (The Bovine HapMap Consortium 2009). Note that identical results were obtained using all SNPs, inclusive of X-linked and unmapped SNPs, either because X-linked SNPs have minimal effect on estimating cattle ancestry or because there are relatively few X-linked SNPs; only results from autosomal SNPs are presented (Fig. 5). These two clusters corresponded clearly to the two breed types (Fig. 4 top): on average, Brahman individuals have 0.92 (±0.05 SD) probability of belonging to one of the two clusters (zebu ancestry) and on average individuals of the 10 taurine breeds have >0.92 (<0.02 SD) probability of belonging to the second cluster (taurine ancestry). This result is consistent across five replicate runs with symmetric similarity coefficient (Rosenberg *et al.* 2002), SSC, of 0.99. The composite Belmont Red and Santa Gertrudis individuals were found to have mixed taurine and zebu ancestry, with respective probabilities of 0.34 (±0.08 SD; 0.21–0.49) and 0.37 (±0.05 SD; 0.28–0.47) zebu ancestry.

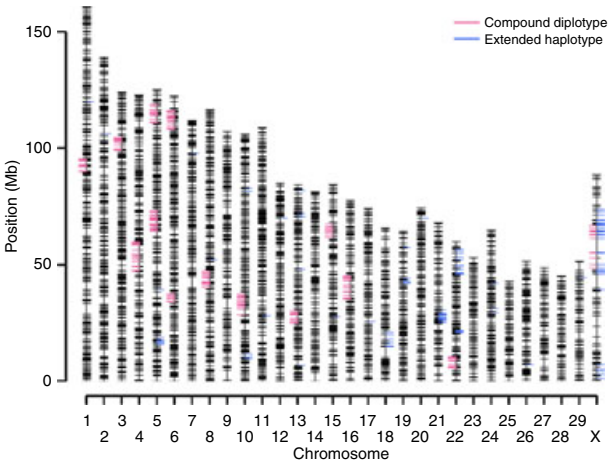

Figure 4 Distribution of SNPs. Each SNP is represented as a horizontal dash on the 30 vertical lines corresponding to the 29 autosomes and the X chromosome. Pink and blue dashes indicate SNPs within the 14 compound diplotypes and 12 extended haplotype homozygosity regions respectively.

Using the same approach, we repeated this analysis using two small subsets of SNPs: (1) 14 SNPs most representative (largest $|z|$) of the 14 compound diplotypes and (2) the 12 SNPs with the largest ln(Rsbi) corresponding to the 12 EHH

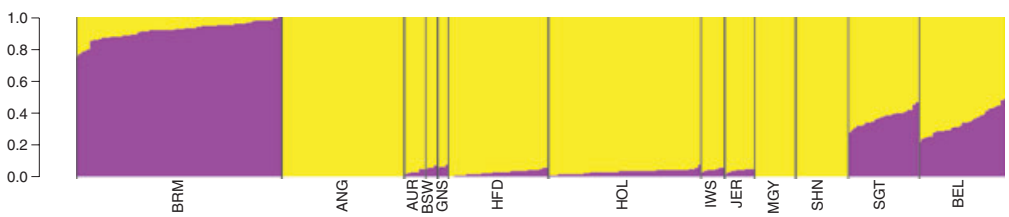

Figure 5 STRUCTURE prediction of the proportion of two ancestral populations (K = 2) corresponding to 317 individuals belonging to 13 breeds using 7821 autosomal SNPs. The result is the average of five Markov chain Monte Carlo replicate runs. Individuals (on the x-axis) have been ordered based on the proportion of taurine (yellow) ancestry within each breed. See Materials and methods for breed code.

regions. The set of 14 compound diplotype SNPs performed in a similar manner to the larger set of autosomal SNPs, where two clusters were identified corresponding to the two breed types, with Brahmans having an average of 97% zebu ancestry and each of the 10 taurine breeds having, on average, >93% taurine ancestry. By contrast, at $K = 2$, the results of the set of 12 EHH SNPs are more suggestive of admixture across all breeds: the average 'zebu' ancestry estimated for the Brahman animals was 80% and the average 'taurine' ancestry for each of the taurine breeds was 65%. The results here again reflect the difference in the definition of these two classes of SNPs (see Discussion; Appendix S1).

Functional genomic analysis of candidate regions

The performance attributes for tropical adaptation in cattle are broadly classified as fertility, growth, carcass composition, heat resistance, parasite resistance and disease resistance. In a bid to identify regions (genes) associated with any of the above characteristics, we combined literature mining, bioinformatics approaches and functional annotation of the cattle genome and carefully studied the 14 compound diplotype (Table 1) and 12 EHH (Table 2) regions. The length of each block of genome varied between 5 Mb and 20 Mb, spanning 12 to 153 genes, including a significant number of genes with unknown function (See Tables S3 & S4 for full list).

In an effort to obtain a broad functional insight for these set of genes, we used Gene Ontologies to find any over-representation in all or a subset of genes. Although we did not observe any over-representation implying heterogeneous nature of genes, we found a number of genes and families of genes that have been reported to be associated with one or more performance attributes for tropical-adaptation (O'Gorman *et al.* 2006, 2009; Wang *et al.* 2007; Piper *et al.* 2009). First, we found a number of keratins on chromosome 19 (42.2–44.2 Mb; Table 2) and where the signature of selection is in the direction of zebu. Second, we found two heat shock proteins: *HSPA14* (Table 1) and *HSPB9* (Table 2). Third, a number of immune system activation genes in response to environmental stress such as interleukins: *IL33, IL16, IL17RB* and *IL17RA*; and CD antigens: *CD9, CD38, CD44, CD59, CD274* and *IL2RG*. Fourth, we found a total of 25 genes from the solute carrier family. Finally, we found a number of genes implicated in tick resistance, including *NADH* dehydrogenases: *NDUFA12, NDUFA9, NDUFAF1* and *NDUFV2* (Piper *et al.* 2009).

A careful observation of the AnimalQTLdb (Hu *et al.* 2007) revealed a specific region in chromosome 4 that reported the presence of QTLs for marbling score in cattle from four independent studies that overlaps with the region we have reported in chromosome 4 (47.4 to 59.9 kb), spanning 30 SNPs.

Discussion

In this study, we examined several techniques to classify the proportion of an animal that could be traced to either a taurine or a zebu origin. Although a breed of composite Sanga-taurine animals was included, none of the SNPs is of Sanga origin, so conclusions for such breeds cannot be categorical because of the inherent ascertainment bias in SNP discovery.

Differences between zebu and taurine cattle, using this sample of animals and SNPs, appear to be more of degree than kind. Given the number of SNPs, it was surprising that only 1% were private, i.e. polymorphic in only taurine or zebu animals. Most of these private alleles were Brahman SNPs and private in Brahman animals, rather than for the taurine SNPs or taurine animals. These results suggest that the ancestral populations of cattle were large, so that large numbers of polymorphisms have been maintained and that most polymorphisms may be ancient and predate the split between the ancestors of cattle that led to the zebu breeds compared with the taurine breeds (The Bovine Genome Sequencing and Analysis Consortium *et al.* 2009; The Bovine HapMap Consortium 2009).

The Brahman originated in the United States of America as a composite of at least four breeds from India and Brazil, as well as the inclusion of taurine cows to increased numbers (Briggs & Briggs 1980). Breeders have subsequently tried to increase the amount of zebu ancestry by using semen from purebred zebu animals, but there would still be a residue of taurine ancestry. The range of zebu breeds used, plus the original use of taurine dams, help to explain the greater variability of the Brahman.

Analysis of population substructure shows that some Brahman animals have a residue of taurine alleles. It also shows that some taurine animals show either an introgression of zebu alleles, or alleles that are now primarily found in zebu animals but that may stem from the common ancestor of the zebu and taurine animals. This is supported by the New South Wales Department of Primary Industries, who claim that Brahman was developed from the progeny of four Indian zebu breeds with some infusion of local British breeds (*Bos taurus*) in the early 1800s in USA (AGFACT A2.3.11; http://www.dpi.nsw.gov.au/agriculture/livestock/beef/breeding/breeds/brahman).

The current set of SNPs classifies the composite animals into proportions of zebu and taurine that agrees with the known ancestry of the Santa Gertrudis, which is a nominally 5/8 Shorthorn and 3/8 Brahman. The interesting comparison of ancestry is the Belmont Red, which shows a similar proportion of zebu and Brahman ancestry. The Belmont Red is nominally ½ Africander and ¼ each of Hereford and Shorthorn. In the Beef CRC cattle, commercial Belmont Red cattle were used, and while those are generally without Brahman ancestry, and there is certainly Brahman ancestry in some research herds of the Belmont Red, the level of zebu ancestry found here (34%) is greater than what

would be expected for these animals to be registered as Belmont Red. Since 1985, the Belmont Red Association has allowed up to 25% *Bos indicus* in their registered animals (http://www.belmontred.com.au/). This suggests that these SNPs are a signal of Sanga ancestry, but because Sanga were not used in the SNP discovery, this ancestry is not recognized as a third group.

Genomic regions of differential extended haplotype homozygosity between two populations are indicative of recent selection or rapid fixation of the alternate allele within a short period of time, thereby preventing recombination at nearby regions in one of the two populations. This is different to compound diplotypes, which are extended regions with differential allele frequencies between two populations, and are therefore indicative of variable selection pressure or genetic drift. The EHH approach is useful when we consider the zebu as an ancestral breed to the taurine: recent selection in the taurine from the zebu will be reflected in the analysis. Conversely, if environmental (climatic) adaptation occurred independently in the two populations (breed types), then one would expect the corresponding genetic regions controlling adaptation to be in drift in both populations with different allele frequencies.

Some compound diplotypes may exhibit more than large differences as a result of drift between zebu and taurine ancestries. Further analyses of these SNPs, particularly in animals such as the Nelore or the Gir breeds, which have essentially no known taurine ancestry, might help resolve whether some of the allele distributions represent zebu-specific effects compared with effects that might be attributable to the multibreed zebu as well as original taurine cow composition of the Brahman breed. These regions may represent those parts of the genome that contribute to the temperate and tropical adaptations of zebu and taurine animals. Specific association tests between these SNPs and trait values for parasite resistance, rectal temperatures and drought tolerance may confirm that these are signatures of adaptive evolution.

From a functional genomics viewpoint, we argue that we have indeed found a number of genes that are either directly or indirectly associated with one or more performance attributes for tropical adaptation. For instance, a number of keratins (heteropolymeric structural proteins) form the basis for structural constituents of epidermis during epidermis development, which in turn plays a role in adaptating to different climatic conditions, including tick resistance (Wang *et al.* 2007; Piper *et al.* 2008). In addition, heat shock proteins have been found to be heavily differentially expressed in a number of gene expression studies (for a recent review, see Collier *et al.* (2008) and references therein), and have independently been shown to be associated with tropical adaptation. Finally, the overlapping region in the QTL database also provides additional evidence of the significance of these genomic regions and requires detailed and directed experiments to obtain a thorough insight into molecular basis of tropical adaptation in cattle.

In conclusion, we anticipate the study presented here to be an effective approach to identifying genomic regions specific to the two cattle land races and subsequently assisting in the discrimination between temperate and tropically adapted cattle. The application of our procedure using larger samples and a denser SNP chip is warranted.

Acknowledgements

The authors would like to thank Dr Warren Snelling for providing the mappings of composite markers on the Btau4.0 assembly, and Dr Gavin Huttley for useful discussions. Dr William Barendse provided useful recommendations about data analysis and insights from the Bovine HapMap Consortium. Sigrid Lehnert and Brian Dalrymple provided helpful reviews. Shivashankar Nagaraj is thankful to the CSIRO OCE Post-Doctoral Fellowship Program. We thank the owners of the Beef CRC database, the Commonwealth Scientific and Industrial Research Organisation, the Queensland Department of Primary Industries and Fisheries, the New South Wales Department of Primary Industries and the University of New England.

References

Barendse W., Harrison B.E., Bunch R.J., Thomas M.B. & Turner L.B. (2009) Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* **10,** 178.

Bradley D.G., MacHugh D.E., Cunningham P. & Loftus R.T. (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 5131–5.

Briggs H.M. & Briggs D.M. (1980) *Modern Breeds of Livestock*. Fourth Edition. Macmillan Publishing Co, New York.

Burrow H.M., Johnston D.J., Barwick S.A., Holroyd R.G., Barendse W., Thompson J.M., Griffith G.R. & Sullivan M. (2003) Relationships between carcass and beef quality and components of herd profitability in Northern Australia. In: *Association for the Advancement of Animal Breeding and Genetics*, pp. 359–62. Association for the Advancement of Animal Breeding and Genetics, Melbourne, Australia. Available at: http://www.aaabg.org.

Chase C.C. Jr, Riley D.G., Olson T.A., Coleman S.W. & Hammond A.C. (2004) Maternal and reproductive performance of Brahman × Angus, Senepol × Angus, and Tuli × Angus cows in the subtropics. *Journal of Animal Science* **82,** 2764–72.

Collier R.J., Collier J.L., Rhoads R.P. & Baumgard L.H. (2008) Invited review: genes involved in the bovine heat stress response. *Journal of Dairy Science* **91,** 445–54.

Evanno G., Regnaut G. & Goudet J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14,** 2611–20.

Flori L., Fritz S., Jaffrezic F., Boussaha M., Gut I., Heath S., Foulley J.L. & Gautier M. (2009) The genome response to

artificial selection: a case study in dairy cattle. *PLoS ONE* **4,** e6595.

Fries R. & Ruvinsky A. (1999) *The Genetics of Cattle*. CAB International, Wallingford, UK.

Gautier M., Flori L., Riebler A., Jaffrézic F., Laloé D., Gut I., Moazami-Goudarzi K. & Foulley J.L. (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* **10,** 550.

Hardenbol P., Yu F., Belmont J. *et al.* (2005) Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Research* **15,** 269–75.

Hawken R.J., Barris W.C., McWilliam S.M. & Dalrymple B.P. (2004) An interactive bovine in silico SNP database (IBISS). *Mammalian Genome* **15,** 819–27.

Hu Z.L., Fritz E.R. & Reecy J.M. (2007) AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Research* **35,** D604–9.

Ibeagha-Awemu E.M., Jann O.C., Weimann C. & Erhardt G. (2004) Genetic diversity, introgression and relationships among West/Central African cattle breeds. *Genetic Selection Evolution* **36,** 673–90.

Kantanen J., Olsaker I., Holm L.E., Lien S., Vilkki J., Brusgaard K., Eythorsdottir E., Danell B. & Adalsteinsson S. (2000) Genetic diversity and population structure of 20 North European cattle breeds. *Journal of Heredity* **91,** 446–57.

Kemenes P.A., Regitano L.C.d.A., Rosa A.J.d.M., Packer I.U., Razook A.G., Figueiredo L.A.d., Silva N.A., Etchegaray M.A.L. & Coutinho L.L. (1999) *k-Casein, b-lactoglobulin* and *growth hormone* allele frequencies and genetic distances in Nelore, Gyr, Guzerá, Caracu, Charolais, Canchim and Santa Gertrudis cattle. *Genetics and Molecular Biology* **22,** 539–41.

Kieffer N.M. & Carwright T.C. (1968) Sex chromosome polymorphism in domestic cattle. *Journal of Heredity* **59,** 35–7.

Lenstra J.A. & Bradley D.G. (1999) Systematics and phylogeny of cattle. In: *The Genetics of Cattle* (Ed. by R. Fries & A. Ruvinsky), pp. 1–14. CABI Publishing, Oxfordshire, UK.

Liu Y., Qin X., Song X.Z. *et al.* (2009) *Bos taurus* genome assembly. *BMC Genomics* **10,** 180.

Loftus R.T., MacHugh D.E., Bradley D.G., Sharp P.M. & Cunningham P. (1994) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 2757–61.

Lunstra D.D. & Cundiff L.V. (2003) Growth and pubertal development in Brahman-, Boran-, Tuli-, Belgian Blue-, Hereford- and Angus-sired F_1 bulls. *Journal of Animal Science* **81,** 1414–23.

MacEachern S., McEwan J., McCulloch A., Mather A., Savin K. & Goddard M. (2009a) Molecular evolution of the Bovini tribe (Bovidae, Bovinae): is there evidence of rapid evolution or reduced selective constraint in Domestic cattle? *BMC Genomics* **10,** 179.

MacEachern S., Hayes B., McEwan J. & Goddard M. (2009b) An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. *BMC Genomics* **10,** 181.

MacHugh D.E., Shriver M.D., Loftus R.T., Cunningham P. & Bradley D.G. (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146,** 1071–86.

Mukasa-Mugerwa E. (1989) *A Review of Reproductive Performance of Female* Bos Indicus *(zebu) Cattle. ILCA Monograph 6*. ILCA, Addis Ababa, Ethiopia.

Nijman I.J., Otsen M., Verkaar E.L. *et al.* (2003) Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity* **90,** 10–6.

Nordborg M., Hu T.T., Ishino Y. *et al.* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology* **3,** e196.

O'Gorman G.M., Park S.D., Hill E.W. *et al.* (2006) Cytokine mRNA profiling of peripheral blood mononuclear cells from trypanotolerant and trypanosusceptible cattle infected with *Trypanosoma congolense*. *Physiological Genomics* **28,** 53–61.

O'Gorman G.M., Park S.D., Hill E.W., Meade K.G., Coussens P.M., Agaba M., Naessens J., Kemp S.J. & MacHugh D.E. (2009) Transcriptional profiling of cattle infected with *Trypanosoma congolense* highlights gene expression signatures underlying trypanotolerance and trypanosusceptibility. *BMC Genomics* **10,** 207.

Paradis E., Claude J. & Strimmer K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20,** 289–90.

Piper E.K., Jackson L.A., Bagnall N.H., Kongsuwan K.K., Lew A.E. & Jonsson N.N. (2008) Gene expression in the skin of *Bos taurus* and *Bos indicus* cattle infested with the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*. *Veterinary Immunology and Immunopathology* **126,** 110–9.

Piper E.K., Jonsson N.N., Gondro C., Lew-Tabor A.E., Moolhuijzen P., Vance M.E. & Jackson L.A. (2009) Immunological profiles of *Bos taurus* and *Bos indicus* cattle infested with the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*. *Clin Vaccine Immunol* **16,** 1074–86.

Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–59.

Rosenberg N.A., Pritchard J.K., Weber J.L., Cann H.M., Kidd K.K., Zhivotovsky L.A. & Feldman M.W. (2002) Genetic structure of human populations. *Science* **298,** 2381–5.

Szpiech Z.A., Jakobsson M. & Rosenberg N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24,** 2498–504.

Tang K., Thornton K.R. & Stoneking M. (2007) A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology* **5,** e171.

The Bovine Genome Sequencing, Analysis Consortium., Elsik C., Tellman R. & Worley K. (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324,** 522–8.

The Bovine HapMap Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324,** 528–32.

Turner J.W. (1980) Genetic and biological aspects of zebu adaptability. *Journal of Animal Science* **50,** 1201–5.

Upton W., Burrow H.M., Dundon A., Robinson D.L. & Farrell E.B. (2001) CRC breeding program design, measurements and database: methods that underpin CRC research results. *Australian Journal of Experimental Agriculture* **41,** 943–52.

Wang Y.H.R.A., Kemp D., McWilliam S.M., Ingham A., Davis C.A., Moore R.J. & Lehnert S.A. (2007) Gene expression profiling of

Hereford Shorthorn cattle following challenge with *Boophilus microplus* tick larvae. *Australian Journal of Experimental Agriculture* **47**, 1397–407.

Weir B.S. & Cockerham C.C. (1984) Estimating F-Statistics for the analysis of population structure. *Evolution* **38**, 1358–70.

Wiener P., Burton D. & Williams J.L. (2004) Breed relationships and definition in British cattle: a genetic analysis. *Heredity* **93**, 597–602.

Wolcott M.L., Johnston D.J., Barwick S.A. & Burrow H.M. (2006) Genetic correlations of steer growth, fatness and Igf-i with feed intake and efficiency in two tropically adapted genotypes. In: *8th World Congress on Genetics Applied to Livestock Production*. Belo Horizonte, MG, Brazil.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Supplementary methods.

Figure S1 Sensitivity analysis of sample size, *n*, on proportion of monomorphic SNPs.

Figure S2 Sensitivity analysis of sample size, *n*, on percentage of monomorphic SNPs.

Figure S3 Neighbour joining trees showing the relationship between all 13 breeds.

Figure S4 Neighbour-joining trees of all 13 breeds.

Figure S5 Comparison of extended haplotype homozygosity between taurine and zebu cattle breeds.

Figure S6 Relationships between K and STRUCTURE's estimated log likelihood of the data, $L(K)$, the first order rate of change of the likelihood with respect to K, $L'(K)$, the second order rate of change of $L(K)$, $L''(K)$ and the ΔK statistic.

Figure S7 STRUCTURE predictions of the proportions of two ancestral populations ($K = 2$) for 317 individuals belonging to 13 breeds using (from top to bottom).

Table S1 F_{ST} estimated for between all 13 breeds, between taurine and zebu animals and between taurine breeds.

Table S2 SNPs that are private in either Brahman (zebu) or taurine breeds.

Table S3 SNPs within the 14 compound diplotypes.

Table S4 Extended haplotype homozygosity regions with significant evidence for recent positive selection between Zebu and Taurine cattle.

Table S5 Proportion of Zebu or Taurine ancestry as estimated by STRUCTURE for $K = 2$ and correlations of estimated Zebu proportions between those from using all autosomal SNPs and the corresponding SNP subset (row).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.