Detection of Chromosome Segments of Zebu and Taurine Origin and Their Effect on Body Weight

S. Bolormaa[†,*], B. Hayes[*], R. Hawken[†,‡], Y. Zhang[†,§], and M. Goddard[†,*,**]

Introduction

Taurine (Bos *taurus*) and zebu (Bos *indicus*) cattle were domesticated independently in the Near East and in India respectively (Beja-Pereira et al., 2006) and are believed to have diverged 117-275,000 years ago (Bradley et al., 2006). Bos *taurus* cattle were introduced to Australia by European settlers in the end of the 18th century. The main *B. indicus* breed used for beef production in Australia is the Brahman which was imported in the 20th century. The Brahman breed originated in the United States of America in the early 1900s where it was developed from progeny of four Indian cattle breeds with some infusion of British-bred cattle (http://www.brahman.com.au/history.html). Since then crossbreeding between Brahman and *B. taurus* breeds has been used to form several composite breeds such as the Santa Gertrudis. Because this crossbreeding has been recent, large chromosome segments of either *B. taurus* or *B. indicus* origin should be segregating in these composite breeds.

Due to their long separation and different selection pressures, *B. indicus* and *B. taurus* cattle may differ substantially in allele frequency at some genes affecting quantitative traits. If we can detect chromosome segments that derive from zebu or taurine origin, we can estimate their effect on quantitative traits and hence detect quantitative trait loci (QTLs) where there is a large difference in allele frequency between the sub-species. In this paper, we present two methods to determine the origin of a chromosome segment based on 50K SNP data and show that, at some positions in the genome, *B. indicus* and *B. taurus* chromosome segments differ in their effect on live weight or carcase weight.

Materials and methods

SNP data. In total, 53,798 SNPs (50K chip) were genotyped and 50,650 were polymorphic. All genotypes had more than 95% quality scores and the proportion of missing genotypes was less than 2.1%. Missing genotypes and haplotypes were imputed using fastPHASE program (Scheet and Stephens, 2006).

Cattle. Two different phenotype datasets of the Beef CRC cattle were used. The Beef CRCI dataset contained 900 steers that were measured for carcass weight (Table 1). These steers were from 7 different pure breeds of three breed types. Four breeds (Angus, Murray Grey, Shorthorn and Hereford) were *Bos taurus* (Bt), one breed (Brahman) was *Bos indicus* (Bi) and two breeds (Santa Gertrudis and Belmont Red) were Bt×Bi synthetic breeds. The Beef

[*] Department of Primary Industries, Bundoora, VIC, Australia
[†] Cooperative Research Centre for Beef Genetic Technologies, Armidale, NSW, Australia
[‡] CSIRO Livestock Industries, Bioscience Precinct, St. Lucia, QLD, Australia
[§] Animal Genetics and Breeding Unit (AGBU), Armidale, NSW, Australia

[**] University of Melbourne, Parkville, VIC, Australia

CRCII dataset contained 1,456 cows with live weight data. These cows were from either Brahman (Bi) or Bt×Bi composite breed.

Estimating the origin of a chromosome segment. Each chromosome was divided into segments consisting of 11 consecutive SNPs. The phase-known haplotypes were used to calculate the probability that each segment was of Bi origin ('b'). Two different estimates of 'b' for each segment were used: one based on SNP genotypes and one based on haplotypes.

The estimate based on SNPs was: $b = \dfrac{\sum_i \dfrac{(x_i - p_{BT})(p_{BI} - p_{BT})}{p_{BTxBI}(1 - p_{BTxBI})}}{\sum_i \dfrac{(p_{BI} - p_{BT})^2}{p_{BTxBI}(1 - p_{BTxBI})}}$

, where x_i is allele at ith SNP for this animal (scored 0 or 1), p_{Bt} is allele frequency in Bt, p_{Bi} is allele frequency in Bi, and $p_{Bt\times Bi}$ is allele frequency in Bt×Bi animals. If the 'b' value is less than zero, it replaced by 0 and when it is more than one, it replaced by one.

The formula for calculating 'b' from the haplotype at the 11 SNPs in a segment is: $b = \dfrac{p_{Bi_{ij}}}{p_{Bi_{ij}} + p_{Bt_{ij}}}$, where p_{Bt} is frequency of the haplotype in Bt and p_{Bi} is frequency of the haplotype in Bi animals. If $p_{Bt_{ij}} = p_{Bi_{ij}} = 0$, then 'b' for that particular segment was treated as missing.

A principle component analysis of the SNP data was also performed using 'R version 2.9.1' package (R Development Core Team, 2005). The first PC distinguished Bi from Bt breeds and so the score for an individual composite animal on the first principle component is another estimate of the proportion of its genome that is BI in origin. PC analysis was performed for each bovine chromosome 1 to 29. For each chromosome the PC score of an animal was compared with the average 'b' value over the chromosome.

The effect of chromosome segment origin on weight. Using the ASReml software (Gilmour et al., 2002), a genome wide association analysis was performed in which the effect of the origin of each chromosome segment was estimated. The mixed model used was: weight ~ mean + fixed effects + βb + animal + error; where β is the regression of weight on 'probability of Bi origin' and animal and error were fitted as random effects. Weight was live weight in the CRCII data and carcase weight in the CRCI data. Fixed effects were different for the CRCI and CRCII datasets. For CRCI dataset, breed, origin of herds, sex, year of measurement, season, market-weight destination and nutritional treatment were fitted as fixed effects, and age deviation from group mean. Whereas for CRCII data the effects of breed, origin of herds, sire group, cohort, calving month and their first degree interactions were fitted as fixed effects. A separate analysis and significance test was performed for each of 4000 chromosome segments and, therefore, we compared the number of segments that were significant to the number expected by chance using a False Discovery Rate (FDR) (Storey 2002). To validate the finding, the segments that were significant in the CRCII composite animals were assessed on the Brahmans and the CRCI animals.

Results and discussion

Detection of segments of zebu and the taurine origin. The correlation between the two 'b' values (one from SNP genotypes and one from haplotypes) was 0.99. For the rest of the analysis we used the average of the two 'b' values for each segment. The distribution of 'b' values in the Bt, Bi and Bt×Bi cattle is shown in Figure 1. Most 'b' values in the Bt animals are close to zero as expected. For Bi animals most 'b' values are approaching 1.0 but for composite Bt×Bi animals there are some segments with low 'b' values, indicating a Bt origin, and some segments with high 'b' values indicating a Bi origin. The uncertainty of Bt origin indicated by 'b' values in the range 0.1 to 0.4 may be due to the fact that there were 4 Bt breeds combined and allele frequency and haplotype frequency may vary between them. In the Brahman cattle most of the segments had a 'b' value close to 1 indicating little uncertainty, but there were a small number of 'b' values below 0.4. This could be due to the presence of chromosome segments of Bt origin in the Brahman as expected from their history.

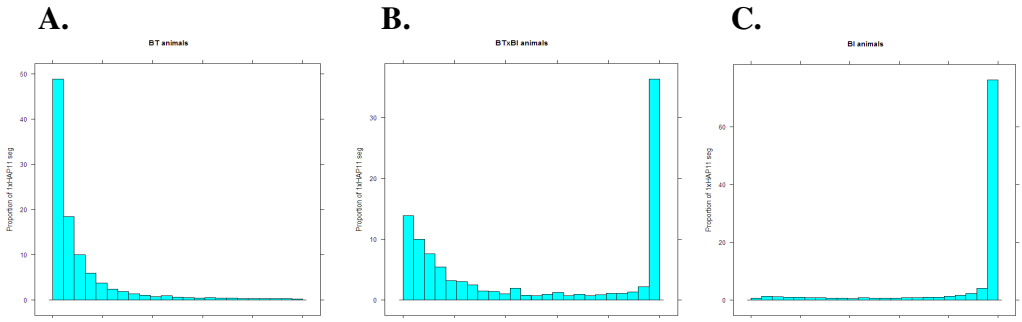

Figure 1: A. Distribution of b values for 11-SNP-SEGs in the taurine animals across bovine genome; B. Distribution of b values 11-SNP-SEGs in crossbred animals across bovine genome; C. Distribution of b values 11-SNP-SEGs in zebu animals across bovine genome

The correlation between the score on the first PC and the average 'b' value over a chromosome ranged from 0.95 to 0.97 across the chromosomes. The PC scores are not independent of the 'b' values because both are based on the SNP genotypes. However, this high correlation supports the interpretation that the 'b' values are estimating the probability that the chromosome segments are of Bi origin.

GWAS using the origin-known segments. Table 1 shows the number of segments that had a significant effect on either live weight or carcase weight. The highest number (26) of significant segments (P<0.001) were detected where the Tropical Composites and Brahmans were combined together because this gives the biggest dataset. When the Bt×Bi CRCII cattle were analysed by themselves only 10 segments were significant and when the CRCII Brahmans were analysed alone only 4 segments were significant. This is perhaps not surprising because within the Brahmans nearly all segments are of Bi origin. The 67 segments that were significant (P<0.01) in the CRCII Bt×Bi cattle were examined in the

CRCII Brahmans and in all the CRCI cattle. In the CRCII Brahmans, 69% of these segments had an effect in the same direction on weight as it did in the Bt×Bi cattle. Only 7 of these segments were significant in the Brahmans but, among these 7, 86% had an effect in the same direction. In the CRCI cattle 66% of the 67 segments had an effect in the same direction and all four of those among the 67 that were significant in the CRCI analysis. If the significant segments discovered in the Bt×Bi cattle were all false positives, we would expect only 50% of segments to have an effect in the same direction in another group of cattle. Therefore these results confirm the FDRs given in Table 1. That is, some at least of the significant effects are real.

Table 1: Number of significant SNPs and FDR at different P thresholds for weights in CRCI and CRCII cattle

	No.	No. SEGs at threshold of P			FDR at threshold of P		
trait(x)	records	0.001	0.01	0.05	0.001	0.01	0.05
CRCII							
w1WGT-all	1100	26	99	371	0.17	0.44	0.58
w1WGT-BB	585	4	38	232	1.11	1.17	0.95
w1WGT-TC	515	10	67	281	0.44	0.66	0.78
CRCI							
CWT	900	10	54	287	0.44	0.82	0.76

FDR = false discovery rate; P = threshold; w1WGT ="end of wet season 1" weight (kg); CWT = carcass weight; BB = Brahman; TC = Tropical Composites; all = both Tropical Composites and Brahmans combined; No. = number

If the Brahman cattle were pure Bi, one would not expect the 'b' value to have any effect. Therefore, the validation in the Brahmans of effects discovered in the Bt×Bi composites, confirms the finding that the Brahmans contain some chromosome segments of Bt origin. The average of 'b' values across genome for our Brahmans was 0.9 indicating that about 10% of their genes are of taurine origin.

Conclusion

The two proposed formulas both provide good estimates of the origin (Bt or Bi) of chromosome segments in composite animals and in 'pure' Brahmans. These estimates can then be used to detect chromosome segments which carry a QTL where Bt and Bi differ greatly in allele frequency. We detected and verified a small number of such segments affecting body weight.

References

Beja-Pereira, A. et al. (2006). Proc. Nat. Acad. Sci. U.S.A., 103: 8113-8118.

Bradley, D., MacHugh, D., Gunningham, P., and Loftus, R. (1996). Proc. Nat. Acad. Sci. U.S.A., 93: 5131-5135.

Gilmour, A. R., Gogel, B.J., Gullis, B.R., Welham, S.J., and Thompson, R. (2002). Asreml user guide release 1.0. VSN International Ltd. Hemel Hempstead, U.K.

R Development Core Team (2005). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Storey, J. D. (2002). J. R. Stat. Soc., Series B 64:479-498.