Proc. Assoc. Advmt. Anim. Breed. Genet. ***17****: 215-218*

EXTENT OF LINKAGE DISEQUILIBRIUM AND HAPLOTYPE BLOCKS IN BOVINE GENOME USING HIGH DENSITY SNP MARKERS

M. S. Khatkar, K. R. Zenger, J. A. L. Cavanagh, R. J. Hawken, M. Hobbs, W. Barris, P. C. Thomson, F. W. Nicholas , B. Tier and H. W. Raadsma

[1]Co-operative Research Centre for Innovative Dairy Products-CRC IDP, [2]ReproGen – Centre for Advanced Technologies in Animal Genetics and Reproduction, Faculty of Veterinary Science, The University of Sydney, Camden, Australia.[3] CSIRO Livestock Industries, Brisbane, Australia, [4] AGBU, University of New England, Armidale, Australia

SUMMARY

A haplotype block map of the bovine genome was constructed based on the analysis of 15,036 SNPs on 1,000 un-related bulls. This map identified a total of 727 haplotype blocks consisting of three or more SNPs. These blocks have a mean length of 69.7 kb and cover 2.2 % of the length of all autosomes. This analysis suggested that approximately 250,000 SNPs would be required to prepare a complete haplotype block map of the bovine genome with 75,000 tag SNPs

INTRODUCTION

With the availability of high-density maps, linkage disequilibrium (LD, or non-random association between alleles at different loci within a population) has recently become important in the context of fine mapping genes or QTL. Recent studies in human have suggested that the underlying structure of LD of the genome can be described into a series of discrete blocks (coldspots; regions of low recombination) delimited by regions of high recombination (hotspots) (International HapMap Consortium 2005). Construction of haplotype blocks and identification of tag SNPs were found quite powerful in identification of specific markers for association mapping in human (International HapMap Consortium 2005; Hinds *et al*. 2005; Barrett *et al*. 2005).

There is little information available on the high resolution LD and haplotype block structure in cattle. The recent availability of high density SNP markers and high throughput genotyping platforms are now making it possible to conduct such studies in cattle. As part of an ongoing gene discovery programme, this study was carried out to study the pattern of LD and haplotype block structure over all of the autosomes (BTA1-29) in Australian Holstein-Friesian dairy cattle.

MATERIAL AND METHODS

Description of DNA samples. A panel of 1,546 Holstein Friesian bulls born between 1955 and 2001 was selected for genotyping. Most of these bulls were born in Australia and more bulls were from recent cohorts than from older cohorts. Based on the coefficient of pair-wise co-ancestry between bulls, the least-related 1,000 bulls were chosen for this analysis from the original 1,546 bulls. The coefficient of co-ancestry among these bulls is 0.012. Genomic DNA was obtained from semen samples sourced through Genetics Australia (Bacchus Marsh, Vic, Australia), and from international contributors. As the yields of some genomic DNA per straw were limited, all DNA samples were amplified using a Whole Genome Amplification (WGA) kit (Repli-G, Molecular Staging Inc. USA) and all genotyping was carried out using WGA DNA (HAWKEN *et al.* 2006).

Description of SNP discovery, genotyping and mapping. A genome-wide high density panel of 15,036 SNPs were used for genotyping using a high-throughput SNP assay service provided by Affymetrix, Inc. (see Raadsma *et al.* for details on the SNPs and Zenger *et al.* for genotyping platform*).* The locations of the SNPs were determined on the bovine sequence assembly Btau 3.1 (ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/fasta/Btau20060815-freeze/). The SNPs were placed on chromosomal linearised scaffolds using sequence similarity. The positions for all the 15,036 genotyping assays on this sequence map could be estimated. However, only 13,705 SNPs were placed on sequence scaffolds which have been assigned to a real chromosome; the rest (1,331 SNPs) were on chromosomally unanchored scaffolds.

Identification of haplotype block and tag SNPs. The haplotype blocks on the bovine autosomes were identified as per the definition of Gabriel et al. (2002) and implemented in Haploview software (Barrett et al. 2005). Under this default algorithm 95% confidence bounds on D' are generated and a block is created if 95% of informative comparisons are in "strong LD" (D' confidence interval bound: 0.7-0.98). The most useful SNPs (tag SNPs) were also identified with the Tagger (de Bakker et al. 2005) option implemented in Haploview.

RESULTS AND DISCUSSION

Of the 15,036 SNPs genotyped, 87 % were polymorphic (minor allele frequency (MAF) > 0) in the 1,000 bulls finally included in this study. The SNPs with MAF <0.05, showing departure from Hardy-Weinberg Equilibrium and typed in less than 50 % of animals were excluded for this analysis. Of the remaining SNPs, 9,195 were able to be located on autosomes in the bovine sequence assembly Btau 3.1 and were included in the present analysis. Of these, 7,057 of SNPs are from the MegAllele 10k SNP panel and 2,138 from the Dairy CRC custom SNP panel. The average inter-marker spacing for the entire genome was 251.8 kb. There were 59 inter-marker intervals above 2 Mb and only 5 intervals above 3 Mb. The overall MAF of the SNPs used in the present analyses was 0.286.

Figure 1 shows the mean D' and r^2 computed between all the syntenic pairs of SNP markers pooled over all the autosomes in the different distance categories. This figure indicates that there is decay in LD with increasing distance. The mean LD over long distance (> 50Mb) is similar to the mean LD between non-syntenic markers. The decline in the value of r^2 is much steeper as compared to D'. On an average r^2 declined to 0.3 within a distance of 50 kb.

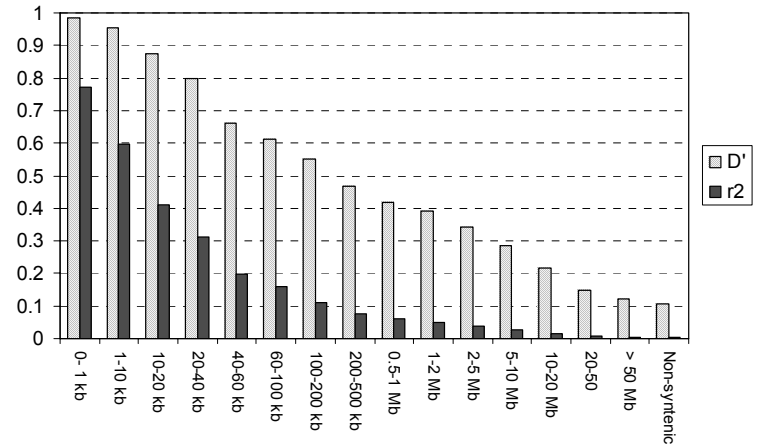

Figure 1. Mean linkage disequilibrium measures (*D'* and r^2) among syntenic SNP pairs over different map distance bins, pooled over all autosomes. 'Non-syntenic' is average LD among a random sample of non-syntenic SNPs pairs.

A total of 727 haplotype blocks consisting of three or more SNPs were identified. There were an additional 1,068 blocks consisting of only two adjacent SNPs, giving a grand total of 1,795 blocks. Figure 2 shows haplotype blocks on a portion of one chromosome as an example of identification of haplotype blocks. The average block length was 69.7 ± 7.7 kb for the blocks consisting of 3 or more SNPs, which is roughly 5 to 10 times larger than in humans. These blocks comprised a total of 2,964 SNPs, and covered 50,638 kb of the sequence map, which constitutes 2.2 % of the length of all autosomes. A set of 1,552 tag SNPs, which will be useful for further fine-mapping studies, has been identified within these blocks. SNPs in most of the bovine genome (~ 98 %) are still present as singletons with no significant LD with adjacent SNPs to form haplotype blocks, suggesting more SNPs are required. Overall, the results suggest that as many as 75-100k tag SNPs would be needed to track all important haplotype blocks in the bovine genome. This would require approximately 250,000 SNPs in the discovery phase.

CONCLUSIONS

This analysis indicates that there are a number of regions in the bovine genome where SNPs in this dataset are present in the form of blocks of high LD. These blocks have been identified when the SNPs are very close to each. The haplotype blocks constructed from the present markers provide very little coverage of the chromosomes. Additional markers and LD analyses are required for identification of tag SNPs for whole-genome population-wide LD studies in cattle. Selection of tag SNPs is important to represent the variability within haplotype blocks for efficient association mapping studies.

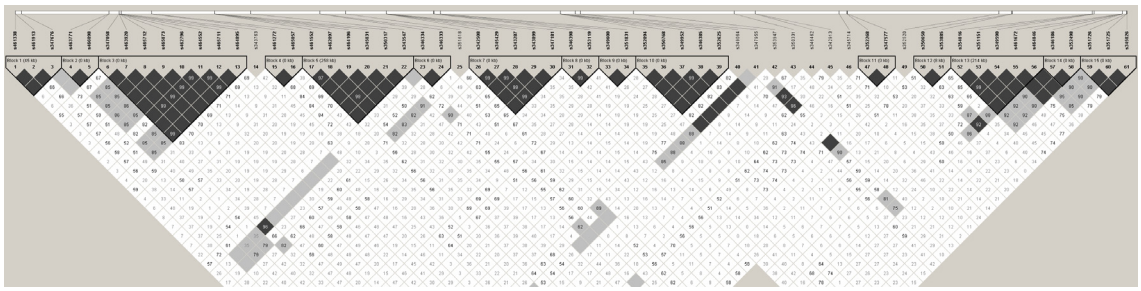

Figure 2. **Haplotype block map of a segment of BTA1 (0.4- 8.3 Mb) in the form of a heatmap of confidence bounds of *D′*. This figure was prepared by Haploview software. Here dark grey colour indicates strong LD, light grey uninformative and white suggests strong evidence of historical recombination.**

ACKNOWLEDGMENTS

We thank Genetics Australia for semen samples, ADHIS for pedigree data. This research is supported by the Cooperative Research Centre for Innovative Dairy Products, Victoria Australia.

REFERENCES

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005) *Bioinformatics* **21**: 263.

de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J., and Altshuler, D. (2005) *Nat Genet* **37**: 1217.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) *Science* **296**: 2225.

Hawken, R.J., Cavanagh, J.A., Meadows, J.R., Khatkar, M.S., Husaini, Y., Zenger, K.R., McClintock, S., McClintock, A.E., and Raadsma, H.W. (2006) *J Dairy Sci* **89**: 2217.

Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A., and Cox, D.R. (2005) *Science* **307**: 1072.

International HapMap Consortium (2005) *Nature* **437**: 1299.

Raadsma, H.W., Zenger, K.R., Khatkar, M.S., Crump, R., Moser, G., Solkner, J. , Cavanagh, J.A.L., Hawken, R.J., Hobbs, M., Barris, W., Nicholas, F.W. and Tier, B. (2007) *Proc. Assoc. Advmt. Anim. Breed. Genet. **17**: 231.*

Zenger, K.R., Khatkar, M.S., Tier, B., Hobbs, M., Cavanagh, J.A.L., Solkner, J., Hawken, R.J., Barris, W. and Raadsma, H.W. (2007) *Proc. Assoc. Advmt. Anim. Breed. Genet. **17**: 123.*