Proc. Assoc. Advmt. Anim. Breed. Genet. 18:247-250

GENOME STRUCTURE IN AUSTRALIAN HOLSTEIN FRIESIAN CATTLE REVEALED BY COMBINED ANALYSIS OF THREE HIGH DENSITY SNP PANELS

M.S. Khatkar[1,2], B. Tier[1,4], M. Hobbs[1,2], D. Khatkar[1,2], J.A.L. Cavanagh[2], R. Crump[1,4], G. Moser[1,3] and H. W. Raadsma[1,2]

[1]The CRC for Innovative Dairy Products, [2]ReproGen, Faculty of Veterinary Science, The University of Sydney, Camden, Australia, [3]Bellbowrie, QLD 4070, Australia, [4] Animal Breeding and Genetics Unit, UNE, Armidale, NSW 2351, Australia.

SUMMARY

We genotyped overlapping samples of Australian dairy bulls using three different SNP chips (15k, 25k and 54k). These chips have different but complementary coverage hence increasing the number of animals and the density and coverage of SNPs to 74k in a combined dataset. A combined analysis of the data from these three SNP chips showed a four fold increase in the coverage of the genome by haplotype blocks over bovine hapmap reported previously (Khatkar *et al.* 2007). An analysis of contiguous runs of homozygosity revealed long stretches (up to 49.39 Mb) of homozygosity on chromosome 1 in many bulls. Distribution of these segments of homozygosity in a sample of bulls is presented. The results for one chromosome are described in detail.

INTRODUCTION

The success of genomic selection (GS) and genome wide association studies (GWAS) largely depends on the number of animals in the design and density of markers in the genome screen. The latter depends on the genome structure in terms of the extent and variation in linkage disequilibrium in the population. With the availability of high density and low cost SNP chips, it is now possible to genotype large numbers of animals with high density SNP markers. Such initial datasets generated in cattle have enabled the elucidation of haplotype block structure (Khatkar *et al.* 2007), selection signatures (Hayes *et al.* 2008), extent of homozygosity (MacLeod *et al.* 2007) and the extent of linkage disequilibrium (McKay *et al.* 2007; Khatkar *et al.* 2008; Sargolzaei *et al.* 2008), in the bovine genome. These studies recommended the requirement of higher density SNP markers for association mapping and genome structure analysis. Due to sparse and unequal SNP coverage it was not possible to explore fine scale LD in the bovine genome.

Recently, there has been a significant and rapid increase in density of SNPs available on diverse sources of bovine SNP genotyping arrays. We genotyped three different SNP chips in overlapping samples of Australian dairy bulls. This combined dataset provides higher density SNP coverage compared to any individual SNP chip presently available. Here we present the results from the analysis of this combined dataset with detailed results focused on chromosome 1.

MATERIALS AND METHODS

Genotypic data. Data from three SNP chips viz. 15k, 25k and 54k genotyped for 1536, 441 and 377 Australian Holstein-Friesian (HF) bulls, respectively, were combined into a single dataset. This combined dataset has 73,569 unique SNPs and 1,943 bulls. After excluding the SNPs with low (<0.05) minor allelic frequency (MAF) and deviation (P<0.0001) from Hardy Weinberg Equilibrium (HWE), 3,102 SNPs on chromosome 1 (BTA1) were used for the present analysis.

Construction of Haplotype block map. Haplotype blocks were identified using Haploview software (Barrett *et al.* 2005) and are defined as detailed in our previous bovine HapMap study (Khatkar *et al.* 2007).

Detection of runs of homozygosity (ROH). A run of homozygosity was defined as a contiguous segment of homozygosity The ROH was detected across the chromosome by inspecting the genotype of an animal starting from the beginning of the chromosome. The segment ended when a heterozygous SNP was encountered and new segment was started at the next homozygous SNP genotype. The only segments consisting of a minimum 6 SNPs and a minimum length of 100 kb were counted as runs of homozygosity. These thresholds were applied to filter out numerous, small segments resulting from local LD.

RESULTS AND DISCUSSION

Haplotype blocks. A total of 3,102 SNPs located on BTA1 were used to construct the haplotype block map of chromosome 1. The individual SNPs are well distributed across the range of MAF (0.05-0.5) (Figure 1a). The mean spacing between SNPs is 51.9 ± 9.7 kb. Figure 1b shows the distribution of spacing between adjacent SNPs which is fairly even. There are only 3 gaps of more than 500 kb between consecutive SNPs. Figure 2 presents the haplotype block map of chromosome 1. The map contains 298 haplotype blocks consisting of more than 2 SNPs. The mean length of these haplotype blocks is 113.4± 9.18 kb. Figure 1c shows that most blocks are relatively short in length (< 100 kb). These blocks cover 11.3 % of BTA1. This is a four fold increase in the coverage as compared to the earlier bovine HapMap of BTA1 constructed based on 528 SNPs (Khatkar *et al.* 2007)). Pair-wise tag analysis selected 81 % of the 3,102 SNPs as tag SNPs, suggesting limited redundancy, at this density of SNP coverage.

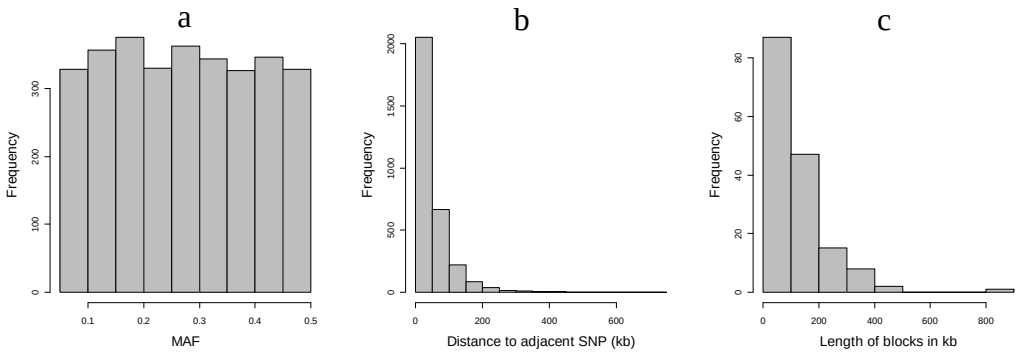

Figure 1. Distribution of a) MAF of 3102 SNPs b) SNP spacing c) length of haplotype blocks on bovine chromosome 1.

Runs of homozygosity. On an average there were 73.2 ± 0.58 (range from 32 to 119) runs of homozygosity on chromosome 1, when a threshold of a minimum length of 100 kb and at least 6 SNPs in a segment was applied. These ROH, on average, cover 28.3 % of the chromosome. The number and coverage of ROH, applying different thresholds for minimum segment length, are presented in Table 1. Increasing the minimum length of segment from 500kb to 5Mb showed a rapid decline in mean number of ROH across the chromosome suggesting that longer ROH occur with low frequency. The longest ROH on chromosome 1 is 49.4 Mb spanning across 861 SNPs. In summary, there is high level of homozygosity in Australian HF bulls. Figure 3 shows the variation

in the proportion of bulls, carrying ROH longer than 500kb, along the chromosome 1 with highest peak in the region ranging from 84Mb to 87 Mb. The figure clearly shows a non-uniform distribution of regions of homozygosity and regions showing strong signs of loss of heterozygosity. To what extend this is a function of population structure and/or influence of selection, remains unclear.

Figure 2. Heatmap of haplotype structure of BTA1 showing confidence bounds of D'. Dark grey colour indicates strong evidence of LD, light grey uninformative and white suggests strong evidence of recombination.

Table 1. Regions of Homozygosity on chromosome 1 in 377 HF bulls.

Threshold for minimum length	Number of contiguous homozygous segments				% of chromosome			
	Mean	SE	min	Max	mean	SE	min	max
100 Kb	73.2	0.58	32	119	28.3	0.42	11	66
250 Kb	55.9	0.46	26	99	26.3	0.42	10.3	64.2
500 Kb	25.9	0.3	11	56	19.5	0.45	8.5	60.3
1 Mb	7	0.14	2	19	11.6	0.49	1.7	56.4
2 Mb	2	0.08	0	10	7.6	0.5	0	51.3
5 Mb	0.6	0.05	0	5	5	0.48	0	51.3

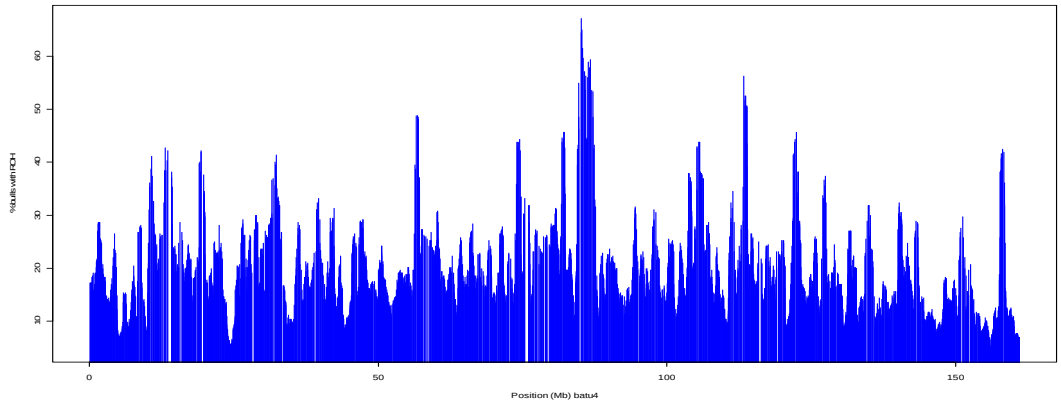

Figure 3. Percentage of bulls carrying runs of homozygosity larger than 500 Kb counted at different points along the chromosome 1.

Animal genomes

Recently, large numbers of structural variations (SVs) including insertion-deletions, , translocations/inversions and copy number variations have been discovered in normal human individuals (Redon *et al.* 2006; Frazer *et al.* 2009). Many of these variations were found associated with phenotypic differences and disease status (Smith 2009). The SNP panels were unable to tag all of the known SVs. It is expected that these types of structural variations will be present in the bovine genome as well. There were 5 % of SNPs with MAF above 0.01, which deviated from HWE in the present study. Some of these deviations from HWE may arise from difficulties in calling the genotype due to the structural variations mentioned above. It is possible to use SNP genotypic data to discover some of structural variations (Simon-Sanchez *et al.* 2007), but it requires very high density SNP panels. However, with the availability of cheaper high throughput whole genome sequencing technology (Kidd *et al.* 2008), it will be possible to discover these structural variations even in livestock species. Studying these new types of DNA variants along with high density SNPs will provide deeper insight into the complex relationship of genotype with phenotype.

ACKNOWLEDGMENTS

The genotype data was provided by bovine genome resources from the Co-operative Research Centre for Innovative Dairy Products.

REFERENCES

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

Frazer, K.A., Murray, S.S., Schork, N.J. and Topol, E.J. (2009) *Nat Rev Genet* **10**:241.

Hayes, B.J., Lien, S., Nilsen, H., Olsen, H.G., Berg, P., Maceachern, S., Potter, S. and Meuwissen, T.H. (2008) *Anim Genet* **39:**105.

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A., Barris, W., Schnabel, R.D., Taylor, J.F. and Raadsma, H.W. (2008) *BMC Genomics* **9:**187.

Khatkar, M.S., Zenger, K.R., Hobbs, M., Hawken, R.J., Cavanagh, J.A., Barris, W., McClintock, A.E., McClintock, S., Thomson, P.C., Tier, B., Nicholas, F.W. and Raadsma, H.W. (2007) *Genetics* **176:**763.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N.A., Tsang, P., Newman, T.L., Tuzun, E., Cheng, Z., Ebling, H.M., Tusneem, N., David, R., et al. (2008) *Nature* **453**:56

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Dias Neto, E., Gill, C.A., Gao, C., Mannen, H., Stothard, P., Wang, Z., Van Tassell, C.P., Williams, J.L., Taylor, J.F. and Moore, S.S. (2007) *BMC Genet* **8**:74.

MacLeod, I.M., Hayes, B.J., Chamberlain, A.J., Savin, K.W., McPartlan,H.,Haile-Mariam, M., Bowman, P. and Goddard, M.E.(2007) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17:** 549.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., et al. (2006) *Nature* **444**:444

Sargolzaei, M., Schenkel, F.S., Jansen, G.B. and Schaeffer, L.R. (2008). J Dairy Sci 91. 2106-17.

Simon-Sanchez, J., Scholz, S., Fung, H.C., Matarin, M., Hernandez, D., Gibbs, J.R., Britton, A., de Vrieze, F.W., Peckham, E., Gwinn-Hardy, K., Crawley, A., Keen, J.C., Nash, J., Borgaonkar, D., Hardy, J. and Singleton, A. (2007) Hum Mol Genet **16:**1

Smith, M. (2009). *Ann N Y Acad Sci* **1151:**1