Proc. Assoc. Advmt. Anim. Breed. Genet. ***17****: 549-552*

USE OF DENSE DNA MARKERS TO MEASURE CHROMOSOME SEGMENT HOMOZYGOSITY IN DAIRY CATTLE

I. M. MacLeod[1,2], B. J. Hayes[1], A. J. Chamberlain[1], K. W. Savin[1], H. McPartlan[1], M. Haile-Mariam[1], P. Bowman[1] and M. E. Goddard[1,2]

[1] Victorian Department of Primary Industries, Attwood, Melbourne, VIC, Australia
[2] Faculty of Land & Food Resources, University of Melbourne, VIC, Australia.

SUMMARY

In livestock it is now possible to use dense genome wide DNA markers to search for favourable mutations, which are in linkage disequilibrium (LD) with one or more of the DNA markers. LD arises because small segments of chromosomes in different animals are inherited from a common ancestor. Knowledge of the length of homozygous segments, and extent of LD, is important in design and analysis of studies searching for favourable mutations. This is a preliminary study of existing chromosome segment homozygosity in outbred cattle. We find the distribution of homozygous segment lengths is exponential. Results show a positive correlation between the level of inbreeding and the average length of segments. Clusters of longer than average homozygous segments in some chromosome regions, may have arisen as a result of recent selection for favourable mutations. Results indicate that the DNA markers are reasonably accurately mapped on the bovine genome (build 3.1), but in small regions, may be more accurately fine mapped on the human genome.

INTRODUCTION

Recent advances in molecular genetics have enabled individual animal genotyping for thousands of DNA markers simultaneously; technology often termed a SNP (single nucleotide polymorphism) chip. With dense markers we can exploit the knowledge that unrelated animals (in recent pedigree) are likely to share small homozygous segments of chromosomes, inherited from a common ancestor many generations ago; ie. segments are identical by descent (IBD). The loci on IBD segments are said to be in linkage disequilibrium (LD). SNP chip data are used to search for favourable mutations (quantitative trait loci – QTL), which are in LD with one or more of the DNA markers. Knowledge of distribution of IBD segments and LD is important for effective QTL studies.

Generally, simulations and analytical methods are used to model the probability of IBD and extent of LD (eg. Meuwissen and Goddard 2001) between unrelated animals (in recent pedigree). However, models rely on assumptions about population size, recombination and mutation rate. Only with the development of SNP chips, is it possible to genotype markers densely enough across the genome to evaluate the existing genome-wide homozygosity patterns in outbred livestock populations. We can measure haplotype homozygosity between SNPs, as an indirect measure of IBD segments or chromosome segment homozygosity (Hayes *et al.* 2003). We use SNP chip genotypes from outbred cattle to measure haplotype homozygosity within animals, as an indirect indicator of IBD segments segregating in the population. It provides an upper bound only on IBD segment length distribution, because we can not differentiate IBD segments from others which are identical at all observed markers, but carry unobserved recombinations or mutations. We use the results to:

1. check that real data is similar to predictions of LD and segment homozygosity

2. investigate the relationship of homozygous segment length with inbreeding.
3. indicate the optimal window of segment length for multiple marker IBD methods and multi-marker haplotypes to test for the presence of QTL.
4. check for indications of selected (natural or artificial) regions within chromosomes.
5. give an indication of the likely accuracy of the SNP locations on the bovine genome.

MATERIALS AND METHODS

Three hundred and eighty two Holstein bulls were genotyped using the Affymetrix 10 000 SNP chip. The position of 9918 typed SNPs was checked using BlastN software, and we were able to position (Mb units) 7705 SNPs on the bovine genome, build 3.1, and 8612 on the human genome. The genotyped SNPs are genome-wide, with an average spacing of approximately 0.3Mb

Using the ordered SNP genotypes, we calculated the base pair length of homozygous segments within animals. We define a "segment" as 2 or more adjacent SNP being homozygous within an animal. Only bulls are genotyped, so the X chromosome data was discarded. We used 2 methods to calculate homozygous segment length; the first ignores missing SNP data (4.5% of total), so that if the locus on either side of a missing value is homozygous, the missing value is also assumed homozygous. The second method only counts a segment as homozygous if the missing values are flanked by 3 homozygous loci on either side. The results are similar with either method, so we report results for the first method only. Inbreeding in each animal was estimated from pedigree data to give the traditional F coefficient (PEDIG software - Boichard 2002), and additionally, by directly calculating the single locus homozygosity for each animal. The 10 most inbred (F=0.07 to 0.13) and the 9 least inbred animals (F<0.02) were used to compare homozygous segment length distributions. These 2 groups will be referred to as "most" and "least" inbred groups.

RESULTS AND DISCUSSION

Distribution of homozygous segment lengths. The bovine positioned SNPs covered a total physical distance of 2300Mb, and on average the homozygous segments covered 49% of this region. Excluding the most inbred animals, the average number of homozygous segments per animal was 848 (approx. 29 per chromosome) with an average length of 1.32Mb. The distribution of homozygous segment lengths from the most inbred animals is shown in Figure 1. The distribution is exponential as predicted from analytical models and simulations (Hayes *et al.* 2003; Stam 1980). The inset in Figure 1 shows the tail of the distribution enlarged, and for comparison there is a second inset showing the distribution of segments from 9 least inbred animals (F<0.02). The insets show that the most inbred animals have a higher proportion of homozygous segments in the tail of the distribution compared to the least inbred. The 2 distributions are otherwise similar in shape.

The 2 most inbred animals have an F coefficient of approximately 0.125; ie. 12.5% probability that alleles on chromosome pairs carried by these animals are IBD. The sire and dam of these 2 animals share a common parent, so chromosome segments can be inherited from a common ancestor only 2 generations back. Assuming on average the chromosomes are 100cM long, the chance of a crossover event (recombination) per chromosome, per generation is about once. Therefore, some homozygous segments should be at least as long as one quarter the entire chromosome length. In these 2 animals (F≥0.125), there are some segments between 20 to 95Mb long. The total length of homozygous segments greater than 20Mb represents 11% of the total genome in both animals.

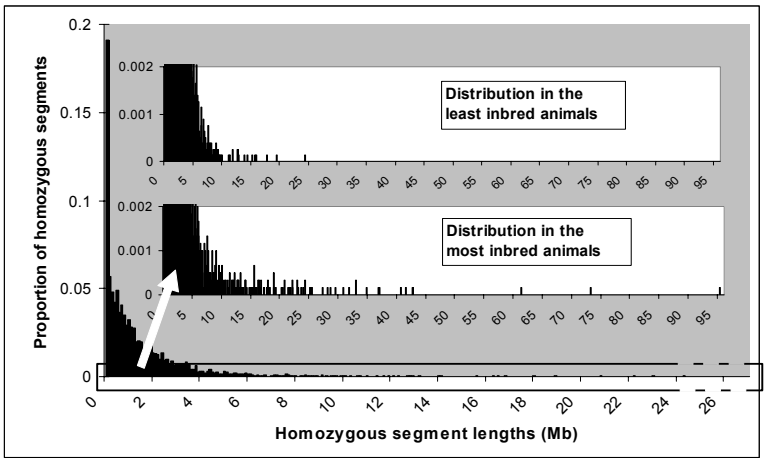

Figure 1. Distribution of the length of homozygous segments in the 10 most inbred animals (F=0.07 to F=0.13). Insets show the enlarged tail of this distribution up to 95Mb segment lengths for most vs. least inbred animals.

Inbreeding. The correlation between F coefficients and average length of homozygous segments was 0.5. The correlation between the more direct inbreeding measure of single locus homozygosity and average homozygous segment length was 0.83 (bovine genome positioned SNPs) and 0.77 (human genome positioned SNPs). Positive correlations indicate that as inbreeding levels increase, so too does the average length of homozygous segments. This is expected given the inbreeding coefficient predicts the probability of co-ancestry of alleles on chromosome pairs within animals. The lower correlation using the F coefficient compared to using single locus homozygosity, indicates that the recorded pedigree is not adequately estimating the full extent of inbreeding, because base animals in the pedigree are assumed to have F=0. The correlation between the F coefficient and the single locus homozygosity was 0.4, also suggesting under-estimation of F coefficients. The high correlation between average homozygous lengths and single locus homozygosity confirms that ancestral population size strongly influences expected segment homozygosity (Hayes *et al.* 2003).

Multi-marker bracket size. Across all 382 animals, 78% of segments are less than 2Mb long and 96% of segments are less than 5Mb long. This suggests that for QTL analysis using multi-marker approaches, marker windows encompassing more than 5Mb may yield very little extra information. However, as noted above, the observed distribution of homozygous segment lengths may have been affected by inaccuracies of SNP positions. If significant numbers of SNP are inaccurately positioned on the bovine map, some segments may be observed as shorter than their true length. However, it is also likely that IBD segment length is generally shorter than these observed homozygous segments, because mutation or recombination events may be missed between the observed SNP markers.

Selection. Figure 2 shows the average length of homozygous segments with their mid-points at 0.5Mb intervals along chromosome 20. It is possible that consistent selection pressure for improved milk production over many generations could result in more conserved areas of the chromosome around loci which affect milk production. Figure 2 shows a cluster of longer than average homozygous segments (circled) with mid-points around 27.8 to 31.8Mb (excluding most inbred

animals). The segments overlap with the position of a mutation in the growth hormone receptor gene (around 31Mb) shown to have a large effect on milk production (Blott *et al.* 2003).

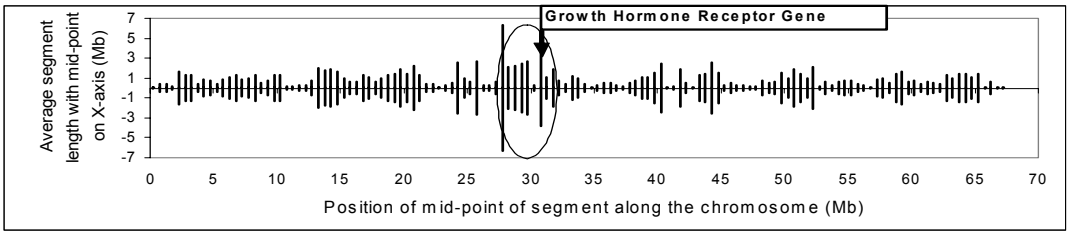

Figure 2. Average segment lengths at 0.5Mb intervals (excluding most inbred animals) across chromosome 20. The circled longer segments overlap the GHR gene region.

Accuracy of SNP positions. The results are a good indication that most SNPs are located on the correct chromosome, and SNP positions within chromosome regions are moderately good. If positions were poorly mapped we would not expect to find the long segments in the 2 most inbred animals, and the correlation between single locus homozygosity and homozygous segment length would be lower. The long homozygous segments do not however, give a strong indication of whether or not the fine ordering of SNPs within regions is very accurate, because even if a large segment is homozygous, the reordering of the SNPs within this segment will not change this result. There may be an over representation of very short segments ≤0.1Mb (Figure 1) due to wrongly positioned SNPs causing homozygous segments to be broken up, and also because SNPs were often clustered within contigs (average of 27Kb long) followed by gaps of > 0.1Mb. Using human map SNP positions, the large homozygous segments in the 2 most inbred animals were found on homologous human chromosomes. Using human mapping (excluding the most inbred animals), we find 15% fewer, but longer segments, with the same proportion of the genome homozygous as with bovine positions (49%). Also, the average number of SNPs per homozygous segment is 6.2 compared with 5.7 using the bovine map. The higher number of SNPs per segment with human mapping, implies that, within small regions, using human mapping may sometimes be more accurate than on bovine build 3.1.

ACKNOWLEDGMENTS

I. Macleod gratefully acknowledges scholarship assistance from Melbourne University, Meat & Livestock Australia and Department of Primary Industries Victoria. We thank laboratory staff in the molecular genetics group for assistance in sample preparation.

REFERENCES

Blott S., Kim J.J., Moisio S., Schmidt-Kuntzel, A., Cornet, A.,Berzi, P., Cambisano, N., Ford, C., Grisart, B., Johnson, D., Karim, L., Simon, P., Snell, R., Spelman, R., Wong, J., Vilkki, J., Georges, M., Farnir, F., Coppieters, W. (2003) *Genetics* **163**: 253

Boichard D. (2002) 7th World Congress on Genetics Applied to Livestock Prod.

Meuwissen T.H.E. & Goddard M.E. (2001) *Genet Sel Evol* **33**: 605

Hayes B.J., Visscher P.M., McPartlan H.C., Goddard M.E. (2003) *Genome Res.* **13**: 635

Stam P (1980) *Genetical Research* **35**: 131