Proc. Assoc. Advmt. Anim. Breed. Genet. 19:375-378

TOWARDS A COMPREHENSIVE RECOMBINATION MAP OF HOTSPOTS IN SHEEP

C. Massault[1], J. Henshall[2], C. Gondro[1] and J.H.J. van der Werf[1]

[1] School of Environmental and rural science, University of New England, Armidale, NSW 2351
[2] CSIRO, Armidale, NSW 2350

SUMMARY

Recombination rate is positively associated with genetic diversity (Spencer *et al*. 2006). Hotspots are regions with a higher recombination rate than average, which could be indicative of diversity in the genome. We used genotype data from Illumina OvineSNP50 Bead Chip on 3908 sheep, to first infer sire haplotypes and then detect the recombination events for the whole genome. We found on average over the genome 0.033 recombination events per mega base pair (Mb). Variance of recombination events between individuals was large, but much smaller between sire groups. Breed was found to have a highly significant effect on recombination events over the genome. Heritability of recombination events was found to be medium (0.25 ±0.09). A recombination map is a useful tool to better understand the mechanisms of diversity, and this information can be used to have more insight in linkage disequilibrium.

INTRODUCTION

Recombination hotspots are regions of the genome where crossover events occur at a much larger rate than on average. The interest for finding recombination hotspots has been growing in the last few years. A number of studies have reported these regions in various species: yeast (Gerton *et al.* 2000), mice (Wu *et al.* 2010), maize (Tenaillon, *et al.* 2002) and human (Meyers *et al.* 2005). However there is little information in the literature about recombination hotspots in livestock, except for chicken (Groenen *et al.* 2009). In sheep, recombination hotspots were reported for a specific region by analysing variation in a single gene (Hickford *et al.* 2007). A complete map of recombination hotspots for the whole sheep genome would help to characterize highly conserved and regions exhibiting high diversity. Recombination hotspots have been mostly identified in regions that correspond to gene promoters, enhancing the diversity, while coldspots were mostly identified in transcribed regions in human (McVean *et al*. 2004). Finding hotspots and coldspots will help to associate features to recombination events.

Genotype data, based on the Illumina OvineSNP50 Bead Chip, comprising 54,241 SNPs, has made it relatively easy to detect recombination. We used 3908 sheep assays on the 50K chip to identify recombination hot and cold spots across the 26 autosomes.

MATERIAL AND METHODS

Animals. For this study, we used data from the Sheep Genomics Project of 11 Merino sire families and 9 families from various breeds (Dorset, white Suffolk, Border Leicester, Coopworth) born on 2005 and 2006 at the Falkiner Memorial Field Station flock (FMFS). The family sizes ranged between 92 and 389 offspring with an average of 195 offspring per sire, summing to a total of

3908 offspring. Sire and offspring were genotyped using the Illumina 50K ovine SNPchip (Illumina Inc., San Diego, CA, USA), corresponding to 48,641 SNPs, after quality control by filtering SNPs that did not pass quality control metrics and, the removal of unmapped SNPs (4150) and SNPs on the sex chromosomes (1450).

Phasing and detection of recombinations. We inferred paternal haplotypes of sires and offspring using information of the population structure (paternal half-sib families). We could infer haplotypes at a specific SNP either with certainty (PWC – phased with certainty) or phased by linkage (PBL). Offspring haplotypes were recoded according to whether each SNP was inherited from one sire haplotype (0 and 2, when inferred by PWC and PBL, respectively) or the other haplotype (1 and 3, when inferred by PWC and PBL, respectively). We limited our analysis to haplotypes inherited from the sire, because there were a large number of maternal families of small size (1-2 offspring).

We determined a recombination event when two adjacent PWC SNPs changed paternal phased inheritance. Recombinations due to genotyping errors are double recombinations with only a single PWC SNP in the central segment that occurs at the same position for a large number of animals. Recombinations due to map errors are double recombination with few PWC SNP in the central segment. Recombination events were detected for each animal and chromosome by chromosome. Recombinations due to genotyping error or map error were ignored. The recombination rate was normalised per Mb over the genome

$\frac{Nrec * 10^8}{dist * N}$ where N_{rec} is the number of recombinations between 2 specific SNPs, N is the number of animals in the data set and *dist* the distance in base pairs between the two specific recombining SNPs.

RESULTS AND DISCUSSION

Table 1 reports the length in Mb, the number of SNPs, the number of recombinations per animal, the number of recombinations per animal per Mb, variance of number of recombination events between individuals and between sire groups, and number of SNPs with high recombination rate . While the number of recombination events per individual varies between chromosomes, the number of recombination events per animal per Mb is rather stable along the genome (between 0.02 and 0.09, mean of 0.033). The normalised average of recombination rate for 100 Mb could have extremely high values, probably due to some mapping errors that have not been detected. We therefore removed the top ranked 5% of SNPs, lowering the average of 1.2 recombinations per SNPs. The new sheep genome map would remove the undetected error map. Looking at 1% SNPs that recombine the most, the number per chromosome vary between 0 recombination hotspots (chromosomes 26) and 70 hotspots (chromosome 3). There seems to be no relation between chromosome length and the number of hotspots, which suggest that they are not randomly distributed along the genome. Figure 1 illustrates the difference between chromosome 26 with few recombination rates above average and chromosome 1, with a large number of SNPs that have a recombination rate included in the highest 1% rate on the same y-axis scale of normalised count over the genome distance. We can observe that high recombination rate occurs more often in chromosome 1 than in chromosome 26.

The average of recombination over the distance varies a little, while the mean and variance of number of recombinations per individual is directly influenced by the chromosome length. The sire group variation could be indicative that number of recombination events might be a heritable feature as mentioned by Coop *et al.* 2008. Wang and Xu (2005), they found a heritability for recombination rate around 0.5. It is an interesting view that implies that underlying genes control diversity within each individual. However, in our data set, we found heritability of recombination events of 0.25 (±0.09) accounting for breed as a fixed effect ($p < 0.001$) for the whole genome. This estimate has a limited accuracy due to the limited number of sires in the dataset.

Table 1. Chromosome length in Mb, number of SNPs in Mb, average number of recombinations per animal (AVG), normalised average number of recombinations per animal per Mb pair (AVG/Mb), variance of recombination events between individuals (V1), variance of recombination events between sires (V2) per chromosome, number of SNPs with a recombination rate belonging to the top 1 % (500 in total).

Chr	Length (Mb)	# SNPs	AVG	AVG/Mb	V1	V2	Hotspots
1	300	5494	5.3	0.03	9.9	0.08	63
2	263	5111	4.3	0.03	10.2	0.11	61
3	243	4647	5.0	0.03	12.4	0.14	70
4	127	2508	2.3	0.03	3.1	0.17	24
5	116	2199	2.7	0.04	4.6	0.12	20
6	129	2413	2.0	0.02	2.7	0.03	17
7	108	2094	2.0	0.03	3.2	0.13	14
8	98	1916	1.4	0.02	2.2	0.06	25
9	101	1983	1.7	0.03	2.4	0.04	13
10	94	1719	1.1	0.02	1.6	0.03	17
11	67	1104	2.0	0.05	3.8	0.07	14
12	86	1583	1.7	0.03	2.6	0.07	11
13	89	1565	1.7	0.03	3.3	0.08	12
14	69	1094	1.7	0.04	2.6	0.02	10
15	90	1555	1.5	0.03	2.3	0.07	13
16	77	1450	1.5	0.04	2.2	0.02	11
17	79	1320	1.6	0.03	3.1	0.04	13
18	72	1318	1.9	0.04	3.8	0.18	19
19	65	1153	1.5	0.04	2.6	0.12	17
20	56	1050	1.3	0.04	1.9	0.02	9
21	55	825	1.2	0.09	1.9	0.16	8
22	56	1005	1.2	0.08	1.4	0.4	3
23	67	1056	1.5	0.03	1.9	0.04	17
24	45	679	1.1	0.05	1.8	0.04	5
25	48	931	1.5	0.04	3.0	0.06	14
26	50	868	1.3	0.04	1.9	0.02	0

By identifying hotspots in the sheep genome, we can determine genes or regions of the genome that are more subject to recombination or on the contrary not at all. The next step will be to compare our hotspots located in regions with known high recombination rate, such as MHC.

The search for recombination hotspots has different objectives. Unravelling the actual process of recombination hotspots occurrences could lead to understand how diversity is created and why

some regions remain conserved. The identification of recombination hotspots results in an improved genetic map, which can be useful to perform association mapping and fine mapping of specific genes (Hey 2004) that are, e.g. responsible for disease. At last, recombination hotspots could be give further information about expected level of linkage disequilibrium. A region prone to higher recombination rate is an indication of faster linkage disequilibrium decay. This could be useful information when performing genome association studies.

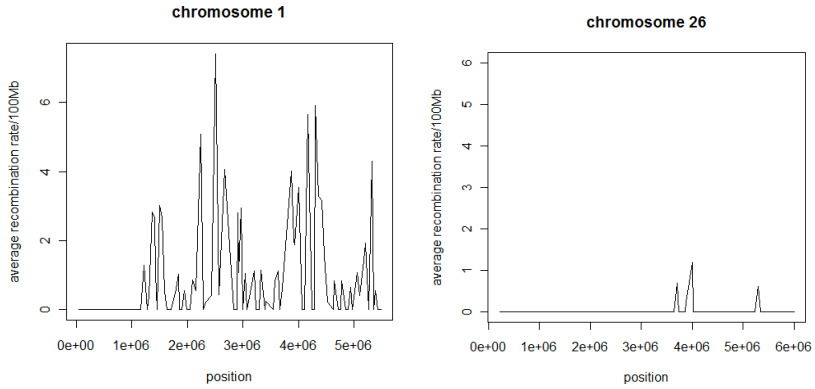

Figure 1. Normalised recombination rate per 100 Mb for chromosome 1 and chromosome 26 on the same scale.

CONCLUSION

This study is a preliminary work to draw a map of hotspot regions in the sheep genome. With the new draft sequence of sheep genome, the map could be refine and further work will be undertaken to find MHC region linked to hotspots identified in this paper.

REFERENCES

Coop G., We X., Ober C., Pritchard J. and Przeworski M. (2008) *Science* **319**: 1395.
Hey J. (2004) *Plos Biology* **2**: 730.
Hickford J, Zhou H. and Fang Q. (2007) *J. Anim. Sc.* **85**:577.
Gerton J., DeRisi J., Shroff R., Lichen M., Brown M. and Petes T. (2000*) PNAS* **97**:11383.
Groenen M., Wahlberg P., Foglio M., Cheng H., Megens H., Crooijmans R., Besnier F., Lathrop M., Muir W., Wong G., Gut I. and Andersson L. (200) *Genome Research* **19**:510.
McVean G., Myers S., Hunt S., Deloukas P., Bentley D. and Donelly P., (2004*) Science* **304**:581.
Meyers S., Bottolo L., Freeman C., McVean G. and Donnelly P. (2005) *Science* **310**: 321.
Spencer C., Deloukas P., Hunt S., Myers S., Silverman B., Donnelly P., Bentley D. And McVean G. (2006) *PLoS Genetics* **2**:e148.
Tenaillon M., Sawkins M., Anderson L., Stack M., Doebley J. and Gaut B. (2002) *Genetics* **162**:1401.
Wang, L. and Xu, X. (2005) BMC Genomics **6**:S58.
Wu Z., Guten I. and Bois P. (2010) *Nucleic Acids Res.* **38**:2346.