

Enhanced mapping tools for the sheep genome - BACs, microsatellites, SNPs and the virtual map

Jill Maddox¹, Brian Dalrymple², Sean McWilliam², Alan McCulloch³, Ewen Kirkness⁴, David Townley², Wes Barris², James Kijas², Abhirami Ratnakumar², John McEwan³, Frank Nicholas⁵ & other members



of the International Sheep Genomics Consortium⁶



¹Department of Veterinary Science, University of Melbourne, Australia; ²CSIRO Livestock Industries, Australia; ³AgResearch, New Zealand; ⁴The Institute for Genomic Research USA; ⁵University of Sydney, Australia/sheepGENOMICS (www.sheepgenomics.com); ⁶www.sheephapmap.org

Background

Sheep play important roles in the production of meat, wool and milk products. They are also used as a model organism for humans for studying the normal development of various physiological systems, as well as abnormal states caused by genetic and acquired diseases. Many of the traits of importance in sheep have been shown by breeding studies to be heritable, and a large number of research groups throughout the world are conducting mapping experiments to identify the loci that influence them. Ideally, there would be a sheep genome sequence map and "cheap" genotyping technology, such as SNP chips, to assist these endeavours. The international sheep genomics community has relatively modest resources, and currently lacks the funding to produce a sheep genome sequence map. In the absence of this, an alternative strategy is being undertaken to produce a virtual sheep genome sequence map and a SNP chip suitable for conducting genome screens.

Mapping Resources

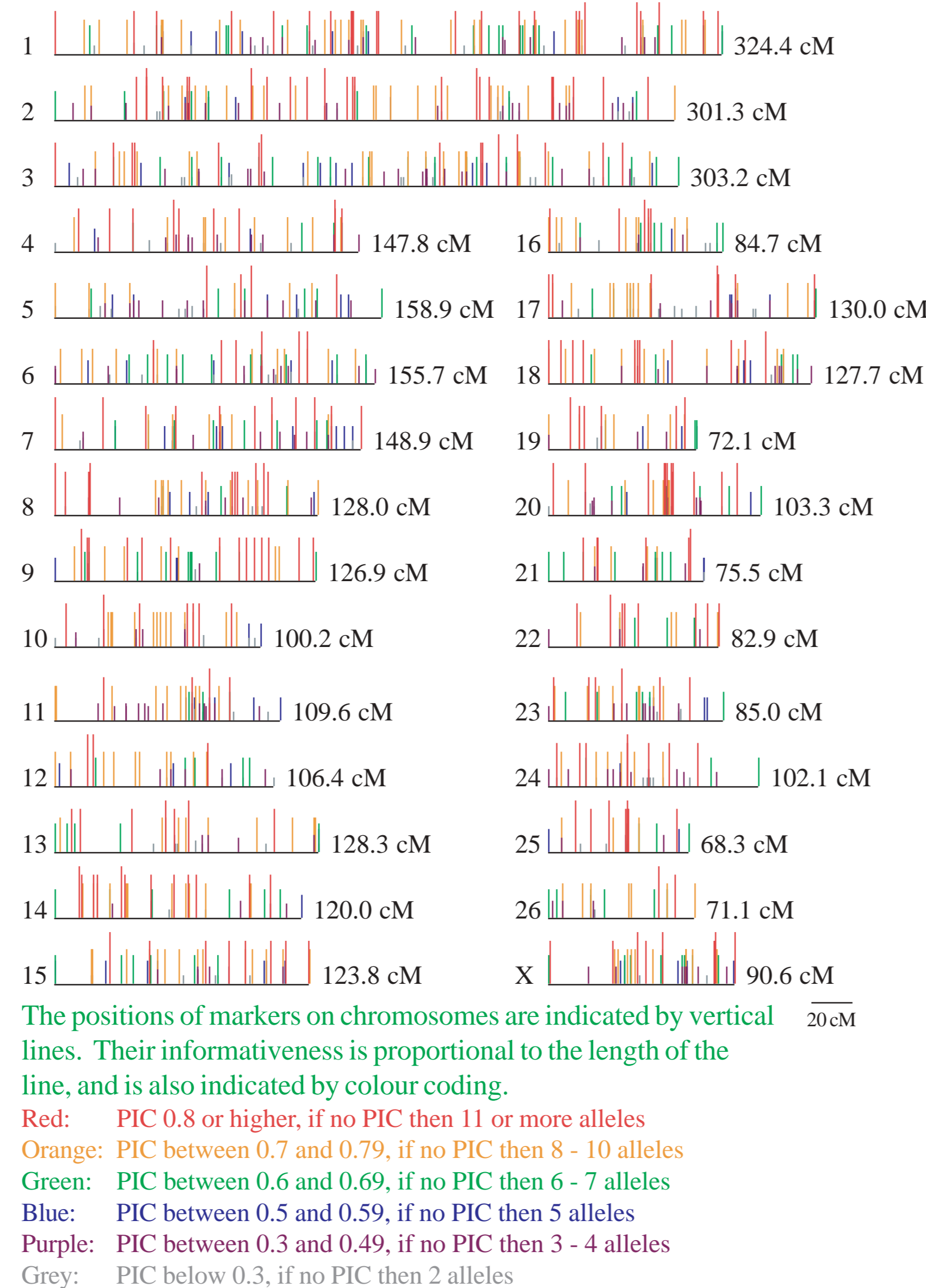
Sheep linkage map

The construction of the sheep linkage map commenced in the early 1990s with the development of the International Mapping Flock by AgResearch. A large number of markers were positioned on the map in the mid to late 1990s, however, the rate of new additions has slowed since then (Figure 1). The current sheep linkage map, version 4.6, has been genotyped for 1,374 markers representing 1,333 loci (Figure 2, Table 1). The map spans ~3,600 cM. One thousand two hundred and nine of the markers (88%) are microsatellites, and 713 (52%) of the markers are based on cattle sequence, 582 (42%) on sheep sequence and 50 (4%) on goat sequence. Further details on the sheep linkage map can be found at rubens.its.unimelb.edu.au/~jillm/jill.htm.

Table 1 Map statistics for SheepMap version 4.6

Chr	Sex Av cM	Female cM	Male cM	Num Markers	Num Loci
1	324.4	290.3	359.3	133	129
2	301.3	280.4	323.4	118	116
3	303.2	289.9	313.2	121	112
4	147.8	131.0	155.5	48	48
5	158.9	158.9	169.2	50	48
6	155.7	132.2	176.9	64	64
7	148.9	139.6	160.9	65	61
8	128.0	124.6	130.4	44	44
9	126.9	117.7	135.5	45	45
10	100.2	96.5	104.2	36	36
11	109.6	103.2	118.3	48	45
12	106.4	88.4	116.6	35	34
13	128.3	128.0	128.7	36	36
14	120.0	97.9	142.1	42	41
15	123.8	101.5	142.8	53	50
16	84.7	77.1	91.4	38	37
17	130.0	113.3	146.6	49	47
18	127.7	109.3	150.1	48	48
19	72.1	65.4	85.0	32	30
20	103.3	89.0	118.8	50	46
21	75.5	65.6	84.9	26	26
22	82.9	60.6	98.7	22	22
23	85.0	73.0	96.8	40	40
24	102.1	99.3	105.3	36	36
25	68.3	56.2	76.6	25	22
26	71.1	57.3	85.9	22	21
Total	3,486.1	3,146.2	3,817.1	1313	1272
X	90.6	131.3	58.6	61	61

Figure 2 Distribution and Informativeness of Markers on Sheep Linkage Map v4.6



12x sheep BAC library

A publicly available 12x sheep BAC library has been constructed from the leukocytes of a Texel ram by Michael Nefedov (CHORI, bacpac.chori.org/library.php?id=162) with funding from AgResearch (NZ), USDA-NRICGP and USDA-ARS (USA) and Meat & Livestock Australia (MLA). Details of the BAC library are shown in Table 2. The library was end sequenced at The Institute for Genome Research with funding from USDA-NRICGP, and sheepGENOMICS (MLA and Australian Wool Innovation). Microsatellites were identified in the BAC end sequences with sputnik, Tandyman, and findpatterns (GCG) (Table 3). The positions of sheep microsatellites that are likely to be more useful for linkage mapping within user defined regions relative to the cattle genome are available via a tool at rubens.its.unimelb.edu.au/~jillm/jill.htm.

The virtual sheep genome

The strategy for constructing the virtual sheep genome is shown diagrammatically in Figure 3. Briefly, blastn and megablast were used to position sheep BAC end sequences relative to the human (hg17), and cattle (Btau_2.0) and dog (canFam2) genomes respectively (Table 4). For each genome comparison, all BACs with overlapping tail to tail matches (matches where both end sequences of the clone relative to a reference sequence have internal 3' ends) with distances between 10-500 kb were used to construct BAC contigs of overlapping clones. These were termed BAC comparative genome contigs (BAC-CGCs). The liftOver utility (genome.ucsc.edu/cgi-bin/hgLiftOver) was then used to transform the coordinates for all match positions to coordinates relative to the human genome, and a new set of BAC-CGCs were produced (termed MegaBAC-CGCs). The MegaBAC-CGC information was then integrated with megablast positional information for markers on the linkage map, and with sheep physical map information to construct virtual sheep chromosomes. The exact position and orientation for many of the MegaBAC-CGCs is still uncertain as many of the MegaBAC-CGCs lack accurate anchoring and orientation information. It is planned to position poorly anchored, or poorly oriented, MegaBAC-CGCs by mapping these on the sheep radiation hybrid and linkage maps.

The virtual sheep chromosomes will be displayed within a GMOD browser at www.livestockgenomics.csiro.au/

Figure 1 Growth of the Sheep Map

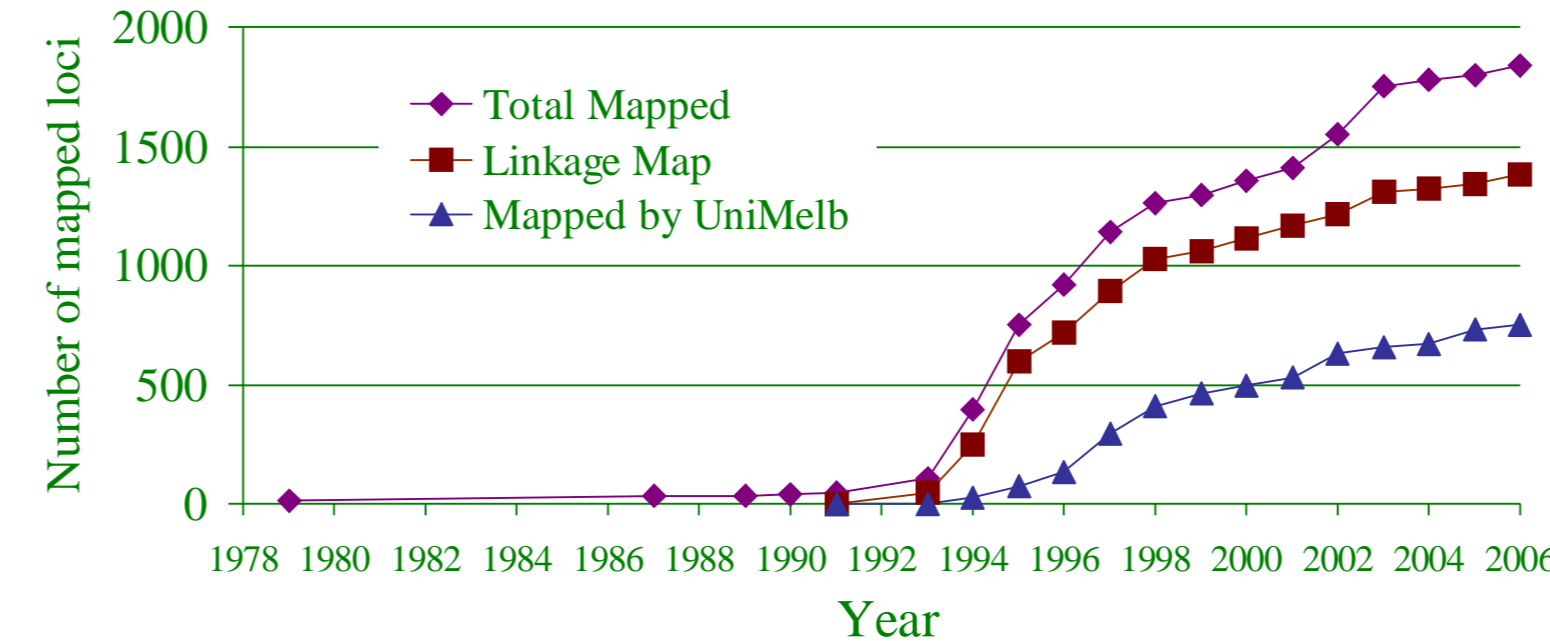


Table 2 Details of the CHORI-243 BAC library

Number of clones	202,752
Average insert size	184 kb
Clones sequencing attempted	199,680 89%
Clones with single end reads	13,546 7%
Clones with no end reads	9,268 4%
Average insert size	687 bp (139-1,010)

Table 3 Microsatellites in sheep BAC end sequences identified by sputnik

Unit	Smallest	Average	Largest	Number
AC	11.0	18.5	127.5	4,080
AG	11.0	18.2	47.0	300
AT	11.0	19.0	63.0	2,031
AAC	8.0	9.4	14.0	70
AAG	8.0	16.9	67.0	26
AAT	8.0	10.1	16.0	62
ACC	8.0	10.7	22.0	24
ACT	8.0	11.3	16.0	11
AGC	8.0	10.4	33.0	818
AGG	8.0	12.0	15.0	8
ATC	8.0	8.8	10.0	5
CCG	8.0	9.5	14.0	4
AAAC	8.0	9.3	12.0	13
AAAG	8.0	19.1	48.0	27
AAAT	8.0	9.0	14.0	55
AATG	8.0	24.3	57.0	3
AATT	12.0	12.0	12.0	1
ACAG	8.0	10.0	13.0	4
ACAT	8.0	9.2	20.0	17
ACCC	8.0	8.0	8.0	2
ACGC	9.0	10.5	12.0	2
ACTC	12.0	12.0	12.0	1
ACTG	8.0	8.0	8.0	2
AGAT	8.0	11.9	21.0	31
AGGC	9.0	11.0	13.0	2
ATCC	8.0	16.4	48.0	16
AAAAC	8.0	9.0	10.0	3
AAAAT	8.0	8.8	10.0	8
AACAC	8.0	8.0	8.0	1
AACTG	8.0	8.9	10.0	15
AAGTG	8.0	8.0	8.0	2
AATAT	8.0	9.0	10.0	2
AATGG	13.0	13.0	13.0	1
CCCCG	30.0	30.0	30.0	1

Smallest, average and largest refer to the number of copies of the repeat unit

Table 4 Construction of sheep BAC-CGCs

Reference genome	Mapped BACs	Number of CGCs	Region Spanned
Bovine	32,602	4,0260	1.59 Gb
Dog	58,757	2,104	2.05 Gb
Human	52,338	2,447	2.35 Gb
Human + Bovine + Dog	84,264	1,172	2.57 Gb

sheep.shtml. The browser will display a range of sheep BAC related and comparative information as well as the positions of mapped sheep and cattle markers, cattle microsatellites, sheep BAC end sequence microsatellites, expressed sequence tag and Refseq sequence, putative genes, SNPs and other features of interest.

It is planned to update the sheep virtual map to incorporate data from Btau_3 by the end of 2006.

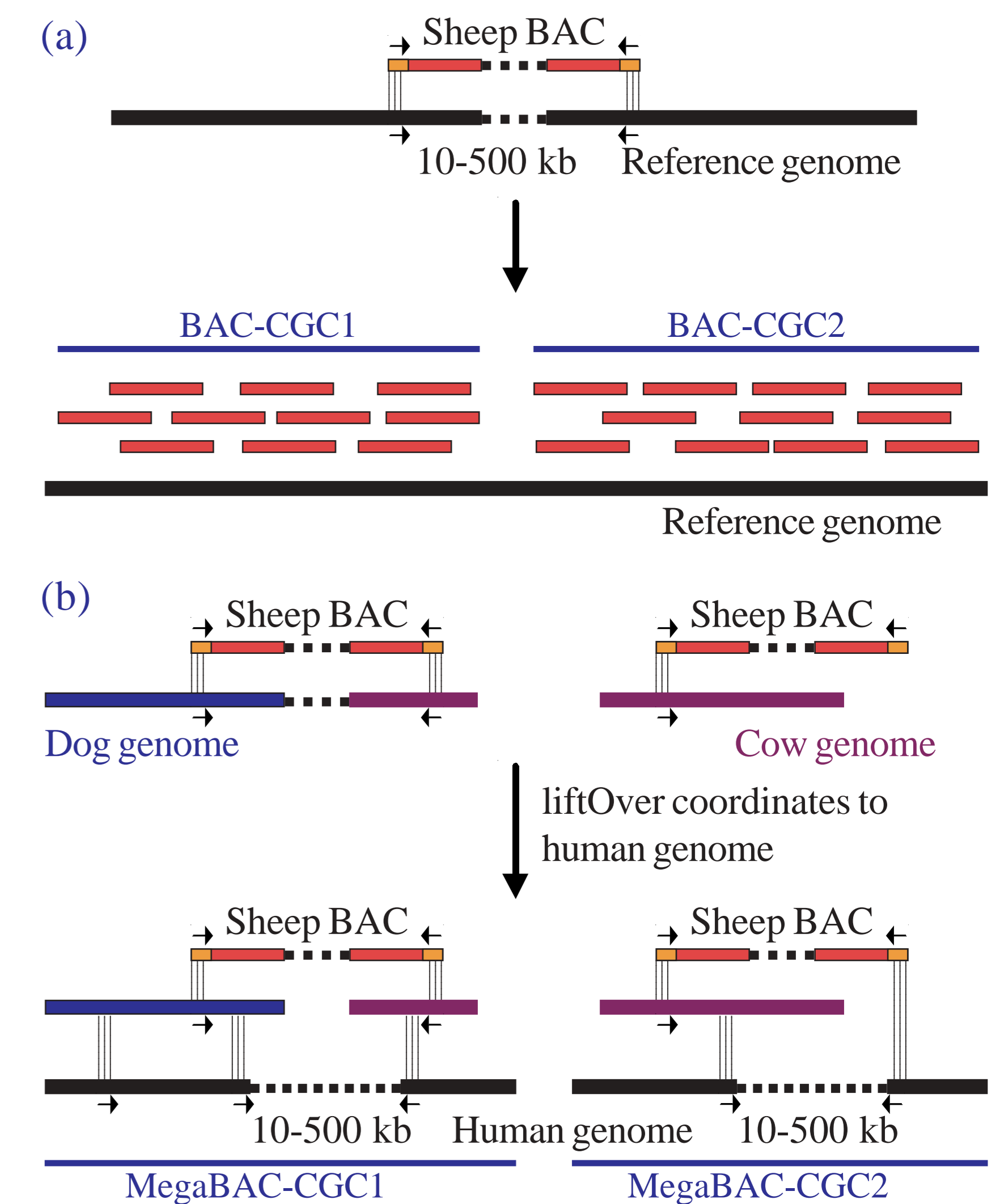
SNP chip

A project is underway under the auspices of the International Sheep Genomics Consortium (primarily with funding from an International Science Linkages grant, Australia) to sequence 80,000 fragments from the sheep genome as a forerunner to the development of a 20k sheep SNP chip. Both the virtual sheep genome and ovine EST sequences are being used to identify the 80,000 equally spaced targets for resequencing. Further details can be found at www.sheephapmap.org.

Acknowledgements

This work was supported by funding from sheepGENOMICS (www.sheepgenomics.com) which is jointly funded by Meat and Livestock Australia and Australian Wool Innovation.

Figure 3 Construction of the Virtual Sheep Genome



(c) HSA17 to OAR11 HSA17 to BTA19

