RESEARCH NOTE

Identification and chromosomal localization of repeat sequences through BAC end sequence analysis in Korean cattle

J. M. HONG[1], S. H. CHAE[2], N. ORIERO[1], D. M. LARKIN[5], C. B. CHOI[1,2,3], J. Y. LEE[4], H. A. LEWIN[5], J. H. BAE[3], I. CHOI[1,2,3] and J. S. YEO[1,2,3,*]

[1]*Institute of Biotechnology and Department of Biotechnology,* [2]*School of Biological Resources,* [3]*Biotechnology, and* [4]*Department of Statistics, Yeungnam University, Gyeongsan 712–749, Korea*
[5]*Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801 USA*

Introduction

The aim of this research was to identify and localise repeat sequences in Korean cattle (Hanwoo), which may be a basis for future genetic studies. To achieve this, a Hanwoo BAC library containing a total of 150,000 clones with an average size of 130–140 kb was constructed using two different restriction enzymes *Eco*RI and *Hind*III. A total of 2,794 clones were sequenced successfully at both 5′ and 3′ terminals using RP2 and T7 primers, generating 5,588 mate-pairs end sequences amounting to a total of 3.3×10^6 base pairs. The generated BAC end sequences were then compared with databases of repeat sequences submitted to GenBank and a total of three hundred and ninety five simple microsatellites (about 60 types) were identified with $(CA)_n$ repeat elements being the most abundant (24.87%). About 1,429 clones that were obtained after repeat masking were subjected to *in silico* analysis using BLASTN and COMPASS software. This analysis resulted in a total of 216 clones being localised with 98 clones exhibiting double hits on the same chromosome and 50 clones exhibited single hits to specific chromosomes with some repeat sequences being found close to a previously known marker. In addition, eighteen Hanwoo BAC clones containing simple repeat sequences were matched to a number of Expressed Sequence Tags (ESTs) containing simple microsatellites. Interestingly, four clones showed polymorphism in their repeated sequences when compared with the sequences submitted from other research groups in GenBank.

Mammalian chromosomes contain considerable amounts of DNA fragments that show great similarity in their nucleotide sequences and randomly occur throughout the chromosomes, commonly known as microsatellite DNA. However, despite their nucleotide sequence similarity, there are significant variations found among the repeat sequences through deletion and addition of base pairs, such that the repetition numbers of these microsatellites are quite variable. Microsatellite DNA is a short fragment of nucleotides (two to six nucleotide base pairs) with tandemly repeated sequences. Microsatellites are currently used as the most useful markers for linkage analysis and gene mapping as well as Quantitative Trait Loci (QTL) mapping of animals, due to their high abundance and polymorphism in their chromosomes (Baron *et al*. 1992; Tozaki *et al*. 2000). $(CA)_n$ dinucleotide repeats are the most popular markers and they have been used to generate linkage maps of human and cattle as well as for other mammalian species. The dinucleotide repeats are known to occur on average once every 30–60 kb in mammalian genomes (Stallings *et al*. 1991) and every 120–150 kb in the bovine genome (Vaiman *et al*. 1994; Moore *et al*. 1994).

Genetic and physical mapping of human and animal genomes have been greatly facilitated by the use of chromosome specific DNA libraries. Microsatellite markers previously placed on genetic maps can be used to screen large insert libraries and provide links between these maps and physical maps (Zhu *et al*. 1999). Barendse *et al*. (1997) showed that a medium-density genetic linkage map provided a basis for cross-over on unique genetic adjacency or the coding sequence with polymorphism based on the microsatellites that had already been known on the cattle map. This resulted in a high-resolution map.

The study reported here was designed to identify microsatellite DNAs located in chromosomes of Korean cattle (Hanwoo). The microsatellites were isolated from

*For correspondence. E-mail: jsyeo@yu.ac.kr.

Keywords. Hanwoo BAC clones; sequence analysis; bovine chromosome.

Hanwoo Bacterial Artificial Chromosome (BAC) clones, through BAC end sequencing and identification *in silico.* (Larkin *et al*. 2003; Zhao *et al*. 2001). The sequence information of microsatellite DNA existing in the Hanwoo chromosome will provide tools to construct the Hanwoo genetic map and eventually enable us to identify genetic markers.

Materials and methods

Hanwoo BAC library construction

A Hanwoo BAC library containing a total of 150,000 clones with an average size of 130–140 kb was constructed from a steer of the Korean native cattle showing high performance in meat quality and meat quantity (Chae *et al*. 2004). The constructed library provided the needed BAC end sequences for further analysis as is described below and summarized in figure 1.

Korean cattle BAC end sequence analysis

The 5′ and 3′ end DNA sequence of Hanwoo BAC clones were amplified using vector-specific universal primers (RP2 and T7) and analysed by an ABI 3700 capillary sequencer. Although the ABI sequencer contains a base calling software with impressive accuracy (Connell *et al*. 1987), we confirmed the accuracy of our generated sequences by analysing the traces from the sequencer using phred software. The obtained Hanwoo BAC end sequences were then analysed with the help of the Repeat Masker program to identify repeat sequences. The flanking nucleotide sequence of each clone after masking without the repeat sequences, were analysed using BLASTN and COMPASS (Comparative Mapping by Annotation and Sequence Similarity) software to localise the position of the each clone on the bovine chromosome.

Results and discussion

BAC library

Some regions of the bovine genome are rarely cleaved by one enzyme or are cut too frequently. Due to this fact, two different restriction enzymes (*Eco*RI *and Hind*III) with different base compositions of their recognition sequences were used to provide an increased coverage of the genome represented in the digested fragment pool. The fragmented genomic DNA was then cloned into BACe3.6 or pIndigoBAC-5 vectors containing the *Eco*RI restriction digestion site and *Hind*III restriction digestion site respectively and about 150,000 of Hanwoo BAC clones containing 130–140 kb average insert size were generated. Upon analysis, less than 2% of the clones were considered as empty (Chae S. H., Larkin D. M., Kim J. W., Ghimeray A. M., Hong J. H., Lewin H. A., Yeo J. S. and Choi I. (2004) Construction and chromosomal localization of Korean native cattle BAC clones through BAC end sequence analysis. *Manuscript in preparation*). Assuming that the bovine genome contains 3×10^9 bp (Eggen *et al*. 2001), the total library constructed therefore corresponds to six genome-equivalents. The Hanwoo BAC clones were constructed in triplicates and are stored as glycerol stock in 384 well plate at – 80°C.

BAC end sequence analysis

Analysis of end sequences from bacterial artificial chromosomes (BACs) has been suggested as a suitable method of selecting overlapping clones in large scale genome sequencing projects (Venter *et al*. 1996; Boysen *et al*. 1997a,b). This is because the end sequences have been proposed to be a source of highly specific sequence markers (Zhao *et al*. 2000). The 9,533 BAC end sequences were analysed from the constructed Hanwoo

Table 1. Repeated sequence summary.
Total length: 3,299,298 bp.
Bases masked 1,150,852 bp (34.88%).

	Number of elements*	Length occupied	Sequence (%)
SINEs	1,912	280,945 bp	8.52
LINEs	2,344	653,391 bp	19.80
LTR elements	500	124,761 bp	3.78
DNA elements	267	46,569 bp	1.14
Total interspersed repeats		**1,105,666 bp**	**33.51**
Small RNA	26	1,902 bp	0.06
Satellites	31	9,734 bp	0.30
Simple repeats	365	15,899 bp	0.48
Low complexity	456	18,075 bp	0.55

*Most repeats fragmented by insertions or deletions have been counted as one element.

BAC clones using T7 and RP2 primers, resulting in approximately 7,079 (74% success rate) high quality BAC end sequencing reads. pIndigo BAC-5 vector offers strict copy number control thus limiting the number of BACs to one or two copies per cell, resulting in sequencing failures due to low DNA yield and in some cases empty wells. Due to this problem it is imperative that the quality of the obtained BAC sequences be examined. This is usually done through base calling using phred software that reports quality values for each base called (Ewing *et al*. 1998a,b). The phred quality score indicated that the obtained DNA sequences were all high quality bases. The BAC end sequences were then trimmed of vector sequences resulting in 6,770 sequences with ≥ 200 bp length, corresponding to an overall success rate of 95.6% which compares favorably with other results on cattle and mouse BAC end sequencing (Larkin *et al*. 2003; Zhao *et al*. 2001). A total of 2794 clones were sequenced successfully at both 5′ and 3′ terminals, generating 5,588 mate-pairs end sequences amounting to a total of 3.3×10^6 base pairs. These were then used to investigate the presence of repeat sequences in the Hanwoo genome.

Repeated sequence analysis

Each of the BAC end sequence mate-pairs was analysed by the RepeatMasker program. It was observed that out of approximately 3.3 million bp that were obtained, about 1.1 million of these contain repeat sequences. This means that about 34.88% of the DNA sequence consists of repeat sequences with interspersed repeats being the most abundant repeat elements amounting to 33.51% of the total 34.88% of repeat DNA sequence. The most abundant class of repeat elements is LINE; 19.8%, followed by SINE; 8.52% and LTR; 1.41% (table 1). The quantity of interspersed repeats found in this study is low when compared to previous work done on the cattle and mouse BAC end sequences (Larkin *et al*. 2003; Zhao *et al*. 2001). This is most likely due to the cutting pattern of different restriction digest enzymes that are used in each study. In this study we used *Eco*RI and *Hind*III enzymes while in the research done by above researchers *Mbo*I was the enzyme of choice. A study carried out by Zhao *et al*. (2001), reports that LINE contents increase in BAC end sequences of clones constructed from *Mbo*I partial digest libraries, which can then increase the overall number of interspersed repeats found in the end sequences.

Simple repeats commonly known as microsatellites were also present and amounted to 0.48% of the total repeat sequence. 60 types composed of 394 microsatellites were found. Most common simple repeats are (CA)n; 24.87%, (TA)n; 15.98%, (A)n; 7.86% and (CTG)n; 4.56% (table 2), consolidating previous research that found (CA)n repeats to be the most abundant simple repeat in the mammalian genome (Stallings *et al*. 1991; Vaiman *et al*. 1994; Moore *et al*. 1994). The entire results of this analysis are presented in table 3. These sequences, including those of the rest of the repeat sequences that were

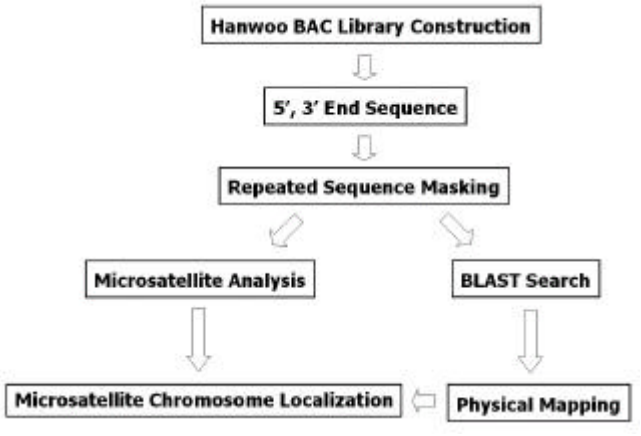

Figure 1. Schematic diagram of the strategy for microsatellite DNA identification and chromosomal mapping.

Table 2. Summary of microsatellites in 2,794 Hanwoo BAC clone.

Repeat sequence	No. of Repeat	Repeat sequence	No. of Repeat
(CA)n	98	(CACCC)n	1
(CAAAAA)n	1	(CTATG)n	1
(CATG)n	3	(TGGA)n	2
(AACTG)n	11	(TATAA)n	1
(CAGTT)n	41	(TTTTG)n	5
(TA)n	63	(AAGTG)n	1
(TTTTA)n	3	(CAATA)n	1
(CAAA)n	1	(GAAAA)n	1
(TTCA)n	3	(TCCA)n	3
(TATATG)n	4	(GGAA)n	1
(CAG)n	13	(CAGT)n	1
(GA)n	6	(CAT)n	3
(A)n	31	(CTA)n	1
(TAAA)n	9	(CAAGA)n	1
(CG)n	1	(TTTAA)n	1
(TGAA)n	6	(TTTG)n	4
(CAA)n	3	(CTTTA)n	1
(CACG)n	1	(CTCA)n	1
(CAGAGA)n	1	(CAAAC)n	1
(CTATT)n	1	(TCTG)n	1
(CATATA)n	5	(TTAGGG)n	1
(TTTC)n	1	(CATAT)n	1
(TTATA)n	1	(TGGGGG)n	1
(TATG)n	2	(CCCA)n	1
(G)n	2	(TAAAAA)n	1
(CTG)n	18	(CCTG)n	1
(CAGAT)n	9	(CGGGGG)n	1
(TCTA)n	2	(CATA)n	1
(TAA)n	9	(GGCTG)n	1
(GAAA)n	2	(ATG)n	2
		Total	**394**

Table 3. Type of repeat sequence and chrosomal localization.

Repeated sequence name	BAC clone ID	No. of repeat	Predicted Chro.	Predicted marker	Repeated sequence name	BAC clone ID	No. of repeat	Predicted Chro.	Predicted marker
CA-1	159–3-F09–RP2	10	16**	EST1013	**TA-11**	159–4–A06–T7	4	5**	MYF5
CA-2	183–1–A04–T7	12	12**	COL4A1	**TA-13**	161–1–D11–RP2	7	4*	CD36
CA-4	186–3–G07–RP2	17	3**	GBP1	**TA-14**	161–1–E12–T7	7	26**	SPF30
CA-5	187–2–H01–RP2	10	14**	D10884	**TA-17**	161–2–F12–RP2	10	3**	FCGR2B
CA-8	204–1–B11–T7	12	13**	CHGB	**TA-18**	161–3–E01–T7	19	14*	TG
CA-10	150–3–A03–RP2	14	24*	BCL2	**TA-19**	161–4–G07–RP2	5	5**	LYZ
CA-12	183–1–D11–RP2	9	3**	JUN	**TA-20**	183–1–B03–T7	13	11*	GGCX
CA-13	184–3–B09–RP2	15	14*	ANGPT1	**TA-22**	183–1–D06–T7	4	1**	PDIR
CA-15	185–3–G03–RP2	6	24**	ME2	**TA-25**	183–1–H11–RP2	7	5**	IGF1
CA-20	200–3–E11–RP2	19	1**	IL12A	**TA-26**	183–2–C03–RP2	18	1*	IL12A
CA-24	184–2–G12–RP2	17	11*	KIAA0235	**TA-27**	184–1–F12–T7	3	13*	VIM
CA-28	186–2–D08–RP2	14	19*	CNP	**TA-29**	184–3–H01–T7	7	1**	IL12A
CA-31	187–4–A07–RP2	5	16*	PLOD	**TA-30**	185–3–B01–T7	3	15**	ADM
CA-32	200–3–G11–T7	5	19**	TP53	**TA-34**	187–2–F04–T7	4	19*	EST1409
CA-33	202–1–B12–RP2	7	16**	EST0804	**TA-35**	187–3–A03–RP2	20	7**	EST1096
CA-35	150–2–E11–RP2	11	27**	NRG1	**TA-36**	187–3–G08–T7	11	8**	CYLC2
CA-48	161–3–B04–RP2	16	11**	UGP2	**TA-39**	188–2–A04–RP2	10	28*	CHS1
CA-49	183–1–B09–T7	11	19*	P4HB	**TA-40**	188–2–H02–RP2	8	1**	AHSG
CA-52	183–2–B02–RP2	11	1**	EST1413	**TA-45**	200–3–H05–RP2	11	8*	TEK
CA-55	184–1–H07–RP2	18	8**	NRP2	**TA-47**	201–1–F09–T7	6	22**	GNAT1
CA-57	184–2–F10–T7	13	8**	TEK	**TA-49**	201–1–H10–RP2	9	8**	VLDLR
CA-66	186–2–H07–RP2	20	3*	MUC1	**TA-55**	202–4–E12–RP2	23	5**	K-ALPHA-1
CA-69	187–2–E05–RP2	17	1**	POU1F1	**TA-60**	204–3–B02–T7	3	3**	RPE65
CA-74	187–2–H09–T7	12	5**	EST1290					
CA-76	187–3–E11–RP2	9	20**	ITGA2	**CAGTT-4**	159–3–B09–T7	3	3*	FCGR2B
CA-80	188–2–F07–RP2	20	12**	ELF1	**CAGTT-5**	159–4–C05–T7	3	15*	C1NH
CA-81	188–3–B07–T7	5	12**	EST2003	**CAGTT-9**	161–3–H07–T7	3	5**	COL2A1
CA-89	202–2–C03–T7	11	12**	COL4A1	**CAGTT-19**	185–1–G06–RP2	4	11**	RRM2
CA-94	204–3–A12–RP2	10	19**	EST1409	**CAGTT-22**	185–2–G05–RP2	2	5**	COL2A1
CA-95	204–3–B08–RP2	13	17**	EST2351	**CAGTT-25**	187–2–E09–RP2	3	12**	SPRY2
CA-96	204–4–A07–RP2	7	28*	SLC25A16	**CAGTT-27**	187–3–A04–RP2	4	7**	GNPI
CA-97	204–4–H07–RP2	11	9*	OPRM1	**CAGTT-28**	187–3–F10–T7	3	9**	EST1223
					CAGTT-31	187–4–B10–T7	3	5*	EST1290
A-1	150–3–A01–RP2	20	1**	P2RY1	**CAGTT-32**	188–2–B01–RP2	3	25*	PRKCB1
A-2	159–2–A02–RP2	21	1**	EST1413	**CAGTT-36**	202–2–C07–RP2	3	1, 22	RPS25
A-3	161–1–A07–RP2	14	2*	FRZB					
A-8	187–2–E02–RP2	22	10**	HSPCA	**CTG-5**	183–1–B12–RP2	6	9**	CTGF
A-9	187–3–C10–RP2	23	21**	BCL2A1	**CTG-9**	184–3–E05–RP2	8	8*	TEK
A-11	188–4–H02–T7	22	1**	AHSG	**CTG-10**	184–3–H01–T7	7	1**	IL12A
A-12	189–1–B12–T7	21	13**	CHGB	**CTG-12**	185–1–D10–RP2	5	4**	D26549
A-14	189–4–E11–T7	22	X	FACL4	**CTG-13**	185–1–H09–T7	9	21*	WARS
A-15	200–3–B11–T7	20	7*	EST1394	**CTG-15**	187–2–E09–T7	6	12**	SPRY2
A-18	159–4–C01–T7	20	19**	PECAM1	**CTG-16**	187–3–A05–RP2	6	16**	DDX9
A-19	161–2–E12–RP2	12	3**	TIE					
A-20	161–2–G08–T7	9	2**	KIAA0137	**CAG-2**	159–4–F02–RP2	6	3*	VCAM1
A-22	184–2–A12–RP2	22	11**	RRM2	**CAG-3**	159–4–F07–RP2	7	2*	CD28
A-23	184–3–E12–T7	20	X	SLC6A8	**CAG-5**	183–2–B01–RP2	8	23**	SFRS3
A-24	184–3–F10–T7	20	X	SLC6A8	**CAG-6**	185–1–C01–T7	6	1**	MX1
A-25	184–4–A04–RP2	23	6*	SPP1	**CAG-8**	187–3–G05–RP2	7	11**	EST1817
A-26	185–3–D10–T7	14	1*	MX1	**CAG-11**	202–1–D03–RP2	7	5**	COL2A1
A-27	186–4–C09–RP2	20	19**	ICT1	**CAG-12**	202–1–H07–T7	7	3**	FCGR2B
A-28	187–4–C01–RP2	23	10**	CCNB2					
A-30	204–1–D12–RP2	9	24*	BCL2	**AACTG-2**	159–1–D02–RP2	3	26**	SPF30
					AACTG-8	188–2–G03–RP2	4	17**	EDNRA
TA-3	150–2–E02–RP2	6	1*	P2RY1	**AACTG-10**	189–3–H09–RP2	4	6**	PTPN13
TA-4	150–3–A10–RP2	6	14**	IL7					
TA-6	150–4–F01–RP2	25	12**	FLT1	**CAGAT-1**	159–1–C04–RP2	3	13**	EST1724
TA-10	159–1–D02–T7	3	26**	SPF30	**CAGAT-5**	187–2–A01–RP2	4	6**	EST1161

Table 3. (*Contd*.)

Repeated Sequence Name	BAC clone ID	No. of Repeat	Predicted Chro.	Predicted marker	Repeated Sequence Name	BAC clone ID	No. of Repeat	Predicted Chro.	Predicted marker
CAGAT-6	188–1–A04–T7	4	24*	AQP4	**CAT-1**	161–4–F02–RP2	4	5*	TST
					CAT-2	184–2–C11–RP2	2	17*	MTMR3
TAAA-3	183–2–G09–T7	4	17**	KRT8					
TAAA-7	159–3–A11–T7	2	3**	JUN	**G-2**	202–1–A08–RP2	20	1*	MX1
TAAA-8	159–3–D09–T7	3	4**	TCRB					
					TCTA-1	150–2–E06–T7	5	6*	WDR1
TAA-1	183-1-D02-T7	2	5**	LUM					
TAA4	188-2-C02-RP2	7	20**	CDH6	**GAAA-1**	184–2–F07–RP2	4	4**	PTN
TAA-8	184–3–H12–T7	4	4*	TCRB					
TAA-9	186–2–G02–T7	2	6**	UGDH	**TGGA-1**	188–3–B07–T7	2	12**	EST2003
TGAA-2	184–2–F07–RP2	2	4**	PTN	**ATG-1**	184–2–C11–RP2	3	17*	MTMR3
TGAA-3	184–2–G05–T7	2	3**	TIE					
TTTTG-2	184–4–H08–RP2	3	16*	EST0339	**CATA-1**	161–2–A04–T7	3	4**	IGFBP3
TTTTG-4	184–2–G12–RP2	4	11*	KIAA0235	**CAGAG-1**	202–3–E11–T7	3	18**	ZFP36
					CACCC-1	188–2–H05–T7	2	15**	ADM
GA-1	150–2–H05–T7	10	21*	PI					
GA-3	161–3–G11–RP2	7	12**	EST2003	**AAGTG-1**	202–2–D02–RP2	4	3**	F3
GA-4	187–2–D09–RP2	5	3**	RPE65					
GA-5	201–1–H11–RP2	5	1*	KIAA0332	**GGAA-1**	161–1–F10–RP2	5	9**	DDO
GA-6	204–1–B02–RP2	8	16**	PLOD					
					CTA-1	161–4–H07–T7	7	12*	COL4A1
CATATA-1	161–1–E11–RP2	2	5**	LUM					
CATATA-3	189–1–G01–T7	–	10**	X73772	**TTTAA-1**	187–2–G02–RP2	–	7**	RASA1
CATG-2	150–2–F09–RP2	5	13*	GNAS1	**CTTTA-1**	187–3–F11–T7	–	10*	CYP19
CATG-3	183–1–B09–T7	–	19*	P4HB	**CATAT-1**	189–3–F11–T7	3	12*	SPRY2
TTTG-1	161–3–C12–RP2	5	4*	COL1A2	**TGGGGG-1**	200-3-E03-RP2	3	24*	LMAN1
TTTG-4	189–4–C01–T7	4	17**	EST1517	**TTTAA-1**	187–2–G02–RP2	3	7**	RASA1
TTTTA-2	184–4–D11–RP2	3	13**	EST0424					
TTCA-3	200–3–H03–T7	3	X	TNFSF5	**CCCA-1**	183–1–H07–RP2	2	X	FACL4
					CGGGGG-1	159–3–B11–T7	2	15**	FDX1
CAA-1	202–2–C02–T7	6	10**	SNAP23					
CAA-2	187–2–H10–T7	6	12*	COL4A1					
					GGCTG-1	161–2–G09–RP2	–	7*	DKFZP56 6B133
TCCA-1	185–3–G01–T7	5	3**	F3					

No of repeats; number of repeat sequences found in each clone, **Predicted chromosome**; the chromosome number each clone is predicted to originate from. **Predicted marker**; previously known markers closely located to the repeat sequence. *; One end sequence of a clone matched to a chromosome. **; both end sequences of a clone match to same chromosome; underlined clones contain ESTs. – repeat sequence is distributed over entire cloned sequence.

found are stored locally in our database and have also been uploaded to GenBank (Accession No: CG83987-CG844908). The finding of these repeat sequences is significant because it allows us to know that more than 60% of the Hanwoo BAC end sequences are completely free of repetitive DNA and can therefore be used as a source of new STS (Sequence tagged site) markers. In fact when the end sequences were analysed using BLASTN, a number of ESTs in 18 clones were found as simple microsatellite sequences, with five of the ESTs being localised on the cattle genome with a high degree of accuracy since they are found in clones that have either both or one of their end sequences showing homology to the same chromosome (table 3). This is exciting as the localisation of ESTs can be a plausible strategy to increase the density of gene markers on the bovine linkage and comparative maps. Inferring bovine EST physical map position is particularly useful for candidate QTL gene identification especially in high-throughput bovine EST sequencing projects (Grosse *et al*. 2000). In fact, further analysis of some of the microsatellites demonstrated the presence of alleles that are closely related to economic traits such as marbling score, daily gain and backfat thickness in Hanwoo (Kim *et al*. 2004). Interestingly, four clones

showed polymorphism in their repeated sequences when compared with the sequences submitted from other research groups in GenBank. This is exciting as it opens an avenue to the identification of possible genes that could be unique to the Korean cattle or the possibility of the identification of unique economic traits in these cattle.

Chromosomal localisation of clones containing repeat sequences

The 2,794 clones that were end sequenced successfully on both 5′ and 3′ terminals had their end sequences analysed for the presence of repeat sequences and the identified repeat sequences were masked using the RepeatMasker program. They were then subjected to further analysis to localise the cloned sequence to bovine chromosomes (table 3). This analysis was facilitated by the help of super computer (W. M. Keck Center, University of Illinois, Urbana-Champaign, IL, USA) using the BLASTN and COMPASS software.

The analysis resulted in 216 clones being localised to bovine chromosomes. Of these, 98 clones were localised with a high degree of accuracy as both end sequences of the clones showed homology to the same chromosome while 50 clones exhibited single end sequence homology to specific chromosomes indicating that they could be from these chromosomes but with a low degree of accuracy (table 3). A further 58 clones were ambiguously localised to 2 different chromosomes each, 10 clones were localised to 3 different chromosomes each and 168 clones were not localised at all (Data not shown). Clones that failed to be localised or were ambiguously localised could have been due to amplification of paralogous sequences that results in high retention frequencies which hinders localisation. They therefore require further analysis to confirm their place in the bovine genome. The above analysis was carried out under the assumption that the overall fractions of DNA of each repeat class will provide some indication that the end sequence dataset and therefore the BAC library are representative of the bovine genome. Only DNA sequences free of repetitive DNA were used in this analysis as previous studies on the human genome show that human BAC end sequences free of repetitive DNA are likely to be useful as a source of new STS.

Taken all together the results indicate that the clones and therefore the repeat sequences obtained are indeed representative of the BAC library as we had an overall success rate of 56.25% in localising them. The results presented in this paper are therefore important as a source of information for comparative genetic studies with other cattle breeds of the world and they do provide a basic structure on which cattle genetic researchers can draw upon in their quest for unique economically important genes. They could also help provide an explanation as to the evolutionary adaptation of these cattle to its local environment. BAC end sequence isolation is still going on and our efforts are now concentrated on the isolation of more microsatellites in the search for sequences that are unique to Korean cattle, and to isolate these for further analysis of the presence of QTLs or for diverse evolutionary studies.

Acknowledgements

This work was supported by a grant from BioGreen 21 Program, Rural Development Administration, Republic of Korea. We thank all our laboratory members especially Amal Kumar Ghimeray for their daily laboratory assistance and Dong Sik Choi for her invaluable help on the manuscript.

References

Barendse W., Vaiman D., Kemp S. J., Sugimoto Y., Armitage S. M., Williams J. L., Sun H. S., Eggen A., Agaba M., Aleyasin S. A., Band M., Bishop M. D., Buitkamp J., Byrne K., Collins F., Cooper L., Coppettiers W., Denys B. and Pomp R. D. 1997 A medium-density genetic linkage map of the bovine genome. *Mammalian Genome* **8**, 21–28.

Baron B., Poirier C., Simon C. D., Barnier C. and Guenet J. L. 1992 A new strategy useful for rapid identification of microsatellites from DNA libraries with large size inserts. *Nucleic Acids Res* **20**, 3665–3669.

Boysen C., Simon, M. I. and Hood L. 1997a Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with bacterial artificial chromosome clones. *Genome Res* **7**, 330–338.

Boysen C., Simon M. I. and Hood L. 1997b Fluorescence-based sequencing directly from bacterial and P1-derived artificial chromosomes. *BioTechniques* **23**, 978–980.

Chae S. H., Larkin D. M., Kim J. W., Ghimeray A. M., Hong J. H., Lewin H. A., Yeo J. S. and Choi I. 2004 Construction and chromosomal localization of Korean native cattle BAC clones through BAC end sequence analysis. *Manuscript in preparation*.

Connell C., Fung S., Heiner C., Bridgham J., Chakerian V., Heron E., Jones B., Menchen S., Mordan W. and Raff M. 1987 Automated DNA sequence analysis. *BioTechniques* **5**, 342–348.

Eggen A., Gautier M., Billaut A., Petit E., Hayes H., Laurent P., Urban C., Pfister-Genskow M., Eilertsen K. and Bishop M. D. 2001 Construction and characterization of a bovine BAC library with four genome-equivalent coverage. *Genet Sel Evol* **33**, 543–548.

Ewing B., Hillier L., Wendl M. C. and Green P. 1998a Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175–185.

Ewing B. and Green P. 1998b Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.

Grosse W. M., Kappes S. M. and McGraw R. A. 2000 Linkage mapping and comparative analysis of bovine expressed sequence tags (ESTs). *Animal Genetics* **31**, 171–177.

Kim J. W., Hong J. M., Lee Y. S., Chae S. H., Choi C. B., Choi I. and Yeo J. S. 2004 Identification of new microsatellite DNAs in the Chromosomal DNA of the Korean cattle (Hanwoo). *Asian-Austrian Journal of Animal Science*, in press.

Larkin D. M., Everts-van der Wind A., Rebeiz M., Schweitzer P. A., Bachman S., Green C, Wright C. L., Campos E. J., Benson L. D., Edwards J., Liu L., Osoegawa K., Womack J. E., de Jong P.J. and Lewin H. A. 2003 A cattle-human comparative map built with cattle BAC-ends and human genome sequence. *Genome Res* **13**, 1966–1972.

Moore S. S., Byrne K., Berger K. T., Barendse W., McCarthy F., Womack J. E. and Hetzel D. J. S. 1994 Characterization of 65 bovine microsatellites. *Mamm. Genome* **5**, 84–90.

Stallings R. L., Ford A. F., Nelson D., Torney D. C., Hildebrand C. E. and Moyzis R. K. 1991 Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. *Genomics* **10**, 807–815.

Tozaki T., Inoue S., Mashima S., Ohta M., Miura N. and Tomita M. 2000 Sequence analysis of trinucleotide repeat microsatellites from an enrichment library of the equine genome. *Genome* **43**, 354–365.

Vaiman D., Mercier D., Moazzmi-Goudarzi K., Eggen A., Ciampolini R., Lepingle A., Velmala R., Kaukinen J., Varvio S. L., Martin P., Leveziel H. and Guerin G. 1994 A set of 99 cattle microsatellites: characterization, synteny mapping and polymorphism. *Mamm. Genome* **5**, 288–297.

Venter J. C., Smith H. O. and Hood L. 1996 A new strategy for genome sequencing. *Nature* **381**, 364–366.

Zhao S., Malek J., Mahairas G., Fu L., Nierman W., Venter J. C. and Adams M. D. 2000 Human BAC ends quality assessment and sequence analyses. *Genomics* **63**, 321–32.

Zhao S., Shatsman S., Ayodeji B., Geer K., Tsegaye G., Krol M., Gebregeorgis E., Shvartsbeyn A., Russell D., Overton L., Jiang L., Dimitrov G., Tran K., Shetty J., Malek J. A., Feldblyum T., Nierman W. C. and Fraser C. M. 2001 Mouse BAC ends quality assessment and sequence analyses. *Genome Res* **11**, 1736–1745.

Zhu B., Smith J. A., Tracey S. M., Konfortov B. A., Welzel K., Schalkwyk L. C., Lehrach H., Kollers S., Masabanda J., Buitkamp J., Fries R., Williams J. L. and Miller J. R. 1999 A 5 × genome coverage bovine BAC library: production, characterization and distribution. *Mammalian Genome* **10**, 706–709.

Received 26 February 2005