Localisation of 5,105 Hanwoo (Korean Cattle) BAC Clones on Bovine Chromosomes by the Analysis of BAC End Sequences (BESs) Involving 21,024 Clones

Jae Min Choi1,4, Sung-Hwa Chae3,4, Se Won Kang2, Dong-Sik Choi2, Yong Seok Lee6
Hong-Seog Park3, Jung-Sou Yeo1,2 and Inho Choi1,2,*

1 Institute of Biotechnology and Department of Biotechnology, Yeungnam University, Gyeongsan, Korea

ABSTRACT: As an initial step toward a better understanding of the genome structure of Korean cattle (Hanwoo breed) and initiation of the framework for genomic research in this bovine, the bacterial artificial chromosome (BAC) end sequencing of 21,024 clones was recently completed. Among these clones, BAC End Sequences (BESs) of 20,158 clones with high quality sequences (Phred score ≥20, average BES equaled 620 bp and totaled 23,585,814 bp), after editing sequencing results by eliminating vector sequences, were used initially to compare sequence homology with the known bovine chromosomal DNA sequence by using BLASTN analysis. Blast analysis of the BESs against the NCBI Genome database for Bos taurus (Build 2.1) indicated that the BESs from 13,201 clones matched bovine contig sequences with significant blast hits (E<e^-40), including 7,075 single-end hits and 6,126 paired-end hits. Finally, a total of 5,105 clones of the Korean cattle BAC clones with paired-end hits, including 4,053 clones from the primary analysis and 1,052 clones from the secondary analysis, were mapped to the bovine chromosome with very high accuracy. (Key Words: Korean Cattle (Hanwoo), BAC End Sequence (BES), Chromosomal Map)

INTRODUCTION

Developing the genetic map for cattle is an expected initial step toward identifying the Economic Trait Loci (ETL). Bovine linkage map construction that is based upon the classical genetic approach has provided a foundation for generating the physical map (Barendse et al., 1997; Ihara et al., 2004). The construction of a bovine whole-genome radiation hybridization (RH) panel (Womack et al., 1997; Itoh et al., 2005) allowed cattle genome researchers to construct subsequent RH panel-based physical maps (Band et al., 2000; Larkin et al., 2003; Everts-van der Wind et al., 2004). Further incorporation of the bovine bacterial artificial chromosome (BAC) clones into the RH map (Larkin et al., 2003; Everts-van der Wind et al., 2005) has prompted the bovine whole genome sequencing project currently being conducted by an international consortium of bovine genome researchers.

The BAC clones have been mostly used for physical mapping due to their capacity to hold large DNA insert in a stable manner (Kim et al., 1996). The BAC library has made it possible to screen multiple clones of interest by using the PCR approach (Yu et al., 2005; Choi et al., 2006). The bovine BAC library (Chae et al., 2007) generated by using the chromosomal DNA of Korean cattle (Hanwoo breed) has been a useful resource for genomic research (Hong et al., 2005). Random analysis of the BAC end sequences (BESs) using universal primers is one of the approaches to obtain chromosomal DNA sequences of the cattle quickly and at relatively low cost. In addition, the BESs can be utilized to localize the precise position of the BAC clones on the bovine chromosomes (Larkin et al., 2003) and to identify a variety of DNA markers (Heaton et al., 2002; Kim et al., 2004; Sasazaki et al., 2006). Here, we report on the mapping of the Korean cattle BAC clones in comparison with currently generated Bos taurus contig map.
Table 1. Data analysis of the BAC end sequences

| Description                                      | No. of clones | Size of BES (bp) | Average BES (bp) |
|--------------------------------------------------|---------------|------------------|------------------|
| Total BAC end sequencing trials<sup>a</sup>      | 21,024        | 23,585,814       | 620              |
| After editing the sequencing data<sup>b</sup>     | 20,158        |                  |                  |
| (base calling (Phred ≥20)/vector masking/BES≥100 bp) | 38,061        |                  |                  |

<sup>a</sup> Number of the BAC end sequencings initially tried.

<sup>b</sup> Sequencing data after editing with base calling (Phred quality score ≥20), vector trimming, and minimum read length (≥100 bp).

**MATERIALS AND METHODS**

**BAC end sequencing**

Single BAC colonies were transferred into 1 ml of Terrific Broth (TB) medium supplemented with antibiotic (chloramphenicol at 12.5 μg/ml) using a 96-deep well plate and incubated at 37°C overnight with gentle rotation (550 rpm; HT-MegaGrow shaking incubator, Bioneer, Daejeon, Korea). The BAC DNA was extracted and purified using a Montage BAC<sub>96</sub> miniprep kit (Millipore, Bedford, MA, USA), or modification of an alkaline lysis method commonly used for isolating plasmid DNA from bacteria (Birnboim et al., 1979; Kelley et al., 1999) designed to optimize DNA yield. BAC end sequencing was performed according to the manufacturer’s instructions with a BigDye Terminator Cycle Sequencing Kit (ver. 3.1, Applied Biosystems, Foster City, CA, USA). Cycle sequencing reaction was performed with 600 ng of BAC DNA, 3 pmoles of primer, 0.87 μl of 5× buffer, 1.38 μl of distilled water and 0.5 μl of BigDye, using a GeneAmp PCR System 9700 (Applied Biosystems). The universal sequences of primers used for the forward primer was T7 and the reverse primer were slightly modified RP2 or M13R. Amplification was conducted at 96°C for 2 min and 80 cycles of denaturation at 96°C for 15 sec, annealing at 50°C for 10 sec, and extension at 60°C for 2.5 min. PCR products were purified by ethanol precipitation and resolved on an ABI 3730XL DNA Analyzer (Applied Biosystems).

**Sequence analysis**

Sequence chromatograms were processed automatically using sequencing analysis software (Sequencing analysis; ver. 3.7, Applied Biosystems). The trace files were trimmed with trim-alt 0.05 (P-score ≥20) using Phred program (Ewing and Green, 1998; Ewing et al., 1998) and the sequence below 100 bp was removed. In addition, vector trimming was conducted using cross-match software (http://www.phrap.org). Repetitive elements were identified using the RepeatMasker program (Jurka et al., 2005) to determine the relative content of interspersed repeats, small RNAs, satellites, simple repeats and low complexity sequences in the nucleotide sequences. Finally cleaned multi-fasta formatted data with the RepeatMasker program and Repbase database (ver 2006.2) were analyzed with NCBI local BLAST (Altschul et al., 1990) and sequences were searched against the NCBI Bos taurus genome data (build 2.1 based on Btau 2.0).

**Comparative mapping**

On the basis of the BLAST results, the only paired-BAC clones with paired-end hits at the same chromosome and less than 300 kb in their estimated insert size were selected for mapping using perl script, GD-library and tgif generation process using post-scripts map make out.

**RESULTS AND DISCUSSION**

Among the over 150,000 BAC clones generated from the chromosomal DNA of a Korean cattle, the sequencing analysis of BAC end sequences (BESs) for 21,024 clones, including those for 4,296 clones previously described (Chae et al., 2007), has been attempted (Table 1). Several thousand BAC clones (20,158) were selected based on quality (phred score ≥20) and length (≥100 base pairs) after base calling and vector trimming. The average and total lengths of the selected sequences were 620 bp and 23,585,814 bp, respectively. A total of 17,903 clones with paired BESs and 2,255 clones with a single BES corresponded to the selection criteria. Approximately 1.54% of the clones revealed a length of BES that was not greater than 100 bp at both ends after vector trimming, which may be due to failed insertion in the clone or a sequencing failure. Thus, on the basis of these results, a complementary estimation can be made that is, at least 98.46% BAC clones have a proper insert sequence.

High quality BESs of the Hanwoo were compared by BLASTN analysis with the B. taurus genomic sequence obtained from a public database (NCBI build 2.1) after repeat masking by RepeatMasker (ver. 3-1-3) software (http://www.repeatmasker.org) with the Repbase database (January 20, 2006). While 6,957 of the 20,158 clones used in the analysis lacked similarity in their BESs to the Bos taurus (Btau 2.1) genome sequence (no hit; E<e<sub>-40</sub>), the remaining 13,201 clones showed a significant similarity (significant blast hit; E<e<sub>-40</sub>) (Table 2). Out of the 13,201 clones with a significant similarity, the end sequences of 7,075 clones showed a single-end hit (only one of both ends of the BESs hits the bovine genome), and 6,126 clones showed paired-end hits (both ends of the BESs hit the bovine genome). There were 4,767 clones of the 7,075 total that produced single hits (hits at only one site of the bovine genome).
2,308 clones produced multiple hits (hits at multiple sites of the bovine genome). Sequence similarity was observed in both end sequences of 6,126 clones. Among these, 1,226 clones revealed each end sequence displayed similarity in a different chromosomal location, but in 4,900 clones, both end sequences showed similarity to only one chromosome. Among these clones with similarity to one chromosome, each end of the 2,447 clones had single hits in the assembled genome and one or both ends of the 2,453 clones had multiple hits.

The location of each clone in a chromosome was determined from the result of BLAST analysis between Hanwoo BES and the\textit{B. taurus} whole genome sequence. To determine the chromosomal location of the BAC clones, the only clones yielding paired-end hits were selected (E<\textit{e}^{-40}). The 4,900 clones with similarity to the\textit{B. taurus} genome in both end sequences (Table 2) were analyzed initially. Of the clones that displayed more than two hits (2,453 clones) in the same chromosome, ones with high E-values in both end sequences were selected for inclusion in the construction of the genome map. The average insert size of Hanwoo chromosomal DNA in BAC clones was 140 kb (Chae et al., 2007). Therefore, clones containing more than 300 kb between the end sequences were excluded from the 4,900 clones with similarities to the chromosome in both end sequences, resulting in 4,053 clones being used for the primary construction of the map (Figure 1).

The clones excluded from the primary genome map, the 6,957 clones that did not display similarity with the\textit{B. taurus} genomic sequences (E<\textit{e}^{-40}) and the 7,075 clones with single-end hits after deletion of repetitive sequences, were used for a secondary comparison against the\textit{B. taurus} genomic sequences with reduced stringency (E<\textit{e}^{-10}). The secondary comparison analyses were performed twice as the first analysis - once with repeat masking and once without. 1,301 clones including repeat sequences and 181 clones without repeat sequences were additionally selected from the second analysis. The removal of clones with end sequences larger than 300 kb resulted in 1,052 clones to be used for additional map construction. The chromosomal location of 5,105 clones in total, therefore, including 4,053 clones from the primary analysis and 1,052 clones from the secondary analysis were revealed (Figure 1).

In this report, we present the physical genome map of the Hanwoo BAC clones based on alignments of selected Hanwoo BES data against the\textit{B. taurus} genomic sequence that was obtained from a public database (NCBI build 2.1) as has been described. Among the 21,024 BAC clones initially used for analysis, a total of 5,105 clones with paired-end hits were mapped on bovine chromosomes. While the highest numbers of BAC clones (316 clones) were mapped on bovine chromosome 1, the lowest numbers of the BAC clones (56 clones) were mapped on chromosome X. This result indicates that the number of clones mapped on each bovine chromosome is in proportion to the physical sizes of each bovine chromosome (Larkin et al., 2003). Only 265 of the 629 clones with paired-end hits from the Hanwoo secondary radiation hybrid map assembled by the COMPASS program (Everts-van der Wind et al., 2004; Chae et al., 2007) were selected to be included in the map construct, implying that our bovine maps are still in the process of being updated. The calculated average insert size of 132.8 kb in the BAC clones used for the map approximates the previously estimated average size of Hanwoo BAC clones.

The completed sequences for the\textit{B. taurus} chromosome remain to be revealed. The current\textit{Bos taurus} genome construct (2.1) is the assembly produced by the HGSC (human genome sequencing center) at the Baylor College of Medicine. The genome was derived from a male of the Hereford breed of cattle. The sequencing strategy produced a 6.2-fold whole-genome shotgun (WGS) assembly wherein the sequence consists of 4,407 contigs and 10,000N (bp) included between each contig. Moreover, there are 97,891 gaps included in the 4,407 contigs. Recent development of a more precise linkage map with more microsatellite DNA

| Table 2. BLAST analysis |
|-------------------------|
| BLAST results | No. of clones |
|----------------|---------------|
| BLAST trials$^a$ | 20,158 |
| no blast hits | 6,957 |
| significant blast hits$^b$ | 13,201 |
| single-end hits$^c$ | 7,075 |
| • single-hits$^d$ | 4,767 |
| • multiple-hits$^e$ | 2,308 |
| paired-end hits$^f$ | 6,126 |
| • each end hit at different chromosome | 1,226 |
| • each end hit at the same chromosome | 4,900 |
| - each end with single hit | 2,447 |
| - one or both of the ends hit with multiple hits | 2,453 |

$^a$Total number of BAC clones used for blast analysis after editing the BESs.
$^b$Blast hits with significant E-value (E<\textit{e}^{-40}). Only one end or both ends of the BESs show a hit. Each BES hits only one position or more than one position of the bovine genome.
ACKNOWLEDGEMENT

The authors would like to thank all members of the molecular biology laboratory for their assistance. This project was supported by a grant (Code#20050301034426) from the Biogreen 21 program, Rural Development Administration, Korea.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Band, M. R., J. H. Larson, M. Rebeiz, C. A. Green, D. W. Heyen, J. Donovan, R. Windish, C. Steining, P. Mahyuddin, J. E. Womack and H. A. Lewin. 2000. An ordered comparative map of the cattle and human genomes. Genome. Res. 10:1359-1368.

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

Birnbaim, H. C. and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513-1523.

Chae, S. H., J. W. Kim, J. M. Choi, D. M. Larkin, A. E. Wind, H. S. Park, J. S. Yeo and I. Choi. 2007. Chromosomal localization of native Korean cattle (Hanwoo) BAC clones via BAC end sequence analysis. Asian-Aust. J. Anim. Sci. 20:316-327.

Choi, S. H., I. C. Kim, D. S. Kim, D. W. Kim, S. H. Chae, H. H. Choi, I. Choi, J. S. Yeo, M. N. Song and H. S. Park. 2006. Comparative genomic organization of the human and bovine PRNP locus. Genomics 87:598-607.

Everts-van der Wind, A., S. R. Kata, M. R. Band, M. Rebeiz, D. M. Larkin, R. E. Everts, C. A. Green, L. Liu, S. Natarajan, T. Goldammer, J. H. Lee, S. McKay, J. E. Womack and H. A. Lewin. 2004. A 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. Genome. Res. 14:1424-1437.

Everts-van der Wind, A., D. M. Larkin, C. A. Green, J. S. Elliott, C. A. Olmstead, R. Chiu, J. E. Schein, M. A. Marra, J. E. Womack and H. A. Lewin. 2005. A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. PNAS. USA. 102:18526-18531.

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

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

Heaton, M. P., G. P. Harhay, G. L. Bennett, R. T. Stone, W. M. Grosse, E. Casas, J. W. Keele, T. P. Smith, C. G. Chitkows-McKown and W. W. Laegreid. 2002. Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. Mamm. Genome. 13:272-281.

Hong, J. M., S. H. Chae, N. Oriero, D. M. Larkin, C. B. Choi, J. Y. Lee, H. A. Lewin, J. H. Bae, I. Choi and J. S. Yeo. 2005. Identification and chromosomal localization of repeat sequences through BAC end sequence analysis in Korean cattle. J. Genetics. 84:329-335.

Ihara, N., A. Takasuga, K. Mizoshita, H. Takeda, M. Sugimoto, Y. Mizoguchi, T. Hirano, T. Itoh, T. Watanabe, K. M. Reed, W. M. Snelling, S. M. Kappes, C. W. Beattie, G. L. Bennett and Y. Sugimoto. 2004. A comprehensive genetic map of the cattle genome based on 3802 microsatellites. Genome. Res. 14:1987-1998.

Itoh, T., T. Watanabe, N. Ihara, P. Mariani, C. W. Beattie, Y. Sugimoto and A. Takasuga. 2005. A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci. Genomics 85:413-424.

Jurka, J., V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany and J. Walichiewicz. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome. Res.
Kelley, J. M., C. E. Field, M. B. Craven, D. Bocskai, U. J. Kim, S. D. Rounsley and M. D. Adams. 1999. High throughput direct end sequencing of BAC clones. Nucleic Acids Res. 27:1539-1546.

Kim, J. W., J. M. Hong, Y. S. Lee, S. H. Chae, C. B. Choi, I. Choi and J. S. Yeo. 2004. Identification of new microsatellite DNAs in the chromosomal DNA of the Korean cattle (Hanwoo). Asian-Aust. J. Anim. Sci. 17:1329-1333.

Kim, U. J., B. W. Birren, T. Slepak, V. Mancino, C. Boysen, H. L. Kang, M. I. Simon and H. Shizuya. 1996. Construction and characterization of a human bacterial artificial chromosome library. Genomics 34(2):213-218.

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

Sasazaki, S., T. Imada, H. Mutoh, K. Yoshizawa and H. Mannen. 2006. Breed discrimination using DNA marker derived from AFLP in Japanese beef cattle. Asian-Aust. J. Anim. Sci. 19:1106-1110.

Womack, J. E., J. S. Johnson, E. K. Owens, C. E. Rexroad. 3rd., J. Schlapfer and Y. P. Yang. 1997. A whole-genome radiation hybrid panel for bovine gene mapping. Mamm Genome. 8(11):854-856.

Yu, S. L., J. E. Kim, H. J. Chung, K. C. Jung, Y. J. Lee, D. H. Yoon, S. H. Lee, I. Choi, C. D. Bottema, B. C. Sang and J. H. Lee. 2005. Molecular cloning and characterization of bovine PRKAG3 gene: structure, expression and single nucleotide polymorphism detection. J. Anim. Breed Genet. 122:294-301.