Korean BAC Library Construction and Characterization of HLA-DRA, HLA-DRB3

Mi-Hyun Park[1,2], Hye-Ja Lee[1], Jeong Bok[1], Cheol-Hwan Kim[3], Seong-Tshool Hong[2], Chan Park[1], KuChan Kimm[1], Bermseok Oh[1] and Jong-Young Lee[1,]*

[1]Center for Genome Science, National Institute of Health, 5 Nokbun-dong, Eunpyung-gu, Seoul 122-701, Korea
[2]Department of Microbiology, Chonbuk National University Medical School, Jeonju 561-756, Korea
[3]Dong-A SeeTech, Co. Ltd., Seoul Build, 81-12 Nonhyeon-2-dong, Gangnam-gu, Seoul, Korea

A human bacterial artificial chromosome (BAC) library was constructed with high molecular weight DNA extracted from the blood of a male Korean. This Korean BAC library contains 100,224 clones of insert size ranging from 70 to 150 kb, with an average size of 86 kb, corresponding to a 2.9-fold redundancy of the genome. The average insert size was determined from 288 randomly selected BAC clones that were well distributed among all the chromosomes. We developed a pooling system and three-step PCR screen for the Korean BAC library to isolate desired BAC clones, and we confirmed its utility using primer pairs designed for one of the clones. The Korean BAC library and screening pools will allow PCR-based screening of the Korean genome for any gene of interest. We also determined the allele types of HLA-DRA and HLA-DRB3 of clone KB55453, located in the HLA class II region on chromosome 6p21.3. The HLA-DRA and DRB3 genes in this clone were identified as the DRA*010202 and DRB3*01010201 types, respectively. The haplotype found in this library will provide useful information in future human disease studies.

Keywords: BAC, HLA, Human genomic library, PCR screening, Pooling

Introduction

Bacterial artificial chromosomes (BACs) are invaluable tools for a variety of genomic and genetic applications, including the physical mapping of chromosomes and large-scale genomic

---

*To whom correspondence should be addressed.
Tel: 82-2-380-2259; Fax: 82-2-354-1063
E-mail: leejy63@nih.go.kr

sequencing, because of their stability over hundreds of generations, capacity to hold large genomic inserts, and ease of manipulation (McPherson *et al.*, 2001). BACs use F-factor-based vectors and thus can propagate DNA segments up to 300 kb in *E. coli* (Shizuya *et al.*, 1992). Several human genomic or chromosome-specific BAC libraries have been constructed, mapped and sequenced (Kim *et al.*, 1996; Osoegawa *et al.*, 1998; Han *et al.*, 2000). In 2001, the International Human Genome Sequencing Consortium announced the completion of the Human Genome Project, and the sequence of the three billion base pairs of the human genome was published (Lander *et al.*, 2001; Venter *et al.*, 2001). As a result of the Human Genome Project, the majority of human genes are available in a characterized and sequenced BAC format. It is now important to systematically identify the function of the genes and the information contained in the intergenic regions of the human genome (Kim, 2004).

No Korean genomic BAC library has previously been available. Thus, we pursued BAC library construction of Korean genomic DNA for two purposes. First, we wished to establish a standard Korean genomic library to be able to compare the nucleotide diversity of the Korean genome with those of other ethnic groups. This is useful for molecular anthropological studies; DNA sequence data from noncoding regions may more accurately reflect human history than data from coding regions, because noncoding regions are not directly subjected to natural selection (Yu *et al.*, 2001). Secondly, we wished to support functional studies of specific genes and to allow BAC-based comparative genomic hybridization (CGH) analyses by generating a genomic resource for the research community. The average size of human genes is 10 to 15 kb, and the average size of intergenic regions is 25-30 kb. Therefore, it is difficult to clone whole human genes by PCR or other methods. The BAC clones will be useful for obtaining information on entire genes. BAC-based CGH has

already had a significant impact in cancer cytogenetics as a powerful tool for detecting chromosome copy number aberrations - even in epithelial solid tumors, in which tumor - specific genome alterations are difficult to detect using conventional cytogenetics (Veltman *et al.*, 2002; Albertson and Pinkel, 2003; Mantripragada *et al.*, 2004). In this respect, BAC clones are useful resources for constructing a tiling resolution BAC array for cancer research.

In this study, a standard Korean genomic BAC (KBAC) library was constructed, and a three-step pooling PCR screen was conducted to identify specific BAC clones. This library will be a public genomic resource for mapped clones for diagnostic and functional studies by Korean scientists and investigators worldwide.

Materials and Methods

Preparation of insert DNA and BAC vector DNA. Genomic DNA obtained from lymphocytes isolated from a single anonymous male donor was embedded in 2% low - melting point agarose plugs (10 μg/plug). The agarose plugs were incubated for 24 h in lysis buffer and then stored in TE buffer at 4°C. All buffers were prepared according to Asakawa *et al.* (1997). The agarose plugs were partially digested with *Hind*III and then subjected to pulsed-field gel electrophoresis (PFGE) using a CHEF DRIII apparatus (Bio-Rad) at 6 V/cm, for 22 h at 14°C in 0.5× TBE buffer to isolate 70-150-kb DNA fragments. The pBAC-lac vector was provided by Prof. N. Shimizu (Keio University of Medicine). To extract a large volume of BAC vector DNA, a 2-liter bottle containing 500 ml LB media with 6.25 μg/ml chloramphenicol was inoculated with a colony of pBAC-lac-transformed bacteria. The culture was grown overnight at 37°C with shaking and then extracted with a Large-Construct kit (Qiagen, (Valencia, CA)) following the manufacturer's instructions. The detailed protocol for construction of a BAC library was described previously (Asakawa *et al.*, 1997).

Generation of BAC clones. Partially digested genomic DNA was ligated to the pBAC-lac vector. Ligation mixtures were transformed into *E. coli* DH10B cells (ElectroMAX DH10B; Life Technologies) with a GenePulser (Bio-Rad). After transformation, the cells were resuspended in 500 μl SOC medium (2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose, pH 7.0) and incubated for 1 h at 37°C with shaking at 250 rpm. The cells were then spread on LB plates containing 12.5 μg/ml chloramphenicol, 50 μg/ml 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 25 μg/ml isopropyl-thiogalactoside (IPTG), and incubated at 37°C overnight. White colonies were picked with a sterile toothpick, inoculated into 800 μl LB containing 7.5% glycerol and the appropriate antibiotic, and incubated for 16-20 h.

Enzyme digestion and end sequencing of isolated BAC clones. To analyze the KBAC clones, 336 clones were selected randomly from the KBAC library. Individual BAC clones were inoculated into 2 ml LB cultures with 12.5 μg/ml chloramphenicol using deep 48-well plates and incubated overnight. Supercoiled BAC DNA

was isolated using a BAC DNA extraction kit (Qiagen). BAC DNAs were digested with *Not*I and *Hind*III to release the BAC insert, and the digested clones were separated by PFGE. BAC ends were sequenced using BigDye Terminators (Perkin-Elmer Applied Biosystems) with an ABI3730 automated sequencer as follows: 0.5 μg DNA, 2 μl BigDye terminator mix, and 50 pmol M13 universal and reverse primers in a total volume of 20 μl (Oh *et al.*, 2004). This mix was denatured at 95°C for 5 min followed by 50 thermal cycles of 95°C for 45 s, 55°C for 30 s, and 60°C for 4 min. As part of the sequence analysis, a computational tool was designed to provide information on the lengths and chromosomal locations of the BAC clones and to manage BLAST search processes. This tool consists of a process manager and map viewer.

Construction of the BAC clone pooling system and high-density replica (HDR) membranes. BAC clones (100,224) were placed in 1,044 96-well plates and then spotted onto 261 384-well plates with a BioRobotics Microgrid II total array system (TAS, Genomic Solutions (Cambridge, UK)). Individual BAC clones were also gridded onto 17 HDR membranes (22.5 × 7 cm) using a 4 × 4 array pin. The libraries containing 12 pooled BAC clones were gridded onto five HDR membranes (11 × 7 cm) using a 6 × 6 array pin. The libraries containing 48 pooled BAC clones were gridded onto two HDR membranes (11 × 7 cm) using a 5 × 5 array pin.

HLA-allele typing of a BAC clone. To analyze the HLA-DRA and HLA-DRB3 allele types in clone KB55453, the entire exons were amplified with the appropriate primer pairs and sequenced.
1. DRA_E1 F: ctcactttaggtgtttccattgatt, R: tttgacttacttcagtttgtggtga
2. DRA_E2 F: cttcacaagactctgggttctttag, R: taagatatccaaggacttcccttct
3. DRA_E3,4 F: tatgggtttgatcctatcttgttgt, R: atacgaatgtctgattcgttctagc
4. DRA_E5(utr) F: aggtaaagcatggttgcttattatg, R: ttaggttttcctgatgagg actaga
5. DRB3_E1 F: aatagctccccaattaaagtgtttt, R: tttattggtggagatttgagaagaa
6. DRB3_E2 F: ggaggtctccagaacaagctg, R: cacagaagtctccaaggataagaag
7. DRB3_E3 F: acctggctcaggtatcagtttaat, R: tttctacacactgttacagggctatc
8. DRB3_E4,5 F: catcaaggctgtaatatttgaatga, R: cactcattggtgtgtttttacagag
9. DRB3_E6 F: ggcatgagacagaaataataggaaa, R: tttatgcactgccttttaaatgtt
The sequences were verified by comparing them with those in the IMGT/HLA Sequence Database using BLAST HLA (http://www.ebi.ac.uk/imgt/hla/blast.html).

Results

A total of 100,224 clones were prepared from the Korean male donor DNA, and a 2.9 genome-equivalents KBAC library was constructed. To analyze the insert sizes of the KBAC clones, we selected 336 random clones and prepared BAC DNAs. We sequenced the BAC clones and then estimated their lengths by homology searches of GenBank with BLAST. Of the 336 clones, 48 clones were not located on chromosomes by BLAST, including the clones were too much matched on chromosomes and the clones that did not localize to any chromosome. The 288 clones were assigned to a single chromosome. The distribution of 288 KBAC clone sizes is shown in Fig. 1. The average clone size was 86 kb,
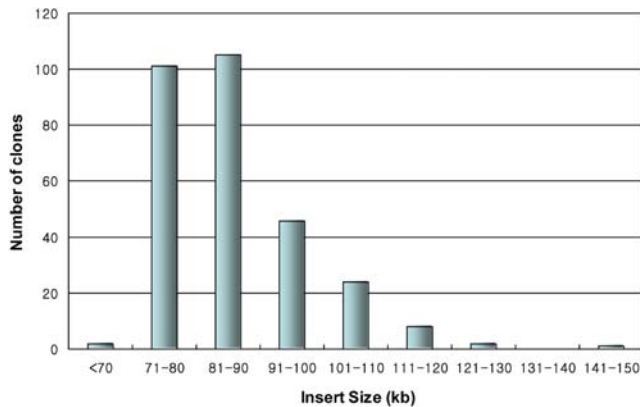
Fig. 1. The size distribution of KBAC clones. Insert size distribution of the 288 end-sequenced KBAC clones. The insert length ranged from 70 to 150 kb, and the average insert size was 86 kb.

with the length ranging from 70 to 150 kb. The sizes of the KBAC clones were also estimated by PFGE after digestion with *Not*I (Fig. 2). The KBAC clones were mapped to chromosomes using a KBAC analysis program developed to analyze the BAC end sequences; the clones were located on all 23 human chromosome pairs (Fig. 3).

To efficiently and more accurately detect a specific BAC clone from the KBAC library, four pooled KBAC libraries were constructed and screened by colony PCR. The four libraries were produced as follows. 1) Twelve KBAC clones were pooled in each well of 87 96-well plates ($12 \times 87 \times 96 = 100,224$ clones). 2) Forty-eight clones were pooled in each well of 21 96-well plates, and 36 clones were pooled in each well of one 96-well plate ($48 \times 21 \times 96 + 36 \times 1 \times 96 = 100,224$ clones). 3) 1,044 clones were pooled in each well of one 96-well plate ($1,044 \times 96 = 100,224$ clones). 4) 4,608 clones were pooled in each of 21 tubes, and 3,456 clones were pooled in one tube ($4,608 \times 21 + 3,456 \times 1 = 100,224$). Thus, the 100,224 KBAC clones were placed into two superpools, either in a pooled 96-well plate or in 22 pooled tubes for the first screen. This allowed faster screening of the library because large numbers of clones could be eliminated with each pool that gave a negative result during a screen, thereby reducing the number of PCRs required.

To test the effectiveness of PCR pooling method for library screening, we performed a screen using PCR primer pairs
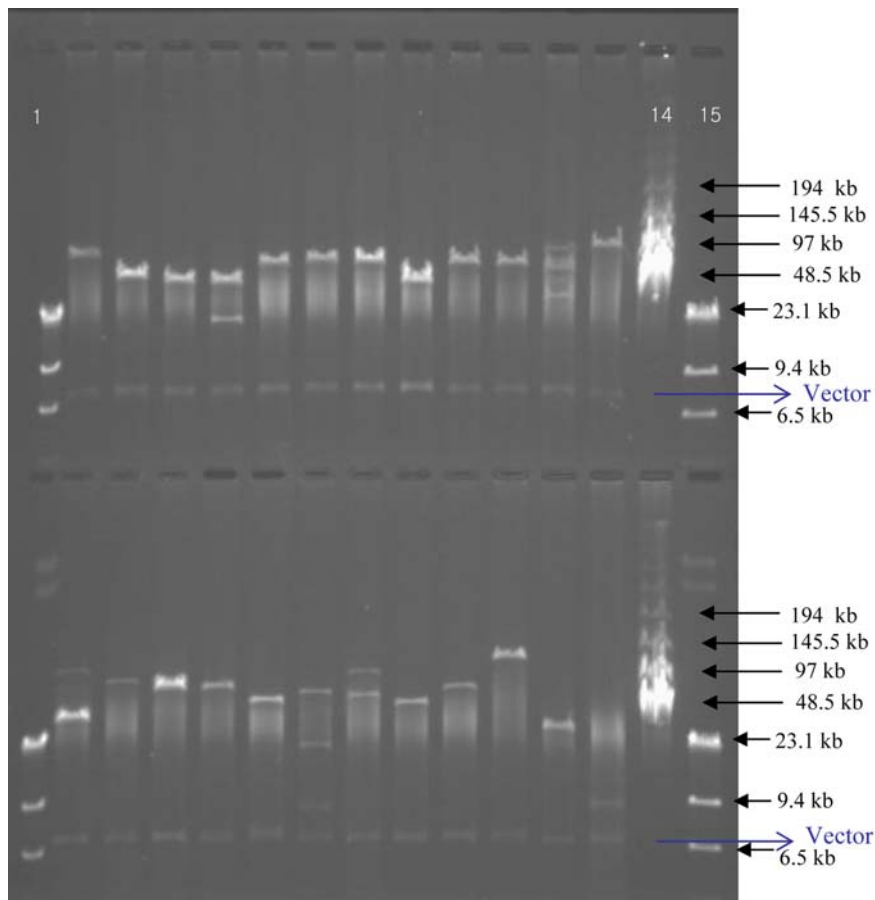
Fig. 2. Analysis of the KBAC clones by enzyme digestion. An ethidium bromide-stained agarose gel shows 24 random BAC clones digested with *Not*I. The sizes of DNA fragments were deduced by comparison with lambda DNA/*Hind*III fragment sizes (lanes 1 and 15) and a lambda DNA ladder (lane 14).
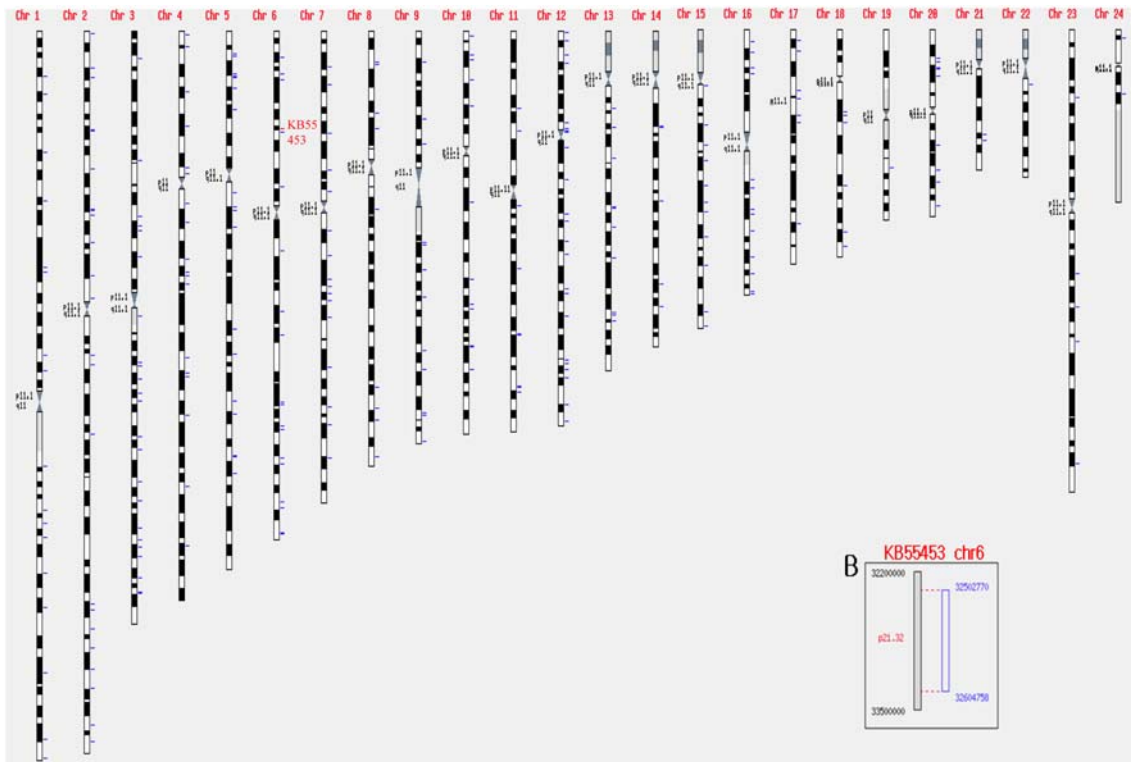
Fig. 3. Chromosome distribution of the 288 end-sequenced KBAC clones. (A) Chromosome mapping of the clones. Chr23 represents the X chromosome, and Chr24 represents the Y chromosome. (B) The exact position and length of clone KB55453 located in chromosome region 6p21.3.

designed for a known clone in the KBAC library (clone KB1105). Figure 4A shows the PCR screening result for KB1105 in the two superpooled libraries. A KB1105-positive PCR product was detected in tube #01 (of 22 tubes) containing 4,608 pooled clones (Fig. 4Aa). A KB1105-positive PCR product was also detected in well E1 of the 96-well plate containing 1,044 pooled clones/well (Fig. 4Ab). We then confirmed the presence of KB1105 in well E5 of plate #01 containing 48 pooled clones/well. The wells in plate #01 (48 pooled clones/well) contained pools from four plates of 12 pooled clones/well (#01~04), and clone KB1105 was detected in plate #01 containing 12 pooled clones/well (Fig. 4Ba). The clones in the wells of the 12-pooled clones/well plate #01 were pooled from 12 plates of single KBAC clones/well (#01~12), and a positive clone was detected in plate #12 (Fig 4Ca). Therefore, we were able to find the KB1105-positive clone located in KBAC_0012_E1.

Next, we screened for the presence of the Notch4 gene in the BAC library. (This gene was not among the 288 randomly selected clones.) In the same manner, we found a KBAC clone containing the Notch4 gene by PCR screening. A Notch4-positive PCR product was detected in tube #05 (of 22 tubes) containing 4,608 pooled clones (Fig. 4Ac). In addition, a Notch4-positive PCR product was detected in well B3 of the 96-well plate of 1,044 pooled clones per well (Fig. 4Ad). We

also confirmed a Notch4-positive clone in well B3 of plate #05 containing 48 pooled clones/well. We detected a positive clone in plate #19 of four plates (#17-#20; 12 pooled clones/well) and a Notch4-positive clone in well #219 of a single clone/well plate (wells #217-228) (Fig. 4Bb, 4Cb). Thus, we were also able to find the Notch4-positive clone located in KBAC_0219_B3. These tests show that the screening pools facilitate an efficient PCR-based screen of this human genomic library for any gene of interest.

One of the 288 clones, KB55453, which is located on chromosome 6, is very interesting because it includes the HLA class II region. KB55453 has 101,988 bp and encompasses the HLA-DRA, -DRB9 and -DRB3 genes (Fig. 3B). We identified the HLA-DRA and HLA-DRB3 allele types in KB55453 (Fig. 5). Each exon region was sequenced with the appropriate primers, and the sequences were compared to those in the IMGT/HLA Sequence Database using BLAST HLA. The HLA-DRA type was identical to DRA*010202, and the HLA-DRB3 type was identical to DRB3*01010201 and DRB3*01010202 in the coding region (Fig. 5A). When we searched for differences between the DRB3*01010201 and DRB3*01010202 allele types in HLA-DRB3, we found a single base deviation (C/G) at nucleotide position -13 in intron 1 (Coquillard *et al.*, 2000). The HLA-DRB3 gene in our clone was the DRB3*01010201 allele (Fig. 5B).
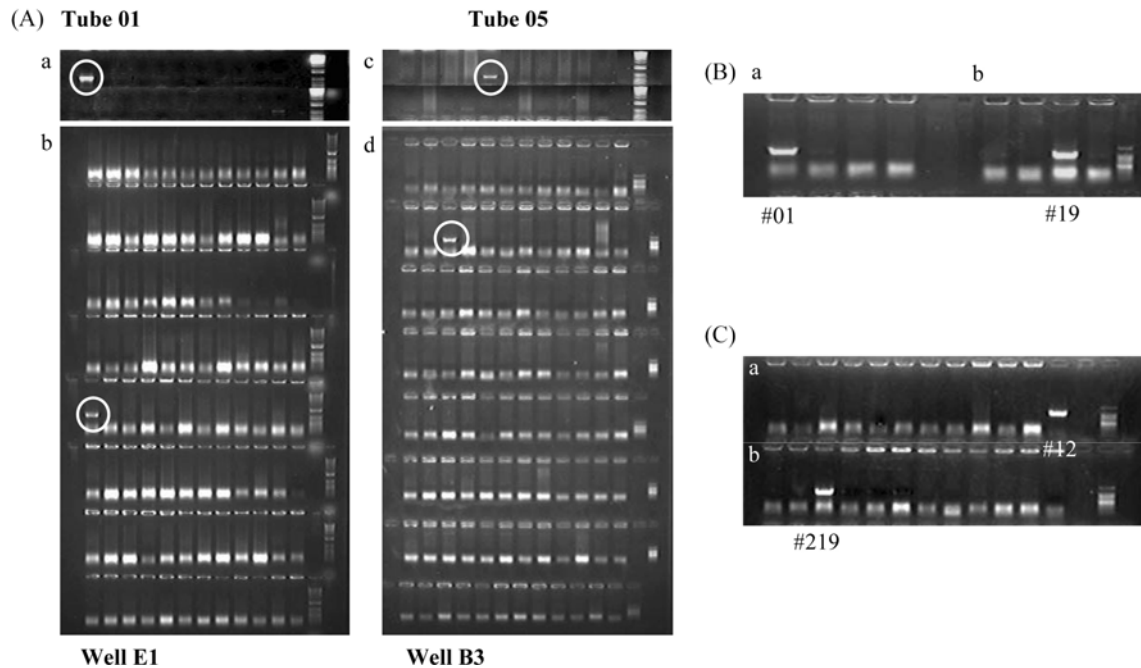
Fig. 4. Identification of clone KB1105 and the Notch4 gene using a three-step PCR screen with optimized PCR primer sets. (A) PCR screen against one 96-well plate containing wells of 1,044 pooled clones/well and against 22 tubes containing 4,608 pooled clones each (the last tube contains 3,456 clones). The KB1105-positive PCR product was detected in tube #01 (a) and in well E1 of the 96-well plate (circle in b). The Notch4-positive PCR product was detected in tube #05 (c) and well B3 of the 96-well plate (circle in d). (B) PCR screening of four 96-well plates containing 12 pooled clones/well. Clone KB1105 was detected in plate #01 among plates #01–#04 (a). Notch4 was detected in plate #19 of #17–#20 96-well plates containing 12 pooled clones/well (b). (C) PCR screening of a single clone. The KB1105 clone was detected in plate #12 of a single-clone/well 96-well plate (out of 12 plates) (a). Notch4 was detected in plate #219 (out of 12 plates) of a single-clone/well 96-well plate (b).
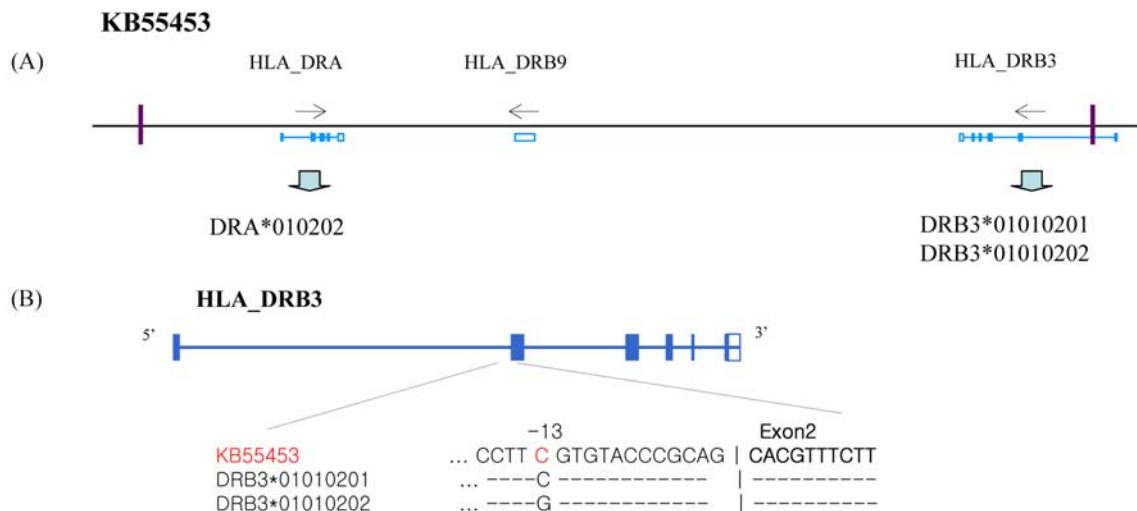
Fig. 5. Identification of the HLA-DRA and HLA-DRB3 allele types in clone KB55453. (A) HLA-DRA and HLA-DRB3 coding region sequences in KB55453 were compared with the IMGT/HLA Sequence Database using BLAST HLA. The HLA-DRA type was identical to DRA*010202, and the HLA-DRB3 type was identical to DRB3*01010201 and DRB3*01010202 in the coding region. (B) The nucleotide sequence (C) at position -13 of intron 1 of KB55453 corresponded to allele DRB3*01010201.

Discussion

We constructed a KBAC library with 2.9× coverage of the genome that contains 100,224 clones of average insert size 86

kb. To identify a faster means of screening the library, we compared two screening methods to detect a specific BAC clone in the KBAC library. Colony hybridization of HDR filters to detect positive clones, although a conventional

procedure, is, in practice, a laborious technique because of non-specific hybridization. However, we prepared HDR filters to accommodate the 100,224 KBAC clones in order to compare this screening method with the three-step PCR pooling strategy we employed. We had to prepare as many as 17 HDR filters (22.5×7 cm) to screen all 100,224 KBAC clones, although we were able to detect positive clones in HDR filters that contained pooled KBAC clones. We also developed a PCR-based screen that is a modification of previously reported methods (Asakawa et al., 1997; Han et al., 2000; He et al., 2004). The clones were pooled into sets of various sizes using different combinations of plates. In this study, we detected a known clone and the Notch4 gene through a three-step PCR screen of the pools. Although this method does not afford faster screening compared with the two-step PCR method using four-dimensional PCR in the previous study, the desired clone can be followed more accurately through each screening step (Asakawa et al., 1997; Suzuki et al., 2000). Regardless, we think that the three-step pooled PCR screen deserves special emphasis because PCR screening for BAC identification is preferable to hybridization.

The HLA gene region is of great interest to many researchers studying a variety of diseases, but some regions are complex and have been difficult to genotype or clone. After a BLAST search of the 288 clones, a BAC clone was found that contains at least 100 kb of chromosomal DNA that was confirmed to include the HLA-DRA, -DRB9 and -DRB3 genes. We typed the HLA alleles of the donor of the KBAC library DNA. The allele type of the HLA-DRB3 gene in clone KB55453 was DRB3*01010201 (Fig. 5). A single-base deviation in intron 1 allowed us to distinguish between alleles DRB3*01010201 and DRB3*01010202 for HLA-DRB3 (Coquillard et al., 2000). In this case, the sequence diversity within introns can be used to separate alleles for unequivocal typing. This implies that an intron-based sequencing approach after exon-based pre-typing can determine exact gene alleles and emphasizes that the sequence information of the KBAC clones, which contain the intron, exon and intergenic regions, is an important resource for genomics researchers. In this study, we have constructed a Korean BAC library that should be useful information for genomic regions that are difficult to analyze by PCR-based methods or by whole-genome shotgun sequencing.

References

Albertson, D. G. and Pinkel, D. (2003) Genomic microarrays in human genetic diseases and cancer. *Hum. Molecular Genet.* **12**, 145-152.

Asakawa, S., Abe, I., Kudoh, Y., Kishi, N., Wang, Y., Kubota, R., Kudoh, J., Kawasaki, K., Minoshima, S. and Shimizu, N. (1997) Human BAC library: construction and rapid screening. *Gene* **191**, 69-79.

Coquillard, G. J., Tang, T. F., Steiner, N., Perlee, L., Ng, J., Hartzman, R. and Hurley, C. K. (2000) DRB3 alleles with variations in the annealing sites of commonly used amplification primers. *Tissue Antigens* **55**, 558-563.

Han, C. S., Sutherland, R. D., Jewett, P. B., Campbell, M. L., Meincke, L. J., Tesmer, J. G., Mundt, M. O., Fawcett, J. J., Kim, U. J., Deaven, L. L. and Doggett, N. A. (2000) Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. *Genome Res.* **10**, 714-721.

He, C. F. and Takao, K. (2004) PCR-based screening BAC library and direct end sequencing of BAC clones. *Yi Chuan Xue Bao* **31**, 1262-1267.

Kim, T. K. (2004) Chemical genomics and medicinal systems biology: chemical control of genomic networks in human systems biology for innovative medicine. *J. Biochem. Mol. Biol.* **37**, 53-58.

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

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A. , Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox,

D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J. (2001) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Mantripragada, K. K., Buckley, P. G., de Stahl, T. D. and Dumanski, J. P. (2004) Genomic microarrays in the spotlight. *Trends Genet.* **20**, 87-94.

McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., Fulton, R., Kucaba, T. A., Wagner-McPherson, C., Barbazuk, W. B., Gregory, S. G., Humphray, S. J., French, L., Evans, R. S., Bethel, G., Whittaker, A., Holden, J. L., McCann, O. T., Dunham, A., Soderlund, C., Scott, C. E., Bentley, D. R., Schuler, G., Chen, H. C., Jang, W., Green, E. D., Idol, J. R., Maduro, V. V., Montgomery, K. T., Lee, E., Miller, A., Emerling, S., Kucherlapati, Gibbs, R., Scherer, S., Gorrell, J. H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garcia, D., de Jong, P. J., Catanese, J. J., Nowak, N., Osoegawa, K., Qin, S., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V. G., Kirsch, I. R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J. F., Hawkins, T., Myers, R. M., Shang, J., Ramirez, L., Schmutz, J., Velasquez, O., Dixon, K., Stone, N. E., Cox, D. R., Haussler, D., Kent, W. J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenberg, J., Chen, X. N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H. S., Sakaki, Y., Shimizu, N., Asakawa, S., Kawasaki, K., Sasaki, T., Shintani, A., Shimizu, A., Shibuya, K., Kudoh, J., Minoshima, S., Ramser, J., Seranski, P., Hoff, C., Poustka, A., Reinhardt, R., Lehrach H. (2001) International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409**, 934-941.

Oh, K. J., Park, Y. S., Lee, K. A., Chung, Y. J. and Cho, T. J. (2004) Molecular characterization of a *thiJ*-like gene in Chinese cabbage. *J. Biochem. Mol. Biol.* **37**, 343-350.

Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**, 1-8.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300-kilobasepair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794-8797.

Suzuki, K., Asakawa, S., Iida, M., Shimanuki, S., Fujishima, N., Hiraiwa, H., Murakami, Y., Shimizu, N. and Yasue, H. (2000) Construction and evaluation of a porcine bacterial artificial chromosome library. *Animal Genetics* **31**, 8-12.

Veltman, J. A., Schoenmakers, E. F., Eussen, B. H., Janssen, I., Merkx, G., van Cleef, B., van Ravenswaaij, C. M., Brunner, H. G., Smeets, D. and van Kessel, A. G. (2002) High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am. J. Hum. Genet.* **70**, 1269-1276.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor, Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine,

J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X. (2001) The sequence of the human genome. *Science* **291**, 1304-1351.

Yu, N., Zhao, Z., Fu, Y. X., Sambuughin, N., Ramsay, M., Jenkins, T., Leskinen, E., Patthy, L., Jorde, L. B., Kuromori, T. and Li, W. H. (2001) Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Mol. Biol. Evol.* **18**, 214-222.