



Analysis of the 1.1-Mb Human $\alpha\delta$ T-Cell Receptor Locus with Bacterial Artificial Chromosome Clones

Cecilie Boysen, Melvin I. Simon and Leroy Hood

Genome Res. 1997 7: 330-338

Access the most recent version at doi:[10.1101/gr.7.4.330](https://doi.org/10.1101/gr.7.4.330)

References This article cites 15 articles, 5 of which can be accessed free at:
<http://genome.cshlp.org/content/7/4/330.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

RESEARCH

Analysis of the 1.1-Mb Human α/δ T-Cell Receptor Locus with Bacterial Artificial Chromosome Clones

Cecilie Boysen,¹ Melvin I. Simon,¹ and Leroy Hood^{2,3}

¹Division of Biology 147-75, California Institute of Technology, Pasadena, California 91125; ²Department of Molecular Biotechnology 357730, University of Washington, Seattle, Washington 98195

Bacterial artificial chromosome (BAC) clones are effective mapping and sequencing reagents. The 1.1-Mb α/δ T-cell receptor locus of humans was mapped and partially sequenced with BAC clones. Seventeen BAC clones covered the 1.1-Mb α/δ locus, with the exception of one small gap that was expected from the coverage that a 3.7-fold BAC library is likely to provide. The end sequences of the BAC inserts could be obtained directly from the BAC DNA by sequencing with the chain terminator chemistry. Five complete BAC inserts were sequenced directly by the shotgun approach. The ends of the 17 BAC inserts were distributed evenly across the locus. By several independent criteria, the BAC clones faithfully represented the genomic DNA, with the exception of a single clone with a 68-kb deletion. These BAC features led to the proposal of a new approach to sequence the human genome.

[The sequenced BAC clones, BAC956, BAC810, BAC480, BAC378, and BAC129, have been submitted to GenBank under accession nos. U85199, U85198, U85197, U85196, and U85195, respectively.]

As the genome project moves from its mapping to its sequencing phase, effective integration of large-scale physical mapping and large-scale DNA sequencing becomes increasingly important. Currently, this process is typically carried out in three discrete steps.

1. A low-resolution physical map is developed from genome-wide or chromosome-specific yeast artificial chromosome (YAC) clones ranging in size from ~100 kb to 1 Mb (Chumakov et al. 1995; Doggett et al. 1995), prepared from genomic DNA that is partially digested by restriction enzymes to generate somewhat random clone overlaps. More recently bacterial artificial chromosome (BAC) clones have been used for physical mapping (Ashworth et al. 1995; Kim et al. 1996). Typically, 2- to 10-fold coverage of the genome or specific chromosomes has been sought.
2. In the past, sequence-ready maps were prepared by subcloning YACs or BACs, after partial restriction digestion, into cosmid vectors with insert sizes typically ranging from 35 to 45 kb. Now many sequencing groups are directly cloning randomly sheared fragments into M13 or plas-

mids for shotgun sequence analysis. Alternatively, cosmid clones can be prepared from genomic or specific chromosomal DNAs. Generally, 5- to 10-fold cosmid clone coverage has been sought. This is perhaps the most difficult step in DNA sequencing because the mapping technologies are neither mature nor robust. Moreover, virtually all projects employing this approach to sequence-ready maps have had significant difficulty with closure. For example, the St. Louis and Sanger groups have now sequenced ~40 Mb of nematode DNA. A significant effort has been put into producing sequence-ready maps. The longest contiguous sequence is 2 Mb, and the average contig is <200 kb (R. Waterston, pers. comm.). Hence, closure is a significant challenge.

3. A minimum tiling (overlap) path of cosmid clones is selected for sequence analysis. Individual cosmid clones are then randomly sheared into ~1-kb fragments for subcloning into bacteriophage M13 vectors. Random M13 inserts are sequenced using M13 primers. Typically 600–900 forward sequence reads are assembled into contigs; contig closure and editing generally requires additional sequences from reverse reads, primer directed sequencing, or selected PCR amplification products.

³Corresponding author.
E-MAIL tawny@u.washington.edu; FAX (206) 685-7301.

This approach has been employed by most large-scale DNA sequencing laboratories (Wilson et al. 1994; Chissoe et al. 1995; Rowen et al. 1996).

The approach described above has several severe limitations. (1) YAC and even cosmid inserts are often chimeric and/or suffer from deletions (Green et al. 1991). The determination of clone fidelity is time consuming, and small deletions can be difficult to identify. (2) YAC clones (or their subclones) must be isolated or purified from yeast DNA. (3) Cosmid inserts may be shorter than tandemly repeated DNA arrays, thus rendering physical mapping and sequence closure across the arrays difficult. For example, the human β T-cell receptor (TCR) locus has five tandemly arrayed 20-kb homology units 90%–92% similar to one another (Rowen et al. 1996). This area was extremely difficult to map and sequence with cosmid clones. (4) The need to subclone DNA into three different vectors and create physical maps at two different levels adds complexity and expense to the potential automation of large-scale DNA sequencing.

BAC inserts appear to offer an interesting alternative to YAC and cosmid inserts as physical mapping and sequencing reagents in that preliminary results indicated that they are relatively stable (infrequent deletions) (Shizuya et al. 1992), are rarely chimeric by analysis with *in situ* hybridization (J. Korenberg, pers. comm.), and can be readily separated from bacterial DNA. To test the advantages of BAC clones both as mapping and sequencing reagents, we generated a physical BAC map of the human α/δ TCR locus from a BAC library with 3.7-fold coverage of the genome.

The human α/δ TCR locus is ideal for testing BAC clones as mapping and sequencing reagents. This locus is present at q11.2 on chromosome 14. First, it is ~1 Mb in length and thus offers a significant mapping and sequencing challenge. Second, the locus encodes ~50 variable (*V*) gene segments, 61 *J_α* gene segments, 3 *D_δ* gene segments, 4 *J_δ* gene segments, and the *C_δ* and *C* genes. The 3' approximately 100 kb of the α/δ locus, encompassing all of the coding elements from *C_δ* to *C_α*, including all the *J_α* gene segments, has been sequenced (Koop et al. 1994), thus offering a superb control for many of the experiments described below. Third, although the region encompassing the *V* elements had not been sequenced, the relative order of many of the *V* gene segments has been determined by deletional mapping (Ibberson et al. 1995); the deletional DNA rearrangement process that joins the *V_α* and *J_α* or *V_δ*, *D_δ* and *J_δ* elements generates T-cell lines or tumors that can be used as mapping reagents because gen-

erally both α/δ loci are rearranged. Fourth, the *V* elements fall into distinct subfamilies whose individual members exhibit 75% or more sequence homology. Members of these subfamilies, ranging in number from 1 to 5, are located across the locus and, hence, probes from a few subfamilies can be used to identify and select BAC clones across the entire *V*-element region. Finally, this region has locus-specific repeats (homology units) that pose challenges to mapping and sequencing of much human DNA.

To test the idea that BAC clones will make suitable reagents for genomic mapping and sequencing, the human TCR α/δ locus was mapped using BAC clones. Most BAC inserts appear to be faithful replicas of genomic DNA. Five of the BAC clones, ranging in size from 86 to 208 kb, were successfully directly sequenced by the shotgun approach after subcloning into M13 bacteriophage vectors. Not only do BAC clones appear to be excellent mapping and sequencing reagents, they also make possible a new approach to sequencing the human genome.

RESULTS AND DISCUSSION

The 1-Mb α/δ Human TCR Was Readily Covered by 17 BAC Clones

Seventeen BAC clones were identified by hybridization to probes for certain *V* and *C* gene segments, probes of cosmid clones mapped previously in this region, or probes obtained by PCR of the ends of BAC inserts. Pulsed field gel electrophoresis (PFGE) suggested that these clones ranged in length from 85 to 240 kb with an average insert size of 137 kb, providing an average 2.3-fold coverage of this locus. The BAC clones were obtained from a 3.7-fold human BAC library with an average insert size of 139 kb (Kim et al. 1996). These BAC clones were analyzed for *V* gene segment arrangement either by hybridization of *V* gene segment probes to restriction enzyme-digested BAC DNA or with *V* gene segment-specific sequence tagged sites (STSS) (Fig. 1). This analysis identified three BAC contigs, two of which ultimately were demonstrated to overlap by only 2 kb. This overlap was not detected until end sequence information from the BAC clones at the ends of the contigs was used to make STSS and probes to hybridize against BAC clones from the other contigs. This leaves the BAC map with a single gap—consistent with what might be expected from a 3.7-fold library. The middle region of the α/δ locus was not screened exhaustively for BAC clones be-

cause we had already obtained a detailed high-resolution (sequence-ready) map of cosmid clones subcloned from YAC clones. The only gap was covered with a P1-derived artificial chromosome (PAC) clone.

BAC Clones Exhibit Striking Genomic Fidelity

The fidelity (e.g., the absence of chimeras, deletions, rearrangements) of the BAC inserts was analyzed by six different methods.

1. All 17 BAC clones were fingerprinted with three different restriction enzymes (*Hind*III, *Eco*RI, and either *Bam*HI or *Pst*I) (Fig. 1). The fingerprints of overlapping clones were compared against one another and against an array of cosmid clones spanning 600 kb in two contigs of the α/δ locus. Almost all fragments in overlapping regions could be matched, except for end fragments, when digested with the restriction enzymes *Eco*RI, *Bam*HI, and *Pst*I. A possible polymorphic *Hind*III site was noted in BAC628. Additional mapping and limited sequence information indicated that BAC196 had undergone an internal deletion of 68.2 kb. The single 3' BAC clone, 705, was checked for fidelity against the 3' 100-kb sequence determined previously. BAC clones 129 and 116 show similar restriction enzyme cleavage patterns at the 5' end. The data indicate that none of the 17 BACs are chimeric and that only BAC196 has a major rearrangement greater than our confidence units of detection, ~1 kb.
2. *V* gene segments were ordered on the BAC clones via PCR or hybridization to Southern blots of BAC DNA digested with the restriction enzymes mentioned above (Fig. 1a). Their locations, with three exceptions, corresponded with those determined previously by deletional mapping and PFGE studies using genomic DNA (Ibberson et al. 1995). We believe that these three exceptions
3. Sequences have been determined for both ends of each BAC clone. This information was used to generate STSs and probes for 15 of the ends, and in all cases they gave appropriately positive results when tested on overlapping BAC clones.
4. Appropriately placed rare cutting restriction sites have been identified across the locus in the BAC clones as well as in YAC and cosmid clones covering half of the region (Fig. 1). The data obtained from the BAC and YAC clones are completely consistent with each other and furthermore agree with PFGE data of genomic DNA determined previously (Satyanarayana et al. 1988; Ibberson et al. 1995).
5. The entire α/δ locus is now sequenced in five BAC, two PAC, and nine cosmid clones (Fig. 1b). When comparing the overlap regions between sequenced BAC (22 kb), BAC and PAC (120 kb), or between BAC and cosmid (78 kb) clones, we have found no discrepancies that cannot be accounted for by polymorphisms.
6. Of the 34 end sequences from the 17 BAC clones, 32 match against the complete sequence of the α/δ locus (Fig. 1b), whereas the last two ends extend outside the sequenced region. The sizes of the BAC inserts determined from DNA sequence comparisons match the sizes determined by

represent mapping errors in the original analysis. Furthermore, in all cases where genomic DNA samples were included on the Southern blot, hybridization bands in the BAC clones matched their genomic counterparts except in cases where the BAC insert ended close to the probe used. For example, BAC363 ends in a *Hind*III site found in the middle of a *V* gene segment, and thus that *V* gene segment probe gave two bands for *Hind*III in BAC378, BAC274, and genomic DNA. BAC363 only showed one band corresponding to one of the two *Hind*III fragments. Likewise, this probe differed in its *Eco*RI pattern from BAC363, as one of the *Eco*RI sites was missing.

Figure 1 Physical maps of the human α/δ TCR locus. (a) Seventeen BAC clones, with sizes as determined by PFGE indicated, spanning the α/δ locus. The BAC clones were characterized in several different ways: restriction enzyme analysis (*Hind*III and *Eco*RI sites shown), by localization of TCR elements by hybridization (●), by STS mapping (■), and by PFGE analysis. These results were compared to similar data obtained for overlapping YAC and cosmid clones. The relative order for most across the locus is known because of a cosmid map of high redundancy. The exact order of several probes could not be determined in several BAC clones not covered by the cosmid map. The gap between the two BAC contigs is indicated as is the deletion in BAC196. (b) The positions of the sequenced BAC, PAC, and cosmid clones are indicated. (c) End sequences obtained from each BAC clone were compared to the final sequence and their positions indicated. The complete names of the BAC and PAC clones presented are as follows: BAC10G10, BAC116B10, BAC129A3, BAC135H10, BAC196E6, BAC259E10, BAC274H3, BAC363D9, BAC378F4, BAC417G3, BAC480F2, BAC540C5, BAC628B12, BAC705G10, BAC810D3, BAC916C11, BAC956H7, PAC161M22, and PAC230I21.

BOYSEN ET AL.

PFGE except for the deleted BAC196. Its insert size was determined to be 100 kb by PFGE and 163.7 kb by sequence comparison, suggesting a 64-kb deletion. We colony purified the BAC196 clone again from the original library. The new BAC196 clone was analyzed by *NotI* digestion and PFGE analysis and gave a band at 170 kb, suggesting that this was the original clone. We digested it further with *EcoRI* and *HindIII* to compare it with the previous BAC196 clone, and clearly, the previous clone was a deleted version of the original.

These observations suggest that there are no chimeras, major rearrangements, or deletions >1 kb in these 17 BAC clones, apart from the single deletion observed in BAC 196.

Although the marker density may be higher in this region than in other regions, if the genome project goal of 100 kb average of STS markers is achieved, then BAC contigs could be constructed across the genome. Closing these contigs will require significant additional physical mapping. This is one of the major reasons that we have recently proposed a new approach to sequencing the human genome that avoids this difficulty (see below).

BAC Inserts Can Be Sequenced Directly at Their 5' and 3' Ends

We have successfully sequenced all 34 insert ends directly from DNA of the 17 BAC clones. In total we have sequenced both ends of 34 BAC and 22 PAC clones using the T7 and SP6 primers. Of the 112 sequences, 110 were successful on the first attempt (98% success). By comparing the end sequences with their final shotgun (high redundancy) sequenced counterparts, the average high-quality read length was determined to be ~500 bp with an error rate of ~0.4%. This read length indicates the number of bases after the first 20–30 bp, which often contains a few errors or ambiguities, and before the error rate rises towards the end of the read.

Of 70 end sequences in the α/δ TCR locus, 23 (33%) contained genome-wide repeats (Table 1). Of these 23 sequences, 15 contained 100 bp or more of unique sequence. Only 11% of the end sequences lacked unique sequences.

The BAC Clone Ends Appear to Be Evenly Distributed Across the α/δ Locus

Of 34 ends from the 17 BAC clones spanning the α/δ locus, 32 fall within the locus. The distribution of

BAC ends across the α/δ locus appears relatively even, keeping in mind that the middle region is underrepresented in BAC clones because it was sequenced previously with cosmid clones (Fig. 1). Accordingly, the restriction enzyme digestion that produced the BAC inserts appears to have randomly cleaved the genomic DNA and random DNA fragments were integrated effectively into the BAC vector. Admittedly the data are quite limited for this analysis. Nevertheless, if the enzyme cleavage sites are random across most of the genome and these fragments can be integrated efficiently into the BAC vector, then BAC libraries will be excellent reagents for large-scale DNA mapping and sequencing.

BAC Inserts Can Be Sequenced Effectively by the Shotgun Approach

We successfully sequenced five BAC clones ranging in size from 86 to 208 kb by the shotgun approach. The BAC clones were randomly sheared and cloned into M13 bacteriophage, and the M13 fragments were sequenced to an average coverage of six- to eight-fold. Closure was generally achieved either by reverse sequence reads from appropriately located M13 clones or by synthesizing DNA primers at the gap edge and using these for walking on appropriate M13 clones. We also employed primer-directed walking directly on the BAC DNA with chain terminator DNA sequencing. No difficulties were experienced in these sequence analyses. The order of the *V* elements in these BAC inserts is completely consistent with that obtained from the physical map analyses mentioned above.

BAC Inserts Exhibit Several Features That Facilitate Physical Mapping

BAC clones are single-copy vectors and, accordingly, exhibit relative stability in clonal growth. BAC clones have several features that are attractive for physical mapping. (1) They faithfully represent chromosomal sequences. By *V* gene segment analysis, restriction enzyme analyses, and end sequence analysis, all 17 BAC clones lying across the human α/δ TCR locus, apart from BAC 196, faithfully reflect genomic sequence to the varying levels of discrimination analyzed. At the base pair level, no discrepancies were found in 220 kb of BAC sequences overlapping other BAC, PAC, or cosmid sequences other than minor polymorphisms, mainly single-nucleotide substitutions and microsatellite variations. This is the first time that the fidelity of BAC

Table 1. Genome-Wide Repeats in the 23 (33%) End Sequences from 70 BAC Ends

Type	No.	Similarity (%)	Some unique	Repeat only
<i>Alu</i>	10	86–93	10/10 (>100 bp)	0
LINE	9	77–92	2/9 (>300 bp)	7
LTR ^a	4	84–93	3/4 (>250 bp)	1

^a(LTR) Long terminal repeat.

inserts has been tested directly on a large scale. (2) BAC inserts appear to be rarely chimeric—none of the BACs that we analyzed were chimeric. In contrast, six of nine YAC clones obtained across this region were clearly chimeric. Moreover, J. Korenberg (pers. comm.) has mapped >2000 BAC clones by chromosomal in situ hybridization; from this she concludes that <5% are chimeric. (3) The BAC clones appear to delete or rearrange rarely (only 1/17 clones exhibited a deletion). In contrast, cosmid libraries appear to contain a higher percentage of defective clones. For example, 13% of the 234 cosmid clones analyzed across the human β TCR locus were found to have defects (deletions, chimeras, or rearrangements), as detected by detailed restriction enzyme mapping and end sequencing data (L. Rowen, pers. comm.). (4) The average BAC clone is of sufficient length to span most locus-specific clusters of tandem repeats. In the human β TCR locus, we identified one tandem cluster of five 20-kb repeats (e.g., a repeat length of 100 kb). Having mapping (and sequencing) reagents that span these tandem clusters facilitates the mapping process significantly. (5) BAC clones seem to be more or less evenly distributed across the 1-Mb TCR α/δ region. One gap of 3 kb was not covered, but this is expected from a library with a 3.7-fold coverage of the human genome. A critical question is whether this successful BAC coverage will extend across the entire genome. We point out that loci, such as the human TCR α/δ locus, with multiple homology units and their tendency to delete probably present one of the most difficult case scenarios for genomic cloning. (6) For shotgun sequencing, BAC clones can be used directly to prepare sequence-ready maps. Thus, only one library construction and one mapping step is necessary, in contrast to most large-scale DNA sequencing projects involving two library constructions and two mapping steps (YACs and cosmids). This increased efficiency will greatly facilitate the automation necessary for large-scale sequencing projects. (7) BAC ends can readily be se-

quenced, thus suggesting a strategy that completely eliminates physical mapping in the large-scale DNA sequencing procedures (see below). (8) An arrayed BAC library allows the easy placement of other landmark features on the BAC clones [e.g., STSs, expressed sequence tags (ESTs), polymorphic microsatellites, etc.]. This will permit the easy transfer of all of the landmarks to BAC clones identified previously.

BAC Clones Are Good Sequencing Reagents

We have been successful in the shotgun sequence analysis of five BAC clones ranging in size from 86 to 208 kb. Furthermore, the 208-kb BAC129 clone has significant genome-wide and locus-specific repeats, yet we were able to sequence this large insert without difficulty. The shotgun clone coverage is distributed quite evenly, suggesting that BAC clones can be sheared randomly. The assembly of long sequence contigs from randomly sequenced M13 inserts has been made possible by new methods developed for base calling, quality assessment, and sequence assembly by P. Green (pers. comm.). Apart from the fact that BAC clones can be sequenced by shotgun analysis, these clones do have several advantages for sequencing, in part similar to those mentioned above for mapping: (1) BAC clones appear to faithfully represent the genome, that is, BAC clones rarely delete, rearrange, or are chimeric; (2) the average BAC clone can readily traverse the largest locus-specific repeats identified to date; (3) the BAC vector is only 7 kb in length, thus representing fourfold lower percentage of the insert than found in the cosmids currently used most frequently for shotgun sequencing (7-kb vector/150-kb insert as compared to an 8-kb vector/30- to 40-kb insert).

In summary, BAC clones are attractive sequencing reagents because of their genomic fidelity, size, stability, and potential for eliminating one cloning and one mapping step in the traditional large-scale shotgun sequencing strategy. BAC inserts also offer

BOYSEN ET AL.

the possibility of sequencing the human genome without the need for physical mapping.

The Human Genome May Be Sequenced by the Sequence-Tagged Connector Approach

A major limitation in conventional large-scale DNA sequencing operations lies in the production of sequence-ready maps with minimum clone overlaps. The YAC or BAC inserts may be derived from genomic DNA or chromosome-specific DNA. Current approaches typically map YAC or BAC inserts, select a minimum tiling path of YAC or BAC clones, subclone each YAC or BAC insert after random restriction enzyme cleavage into cosmid vectors, map the cosmid inserts, and then select a minimum tiling path of cosmid inserts for M13 bacteriophage shotgun sequencing. Thus, three libraries must be constructed and two must be mapped. These efforts have several difficulties: (1) YAC and cosmid inserts tend to delete, rearrange or form chimeras; hence, considerable effort must be spent in initial clone characterization. (2) Creating the sequence-ready cosmid maps employs mapping technologies that are neither robust nor very well automated (Gillett et al. 1996; L. Rowen, K. Wang, P. Charmley, M.E. Ahearn, C. Boysen, D. Seto, B. Paepa, L. Chen, D. Nickerson, and L. Hood, in prep.). Hence, the mapping throughput cannot begin to match the current throughput of the automated DNA sequencers. (3) Finally, the problem of closure of sequence-ready maps is significant (Olson and Green 1993). Extensive nonautomated human effort will be required to provide sequence-ready maps for all of the large-scale sequencing efforts scattered across the world.

We recently proposed that the sequence-tagged connector, or STC, approach eliminates one cloning step and two mapping steps (Venter et al. 1996). We propose to develop a BAC library with 15-fold coverage and average insert size of 150 kb, array the 300,000 clones, fingerprint each, sequence both ends, and make these data immediately available on the World Wide Web, together with access to the BAC library. At this level of redundancy, the end sequences will provide a sequence marker of 300–500 bp, on average, every 5000 bp across the genome. The BAC library and the data pertaining to it will enable the construction of minimal sequence tiling paths of BAC clones in the following way. First, a “seed” BAC clone in an interesting chromosomal region can be sequenced to contiguity. Then, once a seed BAC (150-kb insert) has been sequenced, it is immediately connected to 30 other BACs by one of their end sequences, termed se-

quence tagged connectors, or STCs. A minimum overlapping BAC clone can be selected for sequencing in each direction and the process repeated. Hence, long contiguous stretches of the genome can be sequenced simultaneously by selecting seed BAC clones from many different sites.

The STC approach has two striking advantages. First, any portion of the human genome is immediately accessible. Because the minimum overlap clones are identified in the computer, the rate at which given chromosomal regions can be sequenced is limited only by the throughput of the sequencing procedures. Second, this process can easily be automated. Robots are now being developed that will have the capacity to prepare insert DNAs and 5,000–10,000 sequencing reactions per day.

Major genome centers propose to scale up to sequence 100 Mb or more of finished DNA sequence within the next 2 to 3 years. On average, 2¹/₄ BAC clones (averaging 150-kb inserts) or 11 cosmid clones (averaging 30-kb inserts) must be sequenced each day to finish 100 Mb in 1 year. We suggest that sequence-ready mapping with STCs on the computer is the most efficient way to meet these throughput needs. BAC clones appear to be ideal mapping and sequencing reagents and certainly will play a major role in the large-scale DNA sequencing phase of the Human Genome Project.

METHODS

DNA Source

BAC clones were obtained from a human BAC library at California Institute of Technology, (Pasadena). This library was developed from a normal human male fibroblast cell line, [American Type Culture Collection (ATCC) CRL 1905: CCD-978Sk] (Shizuya et al. 1992) and had a 3.7-fold coverage of the human genome. This cell line was also used in PFGE analysis of human genomic DNA. PAC clones are from a PAC genomic library constructed from a normal male fibroblast cell line maintained by Genome Systems, Inc. (Ioannou et al. 1994).

BAC Library Screening

To obtain BAC clones specific for the TCR α/δ locus, DNA fragments containing TCR variable gene segments were amplified by PCR and used as probes. Then they were labeled with ³²P using a random labeling approach (T7 QuickPrime, Pharmacia or Multiprime DNA Labeling System, Amersham) and hybridized overnight at 65°C to the BAC library membranes in SET [0.6 M NaCl, 0.02 M EDTA, 0.2 M Tris-HCl (pH 8.0), 2% SDS, 0.1% pyrophosphate]. The membranes were washed once for 10 min in 1× saline sodium citrate (SSC) + 0.1% SDS, followed by two or three washes in

ANALYSIS OF THE HUMAN α/δ LOCUS

$0.1 \times \text{SSC} + 0.1\% \text{ SDS}$ at 65°C for 10–20 min each. Positive clones were identified after exposure at -70°C to Kodak X-AR film with an intensifying screen overnight or longer. Probes specific for the ends of the BAC inserts were also used in library screening. These probes were made as described above by labeling PCR products from the known end sequences. Whole cosmid and BAC clones mapped to the TCR α/δ region were also used as probes in hybridization to the BAC library. To prevent vector hybridization, cosmid and BAC DNAs were digested with the restriction enzyme *NotI* and the vector was separated from the inserts by PFGE (see below). Insert DNA was excised from the gel and extracted from the agarose using beads (Sephaglas BandPrep, Pharmacia or Qiaex, Qiagen). When using ^{32}P labeled cosmid or BAC clones as probes, cold vector DNA, human Cot-1 DNA or total placental DNA, and total *Escherichia coli* DNA were used to suppress hybridization of repeat sequences and contaminating vector and *E. coli* DNA.

DNA Preparation

Total human genomic DNA from the same cell line used to make the BAC library was prepared in low melting point (LMP) agarose. Cells were washed twice in phosphate-buffered saline and resuspended to 10^8 cells/ml in PBS. The cells were then warmed to 37°C before they were mixed with an equal volume of melted 1% LMP-agarose and poured into molds. The solidified plugs were incubated for 14–18 hrs at 50°C in a solution of 0.5 M EDTA (pH 9.0), 1% sarcosyl, and proteinase K (0.5 mg/ml), rinsed, and fresh solution was added for an additional 14–18 hr. The plugs were then rinsed and stored at 4°C in 0.5 M EDTA.

BAC DNA was prepared from cell cultures (12.5 $\mu\text{g/ml}$ of Luria broth plus chloramphenicol) using standard alkaline lysis procedures (Sambrook et al. 1989). Mini-preparations were made either by hand without organic extractions or by an automated miniprep machine, the Autogen 740 (Integrated Separation Systems).

Southern Blots and Hybridizations

BAC DNA was digested with various restriction enzymes according to the suggestions of the manufacturer. The resulting fragments were separated in a 0.8% agarose gel and transferred to nylon membranes by capillary action using 0.4 N NaOH. The membranes were rinsed twice in $2 \times \text{SSC}$ before use. PFGE gels were irradiated at 254-nm UV light for 45 sec in the presence of ethidium bromide to nick the DNA before blotting. The blots were prehybridized in hybridization solution (50% formamide, $5 \times \text{SSC}$, 0.02 M sodium phosphate at pH 6.7, 100 $\mu\text{g/ml}$ of denatured salmon sperm DNA, 1% SDS, 0.5% nonfat dry milk, and 10% dextran sulfate) at least 30 min prior to hybridization. DNA probes were labeled and hybridized at 37°C for 14–18 hr followed by washing. Washing conditions were varied in their concentration of SSC from $0.1 \times$ to $2 \times \text{SSC}$, dependent on the desired stringency.

PFGE

Large DNA molecules were separated in 1% agarose in $0.5 \times \text{TBE}$ at 14°C using different PFGE apparatuses, either homemade or from BioRad. Voltage applied was 6 V/cm;

switch times and total time depended on the desired range of fragment lengths to be separated (Birren and Lai 1993).

DNA Sequencing

BAC DNA was completely sequenced by the random or shotgun method (C. Boysen, in prep.). DNA was sonicated using different sonication devices and the ends made blunt with mung bean nuclease (GIBCO, BRL). The end-repaired DNA fragments were separated on a 1% agarose gel and those of 1–3 kb excised and purified with beads as described above. The purified DNA was subcloned into an M13 vector (Sigma) that had been digested with the restriction enzymes *HincII* or *SmaI* and dephosphorylated with calf intestine alkaline phosphatase (Boehringer Mannheim). Single-stranded DNA was prepared from 1-ml cultures of clear plaques, and sequenced using different versions of dye primer cycle sequencing kits following the manufacturer's protocols (Perkin Elmer). The reactions were run on ABI 373 Sequencers. Phred (program by P. Green, pers. comm.) was used to call the bases from the ABI trace data and assign quality values to them. The sequences were then compared to a set of sequences, including *E. coli*, M13- and BAC vector sequences, using cross_match (P. Green, pers. comm.; URL, <http://www.genome.washington.edu/>). Entire sequences or portions thereof with significant hits were not considered in further assembly. Both sequences and quality assignments were taken into consideration during assembly by the program phrap (P. Green, pers. comm.). The assembly was viewed using consed (Chris Abajian and David Gordon, pers. comm.) and extra sequencing reactions to be performed to close gaps or cover ambiguous regions determined. Consed was also used for the final editing.

End sequencing or primer walking-directed sequencing on BAC clones was performed as described by C. Boysen, M.I. Simon, and L. Hood (in prep.). In short, DNA prepared from 3-ml overnight culture was used for each fluorescent terminator sequencing reaction. Each reaction further contained 50 pmoles of primer and 16 μl of terminator sequencing reaction mix (ABI Prism) Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase, FS (Perkin Elmer), for a total of 40 μl . Oligonucleotides were synthesized, deprotected, dried, and resuspended in double distilled water to either 25 or 50 μM . For end sequencing, standard T7 and SP6 primers were used (T7, TAATACGACTCAC-TATAGGG; and SP6, ATTTAGGTGACACTATAG). Cycling was performed in the thermal cycler (GeneAmp 9600, Perkin Elmer, following the instructions in Perkin Elmer's protocol P/N 402078), with the addition of an initial denaturation step. The thermal cycler was heated to 96°C before inserting the tubes. These were kept at 96°C for 4 min, followed by 25 cycles of 10 sec at 96°C , 5 sec at 50°C , and 4 min at 60°C . After cycling, the reaction was purified by passing over a CentriSep spin column (Princeton Separations), dried, and run on a 373 DNA Sequencer, Stretch (Applied Biosystems) using either a 36- or 48-cm well to read plates (4.75% or 4% acrylamide gel, respectively).

ACKNOWLEDGMENTS

We thank Hiroaki Shizuya for useful hints in the handling of BAC clones, Tawny Biddulph for typing, and Lee Rowen and Todd Smith for critically reviewing this manuscript. This work was supported by a grant from the Department of Energy.

BOYSEN ET AL.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Ashworth, L.K., M. Alegria-Hartman, M. Burgin, L. Devlin, A.V. Carrano, and M.A. Batzer. 1995. Assembly of high-resolution bacterial artificial chromosome, P1-derived artificial chromosome, and cosmid contigs. *Anal. Biochem.* 224: 564–571.
- Birren, B. and E. Lai. 1993. *Pulsed field gel electrophoresis: A practical guide*. Academic Press, San Diego, CA.
- Chissoe, S.L., A. Bodenteich, Y.F. Wang, Y.P. Wang, D. Burian, S.W. Clifton, J. Crabtree, A. Freeman, K. Iyer, L. Jian, et al. 1995. Sequence and analysis of the human ABL gene, the BCR gene, and regions involved in the Philadelphia chromosomal translocation. *Genomics* 27: 67–82.
- Chumakov, I.M., P. Rigault, I. Le-Gall, C. Bellann'e-Chantelot, A. Billault, S. Guillou, P. Soularue, G. Guasconi, E. Poullier, I. Gros, et al. 1995. A YAC contig map of the human genome. *Nature* (Suppl. 6547) 377: 175–297.
- Doggett, N.A., L.A. Goodwin, J.G. Tesmer, L.J. Meincke, D.C. Bruce, L.M. Clark, M.R. Altherr, A.A. Ford, H.C. Chi, B.L. Marrone, et al. 1995. An integrated physical map of human chromosome 16. *Nature* (Suppl. 6547) 377: 335–365.
- Gillett, W., L. Hanks, G.K.S. Wong, J. Yu, R. Lim, and M.V. Olson. 1996. Assembly of high-resolution restriction maps based on multiple complete digests of a redundant set of overlapping clones. *Genomics* 33: 389–408.
- Green, E.D., H.C. Riethman, J.E. Dutchik, and M.V. Olson. 1991. Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* 11: 658–669.
- Ibberson, M.R., J.P. Copier, and A.K. So. 1995. Genomic organization of the human T-cell receptor variable α (TCRAV) cluster. *Genomics* 28: 131–139.
- Ioannou, P.A., C.T. Amemiya, J. Garnes, P.M. Kroisel, H. Shizuya, C. Chen, M.A. Batzer, and P.J. de Jong. 1994. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* 6: 84–89.
- Kim, U.-J., H. Shizuya, H.-L. Kang, S.S. Choi, C.L. Garrett, L.J. Smink, B.W. Birren, J.R. Korenberg, I. Dunham, and M.I. Simon. 1996. A bacterial artificial chromosome-based framework contig map of human chromosome 22q. *Proc. Natl. Acad. Sci.* 93: 6297–6301.
- Kim, U.-J., B. Birren, Y.-L. Sheng, 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: 213–218.
- Koop, B.F., L. Rowen, K. Wang, C.L. Kuo, D. Seto, J.A. Lenstra, S. Howard, W. Shan, P. Deshpande, and L. Hood. 1994. The human T-cell receptor TCRAC/TCRDC ($C\alpha/C\delta$) region: Organization, sequence, and evolution of 97.6 kb of DNA. *Genomics* 19: 478–493.
- Olson, M.V. and P. Green. 1993. Criterion for the completeness of large-scale physical maps of DNA. *Cold Spring Harbor Symp. Quant. Biol.* 58: 349–355.
- Rowen, L., B.F. Koop, and L. Hood. 1996. The complete 685 kilobase DNA sequence of the human beta T cell receptor locus. *Science* 272: 1755–1762.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Chapter 1, pp. 1.25–1.28. Cold Spring Harbor, NY.
- Satyanarayana, K., S. Hata, P. Devlin, M.G. Roncarolo, J.E. De Vries, H. Spits, J.L. Strominger, and M.S. Krangel. 1988. Genomic organization of the human T-cell antigen-receptor α/δ locus. *Proc. Natl. Acad. Sci.* 85: 8166–8170.
- Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci.* 89: 8794–8797.
- Venter, J.C., H.O. Smith, and L. Hood. 1996. A new strategy for genome sequencing. *Nature* 381: 364–366.
- Wilson, R., R. Ainscough, and K. Anderson, C. Baynes, M. Berks, J. Bonfield, J. Burton, M. Connell, T. Copsey, J. Cooper, et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* 368: 32–38.

Received July 18, 1996; accepted in revised form February 7, 1997.