We have constructed a complete coverage BAC contig map that spans a 12-Mb genomic segment in the human chromosome 16p13.1–p11.2 region. The map consists of 68 previously mapped STSs and 289 BAC clones, 51 of which—corresponding to a total of 7.721 Mb of genomic DNA—have been sequenced, and provides a high resolution physical map of the region. Contigs were initially built based mainly on the analysis of STS contents and restriction fingerprint patterns of the clones. To close the gaps, probes derived from BAC clone ends were used to screen deeper BAC libraries. Clone end sequence data obtained from chromosome 16-specific BACs, as well as from public databases, were used for the identification of BACs that overlap with fully sequenced BACs by means of sequence match. This approach allowed precise alignment of clone overlaps in addition to restriction fingerprint comparison. A freehand contig drawing software tool was developed and used to manage the map data graphically and generate a real scale physical map. The map we present here is ~3.5 × deep and provides a minimal tiling path that covers the region in an array of contiguous, overlapping BACs.
facilitated initial access to BAC libraries to identify BACs corresponding to the target region. A set of nonoverlapping BACs identified by screening BAC libraries with the STSs were subjected to shotgun sequencing prior to the completion of the map (Loftus et al. 1999). The sequence data were used for the subsequent contig extension and gap closure based on the sequence matches with BAC end sequences that permit precise alignment of clone overlaps. Here we present a complete coverage BAC contig map spanning 12 Mb, drawn to scale, which provides a high-resolution road map for physical and genetic markers and for the complete sequencing of this region.

RESULTS

Initial Framework Contigs

The goal of the project was to generate a BAC contig map with complete coverage of the 16p13.1–11.2 region and provide a minimally redundant BAC set for sequencing. The initial set of BACs were identified using 68 STS markers (Table 1) mapped to the target region by the previous YAC-based mapping (Doggett et al. 1995). These markers are concentrated in ~15 Mb of the target region excluding the centromere and pericentric regions that are poorly covered by STS markers. Pooled human library A was screened using the PCR method as described previously (Kim et al. 1996). A total of 175 positive BACs were identified from the 3.5 x library A. For some STSs that failed to yield positives from PCR–STS screening (D16S732, D16S407, D16S2899, D16S2719, D16S414, D16S497, D16S2893, D16S2828, D16S741, D16S774, D16S519, D16S2746, D16S2852, D16S2891, D16S780, D16S2881, D16S2805, D16S2778, D16S2855, D16S2868, D16S2734), gel-isolated PCR products were used as probes for screening other libraries. As a result, additional BACs including 15 from library D and 49 from the Rosewell Park Cancer Institute (RPCI) library were identified. Inserts were isolated from the initial positive clones by NotI digestion and separation on preparative pulsed field gels for use as probes for further library screening. High-density colony filters were prepared for library BC and D (a synopsis of Caltech BAC libraries is provided on the web site http://www.tree.caltech.edu/lib_status.html) using the Q-Bot robotic work station.

Clone Characterization

All of the clones identified by PCR screening or colony hybridization were picked from the arrayed libraries, streaked on plates for single colony isolation, and characterized by HindIII digestion, sizing, restriction fingerprinting, and clone end sequencing, as described in Methods. At least two single colonies were isolated from each positive BAC and tested for consistency in their HindIII digestion patterns to avoid clone mixtures that occasionally occur in arrayed libraries. Highly unstable clones also showed inconsistencies among different single colonies due to rapid rearrangement or degradation. Of the BACs characterized thus far, ~4% were shown to be unstable (not shown). DNA preparation is often difficult and unsuccessful for some of these unstable BACs due to the partial or complete loss of clones. Chromosomal localization of a total of 76 clones was confirmed by FISH analysis. These BACs, which were FISH mapped to the expected regions, served as anchors for the localization of the associated contigs. A complete list of BACs identified by STS–PCR screening is posted on http://www.tree.caltech.edu/chr16BAC_STS_map.html. Overlaps between clones were determined based on STS contents and restriction fingerprint analysis. A set of nonoverlapping or minimally overlapping BACs was selected from these contigs for sequencing at TIGR (Loftus et al. 1999). BAC end sequence data obtained from chromosome 16-specific BACs and from random BACs from libraries constructed at Caltech and RPCI were used to precisely align the clone overlaps against the completely sequenced BACs through sequence match. Figures 1 and 2 represent examples of the fingerprint gel analysis image and the sequence alignment between a BAC sequence and BAC end sequences, respectively.

Library Walking and Gap Closure

Seventy-seven OVERGO probes derived from BAC end sequences were used for further library screening (Table 2). A total of 20 x coverage Caltech libraries and the 12 x human BAC library (RPCI-11) from RPCI (http://bacpac.med.buffalo.edu) were used for library walking. Approximately 5000 BACs were identified in the initial screening and library walking. This represents BAC coverage of the region in ~40 x redundancy given that the average insert size of BACs is ~130 kb. However, we estimate that nearly 50% of these BACs are false positives resulting from screening errors due to nonspecific hybridization between repetitive elements as suggested by FISH localization of some of the BACs as well as other data (not shown). Newly identified BACs were positioned on the map relative to the initial BACs according to the overlaps determined by using end sequences as well as restriction fingerprint data. Table 3 lists BACs that overlapped with corresponding sequenced BACs based on the sequence matches. Repetitive sequences were suppressed by masking known repeats in BAC end sequences prior to the sequence match using the cross_match program provided by Dr. Phil Green (University of Washington, Seattle, WA); at least 95% matches with >100 bp contiguity were selected. Each of the sequence matches
<table>
<thead>
<tr>
<th>Locus</th>
<th>Marker</th>
<th>Primer sequences</th>
<th>Changes</th>
<th>Corresponding BACs</th>
</tr>
</thead>
<tbody>
<tr>
<td>s3272f</td>
<td>D16S2728</td>
<td>AAAGCTGACATACTGTTAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16ACRCR.0114</td>
<td>D15S540</td>
<td>AATCCAGGAGGCTCACTGAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s297f</td>
<td>D16S759</td>
<td>TCTGAGGAAGGGATGATTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFM172G5</td>
<td>D16S885</td>
<td>AGGTGACCAATACAGAGTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s268f</td>
<td>D16S579</td>
<td>AAAAAAAGAAGGGGGTCCAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29FDH</td>
<td>D16S714</td>
<td>TTCTTCAGCACTGCTTATCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16ACX2</td>
<td>D16S896</td>
<td>GAGTGTTGAGGTTTGTGTTTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s37FD2</td>
<td>D16S863</td>
<td>TTATGGCTATGGGTGACAGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s2951f</td>
<td>D16S866</td>
<td>GATCGAGGAGGATCAGAGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMT070y1</td>
<td>D16S867</td>
<td>AGGTTTCAACTGCTTTATCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s155</td>
<td>D16S868</td>
<td>GAGTGGGAGGCTCACTGCTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>320012c</td>
<td>D16S870</td>
<td>TCACGACCAATACAGACAGCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca41f11</td>
<td>D16S886E</td>
<td>GCTGATTTATCACTGGCAGTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s3450f</td>
<td>D16S897</td>
<td>CCATAGCTGAGGAGGACT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80FD</td>
<td>D16S897</td>
<td>ATGTCCAGCACTGCTGTATCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16ACX81</td>
<td>D16S897</td>
<td>GCCAGCCTGACCACTGCTGTATCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFM11x9a</td>
<td>D16S897</td>
<td>CTTGCTACCTGGACGGACAG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(*) New marker not shown on LANL’s map.
(‡) STS order changed.
was inspected visually, and the overlaps verified by other methods such as restriction fingerprint comparison. Some of the false matches due to repetitive sequences that escaped the masking process were eliminated by restriction fingerprint analysis. Figure 3 represents the final map after gap closure. Although the contig consists of >2000 BACs that were verified and could be placed on the map accurately, most of the redundant clones were not shown in the current map for the sake of clarity and to make map drawing more accurate. All of the supporting data for mapping and clone overlaps, including sequence alignment results and restriction fingerprint gels ideograms, are available from our web site (http://www.tree.caltech.edu).

Figure 1  (A) An example of a digitized restriction fingerprint gel image obtained from a polyacrylamide-based slab gel. Image-2.1 was used for the analysis of the gel. Green lines superimposed on the gel image correspond to the gel bands (bars) detected by the software. (B) FPC-2.5 was used for the analysis and comparison of restriction fragment patterns.
Contig Assembly and Map Drawing

Clones and contigs were placed on the map using the computer software tool AceDraw, which was designed for the organization and management of mapping data and easy map drawing (L. Tang, J. Boulton, B. Liau, H. Zhang, W. Qin, S.H. Huh, X. Xu, Y. Cao, G.A. George, and U.-J. Kim, in prep.; introduction, detailed specification and user manual, and source codes are available from http://www.ugcs.caltech.edu/~genome). Briefly, the program is written C++ for the Unix operating system and allows for freehand drawing of physical contig maps consisting of clones, markers, and other indicators in real scale. The graphic maps thus generated by AceDraw can be dumped into formats that are adequate for porting the map to other databases including AceDB. AceDraw is also able to read AceDB dump files for a graphic display of map data. By using AceDraw, the map (Fig. 3) has been drawn to scale based on the size of the clones, the extent of clone overlaps deduced from sequence matches and fingerprint analysis data, and the order of the markers. Fifty-one BAC sequences were used for sequence matches to align overlapping clones precisely (Fig. 2). The contig consists of 289 BACs with an average insert size of 140 kb that were anchored by 76 BACs embedded in the contig, which have been localized by FISH to relevant loci on the 16p arm. The sequence data from the 51 completely sequenced BACs contain genes and STS markers that have been mapped to this region, confirming the origin of BACs. The order and distribution of STSs in this map is in good agreement with previous YAC–STS maps (e.g., Doggett et al. 1995). Figure 4 summarizes the comparison of the orders and physical spacing of the STSs between the BAC map and the YAC-based map. The overall agreement in the physical organization of the markers suggests that there is no significant gap or internal deletions in clones in either the YACs or the BACs on which the maps were built. The orders of 63 of 67 STS pairs are conserved in both maps. Four minor changes in the local orders of STS pairs may be attributed to the difference in resolution between the two maps.

DISCUSSION

An important problem in genome characterization and sequencing is to provide efficient access to the genomic clones that represent faithful copies of the DNA originated from the region of interest. Identification of a clone or a cluster of clones covering a targeted genomic region is required for physical map development, positional cloning and gene characterization, and large-scale genome sequencing. BACs maintain large genomic DNA inserts with high stability (Kim et al. 1992; Shizuya et al. 1992) and provide reliable templates for accurate genome sequencing. The relatively large insert size makes BACs suitable for large-scale physical map development and sequencing. Deep libraries based on genomic DNA fragments generated by different restriction enzymes and methods are crucial for the development of complete coverage contigs over large genomic regions.

Chromosome 16 was chosen for map development primarily due to the availability of STS markers that were mapped via previous YAC–STS mapping. Mapped STSs are invaluable for accessing the libraries in the beginning. However, the resolution and density of the markers in currently available physical maps are not sufficient for the development of full coverage contig maps. Incremental time-consuming processes such as new marker development and repeated library walking, as well as clone characterization and comparison, are required for contig extension and gap closure. Con-
### Table 2. List of BAC End-Specific OVERGO Primers for Library Walking

<table>
<thead>
<tr>
<th>BAC end</th>
<th>Primers</th>
<th>Corresponding BACs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-2767.1F</td>
<td><strong>AGTATCATGACACATGACTG</strong></td>
<td><strong>A2767.1F</strong>, <strong>A91156</strong>, <strong>S52012</strong></td>
</tr>
<tr>
<td>A-2767.1B</td>
<td><strong>TCTGTTATTTGTGTGATG</strong></td>
<td><strong>A2767.1B</strong>, <strong>S16422</strong>, <strong>S17823</strong>, <strong>S21018</strong>, <strong>S22516</strong></td>
</tr>
<tr>
<td>A-2767.2F</td>
<td><strong>TCTATATATATACATGACTG</strong></td>
<td><strong>A2767.2F</strong>, <strong>S16422</strong>, <strong>S17823</strong>, <strong>S21018</strong>, <strong>S22516</strong></td>
</tr>
<tr>
<td>A-2767.3F</td>
<td><strong>TCTATATATATACATGACTG</strong></td>
<td><strong>A2767.3F</strong>, <strong>S16422</strong>, <strong>S17823</strong>, <strong>S21018</strong>, <strong>S22516</strong></td>
</tr>
<tr>
<td>A-2767.4F</td>
<td><strong>TCTATATATATACATGACTG</strong></td>
<td><strong>A2767.4F</strong>, <strong>S16422</strong>, <strong>S17823</strong>, <strong>S21018</strong>, <strong>S22516</strong></td>
</tr>
<tr>
<td>A-2767.5F</td>
<td><strong>TCTATATATATACATGACTG</strong></td>
<td><strong>A2767.5F</strong>, <strong>S16422</strong>, <strong>S17823</strong>, <strong>S21018</strong>, <strong>S22516</strong></td>
</tr>
<tr>
<td>A-2767.6F</td>
<td><strong>TCTATATATATACATGACTG</strong></td>
<td><strong>A2767.6F</strong>, <strong>S16422</strong>, <strong>S17823</strong>, <strong>S21018</strong>, <strong>S22516</strong></td>
</tr>
<tr>
<td>A-2767.7F</td>
<td><strong>TCTATATATATACATGACTG</strong></td>
<td><strong>A2767.7F</strong>, <strong>S16422</strong>, <strong>S17823</strong>, <strong>S21018</strong>, <strong>S22516</strong></td>
</tr>
<tr>
<td>A-2767.8F</td>
<td><strong>TCTATATATATACATGACTG</strong></td>
<td><strong>A2767.8F</strong>, <strong>S16422</strong>, <strong>S17823</strong>, <strong>S21018</strong>, <strong>S22516</strong></td>
</tr>
</tbody>
</table>

Positive BACs are shown in the contig map (Fig. 3).
tig extension and gap closure would be significantly more time consuming in a region poorly covered by STSs or other markers. In the course of BAC contig construction in our target chromosomal region, we have demonstrated the utility of BAC end sequences as an efficient resource for rapid and precise clone alignment against available sequence contigs such as fully determined BAC sequences. Despite the relatively high density of STSs in the region (1 marker/164 kb of DNA), >24 gaps in the initial map, required repeated screening of libraries to identify additional BACs for the closure. End sequences were determined from all of the BACs identified throughout the project. These and other end sequences from public repositories were used for the determination of the overlaps with the sequences of the “seed” BACs that were being sequenced concurrently in parallel with the map development. In retrospect, a sufficiently deep BAC library with known clone end sequences would have facilitated our map construction dramatically by reducing incremental efforts for repeated library walking and clone characterization. Such end sequence annotated resources are cur-
Figure 3  A complete coverage, sequence-ready BAC contig in the chromosome 16p(13.1–11.2) region. The map is drawn to scale in AceDraw window; each map unit corresponds to 1 kb of DNA. Only the minimal set of confirmed BACs is shown. Fifty-one clones (orange) were sequenced completely; 183 clones (blue) were aligned with the completely sequenced clones by end sequence matches; 55 clones (green) were assembled into the contig based on restriction fingerprint analysis. All of the clones have been sized and fingerprinted. A total of 76 BACs (white diamond) within the clone boxes were FISH mapped and provided anchor points to ensure that BAC clusters are on the correct chromosomal loci.
rrently becoming available (Kelley et al. 1999; http://www.ornl.gov/meetings/bacpac/95bac.html).

Genomes of higher organisms contain myriad repetitive sequences, which differ widely in length and copy number. Previous analyses of chromosome 16 indicated the presence of large duplicated sequence blocks (European Polycyctic Kidney Disease Consortium 1994; Dissing et al. 1998). Recent analysis of DNA sequences from 51 BACs in this contig, which correspond to a total of 7221 kb of genomic sequence, revealed the presence of large, highly conserved sequence blocks in this region (Loftus et al. 1999). These sequences occur in multiple genomic loci and, in some case, can be considerable obstacles to localization and mapping of clones or contigs. FISH data from individual BACs provide an overview of the localization of the clones, as well as the presence of repeat sequences in the clones. Table 4 lists BACs that display positive FISH signals on multiple chromosomal loci. In particular, A-13F4 carries two pairs of large duplicons that appear to occur on both chromosome 16p and 16q arms. A number of STS sequence duplications dispersed throughout the region were also identified from sequence data analysis. These clones were assembled into a current contig on the basis of contextual data such as overlaps with other confirmed clones in the contig. Because of the presence of repeats, BAC end sequence matches often resulted in false alignments. Restriction fingerprint pattern analysis proved critical for the confirmation of true overlaps in many instances.

Table 4. BACs with Positive FISH Signals on Multiple Chromosomal Loci

<table>
<thead>
<tr>
<th>BAC</th>
<th>Band position</th>
<th>Positive STS</th>
<th>Additional FISH signals</th>
<th>Possible reason (sequences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-321F1</td>
<td>16p13.1</td>
<td>D16S2779</td>
<td>1p34</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-29812*</td>
<td>16p13.1</td>
<td>D16S2779</td>
<td>1p34</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-98H8</td>
<td>16p13.1</td>
<td>D16S2714</td>
<td>11q23</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-972D3</td>
<td>16p13.1</td>
<td>D16S2853</td>
<td>random multiple loci</td>
<td></td>
</tr>
<tr>
<td>A-219D1</td>
<td>16p13.1</td>
<td>D16S2586E</td>
<td>16p11.2</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-319E8</td>
<td>16p13.1</td>
<td>D16S501</td>
<td>2q11.2</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-589H1**</td>
<td>16p13.1-1.2</td>
<td>D16S2794</td>
<td>16p11.2</td>
<td></td>
</tr>
<tr>
<td>A-376B3</td>
<td>16p12.3/13.13</td>
<td>D16S2720</td>
<td>16p11.2</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-13F4</td>
<td>16p12-13.2</td>
<td>D16S2696</td>
<td>16p11.2, 16q23</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-685D8*</td>
<td>16p12/13.2</td>
<td>D16S2696/D16S2794</td>
<td>16p11.2, 16q23</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-793D10</td>
<td>16p12/13.1</td>
<td>D16S2696</td>
<td>16p11.2, 16q22</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-613E5</td>
<td>16p12/13.1</td>
<td>D16S410</td>
<td>16p11.2</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-61E3</td>
<td>16p12-13.1</td>
<td>D16S2717</td>
<td>16q22, 16q24, 16p11.2-12</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-101B6</td>
<td>16p11.2</td>
<td>D16S2816</td>
<td>pericentromeric signals on most chromosomes, strongest at 2p11.1 &amp; 16p11.1</td>
<td>repetitive</td>
</tr>
<tr>
<td>A-761H5</td>
<td>16p11.2</td>
<td>D16S2882</td>
<td>16p13.1</td>
<td>duplicated</td>
</tr>
<tr>
<td>A-575C2</td>
<td>16p11.2</td>
<td>D16S298</td>
<td>16p13.1</td>
<td>duplicated</td>
</tr>
<tr>
<td>67N16</td>
<td>16p11.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-280B4</td>
<td>16p11.1-p12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-1811</td>
<td>16p11.1</td>
<td>D16S2816</td>
<td>all centromeres &amp; 2p11.1</td>
<td>repetitive</td>
</tr>
</tbody>
</table>

(*) Not used in contig map.
(**) This clone should be located in c16p11.1.
Currently the contig map is being used to select BACs that cover sequence gaps. These BACs are to be sequenced at the Joint Genome Institute to achieve a 12-Mb contiguity in DNA sequence in this region. Our mapping approach will provide a model system for integrated large-scale genome mapping and sequencing in other human genomic regions and the genomes of other organisms.

**METHODS**

**BAC Library Screening**

Caltech BAC libraries are discussed in our web site (http://www.tree.caltech.edu) and were used for screening by hybridization as described previously (Kim et al. 1995). RPCI 11 human library segments 1 and 2 corresponding to 12 × genome coverage along with high-density filters were purchased from the Caltech website (http://www.tree.caltech.edu) and were used for screening by hybridization as described previously (Kim et al. 1995); RPCI 11 human library D and other human BAC libraries (http://www orcnl.gov/meetings/bacpac95bac.html). These data are available from the BAC end sequence database at TIGR (http://tigr.org/tdb/human/bac_end_search/bac_end_info.html). All currently known human repetitive elements in BAC end sequences were masked using the cross_match program prior to searching for homologies against the individual BAC sequences with a web-based sequence match program available at TIGR (http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html) and GenBank. A minimum of 95% homologies were accepted as sequence matches. Putative overlaps detected by sequence matches were further verified by analyzing restriction fingerprint patterns and STS contents of BACs.

**Contig Assembly and Map Drawing**

Restriction fingerprint data extracted from gels by Image-2.1 were analyzed using contigC and FPC-2.5 developed at the Sanger Centre (Soderlund and Longden 1996; Gregory 1997). The BACs in the initial framework contig clones served as anchors on which new clones were aligned according to the sequence matches and/or fingerprint data. The resulting physical map was drawn with AceDraw (developed at Caltech). The Caltech website also provides experimental data for each of the clones and clone-to-clone relationships.

**ACKNOWLEDGMENTS**

We thank Dr. Phil Green for providing us with sequence analysis softwares. This work has been supported by National Human Genome Research Institute grant HG01464-01 awarded to M.D.A. and U.J.K. N.A.D. was supported by U.S. Department of Energy contract W-7405-ENG-36.

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**


Received April 13, 1999; accepted in revised form May 28, 1999.