A 12-Mb Complete Coverage BAC Contig Map in Human Chromosome 16p13.1–p11.2

Yicheng Cao, Hyung Lyun Kang, Xuequn Xu, Mei Wang, So Hee Dho,[1] Jun Ryul Huh,[1] Byeong-Jae Lee,[1] Francis Kalush,[2] Diana Bocskai, Yan Ding, Judith G. Tesmer,[3] Jonghyeob Lee, Eunpyo Moon,[4] Vesna Jurecic,[5] Antonio Baldini,[5] Heinz-Ulrich Weier,[6] Norman A. Doggett,[3] Melvin I. Simon, Mark D. Adams,[2] and Ung-Jin Kim[7]

Division of Biology, California Institute of Technology, Pasadena, California 91125 USA; [1]Institute for Molecular Biology and Genetics, Seoul National University, Seoul, Korea; [2]The Institute for Genomic Research, Rockville, Maryland 20850 USA; [3]Life Sciences Division and Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA; [4]Department of Biological Sciences, Ajou University, Suwon, Korea; [5]Institute for Molecular Genetics, Baylor College of Medicine, Houston, Texas 77030 USA; [6]Life Science Division, University of California, O.E. Lawrence Berkeley Laboratory, Berkeley, California 94720 USA

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 contigous, overlapping BACs.

A major goal of the Human Genome Project is to provide a complete sequence of the human genome with an accuracy of >99.99% and a high degree of contiguity (Collins et al. 1998). Currently prevailing methods for large-scale genome sequencing include the clone-based approach in which contigs based usually on large insert clones such as BACs are established prior to the initiation of sequencing. The contigs are used for the selection of minimally overlapping clones that are to be sequenced. Alternatively, a set of nonoverlapping or minimally overlapping BACs that have been mapped to target chromosomal loci are selected and shotgun sequenced, leaving the sequence gaps between the clones to be closed by identifying and sequencing additional clones. Restriction fingerprint analysis has been serving an important tool for the detection and quantification of clone overlaps (Coulson et al. 1986; Olson et al. 1986; Sulston et al. 1988, 1989). Recently, a new scheme has been proposed for rapid detection and quantification of BAC overlaps by means of sequence matches using the end sequences generated from a sufficiently large number of BACs that serve as sequence-tagged connectors (STCs) (Venter et al. 1996). In this approach, initiation of sequencing of a large genomic region is not dependent on the completion of a high-quality contig map. Rather, development of physical contig maps and sequencing BAC clones work synergistically, allowing for the early initiation of sequencing on selected BACs. This requires the availability of annotated BAC libraries in which the majority of the clones are tagged with end sequences.

The project was initiated as a part of a publicly funded program to map and sequence large chromosomal regions in human. The centromeric half of human chromosome 16p(13.1–11.2) spans ~20 Mb, includes the 16pCEN as well as the pericentromeric regions, and contains at least 162 expressed sequences (NCBI: GENEMAP98 at http://www.ncbi.nlm.nih.gov/genemap/) that are both biologically and clinically interesting (Mitchison et al. 1993; Stallings et al. 1993; European Polycystic Kindney Disease Consortium 1994; Liu et al. 1996; Dissing et al. 1998). A high-resolution YAC-based STS map is available for chromosome 16 (Doggett et al. 1995). Mapped STS markers

[7]Corresponding author.
E-mail ung@caltech.edu; FAX (626) 796-7066.

9:763–774 ©1999 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/99 $5.00; www.genome.org

facilitated initial access to BAC libaries 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 roadmap 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 × 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 *Not*I 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 *Hin*dIII 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 *Hin*dIII 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 rearangement 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× coverage Caltech libraries and the 12× 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× 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 1. Sixty-Eight (68) STS Markers and Corresponding BAC Clones in the Contig Map

Locus	Marker	Primer sequences		Changes	Corresponding BACs
s327F2	D16S2782	AAGGGTTCTGAGAAATCCT	GGTTTTGGGAGGATGAATTG		A-167B2, A-276F8, A-991G6
16ACCRI.0114	D16S49	AATTCAAGGGAGGCTCAGGTG	CACTCCTCCCTCTATGTTATG	*1	A-777B5
s329F7	D16S2785	TCTCGGAAGGAATCTTTAGC	ACAAACTTGGGTACATTGCC		A-140D10, A-65D3, A-334H1,A-296B11
AFM112xg5	D16S500	AGGTTCCCAGCCTGATTTTT	GGGATGAGCAATACGAAAGA		A-732D3
s326B4	D16S2779	AAAAAGCAGACGGAAATGCC	CAACAGCCAAATCCATAGAC		A-263G1, A-321F1
s29H11	D16S2714	TTTGAATCCACAGCACTGCCAG	GTCCCTCCACTCATCTCTCCTT		A-98H8
16AC2.3	D16S292	GGCATGTCAGGCCAGCCATGTTTT	CTTTGCACAAAAACAGTAGCTATCCAC	*2	A-731F11, A-909D3, A-380G12
s57B2	D16S2853	TTAAGATAGGGTAACTGGAGTC	GAGGAAATCCCATCTAAATCTG	#1	A-972D3, A-1000C7
s35B11	D16S2803	AACAACCTGCCGACTTGGTTTC	TCGTTGAGATGAGGAGACTGTC		A-302H2, A-147A6.A-815A9
AFM070ya1	D16S405	AGTTCTCTGCTGCACCTGGC	TGAAATGGGGACCATGAAGG		A-302H2, A-147A6.A-815A9
s1E5	D16S2696	GAACTGTCCCCTCAACATCC	CGTCAGATATGTGTTTGTCATCC		A-13F4, A-793D10
s302E12	D16S2720	TTCAACCACAAGATGTCCCTGC	CACAGGACCCCTCTTTGCTTAA		A-962B4, A-376B3, A-339G6
Cdalff11	D16S2586E	GTGTTGGTCTCTGTGTTCCTG	GAAAGATGCTCTCTGGGTTTG	*3	A-489D2, A-219D1.A-112B5
s343D3	D16S2799	CCAGTAACCAGAGACATCCACT	GAACTTCACCACAGTCTTACAGG		A-56H6, A-55C6
s80F6	D16S2887	ATGGGTTCAAGGCAAATCCCTC	GAGTCTAACAAGCCCATCAGCA		A-161G4
16ACXE81	D16S287	GCTTGTATTAGTCAGCATTCTCCAG	TACAGACCATAGACTTGACAGTC	*4	A-530H11
AFM113xa9	D16S501	CTGNCCCTACTCTGCCACAC	GGAACCTGCCCATGGAGTGT		A-709E3, A-481E9, A-319E8, A-989F12, A-249G10
AFM165yb6	D16S410	ACGCAGTACGTTTTCGTTGT	ATTGTAAACATTTCTGCCCA		A-613E6,
WI-612	D16S737	AGAGAAGGCCTGAGGAGGTA	GGGATCTGTTCAGATGGTTG		A-334D11, A-496E6
s5F1	D16S2859	GAGATGTCAGTTGTCACTAC	GAGAAGATTCCACAGATACC		A-418G10, A-917C12
s55B4	D16S2847	ATCTGACTTCCCTGCTGATTTG	CTACACATAAAAAACCAACCCC		A-621A12
s41A7	D16S2816	GGTGAAAGACCAGTTGAAGAGC	CACTACTCTGTCCATCCCTCAT		A-1B11, A-101B6
Cdazwb05	D16S2597E	ATGAGGGTGGGTCCTTCTG	TGAAGCGGTCAGTCCTCC		A-69G12
Cdazpg11	D16S2594E	GAAGTATCCCAGCAGAGTCC	TTTGAAAGCAGGTTGCAGG		A-101F10
s308B12	D16S2745	AAGGGAAGGAGAGGGAGAAAAG	CATGATTAGGCAGTAACAAGAG	#2	A-396A4
AFM164th2	D16S3041	AGTCCCTGCCTTGGTTAGTT	CAGCCTAGGTGACAGAGAAA	*5	A-363E6
s40A7	D16S2813	ATTCTGGCAGCAACCTCACTGA	CAGTTTCTGTGGATCAGGAGTGC		A-1000D7
s76AF4	D16S2881	CATGAATACTGTTCCCCTCTGC	CTGGACTACAACATAAGCATCA		A-99D8
s25H11	D16S2707	GTTCAGGTGGTGTCAAAAGG	AGATGGCTAAGGGTACAGAG		A-256A9
s315D8	D16S2767	CTCATTTGGGTATAGTTGCCTGCC	ATGCCTTGGAGAGTCAGAAGTCTC		A-923A4
s35C8	D16S2805	GAGTGCTGTCAGTTGATAGGTC	CTTCGCCAAGTAATCATTCAGC		2149J04, 2206D15, dJ-12L6, dJ-12N6
s33F5	D16S2797	TGAGGACACAGCCATAGGATGC	CTGGCAGCAGAGACATCCAAGA		A-222E3, A-270G1
s301B6	D16S2717	GTGCAATAGTGCAATCTTGG	GAACTGACTTTGAAGGAAGG	#3	A-232D1, A-61E3, A-100B4
WI-1722	D16S734	GGATCCATGCAAACCGAG	CCTAAAACCAGCTGCTCTGG		A-182B7
CdaOpg10	D16S2569E	AAGTATGCAGACTTGACC	AACAATCACTGTTGCTCC		A-399C3
s21B7	D16S2697	AAAGGACTTGGTGACTCGTTCC	GGAAGGGATGAGGGAGCATAAG		A-604G3, A-399C3
s325G11	D16S2778	TGTGGATTCCAGAGGCATCT	ATTGACCAGACTGGTGAGCA		2099F21, 2209L20, 2288E16
s72E12	D16S2876	GATAAAAACCACTTTGGGCCAG	CCACGCCCATCTAATTTCTTAA		A-737B10, A-675B11, A-280A4, A-237H1
s23H6	D16S2704	TCCACAACTGAGTTTGTAAGAG	CTGGACAATTTTCCCCTTCTTT		A-586A7, A-237H1
AFM049xd2	D16S403	GTTTTCTCCCTGGGACATTT	TATTCATTTGTGTGGGCATG		A-237H1, A-413A6, A-570B5
s47F6	D16S2823	AGATAAACATTTTGGAGGCGTC	GAATGTCGTCTCTCTTCCCCTT		A-413A8, A-352H11, A-391G2
AFM191wb10	D16S412	AGCCTGGTTGTTAGAGTGAG	AACAGGTTGATTATTTGGCA		A-413A8, A-352H11, A-391G2
s311B2	D16S2760	TTTGCTCTGGCAGGCTGTGAAG	GGCTTAGAGAGAACCTTGGGCA		A-2760
s33H1	D16S2798	AGGATCTGTGCCTCTGTTGTAC	CCAGACTAGGCTTAACACACAG		A-228B3, A-735G6
WI-12693		TGCACCGTCTTGTGCCTAGT	ATTAGCATCACCATTTTGCAG	*6	A-388D4
AFM220xb10	D16S417	CTGTCCAACATGCAGCC	TGAAGTCAATCCCACTTGAA		A-56B4, A-188E9
s307D4	D16S2744	TGAGTAGTTGTGACAGAGAC	TTAACCTTACCAAGAGCCAG		A-345G4
AFM238xb2	D16S420	ATTTCCTGAGGTCTAAAGCACCC	TTAGGCCCAGTCCACACTCAAG		A-113A6, A-617F1, A-253B10
CdaOnf07	D16S2567E	TGGTGATGATGAGGAAAG	GAGGTGGAGAAGAAATTAAC		A-113A6, A-617F1, A-60H8, A-253B10, A-268E9
AFM025tg9	D16S401	TTCTCTTACAACACTGCCC	ATTTGGATGGCTTGACAGAG		A-372D8, A-698D4
s16A7	D16S2691	CTCTTCACTTATCCTCAACTGC	CATCCCTCCCCATAATCTTTTT		A-249B10, A-41F5, A-67E3, A-557D9
s330F3	D16S2792	CAGCGGTTGTATCTCATTACAG	GGCAAAAGATTTGGCAGATAGTTC		A-88D1
s30A1	D16S2753	GAGCAGTTACTCCATGTCATGC	CCCCTGAAACCATTCTACCCTC		A-6A8, A-218C7
s58F9	D16S2855	TGAGAGAAGCATAGGCACTTTC	GGTCTCATAACCAGGACTCACA		dJ-1061I2, dJ-341O14, dJ-390C18, dJ-541J10
s15H1	D16S2690	GCTGTTTCTTCTGAGGAGTC	CACCATATACATATGTGTGTGTGTGTGC		A-379D8
s67A3	D16S2865	CAGGCTAAGCTCAATTTTGG	GTGTAGGAGTTTTTCTGTGG	#4	A-485G10, A-622B7, A-385H9
s48D6	D16S2868	TTTTCCGATGGGACTATGAGCC	GCTGAGAGTGAGAGTTGCGTAA		2009D16, 2061P14, dJ-340K16
CdaOhc12	D16S2565E	GTGAACACAGCAGGTGTGAG	GAATGGTTCATGGAGTCACAAG		A-481B3, A-153E9
s15C2	D16S2689	CACATTGTCCTTCCTATGCTGC	CAAGTGTCATACCCAGTCACCT		A-670B5
s305D5	D16S2734	CTAAGTACAAATGCTGACAGGC	CAATTGTCCCACCACCCGAGT		A-951C11
16-129N.1/	D16S272	GTGAAGGCTAATAGGCTCTTG	CTCAAAATTCTACTCCATCACC		A-306C7, A-233A7, A-686C6
302G12RF/3	D16S726	CATCTTCGGATAACTGCAGTGTTCAG	CCCAAGTCCCTCACCTGGGCAGT		A-233A7
233E10LA/2	D16S721	CAAGTGTGAGTCATTCACTGCA	AGAGTCTGGCTCTCACTGTGG		A-154B9
16AC6.17	D16S299	TCCAACTGCTGGGATTACAGGCACA	ACAGAGTGAGGACCCCATCTCTATC		A-154B9, A-331G1
85D3-F/85D	D16S714	GGCAATATCAGGCATCCTCCAGGCTAC	ATGGACAGGGTACCAGCACCCTCAAGA		A-154B9, A-331G1, A-373A2
s319C2	D16S2774	TGTTGGAGTCTGGCTCTAAAAC	CTGTTCCACTTTAGAGAGGAGA		A-255G8
s77B10	D16S2882	AGCATTTGAAGGTGAGAAGGTG	CTTGCTGTATCCCCACTTGGTG		A-255G8, A-761H5
16AC3.12	D16S298	TTCCTCATGTATAAATTGGGTGTGGCCA	TGGTCCCGGCCCAATCCCAATGCTT		A-575C2, A-305A8

(*) New marker not shown on LANL's map.
(#) STS order changed.

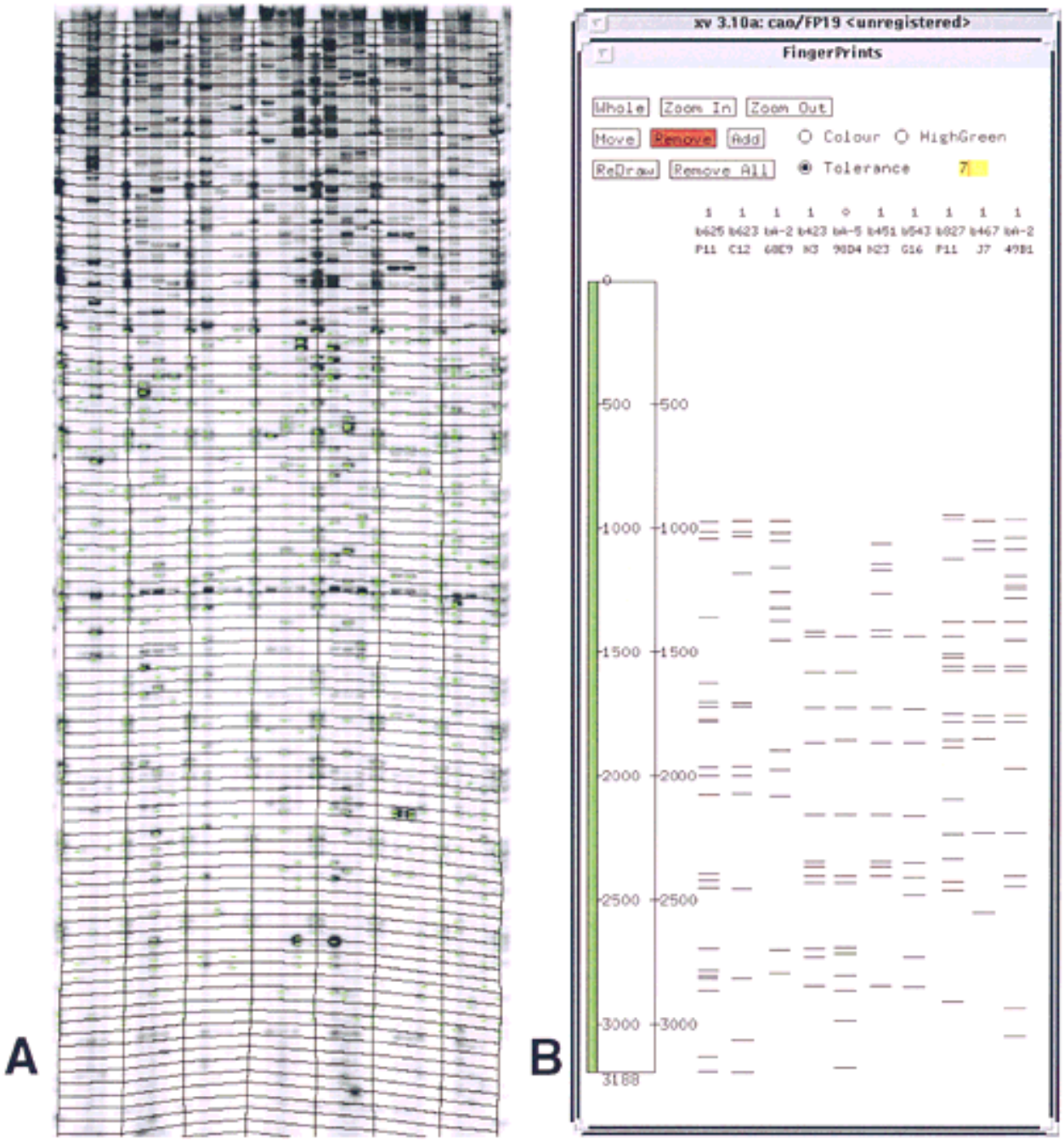


Figure 1 (*A*) An example of a digitized restriction fingerprint gel image obtained from a polyacrylamide-based slab gel. Image-2.1 was used for the analysis of the gel. Green lines superimposed on the gel image correspond to the gel bands (bars) detected by the software. (*B*) FPC-2.5 was used for the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

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).

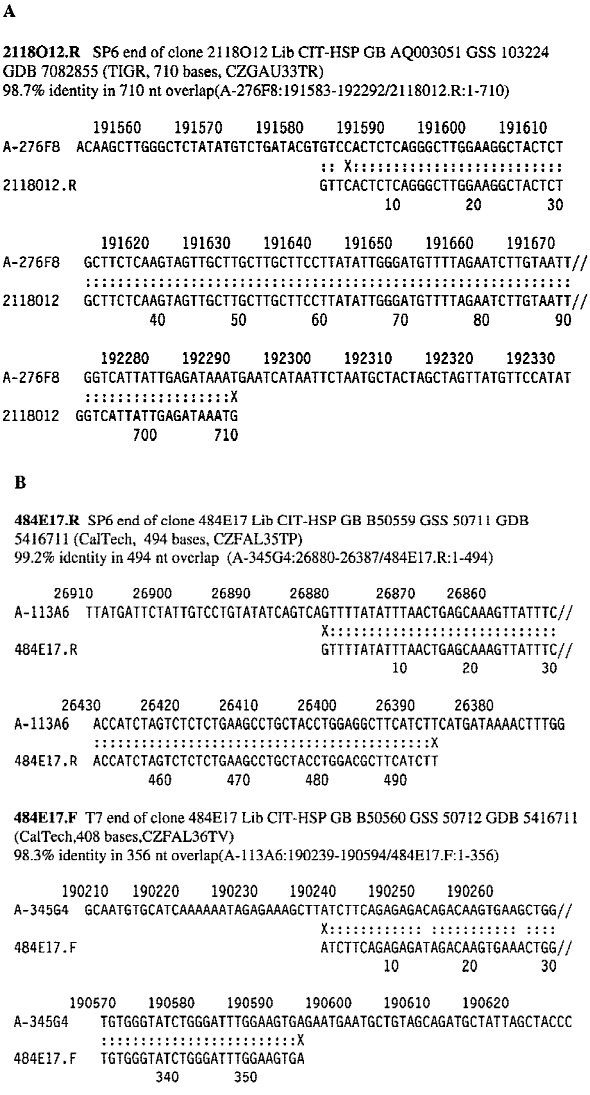


Figure 2 Examples of end sequence matches with completely sequenced BACs.

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 Ace Draw, 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

BAC end	Primers		Corresponding BACs
A-276F8.TP	GCTACATCCAGATCCTCACAC	ATGGGATGATGGTGTGTGAGG	A-276F8, A-991G6, 525D12
A-276F8.TV	GCACCTTTTCTGTTCCAAGCCC	TAGCATCTAGTACAGGGCTTGG	A-276F8, 516A2, 517B3, 67N16, 2118O12
A-731F11.TV	TGAAGAACCCTATCGTGCAGG	ACACGTTTCTTCACCTGCACG	18C6, 27D1O, 516A2, 531E24, 67N16
A-731F11.TP	GCTCTGAGACAGCTTGTCAGAG	TATAAACCATGGTGCTCTGACA	131J18
A-972D3.TV	GGCCCTGGACTCACACAGACCT	CTGGGATTCAGACAAGGTCTGT	A-1OOOC7, 185I23, 193M12, 545O9
A-972D3.TP	CAGGTAGAGACTAGCTATGACC	TAGGAAGTGAGTGGTCATAG	18C6, 516A2, 517B3, 51P6, 531E24
193M12.TVB	GCAGCACATGTGACATGACACC	CGCACACTGTTGGGTGTCAT	A-1OOOC7, A-352E1, A-376B3, A-38OG12, A-489D2, 185I23, 18C6, 193M12, 2J14, 36E8,517B3, 531E24
A-13F4.TP	GCAGAGGGATGTCCAGGGTCAC	GAAATGTGATTTGTGTGACCCT	18C6, 516A2, 517B3, 531E24, 67N16
A-13F4.TV	GTCCTGAAAGTATGTTAAATAG	GTGTCACTGTGACTCTATTTAA	2115J2O, 2196E14, 516A2
118F2.TP	CGTGGGCACCTTCACCTGGGTC	AGAAAGGGCGTGCAGACCCAGG	A-249G1O, A-345G4, A-78A12, 502C3
118F2.TV	GCCTTTTTAAGAACCAACTTTC	GGCTTCAATGAGAAGAAAGTTG	A-1OOOD7, A-793D1O, 18C6, 516A2, 67N16, 6C18
A-56H4.TPB	CTACCCCAAGACAGTATCTTCT	AAACAACATGGAGAAGAAGATA	2196E14, A-1OOOD7, 18C6, 27D1O, 44A1, 516A2, 67N16, 2196E14
A-56H4.TV	GCAGAGAGGTTGAAACTTGGAA	TGACCATTCCCTTCTTCCAAGT	A-1OOOD7, A-345G4,131J18
A-161G4.TP	GCCTGTGGAAGGGGAGCCTACA	GGTAATTGATGTGCTGTAGGCT	A-972D3, 193M12, 319L12, 545O9, 2106I1
A-161G4.TV	GCTACTGCAGCAGGGACTGATG	GACTGATGCAATCTAATCTGCA	A-972D3, 18C6, 44A1, 516A2, 51P6, 2106I1
A-53OH11.TPC	CATCTGTTAAGAACATGGGTCT	ATGGGTCTTGTGAATACCGAAT	A-249G1O, A-352E1, A-376B3, A-38OG12, 18C6, 193M12, 545O9
A-53OH11.TV	CAGCCATGCAGACCGGGCTGTG	CATCTTCACATCATCACAGCCC	A-793D1O, 18C6, 44A1, 516A2, 67N16, 2O65O23
A-249G1O.TPE	GAGTCCTGCCTGTGTGCTGGAG	TCAGCACCCGGCATCTCCAGCA	A-1OOOD7, A-112B5, A-345G4, 319L12, 516A2
A-249G1O.TVC	CAGGGTTCTCTGGTCTTCTTTG	CAAGCTCCAGTACTCAAAGAAG	A-249G1O, A-496E6, A-991G6, 111N12-1, 118F2, 219B15, 5O2C1O, 6C18, 2106I1
A-613E5.TV	CACCTCAGTTTTCTAAAAGTTA	GCAAAGATACATGTTAACTTTT	A-793D1O, 18C6, 219B15, 516A2, 517B3, 531E24, 67N16, 6C18, 728M17
A-613E5.TPC	CTAGGGCTTTGACTCTTCTTCT	CAATCTAGGTGGTGAGAAGAAG	A-617F1, 18C6, 516A2, 67N16
A-334D11.TP	ACCTGAGAGATCACTGCTTGCC	GCCAGTACAGGGCTGGCAAGCA	2099P24, A-345G4, 111N12-1, 118F2, 219B15, 496I23, 5O2C1O, 728M17
A-334D11.TV	GAAACTTAGGAGAAAATCTCTG	AATCTCTGTTACCTTCAAGGGG	2099B22, 111N12-2, 2099B22
A-418G1O.TV	CGCCCGTTGGAAAGTTTTATTA	CTCAAGCATTTGGTTAATAAAA	188D2O, 18C6, 459P15, 516A2, 517B3, 531E24
A-418G1O.TP	CCACCCACCGCCTTCTTTCAAA	CAAATGCAATTTCATTTGAAAG	A-249G1O, A-917C12, 18C6, 26OA22, 33D12, 517B3, 519A13, 531E24, 654M22, 2255G9
A-1O1B6.TV	CATTCCTGAGGGCACCGAGGAG	TAAAGCTATCCTCCCTCCTCGG	A-793D1O, 18C6, 516A2, 517B3, 531E24, 67N16, 2065O23
A-1O1F1O.TVB	CTGGTCTTCCTTGGAGTTTCAT	GTCCTGCCAGTACCATGAAACT	A-112B5, 516A2
A-1O1F1O.TPC	GGATGGCAGCTGCTCATGGAAC	CCCTTTTTCTGTGGGTTCCATG	
254P9.TVB	CGATATTAAACCCATCGAGTTT	TCGAGTTTGGCGTTCTCAGCGC	2025C8
A-1OOOD7.TP	CTCTCCTCCCATTGCCCTTCCC	CATGTGGTTAGTTAGGGAAGGG	2099B22, 2264D9, 516A2
A-1OOOD7.TV	CGATATCAATGAAGATGTAGCC	GGCCTGGCCACCCAGGCTACAT	728M17, 2255G9
A-256A9.TPB	CATAATCCCACTTTCCCATTCA	CTGGAGTTTGTTGTTGAATGGG	A-1OOOC7, 36OO2
A-256A9.TVC	CATGTCTGGACCAGTTCCCTTT	GATGTTGGGGGTTCAAAGGGAA	188D2O, 19OE15, 313)3, 447N19,516A2,517B3, 51P6
A-589H1.TP	CTTGTCTGTGATGTTCCTGCAA	AAATCTAGAACTAGTTGCAGGA	A-1OOOC7, 192F9, 517B3
A-589H1.TV	CTTCCTGGAGTGATCCAAGCCC	GTCAGGAAGTGGCAGGGCTTGG	A-69OA11, 516A2, 517B3
A-923A4.TV	GCCCCGACCCTCCTTGGGACTC	TTGGGGACTCCGCAGAGTCCCA	516A2
A-334D11.TP	CCTGAGAGATCACTGCTTGCCA	TGCCAGTACAGGGCTGGCAAGC	2099P24, A-345G4, 111N12-1, 118F2, 219B15, 5O2C1O, 728M17
A-334H1.TV	CGGTACACAATGACAGTATTTA	GATTTTAGAAATAGTAAATACT	A-1OOOD7, 728M17
A-61E3.TV	GATGCCATTCTCATGTGTAATC	TCCATTTGCCACCTGATTACAC	A-1OOOD7, 44A1, 729D21
A-27OG1.TP	GGGAGCGGGTGGGCGCCCACCG	ATAGGGAGGCCCAGCGGTGGGC	A-1OOOD7, 111N12-1, 728M17
591M7.TP	GTCTGAGTTGCACCAACCTCTG	TAGTCTCTGGACCACAGAGGTT	A-1OOOC7, 111N12-1, 193M12, 36OO2
A-735G6.TPB	GGAGTTTCCAAAGATAAGTTAG	TGCGTACCTAACTCCTAACTTA	2196E14, 177P9, 185M4, 188D2O, 19OE15, 44A1
A-735G6.TV	GGCTGGCTGTTGGCCGATTCAT	CCATGTCAGTCATGATGAATCG	A-1OOOC7, 192F9, 193M12
A-56B4.TP	CTCCATATTGGGGTCCTGCTGT	AGCTAAACTTTTTCACAGCAGG	2196E14 , 44A1, 729D21
A-56B4.TV	CCAAATGTCTGGCTGACTCTTG	ATGCATGCATGGTTCAAGAGTC	A-1OOOD7, A-345G4
A-345G4.TPB	GGTTGCAGTGATTGCTGGGCTT	ACTCCTCCCTAATGAAGCCCAG	A-276F8, A-345G4, 111N12-1, 502C3
A-345G4.TV	GCCAGTGAGACACAAAGGAAGC	GGTCCTGTCTCCCTGCTTCCTT	A-1OOOD7, A-276F8, 529P19, 531E24, 728M17
A-113A6.TV	GCTGCAAGCTGCCCAGAAATCT	CCACGTAAAAGGTAAGATTTCT	A-1OOOD7 a1, 529P19, 531E24
A-113A6.TP	GCTGTTGAGCACCTCCTGCCTC	CAGCATGAAGCATGGAGGCAGG	A-1OOOD7, 111N12-1, 728M17
A-598D4.TPB	CACCAATCCTCATGTCAACTCA	TTCCATGTTAATACTGAGTTGA	A-256A9, A-972D3, 111N12-1, 193M12, 451N23, 467J7, 543G16
A-598D4.TV	AAGTCTAGACCCTAAACCTTCA	GTTGGTATTATGAATGAAGGTT	A-256A9, A-598D4, A-982D3, 31303, 423N3, 44A1, 51P6, 527A19,531E24
A-249B1O.TP	GAGCCACTGGGGCTCGCGTTCC	CGCGTTCCCAGGAATGCCTACC	A-256A9, A-735G6, A-917C12, 193M12, 36E8,7O6K3, 423N3, 519A13, 527A19, 531E24, 543G16, 654M22
A-249B1O.TV	GCCCACTTTTGGCAAAGATTTG	AAGAGCAATTTTGGCAAATCTT	423N3, 44A1, 527A19, 531E24, 543G16
A-88D1.TV	CTGTGCTCAGCACTGTGCGAGT	TTCCTAGGCTGAAAACTCGCAC	A-1OOOD7,A-345G4, 527A19, 2233I12
A-88D1.TP	CGGGTCATGGGATGAGAATATA	CCTGGCCTGGAATCTATATTCT	A-575C2,
A-218C7.TP	GCGTGAACAATGTGGCTTCTAG	TTCCTTTTAGTTTCCTAGAAGC	22N6, 67N16, 723G3, 728M17
A-218C7.TV	ACCCCAAGTGAAGCTGAAGACA	CATTTTAACTACATTGTCTTCA	A-575C2, 22N6, 67N19, 723G3
A-485G1O.TP	ATGCTTTATCCTTATTCCACAG	TTCCACAGCCTCCCCTGTTCTC	728M17
A-485G1O.TV	CCTGATTTTCCAGACTACTATT	AAGGGAACATATATAATAGTAG	319L12
A-67OB5.TV	GGAGCGACACACTCAGGCTGCT	CCAGGCAGGACCACAGCAGCCT	
A-951C11.TP	CGATTTCAGAGCTGGCAGGTGC	ATTAAATATCTCCTGCACCTGC	A-218C7, 2O49O4
A-233A7.TPB	GAGTCTCGCCTCCTAAGAGACA	CTTCAGTTCTCTCCTGTCTCTT	A-233A7, 67N16, 723G3
A-233A7.TV	CTGGTGGCTGCCAGAATTCCCG	GTTTCTTGGTGCCCCGGGAATT	319L12
A-761H5.TP	GCATTGGGGATAGAACCTGGAT	ACATATTCAGCCTTATCCAGGT	111N12-1
A-761H5.TV	GTGGATTGAGGGTCTGCCAATG	GTACGGGTCGCTTTCATTGGCA	2115J2O, 2252H1, 2265F17
A-575C2.TV	CCAGGACGAAAGTCAGCCTCGC	AATGGGTTCCAAATGCGAGGCT	2115J2O, 2252H1, 2265F17
A-575C2.TP	GCACACTTCAAACCTCCTCTCT	CAAGGATATTGAGGAGAGAGGA	
136B12.TV	GCTCCTCAAGGACAAGACCTAA	GAACTGGGCAAGACTTAGGTCT	111N12-1, 193M12
A-685D8.TPB	CGCCAGTGAGGAGGATGTGTTG	GGTGCACTTCCCAACAACACAT	
A-685D8.TV	CCTATATGACAGTTACTAGCTA	CCTGGCTTTCTCTATAGCTAGT	A-218C7, A-38OG12, 193M12
A-635H12.TR	CACAGGAGATGGCTGACTTAGC	TCTGAGAATTAATGGCTAAGTC	A-255G8
A-635H12.TVB	CTAGGCTCTCATGGTGCAGAGG	CATCTTCTTGGAAACCTCTGCA	
A-211C6.TP	CCAAGAGGGATCCACAAAGGTG	CTCATGTGACCCTCCACCTTTG	A-249G1O,A-276F8, A-345G4, 111N12-1
A-211C6.TV	CTTCACACTGTGTCCCTTCTAG	CATGACATGCGGCACTAGAAGG	A-276F8, A-334H1
44M2.TV	GCTGAATGCATGCTCCTCATAT	GAGTGTACTGGCTTATATGAGG	185B4
A-18OG2.TP	CTCGGCACACCACCACCGCCCT	GGTAGACGTCGAAGAGGGCGGT	A-345G4, 111N12-1, 185B4
A-18OG2.TV	GCGCCTCACCTGCATCTCCCAG	CAGCTGGCTGCGGGCTGGGAGA	A-522B7, 111N12-1, 36OO2, 543G16

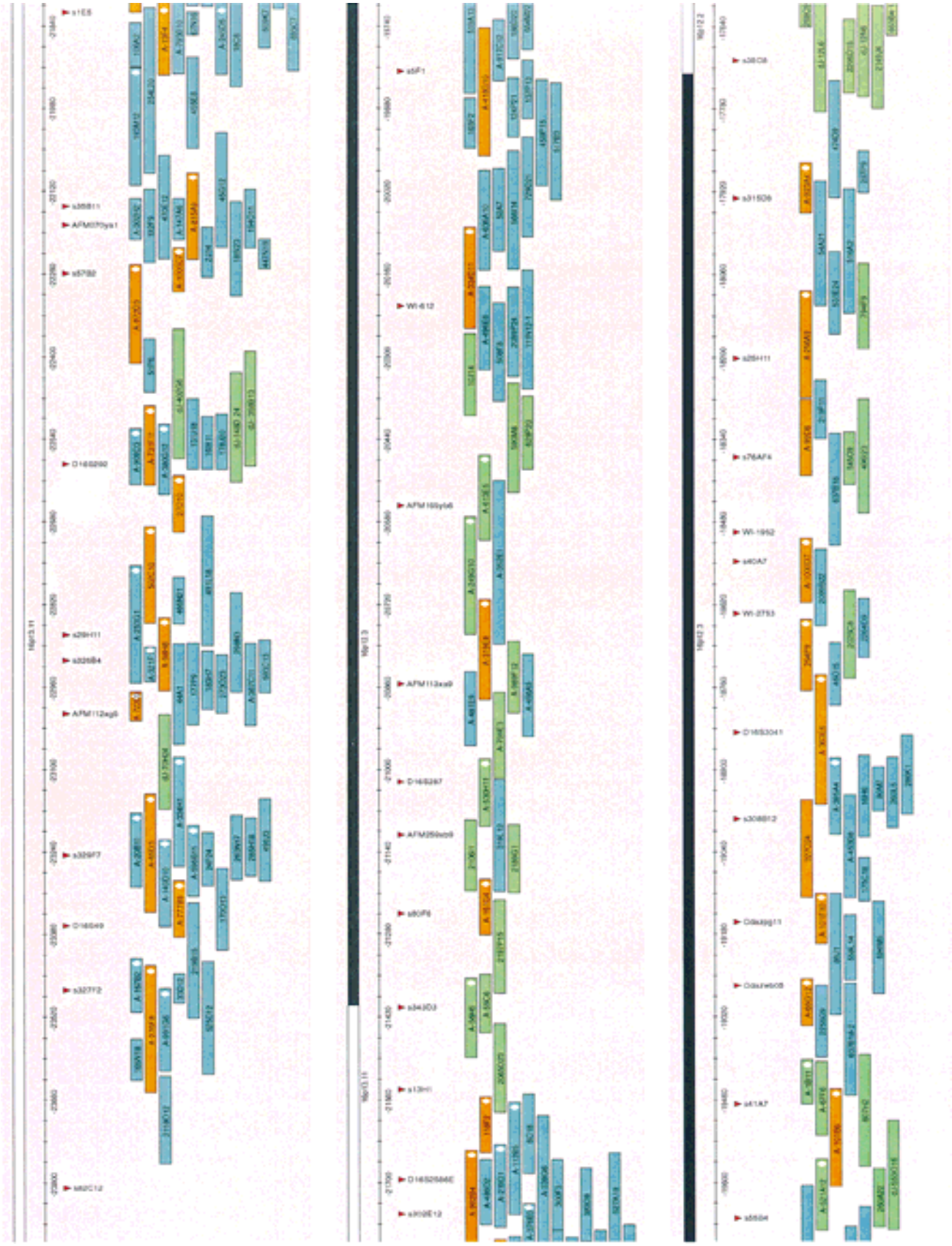
Positive BACs are shown in the contig map (Fig. 3).

Table 3. List of BAC Ends Determined to Overlap with Completely Sequenced BAC by Sequence Match

Complete BAC sequence	Overlapping BAC end (overlap by bp)			Complete BAC sequence	Overlapping BAC end (overlap by bp)		
A-276F8	A-167B2.TV(82239)	219B15.TP(90678)	A-991G6.TV(180435)	A-363E6	2368D22.TF*(18671)	A-453D8.TPC(14824)	45A11.TV(150881)
	525D12.TV(185057)	165N18.TP(196256)	A-167B2.TP(210391)		86K1.TV(118194)	16H6.TV(81252)	2013G22.TR*(81252)
	165N18.TV(199551)	33D12.TPD(199535)			260L5.TV(79046	A-395A4.TV(77676))	80M7.TV(77542)
A-777B5	A-140D10.TP*(73180)	A-65D3.TV(53335)	2019H23.TF*(44860)	254P9	46015.TV*(76799)	2368D22.TF* (121046)	2368D22.TR*(130390)
	A-596B11.TP(38521)	499J3.TV(8980)	A-20B11.TP(8974)	A-1000D7	637B18.TV(12913)	2099B22.TF(106266)	
	24F24.TP(8927)			A-256A9	516A2.TV(35337)	531E24.TP(51916)	2131M2.TR*(16153)
A-65D3	A-777B5.TP(53343)	170013.TV(74897)	A-140D10.TV*(125609)		269P11.TV(70058)		
	285H18.TV(139748)	24F24.TVB(140156)	A-596B11.TV(146594)	A-923A4	297P9.TVB(52765)	474D9.TV(60673)	54A21.TVB(53260)
	263N17.TP(160108)	A-20B11.TV(167702)	A-65D3.TP(92700)	A-270G1	A-61E3.TV(8077)	529P19.TPE(25523)	22N6.TVB(33200)
	A-65D3.TV(192723)	2019H23.TF*(184248)	A-596B11.TP(177909)		54A23.TVB*(33200)	723G3.TV(159690)	25009.TP(202000)
	499J3.TV(148369)	A-20B11.TP(148333)	24F24.TP(148316)		A-222E3.TP(193902)	400K4.TV(105794)	25009.TV(105788)
	263N17.TV(146834)	285H18.TP(139370)	A-334H1.TP(122069)		2265F17.TF(993251)		
A-732D3	A-98H8.TV(186)	298N3.TP(2565)	593C13.TP(3359)	A-61E3	A-232D1.TV(33195	483F10.TVB*(67494)	87P21.TV(70505)
	183H7.TV(31438)	273023.TP(37969)			436J15.TV*(85508)		
A-98H8	502C10.TP(10978)	491L18.TPB(44964)	466N21.TPB(47338)	591M7	36002.TVC(90046)	A-675B11.TV(83748)	A-237H1.TP(10509)
	44A1.TV(80488)	A-321F1.TV(105414)	A-253G1.TP(117319)	A-237H1	A-413A8.TV(16059)	A-570B5.TP(46524)	A-838E11.TV(79867)
	183H7.TP(84689)	273023.TV(82758)	593C13.TV(82726)		690J6.TP(89969)	161J21.TPB(153286)	79L4.TP(188597)
	177P9.TV(82129)	A-367C11.TP(82090)			79L4.TV*(76539)	A-280A4.TP(35271)	A-675B11.TP(31856)
502C10	A-98H8.TP(16831)	298N3.TV(49760)	466N21.TV(70452)	A-142A6	548N1.TVE(5907)	589D12.TV(16220)	A-391G2.TV(32721)
	A-253G1.TV(119280)	491L18.TV*(126432)	27D10.TP(11892)		642D14.TV(32738)	44G5.TV(102638)	A-352H11.TPC(112243)
A-731F11	131J18.TP(108720)	178J20.TV(108733)	169I11.TV(108733)		161J21.TVE(16484)	A-570B5.TV(96027)	A-838E11.TP(24174)
	A-380G12.TP(104215)	A-909D3.TV(96750)		A-279B10	44G5.TP(7922)	A-413A8.TP(11152)	A-352H11.TV(11101)
A-972D3	51P6.TV(42229)	2J14.TV(21472)	185I23.TP(53488)		717B8.TV(14587)	A-391G2.TP(14604)	283J18.TF(29204)
	A-1000C7.TP(46472)				185M4.TV(134733)	548N1.TP(354)	
A-815A9	A-1000C7.TV(16955)	447N19.TR(33853)	2J14.TPB(79701)	A-735G6	2196E14.TR(46227)	A-228B3.TP(71541)	254G22.TV(86713)
	185I23.TV(96032)	A-302H2.TP(118921)	194011.TV(118934)		31303.TP(124622)	190E15.TV(132314)	244H11.TPB(137668)
	192F9.TP(118868)	470E12.TR(145787)	192F9.TV(132489)		A-36E8.TV*(1113)		
	45G12.TP(125127)	A-147A6.TP(112907)	A-302H2.TV(112885)	A-345G4	A-604G3.TV(27783)	484E17.TV(79489)	A-124H10.TV(110377)
	194D11.TP(111878)	193M12.TVB(21749)			A-56B4.TV(110371)	185M4.TP(147288)	A-188E9.TV(180369)
A-13F4	A-245D5.TP(125569)	106A2.TV(114270)	18C6.TV(127434)		484E17.TV(79489)		
	509K7.TV(89897)	711J22.F(47334)	32G16.R*(127428)	A-113A6	A-253B10.TV(48354)	A-60H8.TP(102379)	2281P23.TF*(119500)
	106A2.R(37554)	455E8.TV(32143)			A-600H1.TV(151698)	484E17.TP(20113)	242E2.TP(170662)
A-962B4	118F2.TV(4124)	A-112B5.TP(109570)	A-489D2.TP(126233)		A-113A6.TV(170655)	2281P23.TRB*(167661)	A-600H1.TP(166182)
	A-219D1.TP(135464)	300F3.TV(136599)	A-57C4.TP*(152761)		A-78A12.TP(166145)	A-617F1.TP(166141)	A-60H8.TV(151702)
	A-376B3.TV(152772)	A-339G6.TV(152761)	383D8.TV(164783)		728M17.TV(91860)		
	527A19.TV(172823)	254L20.TP(8112)	383D8.TP(122221	625P11	A-113A6.TV(5350)	242E2.TP(5400)	A-253B10.TP(105763)
	A-376B3.TP(82127)	A-693C7.TV*(46744)	A-245D5.TV(36387)		623C12.TV(176119)		
	A-509K7.TP*(36369)	67N16.TV(20194)	A-793D10.TV(20194)	A-598D4	451N23.TP(59595)	A-372D8.TP(78585)	A-598D4.TP(103881)
	711J22.TP(13626)				A-598D4.TV(103920)	543G16.TP(101292)	A-372D8.TV(80154)
118F2	A-219D1.TV(13126)			A-249B10	706K3.TP(213596)	293M8.TP(65114)	706K3.TV(187344)
A-161G4	319L12.TVB(21468)				A-41F5.TV(103520)	A-67E3.TV(102391)	A-557D9.TP(102353)
A-319E8	A-481E9.TP(46023)	A-456A5.TV(77241)			476J14.TP*(88046)		
A-334D11	2099P24.TR(70908)	A-496E6.TP(71680)	111N12.TV(98776)	A-218C7	A-6A8.TP(64834)	135E10.TP(76582)	2115J20.TF(17518)
	A-636A10.TV(75106)	52A7.TV(42195)	729D21.TV(18421)		185B4.TP(47269)		
A-418G10	654M22.TP(38187)	188D20.TV(53709)	A-917C12.TV(91892)	A-485G10	A-379D8.TPB(45788)	A-522B7.TV(134434)	A-385H9.TV(113872)
	519A13.TV(113113)	260A22.TV(121771)	124P21.TP(160130)		175I4.TP(88611)		
	165F2.TP(213169)	137P13.TP(162240)	124P21.TV(132553)	582J2	A-670B5.TV(36486)		
	459P15.TP(129847)	517B3.TV(125830)	165F2.TV(98721)	A-670B5	582J2.TP(35417)	532B7.TP(39166)	365D10.TP(86916)
A-69G12	558L14.TP(3953)	637B18.TP(73944)	2255G9.TR(77547)		502C3.TP(61885)	12021.TP*(28586)	
A-101F10	2055J8.TF*(4771)			A-951C11	502C3.TV(21744)	204904.TR(31313)	209J2.TV(66259)
327024	80M7.TP(16852)	286K1.TP(24337)	2013G22.TFD*(33103)	A-233A7	A-306C7.TV(131271)	A-306C7.TP(145064)	
	A-395A4.TP(63085)	16H6.TP(76517)	A-453D8.TVB(165438)	A-761H5	A-331G1.TPB(47420)	A-154B9.TV(47441)	A-255G8.TP(66546)
	2055J8.TR*(140227)	175C18.TV(66879)			A-373A2.TV(113930)		
				A-575C2	A-761H5.TP(42743)		

(*) Not shown on map.

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 througout 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-



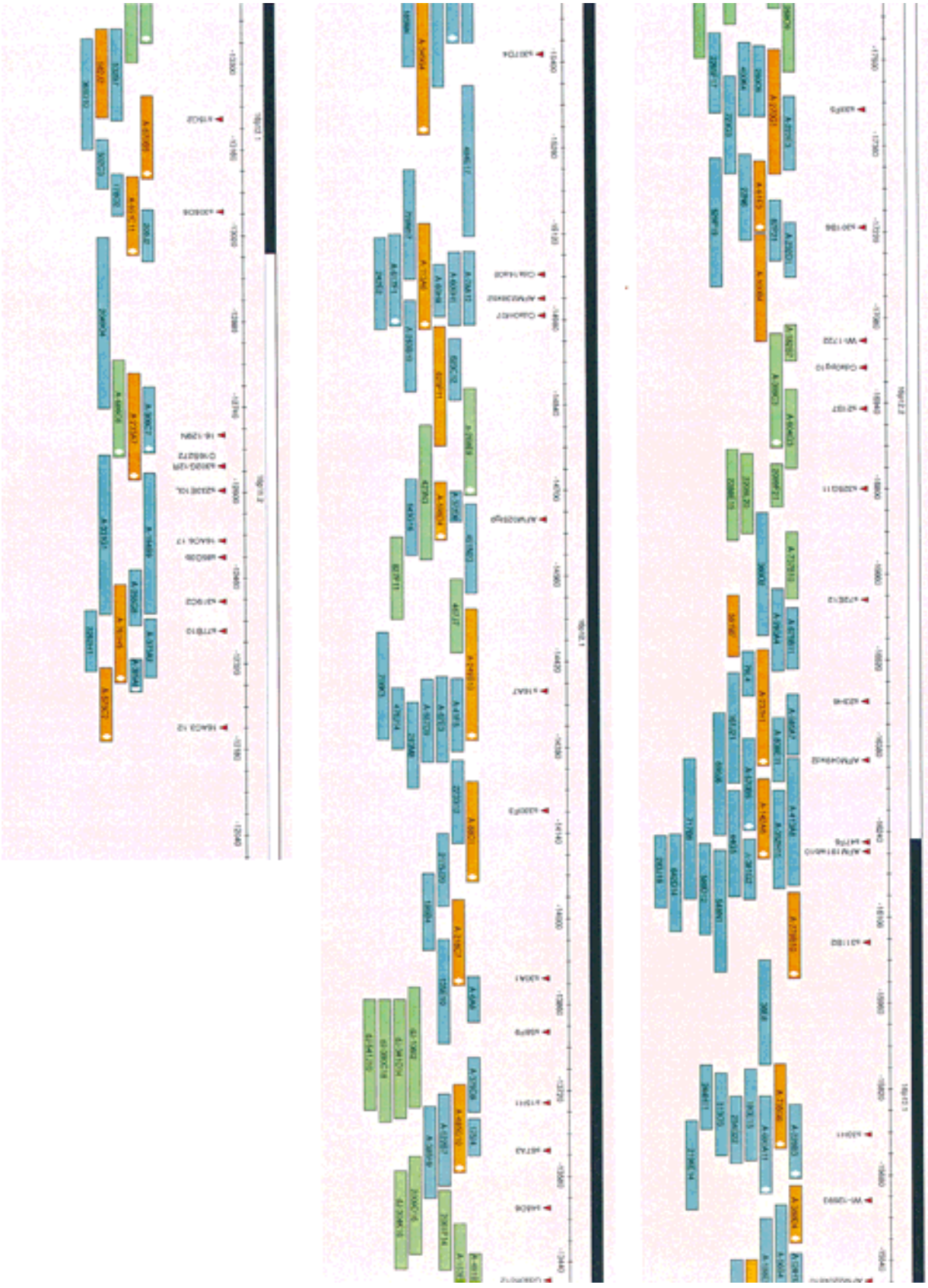


Figure 3 A complete coverage, sequence-ready BAC contig in the chromosome 16p(13.1–11.2) region. The map is drawn to scale in AceDraw window; each map unit corresponds to 1 kb of DNA. Only the minimal set of confirmed BACs is shown. Fifty-one clones (orange) were sequenced completely; 183 clones (blue) 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.

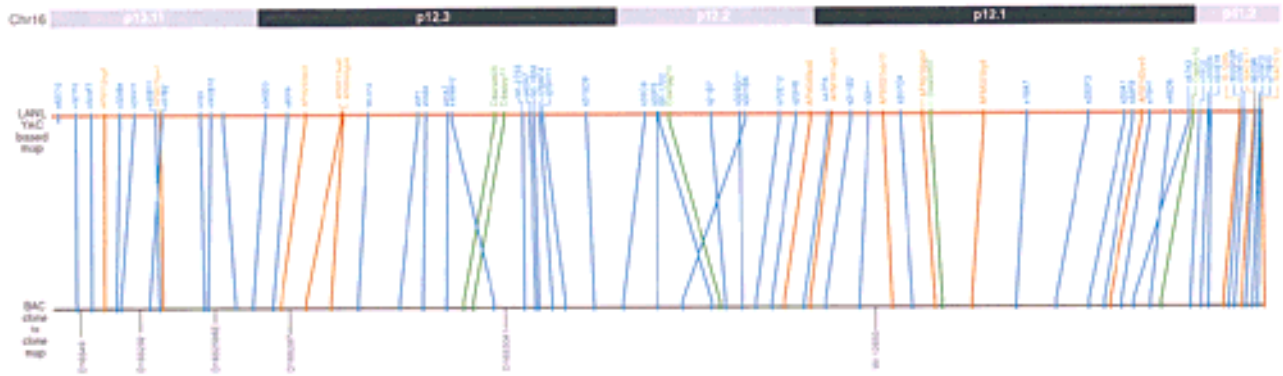


Figure 4 Comparison of the order of STSs in the BAC-based physical map with the order of the same set of STSs in the previously constructed YAC-based map.

rently 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

BAC	Band position	Positive STS	Additional FISH signals	Possible reason (sequences)
A-321F1	16p13.1	D16S2779	1p34	duplicated
A-29B12*	16p13.1	D16S2779	1p34	duplicated
A-98H8	16p13.1	D16S2714	11q23	duplicated
A-972D3	16p13.1	D16S2853	random multiple loci	
A-219D1	16p13.1	D16S2586E	16p11.2	duplicated
A-319E8	16p13.1,	D16S501	2q11.2	duplicated
A-589H1**	16p13.1-12	D16S2794	16p11.2	
A-376B3	16p12.3/13.13	D16S2720	16p11.2	duplicated
A-962B4	16p12/13.1	D16S2586E/D16S2720	16p11.2	duplicated
A-13F4	16p12-13.2	D16S2696	16p11.2, 16q23	duplicated
A-685D8*	16p12-13.2	D16S2696/D16S2794	16p11.2, 16q23	duplicated
A-793D10	16p12/13.1	D16S2696	16p11.2, 16q22	duplicated
A-613E5	16p12/13.1	D16S410	16p11.2	duplicated
A-61E3	16p12-13.1	D16S2717	16q22, 16q24, 16p11.2-12	duplicated
A-101B6	16p11.2	D16S2816	pericentromeric signals on most chromosomes, strongest at 2p11.1 & 16p11.1	repetitive
A-761H5	16p11.2	D16S2882	16p13.1	duplicated
A-575C2	16p11.2	D16S298	16p13.1	duplicated
67N16	16p11.1		16p12-13.1, 16p13.3, 16q22	duplicated
A-280B4	16p11.1-p12		16q11.2	duplicated
A-1B11	16p11.1	D16S2816	all centromeres & 2p11.1	repetitive

(*) 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 Dr. Peter de Jong's laboratory at RPCI (Buffalo NY).

BAC Clone Characterization

Single colonies were isolated from each positive BAC by streaking on agar plates. Clone culture, DNA preparation, and other standard procedures for BAC clone manipulation were performed as described previously (Kim et al. 1996). At least two single colonies were selected from each clone, grown, and the DNA samples prepared and tested for their consistency in *Hin*dIII digestion pattern on agarose gels, as well as the presence of the expected STS markers. Each single colony was kept frozen in glycerol stocks in microtiter plates until further use. BAC end sequencing was performed using miniprep DNA prepared by Autogen 740 automated miniprep machines directly as templates as described elsewhere (Kelley et al. 1999). FISH mapping was performed using miniprep DNA as described previously (Baldini et al. 1994; Weier et al. 1995). The insert sizes of the BAC clones were determined by digesting miniprep DNA with *Not*I and running on pulsed-field gels.

Restriction Fingerprinting Analysis

BAC DNA samples prepared by Autogen 740 were double digested with *Ban*I and *Msp*I (New England Biolabs, Beverly, MA) in the presence of RNase I as described previously (Kim et al. 1995). After ethanol precipitation, the fragments were end labeled by [α^{32}P]dATP using AMV-reverse transcriptase (U.S. Biochemical, Cleveland, OH). Restriction fragments were resolved on commercial precast sequencing gels (4.5% polyacrylamide, 1× TBE, 7 M urea; Stratagene, La Jolla, CA). *Hin*fI-digested λ DNAs were used as markers after end labeling with AMV-reverse transciptase. *Ban*I–*Msp*I fragments from A-334D11 were run on every gel as an internal control to gauge the consistency in electrophoretic behavior of individual gels. Digital gel images were obtained by scanning through a PhosophorImager (Molecular Dynamics, Sunnyvale, CA) and processed using the gel image analysis program (Image-2.5) available from the Sanger Center (http://www.sanger.ac.uk).

Designing BAC End-Specific OVERGOes and Library Walking

OVERGO primer pairs (J. McPherson, pers. comm.; http://www.tree.caltech.edu/protocols/overgo.html) were designed from BAC end sequences. BAC inserts were isolated by *Not*I or *Hin*dIII digestion of the BACs, resolved on 1% low-melting-point pulsed-field agarose gels, and excised of bands after ethidium bromide staining. DNA fragments were extracted from the gel by phenol extraction with 200 µl of buffer-saturated phenol, 200 µl of buffer-saturated phenol/chloroform, and ethanol precipitation. DNA pellets were dissolved in distilled water and labeled by random hexamer labeling kit (Boehringer Mannheim, Indianapolis, IN) as specified by the vendor. Complete details of the protocols for the entire experiments, including high-density filter hybridization, are available from the Caltech web site.

Sequence Match

BAC end sequences were determined for all of the candidate chromosome 16 BACs and the majority of Caltech BAC library D and other human BAC libraries (http://www.ornl.gov/meetings/bacpac/95bac.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

Baldini, A. and E.A. Lindsay. 1994. Mapping human YAC clones by fluorescence in situ hybridization using Alu-PCR from single yeast colonies. In *In situ hybridization protocols* Methods in molecular biology. (ed. K.H.A. Choo), Vol. 33, pp. 75–85. Humana Press, Clifton, NJ.

Collins, F.S., A. Patrinos, E. Jordan, A. Chakravarti, R. Gesteland, L. Walter, and the members of DOE and NIH planning groups. 1998. New goals for the U.S. Human Genome Project: 1998–2003. *Science* **282:** 682–689.

Coulson, A., J. Sulston, S. Brenner, and J. Karn. 1986. Toward a physical map of the genome of the nematode Caenorhapditis elegans. *Proc. Natl. Acad. Sci.* **83:** 7821–7825.

Dissing, M., M.M. Le Beau, and J. Pedersen-Bjergaard. 1998. Inversion of chromosome 16 and uncommon rearrangements of the CBFB and MYH11 genes in therapy-related acute myeloid leukemia: rare events related to DNA-topoisomerase II inhibitors? *J. Clin. Oncol.* **16:** 1890–1896.

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.) **377:** 335–365.

European Polycystic Kidney Disease Consortium. 1994. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* **78:** 725.

Gregory, S. 1997. Contig assembly by fingerprinting. *Gernome mapping: A practical approach* (ed. P.H. Dear), Oxford University Press, Oxford, UK.

Kelley, J.M., C.E. Field, M.B. Craven, D. Bocskai, U.-J. Kim, S.D. Rounsley, and M.D. Adams. 1999. High throughput direct end sequencing of BAC clones. *Nucleic Acids Res.* **27:** 1539–1546.

Kim, U.-J., H. Shizuya, P.J. de Jong, B. Birren, and M.I. Simon. 1992. Stable propagation of cosmid size human DNA inserts in an F factor based vector. *Nucleic Acids Res.* **20:** 1083–1085.

Kim, U.-J., H. Shizuya, J. Sainz, J. Garnes, S.M. Pulst, P. de Jong, and M.I. Simon. 1995. Construction and utility of a human chromosome 22-specific fosmid library. *Gen. Anal. Biomol. Eng.* **12:** 81–84.

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

Liu, P.P., C. Wijmenga, A. Hajra, T.B. Blake, C.A. Kelley, R.S. Adelstein, A. Bagg, J.Recior. J, Cotelingam, C.L. Willman, and F.S.Collins. 1966. Identification of the chimeric protein product of the CBFB-MYH11 fusion gene in inv(16) leukemia cells. *Genes Chromosomes Cancer* **16:** 77–87.

Loftus, B.J. U.-J. Kim, V.P. Sneddon, F. Kalush, R. Brandon, J. Furhmann, T. Mason, M. Barnstead, L. Cronin, A.D. Mays , Y. Cao, R.X. Xu, H.L. Kang, S. Mitchell, E.E. Eichler, P. Harris, J.C. Venter, and M.D. Adams. 1999. *Genomics* (in press).

Mitchison, H.M., P.B.Munroe, A.M., O'Rawe, P.E. Taschner, N. de Vos, G. Kremmidiotis, I.Lensink, A.C. Munk, K.L. D'Arigo, J.W. Anderson et al. 1977. Genomic structure and complete nucleotide sequence of the Hatten disease gene, CLN3. *Genomics* **40:** 346–350.

Olson, M.V., J.E. Dutchik, M.Y. Graham, G.M. Brodeur, C. Helms, M. Frank, M. Maccollin, R. Scheinman, and T. Frank. 1986. Random-clone strategy for genomic restriction mapping in Yeast. *Proc. Natl. Acad. Sci.* **83:** 7826–7830.

Shizuya H., B. Birren, U.-J.Kim, V. Mancino, T.Slepak, Y. Tachiiri, and M.Simon. 1992. A bacterial cloning system for cloning large human DNA fragments. *Proc. Natl. Acad. Sci.* **89:** 8794–8797.

Soderlund, C.A. and I. Longden. 1996. FPC. *Technical report SC-01-96*. The Sanger Centre, Hinxton, Cambridge, UK.

Stallings, R.L., S.A. Whitmore, N.A. Doggett, and D.F. Callen. 1993. Refined physical mapping of chromosome 16-specific repetitive DNA sequences. *Cytogenet Cell Genet.* **63:** 97–101.

Sulston, J., F. Mallett, R. Staden, R. Durbin,T. Horsnell, and A. Coulson. 1988. Software for genome mapping by fingerprinting techniques. *CABIOS* **4:** 125–132.

Sulston, J., F. Mallett, R. Durbin, and T. Horsnell. 1989. Image analysis of restriction enzyme fingerprint autoradiograms. *Comput. Appl. Biosci.* **5:** 101–106.

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

Weier, H.U.G., M. Wang, J.C. Mullikin, Y. Zhu, J.F. Cheng, K.M. Greulich, A. Bensimon, and J.W. Gray. 1995. Quantitative DNA fiber mapping. *Hum. Mol. Genet.* **4:** 1903–1910.

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