MICROBIAL & COMPARATIVE GENOMICS
Volume 4, Number 3, 1999
Mary Ann Liebert, Inc.

Physical Map and Gene Survey of the *Ochrobactrum anthropi* Genome Using Bacterial Artificial Chromosome Contigs

JEFFREY P. TOMKINS[1], HEATHER MIULLER-SMITH[1], MACIEK SASINOWSKI[1], SANGDUN CHOI[1], HEATHER SASINOWSKA[2], MATTHEW F. VERCE[3], DAVID L. FREEDMAN[3], RALPH A. DEAN[1], and ROD A. WING[1]

ABSTRACT

Bacterial artificial chromosome (BAC) clones are effective mapping and sequencing reagents for use with a wide variety of small and large genomes. This report describes research aimed at determining the genome structure of *Ochrobactrum anthropi,* an opportunistic human pathogen that has potential applications in biodegradation of hazardous organic compounds. A BAC library for *0. anthropi* was constructed that provides a 70-fold genome coverage based on an estimated genome size of 4.8 Mb. The library contains 3072 clones with an average insert size of 112 kb. High-density colony filters of the library were made, and a physical map of the genome was constructed using a hybridization without replacement strategy. In addition, 1536 BAC clones were fingerprinted with *Hin*dIII and analyzed using IMAGE and Fingerprint Contig software (FPC, Sanger Centre, U.K.). The FPC results sup- ported the hybridization data, resulting in the formation of two major contigs representing the two major replicons of the *0. anthropi* genome. After determining a reduced tiling path, 138 BAC ends from the reduced tile were sequenced for a preliminary gene survey. A search of the public databases with the BLASTX algorithm resulted in 77 strong hits (E-value < 0.001), of which 89% showed similarity to a wide variety of prokaryotic genes. These results provide a contig-based physical map to assist the cloning of important genomic regions and the potential sequencing of the *0. anthropi* genome.

INTRODUCTION

Ochrobactrum anthropi was originally classified in the Centers for Disease Control and Prevention O(CDC) group Vd and was believed to have many similarities to the genus *Achromobacter*. However, Holmes et al. (1988) proposed classification of group Vd organisms as *0. anthropi,* emphasizing that this new genus and species is actually quite distinct from *Acromobacter* organisms. Much of what is currently known about *0. anthropi* is based on its emergence as an opportunistic human pathogen. Examples include

[1]Clemson University Genomics Institute, [2]Department of Mathematical Sciences, and [3]Department of Environmental Engineering and Science, Clemson University, Clemson, South Carolina.

0. anthropi meningitis (Chang et al., 1996) and infections in patients with permanently installed catheters (Ainor et al., 1994). Both cases involved immunocompromised individuals. 0. *anthropi* also has potential applications for bioremediation. Laura et al. (1996) isolated a strain from activated sludge and demonstrated its ability to biodegrade the pesticide atrazine as a sole source of carbon and energy. Several novel enzymes have also been identified from *0. anthropi,* including a D-stereospecific arninopeptidase (Asano et al., 1992) and a carboxylesterase that specifically hydrolyzes only 1-methyl acetate (Murase et al., 1991).

Relatively little is known about the genome structure of *0. anthropi.* The 16S ribosomal genes from two strains have been reported (Yanagi and Yamasoto, 1993; *http:llwww.ncbi.nlm.nih.govlWebIGenbankl index.html,* along with the sequence of a D-arninopeptidase gene (Asano et al., 1992) and the spacer region between 16S and 23S RRNA (Rijpens et al., 1996). Thus, determining this organism's genome organization and structure will allow for the cloning of genes involved in important biochemical pathways associated with biodegradation. A detailed contig-based map would also be invaluable for whole genome shotgun sequencing by helping to align sequencing contigs and by directing sequence gap closure.

Physical maps for bacterial genomes are generally developed using a restriction mapping approach with rare cutting enzymes and pulsed-field gel electrophoresis (PFGE) (Irazabal et al., 1997; Kundig et al., 1993; Riethman et al., 1997; Ron-ding et al., 1992). However, these maps do not provide sufficient resolution to act as guides for whole or partial genome sequencing. They also do not provide the resources needed for the immediate cloning of genes involved in key biochemical pathways. Recently, bacterial artificial chromosome (BAC) library technology has been exploited to develop high-resolution contig-based physical maps for archael and bacterial genomes (Diaz-Perez et al., 1997; Brosch et al., 1998; Dewar et al., 1998). The BAC system is very appealing as a vehicle for genome analysis in bacteria, as the confounding effects of repetitive DNA are expected to be relatively negligible. Therefore, we have employed BAC technology as a tool to develop a physical framework for the *0. anthropi* genome. The specific goals of the present study were (1) to develop and characterize a BAC library for *0. anthropi,* (2) to develop a high-resolution physical map of the genome using BAC contigs, and (3) to BAC end sequence a tiling path of clones for a preliminary gene survey.

MATERIALS AND METHODS

Culture isolation and identification

An aerobic enrichment culture was developed using an inoculum of activated sludge and vinyl chloride as the sole source of organic carbon and energy (Freedman and Verce, 1997). Samples were streaked on Noble agar and incubated in a stainless steel cylinder charged with 10% vinyl chloride and 90% air. Isolated colonies developed in approximately 1 month and were picked and streaked two more times. Colonies that were transferred back to minimal medium and supplied with only vinyl chloride have not yet grown, but colonies do grow quickly on trypticase soy broth.

The isolate was identified based on its 16S RDNA sequence (Dorsch and Stackebrandt, 1992). DNA was extracted from cell pellets with a silica glass preparation in the presence of chloroform/phenol/isoamyl alcohol. Small subunit rDNA was amplified by the polymerase chain reaction (PCR) using two 16S rDNA primers specific to the bacterial domain (Alm et al., 1996; Kane et al., 1993). The amplified product was cloned into a pUC plasmid (TA cloning kit, Invitrogen Corp., San Diego, CA) and sequenced (ABI Model 377) using M13 forward and reverse primers (Invitrogen Corp.) and two pairs of internal and nested primers (Altschul et al., 1997). The resulting sequence was identified as 16S rDNA from *0. anthropi* based on the Sequence Match program of the Ribosomal Databases Project 11 (Maidak et al., 1999) and the BLAST alignment tool of GenBank (Altschul et al., 1997).

BAC library construction

The single-copy BAC library vector, pBeloBAC II, was obtained from Dr. Hiroaki Shizuya (1992) and prepared as described by Woo et al (1994). Megabase bacterial DNA embedded in agarose plugs was obtained as described by Riethman et at. (1997). Partial digests of megabase DNA were performed as follows. Chopped plugs were distributed in 100-µl aliquots and incubated on ice for 30 n-tin with 14 µl 10X

enzyme buffer, 14 µl 40 mM spermidine, and 1.4 µl bovine serum albumin (BSA). After a second 30-min incubation with 4 U *Hin*dIII on ice, digestion reactions were allowed to proceed at 37'C for 30 min. Digestions were stopped by placing on ice and adding 1:10 vol 0.5 M EDTA. Partially digested megabase DNA was subjected to a single size selection by PFGE (CHEF mapper apparatus, BioRad, Hercules, CA). Size selection conditions were 1% low gelling temperature agarose, 20-40-see linear ramp, 6 V/cm, 14 °C, 20 h run time, and 0.5X TBE buffer. Seven fractions between 100 and 450 kb were cut from the gel based on a 50-kb lambda ladder reference. DNA was removed from the agarose by Gelase (Epicentre, Madison, WI) treatment, and legations were performed in 100-µl reactions using 20 ng vector and 200 ng *0. anthropi* DNA and allowed to proceed for 20 h at 16 °C. After dializing legations against 0.5X TEIO:I, transfonnations were performed using 2 µl ligation and 20 µl competent cells (DH10B, GIBCO-BRL, Gaithersburg, MD). Electroporations were performed on a cell porator with voltage booster (GIBCO-BRL) using 380 V at resistance of 4 kΩ. Transformed cells were diluted immediately with 0.5 ml SOC medium (Sambrook et al., 1989) and incubated at 37 °C for 50 min before being plated on selective medium (Luria-Bertani [LB] medium) with 12.5 µg/µl chloramphenicol, 0.55 mM IPTG, and 80 µg/ml X-Gal. After a 20 h incubation at 37 °C, the plates were placed at ambient temperature in the dark for an additional 20 h to allow stronger color development of nonrecombinant colonies. After determining insert sizes of clones from each of the seven legations, the ligation from the 300-350 kb range was used for additional transformations to construct the library. Recombinant white colonies were picked by hand and stored individually in eight 384-well microliter plates (Genetix, Christchurch, Dorret, UK) containing 50 µl freezing broth (Woo et al., 1994). After incubation overnight, micrometer plates were stored at –80 °C. Two copies of the library were made using a hand-held replicating device and stored in separate –80 °C freezers.

BAC clone characterization

To prepare BAC DNA, 3 ml LB chloramphenicol (12.5 µg/µl) cultures were grown overnight in 5-cell autogen tubes and miniprepped robotically (Autogen 740 plasmid isolation system). To estimate insert size and determine distribution of clone size, a total of 136 BAC preparations were performed from clones selected at random throughout the library. The BAC DNA was digested with 7.5 U (10 h at 37 °C) *Not*I and analyzed by PFGE in 1% agarose gels (6 V/cm, 5-15 sec switch time, 15 h run time, 14 °C). Southern blots of size-separated BAC inserts were performed using standard protocols (Sambrook et al., 1989) after UV nicking the gels (Gene Linker, BioRad).

BAC library screening

High-density colony filters for hybridization-based screening of the library were prepared using a Beck- man Biomek 2000 robotics workstation. Clones were gridded in double spots using a 3 x 3 array on 10 X 12 cm Hybond N^+ filters (Amersham, Arlington Heights, IL). This gridding pattern allows 1536 clones to be represented per filter. Colony filters were treated and hybridized using standard techniques (Sambrook et al., 1989). For the hybridization without replacement experiments, BAC inserts for probes were cut from ethidium bromide-stained CHEF gels, and DNA was extracted using a QIAEX H gel extraction kit (Qiagen, Chatsworth, CA). Radiolabeling of BAC insert DNA and hybridization of colony filters were per- formed using standard techniques (Sambrook et al., 1989). Autoradiograph data generated by the hybridization experiments were classified by signal intensity on a scale of 3, 6, and 9 as determined by weak, moderate, and strong hit intensities, respectively. These data were then used to generate a probe versus hit binary data matrix. A UNIX-based computer program based on a weighted strength algorithm was used to analyze the binary data (Sasinowska and Sasinowski, 1999).

BAC fingerprinting

BAC DNA for fingerprinting was prepared as described previously. The process of fingerprinting, digitizing of gel images, and contig building has been described previously (Marra et al., 1997). The IMAGE software program was used to digitize gel images, and the Fingerprint Contig (FPC) program was used for contig building. A users' guide for FPC may be viewed at the following Sanger Centre website:

http://www.sanger.ac.uk/Users/cari/fpc_ faq.shtml.

BAC end sequencing

To prepare BAC DNA for end sequencing, 4 ml LB chloramphenicol (12.5 μg/μl) was inoculated with 4 μ BAC freezer stock, and cultures were grown for 20 h in 6-cell autogen tubes at 37°C. BAC cultures were miniprepped robotically (Autogen 740 plasmid isolation system), and DNA was resuspended in 25 μl 1 mM Tris. Sequencing reactions were set up according to manufacturer's instructions for the Big Dye Terminator chemistry (Perkin-Elmer, Norwalk, CT). Samples were then loaded onto 48-lane sequencing gels in ABI377 automated sequencers. Gels (250 ml) were composed of the following: 5% Long Ranger polyacrylamide (FMC, Rockland, ME), 6 M urea, 18 μl N,N,N′,N′ = tetramethylethylenediamine (TEMED), 150 μl ammonium persulfate (10% stock), and 1× TBE buffer. Reaction products were electrophoresed using a 3.5-h run. DNA sequence data were analyzed using the ABI base calling software program (Applied Biosystems, Foster City, CA). The resulting file was compared with the GenBank database using the BLASTX algorithm and the BatchBlast script (Baylor College of Medicine). The results were sorted using scripts developed at Clemson University Genomics Institute (CUGI). Search results were grouped according to gene function. The sequence data described in this article were submitted to the GenBank data library under accession numbers AQ242112–AQ242254, Rockland, ME.

RESULTS

BAC library construction and characterization

Prior to BAC library construction, there was concern that cloning bacterial DNA in a bacterial cloning system might create logistical problems because of potential similarities in DNA structure. Therefore, South-

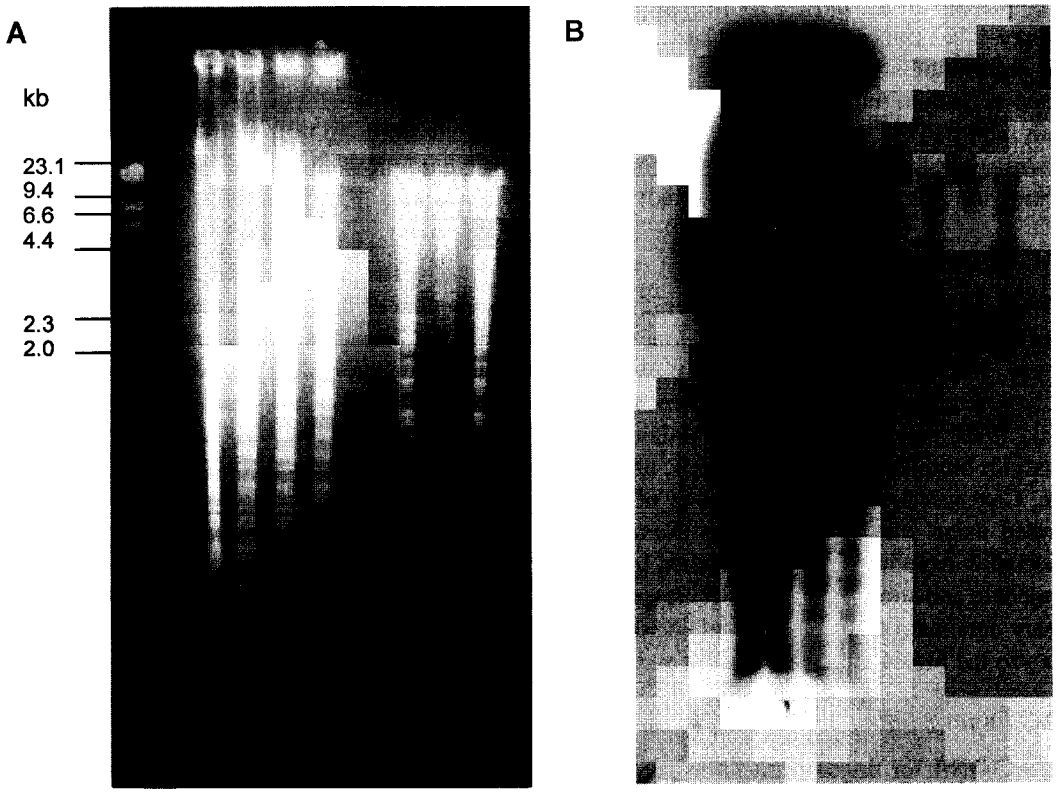

FIG. 1. **(A)** Ethidium bromide-stained agarose gel of size-separated DNA from *O. anthropi* and *E. coli* digested with *Hin*dIII. **(B)** Autoradiograph of a Southern blot of gel A hybridized with radioactively labeled total *O. anthropi* DNA. Lane 1, *Hin*dIII digested-lambda DNA marker; lanes 3–6, *O. anthropi* DNA; lanes 8–10, *E. coli* DNA.

ern blots of restricted *Escherichia coli* and *O. anthropi* genomic DNA were probed with total *O. anthropi* genomic DNA. Results indicated that DNA sequence similarities between the genomes were negligible, as there was no detectable hybridization or very weak hybridization (Fig. 1). A BAC library was constructed for *O. anthropi* that is suitable for physical mapping and cloning genes associated with bioremediation. *Hin*dIII was used as the cloning enzyme because complete digests with *O. anthropi* genomic DNA produced fragments ≤ 30 kbp. The library consists of 3072 clones stored in eight 384-well microtiter plates. Less than 2% of the clones do not contain inserts, as judged by random analysis of BAC sampled from the library. A random sampling of 136 BAC taken from the library indicated an average insert size of 112 kb with a range of 45–254 kb. Figure 2A shows 15 randomly selected clones digested with *Not*I to release the

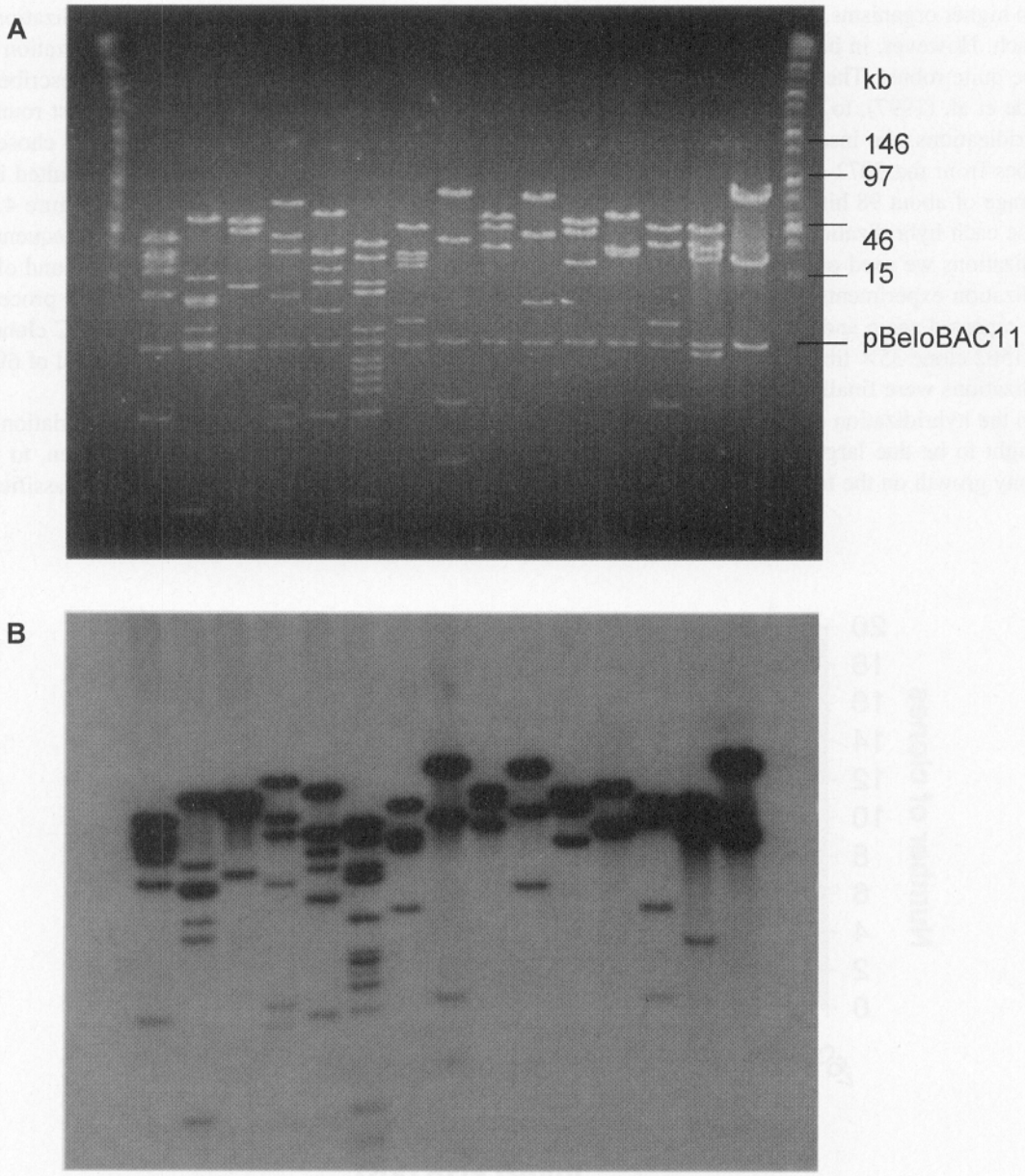

FIG. 2. Fifteen randomly selected *O. anthropi* BAC clones. **(A)** Ethidium bromide-stained CHEF gel (5–15 sec switch time, 14 h) showing insert DNA above and below the 7.5-kb pBeloBAC11 vector band. **(B)** Autoradiograph of gel in **A** after Southern transfer and probing with total *O. anthropi* genomic DNA. Molecular marker is MidRange I (New England Biolabs, Beverly, MA).

insert. The two *Not*I sites in pBeloBAC11 flank the multicloning site. Because *Not*I is a GC-8-base cutter and the *O. anthropi* genome is relatively GC rich, digestion typically generates a vector band plus 1–8 insert bands based on our data. Figure 2B shows a Southern blot of the gel in Figure 2A probed with total *O. anthropi* genomic DNA, indicating that the source of cloned DNA originated from *O. anthropi*.

To determine the size distribution of BAC clones in the library, the 136 BAC analyzed with *Not*I digests were grouped by insert size, and the insert size of each clone was plotted against the frequency of each group of clones represented in the library (Fig. 3). Based on this analysis, >65% of the clones in the library have an average insert size >100 kb.

Contig assembly by hybridization

In higher organisms, repetitive elements can cause problems in contig assembly using a hybridization approach. However, in bacterial genomes, repetitive DNA is relatively minimal, allowing hybridization data to be quite robust. Therefore, a hybridization without replacement strategy was employed, as described by Prade et al. (1997), to generate a complete physical map of the *O. anthropi* genome. In the first round of hybridizations, the inserts from 10 random BAC clones, each about 100–160 kb in size, were chosen as probes from the 3072-clone BAC pool to hybridize to the 70× deep library. Each probing resulted in an average of about 98 hits. An autoradiograph picture of one colony hybridization is shown in Figure 4. Because each hybridization experiment generated such a large amount of redundant data, on subsequent hybridizations we used only half of the library, which provided a 35× genome coverage. Each round of hybridization experiments used inserts from 6–14 randomly selected BAC clones as probes. This procedure was repeated again and again until no new hits were achieved, indicating that nearly all the BAC clones in the 1562 clone 35× library were either used as probes or identified by other BAC probes. A total of 69 hybridizations were finally used to complete the data collection process.

In the hybridization experiments, the intensity of the hits identified by a probe varied. The variation was thought to be due largely to the degree of overlap between a probe and an insert or, less often, to poor colony growth on the filters. Therefore, data generated by the hybridization experiments were classified by

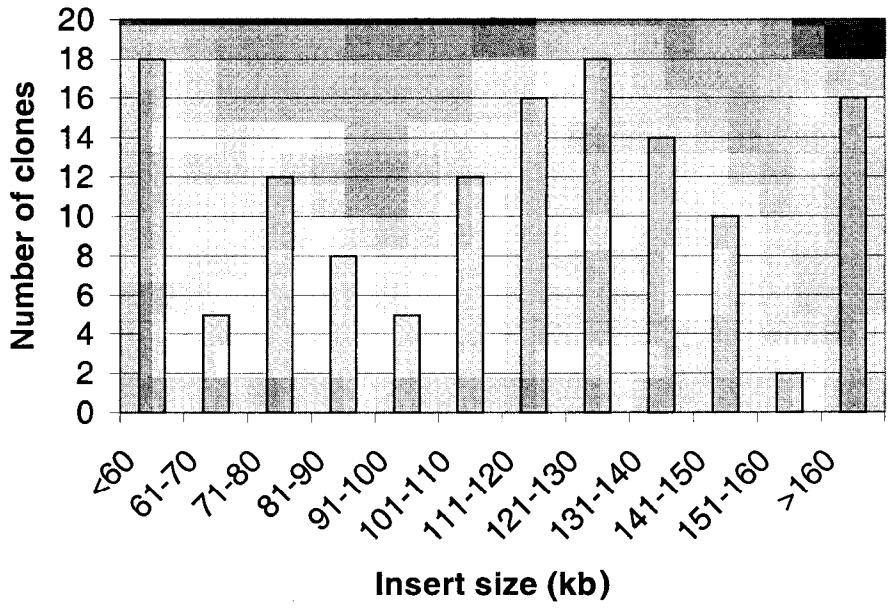

FIG. 3. Insert size distribution of clones in the *O. anthropi* BAC library. To estimate insert size range, BAC DNA from 136 randomly selected clones was analyzed, as shown in Figure 1A. Results indicate that the average insert size is 112 kb, with >65% of the clones > 100 kb.

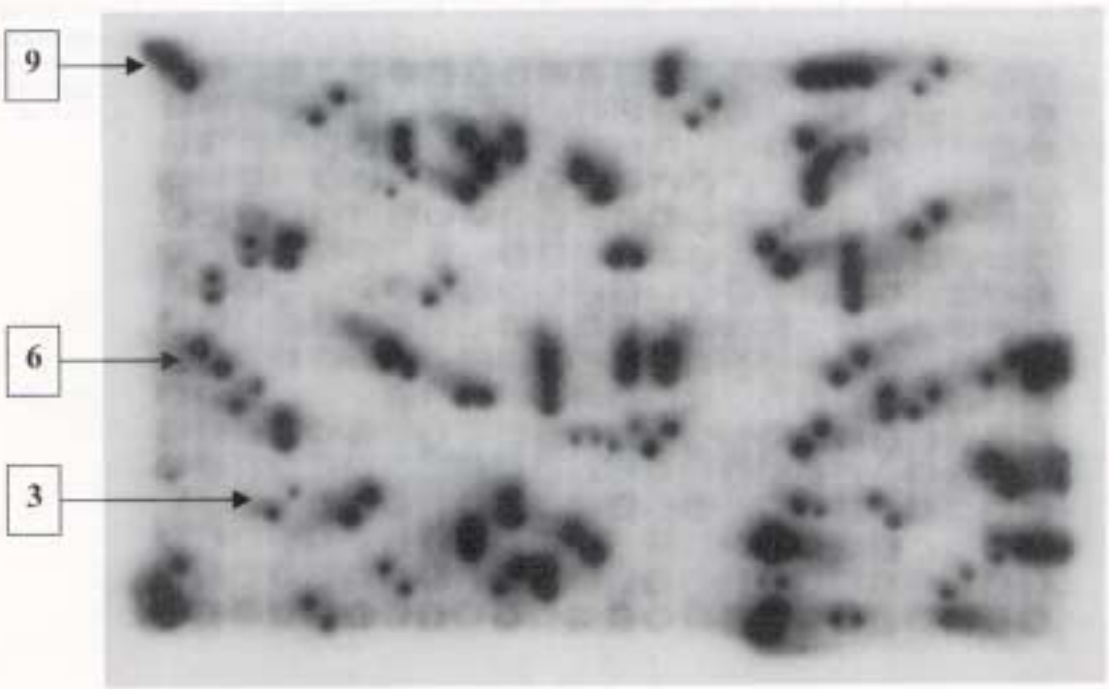

FIG. 4. Example of an *O. anthropi* colony filter hybridization using a randomly selected 100-kb BAC insert as a probe. Each filter allows 1536 BAC to be screened (~40× genome coverage). Arrows point toward hits classified by intensity as 3, 6, or 9 for weak, moderate, and strong signals, respectively.

signal intensity on a scale of 3, 6, and 9 as determined by weak, moderate, and strong hit intensities, respectively (Fig. 4). These data were then used to generate a probe versus hit binary data matrix. A hybridization strength sensitive algorithm was used to convert input consisting of an unordered matrix representing hybridization events into an ordered matrix representing the physical order of clones along the chromosome. This was accomplished by first ordering the rows, which represented the clones, and then ordering the columns, which represented the probes. The algorithm that performed the first task was based on a weighting scheme that combined the frequency of identical probe hits, the number of hits per clone, and the strength of hits to obtain the most appropriate ordering of clones. The algorithm used to order the columns was based on selecting columns with consecutively repeated probe hits (Sasinowska and Sasinowski, 1999).

Analysis of the binary matrix resulted in the development of two large contigs and two small contigs. Arratia et al. (1991) pointed out that two or more hits should support a link in the hybridization contig maps to avoid false joints due to experimental problems. In our analysis, each link was supported by at least two hits. The two large contigs contained 444 and 283 clones, and the two small contigs contained 12 and 10 clones. It is likely that these contigs represent the two major replicons and two plasmids of *O. anthropi* that Jumas-Bilak et al. (1998) characterized using PFGE of linearized chromosomes. The remaining clones were either unidentified because of poor colony growth on the filters, formed numerous small contigs, or represented empty vectors. Redundant clones in the contigs were then removed to form a reduced tiling path containing a total of 69 clones, which were used for a gene survey via BAC end sequencing. The 69 clones used for BAC end sequencing were further reduced by removing additional redundant clones. This resulted in a sequence-ready minimum tile of only 38 and 32 clones for contigs one and two, respectively, whereas the two small contigs contained six clones each. Table 1 shows the minimum tiles as ordered binary data matrices produced by the weighted strength algorithm program.

TABLE 1. ORDERED DATA MATRIX SHOWING MINIMUM TILES FOR HYBRIDIZATION DATA AMONG FOUR CONTIGS. BAC CLONES USED AS PROBES ARE COLUMN HEADINGS AND BAC CLONES HIT BY PROBES ARE ROW HEADINGS

Contig 1	1E01 (i)	1B18 (2)	1B01 (2)	3G06 (m)	1I10 (2)	1K15 (5)	1A13 (i)	1B21 (5)	1A20 (5)	1C21 (i)	1L11 (5)	1F23 (5)	1H17 (5)	1H16 (5)	1O18 (5)	5A06 (m)	1E13 (i)	1C05 (5)	1A02 (5)	2A22 (5)	1I16 (6)	1B09 (6)	1B13 (6)	1D06 (i)	1C06 (7)
1O20 (2)	9	9																							
1F20 (2)		6	6	9	9	3																			
2L18 (5)						9	9	9	6																
3F13 (5)									9	9	6	6													
3O07 (m)											6	6	9	6											
1I11 (5)													6	9	6	9	6								
1H12 (5)																	9	9							
1N17 (5)																		9	9	6					
1L09 (6)																				9	9	9	9	6	
4H01 (7)																								9	9

Contig 2	1D15 (4)	1H20 (4)	1B14 (4)	1A01 (4)	5A04 (m)	1A22 (4)	1D13 (1)	4C18 (1)	1B07 (1)	4A01 (1)	4A04 (1)	1B11 (1)	1H06 (1)	1B02 (1)	1B09 (1)	1C12 (i)	1C13 (i)	6A04 (m)
3O01 (m)	9	9	9	9														
3I23 (m)				9	9													
2P10 (4)					9	9												
4O24 (1)						9	9											
2K20 (1)							9	6	6									
2A01 (m)								9	9	9	9							
1D24 (1)										6	6	9						
3P02 (1)												9	9	6	6			
4L10 (1)															6	6		
1N24 (1)													•			9	9	6

Contig 3	1D19 (4)	1B03 (i)	4C19 (i)	1H14 (i)	1A16 (4)	5A09 (m)
4E01 (4)	6	9	9			
2M01 (m)		6	9	9		
3L01 (4)				6	9	9
2B21 (4)						3

Contig 4	1J02 (i)	1B05 (4)	2C02 (4)	1B20 (4)	1B12 (5)	6A03 (m)
4P12 (4)	9	6	6			
1J07 (5)			3	9	9	
4M03 (5)					9	9
4D05 (5)						9

Numbers and letters in parentheses: number = anchored FPC contig number, i = independent FPC clone, m = missing FPC clone.

Contig assembly by DNA fingerprinting

To augment the hybridization-based physical mapping data, 1536 BAC clones (35× genome coverage) obtained consecutively from the first four 384-well plates of the library were subjected to *Hin*dIII fingerprinting. After the BAC digests were separated on analytic agarose gels, followed by staining, gel pictures were imaged using a fluorimaging system (Bio-Rad), saved as TIFF files and transferred to a UNIX workstation. Restriction bands were called and digitized using IMAGE software (Sanger Centre, U.K.). Bands <1 kb were ignored because they were not well resolved and accounted for <1% of the total length of the BAC inserts. Resulting digitized bands that contained the relative mobilities of all the restriction fragments from the BAC clones were then transferred to a database.

Contig assembly using FPC was performed at a tolerance value of 7 and a cutoff score of 10^{-12}. These parameter thresholds provided the fewest contigs with a minimal amount of error. For a detailed discussion of FPC parameter thresholds, see the FPC users' guide at *http://www.sanger.ac.uk/Users/cari/fpc*_faq.shtml. Compared to the hybridization-based data, the FPC analysis produced more contigs. In summary, results of

the analysis provided 13 high-quality FPC contigs (Table 2). Seven of the contigs were quite large, containing between 25 and 328 clones and ranging in size between 216 kb and 1288 kb. The remaining contigs were small, containing between 2 and 6 clones and did not exceed 100 kb in size.

Physical map construction by combined methods

Results obtained using the hybridization data and the fingerprint data were combined to produce a physical map. With uncharacterized bacterial genomes, there is a lack of genetic markers to assist the alignment of contigs. However, there is also a lack of large repetitive elements in bacterial genomes, which allows hybridization data to be quite reliable. Therefore, the hybridization data were used as the foundational contig assembly, and the fingerprint data were used as a verification tool. By manually comparing the FPC contigs with the hybridization-based contigs, we were able to verify the physical associations among clones. In addition, the positions of the BAC clones from the hybridization-based minimum tiles were mapped in the FPC database. Not only did this allow verification of the hybridization data, but it also allowed placement of the FPC contigs within the hybridization-based contigs (Table 1).

Using PFGE of linearized chromsomes, Jumas-Bilak et al. (1998) estimated that *O. anthropi* contained two major replicons of about 2700 kb and 1900 kb in size in addition to two smaller plasmids varying in size between 50 and 150 kb depending on the strain tested. Our contig-based physical mapping data produced two large contigs that we estimated to be 2313 kb and 1882 kb for contigs one and two, respectively, based on the fingerprints of the clones within the minimum tiles, taking into account overlap between clones. When adding up the sizes of the FPC contigs that comprised the two large hybridization contigs, nearly similar estimates of 2028 kb and 2197 kb for contigs 1 and 2, respectively, were obtained. These large contigs probably correspond to the two major replicons characterized previously by Jumas-Bilak et al. (1998). Any discrepancies in contig sizes between the two studies may be a result of the methods used to obtain sizes.

TABLE 2. RESULTS OF FPC ANALYSIS OF 1536 FINGERPRINTED *O. anthropi* BAC CLONES[a]

FPC contig	*No. of clones*	*FPC score*[b]	*Q*[c]	*Estimated contig size (kb)*
1	222	0.900	0	909
2	57	0.959	0	316
3	30	1.000	0	77
4	328	0.918	1	1288
5	305	0.890	3	1265
6	46	9.943	0	232
7	25	0.973	0	216
8	6	1.000	0	50
9	3	1.000	0	31
10	2	1.000	0	39
11	2	1.000	0	42
12	2	1.000	0	46
13	2	1.000	0	35

[a]For further description of FPC terms and protocols, see the FPC users' guide at *http://www.sanger.ac.uk/Users/cari/fpc_faq.shtml.*

[b]*FPC score* = *((P + left + right) − n)/(total ends)* where *n* is the number of negative bands, *P* is the number of positive bands, *left* and *right* are the maximum number of extra bands to left and right ends, *ends* is an approximate number of end fragments, and *total* is the total number of bands.

[c]*Q*, clones that incorrectly map to the same space due to similar fingerprints, have poor fingerprints, or have a suboptimal solution.

TABLE 3. SUMMARY OF BAC END SEQUENCES (BES) SHOWING THE STRONGEST SIMILARITY (E-VALUE < 0.001 ONLY) TO DATABASE ACCESSIONS FROM THE BLASTX SEARCH

BES	BES ID No.	Best match ID No.	Gene description	Organism	AA identity	AA similarity
Unclassified/Miscellaneous						
1F20-F	AQ242131	M93187	ORFA	*Arthrobacter sp.*	38% (39/103)	56% (58/103)
1F20-R	AQ242132	AE000203	41.1 KD protein in putphoh intergenic region	*Escherichia coli*	31% (71/211)	54% (114/211)
1N24-F	AQ242193	Z69368	57.6 KD protein in chromosome I	*Schizosaccharomyces pombe*	48% (60/124)	61% (76/124)
1N24-R	AQ242194	AE000239	CPXA_PSEPU	*Escherichia coli*	64% (90/140)	74% (104/140)
2A18-R	AQ242245	AE000117	59.6 KD protein in ARAC-TBPA intergenic region	*Escherichia coli*	48% (72/147)	70% (103/147)
2B21-F	AQ242228	AL021411	Hypothetical protein	*Streptomyces coelicolor*	45% (18/40)	62% (25/40)
2D09-R	AQ242207	AE000079	16.7 KD protein	*Rhizobium sp.* NGR234	62% (58/93)	75% (70/93)
2K05-R	AQ242168	Z96797	Hypothetical protein	*Mycobacterium tuberculosis*	61% (72/118)	74% (88/118)
2L23-F	AQ242151	AJ224445	orF20	*Sinorhizobium meliloti*	62% (86/137)	73% (101/137)
2M03-F	AQ242201	AE000064	34.7 KD PROTEIN	*Rhizobium sp.* NGR234	57% (75/130)	73% (95/130)
2P18-R	AQ242198	Z78418	Hypothetical protein	*Caenorhabditis elegans*	40% (33/82)	52% (43/82)
3I23-R	AQ242118	Z74697	139.6 KD protein	*Mycobacterium*	48% (62/130)	62% (80/130)
3O01-F	AQ242123	D90913	Hypothetical protein	*Synechocystis sp.*	58% (18/31)	71% (22/31)
3P09-F	AQ242232	AL021246	Hypothetical protein	*Mycobacterium tuberculosis*	41% (40/97)	66% (64/97)
3P15-F	AQ242155	D64000	Hypothetical protein	*Synechocystis sp.*	56% (93/164)	75% (124/164)
4P12-R	AQ242235	Z26318	Royal jelly protein	*Apis mellifera*	23% (21/88)	47% (42/88)

F = Forward sequence, R = reverse sequence.

TABLE 3. *Continued*

Cell division/DNA replication and repair						
1A08-F	AQ242129	U82225	GyrB Topoisomerase	*Bartonella bacilliformis*	74% (115/156)	88% (137/156)
1A24-R	AQ242188	D90888	radB	*Escherichia coli*	37% (46/122)	59% (73/122)
1E13-F	AQ242165	AE000068	Putative replication protein A	*Rhizobium sp.* NGR234	59% (69/116)	75% (87/116)
1L09-R	AQ242209	U80928	RepA	*Rhizobium etli*	74% (78/106)	81% (86/106)
4A02-F	AQ242191	M30255	DNA ligase	*Escherichia coli*	58% (86/148)	76% (113/148)
4E01-F	AQ242218	U37793	DNA polymerase III, beta chain	*Caulobacter*	46% (52/111)	63% (70/111)
4E01-R	AQ242219	U32760	Cell division protein	*Haemophilus influenza*	50% (57/113)	68% (77/113)
Cell signalling						
2D24-F	AQ242242	AE000347	Signal recognition particle protein	*Escherichia coli*	48% (61/127)	64% (82/127)
2C07-R	AQ242120	AE000064	Mannose-1-phosphate guanyltransferase	*Rhizobium sp.* NGR234	50% (103/205)	64% (132/205)
2L16-R	AQ242231	Z70692	Ketopantoate hydroxymethyl transferase	*Mycobacterium tuberculosis*	51% (38/74)	60% (45/74)
3P07-F	AQ242195	AE000098	Hydrolase/peptidase	*Rhizobium sp.*	30% (43/146)	54% (79/146)
Cell structure/membrane proteins						
1F24-R	AQ242174	U26443	Porin	*Brucella suis*	66% (60/91)	77% (70/91)
1G24-F	AQ242145	L29029	VSP-3 cell wall protein	*Chlamydomonas reinhardtii*	28% (37/132)	44% (58/132)
2D24-R	AQ242243	AJ006770	Extensin	*Cicer arietinum*	51% (36/71)	58% (41/71)
Metabolism						
1A09-F	AQ242127	AF035413	AgaE opine catabolism	*Agrobacterium tumefaciens*	46% (93/203)	62% (125/203)
1B11-F	AQ242183	P70865	Alanyl tRNA synthetase	*Bartonella bacilliformis*	66% (32/48)	77% (37/48)
1C18-R	AQ242122	AF054609	ATP synthase subunit B	*Brucella melitensis*	83% (47/56)	89% (50/56)
2A10-R	AQ242239	X99599	ATP synthase delta chain	*Rhodobacter capsulatas*	43% (53/123)	65% (81/123)
2D09-F	AQ242206	U39940	Choline dehydrogenase	*Sinorhizobium meliloti*	45% (57/124)	63% (79/124)
1F24-F	AQ242173	NUO9	NADH dehydrogenase subunit 8	*Reclinomonas americana*	78% (83/106)	84% (89/106)
2G14-F	AQ242222	U84888	Phosphoglucomutase	*Mesembryanthemum crystallinum*	55% (47/85)	81% (69/85)
2I06-F	AQ242116	X70420	Molybdenum cofactor biosynthesis	*Escherichia coli*	51% (53/103)	70% (72/103)
2K20-F	AQ242171	U25800	NADH dehydrogenase	*Rhodobacter capsulatas*	80% (85/106)	87% (92/106)
2L18-R	AQ242200	L36823	Cinnamyl-alcohol dehydrogenase	*Stylosanthes humilis*	67% (67/114)	68% (78/114)
2L23-R	AQ242152	U38543	Oxidoreductase	*Escherichia coli*	31% (32/101)	53% (54/101)
2P10-F	AQ242114	Z99120	Allantoinase	*Bacillus subtillis*	36% (19/52)	59% (31/52)
3C07-F	AQ242185	E64180	Lipid A disaccharide synthetase	*Haemophilus influenzae*	39% (28/72)	64% (46/72)
3C07-R	AQ242186	M59236	Cobyrinic acid A,C-diamide synthase	*Pseudomonas dinitrificanus*	55% (54/97)	69% (67/97)

TABLE 3. ***Continued***

BES	BES ID No.	Best match ID No.	Gene description	Organism	AA identity	AA similarity
3D07-F	AQ242125	D31783	carboxynorspermidine decarboxylase	*Vibrio alginolyticus*	54% (41/75)	77% (58/75)
3I23-F	AQ242117	U29587	moaE gene phopholipid synthase	*Escherichia coli*	50% (56/112)	67% (75/112)
3L01-R	AQ242227	AF033856	Phosphoglucomutase	*Agrobacterium tumefaciens*	81% (84/104)	88% (92/104)
3L23-F	AQ242157	AF031709	Cystathionine gamma-lyase-like protein	*Stenotrophomonas maltophilia*	43% (40/92)	63% (58/92)
3O07-F	AQ242153	Z22636	exoN protein, polysaccharide biosynthesis	*Rhizobium meliloti*	67% (116/171)	80% (137/171)
3PO2-R	AQ242254	Q08113	NIFR3 protein	*Rhodobacter capsulatus*	59% (104/176)	70% (124/176)
3P09-R	AQ242233	U51905	Phosphoserine aminotransferase	*Methanosarcina*	58% (58/99)	69% (69/99)
3P15-R	AQ242156	AF031709	Cystathionine gamma-lyase	*Stenotrophomonas maltophilia*	45% (74/163)	66% (107/163)
4A02-R	AQ242192	M12893	2-isopropylmalate synthase	*Saccharomyces cerevisiae*	56% (18/32)	65% (21/32)
Phage remnant						
4M07-F	AQ242163	AE000751	Phage SPO1 DNA polymerase	*Aquifex aeolicus*	43% (21/49)	73% (36/49)
Transcriptional regulation/post translational regulation						
1D03-R	AQ242217	AE000728	Transcriptional regulator (FurR Family)	*Aquifex aeolicus*	34% (18/52)	51% (27/52)
1G24-R	AQ242146	B64189	Glycine cleavage transcriptional activator	*Haemophilas influenza*	27% (28/101)	49% (50/101)
2A18-F	AQ242244	D90899	Methionine sulfoxide reductase	*Synechocystis sp.*	41% (23/56)	51% (29/56)
2E11-F	AQ242133	U32867	PCA operon transcriptional activator	*Agrobacterium tumefaciens*	77% (127/165)	89% (147/165)
3F13-R	AQ242204	AF017747	Heavy metal dependent transcription regulator	*Proteus mirabilis*	51% (41/81)	69% (56/81)
4C09-R	AQ242150	X82174	DNA binding protein	*Bacillus subtilis*	27% (33/122)	52% (64/122)
4H18-R	AQ242211	AF001171	PehR	*Ralstonia solanacearum*	47% (43/92)	65% (60/92)
Transport/binding proteins						
1A09-R	AQ242128	AF012537	Arginine and ornithine binding protein	*Pseudomonas aeruginosa*	53% (47/88)	71% (63/88)
2IO6-R	AQ242253	P04983	Ribose transport ATP binding protein	*Escherichia coli*	44% (106/241)	64% (154/241)
2K05-F	AQ242167	AE00283	L-arabinose transport ATP-binding protein	*Escherichia coli*	41% (31/74)	71% (53/74)
2L16-F	AQ242230	D90769	ABC transporter permease protein	*Escherichia coli*	32% (54/156)	56% (93/166)
2A10-F	AQ242238	U42410	Heavy-metal transporting P-type ATPase	*Proteus mirabilis*	77% (78/100)	84% (85/101)
3F13-F	AQ242203	S47788	D-xylose transport ATP binding protein	*Escherichia coli*	46% (101/221)	63% (140/221)
3L08-F	AQ242135	AE000185	ABC transporter ATP-binding protein	*Escherichia coli*	50% (46/92)	71% (65/92)
4P23-F	AQ242190	A32349	ATP-binding protein, chvD	*Agrobacterium tumefaciens*	73% (60/82)	89% (73/82)
Pathogenic/cell defense						
1E13-R	AQ242166	AF043471	Hemolysin	*Escherichia coli*	46% (58/127)	56% (71/127)
2C07-F	AQ242119	Z98268	TlyA, cytotoxin /haemolysin homologue	*Mycobacterium tuberculosis*	51% (74/145)	65% (94/145)
2M03-R	AQ242202	D88802	Chloramphenicol resistance protein	*S. lividans*	35% (29/82)	57% (47/82)
4C09-F	AQ242149	AB000617	LmrB, drug-export protein	*Bacillus subtilis*	30% (25/83)	50% (42/83)
3PO2-F	AQ242247	P37594	Methyl viologen resistance protein	*Salmonella typhimurium*	35% (30/84)	64% (54/84)
4H18-F	AQ242210	B47301	VirB4 homolog	*Bordetella pertussis*	50% (58/117)	68% (79/117)
4O24-R	AQ242113	X68232	Multidrug exporter protein	*Klebsiella pneumoniae*	63% (53/84)	79% (67/84)
4P12-F	AQ242234	AE000538	Cell binding factor	*Helicobacter pylori*	43% (35/80)	66% (53/80)

In our study, we estimated contig sizes from the fingerprint database, whereas Jurnas-Bilak et al. (1998) used estimates from PFGE of whole linearized replicons. The two smaller hybridization contigs obtained in our study may represent the plasmids that, based on the FPC data, appear to be derived from recombination events occurring within or between the major replicons. One of the plasmids (hybridization contig 4) appeared to be a chimera containing segments from both of the major replicons. Further, when the clones from the smaller hybridization contigs were mapped to the FPC database, their positions were located within the larger contigs rather than at the ends of the physical maps. The sizes of the small hybridization contigs based on fingerprints of clones in the minimum tile were estimated to be 175 and 165 kb for contigs 3 and 4, respectively. In one strain of *0. anthropi,* Jumas-Bilak et al. (1998) found two distinct plasmids, each about 150 kb in size.

Gene survey by BAC end sequencing

Sequencing of BAC ends using both forward and reverse primers was performed on the 69 clones from the reduced tiles, generating a total of 138 sequences. We estimate that a BAC end sequence (BES) was obtained on average every 35 kb throughout the *0. anthropi* genome. The resulting file was then compared with the GenBank database using the BLASTX algorithm and the BatchBlast script (Baylor College of Medicine). The results were sorted using scripts developed at the CUGI. Overall, I 10 of 138 BES resulted in hits during the database search. Of the I 10 BES, 77 hits gave an E-value < 0.001 as the top score. A summary of these hits describing best match accession number, gene type, organism, amino acid identity, and amino acid similarity is shown in Table 3.

All of the BES identified and listed in Table 3 are sorted by function of their putative protein product. As a result of sorting the sequences by function, the largest group of sequences (30%) was associated with metabolism and biosynthesis. Sequences associated with transport/binding proteins comprised 10%, and sequences associated with pathogenic and cell defense responses also represented 10%. Sequences associated with DNA and protein regulation represented 9%. Another 9% of the sequences were associated with cell division and DNA replication/repair. Cell signaling sequences represented 5%, and cell structure/membrane sequences were 4%. Only one phage remnant sequence was found, giving 1% of the total. The number of unclassified sequences with GenBank hits totaled 21%, and the number of unclassified sequences with no database hits represented 44%. Unidentified and unclassified sequences may represent a subset of those proteins that are specific to *0. anthropi* and Proteobacteria in general. The *E. coli* genome, which is 4.6 Mbp in size, contains 4288 genes (Blattner et al., 1997). As *0. anthropi* has a comparable genome size, we estimate that the limited gene survey presented in this study allowed identification and physical placement of approximately 3% of the genes in the genome of *0. anthropi.*

DISCUSSION

We have successfully constructed a BAC library for the bacterial genome of *0. anthropi,* an important bacterium containing genes involved in opportunistic pathogenesis, biodegradation, and novel enzymatic reactions. The library has an average insert size of 112 kb and provides an estimated 70 genome equivalents. Further, >65% of the clones are > 100 kb. The library is an ideal substrate for a number of genomics- based applications, including physical mapping, DNA sequencing, and the cloning of important genes and pathways involved in biodegradation.

Using this new BAC library, we successfully constructed a physical map of the *0. anthropi* genome by combining two distinct methods of contig assembly, hybridization and fingerprinting. This approach was accomplished by using the hybridization-based data as the foundation and verifying the resulting contigs by comparison to the FPC data. The hybridization-based data were considered to be quite robust because bacterial genomes are relatively void of repetitive sequences. Repetitive DNA in eukaryotic genomes can confound a hybridization without replacement strategy by giving false physical associations among unrelated clones. Furthermore, it is difficult to develop extensive contigs from fingerprint data without the use of molecular markers. In newly characterized bacterial genomes, there is a paucity of molecular markers available for aligning contigs. In the case of *0. anthropi,* there were essentially no molecular markers avail-

able. It is noteworthy that although it was difficult to completely reconstruct the replicons using the fingerprint data, the physical associations accurately verified the data obtained by hybridization.

Using the 69 clones from reduced tiles, BAC end sequencing was performed to provide a gene survey. Based on an estimated genome size of 4.8 Mb (Jurnas-Bilak et al., 1998), DNA sequences should have been obtained on average every 35 kb throughout the genome. Database searches with these sequences resulted in a majority (92%) of the 77 strong hits (E-value < 0.001), showing similarity to bacterial genes from a diversity of species. Some of the strongest hits were from closely related proteobacterial genomes, such as *Agrobacterium tumefaciens* and various species of *Rhizobium.* Although virtually nothing was previously known about gene structure in *0. anthropi*, the 16S ribosomal RNA genes from two strains have been sequenced (Yanagi and Yarnasoto, 1993), resulting in the phylogenetic placement of *0. anthropi* as follows: Ochrobactrum, Rhizobiaceae, alpha subdivision, Proteobacteria, Eubacteria. The Proteobacteria are the most physiologically diverse of all the bacteria, and at the present time, 16 genomes within this group are being sequenced or have been sequenced *(http://www.tigrorgltdblmdblmdb.html).* However, no members of the Rhizobiaceae group, which includes *0. anthropi* and A. *tumefaciens,* are being sequenced by public research groups at this time. Interestingly, from the 77 hits on the GenBank database with the BES data, 14% were from Rhizobiaceae genomes.

We have effectively demonstrated the utility of using BAC clones in bacterial genomics research. These techniques have allowed the construction of a detailed physical map combined with a gene survey of a bacterial genome for which little information currently existed. This framework approach can provide essential information related to genome structure and organization that can serve as a guide prior to a full-scale sequencing effort. Furthermore, if resources for a full-scale genome sequencing effort were not available, this system would still provide the necessary tools needed for the cloning and analysis of important genes and whole operons or genetic pathways.

ACKNOWLEDGMENTS

The research was supported by state and Hatch funds allocated to the South Carolina Agricultural Experiment Station, funds from the Coker Endowment to R.A.W., and by NSF MRI grant 9724557 to R.A.W. and R.A.D. This is technical contribution No. 4568 of the South Carolina Agricultural Experiment Station.

REFERENCES

ALM, E.W., OERTHER, D.B., LARSEN, N., STAHL, D.A., and RASKIN, L. (1996). The Oligonucteotide Probe Data- base. Appl Environ Microbiol **62,** 3557-3559.

ALNOR, D., FRIMODT-MOLLER, N., ESPERSEN, F., and FREDERIKSEN, W. (1994). Infections with the unusual human pathogens *Agrobacterium* species and *Ochrobactrum anthropi.* Clin Infect Dis 18, 914-920.

ALTSCHUL, S.F., MADDEN, T.L., SCHAFFER, A.A., ZHANG, J., ZHANG, Z., MILLER, W., and LIPMAN, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25,3389-3402.

ARRATIA, R., LANDER, E.S., TAVARE, S., and WATERMAN, M.S. (1991). Genomic mapping by anchoring ran- dom clones: A mathematical analysis. Genomics 11, 806-827.

ASANO, Y., KATO, Y., YAMADA, A., and KONDO, K. (1992). Structural similarity Of D-aminopeptidase car- boxypeptidase DD and beta-lactamases. Biochemistry **31,** 2316-2328.

BLATTNER, F.R., PLUNKET, G., HI, BLOCK, C.A., PERNA, N.T., BURLAND, V., RILEY, M., et al. (1997). The complete genomic sequence of *Escherichia coli* K-12. Science 277,1453-1462.

BROSCH, R., GORDON, S.V., BILLAULT, A., GARNIER, T., EIGLMEIER, K., SORAVITO, C., et al. (1998). Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect Immun **66,** 2221-2229.

CHANG, H.J., CHRISTENSON, J.C., PAVIA, A.T., BOBRIN, B.D., BIAND, L.A., CARSON, L.A., et al. (1996). *Ochrobactrum anthropi* meningitis in pediatric pericardial allograft transplant recipients. J Infect Dis **173,** 656-660.

DEWAR, K., SABBAGH, L., CARDINAL, G., VEILLEUX, F., SANSCHAGRIN, F., BIRREN, B., and LEVESQUE,

R.C. (1998). *Pseudomonas acruginosa* PAOI bacterial artificial chromosomes: Strategies for mapping, screening, and sequencing 100 kb l@i of the 5.9 Mb genome. Microb Comp Genomics 3, 105-117.

DIAZ-PEP,EZ, S.V., ALATRISTE-MONDPAGON, F., HERNANDEZ, R., BIPREN, B., and GUNSALUS, R.P. (1997). Bacterial artificial chromosome library as a tool for physical mapping of the archaen *Methanosarcina ther- mophila TM-*1. Miemb Comp Genomics 2, 275-286.

DORSCH, M., and STACKEBRANDT, E. (1992). Some modifications in the procedure of direct sequencing of PCR amplified 16S RDNA. J Microbiol Methods 16,271-279.

FREEDMAN, D.L., and VERCE, M.F. (1997). Ethene- and ethane-promcled biodegradation of vinyl chloride. In: *In Situ and On-Site Rioremediation,* Volume 3. B. C. Alleman and A. Leeson, eds. (Battelle Press, Columbus, OH).

HOLMES, B., POPOFF, M., KIREDJIAN, M., and KERSTTERS, K. (1988). *Ochrotactram anthropi* gen. nov., sp. nov. from human clinical specimens and previously known as group Vd. Int J Systematic Bacteriol 38, 406-416.

IRAZABAL, N., MARIN, I., and AMILS, R. (1997). Genomic organization of the acidophilic chemolithoautotrophic *bacterium Thiobacillusferrooxida@* ATCC 21834. J Bacteriol 179, 1946-1950.

JUMAS-BTLAK, E., MICHAUX-CHARACHON, S., BOURG, G., RAMUZ, M., and ALLARDET-SERVENT, A. (1998). Unconventional genorr@c orgar@zation in the alpha subgroup ofthe Protcobacteria. J Bacteriol 180,2749-2755.

KANE, M.D., POULSEN, L.K., and STAHL, D,A. (1993). Monitoring the enrichment and isolation of sulfate-reduc- ing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S RRNA se- quences. Appl Environ Microbiol 59, 682-686.

KUNDIG, C., HENNECKE, H., and GOTTFERT, M. (1993). Correlated physical and genetic map of the *Bradvhizh,-biumjaponicum I* 10 genome. J Bacteriol 175, 613-622.

LAURA, D., DE SOCIO, G., FRASSANITO, R., and ROTILIO, D. (1996). Effects of atrazine on *Ochrobactmm an- thropi* membrane fatty acids. Appl Environ Micmbiol 62, 2644-2046.

MAIDAK, B.L., COLE, JR., PARKER, C.T., JR., GARRITY, G.M., LARSEN, N., Li B., et al. (1999). A new ver- sion of the RDP (Ribosomal Database Project). Nucleic Acids Res 27, 171-173.

MARRA, M.A., KUCABA, T.A., DIETRICH, N.L., GREEN, E.D., BOWNSTEIN, B., WILSON, R.K., et al. (1997). High throughput fingerprint analysis of luge-insert clones. Genome Res 7, 1072 1084.

MURASE, H., SUGIHAPA, A., MORO, T., SHIMADA, Y., and TOMINAGA, Y. (1991). Purification and properties of cuboxylesterase from *Ochrobactmm anthropi.* Agric Biol Chem 55, 2579-2584.

PRADE, R.A., GRIFFITH, J., KOCHUT, K., ARNOLD, J., and TIMBERLAKE, WE. (1997). In vitro reconstruction *of Aspegillis (= Eme,icelia) nidulans* genome. Proc Natl Acad Sci USA 94, 14564-14569.

RIETUMAN, H., BIRREN, B., and GNIRKE, A. (1997). Preparation, manipulation, and mapping of HMW DNA. In *Genome Analysis, Volume 1, Analvzing DNA.* B. Birren, ED., Green, S. Klapholz, R.M. Myers, and J. Roskams, eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), Chapter 2,

RIJPENS, N.P., JANNES, G., VAN ASEROECK, M., ROSSAU, R., and HERMAN, L.M. (1996). Direct detection of *B,ucelia* spp. in raw milk by PCR and reverse hybridization with 16S-23S RRNA spacer probes. Appl Environ Mi- crobiol 62, 1683-1688.

ROMLING, U., GROTHUES, D., HEUER, T., and TUMMLER, B. (1992). Physical genorne analysis ofbacteria. Elec-trophoresis 13, 626-63 1.

SAMBROOK, J., FRISCH, E.F., and MANTATUS, T. (1989). *Molecular Cloning: A Laboratory Manual,* 2nd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

SASINOWSKA, H., and SASINOWSKI, M. (1999). An algorithm for the assembly of robust physical maps based on a combination of multi-level hybridization data and fingerprinting data. Computers Chem 23, 251-262.

SHIZUYA, H., BIRREN, B., KIM, U.J., MANCINO, V., SLEPAK, T., TACHURI, Y., and SIMON, M. (1992). Cloning and stable maintenace of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA 89, 8794-8797.

WOO, S.S., JIANG, J., GILL, B.S., PATERSON, A.H., and WING, R.A. (1994). Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor.* Nucleic Acids Res 22, 4922-4931.

YANAGI, M., and YAMASATO, K. (1993). Pylogenetic analysis of the family Rhizobiaccae and related bacteria by sequencing of 16S RRNA gene using PCR and DNA sequencing. FEMS Microbiol Lett 107, 115-120.

Address reprint requests to:
Jeffrey P. Tomkins, Assistant Professor
Clemson University Genomics Institute
Room 100 Jordan Hall

Clemson, SC 29634
E-mail: jtmkns@clemson.edu