

Genetic Variation at the O-Antigen Biosynthetic Locus in *Pseudomonas aeruginosa*

Christopher K. Raymond,* Elizabeth H. Sims, Arnold Kas, David H. Spencer, Tanya V. Kutuyavin,
Richard G. Ivey, Yang Zhou, Rajinder Kaul, James B. Clendenning,
and Maynard V. Olson

*University of Washington Genome Center, Department of Medicine,
University of Washington, Seattle, Washington 98195*

Received 11 January 2002/Accepted 3 April 2002

The outer carbohydrate layer, or O antigen, of *Pseudomonas aeruginosa* varies markedly in different isolates of these bacteria, and at least 20 distinct O-antigen serotypes have been described. Previous studies have indicated that the major enzymes responsible for O-antigen synthesis are encoded in a cluster of genes that occupy a common genetic locus. We used targeted yeast recombinational cloning to isolate this locus from the 20 internationally recognized serotype strains. DNA sequencing of these isolated segments revealed that at least 11 highly divergent gene clusters occupy this region. Homology searches of the encoded protein products indicated that these gene clusters are likely to direct O-antigen biosynthesis. The O15 serotype strains lack functional gene clusters in the region analyzed, suggesting that O-antigen biosynthesis genes for this serotype are harbored in a different portion of the genome. The overall pattern underscores the plasticity of the *P. aeruginosa* genome, in which a specific site in a well-conserved genomic region can be occupied by any of numerous islands of functionally related DNA with diverse sequences.

Microbes occupy virtually every habitable niche in the biosphere, highlighting the underlying capacity for genetic adaptability in these organisms. *Pseudomonas aeruginosa* in particular is notable for its ability to thrive in diverse habitats (44). Consistent with a genetic basis for this environmental adaptability, characterization of intraspecies differences between strains of this organism has revealed extensive variation both in the gross overall structural organization of the genome (41) and in sequence variation of specific genes (20; M. V. Olson, A. Kas, and D. H. Spencer, unpublished data). Of particular interest is the observation that a rather large island (~50 kbp) of genomic sequence, containing dozens of potential genes, is substituted in different *P. aeruginosa* isolates (25). This suggests that modular blocks of genes that are swapped between different strains is one mechanism that mediates diversity among these different bacterial lineages.

The description of the full genomic sequence of *P. aeruginosa* strain PAO1 (46) provides a reference sequence map with which to initiate studies of genetic diversity at the whole-genome level. We employed whole-genome shotgun sequence analysis of three different *P. aeruginosa* strains as a means to detect both local and more global differences among strains isolated from different sources (Olson et al., unpublished). This study revealed a region with prominent differences between strains that proved to encode the O-antigen biosynthesis genes involved in the creation and assembly of the bacterial outer carbohydrate lipopolysaccharide coat (reviewed in reference 39). In various *P. aeruginosa* strains, the outer carbohydrate polymer, referred to as the B band, is composed of

chemically diverse chains of repeating polysaccharides. The studies to date have indicated that a major set of enzymes responsible for O-antigen synthesis and assembly are encoded in single, large gene clusters (2, 5, 12). The genes within these operons mediate chemical modification of various sugars, sequential assembly of these sugars into polysaccharide subunits, translocation of these polysaccharide subunits from the cytoplasm to the periplasm, and ligation of subunits to form the repetitive polysaccharide chains of the O antigen (39).

Twenty unique serotype strains of *P. aeruginosa*, the so-called International Antigenic Typing System (IATS) strains (28, 29), have been characterized extensively, although it seems likely that additional serogroups are present in the natural environment (45). Reference strain PAO1 is an O5 serotype strain. The B-band biosynthetic clusters from strains of serotype O5 and two other serotypes, O6 and O11, have been characterized at the DNA sequence level (2, 5, 12). Many of the genes in the O5, O6, and O11 B-band biosynthetic clusters have been assigned functional roles in B-band biosynthesis based on either direct biochemical characterization or sequence similarity to genes whose functions are known (7, 8, 11, 39, 52, 53). The data have revealed two features. First, there is little DNA sequence conservation between B-band biosynthetic genes. Second, all three sets of B-band clusters are in the same highly conserved region of the *P. aeruginosa* genome.

These observations concerning the O5, O6, and O11 biosynthetic gene clusters, which revealed the presence of functionally and genetically diverse DNA sequences nestled in a conserved locus in the genome, motivated us to examine the DNA sequences at this locus for the 20 IATS *P. aeruginosa* strains. We used yeast recombinational cloning technology to clone this region (3, 23, 37, 51). This technology appears to be well suited to targeted cloning of any region of genomic DNA of microbes provided that the flanking sequences of the region

* Corresponding author. Mailing address: University of Washington Genome Center, Box 352145, University of Washington, Seattle, WA 98195. Phone: (206) 221-6954. Fax: (206) 685-7344. E-mail: craymond@u.washington.edu.

TABLE 1. PCR primers used in this study

Primer	Comment
Primers used to construct targeting plasmids	
AATTATAATTATTTTTATAGCACGTGATGAAAAGGACCGCACGAG AAGGACGCAATGAAAGAAGT.....	Primer 1, amplification of <i>himD/ihfB</i> region targeting sequence
CGCACATTTCCCGAAAAGTGCCACCTGACGTGCCGGGCGCGGC ACGAACTTGCCGTCGAGGCG.....	Primer 2, amplification of <i>himD/ihfB</i> region targeting sequence
ACTTCGTATAGCATACATTATACGAAGTTATATTCGATGCGCCCA GCAGTCCAGCGATGTCCAC.....	Primer 1, amplification of <i>wbpM</i> region targeting sequence
GCTTTTTTAAAAGATTTTCAAATCCATATATAAACATAGTGGTTG AGCATGCTGCTGATCGGCG.....	Primer 2, amplification of <i>wbpM</i> region targeting sequence
ACTTCGTATAGCATACATTATACGAAGTTATATTCGATGCCCTTCT CGCTCTATGGCGAACGCGT.....	Primer 1, amplification of <i>tyrB</i> region targeting sequence
GCCTCTCCCGCGCGTTGGCCGATTCATTAATGCAGGCCCGGGCG ATTA AAAAGGGGAAGTCAAGCCTC.....	Primer 2, amplification of <i>tyrB</i> region targeting sequence
Primers used to detect recombinant plasmids	
TATAGCACGTGATGAAAAGGACCGC.....	Vector primer 1
CGAGCTCATCGCTAATAACTTCGTA.....	Vector primer 2
CGCAACTCCTTGCCCGGCTTGAAGT.....	Insert primer 1
GCGTTCCCTGGTGTTCAACTACTGG.....	Insert primer 2
ACCGACTGAGCTATCGCGGAACAGC.....	Insert primer to detect ATCC O15 clone
Primers used to cross-check amplified genomic fragments with cloned sequences	
TGCTCTTTGGCTTCTTGATTCTTG.....	Primer 1, amplifies O1-specific product
TCTACACCGCCAGAACTACCTAGCT.....	Primer 2, amplifies O1-specific product
GCGGCTTTTATCAACCGTGTCCGA.....	Primer 1, amplifies product from O2, O5, O16, O18, and O20
TCCCTTGGCACACTGGAAGGACAT.....	Primer 2, amplifies product from O2, O5, O16, O18, and O20
GCGTCGTTGTTTCAGTTTGGACGTG.....	Primer 1, amplifies product from O3 and O15
TCTGGAAAACCTGAGCAGCCGTC.....	Primer 2, amplifies product from O3 and O15
GCTAATAACGGAAGGACCTTGAAT.....	Primer 1, amplifies O4-specific product
TAAAAGCTCGGCGTAACGCTTATG.....	Primer 2, amplifies O4-specific product
ATTGGTAGTGTACACGAGTGCA.....	Primer 1, amplifies O6-specific product
CGAATTAGCTTGTCTTTCAGGAAAG.....	Primer 2, amplifies O6-specific product
GGAATGTCGCTCTCGTTTCAAGTG.....	Primer 1, amplifies product from O7 and O8
CTAGTATTATCAACTGCTGTAC.....	Primer 2, amplifies product from O7 and O8
GGAACATTGGAATCAAGAGGTTATG.....	Primer 1, amplifies O9-specific product
CCACTAAATACCAGGCATACAATA.....	Primer 2, amplifies O9-specific product
CAGCAGGGAATATCGCTTGAACAGT.....	Primer 1, amplifies product from O10 and O19
CTATAATGCATTAGCGACTACCG.....	Primer 2, amplifies product from O10 and O19
TACTTACCCATAGCTAGCGCTCTA.....	Primer 1, amplifies product from O11 and O17
TTCTCTCAACTTAACCGTGGCC.....	Primer 2, amplifies product from O11 and O17
ACTATTAGTCGATGATCCTTGTG.....	Primer 1, amplifies product from O13 and O14
AGCAGTTAGCAAATTTCACTCTCCAGCC.....	Primer 2, amplifies product from O13 and O14

are known and well conserved. The DNA sequences of the 20 isolated segments revealed rich genetic diversity within the context of the conserved flanking sequences.

MATERIALS AND METHODS

Strains, media, and reagents. The yeast strain used in this study was CRY1-2 (*MAT α ura3 Δ cyh2^R* [37]). The *Escherichia coli* host strain was DH10B (17). The 20 IATS *P. aeruginosa* strains were a gift from Stephen Lory. The same set of strains was also purchased from the American Type Culture Collection (ATCC) (Rockville, Md.). Yeast transformants harboring recombinant plasmids were selected on standard uracil-deficient media (42) containing 2.5 μ g of cycloheximide (Spectrum Chemicals, Gardena, Calif.) per ml. Chloramphenicol-resistant *E. coli* was selected on Luria-Bertani medium containing 6 μ g of chloramphenicol per ml.

Recombinational cloning. Recombinational cloning vectors were prepared by yeast recombination methods (34, 37, 38). Targeting segments were amplified from *P. aeruginosa* strain PAO1 genomic DNA by using tailed primers (Table 1) that create overlaps with standard vector sequences. Targeting plasmids were assembled in a single step by recombining a yeast-*E. coli* shuttle vector (37) (Fig. 1), the two PCR-amplified targeting segments, and a central fragment that carries the yeast wild-type *CYH2* gene and an Amp-ori stuffer fragment (Fig. 1) (37). Genomic DNA was prepared from *P. aeruginosa* strains by the method of Liang et al. (25). For recombinational cloning, the genomic DNA was sheared by 40 passages through a 26.5-gauge needle. One microgram of genomic DNA and 100 ng of linearized recombinational cloning plasmid were cotransformed into

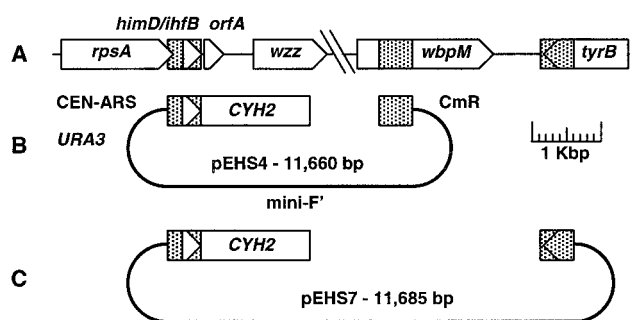


FIG. 1. Plasmids used for yeast recombinational cloning of the O-antigen region. (A) Genomic organization of the conserved *P. aeruginosa* genes at the 5' and 3' ends of the O-antigen operons. ORFs, which are drawn to scale, are shown as arrows oriented in the directions in which they are transcribed. Five-hundred-base-pair targeting elements used for cloning are shown as stippled boxes. (B) Yeast-*E. coli* shuttle plasmid used for recombinational cloning. Yeast *CEN-ARS* and *URA3* sequences facilitate plasmid segregation-replication and selection, respectively, in yeast. Plasmid maintenance in *E. coli* relies on the single-copy mini F' origin (43), and chloramphenicol resistance (*CmR*) is used for plasmid selection. (C) Recombinational cloning plasmid used to isolate the *himD/ihfB*-to-*tyrB* region from the ATCC O15 strain.

TABLE 2. GenBank accession numbers for the *himD/ihfB*-to-*wbpM* DNA sequences of the 20 IATS *P. aeruginosa* strains

Clone	Accession no.	Comment
O1	AC104719	
O2	AC104731	
O3	AC104733	
O4	AC104734	
O5	AC104735	
O6	AC104736	
O7	AC104737	Derived from ATCC O7 strain
O8	AC104738	
O9	AC104739	
O10	AC104720	
O11	AC104721	
O12	AC104722	Derived from ATCC O12 strain
O13	AC104723	
O14	AC104724	
O15 ATCC	AC104725	Derived from ATCC O15 strain
O15 Lory	AC104726	
O16	AC104727	
O17	AC104728	
O18	AC104729	
O19	AC104730	
O20	AC104732	

lithium acetate-treated yeast cells (15). Typically, 5 to 20 yeast colonies were recovered. These colonies were screened by whole-cell PCR (27) by using the vector and insert primers described in Table 1. The percentage of recombinant clones that contained both expected vector-insert junctions varied from 10 to 80% of the colonies screened. Plasmid DNA was transferred to *E. coli* by the method of Hoffman and Winston (19).

For DNA sequencing, shotgun libraries were prepared from recombinant clones in Bluescript KS+ (Stratagene, La Jolla, Calif.), and standard shotgun sequencing methods were used to generate a Q20 base coverage of more than eightfold per clone. As many as three rounds of AUTOFINISH (16) were used to create a finished sequence, and final advanced finishing and fingerprint-to-sequence comparisons were performed by well-established methods. DNA sequences were analyzed by using several publicly available software packages. Open reading frames (ORFs) were identified by using GENEMARK (30), protein families were classified by using PFAM (1), and multiple alignments were created with MULTALIN (6). Potential transmembrane domains were characterized by using the DAS program server (9).

***P. aeruginosa* O-antigen serotyping.** Strains were serotyped by using commercially available antibodies to serotypes O1 through O17 (ERFA, Westmount, Quebec, Canada). Strains were streaked onto Luria-Bertani agar, and colonies were resuspended in phosphate-buffered saline. Ten microliters of cell suspension was mixed with 10 μ l of antibody and incubated at room temperature for 5 to 10 min. Serotype-positive clones exhibited pronounced aggregation and settling, whereas control cells remained uniformly suspended in solution.

Nucleotide sequence accession numbers. The DNA sequences obtained for the 20 IATS strains have been deposited in the GenBank database under the accession numbers shown in Table 2. These sequences are also available at a supplemental website (<http://www.genome.washington.edu/uwgc/O-Antigen/>).

RESULTS

Yeast recombinational cloning. The analysis of whole-genome sequences from four different *P. aeruginosa* strains revealed sharp, well-defined boundaries between conserved and divergent sequences on either side of the O-antigen B-band biosynthetic gene cluster (Olson et al., unpublished). This finding was consistent with the previously published reports showing the conserved *himD/ihfB* gene upstream and the C-terminal coding region of the *wbpM* gene downstream of the divergent B-band gene cluster sequences (2, 5, 12, 39). Moreover, Southern blot analysis has shown that *wbpM* (presumably the C-terminal coding sequence) is present in all 20 IATS

strains (5). We therefore elected to target our cloning efforts to the region between *himD/ihfB* and the conserved, C-terminal coding region of *wbpM*. A yeast-*E. coli* shuttle vector was constructed that contained 500-bp blocks of the *himD/ihfB* sequence on one side and the *wbpM* sequence on the other side of the recombinational cloning site (Fig. 1). The cloning vector also contains the yeast wild-type *CYH2* gene. This gene serves as a counterselection agent against nonrecombinant plasmids when transformed yeast cells are plated in the presence of the drug cycloheximide (37). Recombination between the targeting plasmid and genomic DNA leads to loss of *CYH2* and hence to cycloheximide resistance. To clone the *himD/ihfB*-to-*wbpM* region from a particular IATS strain, lithium acetate-competent yeast cells (15) were cotransformed with linearized targeting plasmid and total *P. aeruginosa* genomic DNA. Individual yeast transformants were screened by PCR for the presence of genomic inserts. Recombinant plasmids were shuttled into *E. coli*, analyzed by restriction enzyme analysis, and used to prepare shotgun sequencing libraries. Inserts whose sizes ranged from ~14 to ~25 kbp were observed (Fig. 2).

Several quality control measures were used to ensure the accuracy of the data. The 20 IATS strains were originally obtained as a gift from Stephen Lory, and all recombinant clones were obtained from this collection except as noted below. Later, the same collection was purchased from the ATCC. We used an antibody agglutination serotype assay, available for serogroups O1 through O17, to confirm the identity of each strain. The serotypes of most strains were as expected; the only exception was the Lory O7 strain, which failed to agglutinate. We therefore isolated and sequenced the ATCC O7 strain and deposited the sequence of this strain in the GenBank database. Cloning of O12 failed initially, but a repeat effort with the ATCC O12 clone was successful; this clone was sequenced, and the sequence was deposited in the GenBank database. Following DNA sequence determination, we designed PCR primers to amplify a unique segment of each O-antigen DNA sequence (Table 1). These primers were used to amplify and then sequence the corresponding regions from genomic DNA (the O12 clone was inadvertently omitted from this analysis). In all cases the amplified genomic sequences matched the sequences of the isolated clones, except as noted for O15 below. These results demonstrated that the Lory strain collection and the ATCC collection were identical in terms of their O-antigen sequences. Finally, we observed that the sequences which we determined for O5, O6, O11, and O17 are identical to the published sequences (2, 5, 10, 12). These cross-checking efforts made us confident that our data set is an accurate reflection of the underlying biological and genetic diversity.

The *P. aeruginosa* O15 strains proved to be somewhat paradoxical. Both the Lory strain and the ATCC strain were serotyped appropriately with the monoclonal anti-O15 antibody. Cloning and sequencing of the *himD/ihfB*-to-*wbpM* region from the Lory strain yielded a 20-kbp segment that shared sequence identity with O3 (Fig. 2). However, the O15 sequence was interrupted at three locations with insertion sequence elements. PCR primers designed by using an uninterrupted segment of O3/O15 amplified the expected products with the anticipated sequence from the Lory O15 strain; however, they failed to amplify a product from the ATCC O15 strain. Moreover, PCR analysis suggested that the ATCC O15 strain did

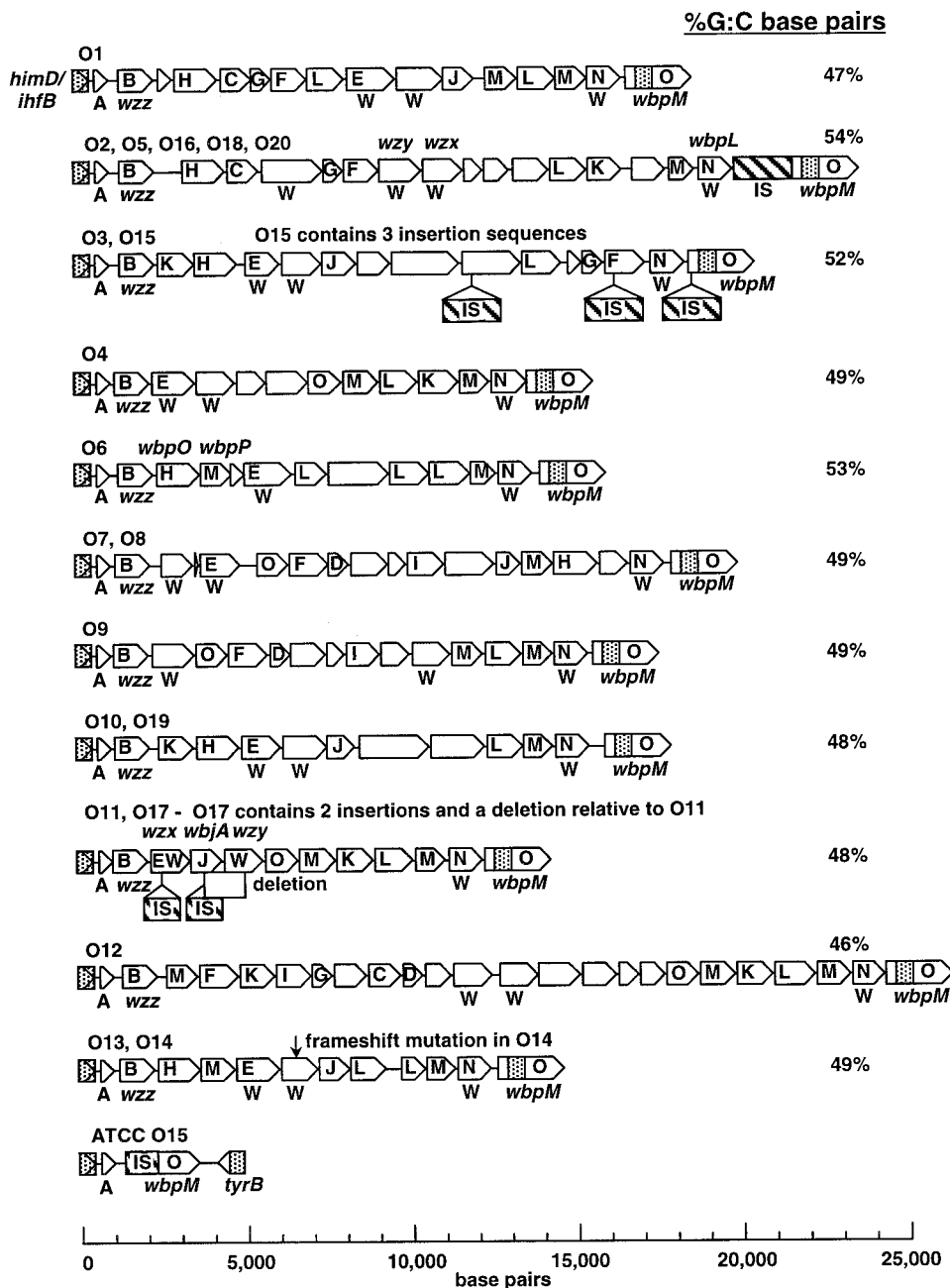


FIG. 2. Eleven groups of *P. aeruginosa* O-antigen biosynthetic gene clusters. Genes within the DNA sequences were identified by the GENEMARK algorithm (30) and are represented by arrows drawn to scale. Protein families, identified in queries to the PFAM database (1) and represented three or more times in the overall data set, are represented by single-letter designations. W indicates an ORF with several potential membrane-spanning domains. A detailed description of the families is provided in Table 3. Specific genes described previously in gene clusters O5, O6, and O11 are shown above the clusters. The stippled boxes represent the targeting elements used in recombinational cloning. The C-terminal coding region of *wbpM*, shown as an open arrow, is presumed to extend rightward of the cloned region, as shown. A denotes *orfA*, as described in the text. The G+C content of each cluster is shown.

not possess the *wbpM* segment used to target recombinational cloning. We created a second recombinational cloning plasmid that carried the same upstream *himD/ihfB* targeting region but a downstream 500-bp targeting element from the *tyrB* gene that is 3' to the B-band gene clusters (Fig. 1). Using this vector, we cloned and then sequenced a short segment from the ATCC O15 genome that proved to be devoid of a B-band biosynthetic gene cluster (Fig. 2).

Eleven groups of DNA sequences. Our goal in this investigation was to study the diversity, and hence the relationships, among DNA sequences found in the *himD/ihfB*-to-*wbpM* region in the 20 well-defined IATS strains. To this end, we found 11 groups of gene clusters that are highly divergent from one another at the DNA sequence level (Fig. 2). Within each group we observed a high degree of sequence conservation. The largest group contains strains of serotypes O2, O5, O16, O18,

TABLE 3. PFAM-annotated protein families with at least three members in the *P. aeruginosa* O-antigen biosynthesis operons

Protein family	PFAM description	Reference(s)
B	wzz, chain length determinant protein	14
C	Oxidoreductase family, NAD-binding Rossmann fold	21
D	Cytidyltransferase	31
E	Polysaccharide biosynthesis protein	50
F	DegT/DnrJ/EryC1/StrS family	47
G	Bacterial transferase hexapeptide (four repeats)	35
H	UDP-glucose/GDP-mannose dehydrogenase family	40, 52, 53
I	NeuB family	32
J	Glycosyltransferase group 2	11
K	UDP-N-acetylglucosamine 2-epimerase	48
L	Glycosyltransferase group 1	2
M	NAD-dependent epimerase/dehydratase family	8, 49
N	Glycosyltransferase group 4	24
O	Polysaccharide biosynthesis protein	7, 26

and O20. This cluster was not unanticipated as these strains are serologically related and have similar chemical structures in their O-antigen polysaccharides (33). All other groups have only two members. These groups are the groups that include O3 and O15 (Lory), O7 and O8, O10 and O19, O11 and O17, and O13 and O14. The cloned regions exhibited a pronounced shift in base composition and had G+C contents ranging from 46 to 55% (Fig. 2). This is in contrast to the overall G+C content of the *P. aeruginosa* genome, which is 67% (46). This unusual disparity in the base composition of B-band gene clusters has been noted previously, prompting speculation that these clusters were acquired by horizontal transfer from diverse bacterial species (12, 39).

Predicted proteins cluster in protein families. We anticipate that the *himD/ihfB*-to-*wbpM* DNA regions which we cloned in most cases encode genes involved in the synthesis of the bacterial O-antigen outer polysaccharide coat. This has been experimentally demonstrated for O5, O6, and O11 gene clusters (2, 5, 12). To this end, we analyzed the major ORFs for the 11 sequence groups, as shown in Fig. 2. These potential proteins were identified with the gene prediction program GENEMARK (30). Each DNA sequence analyzed except the ATCC O15 sequence encoded multiple, closely spaced, nonoverlapping ORFs that were oriented in the same direction.

Protein sequences were analyzed by using the PFAM protein analysis package (1). Many of the predicted proteins exhibited significant relationships with previously characterized protein families. We considered relationships to be valid in cases where the expected values were less than 0.02. In Fig. 2, the ORFs are annotated according to these family relationships. The protein families to which the annotated ORFs belong are shown in Table 3. Table 3 shows only those families that were represented by potential proteins three or more times in the 11 groups of DNA sequences analyzed. In all, 14 protein families with at least three representatives were identified. Consistent with a role in B-band biosynthesis, most of the protein families listed in Table 3 have been described previously in analyses of the O5, O6, and O11 operons (2, 5, 10, 12, 39), and almost all of them appear to play some role in polysaccharide metabolism. A complete list of the annotated protein families revealed by this analysis, including those with less than three repre-

sentatives, is available at our website (<http://www.genome.washington.edu/uwgc/O-Antigen/>).

There is some conservation of gene order within the putative operons. Despite the lack of DNA sequence conservation among gene clusters, *wzz* homologs are invariably found at the 5' end of each operon. In addition, the GENEMARK program predicts the presence of a small protein consisting of approximately 100 amino acids just upstream of *wzz*. Unlike the *himD/ihfB* gene that is encoded by sequences common to the 20 IATS strains, this putative gene is found in the region of highly diverged DNA sequences unique to each DNA cluster. Despite this nucleotide sequence divergence, a multiple alignment of the translation products (6) revealed a high degree of amino acid sequence conservation (Fig. 3), suggesting that the encoded protein may be expressed and possess functional significance. We have designated this potential gene, which has no sequence homologs in the PFAM database, *orfA* (Fig. 2). On the downstream end of each operon, proteins belonging to the L-M-N (*wbpL*)-O (*wbpM*) families were found in this order in 8 of 11 sequence groups. Apart from these relationships, there were only scattered instances of conservation of gene order among the DNA sequence groups.

One characteristic of O-antigen biosynthesis gene clusters is the presence of two genes that encode hydrophobic proteins with multiple membrane-spanning domains. One of these genes, *wzx*, encodes an oligosaccharide transporter protein, or flippase, that translocates oligosaccharide subunits from the cytoplasm to the site of B-band synthesis in the periplasm of the cell (4). A second hydrophobic gene, *wzy*, encodes an O-antigen polymerase that assembles the oligosaccharide subunits into the repetitive B-band polysaccharide (12, 13). Neither protein exhibits significant sequence identity with other family members, and the functional roles of these proteins have been assigned based on knockout phenotypes (4, 12, 13). We performed a hydropathy scan of all identified ORFs using the DAS algorithm (9) and annotated the ORFs that appear to encode proteins with multiple potential transmembrane domains with the letter W (Fig. 2). All members of family N, which are *wbpL* homologs encoding the glycosyltransferase group 4 proteins (24), appear to contain genes that encode several membrane-spanning domains. In all but the O6 cluster, we found at least two other ORFs which encode proteins that have the features of a hydrophobic, multiple-membrane-spanning protein; as previously noted (2), the O6 gene cluster has no obvious *wzy*, which is the ORF that encodes O polymerase. The well-characterized O5 cluster has four highly hydrophobic ORFs encoding the N family protein (*wbpL*), O polymerase (*wzy*), flippase (*wzx*), and an acetylase of unknown function (*wbpC*) (33, 39). The latter is not found in other O-antigen clusters. Furthermore, we found that the putative flippases in O6 and O11 were grouped in family E, the polysaccharide biosynthesis protein family, suggesting that other members of this group may also be flippases. Finally, in 8 of 11 operons, the two hydrophobic ORFs that may encode flippase and O polymerase are contiguous.

Boundaries of sequence diversity. The diversity of sequences present at the O-antigen-B-band biosynthetic locus underscores the plasticity of genome sequences in different isolates of *P. aeruginosa*. Previous studies have described an island of strain-specific sequence that punctuates the genome (25), yet a

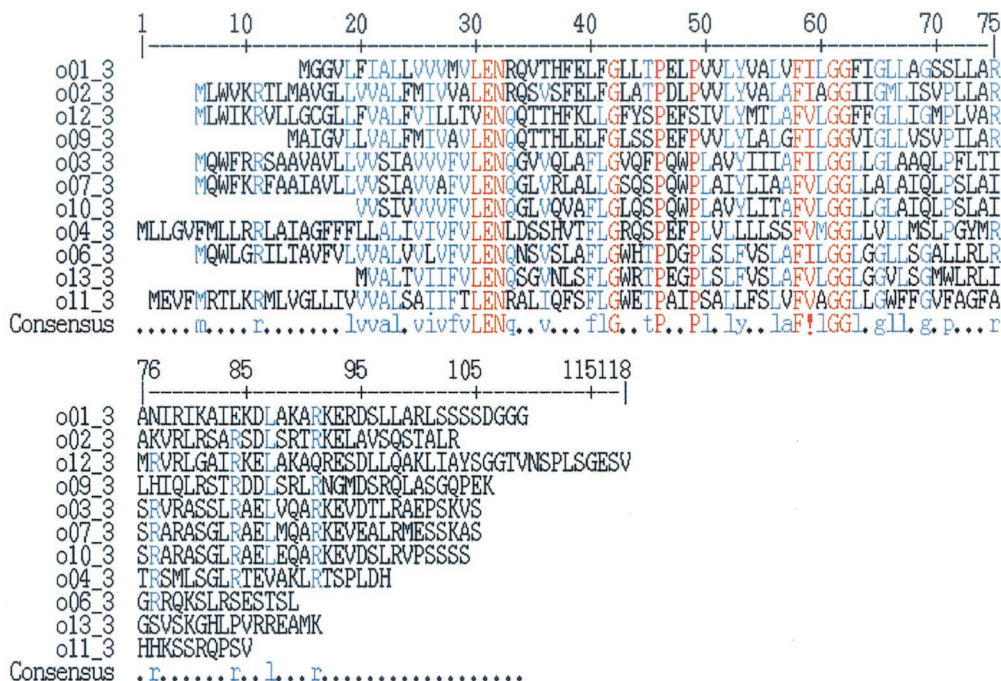


FIG. 3. Multiple alignment of *orfA* amino acid sequences (6). Red residues are >90% conserved, blue residues are 50 to 90% conserved, and the exclamation point indicates an I or V residue. *orfA* has no detectable homolog in the PFAM database.

striking feature of the present work is that many different sequence groups occupy the same locus in the genome. Nucleotide alignments were used to characterize the boundaries of the B-band biosynthetic islands at their 5' and 3' ends (Fig. 4). The 5' boundary occurs within a sharp, ~20-bp sequence window. The 3' boundary, which occurs within the *wbpM* coding region and therefore is presumably constrained by selection on the protein sequence, is less well-defined. A gradual decay from a clear consensus sequence to marked sequence diversity is observed to occur over a span of 450 nucleotides. Nonetheless, the overall picture is one in which many sequence-diverse gene clusters with similar functions (O-antigen synthesis) occupy a relatively precise location in the *P. aeruginosa* genome.

DISCUSSION

Our laboratory is applying several experimental strategies to pursue the study of genetic diversity in *P. aeruginosa*. Whole-genome shotgun resequencing of different *P. aeruginosa* isolates (Olson et al., unpublished) revealed that the O-antigen-B-band biosynthetic locus is a region where there is significant genetic heterogeneity. Previous investigators appreciated that the phenotypic diversity of surface antigens in this bacterium has an underlying genetic basis at this locus, and DNA sequencing efforts have demonstrated that divergent gene clusters occupy a common site in the genome (2, 5, 10, 12). We endeavored to provide a more comprehensive overview of this region by cloning and sequencing this region from the 20 serotypically distinct IATS strains. To achieve this goal, yeast recombinational cloning was used for targeted isolation of this segment from whole genomic DNA. Several successful variations of this technique have been described (3, 23, 37, 51). We

view yeast-based cloning as a powerful method for manipulating large pieces of DNA (37) and as a general tool for targeted cloning of specific regions from bacterial genomes.

Cloning and sequencing of the O-antigen biosynthetic locus from the 20 IATS strains revealed 11 distinct gene clusters. Annotation of the gene products of these clusters according to protein families revealed significant overlaps with gene fami-

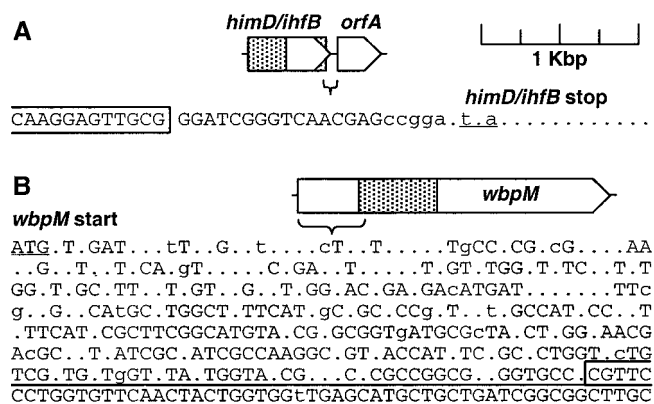


FIG. 4. Degeneration of DNA sequence conservation in the O-antigen region. (A) Sequence boundary at the 5' end of the O-antigen region. Targeting sequences used for cloning are enclosed in a box. Uppercase letters indicate nucleotides that are fully conserved, lowercase letters indicate nucleotides that are 90% conserved, and positions with two or more nucleotide differences are indicated by dots. Sequence conservation degenerates immediately at the end of the *himD/ihfB* coding region. (B) DNA sequence conservation in the 5' coding region of *wbpM* degenerates in the antisense direction. The recombinational targeting sequence is enclosed in a box.

lies known to be involved in the biosynthesis of cell surface polysaccharides, suggesting that all 11 groups are involved in the O-antigen biosynthetic pathway. The tentative assignments must be confirmed by using bacterial genetics. One interesting superficial feature of these clusters is related to the content and ordering of genes. The number of predicted genes varies from as many as 24 predicted ORFs for O12 to as few as 12 ORFs for the O11-O17 and O13-O14 groups. The ordering of genes is well-conserved at the 5' end (*orfA-wzz*) and at the 3' end (L-M-N-O in 8 of 12 groups), but otherwise gene order appears to be relatively random in the central portions of these clusters.

The sequences which we have determined provide an opportunity to develop DNA sequence-based PCR methods to serotype *P. aeruginosa* strains. Such methods may be particularly useful in the health industry for tracking infectious outbreaks and for characterization of untypeable strains, which are frequently encountered in advanced cystic fibrosis patients (18, 22).

Of the 11 DNA sequence groups which we characterized, 6 groups have more than one member. Within these groups there is a very high degree of DNA sequence conservation (>98%). This raises the question of how chemically distinct O-antigen structures are synthesized by closely related gene products. One possibility is that amino acid differences among related gene products alter the chemical specificities of the biosynthetic enzymes, leading to production of similar, yet distinct O-antigen structures. To facilitate experimental tests of this hypothesis, a comprehensive table of amino acid variation within the encoded gene products of each group is provided at our supplemental website (<http://www.genome.washington.edu/uwgc/O-Antigen/>). A more likely explanation for the chemical variability of O-antigen structures is that additional genes that are not linked to the O-antigen-B-band biosynthesis cluster contribute to assembly of the coat polysaccharide. Consistent with this idea, Newton and coworkers (33) demonstrated that three genes encoded in bacteriophage D3 were capable of converting the O5 serotype structure into an O16-like chemical structure. Interestingly, two of the genes responsible for seroconversion were found by Southern blot analysis to reside in O2 and O16 strains. Additionally, Rahim and coworkers recently described four contiguous genes (*rmlBDAC*) responsible for the synthesis of dTDP-L-rhamnose in *P. aeruginosa* (36). L-Rhamnose is a constituent of the conserved A-band core structure, and it is also found in the B-band oligosaccharides of O3 and O6 strains. It is reasonable to speculate that similar unlinked elements are responsible for serotype differences between strains in a given sequence group. This must be true for the O10 and O19 groups, for which no DNA sequence differences were found in 16 kbp of sequence. Similarly, for the O7-O8 group only two conservative amino acid changes were found, one in each of two different ORFs.

The overall percentages of DNA sequence differences, single-nucleotide polymorphisms, and insertions and deletions in the O10-O19 and O7-O8 sequence groups are 0 and 0.1% (22 changes in 18,703 bp), respectively. A whole-genome survey of the single-nucleotide polymorphisms and insertions and deletions in different *P. aeruginosa* strains revealed an average overall frequency of differences of 0.5% (Olson et al., unpublished). These data suggest that the O10-O19 and O7-O8 se-

quence groups are much younger in an evolutionary sense than the overall genome.

Dean and Goldberg (10) were the first investigators to describe a nonfunctional (with the exception of *wbpM*) B-band gene cluster in serotype O17 strains. The implication of this finding is that the genes responsible for core B-band synthesis reside elsewhere in the genome. We appear to have found a second example of this phenomenon with serotype O15. The Lory O15 strain bears a copy of the O3 gene cluster that is punctuated with insertion elements. The ATCC O15 strain carries a nearly complete deletion of the B-band cluster. However, this strain has *orfA*, which is located at the extreme 5' end of the B-band gene clusters. Surprisingly, the *orfA* from the ATCC O15 sequence is most closely related to the O11 gene. This result indicates that the two O15 strains arose via different cell lineages and that *P. aeruginosa* strains of the same serotype do not necessarily have the same overall genotype. It is noteworthy that the O17 genotype also resides in a strain with vestigial O11 sequences (10).

In conclusion, we used yeast recombinational cloning to access the genetically diverse O-antigen-B-band biosynthetic gene cluster regions from the 20 serotypically distinct IATS *P. aeruginosa* strains. The DNA sequences from these regions lay the groundwork for DNA-based serotyping methods and further investigation into the mechanisms by which chemically diverse surface polysaccharides are assembled. One challenge for the future is to relate the chemical structures of the O-antigen lipopolysaccharides to the gene products involved in their synthesis. We found a sharp delineation in the boundaries between the ubiquitously conserved genomic sequence and the regions of genetic diversity. Finally, we provide evidence that the entire O15 gene cluster resides elsewhere in the genome and that strains with the same serotype do not necessarily have the same genomic organization.

Our description of the genomic sequences at this locus focuses on a phenomenon that is likely to be widespread among bacteria: the maintenance in bacterial populations, by balancing selection, of a large number of highly divergent, yet functionally related gene clusters at specific genomic loci. The methods which we describe provide an efficient, general approach for analysis of such genetic systems. A major challenge for the future is to develop an improved understanding of how these divergent gene clusters evolve. The anomalous base composition of the O-antigen-B-band biosynthetic gene cluster relative to the base composition of the remainder of the *P. aeruginosa* chromosome suggests that the clusters evolve in huge, taxonomically diverse bacterial populations, subpopulations of which exchange genetic information infrequently. It remains to be determined how these gene clusters recombine in the same site of the *P. aeruginosa* chromosome and what combination of processes accounts for the abrupt boundaries between the highly conserved framework of the *P. aeruginosa* genome and the divergent sequences present in the O-antigen-B-band biosynthetic gene cluster in different strains.

ACKNOWLEDGMENTS

We thank the entire staff of the University of Washington Genome Center for their indispensable contributions to this work. In particular, we thank Peter Chapman for coordinating sequencing efforts, Eric Haugan and David Waring for annotation and submission of O-anti-

gen sequences, and Gregory Alexander for creating supplemental web-site content associated with this work.

This work was supported by Cystic Fibrosis Foundation grants to Maynard V. Olson (*P. aeruginosa* genomic sequence) and Sam Miller (program project) and by Center of Excellence for the Study of Natural Genetic Variation grant HG02351 from the NHGRI to Maynard V. Olson.

REFERENCES

- Bateman, A., E. Birney, R. Durbin, S. R. Eddy, K. L. Howe, and E. L. Sonnhammer. 2000. The Pfam protein families database. *Nucleic Acids Res.* **28**:263–266.
- Belanger, M., L. L. Burrows, and J. S. Lam. 1999. Functional analysis of genes responsible for the synthesis of the B-band O antigen of *Pseudomonas aeruginosa* serotype O6 lipopolysaccharide. *Microbiology* **145**:3505–3521.
- Bhargava, J., C. S. Shashikant, J. L. Carr, H. Juan, K. L. Bentley, and F. H. Ruddle. 1999. Direct cloning of genomic DNA by recombinogenic targeting method using a yeast-bacterial shuttle vector, pClasper. *Genomics* **62**:285–288.
- Burrows, L. L., and J. S. Lam. 1999. Effect of *wzx* (*rfbX*) mutations on A-band and B-band lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa* serotype O5. *J. Bacteriol.* **181**:973–980.
- Burrows, L. L., D. F. Charter, and J. S. Lam. 1996. Molecular characterization of the *Pseudomonas aeruginosa* serotype O5 (PAO1) B-band lipopolysaccharide cluster. *Mol. Microbiol.* **22**:481–495.
- Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**:10881–10890.
- Creuzenet, C., and J. S. Lam. 2001. Topological and functional characterization of WbpM, an inner membrane UDP-GlcNAc C6 dehydratase essential for lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **41**:1295–1310.
- Creuzenet, C., M. Belanger, W. W. Wakarchuk, and J. S. Lam. 2000. Expression, purification, and biochemical characterization of WbpP, a new UDP-GlcNAc C4 epimerase from *Pseudomonas aeruginosa* serotype O6. *J. Biol. Chem.* **275**:19060–19067.
- Cserzo, M., E. Wallin, I. Simon, G. von Heijne, and A. Elofsson. 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**:673–676.
- Dean, C. R., and J. B. Goldberg. 2000. The *wbpM* gene in *Pseudomonas aeruginosa* serogroup O17 resides on a cryptic copy of the serogroup O11 O antigen gene locus. *FEMS Microbiol. Lett.* **187**:59–63.
- Dean, C. R., A. Datta, R. W. Carlson, and J. B. Goldberg. 2002. *wbjA* adds glucose to complete the O-antigen trisaccharide repeating unit of the lipopolysaccharide of *Pseudomonas aeruginosa* serogroup O11. *J. Bacteriol.* **184**:323–326.
- Dean, C. R., C. V. Franklund, J. D. Retief, M. J. Coyne Jr., K. Hatano, D. J. Evans, G. B. Pier, and J. B. Goldberg. 1999. Characterization of the serogroup O11 O-antigen locus of *Pseudomonas aeruginosa* PA103. *J. Bacteriol.* **181**:4275–4284.
- de Kievit, T. R., T. Dasgupta, H. Schweizer, and J. S. Lam. 1995. Molecular cloning and characterization of the *rfc* gene of *Pseudomonas aeruginosa* (serotype O5). *Mol. Microbiol.* **16**:565–574.
- Franco, A. V., D. Liu, and P. R. Reeves. 1998. The Wzz (Cld) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. *J. Bacteriol.* **180**:2670–2675.
- Gietz, R. D., and R. A. Woods. 2001. Genetic transformation of yeast. *Bio-Techniques* **30**:816–831.
- Gordon, D., C. Desmarais, and P. Green. 2001. Automated finishing with autofinish. *Genome Res.* **11**:614–625.
- Hanahan, D., J. Jessee, and F. R. Bloom. 1991. Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol.* **204**:63–113.
- Hancock, R. E., L. M. Mutharia, L. Chan, R. P. Darveau, D. P. Speert, and G. B. Pier. 1983. *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis: a class of serum-sensitive, nontypable strains deficient in lipopolysaccharide O side chains. *Infect. Immun.* **42**:170–177.
- Hoffman, C. S., and F. Winston. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57**:267–272.
- Kiewitz, C., and B. Tummler. 2000. Sequence diversity of *Pseudomonas aeruginosa*: impact on population structure and genome evolution. *J. Bacteriol.* **182**:3125–3135.
- Kingston, R. L., R. K. Scopes, and E. N. Baker. 1996. The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure* **4**:1413–1428.
- Lam, J. S., L. A. MacDonald, A. M. Koprinski, and D. P. Speert. 1988. Characterization of nontypable strains of *Pseudomonas aeruginosa* from cystic fibrosis patients by means of monoclonal antibodies and SDS-polyacrylamide gel electrophoresis. *Serodiagn. Immunother. Infect. Dis.* **2**:365–374.
- Larionov, V., N. Kouprina, J. Graves, X. N. Chen, J. R. Korenberg, and M. A. Resnick. 1996. Specific cloning of human DNA as yeast artificial chromo-
somes by transformation-associated recombination. *Proc. Natl. Acad. Sci. USA* **93**:491–496.
- Lehrman, M. A. 1994. A family of UDP-GlcNAc/MurNAc:polyisoprenol-P GlcNAc/MurNAc-1-P transferases. *Glycobiology* **4**:768–771.
- Liang, X., X. Q. Pham, M. V. Olson, and S. Lory. 2001. Identification of a genomic island present in the majority of pathogenic isolates of *Pseudomonas aeruginosa*. *J. Bacteriol.* **183**:843–853.
- Lin, W. S., T. Cunneen, and C. Y. Lee. 1994. Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. *J. Bacteriol.* **176**:7005–7016.
- Ling, M., F. Merante, and B. H. Robinson. 1995. A rapid and reliable DNA preparation method for screening a large number of yeast clones by polymerase chain reaction. *Nucleic Acids Res.* **23**:4924–4925.
- Liu, P. V., H. Matsumoto, H. Kusama, and T. Bergan. 1983. Survey of heat-stable major somatic antigens on *Pseudomonas aeruginosa*. *Int. J. Syst. Bacteriol.* **33**:256–264.
- Liu, P. V., and S. Wang. 1990. Three new major somatic antigens of *Pseudomonas aeruginosa*. *J. Clin. Microbiol.* **28**:922–925.
- Lukashin, A. V., and M. Borodovsky. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**:1107–1115.
- Munster, A. K., M. Eckhardt, B. Potvin, M. Muhlenhoff, P. Stanley, and R. Gerardy-Schahn. 1998. Mammalian cytidine 5'-monophosphate N-acetylneuraminic acid synthetase: a nuclear protein with evolutionarily conserved structural motifs. *Proc. Natl. Acad. Sci. USA* **95**:9140–9145.
- Nakata, D., B. E. Close, K. J. Colley, T. Matsuda, and K. Kitajima. 2000. Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosphate synthase which does not have deaminoneuraminic acid (KDN) 9-phosphate synthase activity. *Biochem. Biophys. Res. Commun.* **273**:642–648.
- Newton, G. J., C. Daniels, L. L. Burrows, A. M. Kropinski, A. J. Clarke, and J. S. Lam. 2001. Three-component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage D3. *Mol. Microbiol.* **39**:1237–1247.
- Oldenburg, K. R., K. T. Vo, S. Michaelis, and C. Paddon. 1997. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.* **25**:451–452.
- Raetz, C. R., and S. L. Roderick. 1995. A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. *Science* **270**:997–1000.
- Rahim, R., L. L. Burrows, M. A. Monteiro, M. B. Perry, and J. S. Lam. 2000. Involvement of the *rml* locus in core oligosaccharide and O polysaccharide assembly in *Pseudomonas aeruginosa*. *Microbiology* **146**:2803–2814.
- Raymond, C. K., E. H. Sims, and M. V. Olson. 2002. Linker-mediated recombinational subcloning of large DNA fragments using yeast. *Genome Res.* **12**:190–197.
- Raymond, C. K., T. A. Powder, and S. L. Sesson. 1999. General method for plasmid construction using homologous recombination. *BioTechniques* **26**:134–141.
- Rocchetta, H. L., L. L. Burrows, and J. S. Lam. 1999. Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*. *Microbiol. Mol. Biol. Rev.* **63**:523–553.
- Roychoudhury, S., T. B. May, J. F. Gill, S. K. Singh, D. S. Feingold, and A. M. Chakrabarty. 1989. Purification and characterization of guanosine diphospho-D-mannose dehydrogenase. A key enzyme in the biosynthesis of alginate by *Pseudomonas aeruginosa*. *J. Biol. Chem.* **264**:9380–9385.
- Schmidt, K. D., B. Tummler, and U. Romling. 1996. Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats. *J. Bacteriol.* **178**:85–93.
- Sherman, F. 1991. Getting started with yeast. *Methods Enzymol.* **194**:3–21.
- Shizuya, H., B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**:8794–8797.
- Spiers, A. J., A. Buckling, and P. B. Rainey. 2000. The causes of *Pseudomonas* diversity. *Microbiology* **146**:2345–2350.
- Stanislavsky, E. S., and J. S. Lam. 1997. *Pseudomonas aeruginosa* antigens as potential vaccines. *FEMS Microbiol. Rev.* **21**:243–277.
- Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warren, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. W. Hancock, S. Lory, and M. V. Olson. 2000. Complete genome sequence of *Pseudomonas aeruginosa*: a versatile bacterium and an opportunistic pathogen. *Nature* **406**:959–964.
- Stutzman-Engwall, K. J., S. L. Otten, and C. R. Hutchinson. 1992. Regulation of secondary metabolism in *Streptomyces* spp. and overproduction of daunorubicin in *Streptomyces peucetius*. *J. Bacteriol.* **174**:144–154.
- Swartley, J. S., L. J. Liu, Y. K. Miller, L. E. Martin, S. Edupuganti, and D. S. Stephens. 1998. Characterization of the gene cassette required for biosyn-

- thesis of the (α 1 \rightarrow 6)-linked *N*-acetyl-D-mannosamine-1-phosphate capsule of serogroup A *Neisseria meningitidis*. *J. Bacteriol.* **180**:1533–1539.
49. **Thoden, J. B., A. D. Hegeman, G. Wesenberg, M. C. Chapeau, P. A. Frey, and H. M. Holden.** 1997. Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli*. *Biochemistry* **36**:6294–6304.
 50. **Yao, Z., and M. A. Valvano.** 1994. Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (*rfb*) of *Escherichia coli* K-12 W3110: identification of genes that confer group 6 specificity to *Shigella flexneri* serotypes Y and 4a. *J. Bacteriol.* **176**:4133–4143.
 51. **Zeng, C., N. Kouprina, B. Zhu, A. Cairo, M. Hoek, G. Cross, K. Osoegawa, V. Larionov, and P. de Jong.** 2001. Large-insert bac/yac libraries for selective re-isolation of genomic regions by homologous recombination in yeast. *Genomics* **77**:27–34.
 52. **Zhao, X., C. Creuzenet, M. Belanger, E. Egbosimba, J. Li, and J. S. Lam.** 2000. WbpO, a UDP-N-acetyl-D-galactosamine dehydrogenase from *Pseudomonas aeruginosa* serotype O6. *J. Biol. Chem.* **275**:33252–33259.
 53. **Zhao, X., C. Creuzenet, M. Belanger, E. Egbosimba, J. Li, and J. S. Lam.** 2000. WbpO, a UDP-N-acetyl-D-galactosamine dehydrogenase from *Pseudomonas aeruginosa* serotype O6. *J. Biol. Chem.* **275**:39802.