

Identification of a Genomic Island Present in the Majority of Pathogenic Isolates of *Pseudomonas aeruginosa*

XIAOYOU LIANG,¹† XUAN-QUYNH T. PHAM,² MAYNARD V. OLSON,² AND STEPHEN LORY¹†*

Departments of Microbiology¹ and Medicine,² University of Washington Genome Center, University of Washington, Seattle, Washington 98195

Received 8 August 2000/Accepted 10 November 2000

Pseudomonas aeruginosa, a ubiquitous gram-negative bacterium, is capable of colonizing a wide range of environmental niches and can also cause serious infections in humans. In order to understand the genetic makeup of pathogenic *P. aeruginosa* strains, a method of differential hybridization of arrayed libraries of cloned DNA fragments was developed. An M13 library of DNA from strain X24509, isolated from a patient with a urinary tract infection, was screened using a DNA probe from *P. aeruginosa* strain PAO1. The genome of PAO1 has been recently sequenced and can be used as a reference for comparisons of genetic organization in different strains. M13 clones that did not react with a DNA probe from PAO1 carried X24509-specific inserts. When a similar array hybridization analysis with DNA probes from different strains was used, a set of M13 clones which carried sequences present in the majority of human *P. aeruginosa* isolates from a wide range of clinical sources was identified. The inserts of these clones were used to identify cosmids encompassing a contiguous 48.9-kb region of the X24509 chromosome called PAGI-1 (for “*P. aeruginosa* genomic island 1”). PAGI-1 is incorporated in the X24509 chromosome at a locus that shows a deletion of a 6,729-bp region present in strain PAO1. Survey of the incidence of PAGI-1 revealed that this island is present in 85% of the strains from clinical sources. Approximately half of the PAGI-1-carrying strains show the same deletion as X24509, while the remaining strains contain both the PAGI-1 sequences and the 6,729-bp PAO1 segment. Sequence analysis of PAGI-1 revealed that it contains 51 predicted open reading frames. Several of these genes encoded products with predictable function based on their sequence similarities to known genes, including insertion sequences, determinants of regulatory proteins, a number of dehydrogenase gene homologs, and two for proteins of implicated in detoxification of reactive oxygen species. It is very likely that PAGI-1 was acquired by a large number of *P. aeruginosa* isolates through horizontal gene transfer. The selection for its maintenance may be the consequence of expression of any one of the genes of unknown function or the genes which allow *P. aeruginosa* to survive under the conditions that generate reactive oxygen species. Alternatively, one or both of the transcriptional regulators encoded in PAGI-1 may control the expression of genes in the *P. aeruginosa* chromosome, which provides a selective advantage for strains that have acquired this genomic island.

Although it is a common inhabitant of soil and water, *Pseudomonas aeruginosa* causes a variety of human infections ranging from superficial skin infection to acute infections of damaged sites such as eyes and invasion of tissues through severe burns and wounds (4, 8, 24). This organism causes a number of infections of mucosal tissues, such as those of the urinary or respiratory tract (10, 25). The common predisposing condition in patients that are infected by *P. aeruginosa* is some form of breach in host defenses or the specialized nature of the underlying disease, such as cystic fibrosis (2, 6, 11). It is also an important nosocomial pathogen, and it contributes significantly to the incidence of urinary tract infections (UTI) in patients with indwelling catheters (2, 25). Many mucosal infections are difficult to eradicate due to several factors, the most important of which is the relative poor activity of antibiotics against *P. aeruginosa* due to multiple resistance mechanisms expressed by the bacterium (13).

One of the many unanswered questions about *P. aeruginosa*

infections relates to the genetic makeup of strains that are responsible for the specific infections. Do all strains of *P. aeruginosa* have a similar or identical genetic repertoire, in which case the types of infections are solely determined by the predisposing conditions of the patients? Alternatively, is the genome of *P. aeruginosa* relatively plastic, in which case various horizontal gene transfer mechanisms might have led to the evolution of strains with specific determinants which allow them to infect different compromised patients? Extensive genomic rearrangements, as well as acquisition or loss of large blocks of DNA, have been demonstrated in several surveys of natural *P. aeruginosa* isolates (20). Evolution of pathogenic variants from nonpathogenic or less virulent strains is a well-documented phenomenon in many bacterial species. Lysogenic conversion of bacteria by bacteriophages often enhances the virulence of some organisms (26). Similarly, distinct disease mechanisms operate in several bacterial species, and these depend on the ability of commensal strains to acquire virulence genes located on plasmids or on large DNA segments (5, 12).

Comparisons of several genomes of *P. aeruginosa* from different clinical sources may provide answers about the association of specific genes with particular diseases. Moreover, completion of the sequencing project of the genome of *P. aeruginosa* PAO1 (22) provides a reference for understanding

* Corresponding author. Mailing address: Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115. Phone: (617) 432-5099. Fax: (617) 738-7664. E-mail: stephen_lory@hms.harvard.edu.

† Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115.

TABLE 1. Strains and plasmids^a

Strain or plasmid	Description	Source
<i>E. coli</i>		
DH5 α	<i>supE44</i> Δ <i>lacU169</i> (ϕ 80 <i>lacZ</i> Δ M15) <i>hsdR17 recA1 endA1 gyrA96 thi-1 relA1</i>	Gibco BRL
LE392	F ⁻ <i>hsdR514</i> (<i>r_K</i> ⁺ <i>m_K</i> ⁺) <i>supE44 supF58 lacY1 galK2 galKT22 metB1 trpL</i>	Lab collection
TG1	<i>supE hsd</i> Δ 5 <i>thi</i> Δ (<i>lac-proAB</i>) F ⁺ [<i>traD36 proAB</i> ⁺ <i>lacI</i> ^q <i>lacZ</i> Δ M15]	Lab collection
<i>P. aeruginosa</i>		
PAO1	Wild-type laboratory strain	Lab collection
X13273	Blood isolate from UW clinical lab	Lab collection
X13397	Blood isolate from UW clinical lab	Lab collection
X16529	Blood isolate from UW clinical lab	Lab collection
S29712	Blood isolate from UW clinical lab	Lab collection
S35004	Blood isolate from UW clinical lab	Lab collection
H21651	Blood isolate from UW clinical lab	Lab collection
CF2	CF isolate from Boston, Mass.; nonmucoid	Lab collection
CF3	CF isolate from Gainesville, Fla.; nonmucoid	Lab collection
CF4	CF isolate from Tulsa, Okla.; nonmucoid	Lab collection
CF5	CF isolate from Columbia, Mo.; nonmucoid	Lab collection
CF18	CF isolate from a 24-month-old patient; nonmucoid	Lab collection
CF29	CF isolate from San Francisco, Calif.; mucoid	Lab collection
PA15	UTI isolate from Harbview	Lab collection
U2504	UTI isolate from R. Ramphal	Lab collection
U779	UTI isolate from R. Ramphal	Lab collection
UTI125	UTI isolate from Brad Cookson	Lab collection
W57761	UTI isolate from UW clinical labs	Lab collection
X24509	UTI isolate from UW clinical labs	Lab collection
Plasmids		
M13mp18	Single-stranded filamentous phage vector	Novagen
LX1 to LX744	M13 clones containing UTI strain X24509-specific sequences	This study
Supercos-DBI	Cosmid vector	Stratagene
pDX4F8	Cosmid containing the left part of PAGI-1	This study
pDX9E8	Cosmid containing the left part of PAGI-1	This study
pDX11G5	Cosmid containing the middle part of PAGI-1	This study
pDX11E6	Cosmid containing the right part of PAGI-1	This study

^a UW, University of Washington; CF, cystic fibrosis.

changes at the genomic level which allow this organism to thrive in such a wide range of environments. Although PAO1 was originally isolated from a clinical source, it has been passaged extensively on laboratory media for more than 3 decades and is therefore very likely adapted to laboratory conditions. Here we report the preliminary identification of DNA segments present in a strain of *P. aeruginosa* isolated from a patient with a UTI and absent in strain PAO1. One such segment was sequenced and was shown to be present in the majority of clinical isolates. The analysis of this DNA segment provides several clues about the potential virulence properties that the gene products of this island confer on *P. aeruginosa*.

MATERIALS AND METHODS

Strains, plasmids, and growth of bacteria and M13 phages. The bacterial strains, plasmids, and phages used in this study are listed in Table 1. *P. aeruginosa* and *Escherichia coli* were grown at 37°C on Luria-Bertani (LB) medium (21). For plasmid maintenance, media were supplemented with 100 μ g of ampicillin per ml for *E. coli* and 150 μ g of carbenicillin per ml for *P. aeruginosa*. M13 phages were propagated at 37°C in LB medium in the presence of *E. coli* TG1.

Preparation of *P. aeruginosa* chromosomal DNA for the construction of libraries and Southern blot analyses. A 20-ml overnight culture from a single colony was centrifuged at 1,700 \times g for 10 minutes; the pellet was then washed with an equal volume of 0.85% NaCl followed by a second wash with TES (10 mM Tris-HCl, 25 mM EDTA, 150 mM NaCl [pH 8.0]). The bacteria were suspended in 10 ml of TE (10 mM Tris-HCl, 25 mM EDTA). Then, 0.5 ml of 20% sodium dodecyl sulfate was added to the suspension and the cells were incubated at 37°C for a few minutes until the suspension cleared. The lysate was extracted with

Tris-buffered phenol (pH 8.0), followed by Tris-buffered phenol-chloroform. The DNA in the aqueous phase was precipitated with 1/10 volume of 3 M sodium acetate (pH 4.8) and 2 volumes of ethanol. The DNA was spooled out with a glass rod, rinsed with 1 ml of 70% ethanol, air dried, and then resuspended in TE containing 20 μ g of RNase per ml.

Construction of an M13 library of *P. aeruginosa* chromosomal DNA fragments. Approximately 100 μ l of chromosomal DNA was sheared for 5 s on a W380 sonicator (Ultrasonics, Inc.) using the following settings: duty cycle, 68%; output control, 41/2 position; cycle time, continuous. The sonicated DNA was repaired by T4 DNA polymerase (Boehringer Mannheim) in 20 μ l of 5 \times buffer, 2 μ l of 10 mM deoxynucleotide triphosphate, 3 μ l of T4 DNA polymerase (10 U/ μ l), and 75 μ l of sonicated DNA solution at 37°C for 1 h. The blunt-ended DNA was size fractionated by agarose gel electrophoresis. Fragments between 1.5 and 3.0 kb were recovered using a gel extraction kit (Qiagen). The fragments were treated with T4 DNA polymerase and purified using the PCR product purification kit (Qiagen). The DNA concentration was determined spectrophotometrically and adjusted to 50 ng/ μ l. The ligation mixture (10 μ l) consisted of 50 ng of the *P. aeruginosa* DNA, 50 ng of M13mp18 replicative-form DNA which had been predigested with *Sma*I and dephosphorylated (Novagen), and 200 U of T4 DNA ligase (New England Biolabs) and was incubated at room temperature overnight. Aliquots of this ligation mixture were used to transform *E. coli* TG1 and plated in soft agar, overlaid on L-agar, plates containing X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) and IPTG (isopropyl- β -D-thiogalactopyranoside) as described by Sambrook et al. (21). White plaques were recovered following overnight incubation at 37°C.

Preparation of the arrays for differential genomic DNA hybridization. White plaques from the M13 library plates were inoculated into 1 ml of a suspension of *E. coli* TG1 cells (grown overnight and resuspended 1:200 in LB medium) in 96-well plates (Beckman) containing two 3-mm glass beads (VWR Scientific). The plates were covered with sterile filter paper and incubated 12 to 14 h with shaking at 37°C. The following day, the plates were centrifuged at 1,700 \times g for

30 min and 100 μ l of the phage-containing supernatant was transferred to polypropylene microtiter plates.

A 96-pin metal spotting device (Nalgene/Nunc) was used to spot aliquots of the M13 phage mixture in duplicate onto two sets of 7.9- by 11-cm nylon membrane sheets (Hybond-N+; Amersham). The sheets were treated by sequential 5-min exposure to filter paper saturated with a solution of 1.5 M NaCl–0.5 M NaOH, followed by a neutralization step on sheets impregnated with 1.5 M NaCl–0.5 M Tris-HCl (pH 7.5). The nylon sheets were then rinsed in 5 \times SSC (20 \times SSC is 0.3 M sodium citrate and 3 M NaCl), air dried, baked at 80°C for 2 h, and soaked in 5 \times SSC prior to hybridization. Chromosomal DNA probes were prepared by sonication of total DNA from PAO1 and from the UTI strain X24509, followed by purification using a PCR purification kit (Qiagen). This DNA was labeled with a Gene Image random primer labeling kit (Amersham). The filters were hybridized with the probes and detected using the instructions and conditions provided by the manufacturer except that the washing temperature was increased to 62°C. One set of the membranes was hybridized with the PAO1 chromosomal DNA probe, while the other set was hybridized with the X24509 chromosomal DNA probe. M13 clones which were positive for the X24509 probe hybridization but negative for PAO1 probe hybridization were recovered and used as templates for DNA sequencing of the inserts. To assess the distribution of the insert sequences among different *P. aeruginosa* isolates, M13 clones that did not react with the PAO1 probe were rearranged on fresh nylon sheets and probed with labeled chromosomal DNA from 19 different *P. aeruginosa* strains.

Construction and screening of a cosmid library of *P. aeruginosa* X24509. The large insert DNA fragments were prepared by limited digestion of purified X24509 chromosomal DNA with 0.8 U of *Sau*3A (New England Biolabs) in a 100- μ l volume at 37°C for 1 h. The enzyme was inactivated by heat treatment at 67°C for 10 min. The DNA was extracted with 1 volume of phenol-chloroform (1 volume/1 volume) followed by ethanol precipitation. The DNA fragments were resuspended in 80 μ l of water and dephosphorylated with 10 U of calf intestinal alkaline phosphatase (Promega) at 37°C for 1 h. The phosphatase was heat inactivated and extracted with phenol-chloroform, and the DNA was precipitated with ethanol. Vector DNA (Super-Cos-DBI) was digested with *Bam*HI, dephosphorylated, and purified as described for the preparation of chromosomal insert DNA. Ligation was performed in a total volume of 10 μ l containing 5 μ g of insert DNA, 1 μ g of vector DNA, and 200 U of T4 DNA ligase at room temperature overnight. Packaging into lambda particles and recovery of cosmid clones was done as described in the protocol provided by the supplier of the packaging extracts (Stratagene). Clones were grown in 96-well plates, and following addition of glycerol to 25%, they were stored frozen at –80°C. The library consisted of a total of 19 copies each of 96 individual cosmid clones. For hybridization with labeled DNA probes, aliquots of the library were spotted onto nylon membranes with a 96-pin spotting tool and hybridized with specific DNA probes derived from selected M13 clones that contained X24509-specific segments.

Sequencing of M13 clones and cosmids. Single-stranded DNA for a small number of M13 clones were prepared as described by Sambrook et al. (21). One end of the single-stranded DNA was sequenced with the reverse primer GTTT TCCAGTACGACGTTG. The other ends of some clones were sequenced by first PCR amplifying the entire M13 insert and then sequencing the product with the forward primer CAGCTATGACCATGATTACG. The cosmids were propagated in *E. coli* DH5 α and were purified using the Qiagen midiprep kit. Ends of the cosmids were sequenced with the T3' primer GGCGTATCAGAGGCC TTTC and the T7' primer CGATGATAAGCGGTCAAAC. The DNA sequence of the entire cosmid was determined following the construction of an M13 library using sonicated DNA as described for the construction of the M13 library from chromosomal DNA. The individual reads from the shotgun sequencing of the M13 clones were assembled into contigs using the base-caller program Phred and the assembly program Phrap (7).

Nucleotide sequence accession number. The PAGI-1 sequence determined in this study was submitted to the National Center for Biotechnology Information (NCBI) gene bank under accession no. AF241171.

RESULTS

Identification of sequence present in the genome of a UTI isolate of *P. aeruginosa* and absent from PAO1. A library of random fragments of chromosomal DNA from UTI isolate X24509 was constructed in M13mp18 and plated onto media with X-Gal. Phages from white (i.e., insert-containing) plaques

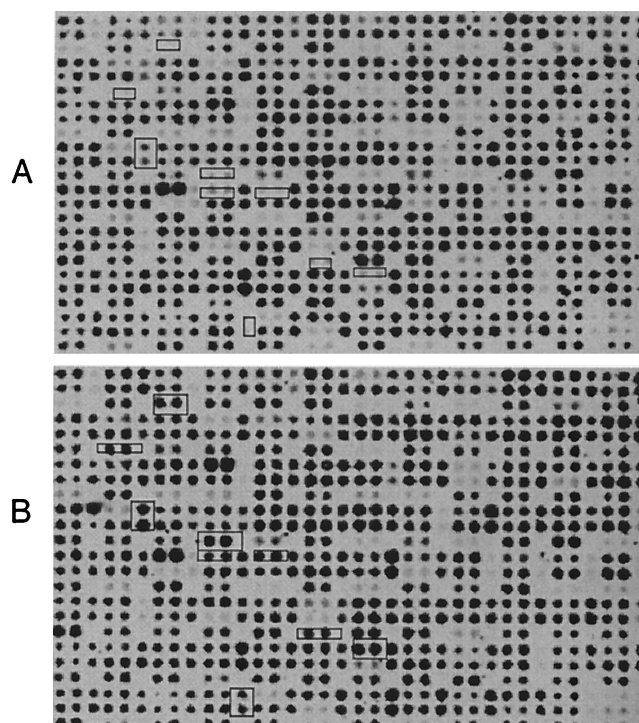


FIG. 1. Differential genomic DNA hybridization array. A library of DNA fragments from *P. aeruginosa* X24509 in M13 was spotted, in duplicate, onto two sets of nylon membranes. One membrane was probed with genomic DNA from *P. aeruginosa* PAO1 (A), and the other was probed with genomic DNA from X24509 (B). The boxed spots show the clones that carry DNA inserts that show significantly reduced hybridization with the PAO1 probe and therefore are likely to be X24509 specific.

were propagated in 96-well plates, and small aliquots were arrayed onto nylon sheets in duplicate. One set of the arrayed clones was probed with labeled DNA from X24509, while the second set was probed with PAO1 DNA. The hybridization results are shown in Fig. 1. The phages that show X24509-specific hybridization are indicated, and the corresponding clones should carry DNA fragments that are present in the X24509 genome but absent from the PAO1 genome. Approximately 2,000 M13 clones were screened by differential genomic DNA hybridization, and 74 putative clones of DNA unique to X24509 were obtained. A number of clones that reacted weakly with the PAO1 probe may contain inserts which contain not only unique X24509 segments but also those that may be present in PAO1, such as junction sequences of inserted elements. Alternatively, they may contain polymorphic genes which react poorly with the PAO1 probe. A single sequencing read was generated from all 74 M13 clones using the M13 reverse primer, and for a number of genes (see Table 2) a sequencing read was obtained from the complementary strand following PCR amplification of the insert DNA. When these sequences were compared to the PAO1 genome using the Basic Local Alignment Search Tool (BLAST) at the NCBI (http://www.ncbi.nlm.nih.gov/Microb_blast/unfinishedgenome.html), 54 out of 74 showed no corresponding sequences in the PAO1 genome. The sequence analysis is shown in Table 2. One of the striking features of these sequences is the relatively

TABLE 2. X24509-specific sequence information^a

Name	GC %	Homolog product (range)	Identity (%)
LX4	53.11	tRNA-splicing endonuclease positive effector of <i>Schizosaccharomyces pombe</i> (1512–1707)	28
LX5	55.00	Posterior end mark of <i>Ciona savignyi</i> (224–325)	26
LX5R		Integrase of <i>Actinobacillus actinomycetemcomitans</i> (7–77)	25
LX7	66.05	Lytic enzyme of <i>Pseudomonas aeruginosa</i> (1–209)	62
LX9	60.41	Gag of human endogenous retrovirus K (168–352)	26
LX10	50.34	Cap5F of <i>Staphylococcus aureus</i> (150–367)	42
LX12	59.78	1–400, PAO1 sequence, others, putative homeobox protein of <i>Macaca mulatta</i> (56–89)	43
LX12R		Protein kinase B of <i>Pseudomonas aeruginosa</i> (106–151)	45
LX15	59.05	FGF homolog of <i>Drosophila melanogaster</i> (509–598)	32
LX16	63.82	Calcium channel alpha1 subunit of <i>Homo sapiens</i> (907–1003)	33
LX17	63.79	Polymerase of rice ragged stunt virus (453–512)	37
LX18	63.96	Tn4652 transposase of <i>Pseudomonas putida</i> (820–876)	31
LX19	54.79	Heat shock protein HtrC of <i>Escherichia coli</i> (48–156)	27
LX21	64.59	Homeodomain protein of <i>Homo sapiens</i> (172–213)	38
LX23	57.89	Probable membrane protein YDR221w of <i>Saccharomyces cerevisiae</i> (453–546)	23
LX24	61.65	Putative oxidoreductase of <i>Streptomyces coelicolor</i> (140–337)	31
LX24R		Polyketide synthase of <i>Streptomyces ambofaciens</i> (499–558)	35
LX26	46.09	MTG8-related protein of <i>Homo sapiens</i> (413–448)	38
LX29	65.16	Homeobox protein of <i>Homo sapiens</i> (73–140)	49
LX29R		No hit	
LX30	59.46	Hypothetical protein 2 (PicA 5' region) of <i>Agrobacterium tumefaciens</i> (5–80)	31
LX30R		PAO1 sequence	
LX31	49.65	PAO1 sequence	
LX31R		Peptide synthetase of <i>Claviceps purpurea</i> (51–110)	30
LX32	56.79	51–377, pilin biogenesis protein (PilB) of <i>Pseudomonas aeruginosa</i>	97
		384–701, pilin biogenesis protein (PilC) of <i>Pseudomonas aeruginosa</i>	76
LX32R		22–753, pilin biogenesis protein (PilC) of <i>Pseudomonas aeruginosa</i>	73
LX35	56.74	Outer membrane protein of <i>Campylobacter jejuni</i> (82–242)	27
LX36	52.76	Probable membrane protein YMR317w of <i>Saccharomyces cerevisiae</i> (424–512)	42
LX37	50.31	Putative cell division protein of <i>Escherichia coli</i> (7–289)	38
LX38	48.99	20- β -hydroxysteroid dehydrogenase of <i>Streptomyces exfoliatus</i> (30–153)	33
LX39	52.24	Regulatory protein mnt of Bacteriophage P22 (1–36)	50
LX40	60.20	Hypothetical protein Rv3633 of <i>Mycobacterium tuberculosis</i> (120–262)	38
LX40R		Putative oxidoreductase of <i>Streptomyces coelicolor</i> (286–398)	33
LX41	61.65	Hypothetical protein Rv3854c of <i>Mycobacterium tuberculosis</i> (188–413)	53
LX41R		PpdK of <i>Mycobacterium tuberculosis</i> (94–171)	35
LX42	53.24	Putative DNA-binding protein of <i>Arabidopsis thaliana</i> (359–390)	50
LX43	59.59	Intestinal mucin of <i>Homo sapiens</i> (156–206)	44
LX46	61.78	LipG of <i>Mycobacterium tuberculosis</i> (218–382)	47
LX46R		LipG of <i>Mycobacterium tuberculosis</i> (11–159)	41
LX47	54.66	Reverse gyrase of <i>Pyrococcus furiosus</i> (93–148)	39
LX48	50.07	20- β -hydroxysteroid dehydrogenase of <i>Streptomyces exfoliatus</i> (8–189)	31
LX49	54.14	Receptor interacting protein 140 of <i>Mus musculus</i> (513–600)	26
LX50	61.84	Polyketide synthase of <i>Streptomyces hygroscopicus</i> (4400–4489)	29
LX51	57.08	Hypothetical protein Rv3574 of <i>Mycobacterium tuberculosis</i> (32–182)	28
LX52	62.00	Late expression factor 8 of <i>Lymantria dispar</i> nucleopolyhedrovirus (614–637)	50
LX53	49.70	Putative protein of <i>Arabidopsis thaliana</i> (167–196)	46
LX54	64.37	6-Deoxyerythronolide B synthase II of <i>Saccharopolyspora erythraea</i> (2678–2745)	36
LX55	49.67	KIAA0350 of <i>Homo sapiens</i> (352–389)	43
LX57	69.67	DNA topoisomerase I of <i>Physarum polycephalum</i> (81–170)	35
LX58	61.25	VgrG protein of <i>Escherichia coli</i> (378–585)	34
LX59	52.19	Annexin XI of <i>Mus musculus</i> (132–187)	42
LX60	52.38	Etr-3 of <i>Xenopus laevis</i> (204–334)	21
LX61	62.38	Lactate dehydrogenase of <i>Neisseria meningitidis</i> (314–358)	36
LX62	56.11	SLM-2 of <i>Mus musculus</i> (219–266)	39
LX63	66.4	Chitinase of <i>Emericella nidulans</i> (426–604)	24
LX64	50.00	No hit	
LX66	59.00	Long ORF of TT virus (26–81)	34
LX67	56.56	2-Isopropylmalate synthase of <i>Schizosaccharomyces pombe</i> (15–201)	54
LX68	56.86	IS911 hypothetical protein, variant (IS911A) of <i>Escherichia coli</i> (14–108)	49
LX69	59.14	Putative transposase subunit of <i>Pseudomonas alcaligenes</i> (45–130)	33
LX70	57.00	Chitinase of <i>Emericella nidulans</i> (362–446)	30
LX72	58.50	Glycosyltransferase of <i>Pseudomonas aeruginosa</i> (157–237)	59
LX73	67.57	Uptake hydrogenase of <i>Rhodobacter sphaeroides</i> (136–220)	31
LX74	41.57	Ribosomal maturase of <i>Rhododendron edgeworthii</i> (159–274)	29

^a Bold type indicates the sequences of the four clones which were used to screen the cosmid library. Ranges show positions in base pairs. R, sequence from the other end of M13 clone.

low G+C content of most of the sequences, compared to that of *P. aeruginosa* (66%) (22). This suggests that these segments were very likely acquired by X24509 by horizontal gene transfer.

Distribution of X24509 DNA segments among different *P. aeruginosa* isolates. In order to assess whether any of the X24509 segments that were identified by differential genomic DNA hybridization are present in other *P. aeruginosa* strains, the 54 M13 clones of genes present in X24509 were arrayed, in multiple copies, onto nylon sheets and probed with labeled chromosomal DNA from a collection of *P. aeruginosa* strains. A total of 18 different probes were used, representing six UTI isolates, six blood isolates, and six strains obtained from patients with cystic fibrosis. The arrays were also probed with a PAO1 probe. The summary of the hybridization results is shown in Table 3. Certain M13 clones reacted only with the X24509 probe, and they include LX4, LX19, LX26, LX42, LX49, LX64, and LX67. Based on our limited survey, these are X24509-specific segments that are not present in any other isolates. Clones LX12, LX29, LX30, LX31, LX32, LX39, LX51, LX53, LX57, and LX68 reacted weakly with the PAO1 chromosomal probe. All but one very likely contain homologous sequences from PAO1 at one end of the insert. Clone LX32 carries the portion of the *pilB* and *pilC* chromosomal segment, and the weak hybridization of the PAO1 probe is almost certainly due to the polymorphic *pilC* gene. Finally, certain M13 clones (LX24, LX40, LX41, and LX46) reacted with the majority of probes from clinical strains. This group of M13 clones was then used to identify the contiguous regions in the *P. aeruginosa* X24509 chromosome in cosmid clones.

Cloning and sequencing of the pathogen-associated DNA from *P. aeruginosa*. A cosmid library containing chromosomal segments of *P. aeruginosa* X24509 was prepared, and it was analyzed using DNA probes generated by PCR amplification of inserts in M13 clones LX24, LX40, LX41, and LX46. All of the probes identified the same four cosmids, indicating that the probes were derived from a single contiguous region of the X24509 chromosome. The sequence of the ends of four cosmid clones (pDX4E8, pDX9F8, pDX11G5, and pDX11E6) was obtained using the primers T3' and T7', which flank the cloning site of the vector. The alignment of the cosmids and the location of the four M13 clones used for screening of the cosmid library is shown in Fig. 2. The T7 end of the insert in cosmid pDX4E8 showed identity to the sequence of the PAO1 genome (<http://www.pseudomonas.com/>; release date, 15 December 1999) increasing from bp 2438028, while the sequence obtained from the T3 end was unique. The T7 end of cosmid DX9F8 is identical to the PAO1 sequence, increasing from bp 2437159, and the T3 end is unique. This indicates that the T7 ends of the inserts in both DX9E8 and DX4F8 end within the same region with a difference of ca. 900 bp. For the insert in cosmid pDX11E6, its T3 end is identical to the PAO1 sequence, decreasing from bp 2456045, and its T7 end is a previously unknown sequence. For pDX11G5, its T3 end was a unique sequence, but its T7 end is identical to the PAO1 sequence, decreasing from bp 5263148. The discrepancies of the similar location of the ends of cosmids pDX4F8, pDX9E8, and pDX11E6 and the distal location of pDX11G5, relative to the PAO1 genome, were resolved during sequencing of the cosmids (see below).

The sequence of cosmids pDX4F8 and pDX11E6 was de-

termined by a shotgun approach, following preparation of random M13 libraries and obtaining single sequence reads from ca. 1,200 M13 clones. Given an average read length of 500 bp, approximately 600 kb of unique reads were generated, representing a 10-fold coverage of the sequenced DNA estimated to be ca. 60 kb in size. The final contig was 60,180 bp in size from the T7 end of cosmid pDX4F8 to the T3 end of cosmid pDX11E6. The BLAST search of the contig against the PAO1 genome sequence revealed that both ends of the contig contain sequences identical to that of PAO1. The X24509-specific region designated PAGI-1 (for "*P. aeruginosa* genomic island 1", which is absent from PAO1, starts at 925 bp and ends at 49,818 bp of the assembled contig. The length of the island sequence is therefore 48,893 bp. The island sequence flanked by junction sequences, i.e., those identical to the PAO1 sequences (1,575 bp on the left and 833 on the right), which correspond to the sequences present in the PAO1 genome was submitted to the gene bank of NCBI. Cosmid pDX11E6 contains bp 12311 to 48893 of PAGI-1, pDX11G5 contains bp 801 to 43891 of PAGI-1, pDX4F8 contains bp 1 to 43474 of PAGI-1, and pDX9E8 contains bp 1 to 48820 of PAGI-1. The locations of the M13 inserts that identified the island, relative to the PAGI-1 sequence, are as follows: LX24 is located at bp 25502 to 29068 of PAGI-1, LX40 is at bp 24392 to 26315 of PAGI-1, LX41 is at bp 21478 to 23497 of PAGI-1, and LX46 is at bp 20192 to 21236 of PAGI-1.

Although both ends of PAGI-1 connect with PAO1 sequence from the same region, these sequences are not contiguous in PAO1. The left end of PAGI-1 was contiguous with the PAO1 genome sequence at bp 2438952, while the right end was at bp 2445682. This shows that a 6,729-bp PAO1 sequence from bp 2438953 to 2445681 is replaced by PAGI-1 in strain X24509. The insertion position of PAGI-1 and the replacement of the PAO1 sequence are diagrammed in Fig. 3.

Sequence analysis and annotation of PAGI-1. The complete nucleotide sequence of PAGI-1 was determined and the sequence was annotated. PAGI-1 is 48,893 bp in length. The open reading frames (ORFs) of PAGI-1 were predicted using the GeneMark method based on Markov chain models for coding and noncoding regions of *P. aeruginosa* (14, 17). A total of 51 putative ORFs were predicted, and their organization is shown in Fig. 4. The protein sequence similarities of the PAGI-1 genes were analyzed using the program based on the BLAST algorithm (NCBI) and the e-motif search developed by Craig G. Nevill-Manning, Thomas D. Wu, and Douglas L. Brutlag, Biochemistry, Stanford University, and the results of this analysis are shown in Table 4. Approximately one-half of the genes are either hypothetical unknown or conserved hypothetical unknown genes. Among the genes that could be assigned putative functions, the most notable are the remnants of two transposable elements. Two genes, delineated by *orf32-orf33* and *orf37-orf38*, share high levels of similarity with elements encoding transposases A and B, respectively, which belong to the IS3 family of insertion sequences (19).

Annotation also revealed the presence of two putative transcriptional regulators. *orf5* encodes a polypeptide which shows significant similarity to the RpoN-dependent subfamily of regulatory elements and is most closely related to the PrpR of *Salmonella enterica* serovar Typhimurium (15). The predicted product of *orf30* shares similarity with putative regulatory pro-

TABLE 3. Distribution of 55 X24509-specific sequences^a

Name	Result for:																		
	PAO1	X24509	UTII125	U779	U2504	W57761	U126	X13273	X13397	X1259	S29712	S35004	H21651	CF18	CF2	CF3	CD4	CF5	CF29
LX4		P																	
LX5		P	P		P	P	P	P	P	P	P	P	P	P	P				P
LX7		P		P	P	P									P	P	P	P	P
LX9		P			P	P	P							P					
LX10		P	P	P			P												P
LX12	P	P	P		P		P	P	P	P	P	P	P	P	P				P
LX15		P	P	P			P	P					P						
LX16		P	P		P		P								P			P	P
LX17		P	P	P	P	P	P							P	P	P	P	P	P
LX18		P	P		P	P	P	P			P	P	P		P				P
LX19																			
LX21		P	P		P		P	P	P	P		P	P	P	P	P	P	P	P
LX23		P	P				P	P	P	P	P	P	P	P	P				
LX24	P	P	P	P	P	P	P	P	P	P	P		P	P	P	P	P	P	
LX26																		P	
LX29	P	P	P	P	P	P	P	P	P			P			P	P	P	P	
LX30	P	P	P	P	P	P	P						P		P	P	P		
LX31	P	P	P	P	P	P	P	P	P	P			P				P	P	P
LX32	P	P	P	P	P	P	P	P	P		P				P	P	P	P	P
LX35		P	P																
LX36		P	P																
LX37		P	P																
LX38		P	P				P												
LX39	P	P				P								P					P
LX40		P	P	P	P	P	P	P	P	P	P	P		P	P	P	P	P	P
LX41		P	P	P	P	P	P	P	P	P	P	P		P	P		P	P	P
LX42		P																	
LX43		P	P		P		P	P		P	P	P	P		P				P
LX46		P	P	P	P	P	P	P	P	P	P		P	P	P	P	P	P	
LX47		P	P				P	P										P	
LX48		P	P				P											P	
LX49		P																	
LX50		P	P				P				P	P	P		P				
LX51	P	P	P	P	P	P	P				P			P	P	P	P	P	
LX52		P	P				P	P			P				P			P	
LX53	P	P	P	P	P		P	P	P				P				P	P	
LX54		P	P	P	P	P	P		P			P	P	P	P	P	P		
LX55		P	P				P												
LX57	P	P	P	P	P	P	P	P							P	P	P	P	P
LX58		P				P	P							P	P	P	P	P	
LX59		P	P				P	P	P	P	P	P	P		P		P		
LX60		P	P				P	P	P	P	P	P	P		P		P		P
LX61		P	P		P	P					P	P		P	P				P
LX62		P	P				P												
LX63		P	P	P	P	P	P		P			P		P	P	P	P	P	P
LX64		P																	
LX66		P	P				P	P	P	P	P	P	P	P	P	P	P	P	
LX67		P																	
LX68	P	P	P	P	P	P	P	P	P	P			P		P	P	P	P	P
LX69		P	P								P				P				
LX70		P	P				P		P		P				P				
LX72		P	P	P			P												P
LX73		P	P	P	P	P	P	P	P						P	P	P	P	P
LX74		P				P		P	P										

^a P, positive hybridization of M13 clones with the corresponding chromosomal probes; empty cell, negative hybridization result. Bold type shows the hybridization profiles of the M13 clones which reacted with the largest number of chromosomal probes and were selected for the screen of the X24509 cosmid library.

teins in various *Mycobacterium* and *Streptomyces* species. This protein, SrmR, regulates expression of polyketide biosynthetic genes in *Streptomyces ambofaciens* (9). A number of genes encode various dehydrogenases (*orf8*, -9, -14, -16, -18, and -23). The left end of PAGI-1 also contains the coding sequence for two proteins that are homologous to paraquat-inducible proteins of *E. coli* (16). Paralogues of these genes were also iden-

tified in the *P. aeruginosa* genome, PAO4689 and PAO4690, which share 89 and 96% identity with *orf2* and *orf3*, respectively.

The insertion of PAGI-1 is accompanied by deletion of a 6,729-bp region, relative to the PAO1 genome (Fig. 3). The five genes which have been deleted (Table 5) include a gene encoding a putative transcriptional regulator of the LysR fam-

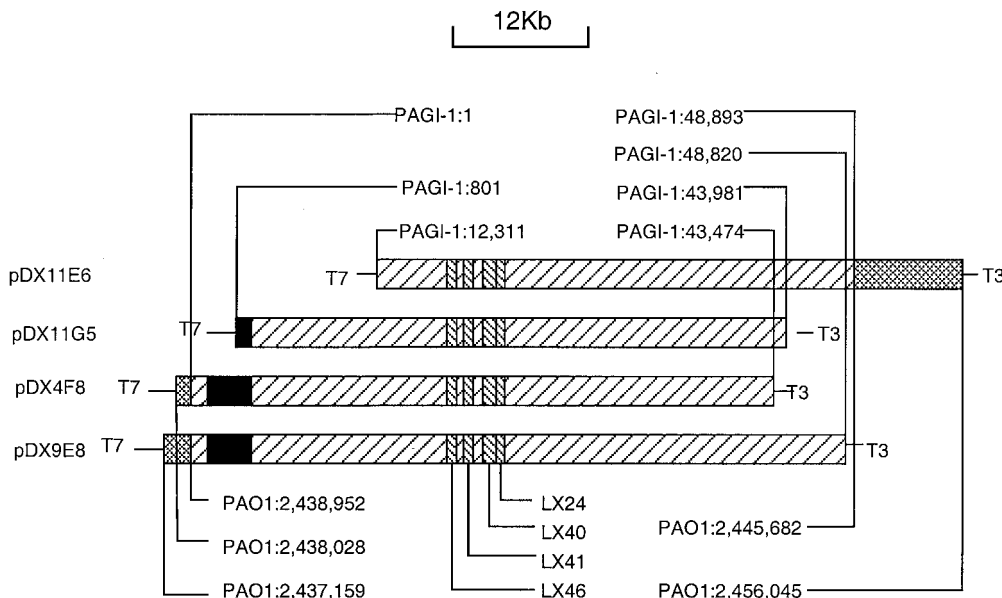


FIG. 2. Overlap of four cosmids that contain the X24509 genomic island PAGI-1. The cross-hatched regions indicate the junction sequences that are also present in PAO1. The left-hatched regions represent PAGI-1 sequences. The black boxes represent regions containing sequences with high degrees of similarity to a region in the PAO1 genome. The right-hatched boxes indicate the locations of the sequences of the M13 clones that were used to screen the X24509 cosmid library. Also indicated are the locations of the junction sequences in PAGI-1 and the sequences in the X24509 genome, with corresponding coordinates of the genome of strains PAO1. The locations of the T3 and T7 sites in the cosmid vector, used for sequencing, are also shown.

ily (PAO2220) and a gene encoding a transposase (PAO2221) which is homologous to a similar enzyme in *Rhizobium* sp. strain NGR234. The insertion of PAGI-1 results in the disruption of a gene which specifies a hypothetical protein (PAO2223) at the right junction, leading to the premature termination of the coding sequence of PAO2223 by 187 bp. The left junction between PAGI-1 and the rest of the chromosome is within a gene encoding a homologue of α -ketoglutarate

semialdehyde dehydrogenase (PAO2217). This protein is probably unaffected by the presence of PAGI-1, because the PAGI-1 sequence replaces two codons of PAO2217 (GTTGCT) with ATTGAC followed by a termination codon, which gives a polypeptide of identical size but with two different carboxy-terminal amino acids (Val Ala replaced with Ile Asp).

Location of PAGI-1 sequences in different *P. aeruginosa* strains. In order to assess the general pattern of PAGI-1 ac-

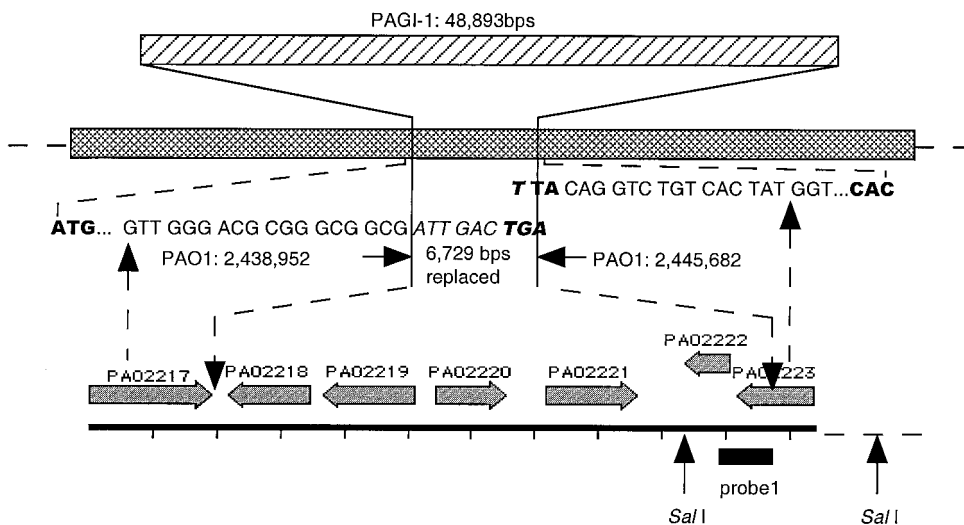


FIG. 3. Location of PAGI-1 in the genome of *P. aeruginosa*. Shown are PAGI-1, the replaced 6,729-bp chromosomal segment, and the junction sequences of the two flanking genes and PAGI-1. The cross-hatched bar represents the *P. aeruginosa* PAO1 genome. The genes within the replaced region are indicated, based on the annotation of the PAO1 sequence. The DNA sequences shown in italics, at the junctions, are derived from PAGI-1.

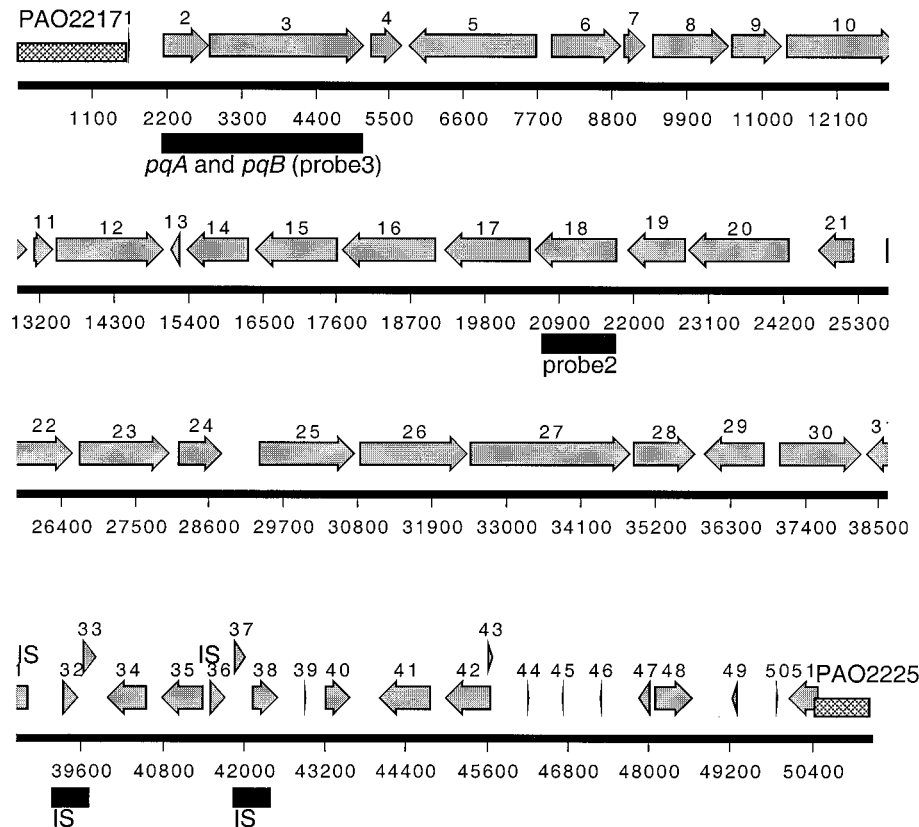


FIG. 4. Annotation of the PAGI-1 sequence. The cross-hatched region indicates the junction of the island with the chromosome relative to the coordinates in the PAO-1 genome. Also indicated are the locations of the homologues of the *pqiA* and *pqiB* genes and probe 2 and probe 3, which were used in the Southern hybridization. The gray arrows indicate the open reading frames in PAGI-1.

quisition in different *P. aeruginosa* isolates, we carried out a Southern blot analysis of different *P. aeruginosa* strains using probe 2 and probe 3 from two parts of PAGI-1 (Fig. 4) as well as probe 1 from a portion of the 6,729-bp region of PAO1 (Fig. 3), which is replaced in X24509 by PAGI-1. When probe 1 (corresponding to the 5' portion of PAO2222 and the 3' portion of PAO2223) (Fig. 3) was used, several strongly hybridizing bands of the same size as seen in PAO1 were detected in two strains from UTI patients (UTI125 and W57761) and several blood isolates (X13273, X13397, S35004, and H21651). The strongly hybridizing band indicates the presence of the 6,729-bp PAO1 segment in all of these strains, and additional DNA bands of weaker hybridization can be detected in the blots of these strains as well. Since one recognition site for *SalI*, which was used for the digestion of chromosomal DNA, is located outside the 6,729-bp region (Fig. 3), the identical size of the strong band in Fig. 5A indicates that this region is most likely at the same location in all of these strains. Probe 2, corresponding to the middle portion of PAGI-1 (corresponding to *orf18*) (Fig. 4), detected the presence of hybridizing sequences in the majority of strains, including some that showed hybridization to the PAO1-specific 6,729-bp probe 1 (Fig. 5B). These results indicate that in these latter strains (UTI125, W57761, X13273, X13397, and S35004) the acquisition of PAGI-1 was not accompanied by deletion of the same region, as was shown for X24509. Because the location of the

6,729-bp region is conserved, PAGI-1 in these strains is probably present at a location that is different from that of X24509. Strain H21651 is similar to PAO1, and it lacks PAGI-1 but contains the PAO1 band. Finally, strain PA15 shows no hybridization with either PAGI-1 probe 2 or probe 1 from the deleted 6,729-bp region of the PAO1 chromosome. Therefore, in this strain, a different genetic island is present at the same location as PAGI-1. All strains hybridized with probe 3 from the left portion of PAGI-1, encompassing *orf2* and *orf3* (Fig. 4), and either one or two bands were observed (Fig. 5C). All of the strains that were shown to carry PAGI-1 in Fig. 5B displayed two bands. These results are consistent with the presence of a highly homologous copy of these genes in PAO1, and apparently in all other *P. aeruginosa* strains as well.

DISCUSSION

In this study, we developed an array-based differential genomic DNA hybridization method to compare genomes of different bacterial strains. This method is similar to the representational-difference analysis that was used to detect genetic differences between virulent and avirulent *Mycobacterium bovis* strains (18) and between *Neisseria gonorrhoeae* and *Neisseria meningitidis* (23). By identifying those segments of DNA in a random-fragment phage library of *P. aeruginosa* X24509 DNA which are absent from the genome of PAO1, we have shown a

TABLE 4. Annotation of PAGI-1^a

No.	Identification ^b	Left end	Right end	Length (bp)	GC %	Homolog product	Score or stringency	E value ^c
	Dt	1	1584	1,584		PAO2217 of PAO1		
1	Dt	1636	1692	57	66.67	No significant similarity		
2	Dt	2140	2853	714	63.31	Paraquat-inducible protein A (<i>Escherichia coli</i>)	183	1e ⁻⁴⁵
3	Dt	2846	5149	2,304	65.97	Paraquat-inducible protein B (<i>Escherichia coli</i>)	285	9e ⁻⁷⁶
4	Dt	5208	5711	504	59.92	No significant similarity		
5	Rt	5777	7729	1,953	64.93	Propionate catabolism operon regulator prpR (<i>Escherichia coli</i>)	285	9e ⁻⁷⁶
6	Dt	7899	8981	1,083	67.93	FadE36 (<i>Mycobacterium tuberculosis</i>)	48	1e ⁻⁰⁴
7	Dt	8978	9334	357	66.95	Putative acyl-CoA dehydrogenase (<i>Streptomyces coelicolor</i> A3(2))		
8	Dt	9385	10557	1,173	68.37	Putative acyl-CoA dehydrogenase (<i>Streptomyces coelicolor</i> A3(2))	225	7e ⁻⁵⁸
9	Dt	10557	11324	768	71.61	2-Deoxy-D-gluconate 3-dehydrogenase (<i>Escherichia coli</i>)	140	1e ⁻³²
10	Dt	11366	13057	1,692	68.09	AMP-binding protein domain (<i>Oryza sativa</i>)	496	1e ⁻¹³⁹
11	Dt	13114	13437	324	63.27	Hypothetical protein (<i>Escherichia coli</i>)	60	5e ⁻⁰⁹
12	Dt	13434	15092	1,659	64.86	Putative amino acid transport protein (<i>Escherichia coli</i>)	508	e ⁻¹⁴³
13	Rt	15156	15308	153	58.17	Amino acid permeases	10 ^{-5.3d}	
14	Rt	15382	16341	960	64.90	Hypothetical dehydrogenase (<i>Streptomyces coelicolor</i> A3(2))	74	2e ⁻¹²
15	Rt	16391	17644	1,254	65.79	Diaminobutyric acid aminotransferase (<i>Halomonas elongata</i>)	359	2e ⁻⁹⁸
16	Rt	17690	19114	1,425	66.25	Putative aldehyde dehydrogenase (<i>Escherichia coli</i>)	450	1e ⁻¹²⁵
17	Rt	19193	20524	1,332	64.56	Putative amino acid/amine transport protein (<i>Escherichia coli</i>)	351	6e ⁻⁹⁶
18	Rt	20537	21784	1,248	62.58	1,3-Propanediol dehydrogenase	259	2e ⁻⁶⁸
						Iron-containing alcohol dehydrogenase motif	10 ^{-10.3d}	
19	Rt	21925	22800	876	63.81	Hydrolase motif II	10^{-8d}	
20	Rt	22797	24323	1,527	60.18	Aromatic-ring hydroxylase motif	10^{-10d}	
21	Rt	24707	25288	582	59.62	No significant similarity		
22	Dt	25737	26624	888	59.68	Conserved hypothetical protein (<i>Mycobacterium tuberculosis</i>)	84	2e⁻¹⁵
23	Dt	26686	28053	1,368	61.26	Bacterial phytoene dehydrogenase protein motif	10^{-10.2d}	
24	Dt	28149	28811	663	52.49	No significant similarity		
25	Dt	29342	30790	1,449	61.49	Conserved hypothetical protein (<i>Pseudomonas fluorescens</i>)	256	4e⁻⁶⁷
						Glycosyl hydrolase family 15 motif I	10^{-7.5d}	
26	Dt	30814	32460	1,647	62.72	Conserved hypothetical protein (<i>Pseudomonas fluorescens</i>)	182	6e ⁻⁴⁵
27	Dt	32469	34871	2,403	63.17	Transport protein (<i>Pyrococcus abyssi</i>)	86	1e ⁻¹⁵
28	Dt	34882	35844	963	66.25	Putative oxidoreductase (<i>Streptomyces coelicolor</i> A3(2))	59	5e ⁻⁰⁸
29	Rt	35929	36858	930	61.40	Conserved hypothetical protein (<i>Mycobacterium tuberculosis</i>)	90	2e ⁻¹⁷
30	Dt	37035	38288	1,254	64.04	SrmR, possible transcriptional activator (<i>Mycobacterium leprae</i>)	296	2e ⁻⁷⁹
31	Rt	38311	38853	543	54.33	No significant similarity		
32	Dt	39323	39586	264	56.06	Transposase A, IS3 family (<i>Yersinia enterocolitica</i>)	144	2e ⁻³⁴
33	Dt	39619	39850	231	59.48	Transposase B, IS3 family, defective		
34	Rt	39984	40610	627	55.66	No significant similarity		
35	Rt	40788	41429	642	58.66	No significant similarity		
36	Dt	41503	41751	249	54.22	No significant similarity		
37	Dt	41841	42047	207	62.80	Transposase A IS3 family (<i>Yersinia enterocolitica</i>)	35	0.069
38	Dt	42117	42530	414	62.32	Transposase B, IS3 (<i>Yersinia enterocolitica</i>)	83	8e ⁻¹⁶
39	Dt	42906	42959	54	44.44	First half of finger motif	10 ^{-5.9d}	
40	Dt	43193	43612	420	54.29	No significant similarity		
41	Rt	43996	44802	807	47.96	No significant similarity		
42	Rt	44975	45712	738	52.98	No significant similarity		
43	Dt	45612	45734	123	51.22	No significant similarity		
44	Dt	46193	46273	81	61.73	No significant similarity		
45	Rt	46706	46759	54	55.56	TAT motif III	10 ^{5.1d}	
46	Rt	47268	47300	33	48.48	No significant similarity		
47	Rt	47839	48042	204	53.92	No significant similarity		
48	Dt	48072	48686	615	56.71	DNA polymerase III, epsilon subunit, putative (<i>Deinococcus radiodurans</i>)	80	1e ⁻¹⁴
49	Rt	49227	49325	99	55.56	No significant similarity		
50	Dt	49871	49924	54	59.26	No significant similarity		
51	Rt	50058	50219	162	53.16	No significant similarity		

^a Bold type shows the sequences of the four clones which were used to screen the cosmid library. CoA, coenzyme A. Right end and left end positions are base pair numbers.

^b Dt, direct typical; Rt, reverse typical.

^c E value, expectation value.

^d e-motif search.

3% difference between these two strains. A similar survey of several randomly selected UTI strains and those obtained from patients with cystic fibrosis suggests that these strains contain in their genomes approximately 3% unique DNA which is not present in PAO1 (data not shown). This suggests that the genomic variation among different *P. aeruginosa* isolates is less pronounced than that in different *E. coli* isolates. For example,

using an M13 library from a UTI strain of *E. coli* (R45), we have shown that 25% of these M13 clones do not correspond to sequences in the genome of *E. coli* DH5 α (unpublished observations), which is consistent with the observed size variation among the genomes of different *E. coli* isolates (1).

The ability to detect strain-specific DNA in a cloned library can be further exploited for the identification of contiguous

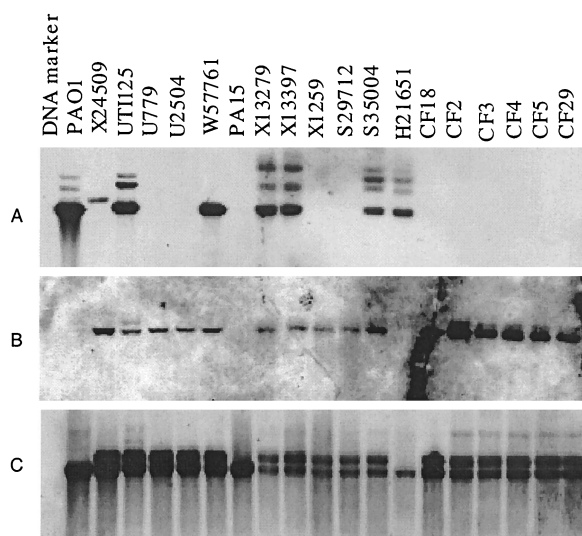


FIG. 5. Southern hybridization analysis of *P. aeruginosa* chromosomal DNA. Chromosomal DNA from different clinical isolates was digested with *SalI*, fractionated by gel electrophoresis, and blotted onto nylon membranes. The probes used were probe 1, a DNA fragment corresponding to the 5' portion of PAO2222 and the 3' portion of gene PAO2223 (Fig. 3) (A), probe 2, an internal PAGI-1 segment (corresponding to *orf18* in Fig. 4) (B), and probe 3, an internal region of *orf2* and *orf3* (C).

segments, called genomic islands, which are unstable DNA segments that usually carry determinants important for survival of the bacteria in unique environmental niches. Best characterized of these strain-specific chromosomal regions are the pathogenicity islands, which carry virulence genes (12). The presence of pathogenicity islands is often the only genetic difference between virulent and avirulent strains of the same species. Using differential DNA hybridization analysis we were able to identify DNA segments that are widely distributed among clinical isolates of *P. aeruginosa*. One such segment was sequenced and was termed PAGI-1. In our limited survey of various *P. aeruginosa* isolates, 85% of strains carried PAGI-1, which was absent from the genome of PAO1, a strain whose genome was recently sequenced (22). The sequence of PAGI-1 and its flanking sequences showed that the island is not located near any tRNA genes, a common property of pathogenicity islands. Moreover, PAGI-1 was apparently incorporated into the *P. aeruginosa* chromosome through recombination, replacing a preexisting 6,729-bp sequence that is present in most strains lacking the island, including PAO1. PAGI-1 contains two copies of insertion sequences that may facilitate the inter-bacterial transfer of the island. Interestingly, the deleted

6,729-bp region also specifies a transposase related to a homologue in *Rhizobium* sp. strain NGR234. Therefore this region, as well as PAGI-1, may have the capability to translocate into different regions of the chromosome. Our survey of various *P. aeruginosa* isolates by Southern blot analysis showed that in a number of strains (UTI125, W57761, X13273, X13397, and S35004) PAGI-1 is in different locations. However, the locus containing the 6,729 bp is a "hot spot" for integration of genomic islands, since we identified at least one strain (PA15) which contains another unique insert at this location. A 2,793-bp sequence (64.4% G+C) containing three ORFs of unknown function is present at this site in PA15 (data not shown).

PAGI-1 also encodes two transcriptional regulators. If expressed, the products of these genes can regulate the expression of not only genes in PAGI-1 but also those that are located in the chromosome, analogous to plasmid-encoded regulators of chromosomal genes (3). The product of *orf5* is a homologue of RpoN-dependent transcriptional activators. Examination of putative regulatory regions of PAGI-1 did not indicate the presence of any RpoN-dependent promoter sequences. Therefore, the putative *orf5*-encoded transcriptional activator does not regulate any genes on PAGI-1; its targets are very likely outside of the island. The acquisition of PAGI-1 by strain X24509 and several other *P. aeruginosa* strains involved the replacement of the 6,729-bp region with a concomitant loss of several genes, one of which encodes a transcriptional regulator of the LysR family (PAO2220 in the genome of *P. aeruginosa* PAO1). It is conceivable that the loss of this regulator may also influence the expression of genes which are located outside of the replaced segment.

Two genes, *orf2* and *orf3*, are homologues of *pqiA* and *pqiB*, respectively, a pair of paraquat-inducible genes of *E. coli* (16). In *E. coli*, *pqiA* and *pqiB* are under the control of the SoxS and SoxR regulators, which respond to redox-cycling agents capable of generating superoxide radicals in the cell. The precise function of PqiA and PqiB in the detoxification of superoxide radicals is not known. The genome of *P. aeruginosa* contains a pair of *pqiA* and *pqiB* homologues (PAO4690 and PAO4689). These genes show high levels of DNA sequence identity to *orf2* and *orf3* of PAGI-1. Southern blot analysis confirmed that all strains carry at least one copy of the sequence homologous to *orf3* (*pqiB* homologue), with all strains containing PAGI-1 showing the presence of an additional copy of this gene (Fig. 5). The homologues of *pqi* genes in PAGI-1 were most likely acquired from *P. aeruginosa* through a gene duplication event.

The G+C content of PAGI-1 showed a highly asymmetric distribution pattern (Table 4). The first 75% of the sequence,

TABLE 5. Gene annotation of PAO1 sequence replaced or disrupted by PAGI-1

Gene	Start site	End site	Direction	Homolog product
PAO2217	2437428	2439011	→	α-Ketoglutarate semialdehyde dehydrogenase
PAO2218	2440253	2439150	←	Conserved hypothetical
PAO2219	2441555	2440347	←	OpdE
PAO2220	2441771	2442691	→	OprR
PAO2221	2443161	2444366	→	Transposase
PAO2222	2445533	2444886	←	Hypothetical
PAO2223	2446564	2445545	←	Hypothetical

from *orf1* to *orf30*, has a G+C content of 63.7%, which is slightly less than that reported for most strains, including PAO1 (66%) (22). The remaining 25% of PAGI-1 sequence has an average G+C content of 54.9%, which is significantly below the chromosomal content. This analysis suggests that the portions of PAGI-1 have at least two different origins and that the island was very likely assembled in an ancestral bacterium, which was not necessarily *P. aeruginosa*. The final capture of PAGI-1 by *P. aeruginosa* through a process of recombination and replacement fixed PAGI-1 in the chromosome of the majority of strains.

The annotation of PAGI-1 provided several clues about the function of a number of gene products and the basis of selection for the acquisition of the island by the majority of *P. aeruginosa*. The presence of two homologues of the paraquat-inducible genes whose products may function in detoxification of reactive oxygen species and the large number of dehydrogenases which may provide reduced cofactors for the detoxification reactions suggest a role for the PAGI-1 genes in the protection of the bacteria against oxidative damage. Interestingly, there are two homologous copies of the paraquat-inducible genes in all strains carrying PAGI-1 and at least one copy in strains lacking the island, such as PAO1. It is conceivable that one or both of the transcriptional factors encoded in PAGI-1 may be involved in regulating the expression of a number of genes, including the duplicated copies of the paraquat-inducible genes. A comprehensive examination of global gene expression in strains carrying PAGI-1, and the potential involvement of the transcriptional factors encoded by the island and by the 6,729-bp replaced segment, is currently under way.

ACKNOWLEDGMENTS

We thank Steve Moseley for critical reading of the manuscript.

This work was supported by grant DK53369 from the National Institutes of Diabetes and Digestive Kidney Diseases and a grant from the Cystic Fibrosis Foundation to the University of Washington Genome Center.

REFERENCES

- Bergthorsson, U., and H. Ochman. 1995. Heterogeneity of genome sizes among natural isolates of *Escherichia coli*. *J. Bacteriol.* **177**:5784–5789.
- Bodey, G. P., R. Bolivar, V. Fainstein, and L. Jadeja. 1983. Infections caused by *Pseudomonas aeruginosa*. *Rev. Infect. Dis.* **5**:279–313.
- Caron, J., L. M. Couffield, and J. R. Scott. 1989. A plasmid-encoded regulatory gene, *ms*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *E. coli*. *Proc. Natl. Acad. Sci. USA* **86**:963–967.
- Cheng, K. H., S. L. Leung, H. W. Hoekman, W. H. Beekhuis, P. G. Mulder, A. J. Geerards, and A. Kijlstra. 1999. Incidence of contact-lens-associated microbial keratitis and its related morbidity. *Lancet* **354**:181–185.
- Cornelis, G. R., A. Boland, A. P. Boyd, C. Geuijen, M. Iriarte, C. Neyt, M. P. Sory, and I. Stainier. 1998. The virulence plasmid of *Yersinia*, an antihost genome. *Microbiol. Mol. Biol. Rev.* **62**:1315–1352.
- Cryz, S. J., Jr. 1984. *Pseudomonas aeruginosa* infections. Academic Press, Inc., New York, N.Y.
- Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**:175–185.
- Fleiszig, S. M., E. J. Lee, C. Wu, R. C. Andika, V. Vallas, M. Portoles, and D. W. Frank. 1998. Cytotoxic strains of *Pseudomonas aeruginosa* can damage the intact corneal surface *in vitro*. *CLAO J.* **24**:41–47.
- Geistlich, M., R. Losick, J. R. Turner, and R. N. Rao. 1992. Characterization of a novel regulatory gene governing the expression of a polyketide synthase gene in *Streptomyces ambofaciens*. *Mol. Microbiol.* **6**:2019–2029.
- Gilligan, P. H. 1991. Microbiology of airway disease in patients with cystic fibrosis. *Clin. Microbiol. Rev.* **4**:35–51.
- Govan, J. R., and V. Deretic. 1996. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol. Rev.* **60**:539–374.
- Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* **23**:1089–1097.
- Hancock, R. E. 1998. Resistance mechanisms in *Pseudomonas aeruginosa* and other nonfermentative gram-negative bacteria. *Clin. Infect. Dis.* **27**(Suppl. 1):S93–S99.
- Hayes, W. S., and M. Brodovskiy. 1998. How to interpret an anonymous bacterial genome: machine learning approach to gene identification. *Genome Res.* **8**:1154–1171.
- Horswill, A. R., and J. C. Escalante-Semerena. 1997. Propionate catabolism in *Salmonella typhimurium* LT2: two divergently transcribed units comprise the *ppp* locus at 8.5 centisomes, *pppR* encodes a member of the sigma-54 family of activators, and the *pppBCDE* genes constitute an operon. *J. Bacteriol.* **179**:928–940.
- Koh, Y. S., and J. H. Roe. 1995. Isolation of a novel paraquat-inducible (*pqi*) gene regulated by the *soxRS* locus in *Escherichia coli*. *J. Bacteriol.* **177**:2673–2678.
- Krogh, A., I. S. Mian, and D. Haussler. 1994. A hidden Markov model that finds genes in *E. coli* DNA. *Nucleic Acids Res.* **22**:4768–4778.
- Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh, and S. K. Stover. 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J. Bacteriol.* **178**:1274–1282.
- Mahillon, J., and M. Chandler. 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**:725–774.
- Romling, U., K. D. Schmidt, and B. Tummeler. 1997. Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. *J. Mol. Biol.* **271**:386–404.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrenner, M. J. Hickey, F. S. L. 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.-S. Wong, Z. Wu, 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.
- Tinsley, C. R., and X. N. Nassif. 1996. Analysis of the genetic differences between *Neisseria meningitidis* and *Neisseria gonorrhoeae*: two closely related bacteria expressing two different pathogenicities. *Proc. Natl. Acad. Sci. USA* **93**:11109–11114.
- Vindenes, H., and R. Bjerknes. 1995. Microbial colonization of large wounds. *Burns* **21**:575–579.
- Visca, P., F. Chiarini, A. Mansi, C. Vetriani, L. Serino, and N. Orsi. 1992. Virulence determinants in *Pseudomonas aeruginosa* strains from urinary tract infections. *Epidemiol. Infect.* **108**:323–336.
- Waldor, M. K. 1998. Bacteriophage biology and bacterial virulence. *Trends Microbiol.* **6**:295–297.