Acquisition and Evolution of the *exoU* Locus in *Pseudomonas aeruginosa*

Bridget R. Kulasekara,^{1†} Hemantha D. Kulasekara,^{1‡} Matthew C. Wolfgang,^{1§} Lisa Stevens,¹
Dara W. Frank,² and Stephen Lory^{1*}

Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115,¹ and Department of Microbiology and Molecular Genetics, Medical College of Wisconsin, 8701 Watertown Plank Rd., Milwaukee, Wisconsin 53226²

Received 31 December 2005/Accepted 22 March 2006

ExoU is a potent *Pseudomonas aeruginosa* cytotoxin translocated into host cells by the type III secretion system. A comparison of genomes of various *P. aeruginosa* strains showed that the ExoU determinant is found in the same polymorphic region of the chromosome near a tRNA^{Lys} gene, suggesting that *exoU* is a horizontally acquired virulence determinant. We used yeast recombinational cloning to characterize four distinct ExoU-encoding DNA segments. We then sequenced and annotated three of these four genomic regions. The sequence of the largest DNA segment, named ExoU island A, revealed many plasmid- and genomic island-associated genes, most of which have been conserved across a broad set of β - and γ -Proteobacteria. Comparison of the sequenced ExoU-encoding genomic islands to the corresponding PAO1 tRNA^{Lys}-linked genomic island, the pathogenicity islands of strain PA14, and pKLC102 of clone C strains allowed us to propose a mechanism for the origin and transmission of the ExoU determinant. The evolutionary history very likely involved transposition of the ExoU determinant onto a transmissible plasmid, followed by transfer of the plasmid into different *P. aeruginosa* strains. The plasmid subsequently integrated into a tRNA^{Lys} gene in the chromosome of each recipient, where it acquired insertion sequences and underwent deletions and rearrangements. We have also applied yeast recombinational cloning to facilitate a targeted mutagenesis of ExoU island A, further demonstrating the utility of the specific features of the yeast capture vector for functional analyses of genes on large horizontally acquired genetic elements.

Pseudomonas aeruginosa is a ubiquitous gram-negative bacterium with the ability to survive in diverse ecological niches. It is also an opportunistic pathogen that causes a wide variety of acute and chronic infections in individuals with compromised host defenses and those with cystic fibrosis (7). Infections by *P. aeruginosa* are important in a variety of clinical settings and account for 10% to 20% of nosocomial infections. The outcome of hospital-acquired pneumonia caused by *P. aeruginosa* is especially severe as it has a high mortality rate (3). The genome of this organism encodes a variety of secreted virulence factors utilizing dedicated secretion systems (46). Those secreted by the type III secretion system (TTSS) appear to play a major role in the outcome of infection, particularly for acute pneumonias (21, 41). The TTSS is a specialized protein-targeting system whereby the bacterial secreted factors (the so-called effectors) are directly introduced into the host cell through an injection-like apparatus. A number of gram-negative bacterial pathogens use this secretion mechanism, and based on the conserved sequences of the components of the secretion machineries, they appear to function using a conserved mechanism (37). The consequences of the action of the

TTSS are determined by the specific biochemical activities of the individual effectors on the host cell. Most strains of *P. aeruginosa* carry genes encoding a functional TTSS and a combination of three effectors (13, 48). Among these, genes encoding exoenzyme T (ExoT) and exoenzyme Y (ExoY) are conserved in the genomes of nearly all *P. aeruginosa* strains regardless of their origin, while most strains carry the gene for either ExoS or ExoU but not both (4, 13, 48). The carriage of *exoS* is more prevalent among *P. aeruginosa* isolates. ExoS is a bifunctional cytotoxic protein, with the carboxy-terminal half specifying an ADP-ribosyltransferase activity targeting the Ras family of small G proteins that causes cell death (17, 18). The amino-terminal domain is a GTPase-activating protein directed toward the Rho family of low-molecular-weight GTPase proteins (16). The *exoU* gene has been shown to occur in a significant fraction of strains responsible for hospital-acquired pneumonia, keratitis, and ear infections (21, 31, 45). ExoU is a potent cytotoxin with phospholipase A2 activity (43). Although the precise role of ExoU in the various infections by *P. aeruginosa* is not known, this effector has been associated with the development of septic shock in an animal model (29).

The basis for the incompatibility of *exoU* and *exoS* within the same *P. aeruginosa* genome is not known. Interestingly, *exoU* and *exoS* do not reside at the same locus (14, 22, 46). Several lines of evidence have indicated that the determinants for expression and secretion of ExoU (the *exoU* gene and the *spcU* gene encoding its cognate chaperone [15]) are located within a genomic island. Genomic islands are segments of bacterial genomes that have been acquired through horizontal gene transfer. They often are integrated at sites adjacent to tRNA genes, are flanked by direct repeats, display a percent G+C

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

† Present address: Seattle Biomedical Research Institute, 307 Westlake Avenue N, Seattle, WA 98109.

‡ Present address: Department of Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195.

§ Present address: Department of Microbiology and Immunology and Cystic Fibrosis/Pulmonary Research and Treatment Center, University of North Carolina, Chapel Hill, NC 27599.

and codon usage that are different from those of the core genome, and contain mobile genetic elements (9, 44). Genomic islands are subject to rearrangements, and over time, many of the genes originally present may be absent or mutated. Rearrangements may also result in translocation of a portion of an island to a different location on the chromosome (26).

Southern blot analysis of *exoU*-containing strains suggested that the *exoU/spcU* locus is linked to a highly polymorphic region (14). This observation implied that the *exoU* and *spcU* genes are located within a region of the chromosome associated with genomic plasticity, a conjecture that was also supported by microarray analysis of genomic differences between various *P. aeruginosa* strains (48). The percent G+C of *exoU/spcU* is 58.8, well below 66.7, the mean percent G+C of the PAO1 genome. Finally, in some isolates not carrying ExoU, the region of the core genome *exoU/spcU* borders (adjacent to the PA0988 homolog) is separated by less than 1 kbp from a tRNA gene (25).

Association of horizontally transferred DNA with tRNA or tmRNA genes is believed to be a result of the activities of specific integrases that carry out site-specific recombination between *attP* of the acquired genetic element and the 3' end of a tRNA gene (*attB*). As a result, the newly integrated genetic element is flanked by a tRNA gene (*attL*) and the 3' end of a tRNA gene (*attR*) (24). In *P. aeruginosa*, two identical tRNA^{Lys} genes adjacent to PAO1 homologs PA0976 and PA4541 can be sites for plasmid integration and are often associated with genomic islands (23, 25, 26). In strain PAO1, the 3' region adjacent to the open reading frame (ORF) PA0976 contains an 8.9-kbp genomic island (from PA0977 to PA0987) encoding nine hypothetical unknown proteins, a colicin protein, and a neighboring immunity protein (25, 46). We have previously used yeast recombinational cloning to demonstrate that in two unrelated *P. aeruginosa* strains, the *exoU/spcU* pair is found on an 81-kbp DNA segment flanked by tRNA^{Lys} and homologs of *P. aeruginosa* PAO1 genes PA0976 and PA0988. In strain PA14, this region is occupied by a 14-kbp genomic (pathogenicity) island called PAPI-2, containing the coding sequences for ExoU and SpcU (23).

In certain isolates of *P. aeruginosa*, this tRNA^{Lys} site proximal to the PA0976 homolog serves as a site for the reversible integration of plasmids. For example, in strains of lineage K (a common clone found in Europe), plasmid pKLK106 can be integrated either at this tRNA^{Lys} site or at an identical tRNA^{Lys} gene located adjacent to the PA4541 homolog. In *P. aeruginosa* clone C strains (another common clone found in Europe), this tRNA^{Lys} site proximal to the PA0976 homolog is occupied by the genomic island PAGI-4. However, it was found that in these same strains, plasmid pKLC102, closely related to pKLK106, can integrate into the tRNA^{Lys} gene adjacent to the PA4541 homolog (26). The integration event of pKLK106 or pKLC102 in all strains is accompanied by a direct repeat flanking the integrated plasmids, consisting of the 3' end of the tRNA^{Lys} gene (25).

In this paper, we have probed the chromosomal regions that *exoU* and *spcU* occupy in four strains isolated from geographically distant locations. Using yeast recombinational cloning, four additional unique *exoU/spcU*-containing genomic islands were identified, and three were sequenced and annotated. This allowed us to deduce the evolutionary history of these islands

that may have originated from a single ancestral mobile genetic element.

MATERIALS AND METHODS

Strains and growth conditions. The source of each *P. aeruginosa* strain, containing *exoU*, is as follows: strains 6077 and 19660 are from ocular infections from Berkeley, CA; strain X13273 is a blood isolate from Seattle, WA; strains S54485, JJ692, and U2504 are urinary tract isolates from Seattle, Washington, Minneapolis, Minnesota, and Gainesville, Florida, respectively. Antibiotic concentrations used for *Escherichia coli* were 10 µg/ml for gentamicin, 5 µg/ml for tetracycline, and 50 µg/ml for streptomycin. For *P. aeruginosa*, gentamicin was used at a concentration of 30 µg/ml and tetracycline was used at a concentration of 60 µg/ml. Luria broth (LB) and LB agar were used if the medium used is not indicated.

Yeast recombinational cloning. The recombinational cloning vector pLLX13 and techniques used have been described by Wolfgang et al. (48). pLLX13 was modified to contain the flanking targeting sequences of the tRNA^{Lys}-associated hypervariable region, including PA0975 and PA0989, and is called p0975-0989capture (48). Following successful recombinational cloning, the vectors were referred to by the name of the strain used for cloning, followed by "pCap0976.1."

Generation of transposon insertions in pCap0976.1:6077 and introduction of mutated genes into the JJ692 chromosome. The *mariner*-based transposon, on plasmid pBT20, was used to generate transposon insertions (28). A library of random transposon insertions in pCap0976.1:6077 was generated by combining *Escherichia coli* GeneHogs/pCap0976.1:6077 in phosphate-buffered saline (PBS) at an optical density at 600 nm (OD₆₀₀) of 20, with *E. coli* SM10 λ pir/pBT20 at an OD₆₀₀ of 40. Five 50-µl amounts were incubated on dry LB agar plates at 37°C for 2 h. Spots were resuspended in 1 ml of 1× PBS, and 200 µl was plated onto LB plates containing streptomycin and gentamicin. A total of 40,000 colonies were recovered and used for purification of plasmid DNA. The resulting plasmids were electroporated into *E. coli* GeneHogs, and transformants were plated on gentamicin- and tetracycline-containing media to select for pCap0976.1:6077-containing transposon insertions. Thirty-two *E. coli* GeneHogs p0976.1:6077::BT20 isolates were individually mated into *P. aeruginosa* JJ692/pSW(I-SceI) (50) and resuspended in 1× PBS. Dilutions were plated onto minimal medium A with gentamicin (30 µg/ml) agar plates (6) in order to isolate single colonies. Double recombination events were verified by loss of tetracycline resistance present on the vector. Insertions were mapped using semirandom PCR (28) and sequencing of each amplicon.

Sequencing of the ExoU islands. ExoU island B was sequenced using randomly subcloned pUC19 libraries, generated from the *P. aeruginosa*-specific portion of pCap0976.1:19660. Sequence gaps were closed using custom primers. To sequence ExoU island A, a cosmid library of 6,077 genomic DNAs was first constructed, consisting of ~40-kbp inserts in the sCos-DBI vector (49). Cosmids were screened for the presence of *exoU* and EXA24 by using colony hybridization of PCR products specific for these two chromosomal regions. Two cosmids hybridizing to the probes were identified that spanned from 31,460 bp to 80,000 bp of ExoU island A. Two random M13 libraries were constructed from these two cosmids and were subsequently sequenced. The remaining portion was sequenced by creating a pUC19 library using the DNA from pCap0976.1:6077. As before, sequencing gaps were closed using custom primers. The sequence of ExoU island C was generated by sequencing EcoRI and PstI fragments of pCap0976.1:X13273 subcloned into pUC19 and by using custom primers. Sequencing reads were assembled and viewed using Phred, Crossmatch, Phrap, and Consed (10, 11, 19).

ORF prediction and annotation. Putative ORFs were predicted using, first, Glimmer (42) trained on the *P. aeruginosa* PAO1 complete genome sequence (46) and, subsequently, GeneMark (32). The minimum number of amino acids used for a predicted coding sequence was 30. The first possible start codon was used for all ORFs except when an alternate start codon was used by homologous coding sequences. Predicted protein sequences were annotated using the BLAST algorithm (1) to search the complete NCBI database on 5 March 2006. CDD (33) was used to find putative domains. Only conserved domains producing alignments greater than 65% are included. If more than one significant alignment was produced per protein sequence, then the domain with the smallest *E* value (expected number of high-scoring segmented pairs with a score greater than or equal to 5 [1]) is listed.

PFGE. *P. aeruginosa* strains were grown to an OD₆₀₀ of 0.5 to 1.5 in LB. Cultures were resuspended in an equal volume of phosphate-buffered saline and mixed with equal volumes of 1% low-melting-point agarose dissolved in phosphate-buffered saline. Eighty-microliter volumes were added to individual plug molds. Once solidified, the plugs were lysed as described by Romling et al. (40), after which the lysis solution was replaced with 10 mM Tris-Cl and 10 mM EDTA. The plugs were washed once in water and three times in SpeI digestion

buffer (New England Biolabs) at 55°C for 1 h each, after which 30 units of SpeI was added to the plug combined with 300 μ l of SpeI buffer and incubated overnight. Half of a plug was used for pulsed-field gel electrophoresis (PFGE). The marker used was the Lambda Ladder PFG marker from New England Biolabs. Conditions for PFGE were 1% agar and 0.5 $\times$ Tris-borate-EDTA buffer at 14°C and 6 V per cm. The initial switch time was 0.22 s, and the final switch time was 54.17 s, with an angle of 120° and a total time of 25.6 h.

Nucleotide sequence accession numbers. The nucleotide sequences of ExoU islands A, B, and C have been deposited in the GenBank database and assigned the accession numbers DQ437742, DQ437743, and DQ437744, respectively.

RESULTS

Genomic variability among ExoU/SpcU-containing *P. aeruginosa* isolates. Six *P. aeruginosa* strains, 19660 (strain 1), X13273 (strain 2), S54485 (strain 3), 6077 (strain 4), JJ692 (strain 5), and U2504 (strain 6), were chosen for analysis of their ExoU/SpcU-encoding loci. These six strains were the ExoU-containing representatives from an 18-member panel, analyzed previously for genome variability by hybridization of genomic DNA to a *P. aeruginosa* PAO1 whole-genome DNA microarray (48). The phylogenetic relationships of the 18 strains used in this study showed a limited clonal association between the *exoU/spcU*-containing isolates and the rest of the strains carrying the *exoS* gene. Five of the six *exoU/spcU*-containing strains grouped within a distinct cluster, and three strains within this branch (strains 4 to 6) belonged to a separate cluster (with strains 4 and 6 being the most closely related). To further assess the relatedness of these strains, we employed PFGE of chromosomal DNA digested with the restriction endonuclease SpeI (Fig. 1A). Based on this analysis, strains 4 and 6 appear to be clonal variants as the majority of the restriction fragments are conserved. This result is thus in close agreement with the earlier microarray analysis (48).

Capture of the *exoU/spcU*-containing hypervariable segments by yeast recombinational cloning. We have previously shown that in strains 4 and 5, the 81-kbp DNA segment, designated ExoU island A, occupies the PAO1 PA0977-to-PA0987 chromosomal region.

We employed yeast recombinational cloning (39, 48) with a capture vector carrying 1-kbp targeting sequences from PA0975 to PA0976 and PA0988 to PA0989 to isolate the *exoU/spcU* locus from four *P. aeruginosa* strains (strains 1, 2, 3, and 6). Restriction endonuclease (NcoI) fingerprinting revealed that strain 6 (in addition to strains 4 and 5) contains ExoU island A.

These data, together with those of the earlier hybridization study, indicate that strains 4 to 6 may have evolved from the same ancestral strain in spite of their origins from three geographically distant locations within the United States. The remaining three strains did not carry the 81-kbp ExoU island A and instead carried insertions of 60 kbp (strain 3), 32 kbp (strain 1), and 6 kbp (strain 2).

Sequence analysis of ExoU island A. We sequenced and annotated ExoU island A from strain 4. The 81.17-kbp segment is predicted to encode 77 open reading frames (Table 1 and Fig. 2). As with other genomic islands, the percent G+C (57.0) differed from that of the core genome (PAO1, 66.7). Another feature of ExoU island A which is common in genomic islands that are integrated next to a tRNA gene is the presence of a gene carrying an integrase (*int*), presumed to be responsible for incorporation of genetic elements into the tRNA^{Lys} gene (47). The ExoU island A *int* gene is almost

FIG. 1. Molecular analysis of the genome content of six ExoU-encoding strains. (A) Pulsed-field gel electrophoresis of SpeI-treated chromosomal DNA from the six strains. The assigned number of each strain is listed above the respective lane. (B) NcoI fingerprints of the recombinational cloning vector containing the captured regions between PA0975 and PA0989 homologs from six strains. Lanes are labeled with the assigned number of the strain that was used for cloning. The bands from vector DNA are indicated with an asterisk.

identical to the integrase gene that was identified at the same location on the reversibly integrating plasmids pKLC102 and pKLC106 (25, 26). Almost identical *int* genes also exist at the same location on genomic islands PAPI-1 and PAPI-2, both integrated at tRNA^{Lys} genes in PA14 (23). A genetic pattern common to the tRNA-associated genomic islands of *P. aeruginosa* is created following integration of pKLC102 into the tRNA^{Lys} gene in clone C strains. On the episomal form of pKLC102, *attP* is located between *xerC* (*int* gene) and the gene encoding a homolog of the chromosomal partitioning factor Soj. Following incorporation, *xerC* borders the end of the integrated plasmid proximal to the tRNA gene (*attL*), and *soj* is located at the opposite end, bordered by a duplicated 3' end of the tRNA gene (*attR*). This genetic organization is preserved

TABLE 1. Description of the features of ExoU island A

EXA ORF	Annotation result(s)	Orientation	Start codon position	Stop codon position	Amino acid length	G+C (%)	Accession no. of closest homolog	Conserved domain search result
	tRNA ^{Lys}	N/A ^a	1	76	N/A	N/A	N/A	N/A
	TAGI repeat	N/A	202	221	N/A	45	N/A	N/A
1a	Putative integrase	—	1729	446	427	61	AAP84129	None
1b	Putative regulator of excisionase activity	+	1113	1822	229	61	AAP94702	None
2	Conserved hypothetical protein	—	3657	1726	643	58	AAP84130	None
3	Conserved hypothetical protein	—	5855	3894	653	51	ABA73546	None
4	Hypothetical protein	—	6280	5891	129	49	AAN29662	None
5	Hypothetical protein	—	6530	6249	93	52	None	None
6	UvrD/REP DNA helicase	—	9050	6768	760	59	EAR47340	COG0210
7	Hypothetical protein	—	10423	9239	394	58	ABC92876	None
8	DNA helicase-related protein	—	17105	10440	2,221	60	AAM41383	cd00221
	IS407 left-terminal inverted repeat (67% identical)	N/A	17512	17560	N/A	N/A	N/A	N/A
9	IS407 transposase OrfA similar to PA0986	+	17580	17843	87	54	AAG04375	Pfam01527
	IS407-associated repeat conserved with positions 53587-53974 (inverted) and 77997-78377	N/A	18358	18688	N/A	57	N/A	N/A
10	N terminus of IS407 transposase OrfB similar to the N terminus of PA0987	+	17867	18388	173	61	ZP_00970131	None
11	C terminus of IS407 transposase OrfB similar to the C terminus of PA0987	+	18421	18717	98	55	ZP_00968899	None
12	Hypothetical protein	—	19245	19015	76	52	None	None
13	Hypothetical protein with homology to the C terminus of colicin M	+	19450	20319	289	47	AAO54114	None
14	Hypothetical protein	+	20375	20803	142	42	None	None
15	Putative plasmid stabilization factor	—	21214	20840	124	60	ZP_00971292	Pfam05016
16	Putative transcriptional regulator	—	21550	21218	110	59	AAP22603	COG3609
17	Conserved hypothetical protein	—	23472	21958	504	59	ZP_00965522	None
18	Hypothetical protein	—	23822	23469	117	61	ZP_00965523	None
19	Conserved hypothetical protein	—	25204	23822	460	64	AAP22580	None
20	Conserved hypothetical protein	—	26160	25222	312	64	ZP_00965525	None
21	Conserved hypothetical protein	—	26591	26160	143	61	AAP84145	None
22	Hypothetical protein similar to PA0980	+	26972	27256	94	46	AAP82951	None
23	Conserved hypothetical protein similar to PA0981	+	27287	27916	209	46	AAG04370	None
	Repeat conserved with positions 34864-34902 (inverted)	N/A	28466	28504	N/A	51	N/A	N/A
24	Tn721 transposase, truncated	—	31463	28500	987	65	CAG15097	Pfam01526
25	Hypothetical protein similar to PA2223	+	31697	32725	342	48	AAP84179	None
26	Hypothetical protein similar to PA2222	+	32718	33404	228	50	AAP84180	None
27	Hypothetical protein similar to PA2224	+	33436	34161	241	50	AAP84181	None
28	Tn721 resolvase, C terminus	—	34470	34216	84	66	CAG15096	Pfam02796
29	Tn721 resolvase, N terminus	—	34774	34505	89	63	CAG15096	None
	Repeat conserved with positions 28466-28504 (inverted)	N/A	34864	34902	None	54	None	None
30	Hypothetical protein	+	35174	35392	72	56	AAP82955	None
31	Conserved hypothetical protein similar to PA0982	—	36048	35389	219	61	AAP84147	None
32	Conserved hypothetical protein	—	36329	36045	94	58	ZP_00971301	None
33	Conserved hypothetical protein	—	39268	36326	980	63	AAP84149	None
34	Conserved hypothetical protein	—	39711	39268	147	64	AAP84150	None
35	Conserved hypothetical protein	—	41194	39689	501	63	ZP_00971305	None
36	Conserved hypothetical protein	—	42062	41178	294	67	AAP22570	None
37	Conserved hypothetical protein	—	42718	42059	219	60	AAP84153	None
38	Conserved hypothetical protein	—	43101	42715	128	66	AAP84154	None
39	Conserved hypothetical protein	—	43468	43112	118	59	AAP84155	None
40	Conserved hypothetical protein	—	43725	43486	79	65	AAP84156	None
41	Conserved hypothetical protein	—	44060	43722	112	67	AAP84157	None
42	Conserved hypothetical protein	—	44453	44154	99	57	ZP_00971311	None
43	Conserved hypothetical protein	+	44626	44937	103	49	AAP84159	None
44	Hypothetical protein	—	46090	44981	369	47	AAP22563	None
45	Conserved hypothetical protein	—	47702	46221	493	59	AAP84161	None
46	Conserved hypothetical protein	—	48459	47713	248	62	AAP84172	None
47	TraG/TraD family protein	—	50690	48459	743	64	ZP_00965538	Pfam02534
48	Hypothetical protein	—	50963	50694	89	61	ZP_00965539	None
49	Conserved hypothetical protein	—	51472	50972	166	66	AAP22559	None
50	Conserved hypothetical protein	—	52050	51469	193	64	AAP84175	None
51	Conserved hypothetical protein	—	52790	52035	251	65	ZP_00965541	None
52	Conserved hypothetical protein	—	53481	52801	226	63	AAP84178	None
	IS407-associated repeat (inverted) conserved with positions 18358-18688 and 77997-78377	N/A	53587	53974	None	56	None	None
53	Transposase-derived hypothetical protein	—	54557	53610	315	58	CAI46956	None
54	Hypothetical protein	+	54534	54707	57	56	AAO64283	None
55	KatA, catalase isozyme A	+	55063	56511	482	61	AAO7624	cd00328
56	Conserved hypothetical protein	+	56582	56755	57	43	AAN67283	None
57	Alpha/beta hydrolase family protein	—	57751	56894	285	44	EAR24768	COG0596
58	Putative glutathione S-transferase	+	57782	58393	203	44	EAN67479	COG0625
59	Pirin-related protein	+	58761	59501	246	44	CAD17595	COG1741

Continued on following page

TABLE 1—Continued

EXA ORF	Annotation result(s)	Orientation	Start codon position	Stop codon position	Amino acid length	G+C (%)	Accession no. of closest homolog	Conserved domain search result
60	Amidohydrolase family protein	+	59588	60187	199	49	EAN17461	CD01012
61	Putative outer membrane channel lipoprotein	+	60378	61859	493	46	CAD14996	COG1538
62	Putative multidrug resistance protein A	+	61849	63057	402	45	ABC37909	COG1566
63	Putative multidrug resistance protein B	+	63073	64593	506	42	AAQ58753	None
64	Putative LysR binding protein	+	64991	65881	296	40	BAC51446	Pfam03466
65	Pseudogene, halogenase PltM	—	66762	66173	None	62	None	None
66	Pseudogene, transcriptional regulator PltR	—	67497	66898	None	61	None	None
	5' portion of PA0979	N/A	67616	67811	None	58	None	None
67	Conserved hypothetical protein	—	69988	67913	691	64	ABB05606	None
68	HlyD/EmrA family membrane protein	—	71012	69933	359	68	ABB05605	COG1566
69	Hypothetical protein	—	71294	71025	89	62	ABB05604	None
70	Hypothetical protein	+	71550	72785	411	64	ZP_00283098	None
71	Hypothetical protein with homology to the C terminus of PA0710	+	73307	73495	62	47	AAG04099	None
72	Pseudogene, helicase family protein	—	75860	73596	None	63	ZP_00965545	
73	O-methyl transferase family protein	—	77404	75959	481	64	AAP84189	COG4123
74	Conserved hypothetical protein, C terminus	—	77646	77434	70	64	ZP_00965547	None
	IS407 left-terminal inverted repeat (65% identical)	N/A	77714	77762	None	51	None	None
75	Transposase-derived hypothetical protein	+	77667	78344	225	61	AAC16022	None
	IS407-associated repeat conserved with positions 18358–18688 and 53587–53974 (inverted)	N/A	77997	78377	None	59	None	None
76	ExoU, type III secretion effector protein	+	78481	80544	688	59	AAC16023	Pfam01734
77	SpcU, ExoU chaperone	+	80541	80954	137	56	AAC16024	None
	TAGI repeat	N/A	81117	81136	N/A	45	N/A	N/A
N/A	PA0988	+	81347	N/A	N/A	N/A	N/A	N/A

^a N/A, not applicable.

in the highly homologous PAPI-1 and, additionally, in PAGI-2 and PAGI-3 (23, 26, 30). In ExoU island A, a duplication of the 3' end of the tRNA^{Lys} gene at the opposite border and *soj* are absent, suggesting that the ancestral form of ExoU island A had undergone a deletion event following its chromosomal integration, resulting in the loss of the partial tRNA gene and *soj*. Deletion of these features may have had a stabilizing effect on the ancestral ExoU island A, fixing it permanently into the chromosome. The absence of a repeated partial tRNA gene would make excision from the chromosome, dependent on the partial tRNA duplication, impossible (47). Additionally, *soj* is homologous with a chromosomal partitioning gene that is likely important for the maintenance of a large self-replicating DNA element, such as the ancestral ExoU island plasmid.

A unique 20-bp sequence flanks ExoU island A (Fig. 3). This sequence is 125 bp downstream of the tRNA^{Lys} gene next to the PA0976 homolog (Fig. 2) and, on the opposite end of the island, is 172 bp downstream of *spcU*. A similar flanking sequence is observed in numerous genomic islands associated with the tRNA^{Lys} gene. A sequence 90% identical to this repeat flanks PAPI-1 and pKLC102. In both cases, it is approximately 125 bp downstream from tRNA^{Lys} (*attL*) and, on the opposite end, is only 29 bp downstream from *attR*. A sequence highly similar to the above-mentioned repeats also flanks the PAO1 tRNA^{Lys}-associated genomic island, where it is also approximately 125 bp downstream from tRNA^{Lys} (*attL*) but is 123 bp downstream from *attR*. A similar sequence is present 123 bp downstream of an unoccupied tRNA^{Lys} (*attB*) that is proximal to the PA0976 homolog of strain K2 and that has served as an integration site for pKLC106 (accession number AF285422). This sequence also overlaps with 1 bp of the 3' portion of tRNA^{Lys} proximal to PA4541 in strain PAO1. Given the ubiquitous presence of this repeat in tRNA^{Lys}-associated genomic islands, we have named it the tRNA^{Lys}-associated genomic island (TAGI) repeat. We speculate that the TAGI

repeat sequence may be an accessory element necessary for integration and/or excision of these islands.

The relationship between ExoU island A and other *P. aeruginosa* tRNA^{Lys}-associated genomic islands is further verified by examining ORFs encoded by ExoU island A. Homology searches of predicted protein sequences encoded by ExoU island A reveal a number of encoded proteins with unknown function, specifically EXA32 to EXA40, EXA43 to EXA44, and EXA46 to EXA48, that share both homology and synteny with a subset of genes located on the *P. aeruginosa* plasmid pKLC102 (Fig. 4) and genomic islands PAPI-1, PAGI-2, and PAGI-3 (23, 26, 30). Moreover, this set of genes is widely conserved in at least 10 species belonging to β - and γ -*Proteobacteria*, ranging from *Ralstonia metallidurans* to *Haemophilus influenzae*, and belongs to a larger set, comprised of 33 ORFs found among this group of *Proteobacteria* (34). Additionally, these conserved coding sequences are often located on genomic islands that in most cases share other characteristic features, such as association with a tRNA gene and the presence of an integrase. It has been postulated that this set of 33 ORFs may be involved in plasmid maintenance or horizontal gene transfer (26). Additional proteins encoded by ExoU island A are clearly associated with plasmid maintenance and transmission, based on their sequence similarity to proteins with known plasmid-associated functions. These include a putative plasmid stabilization factor (EXA15), several putative helicases (EXA3, EXA6, EXA8, EXA45, and EXA72), and a TraG/TraD family protein (EXA47).

In addition to the absence of an *attR* site and *soj*, the presence of pseudogenes and partial components of operons also suggest that rearrangements and deletions have occurred. For example, two genes in ExoU island A are homologs of the *Pseudomonas fluorescens* biosynthesis operon for the antifungal compound pyoluteorin (36). As in *P. fluorescens*, these two genes are located next to each other and are transcribed in the

FIG. 2. A schematic representation of the features of ExoU island A. Predicted genes and their relationships to horizontally transmissible genetic elements discussed in the text are depicted by arrows, pointing in the direction of transcription. Patterns and colors are assigned based on homology. Percent G+C is shown, based on a sliding 75-bp window. The gray shaded area indicates sequence conservation of the island borders with the core genome.

FIG. 3. Alignment of the TAGI repeats with repeats flanking *exoS* and the sequence marking the *exoS* deletion in *exoU*-carrying strains. *, K2 *attB* is located next to the PA0976 homolog. **, In PAO1, this putative *attB* site is proximal to PA4541. Gray shading denotes nucleotides that are not conserved with the consensus sequence. The boxed region indicates the repeated sequence associated with *exoS*. Sequences for PAPI-1 were taken from a sequence with accession number AY273869. The coordinates for the left repeat are from 925 to 944, and the coordinates for the right repeat are from 108,786 to 108,805. The sequence for the left repeat of pKLC102 was taken from the sequence with accession number AF285426, and the coordinates are from 256 to 275. The sequence for the right repeat of pKLC102 was taken from the sequence with accession number AF285425, and the coordinates are from 507 to 526. The sequence from the PAO1 *attB* site was taken from the sequence with accession number AE004868 and the coordinates are from 117 to 98 (reverse and complement). The sequences from the sites flanking *exoS* in PAO1 were taken from accession number AE004801. For the left repeat, the coordinates are from 8,661 to 8,642 (reverse and complement), and for the right repeat, the coordinates are from 7,210 to 7,191 (reverse and complement). The remaining PAO1 sequences from the tRNA^{Lys}-associated genomic island were taken from accession number AE004531. The coordinates for the left repeat are 1,812 to 1,831, and the coordinates for the right repeat are from 10,759 to 10,778. The sequence from PA14 was taken from ABQ07000001 and had coordinates from 162,446 to 162,475.

same direction; however, the eight remaining ORFs present in *P. fluorescens* are absent in ExoU island A. Genomic islands often encode several transposons and insertion sequences. Duplicated insertion sequences facilitate rearrangements and deletions in genomic islands (20) and have

been shown to cause large-scale chromosomal inversions in *P. aeruginosa* (27). Several genes with homology to transposons or insertion sequences were identified in ExoU island A. Three ORFs, EXA9, EXA10, and EXA11, are homologous to the two genes comprising the IS407 element whose sequences in

FIG. 4. An illustration of the homology and synteny between PAPI-1, pKLC102, and ExoU island A. Homologies were determined using the Pustell Matrix feature of MacVector, using a similarity score of 70% and a window size of 60 or 70 bp.

TABLE 2. Description of the features of ExoU island B

EXB ORF	Annotation result(s)	Orientation	Start codon position	Stop codon position	Amino acid length	G+C (%)	Accession no. of closest homolog	Conserved domain search result
1	Hypothetical protein similar to PA0977 TAGI repeat	— N/A ^a	476 200	198 219	92 N/A	57 45	AAG04366 N/A	None N/A
2	Hypothetical protein similar to the C terminus of an integrase	—	952	443	169	61	AAP84129	None
3	Transposase-derived hypothetical protein	—	1142	957	61	54	ZP_00971320	None
4	Putative IS30 family transposase, N terminus	+	1498	2232	244	57	ZP_00463497	None
5	Putative IS30 family transposase, C terminus	+	2266	2847	193	55	ZP_00463497	Pfam00665
6	Transposase-derived hypothetical protein	—	3187	2813	124	61	ZP_00971320	None
7	Transposase-derived hypothetical protein	—	3435	3187	82	66	AAV82054	None
8	Putative transposase	—	3731	3432	99	60	ZP_00971321	Pfam01527
9	Hypothetical protein similar to PA0980	+	3918	4202	94	47	AAP82951	None
10	Hypothetical protein similar to PA0981	+	4239	4862	207	46	AAG04370	None
11	Hypothetical protein similar to RS08 from PAPI-2	—	5288	4968	106	50	AAP82953	None
12	Hypothetical protein similar to RS09 from PAPI-2	+	5400	5603	67	57	AAP82954	None
13	Hypothetical protein similar to RL019 from PAPI-1	+	5647	5877	76	55	AAP84146	None
14	Hypothetical protein similar to the C terminus of PA0982	—	6047	5874	57	60	AAP22574	None
15	GNAT family acetyltransferase	+	6091	6540	149	53	AAP82956	Pfam00583
16	ISPPu14 transposase Orf3, C terminus	—	7616	6543	357	59	NP_746094	Pfam03050
17	ISPPu14 transposase Orf3, N terminus	—	8122	7577	181	59	NP_746094	None
18	ISPPu14 transposase Orf2	—	8482	8142	111	64	NP_746095	Pfam05717
19	ISPPu14 transposase Orf1	—	8724	8479	81	58	NP_746109	None
20	Hypothetical protein similar to the N terminus of PA0982	—	9305	8859	178	61	AAG04371	None
21	Putative transposase similar to PA0983	+	9340	9627	95	57	AAG04372	Pfam01527
22	Colicin immunity protein similar to PA0984	+	10168	10494	108	41	AAP84212	None
23	Putative pyocin S5 similar to PA0985	—	12015	10519	498	46	AAP84213	None
24	Hypothetical protein	+	12348	12752	134	58	None	None
	IS407 left-terminal inverted repeat (69% identical)	N/A	13022	13070	N/A	53	N/A	N/A
25	Hypothetical protein similar to IS407 transposase OrfB and PA0987	+	13134	13856	240	62	ZP_00138569	Pfam00665
26	ExoU, type III effector protein	+	13981	16044	687	59	AAC16023	Pfam01734
27	SpcU, ExoU chaperone	+	16041	16454	137	56	AAC16024	None
28	RNA polymerase sigma-70 factor, extracytoplasmic function family	—	17275	16691	194	59	EAP51975	COG1595
29	Hypothetical protein	—	17705	17583	40	47	None	None
30	Hypothetical protein	—	18363	18157	68	54	None	None
31	Hypothetical protein	—	18670	19020	116	66	None	None
32	Putative oxygen-independent coproporphyrinogen III oxidase	+	19438	20862	474	63	EAP35559	COG0635
33	Putative nitric oxide reductase transcriptional regulator	—	22498	20864	544	68	NP_522520.1	COG3604
34	Putative quinol-dependent nitric oxide reductase	+	22625	24901	758	65	ABC27611	COG3256
35	Transposase-derived hypothetical protein	+	25262	25504	80	56	AAP82949	None
36	Conserved hypothetical protein	—	26346	25681	221	61	BAE50869	None
37	Putative regulator of mercury resistance proteins	+	26375	26902	148	57	AAD40337	cd01108
38	Hypothetical protein	—	27321	27163	164	60	None	None
39	Putative transcriptional regulator	—	27983	27558	141	49	EAP09715	COG1733
40	Hypothetical protein	+	27996	28118	40	52	None	None
41	NADH:flavin oxidoreductase family protein	+	28326	29462	378	54	ZP_00136272	cd02933
	TAGI repeat	N/A	29794	29813	N/A	45	N/A	N/A
42	PA0988	+	30024	N/A	N/A	N/A	N/A	N/A

^a N/A, not applicable.

the PAO1 genome are represented by PA0986 and PA0987. EXA9 is homologous to PA0986, EXA10 is homologous to the 5' portion of PA0987, and EXA11 is homologous to the 3' portion of PA0987. IS407 has been identified in other species, such as *Burkholderia cepacia* (51) and *Burkholderia mallei* (8), in addition to being found in locations such as *P. aeruginosa* O-antigen biosynthetic clusters (39, 46). A 392-bp sequence, inclusive of the segment homologous to the 3' portion of PA0987, is repeated three times within this island (Fig. 2). The sequence is present as two direct repeats, inclusive of EXA11 and EXA75, and one inverted repeat, inclusive of EXA53. The three repeated sequences together share 74% nucleotide identity. Additionally, a 6.4-kbp sequence that apparently contains a remnant of a transposon (EXA24 to EXA29) is present in ExoU island A

and is flanked by a 39-bp, 90%-conserved, inverted repeat. The region between the repeats includes a truncated transposase gene, EXA24, and two ORFs with homology to the 5' and 3' portions of a resolvase, EXA28 and EXA29.

Sequence analysis of ExoU island B. To determine whether *exoU* is at a conserved location, we used PCR to amplify the region that spans *exoU* and PA0988. All six *exoU*-carrying strains, with the exception of strain 1, gave identically sized products (data not shown). The absence of an amplicon suggests that in strain 1, the *exoU* gene is located at a unique site. We therefore sequenced the captured island from this strain and named it ExoU island B. Sequencing of ExoU island B showed that the 29.85-kbp segment contains 41 predicted ORFs and has a mean percent G+C of 56.8 (Table 2 and Fig. 5).

FIG. 5. An illustration of the features of ExoU island B and ExoU island C. The organization of this figure is the same as that described in the legend to Fig. 2.

In strain 1, the *exoU/spcU* pair resides in a different location relative to the PA0988 homolog from that of most *exoU*-containing strains, and *exoU* and *spcU* are located at approximately equal distances from either junction with PAO1 homologous DNA. Comparison of ExoU island B to the other ExoU-encoding islands reveals that ExoU island B is the most similar to PAPI-2, as the two islands contain blocks of homologous DNA segments. As with PAPI-2 and ExoU island A, ExoU island B contains many genes that are homologous to the region between PA0976 and PA0988 in PAO1. ExoU island B contains a truncated homologue of PA0977, EXB1, and homologues of PA0980 through PA0985 (EXB9, EXB10, EXB14, and EXB20-23). The homologue of PA0982 has apparently been disrupted by an IS_{Ppu14}-type transposon common in *Pseudomonas putida* (35), resulting in EXB14 and EXB20. Also, less

conserved is a portion of IS407, homologous to PA0987, that is present in ExoU island A.

ExoU island B encodes the C-terminal end of an integrase (EXB2) highly similar to the integrase located on other tRNA^{Lys}-associated genomic islands, pKLC102, pKLC106, PAPI-2, and ExoU island A. This portion of an integrase is approximately 30 amino acids longer than the C-terminal portion of the integrase located on the analogous PAO1 genomic island. Other genes that have DNA-associated or plasmid-related functions or are conserved hypothetical genes located on other *P. aeruginosa* tRNA-associated genomic islands are not located on this island. However, the unique 20-bp TAGI repeat that flanks ExoU island A, PAPI-1, and pKLC102 also flanks ExoU island B.

ExoU island B encodes several proteins that share sequence

FIG. 6. (A) A schematic description of the technique used to introduce captured segment-specific *mariner* transposon insertions into the bacterial chromosome. TS1 and TS2 are targeting sequences used to capture a specific chromosomal segment by yeast recombinational cloning, *bla* is the carbenicillin resistance determinant, *tetR* is the tetracycline resistance determinant, *cyh* is the cycloheximide resistance determinant, *oriT* is the origin of transfer, and pI-SceI is the plasmid carrying the gene for the I-SceI restriction endonuclease. (B) Illustration of the *mariner* insertions in ExoU island A marked by arrows.

similarity to proteins of known function. A putative nitric oxide reductase, NorB (EXB34), dependent on quinol for the passage of electrons, and its regulator NorR (EXB33) are encoded by ExoU island B. A nitric oxide reductase is encoded within the core genome; however, this enzyme is dependent on cytochrome *c* for the passage of electrons (2, 46). Cytochrome *c*-dependent NorB is part of the denitrification pathway that sequentially reduces nitrate to diatomic nitrogen and is used by *P. aeruginosa* during growth under anaerobic conditions (52). A quinol-dependent nitric oxide reductase in *P. aeruginosa* has not been previously identified. Expression of quinol-dependent NorB is regulated by the nitric oxide-activated response regulator NorR in *Ralstonia eutropha* (38). Quinol-dependent nitric oxide reductase is encoded by the genomes of some pathogenic bacteria that do not have denitrification ability, such as *Neisseria gonorrhoeae* and *Staphylococcus aureus* (5).

Sequence analysis of ExoU island (islet) C. The sequence encoding ExoU/SpcU from strain 2 consists of a 3.89-kbp segment located between genes homologous to PA0976 and PA0988 (Fig. 5). This region contains *exoU*, *spcU*, and an additional ORF and is very similar to the extremities of ExoU island A. The island is 85.2% identical to ExoU island A for the first 256 bp. The remaining nucleotide sequence is 99.8% identical to the sequence from the right border of ExoU island A, upstream of the PA0988 homolog. Most likely, a repeat

similar to the 392-bp repeat present in ExoU island A associated with IS407 facilitated a deletion in an ancestral island in strain 2, as this repeated sequence in ExoU island A forms the junction between the two different sequence homologies. The 20-bp TAGI repeat mentioned above also flanks ExoU island C.

A host-vector system for functional analysis of genomic islands. We have incorporated several features into the design of the vector p0975–0989capture used for recombinational cloning to allow targeted insertional mutagenesis of defined regions of a bacterial chromosome and facilitate generating libraries of mutants in a specific genomic island. The outline of this approach is shown schematically in Fig. 6A. Following capture of a genomic island by yeast recombinational cloning, the recombinant plasmid is propagated in *E. coli*, where it is subjected to transposon mutagenesis. The library of transposon insertions is then transferred into *P. aeruginosa* by conjugation. The *P. aeruginosa* strain carries a plasmid, expressing the I-SceI restriction endonuclease, which efficiently excises the insert DNA from the plasmid by cleavage at the two flanking I-SceI sites included in the capture vector. The linear DNA is then capable of integrating into the chromosome by double reciprocal recombination (50). Following selection for the resistance determinant carried by the transposon, a library of mutants is obtained which carries insertions limited to the

genomic island. A number of different *P. aeruginosa* recipients can be used, including the strain which was the source of the captured genomic island and a different strain which carries an identical island.

We have tested the utility of this system by using the captured ExoU island A from strain 4. We have generated a library of *mariner* transposon insertions in pCap0976.1:6077 that was subsequently introduced into *P. aeruginosa* strain 5 carrying plasmid pSW(I-SceI). Following selection on gentamicin-containing medium, individual clones were screened for the loss of tetracycline determinant, which indicated insertion of the island into the chromosome by double reciprocal recombination following excision of the entire insert fragment by I-SceI. Fifty to 90 percent of gentamicin-resistant colonies were sensitive to tetracycline. We isolated 38 clones and successfully amplified 31 inserts, which were sequenced to determine the junctions between the transposon and genomic DNA. The distribution of *mariner* insertions is shown in Fig. 6B. The relatively random distribution of the transposon insertions is the result of low target specificity of the *mariner* element. This strategy could be used as a general method for functional analysis of proteins encoded within genomic islands.

DISCUSSION

Acquisition of genes encoding virulence traits through horizontal gene transfer is an important mechanism in the evolution of pathogenic bacteria. The cytotoxin ExoU and its cognate chaperone SpcU are encoded by the genomes of a significant fraction of *P. aeruginosa* isolates, particularly those associated with acute pneumonia and ocular infections. We have investigated the composition of the genomic environment associated with the *exoU/spcU* gene pair. We used yeast recombinational cloning to examine the *exoU/spcU* locus from four isolates and sequenced three of the four captured islands. A comparison of the ExoU island A, B, and C sequences to published sequences provided the basis for a conclusion that *exoU* and *spcU* were acquired through a mobile genetic element. A comparison of the relatedness and synteny of the genes in the largest ExoU-encoding island with plasmid pKLC102 and pathogenicity island PAPI-1 indicates a close evolutionary relationship. A schematic illustrating the homologous regions shared between ExoU island A, PAPI-I, and pKLC102 is shown in Fig. 4. Because conserved hypothetical gene clusters in ExoU island A appear to be a subset of those found in pKLC102, we hypothesize that ExoU island A was derived from an integrative plasmid related to pKLC102.

Sequence analysis indicates that the ancestral *exoU/spcU*-containing plasmid acquired several insertion sequences and transposons. Following chromosomal integration, the ancestral element also underwent deletions, resulting in elimination of a number of genes required for transmissibility and episomal maintenance. Moreover, the number of conserved segments in the ExoU islands varies greatly, yet each strain retains full-length *exoU* and *spcU* genes, suggesting that these strains were subjected to environmental pressures selecting for the ability to secrete functional ExoU. Since the strains selected for this study were previously identified as *exoU* carriers, we cannot exclude the possibility that in strains lacking *exoU*, islands or plasmids related to ExoU island A may carry other genes that

are maintained in the chromosome due to different environmental selective pressures. Indeed, in strain PAO1, the location occupied by the various ExoU islands contains a segment apparently derived from the same ancestral plasmid element. Moreover, the pathogenicity island PAPI-1, although integrated at a different tRNA^{Lys} locus, appears to be derived from the same ancestral plasmid related to plasmid pKLC102.

A model describing the evolutionary history of the various ExoU islands and their relationship to the ancestral integrative plasmid is shown in Fig. 7. The ancestral plasmid, related to pKLC102, initially acquired *exoU/spcU* likely through a horizontal gene transfer event. The invariant association of *exoU/spcU* with IS407 suggests that this set of genes may have transposed as part of a mobile element. Insertion of *exoU/spcU* into the pKLC102-like ancestral plasmid may have occurred in a different bacterial species belonging to β - and γ -Proteobacteria, followed by a subsequent transfer to a *P. aeruginosa* recipient, where it integrated into the tRNA^{Lys} site. Certain strains of *P. aeruginosa*, such as PAO1, probably acquired a form of the ancestral pKLC102 lacking *exoU/spcU* that integrated into one of the two tRNA^{Lys} genes. This explains the existence of two different lineages of genomic islands among different *P. aeruginosa* strains differing in the presence or absence of *exoU*. Once these elements were integrated, recombination events could have deleted portions of the plasmid, eliminating factors needed for autonomous plasmid maintenance, thus fixing them as permanent genomic islands in the chromosome. Although this model predicts acquisition of *exoU* and *spcU* through transposition onto the plasmid element, we cannot exclude the possibility that they were acquired after integration of the pKLC102-like plasmid into the *P. aeruginosa* chromosome, possibly as part of a mobile genetic element with IS407 as depicted in Fig. 7.

Sequence comparison of the tRNA^{Lys}-associated genomic islands also provided evidence enabling us to speculate on the absence of *exoS* in *exoU/spcU*-containing strains. We hypothesize that prior to acquisition of *exoU/spcU*, the recipient strains contained *exoS*, which was subsequently deleted due to genetic determinants encoded by the ancestral ExoU island A. The *exoS* gene appears to have been acquired by the *P. aeruginosa* genome prior to *exoU/spcU* because of its lack of association with a variable portion of the chromosome and relatively typical percent G+C content. Currently, less than 1% of *P. aeruginosa* isolates lack both *exoS* and *exoU/spcU* (13), leading us to believe that recipient strains, prior to acquisition of *exoU/spcU*, contained *exoS* due to evolutionary pressures. Because *exoS* currently shows no association with features of a genomic island, it appears that at the time of *exoU/spcU* acquisition, plasmid incompatibility factors could not have served as a basis for mutual exclusion of *exoS* and *exoU/spcU*. Thus, it appears that *exoU/spcU*-containing strains lack *exoS* due to a targeted deletion event caused by a product of a gene linked to *exoU/spcU* at the time of acquisition. We compared the 10-bp direct repeat flanking *exoS* that also marks the site of deletion in strains lacking *exoS* (data not shown) with the TAGI repeat and noticed a slight homology between the two sets of repeated sequences (Fig. 3). Because the TAGI repeat may be involved in site-specific recombinase-mediated excision and integration of tRNA^{Lys} islands, we postulate that the recombinase acting on the TAGI repeat may also play a role in the excision of *exoS*.

FIG. 8. Single nucleotide polymorphisms of *exoU/spcU*.

We have compared the distributions of single nucleotide polymorphisms in *exoU* and *spcU* (Fig. 8). Although the sequences of these genes are highly conserved, a specific distribution of nucleotide polymorphisms can be identified. Clearly, *exoU/spcU* in ExoU island A and island C represents a clonal variant, with only one nucleotide difference. In contrast, the same gene pair in PAPI-2 and ExoU island B is more polymorphic, with numerous substitutions at different locations within the coding sequence representing a more evolved lineage of the ancestral strain.

The ExoU/SpcU pair is encoded by a genomic island that is hypothesized to have evolved from an ancestral plasmid similar to pKLC102. The integrase gene on this plasmid is 93% identical to this gene on ExoU island A and has been demonstrated to have an ability to integrate into both tRNA^{Lys} genes adjacent to PA0976 or PA4541 homologs. However, *exoU/spcU* has not been found on genomic islands integrated at the tRNA^{Lys} gene next to the PA4541 homolog. Possibly, in strains that had the ability to act as recipients of the ancestral ExoU island A, this site was not available for integration. It is also equally likely that acquisition of the determinants of ExoU/SpcU expression may have been an extremely rare event, meaning that the majority of approximately 25% of strains containing this gene may be the result of clonal expansion from the original recipient strain of the ancestral plasmid. Transfer of *exoU* as a rare event supports the idea that the ancestral island was rapidly rendered irreversibly integrated into the chromosome, as this would prevent further transmission of the *exoU*-carrying element. ExoU/SpcU-carrying strains show many signs of being clonally related. Strains carrying ExoU have been shown to have a relatively conserved genetic repertoire, as they are usually serotypes O1, O10, and O11 (4, 12). We have shown that strains 4, 5, and 6, isolated from geographically diverse locations, are very closely related and have most likely diverged from the same ancestral isolate. In addition, strains 2 to 6 were clustered as being more closely related to each other than to other ExoS-containing isolates, and at least one strain, strain 2, has an *exoU/spcU* locus that appears to have had a large-scale deletion stemming from an ancestral island that appeared to be very similar to ExoU island A.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM068516 to S.L., and work in D.F.'s laboratory was supported by NIH grant AI49577.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Arai, H., Y. Igarashi, and T. Kodama. 1995. The structural genes for nitric oxide reductase from *Pseudomonas aeruginosa*. *Biochim. Biophys. Acta* **1261**:279–284.
- Bergogne-Berezin, E. 2003. Pseudomonads and miscellaneous gram-negative bacilli, p. 2203–2217. In J. Cohen and W. Powderly (ed.), *Infectious diseases*, 2nd ed. Mosby, New York, N.Y.
- Berthelot, P., I. Attree, P. Plesiat, J. Chabert, S. de Bentzmann, B. Pozzetto, and F. Grattard. 2003. Genotypic and phenotypic analysis of type III secretion system in a cohort of *Pseudomonas aeruginosa* bacteremia isolates: evidence for a possible association between O serotypes and *exo* genes. *J. Infect. Dis.* **188**:512–518.
- Busch, A., B. Friedrich, and R. Cramm. 2002. Characterization of the *norB* gene, encoding nitric oxide reductase, in the nondenitrifying cyanobacterium *Synechocystis* sp. strain PCC6803. *Appl. Environ. Microbiol.* **68**:668–672.
- Davis, B. D., and E. S. Mingioli. 1950. Mutants of *Escherichia coli* requiring methionine or vitamin B₁₂. *J. Bacteriol.* **60**:17–28.
- Deretic, V., M. J. Schurr, and H. Yu. 1995. *Pseudomonas aeruginosa*, mucoidy and the chronic infection phenotype in cystic fibrosis. *Trends Microbiol.* **3**:351–356.
- DeShazer, D., D. M. Waag, D. L. Fritz, and D. E. Woods. 2001. Identification of a *Burkholderia mallei* polysaccharide gene cluster by subtractive hybridization and demonstration that the encoded capsule is an essential virulence determinant. *Microb. Pathog.* **30**:253–269.
- Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2**:414–424.
- Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**:186–194.
- Ewing, B., L. Hillier, M. C. Wendt, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**:175–185.
- Faure, K., D. Shimabukuro, T. Ajayi, L. R. Allmond, T. Sawa, and J. P. Wiener-Kronish. 2003. O-antigen serotypes and type III secretory toxins in clinical isolates of *Pseudomonas aeruginosa*. *J. Clin. Microbiol.* **41**:2158–2160.
- Feltman, H., G. Schultert, S. Khan, M. Jain, L. Peterson, and A. R. Hauser. 2001. Prevalence of type III secretion genes in clinical and environmental isolates of *Pseudomonas aeruginosa*. *Microbiology* **147**:2659–2669.
- Finck-Barbancon, V., J. Goranson, L. Zhu, T. Sawa, J. P. Wiener-Kronish, S. M. Fleiszig, C. Wu, L. Mende-Mueller, and D. W. Frank. 1997. ExoU expression by *Pseudomonas aeruginosa* correlates with acute cytotoxicity and epithelial injury. *Mol. Microbiol.* **25**:547–557.
- Finck-Barbancon, V., T. L. Yahr, and D. W. Frank. 1998. Identification and characterization of SpcU, a chaperone required for efficient secretion of the ExoU cytotoxin. *J. Bacteriol.* **180**:6224–6231.
- Fraylick, J. E., J. R. La Rocque, T. S. Vincent, and J. C. Olson. 2001. Independent and coordinate effects of ADP-ribosyltransferase and GTPase-activating activities of exoenzyme S on HT-29 epithelial cell function. *Infect. Immun.* **69**:5318–5328.
- Fraylick, J. E., M. J. Riese, T. S. Vincent, J. T. Barbieri, and J. C. Olson. 2002. ADP-ribosylation and functional effects of *Pseudomonas* exoenzyme S on cellular RalA. *Biochemistry* **41**:9680–9687.
- Fraylick, J. E., E. A. Rucks, D. M. Greene, T. S. Vincent, and J. C. Olson. 2002. Eukaryotic cell determination of ExoS ADP-ribosyltransferase substrate specificity. *Biochem. Biophys. Res. Commun.* **291**:91–100.
- Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
- Hacker, J., and J. B. Kaper. 2000. Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**:641–679.
- Hauser, A. R., E. Cobb, M. Bodi, D. Mariscal, J. Valles, J. N. Engel, and J. Rello. 2002. Type III protein secretion is associated with poor clinical outcomes in patients with ventilator-associated pneumonia caused by *Pseudomonas aeruginosa*. *Crit. Care Med.* **30**:521–528.
- Hauser, A. R., P. J. Kang, and J. N. Engel. 1998. PepA, a secreted protein of *Pseudomonas aeruginosa*, is necessary for cytotoxicity and virulence. *Mol. Microbiol.* **27**:807–818.
- He, J., R. L. Baldini, E. Deziel, M. Saucier, Q. Zhang, N. T. Liberati, D. Lee, J. Urbach, H. M. Goodman, and L. G. Rahme. 2004. The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc. Natl. Acad. Sci. USA* **101**:2530–2535.
- Hou, Y. M. 1999. Transfer RNAs and pathogenicity islands. *Trends Biochem. Sci.* **24**:295–298.
- Kiewitz, C., K. Larbig, J. Klockgether, C. Weinelt, and B. Tummler. 2000. Monitoring genome evolution ex vivo: reversible chromosomal integration of a 106 kb plasmid at two tRNA(Lys) gene loci in sequential *Pseudomonas aeruginosa* airway isolates. *Microbiology* **146**:2365–2373.
- Klockgether, J., O. Reva, K. Larbig, and B. Tummler. 2004. Sequence anal-

- ysis of the mobile genome island pKLC102 of *Pseudomonas aeruginosa* C. J. Bacteriol. **186**:518–534.
27. Kresse, A. U., S. D. Dinesh, K. Larbig, and U. Romling. 2003. Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Mol. Microbiol.* **47**: 145–158.
 28. Kulasekara, H. D., I. Ventre, B. R. Kulasekara, A. Lazdunski, A. Filloux, and S. Lory. 2005. A novel two-component system controls the expression of *Pseudomonas aeruginosa* fimbrial *cup* genes. *Mol. Microbiol.* **55**:368–380.
 29. Kurahashi, K., O. Kajikawa, T. Sawa, M. Ohara, M. A. Gropper, D. W. Frank, T. R. Martin, and J. P. Wiener-Kronish. 1999. Pathogenesis of septic shock in *Pseudomonas aeruginosa* pneumonia. *J. Clin. Investig.* **104**:743–750.
 30. Larbig, K. D., A. Christmann, A. Johann, J. Klockgether, T. Hartsch, R. Merkl, L. Wiehlmann, H. J. Fritz, and B. Tummeler. 2002. Gene islands integrated into tRNA(Gly) genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J. Bacteriol.* **184**:6665–6680.
 31. Lomholt, J. A., K. Poulsen, and M. Kilian. 2001. Epidemic population structure of *Pseudomonas aeruginosa*: evidence for a clone that is pathogenic to the eye and that has a distinct combination of virulence factors. *Infect. Immun.* **69**:6284–6295.
 32. Lukashin, A. V., and M. Borodovsky. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**:1107–1115.
 33. Marchler-Bauer, A., A. R. Panchenko, B. A. Shoemaker, P. A. Thiessen, L. Y. Geer, and S. H. Bryant. 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* **30**:281–283.
 34. Mohd-Zain, Z., S. L. Turner, A. M. Cerdano-Tarraga, A. K. Lilley, T. J. Inzana, A. J. Duncan, R. M. Harding, D. W. Hood, T. E. Peto, and D. W. Crook. 2004. Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. *J. Bacteriol.* **186**:8114–8122.
 35. Nelson, K. E., C. Weinle, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. A. Martins dos Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, L. Brinkac, M. Beanan, R. T. DeBoy, S. Daugherty, J. Kolonay, R. Madupu, W. Nelson, O. White, J. Peterson, H. Khouri, I. Hance, P. Chris Lee, E. Holtzapple, D. Scanlan, K. Tran, A. Moazzes, T. Utterback, M. Rizzo, K. Lee, D. Kosack, D. Moestl, H. Wedler, J. Lauber, D. Stjepandic, J. Hoheisel, M. Straetz, S. Heim, C. Kiewitz, J. A. Eisen, K. N. Timmis, A. Dusterhoft, B. Tummeler, and C. M. Fraser. 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**:799–808.
 36. Nowak-Thompson, B., N. Chaney, J. S. Wing, S. J. Gould, and J. E. Loper. 1999. Characterization of the pyoluteorin biosynthetic gene cluster of *Pseudomonas fluorescens* Pf-5. *J. Bacteriol.* **181**:2166–2174.
 37. Pallen, M. J., R. R. Chaudhuri, and I. R. Henderson. 2003. Genomic analysis of secretion systems. *Curr. Opin. Microbiol.* **6**:519–527.
 38. Pohlmann, A., R. Cramm, K. Schmelz, and B. Friedrich. 2000. A novel NO-responding regulator controls the reduction of nitric oxide in *Ralstonia eutropha*. *Mol. Microbiol.* **38**:626–638.
 39. Raymond, C. K., E. H. Sims, A. Kas, D. H. Spencer, T. V. Kutayin, R. G. Ivey, Y. Zhou, R. Kaul, J. B. Clendenning, and M. V. Olson. 2002. Genetic variation at the O-antigen biosynthetic locus in *Pseudomonas aeruginosa*. *J. Bacteriol.* **184**:3614–3622.
 40. Romling, U., J. Greipel, and B. Tummeler. 1995. Gradient of genomic diversity in the *Pseudomonas aeruginosa* chromosome. *Mol. Microbiol.* **17**:323–332.
 41. Roy-Burman, A., R. H. Savel, S. Racine, B. L. Swanson, N. S. Revadigar, J. Fujimoto, T. Sawa, D. W. Frank, and J. P. Wiener-Kronish. 2001. Type III protein secretion is associated with death in lower respiratory and systemic *Pseudomonas aeruginosa* infections. *J. Infect. Dis.* **183**:1767–1774.
 42. Salzberg, S. L., A. L. Delcher, S. Kasif, and O. White. 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **26**:544–548.
 43. Sato, H., D. W. Frank, C. J. Hillard, J. B. Feix, R. R. Pankhaniya, K. Moriyama, V. Finck-Barbancon, A. Buchaklian, M. Lei, R. M. Long, J. Wiener-Kronish, and T. Sawa. 2003. The mechanism of action of the *Pseudomonas aeruginosa*-encoded type III cytotoxin, ExoU. *EMBO J.* **22**:2959–2969.
 44. Schmidt, H., and M. Hensel. 2004. Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.* **17**:14–56.
 45. Schultert, G. S., H. Feltman, S. D. Rabin, C. G. Martin, S. E. Battle, J. Rello, and A. R. Hauser. 2003. Secretion of the toxin ExoU is a marker for highly virulent *Pseudomonas aeruginosa* isolates obtained from patients with hospital-acquired pneumonia. *J. Infect. Dis.* **188**:1695–1706.
 46. Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrenner, 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, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**:959–964.
 47. van der Meer, J. R., R. Ravatn, and V. Sentchilo. 2001. The *clc* element of *Pseudomonas* sp. strain B13 and other mobile degradative elements employing phage-like integrases. *Arch. Microbiol.* **175**:79–85.
 48. Wolfgang, M. C., B. R. Kulasekara, X. Liang, D. Boyd, K. Wu, Q. Yang, C. G. Miyada, and S. Lory. 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA* **100**:8484–8489.
 49. Wong, G. K., J. Yu, E. C. Thayer, and M. V. Olson. 1997. Multiple-complete-digest restriction fragment mapping: generating sequence-ready maps for large-scale DNA sequencing. *Proc. Natl. Acad. Sci. USA* **94**:5225–5230.
 50. Wong, S. M., and J. J. Mekalanos. 2000. Genetic footprinting with *mariner*-based transposition in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA* **97**:10191–10196.
 51. Wood, M. S., A. Byrne, and T. G. Lessie. 1991. IS406 and IS407, two gene-activating insertion sequences for *Pseudomonas cepacia*. *Gene* **105**: 101–105.
 52. Ye, R. W., D. Haas, J. O. Ka, V. Krishnapillai, A. Zimmermann, C. Baird, and J. M. Tiedje. 1995. Anaerobic activation of the entire denitrification pathway in *Pseudomonas aeruginosa* requires Anr, an analog of Fnr. *J. Bacteriol.* **177**:3606–3609.