Dynamics of *Pseudomonas aeruginosa* genome evolution

Kalai Mathee*†, Giri Narasimhan‡, Camilo Valdes‡, Xiaoyun Qiu§, Jody M. Matewish§, Michael Koehrsen¶, Antonis Rokas¶‖, Chandri N. Yandava¶, Reinhard Engels¶, Erliang Zeng‡, Raquel Olavarietta†, Melissa Doud†, Roger S. Smith§, Philip Montgomery¶, Jared R. White¶, Paul A. Godfrey¶, Chinnappa Kodira¶, Bruce Birren¶, James E. Galagan¶, and Stephen Lory§**

*Department of Molecular Microbiology and Immunology, College of Medicine, †Department of Biological Sciences, College of Arts and Sciences, and ‡Bioinformatics Research Group (BioRG), School of Computing and Information Sciences, College of Engineering, Florida International University, Miami, FL 33199; §Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115; and ¶The Broad Institute of MIT and Harvard, Cambridge, MA 02142

Communicated by John J. Mekalanos, Harvard Medical School, Boston, MA, December 19, 2007 (received for review November 21, 2007)

One of the hallmarks of the Gram-negative bacterium *Pseudomonas aeruginosa* is its ability to thrive in diverse environments that includes humans with a variety of debilitating diseases or immune deficiencies. Here we report the complete sequence and comparative analysis of the genomes of two representative *P. aeruginosa* strains isolated from cystic fibrosis (CF) patients whose genetic disorder predisposes them to infections by this pathogen. The comparison of the genomes of the two CF strains with those of other *P. aeruginosa* presents a picture of a mosaic genome, consisting of a conserved core component, interrupted in each strain by combinations of specific blocks of genes. These strain-specific segments of the genome are found in limited chromosomal locations, referred to as regions of genomic plasticity. The ability of *P. aeruginosa* to shape its genomic composition to favor survival in the widest range of environmental reservoirs, with corresponding enhancement of its metabolic capacity is supported by the identification of a genomic island in one of the sequenced CF isolates, encoding enzymes capable of degrading terpenoids produced by trees. This work suggests that niche adaptation is a major evolutionary force influencing the composition of bacterial genomes. Unlike genome reduction seen in host-adapted bacterial pathogens, the genetic capacity of *P. aeruginosa* is determined by the ability of individual strains to acquire or discard genomic segments, giving rise to strains with customized genomic repertoires. Consequently, this organism can survive in a wide range of environmental reservoirs that can serve as sources of the infecting organisms.

comparative genomics | horizontal gene transfer | core and accessory genomes

Pseudomonads are environmental saprotrophs, and only *P. aeruginosa* is capable of causing serious human infections. This microorganism has the ability to express a variety of virulence determinants, and it is not surprising that it can cause experimental infections in plants, nematodes, insects, and animals. Moreover, the ability to cause a wide range of acute and chronic human infections suggests an evolutionary mechanism whereby the repertoire of genes facilitating its survival in a particular environment is highly flexible showing a great deal of diversity among natural isolates thriving in different habitats. It is also likely that the relatively large genome of most *P. aeruginosa* strains (5–7 Mb) could include a substantial set of conserved genes, coding for functions necessary for survival in most environments (1). Human infections would therefore be a consequence of genome evolution in the environment where *P. aeruginosa* could encounter various hosts such as *Caenorhabditis elegans*, *Drosophila melanogaster*, *Galleria mellonella*, and *Dictyostelium discoideum* (2). This pattern of evolution is in contrast to the decay of genomes seen during the adaptation of pathogens to a parasitic existence (3).

Respiratory tract infections by *P. aeruginosa* include acute pneumonia particularly in patients with neutropenia, immunosuppression or undergoing mechanical ventilation (4, 5). Diffuse panbronchiolitis and respiratory disease in individuals with cystic fibrosis (CF) are two common chronic respiratory infections caused by *P. aeruginosa* (6). CF patients often succumb to infections from organisms found in their immediate environment. The early course of the disease combines features of acute infection, followed by chronic adaptation that includes numerous genetic changes such as overproduction of extracellular alginate polysaccharides, loss of the ability to add O-side chains to lipopolysaccharide core sugars, and mutations in regulatory genes (7).

To identify potential genomic determinants of *P. aeruginosa* adaptability, we determined the complete nucleotide sequence of two CF strains of different origin. Strain PA2192 is an isolate from a chronically infected patient in Boston and has undergone significant phenotypic adaptation characteristic of a majority of CF isolates, including conversion to mucoidy (due to a nonsense mutation in its *mucA* gene), production of lipopolysaccharide devoid of O-side chains, and lack of motility (8). Strain C3719, the so-called Manchester epidemic strain, has been associated with enhanced virulence and transmissibility (9). With the exception of mucoid conversion, it has also undergone similar adaptations associated with chronic CF isolates (9). Comparison of these two genomes with published sequences of clinical isolates PAO1 (10, 11), PA14 (12, 13), and PACS2, show that the genetic repertoire of each *P. aeruginosa* strain has unique components, consisting of specific insertions of blocks of genes acquired by horizontal gene transfer (HGT), with simultaneous discarding of genetic information, all taking place at a limited number of chromosomal loci. It is apparent that the evolution of the *P. aeruginosa* genome is driven by selection for an ability to expand its environmental niches. To infect eukaryotic hosts,

Author contributions: K.M., G.N., C.V., X.Q., J.M.M., M.K., A.R., C.N.Y., R.E., E.Z., R.O., M.D., P.M., J.R.W., P.A.G., C.K., B.B., J.E.G., and S.L. designed research; G.N., X.Q., J.M.M., A.R., R.E., R.O., M.D., R.S.S., P.M., J.R.W., and P.A.G. performed research; K.M., G.N., C.V., X.Q., M.K., A.R., C.N.Y., R.E., E.Z., R.S.S., P.M., J.R.W., P.A.G., C.K., J.E.G., and S.L. analyzed data; and K.M., G.N., M.K., P.A.G., C.K., and S.L. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposted in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo [accession nos. NZ_AAKW00000000 (PA2192) and NZ_AAKV00000000 (C3719)].

‖Present address: Department of Biological Sciences, Vanderbilt University, VU Station B 351634, Nashville, TN 37235.

**To whom correspondence should be addressed. E-mail: stephen_lory@hms.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0711982105/DC1.

including humans, the bacteria use virulence determinants that are conserved in their core genomes.

Results and Discussion

Features of *P. aeruginosa* PA2192 and C3719 Genomes. The single circular 6,222,097-bp chromosome of *P. aeruginosa* C3719 contains 5,578 ORFs, whereas the genome of PA2192 is significantly larger (6,905,121 bp), with a predicted coding capacity of 6,191 ORFs. The % (G+C) content of 66.5 and 66.2 for C3719 and PA2192, respectively, was similar to those of other *P. aeruginosa* genomes [supporting information (SI) Table 1]. It has been proposed that the preferential location of essential genes on the leading strand drives the observed G/C strand bias seen in different bacterial genomes (14). The *dif* sequence (GATTCGCATAATGTATATTATGTTAAAT) for replication termination was found at loci predicted from GC-skew maps (SI Fig. 4; ref. 15). The leading strands show the expected strand bias for both coding sequences and essential genes (PA2192, 54.7% and 65%, respectively; C3719, 54.5% and 66%, respectively), similar to that seen in other *P. aeruginosa* strains (PAO1, 56.1% and 67%, respectively; PA14, 54.4% and 66%, respectively; refs. 16 and 17). Comparison of the corresponding biases for coding sequences and essential genes in *Escherichia coli* (55% and 75%, respectively) and *Bacillus subtilis* (75% and 96%, respectively) shows that the leading strand bias for essential genes is significantly higher in both genomes (18).

The relationship of PA2192 and C3719 to other sequenced genomes was also determined by concatenating 1,836 ORFs with orthologs in all strains including *P. fluorescens* ORFs as outgroup (see *SI Text* and SI Fig. 5). The phylogenetic tree for the six *P. aeruginosa* strains agreed reasonably well with the origin and phenotypes of the genomes analyzed. The multihost pathogen, PA14, followed by PA2192, appears to be less related to the others, perhaps reflecting their large genome sizes. The two epidemic strains C3719 and LES were most closely related to each other. A list of 144 phylogenetically useful genes (PUG) was compiled from the ORFs whose individual phylogenetic trees agreed with the concatenated tree from above and that had at least three parsimony informative sites (SI Table 2). The PUG list provides good candidates for typing of strains, useful in high-resolution phylogenic analyses of closely related strains that are often needed for epidemiological studies of clonal variants.

A skew in strand-specific G/C revealed a significant asymmetry implying ancestral rearrangements, similar to that observed for the genome of PAO1 (11) but not in PA14 (13). G/C skew asymmetry is also found in the genome of PA2192 but not in C3719 (SI Fig. 4). Therefore, an inversion has very likely occurred in the genome of PA2192. In addition to PAO1 and PA2192, strain PACS2 has also undergone a similar genomic rearrangement supported by its asymmetrical G/C skew plot (data not shown). A comparison of gene order (synteny) in the *P. aeruginosa* genomes is shown in Fig. 1. The site of the rearrangements of the PAO1, PA2192, and PACS2 genomes suggests that the inversions resulted from recombination between the ribosomal RNA operon transcribed from the lagging strand (*rrnA*) and one of the three other ribosomal RNA operons transcribed in the opposite direction from the leading strand (*rrnB–D*). In strain PAO1 the recombination occurred between *rrnA* and *rrnB* as shown in ref. 13, in PA2192 between *rrnA* and *rrnD*, and in PACS2 between *rrnA* and *rrnC* (Fig. 1). The relatively common occurrence of inversions of large segments of the chromosome by recombination between two conserved *rrn* genes seen here and before (19) suggest that any physiological consequences in gene expression or in genome stability, when the location of a significant number of genes is changed, are minimal.

The Genome Architecture. Comparison of the annotated genes in the *P. aeruginosa* PA2192 and C3719 genomes with those in other *P. aeruginosa* genomes revealed an extensive conservation

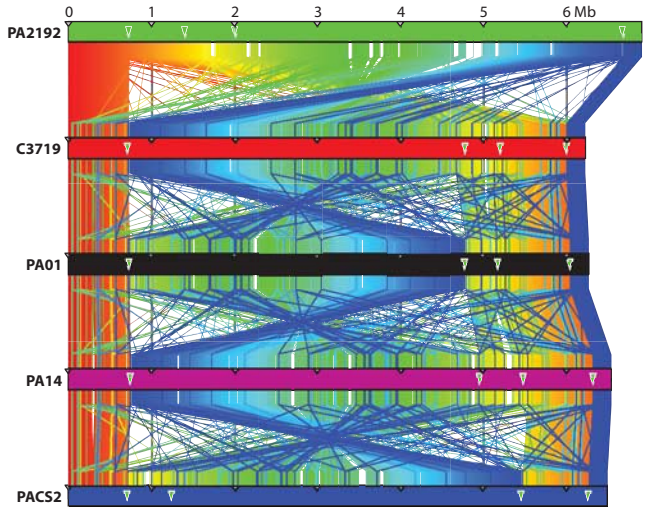

Fig. 1. Large inversions in various *P. aeruginosa* chromosomes. Whole genome alignments for all five strains of *P. aeruginosa* were generated using an anchor-alignment software Murasaki (http://murasaki.dna.bio.keio.ac.jp/; ref. 27). Scale in Mb is shown at top. Green, vertical arrowheads represent the rRNA operons. All alignments were performed on genomic sequences reset to the location of the origin of replication (see *SI Text*). PAO1, PACS2, and PA2192 have large inversions that appear to be anchored at rRNA operons [between *rrnA* and *rrnB* (PA01) or *rrnC* (PACS2) or *rrnD* (PA2192)]. PA2192 is given a color gradient so that corresponding alignment blocks may be easily visualized in the compared genomes.

of a set of genes that are shared by all of the strains (SI Fig. 6). There are 5,021 genes that are conserved across all five genomes analyzed, with at least 70% sequence identity. This set of genes will, henceforth, be referred to as the *P. aeruginosa* core genome. Of these, >90% of the genes share at least 98% sequence identity (data not shown). The genome for each strain carries a relatively modest number of unique sequences, 10% or less of the number in the core genome, ranging from 79 in C3719, to 507 in PA2192. Less than 100 genes are unique to pairs of genomes, whereas the number decreases significantly when comparison is carried out among sets of three or four genomes.

Given the conservation of synteny and the large number of core genes, the overall architecture of the pangenome of *P. aeruginosa* PA2192, C3719, PAO1, PACS2, and PA14 can be represented as a circular chromosome with dispersed polymorphic strain-specific segments, flanked by conserved genes referred to as anchors (Fig. 2 and SI Table 3). We designated these strain-specific segments as regions of genomic plasticity (RGP) and they include any region of at least four contiguous ORFs that are missing in at least one of the genomes analyzed. We designate the accessory genome as the set of genes comprising the sum of unique RGPs and the individual genes or gene clusters of less than four genes that are missing in at least one strain. Thus, any individual strain consists of the core genome plus a select subset of genes from the accessory genome. The RGPs can be sites of insertions of common or unique genomic islands and bacteriophage genomes, or the result of deletions of particular segments of DNA in one or more strains. Therefore, the RGP designation does not make any assumption about the evolutionary origin or genetic basis of these variable chromosomal segments.

The number of RGPs in individual genomes varies from 27 to 37. A total of 52 RGPs were identified in the five genomes analyzed (Fig. 2). The genetic organization of each RGP and their annotation is shown (SI Fig. 7) with accompanying functional annotations for each gene (SI Table 3). Occasionally, certain DNA segments seen in one RGP in one strain can be seen in a different RGP in another strain. Most frequently, this

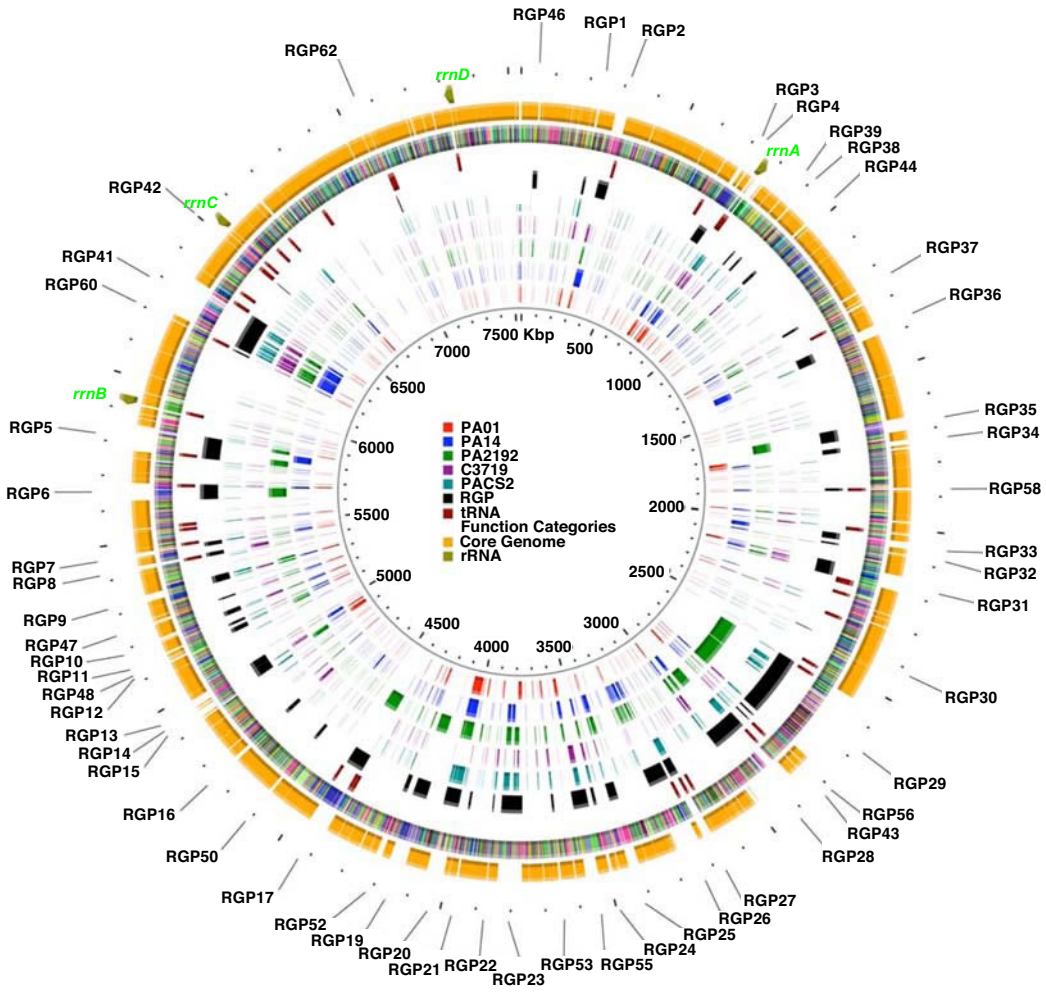

Fig. 2. The *P. aeruginosa* pangenome. Using the core genes from PA14 as the template, all of the accessory genes from PA2192, C3719, PAO1, and PACS2 were integrated. The outer circle (gold) indicates the core genes, the second circle shows the functional annotations. The third circle indicates the position of tRNAs. The accessory genes in the inner circles are from PA14 (blue) PA2192 (green), C3719 (purple), PAO1 (red), and PACS2 (teal). The outer green arrows show the positions of rRNAs.

involves insertion of mobile genetic elements into different sites in different strains. For example, RGP42 is the location of the Pf1 prophage in the genome of PA2192, whereas in C3719, the same element is in a different location (RGP8), where it is inserted into a seven gene cluster appearing to be remnants of an integrated plasmid or an integrative conjugative element.

Functional Content of *P. aeruginosa* Genomes. The analysis and comparison of RGP gene content reveals a diversity between strains. PA14 contains the largest number (seven) of unique RGPs not found in any of the other genomes, whereas the genome of PACS2 has one unique RGP (SI Table 1). PAO1 contains a single unique RGP, whereas PA2192 and C3719 each have six and three strain-specific RGPs, respectively. The rest of the RGPs are shared by at least two strains and 12 RGPs are found in four of the five strains. What is remarkable for this group of strains is the highly customized nature of their genomes, beyond the basic core set. As all of the strains analyzed are pathogenic, the virulence mechanism must be conserved and maintained through evolutionary pressure as suggested by our previous analysis of gene content using DNA microarrays (20).

The distribution of ORFs specifying different functional categories between the core and accessory genome was compared between PA2192, C3719, PAO1 and PA14 (SI Fig. 8). In general, the genes in the core genome contain the majority of housekeeping genes. For a number of functional categories, the percentage of ORFs in the core and accessory genomes was not significantly different. The accessory genome is significantly enriched in ORFs of unknown function or those related to transposons and bacteriophages (particularly in strain PA2192). In the category of genes encoding carbon compound catabolism, strain PA2912 is an exception, with a disproportionately high fraction of such ORFs in its accessory genome, drawn partially from a genomic island encoding enzymes of abietane diterpenoid (Dit) degradation. In contrast, the proportional contribution of functional categories that include translational and posttranslation modification, degradation and transcription/RNA processing and degradation, compared with the core and the accessory genomes of the other three strains, is significantly reduced.

The overall analysis of the total accessory genome is consistent with the examination of functional groups of ORFs encoded in individual RGPs (SI Fig. 9), although the distribution of various ORFs belonging to different functional categories is highly variable between individual RGPs. For example, ORFs for hypothetical unknown proteins account for the entire coding capacity of RGP24, -25, -43, -58, and -62 and they represent >70% of the genes in RGP2, -22, -33, -44, -50, and -52. In

contrast, genes coding for proteins of unknown function are lacking in RGP1, -12, -16, -39, -55, and -56; and are minor components of RGP3, -11, -17, -19, -31, -42, -45, and -46. Other RGPs show a strong bias in function, with a disproportional coding capacity occupied by ORFs belonging to single functional category. For example, RGP3, -10, -17, -42, and -46 harbor ORFs almost exclusively related to phages, transposons or plasmids, whereas all of the coding capacity of RGP12 is dedicated to functions related to DNA replication, recombination, modification, and repair. Significantly, most of the RGPs specify a limited number of ORFs in any one category. For example, RGP22, -33, -42, -44, -46, -50, -52, and -60 encode for proteins of unknown function and one additional functional category that is different between each of these five RGPs. Although RGPs show a dramatic decrease in functional diversity of their protein products in individual genomes, they become the source of specialized functions that very likely benefit the survival of a particular strain in its environmental niche.

Evolution of the Mosaic *P. aeruginosa* Genome. Analysis of orthologs was used to estimate the evolutionary origin of individual genes. BLASTX hits (against the NR database) showed that ≈42% of the ORFs in the core genome had best hits outside the genus *Pseudomonas*, whereas the corresponding number for the accessory genome was 66%. These genes may have been acquired by HGT (SI Table 4). The hits were to the genus *Azotobacter*, a closely related γ-*proteobacterium*, and to *Burkholderia* and *Cupriavidus* (formerly *Ralstonia*), two β-*proteobacteria*. It is therefore apparent that the genetic relatedness as well as potential coinhabitation of the same environmental reservoirs by *Pseudomonas*, *Azotobacter*, *Burkholderia* and *Cupriavidus* significantly enhances the likelihood of gene acquisition by HGT.

Several RGPs very likely arose as the result of deletions of blocks of DNA. For example, RGP20 contains a cluster of 45 genes in strain PAO1, and it is highly conserved in the other genomes (20) but is completely absent from the genome of C3719 (SI Fig. 7). Closer examination revealed that the two genes flanking this RGP in the other strains, were fused in C3719 into a single ORF (fusion of PACG_00990 and PACG_00991) creating a chimeric gene encoding a putative 937-residue protein that consists of the 851 N-terminal amino acids derived from the coding sequence of the right anchor (PA2070 in the PAO1 genome) and an 86-aa C-terminal segment derived from the sequence of the left anchor (PA2024). Although the majority (91%) of its coding sequence is derived from PA2070, a putative outer membrane siderophore receptor, it is unclear whether this newly created protein in C3719 retains its biological function.

A number of individual RGPs show highly variable sequences indicative of hotspots for the integration of horizontally acquired genes. For example, in different strains, RGP5 inserted in a $tRNA^{Gly}$ gene, contains several blocks of unrelated genes. In PAO1, RGP5 consists of 16 genes, of which 10 are related to bacteriophage Pf1. In strain PA2192, the 33 genes encode largely hypothetical unknown proteins, with a few related to bacteriophages and enzymes of DNA metabolism. In contrast, of 23 genes in RGP5 of PA14, all specify proteins that can be placed into well defined functional categories, related to various transporters and iron metabolism. In addition to incorporating heterologous DNA sequence into conserved chromosomal sites, some RGPs can arise by the decay (deletions) of portions of an ancestral element, leaving nonoverlapping segments that would appear as unique horizontally acquired segments. This pattern was previously described for the evolution of variants of the genomic island (RGP7) in certain strains carrying the gene for the cytotoxin ExoU (21).

Excision of RGPs. Many of the RGPs are located adjacent to tRNA genes, which frequently serve as insertion sites (*attB*) for bacteriophages or integrative and conjugative elements (ICE). The transmission of these mobile elements requires excision from the chromosome and formation of a circular intermediate. We screened 14 RGPs (RGP2, -5, -7, -8, -17, -26, -27, -28, -29, -41, -42, -58, -60, and -62) that are adjacent to various tRNA genes, and the tmRNA gene-linked RGP6, for the excision and formation of circular intermediates (SI Fig. 7). We designed PCR primers (SI Table 5 and Fig. 3*A*) to target the left and right junctions of each RGP (primers P1/P3 and P2/P4), the excision event (primers P1/P2) and the circularized form of the element (primers P3/P4). For those elements where excision or circularization was detected, the PCR products were sequenced. From the sequence of the restored *attB* sites, the sequence of the DR ends of the element was identified (SI Table 6).

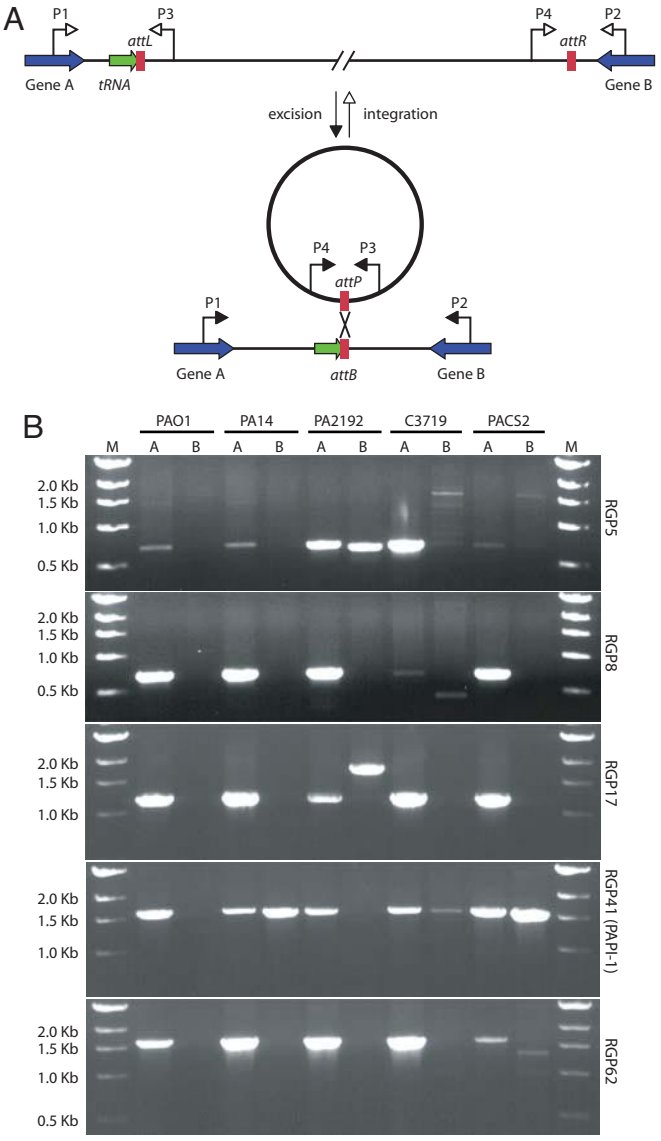

Fig. 3. Detection of RGP excision in various *P. aeruginosa* strains. (*A*) A generic model for RGP excision. Primers P1and P2 were designed to detect excision of RGPs and primers P3 and P4 to detect the presence of the circular form of the RGP following excision. Gene A and Gene B represent genes located adjacent to the RGP (anchors). The location of *att* sites is shown in red. (*B*) Agarose gel electrophoresis of PCR products generated by amplification for RGP5, RGP8, RGP17, RGP41 (PAPI-1) and RGP62 in strains PAO1, PA14, PA2192, C3719, and PACS2, respectively. Lane A shows the results of amplifications with primers P1 and P2 and B with primers P3 and P4.

MICROBIOLOGY

RGP41 is the Pathogenicity Island 1 (PAPI-1) that has been previously shown to excise and circularize in strain PA14 (22). The excision of this island was observed in strains PA14, PA2192, C3719, and PACS2; however, the circular intermediate was not detected in PA2192. The absence of circular form in PA2192 could be due to lack of expression of the integrase gene, or the consequence of two mutations in the *att* sites of the integrated element, preventing its excision.

Excision of RGP5 was observed in strains PAO1, PA14 and PA2192 but a circular intermediate was only detected in PA2192. The weak bands observed with primers P1/P2 and P3/P4 in strain PACS2 appear to be nonspecific amplifications (Fig. 3*B*). RGP8 and RGP17 are two bacteriophage-like elements, found in strains C3719 and PA2192, respectively. RGP8 in C3719 carries an integrase gene; its product may be responsible for the excision of the element, however RGP17 does not carry a gene encoding integrases or excisionases. RGP62, unique to PACS2, is a cluster of 11 genes mostly of unknown function, but with two genes related to phage sequences and a coding sequence for a putative integrase. The excision of this element and a circular intermediate was detected in PACS2 (Fig. 3*B*).

Evolution of RGP29 in Strain PA2192 and Acquisition of New Metabolic Capabilities. The large size of the chromosome of PA2192 is due in part to the presence of a variant of RGP29, a 224-kb complex genetic element integrated into a $tRNA^{Gly}$ gene, the last of a cluster of three tRNA genes ($tRNA^{Gly}$- $tRNA^{Gly}$ - $tRNA^{Glu}$; SI Fig. 10). The last $tRNA^{Gly}$ gene is missing in C3719 as a result of an apparent 160-bp deletion.

The complex arrangements of the RGP29 in PA2192 appear to be the consequence of sequential insertion of several mobile genetic elements. We have identified pairs of direct repeats that define the *att* sites that could be used during integration into the chromosome in this region (SI Fig. 11). The 3′ end of the last $tRNA^{Gly}$ gene of the three tRNA gene cluster contains sequences that are duplicated within RGP29, suggesting that this tRNA gene contained the target (*attB*) site for the integration of two elements acquired by HGT. A 15 bp sequence, TTGGAGCGGGAAACG, is repeated at the boundary with the sequences found in PA14, C3719 and PACS2 (SI Fig. 11). A shorter 12-bp segment (TTGGAGCGGGAA) of the 3′ end of the same tRNA gene is also found approximately in the middle of the element. The portion of RGP29 between these shorter repeats flank a 105-kb segment of DNA that is identical to the genomic island PAGI-2 described in the C-clones of *P. aeruginosa* (ref. 23; SI Fig. 10). To the left of PAGI-2 is a 112-kb DNA segment, and we refer to this element as the Dit Island (see below). We hypothesize that PA2192 acquired these elements by two consecutive integration events using the nearly identical *attB* sequence found in the 3′ end of the $tRNA^{Gly}$ gene. The direct repeats that flank these elements share a high level of sequence conservation, yet because they apparently originated by duplication of the same 3′ portion of the $tRNA^{Gly}$ gene, we can predict the sequence of events that resulted in acquisition of these two elements. First, the Dit Island was integrated into the 3′ end of $tRNA^{Gly}$ gene duplicating the 14-bp *att* site. This was followed by insertion of PAGI-2 into the same site of the $tRNA^{Gly}$ gene, resulting in duplication of the 11-bp sequence (SI Fig. 11). Because the *att* sites are similar, it is conceivable that integrases responsible for site-specific recombination between the *att* sites on the chromosome and the island are related. An integrase gene (PA2G_02184) is located on PAGI-2, whereas the Dit Island carries two integrase genes (PA2G_02064 and PA2G_02069) (SI Fig. 10). Sequence alignment of these three proteins revealed that the integrase on the Dit Island shares 57% sequence identity with PA2G_02069, whereas no similarity was found between PA2G_02064 and any of the other two integrases. This would suggest that the two related integrases independently mediated the insertion of the Dit Island and PAGI-2 into the same *attB* site in the $tRNA^{Gly}$ gene.

The Dit Island contains a cluster of 95 genes (PA2G_01975 to PA2G_02069) related to *dit* genes in other bacteria, encoding proteins of abietane diterpenoids metabolism (SI Fig. 12). These compounds are synthesized by trees as defense molecules and several bacteria, including *Pseudomonas abietaniphila* and *Burkholderia xenovorans*, can use abietane diterpenoids as the sole carbon source (24, 25). Twenty-three of the *P. aeruginosa dit* genes (PA2G_01983 to PA2G_02008) are highly homologous to genes in *P. abietaniphila* BKME-9 (62–91% sequence identity of their protein products) with conserved synteny (26). Several orthologs are found in the *B. xenovorans* genome, although synteny with the PA2192 is less conserved. Conservation of the majority of genes associated with diterpenoid metabolism suggests that PA2192 could use these compounds as a sole source of carbon and energy.

One of the striking features of the *P. aeruginosa* PA2192 Dit Island is the extent to which certain gene families appear to have expanded. The genes encoding the catalytic subunits of the dioxygenase (DitA1a, A2a; DitA1b, A2b; SI Fig. 12) and the P450 monoxygenase (DitQa, DitQb) are duplicated within the island. The PA2192 Dit Island encodes three genes for ferredoxins, with one (PA2G_02042) sharing significant identity (81%) with DitA3 of *P. abietaniphila*. A total of eight potential transporters of abietanes are encoded on the Dit Island. Three are paralogous proteins (PA2G_02008, PA2G_02018, PA2G_02045), having one homologue in *P. abietaniphila* (Orf2) and two in *B. xenovorans* (Bxe_C0615 and Bxe_C0580). The ortholog of PA2G_02002 in *P. abietaniphila* is DitE (absent in *B. xenovorans*). PA2G_02028 shares significant similarity (57%) with *B. xenovorans* Bxe_C0645 whose expression was shown to be induced by growth of the bacteria on dehydroabeitic acid (25).

Concluding Remarks

The comparative analysis of *P. aeruginosa* genomes presented here supports one of the recently recognized principles of prokaryotic evolution. In contrast to highly host adapted pathogens and symbionts undergoing genome reduction, a number of environmental organisms continually expand their genomic repertoires. In *P. aeruginosa* genomes, there is a conserved core component and an accessory component consisting of additional strain-specific regions formed by acquisition of blocks of genes by HGT in some strains and deletions of specific chromosomal segments in others. Using genetic mechanisms that facilitate the movement and alteration of genetic material in bacteria, *P. aeruginosa* is able to customize its genome to fit the needs for survival in virtually any environment. One of the more striking examples of genome expansion is the acquisition by PA2192 of a large cluster of genes involved in diterpenoid metabolism. A set of genes (found on the Dit Island) was added *en bloc* to its genomic repertoire allowing this particular strain to grow in environments rich in abietane diterpenoid resins that are produced primarily by trees. The same strain also infected a CF patient and was able to establish a persistent infection lasting for years. Expansion of nutritionally restricted niches by acquisition of novel metabolic capabilities through HGT appears to be a key evolutionary force shaping the *P. aeruginosa* genome and is reflected in the genome plasticity of individual strains. Importantly, the acquisition of new genetic elements, and consequently new traits does not eliminate others, and the organism retains its ability to thrive in the widest possible range of environments, including an infected human host.

Materials and Methods

Sequencing, Assembly, and Annotation. Information on sequencing, assembly and annotation can be found in the *SI Text*. Subsequent to the generation of the draft assembly and annotation, PCR-based screening of PA2192 was performed to verify locations of rRNA operons that were in the gaps.

Analysis of RGP Excision by PCR. For each RGP, primers specific to two conserved genes flanking that RGP (P1 and P2 in the Fig. 3*A*) were designed (SI Table 5) and were used to amplify the junction region after the island was excised from the chromosome. For those islands that undergo excision, primers specific to the first gene at both ends of each island (P3 and P4 in the Fig. 3*B*) were designed to detect a potential circular form. The junctions, formed by excision and the formation of circular forms, were confirmed by sequencing the PCR products.

Supplemental Web Sites. Detailed information of the genome annotations and analyses can be found at the BROAD web site (www.broad.mit.edu/annotation/genome/pseudomonas_group/MultiHome.html) and the BioRG web site (http://biorg.cis.fiu.edu/genomics/PA/supplemental).

ACKNOWLEDGMENTS. We thank Eshwar Mahenthiralingam (Cardiff School of Biosciences, Cardiff University, Cardiff, Wales) and Gerald Pier (Brigham and Women's Hospital, Harvard, Medical School, Boston, MA) for providing the strains used for sequencing and Timothy M. Collins (Florida International University, Miami) for helpful discussions. We also thank the University of Washington Genome Center for making available the sequence of strain PACS2. This work was supported by National Institutes of Health (NIH)/National Center for Complementary and Alternative Medicine Grant 1-R15-AT002626–01 (to K.M.), NIH/Minority Biomedical Research Support Support of Continuous Research Excellence Grant S06 GM08205 (to K.M. and G.N.), NIH Grant GM068516 (to S.L.), National Institute of General Medical Sciences Research Initiative for Scientific Enhancement Program Grant R25 GM61347 (to M.D.), American Cystic Fibrosis Foundation postdoctoral fellowship QIU06F0 (to X.Q.), a Canadian Cystic Fibrosis Foundation postdoctoral fellowship (to J.M.M.), and Florida International University graduate assistantship (to E.Z.). The genome sequencing of *P. aeruginosa* strains PA2192 and C3719 was funded by the National Institute of Allergy and Infectious Diseases (NIAID) Contract for Microbial Genome Centers HHSN26620040001C.

1. Schmidt KD, Tummler B, Romling U (1996) Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats. *J Bacteriol* 178:85–93.
2. Tan MW (2002) Cross-species infections and their analysis. *Annu Rev Microbiol* 56:539–565.
3. Moran NA (2002) Microbial minimalism: Genome reduction in bacterial pathogens. *Cell* 108:583–586.
4. Elting LS, Rubenstein EB, Rolston KV, Bodey GP (1997) Outcomes of bacteremia in patients with cancer and neutropenia: Observations from two decades of epidemiological and clinical trials. *Clin Infect Dis* 25:247–259.
5. Diaz E, Munoz E, Agbaht K, Rello J (2007) Management of ventilator-associated pneumonia caused by multiresistant bacteria. *Curr Opin Crit Care* 13:45–50.
6. Lyczak JB, Cannon CL, Pier GB (2002) Lung infections associated with cystic fibrosis. *Clin Microbiol Rev* 15:194–222.
7. Smith EE, *et al.* (2006) Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci USA* 103:8487–8492.
8. Pier GB, Matthews WJ, Jr, Eardley DD (1983) Immunochemical characterization of the mucoid exopolysaccharide of *Pseudomonas aeruginosa. J Infect Dis* 147:494–503.
9. Jones AM, *et al.* (2001) Spread of a multiresistant strain of *Pseudomonas aeruginosa* in an adult cystic fibrosis clinic. *Lancet* 358:557–558.
10. Holloway BW (1955) Genetic recombination in *Pseudomonas aeruginosa. J Gen Microbiol* 13:572–581.
11. Stover CK, *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 406:959–964.
12. Schroth MN, Cho JJ, Green SK, Kominos SD, eds (1977) *Epidemiology of Pseudomonas aeruginosa in agricultural areas* (Raven, New York).
13. Lee DG, *et al.* (2006) Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol* 7:R90.
14. Rocha EP, Danchin A (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet* 34:377–378.
15. Hendrickson H, Lawrence JG (2007) Mutational bias suggests that replication termination occurs near the *dif* site, not at *ter* sites. *Mol Microbiol* 64:42–56.
16. Jacobs MA, *et al.* (2003) Comprehensive transposon mutant library of *Pseudomonas aeruginosa. Proc Natl Acad Sci USA* 100:14339–14344.
17. Liberati NT, *et al.* (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci USA* 103:2833–2838.
18. Rocha EP, Danchin A (2003) Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res* 31:6570–6577.
19. Romling U, Schmidt KD, Tummler B (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.
20. Wolfgang MC, *et al.* (2003) Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa. Proc Natl Acad Sci USA* 100:8484–8489.
21. Kulasekara BR, *et al.* (2006) Acquisition and evolution of the *exoU* locus in *Pseudomonas aeruginosa. J Bacteriol* 188:4037–4050.
22. Qiu X, Gurkar AU, Lory S (2006) Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa. Proc Natl Acad Sci USA* 103:19830–19835.
23. Larbig KD, *et al.* (2002) Gene islands integrated into tRNA(Gly) genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J Bacteriol* 184:6665–6680.
24. Chain PS, *et al.* (2006) *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci USA* 103:15280–15287.
25. Smith DJ, Park J, Tiedje JM, Mohn WW (2007) A large gene cluster in *Burkholderia xenovorans* encoding abietane diterpenoid catabolism. *J Bacteriol* 189:6195–6204.
26. Martin VJ, Mohn WW (2000) Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. *J Bacteriol* 182:3784–3793.
27. Popendorf K (2007) Murasaki: Language-theory based homology detection tool across multiple large scale genomes, http://murasaki.dna.bio.keio.ac.jp.