The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes

Jianxin He*, Regina L. Baldini*, Eric Dëziel*, Maude Sauzier*, Quinhao Zhang*, Nicole T. Liberati†, Daniel Lee†, Jonathan Urbach†, Howard M. Goodman†, and Laurence G. Rahme*‡

*Department of Surgery, Harvard Medical School and Massachusetts General Hospital, and †Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114

Edited by Eugene W. Nester, University of Washington, Seattle, WA, and approved December 22, 2003 (received for review July 22, 2003)

The ubiquitous bacterium *Pseudomonas aeruginosa* is the quintessential opportunistic pathogen. Certain isolates infect a broad range of host organisms, from plants to humans. The pathogenic promiscuity of particular variants may reflect an increased virulence gene repertoire beyond the core *P. aeruginosa* genome. We have identified and characterized two *P. aeruginosa* pathogenicity islands (PAPI-1 and PAPI-2) in the genome of PA14, a highly virulent clinical isolate. The 108-kb PAPI-1 and 11-kb PAPI-2, which are absent from the less virulent reference strain PAO1, exhibit highly modular structures, revealing their complex derivations from a wide array of bacterial species and mobile elements. Most of the genes within these islands are homologous to known genes occur in other human and plant bacterial pathogens. For example, PAPI-1 carries a complete gene cluster predicted to encode a type IV group B pilus, a well known adhesin absent from strain PAO1. However, >80% of the PAPI-1 DNA sequence is unique, and 75 of its 115 predicted ORF products are unrelated to any known proteins or functional domains. Significantly, many PAPI-1 ORFs also occur in several *P. aeruginosa* cystic fibrosis isolates. Twenty-three PAPI ORFs were mutated, and 19 were found to be necessary for full plant or animal virulence, with 11 required for both. The large set of “extra” virulence functions encoded by both PAPIs may contribute to the increased promiscuity of highly virulent *P. aeruginosa* strains, by directing additional pathogenic functions.

*P. aeruginosa* possesses a large and diverse genome that is both highly conserved and plastic (14–17). Although clinical and environmental isolates have a remarkably conserved genome (14–19), distinct *P. aeruginosa* strains carry additional specific sequences, interspersed as genomic blocks in the conserved core genome, that account for 10% or more of their DNA (14, 18). To date, no distinct *P. aeruginosa* genomic island that carries virulence genes has been identified, with the exception of a recently identified genomic island that harbors the *exoU* virulence gene (14–17, 19, 20). Here we present the identification and characterization of two *P. aeruginosa* pathogenicity islands (PAPI-1 and PAPI-2) in the genome of the broad host range and highly virulent strain PA14. The large PAPI-1 island is absent from the reference strain PAO1, whereas only part of the smaller PAPI-2 element is found in this isolate. Although PA14 and PAO1 share the same host range, PA14 is considerably more virulent in several model organisms (2, 6). Most of the predicted proteins encoded by the PAPI genes exhibit no homology with any proteins of known function, and mutation of several of these genes reveals their relevance in both plant and animal pathogenicity. We speculate that PAPI-1 and PAPI-2 virulence functions promote the broad host promiscuity of highly virulent *P. aeruginosa* strains, such as PA14.

Materials and Methods

Strains, Plasmids, and Media. All *P. aeruginosa* strains are human isolates (6, 15, 16). The TnphoA mutant 3349 has been described (21). The PA14 genomic cosmid library, constructed in pJSR (6), was grown in *Escherichia coli* VCS257, and subcloned in DH5α. pRK2013 and pEX18Ap (22) served as the *P. aeruginosa* conjugation helper plasmid and marker exchange suicide vector, respectively. Bacteria were grown in LB plus 100 μg/ml ampicillin (*E. coli*), 100 μg/ml rifampicin (PA14), and 250 μg/ml carbenicillin (PA14 transconjugants).

DNA Methods and Library Construction. Probes were labeled with [32P]dCTP (NE) by using Rediprime II (Amersham Pharmacia). Genomic, cosmid, and plasmid DNA extractions followed standard procedures (23). To construct the PA14 cosmid library, a 30- to 50-kb partial Sau3AI digest of total PA14 DNA was size-fractionated in a 10–40% sucrose gradient, cloned into the BamHI site of pJSR, and packaged by using Gigapack III XL (Stratagene).

PA14 Mutants. In-frame deletions in 10 PAPI-1 ORFs and one PAPI-2 ORF were generated by PCR: 1.0- to 1.6-kb 5’ and 3’ segments were amplified from target PA14 genomic or cosmids.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CF, cystic fibrosis; DR, direct repeats.

†To whom correspondence should be addressed. E-mail: rahme@molbio.mgh.harvard.edu.

© 2004 by The National Academy of Sciences of the USA
DNA, and each amplicon, which included the first or last 10–20 aa of the target ORF plus an engineered restriction site, was ligated into pEX18Ap to produce replacement plasmids. In-frame deletion mutants were generated in PA14 via homologous recombination by sucrose resistance selection, and confirmed by hybridization.

Twelve TnphoA transposon insertion mutants of PA14 were obtained from a library that includes transposon–insertion mutants of all nonessential PA14 ORFs. Access to mutants and information about this library is currently available via a web interface (Massachusetts General Hospital Parabiosys: National Heart, Lung, and Blood Institute Program for Genomic applications, Massachusetts General Hospital and Harvard Medical School: http://pga.mgh.harvard.edu/cgi-bin/pa14/mutants/retrieve.cgi).

To evaluate pyocyanin production, bacteria were cultivated in King’s A broth and the relative phenazine concentration was measured at 695 nm. Pyoverdine concentrations in culture supernatants grown in King’s B medium were measured at 380 nm. Total proteolytic activity was estimated by spotting the genomic block.

**Fig. 1.** Correspondence of the PA14 PAPI-1 (A) and PAPI-2 (B) elements with the PA01 genome. Gene designations and linear coordinates (bp) are presented above and below the lines, respectively. Light and dark gray shading represent the conserved left and right boundaries, respectively. The figure is not drawn to scale.

**DNA Sequencing and Annotation.** The nucleotide sequences of the PA14 pl48, pG22, pSK91, pSK24, pF62, and pH44 cosmids were determined by shotgun sequencing. Cosmid fragments subcloned into pBluescript SK(−) and pDN19 were sequenced by primer walking to cover gaps. Individual reads were aligned and assembled by using DNAstar and CAP (http://pbl.univ-lyon1.fr/cap3.html). The sequence was compared to the PA01 annotated genome (www.pseudomonas.com) (26). tRNA genes were identified by using tRNA-SACN-SE (www.genetics.wustl.edu/eddy/IRNAScan-SE). ORFs were predicted by using GENEMARK.HMM (http://opal.biology.gatech.edu/GeneMark/gmmhmm2.prok.cgi) (27) and ORF finder (www.ncbi.nlm.nih.gov/orf/orf.html). Each predicted ORF of >200 bp was analyzed for homologies and conserved motifs by using BLASTN, BLASTP, and BLASTX. A full array of parameters was used. PSORT (http://psort.nibb.ac.jp/form.html) and TMPRED (www.ch.embnet.org/software/TMPRED.form.html) were used to predict cellular localization and transmembrane regions, respectively.

**Sequence Accession.** GenBank accession numbers for PAPI-1 and PAPI-2 are AY273869 and AY273870, respectively.

**Results**

**RL003 Defines Two Genomic Islands Present in PA14 and Absent from PA01.** The PA14 isogenic TnphoA insertion mutant 33A9, which carries an RL003 gene mutation, exhibits reduced plant and mouse pathogenicity (21). Because RL003 is absent from PA01, suggesting that it might occur within a P. aeruginosa pathogenicity island, we screened a PA14 cosmid library with a 300-bp RL03 probe. Initial results with the cosmids pA113, pB104, pl48, pH44, and pG68 showed that only pA113, pB104, and pl48 overlap. Although both borders of pH44 and pG68 contain PA01 sequences, only the left borders of pA113, pB104, and pl48 carry PA01 DNA, indicating that RL003 occurs in at least two sites in the PA14 genome, one of which includes a larger genomic region.

To further define this block, we performed a progressive cosmwalk, starting with a pl48 probe that contains the PA01/PA41 left junction, and identified a cosmid carrying the right PA14/PA01 junction (Fig. 4A, which is published as supporting information on the PNAS web site). A set of five cosmids clones, pl48, pG22, pSK91, pSK24, and pF62, were assembled to define a contiguous 150-kb region, designated PAPI-1, found in PA14 and absent from PA01 (Figs. 1A and 1B). Similarly, we used probes that correspond to the right and left borders of pH44 to confirm that this cosmid does not overlap PAPI-1, demonstrating that a second copy of RL003 occurs on a smaller PA14 genomic block, designated PAPI-2 (Figs. 1B and 1B).

**PAPI-1 Is a Pathogenicity Island.** Comparison of the nucleotide sequence of the region defined by the five PAPI-1 cosmids (annotated in Fig. 2 and in Table 2, which is published as supporting information on the PNAS web site) with the PA01 genome (26), shows that the 20-kb left end of pl48 (GenBank accession no. AY273871) is collinear with the PA01 genome, whereas the internal 107,899 bp are unique to PA14 (Figs. 1A and 2A). This 108-kb region has all of the features of a genomic island: it occupies a block absent from several P. aeruginosa strains; its G+C content (59.7%) is different from that of the core genome (66.6%); it is associated with tRNA genes, because a tRNA<sup>A<sub>Glu</sub></sup>, tRNA<sup>Phe</sup>, and tRNA<sup>Glu</sup> gene cluster (annotated as PA4541.1–3 in PA01) occurs at its leftward PA01/PA14 junction, and a 58-bp direct repeat of the 3′ half of the tRNA<sup>Glu</sup> gene, designated attR, occurs just within its right border, so that it is bounded by 58-bp direct repeats (Figs. 1A and 2A); it contains seven mobility factor genes that encode integrases and transposases, plus four related pseudogenes and direct and inverted repeat sequences (Table 3, which is published as supporting information on the PNAS web site); it appears to have undergone deletions in different P. aeruginosa strains and/or additional insertions have occurred in PA14 (Fig. 3 and described below); and finally, it carries at least 19 virulence factors that occur on genomic islands found in a wide spectrum of other pathogenic bacteria (Tables 1 and 2).

**Functional Organization and Predicted ORFs of PAPI-1.** Data in Fig. 2 and in Table 2 illustrate the highly modular organization and complex origin of PAPI-1, which is inserted in a hypervariable region of the P. aeruginosa genome near the pilA gene (14, 15,
Remarkably, >80% of its DNA sequence is unique and has no similarity with any GenBank sequences. Furthermore, 75 of its 115 predicted ORFs are unrelated to any previously identified proteins or functional domains, and thus cannot be assigned any function by homology. Conversely, 40 PAPI-1 ORF-translated sequences show homology to proteins from several bacterial species, demonstrating the modular evolution of PAPI-1. For instance, 18 PAPI-1 genes display significant homology to pathogenicity-related genes, including a putative type III effector (RL030), a type IVB-like pilus gene cluster (RL077–86), and a chaperone/usher pathway (cup) gene cluster (RL040–44) (Fig. 2 and Table 2, and described below). The majority of the remaining related PAPI-1 genes encode functions related to DNA mobilization, integration, and partition activities. Many PAPI-1 predicted proteins are related to sequences found in Salmonella, pathogenic E. coli, Haemophilus somnus, Yersinia pestis, P. aeruginosa, Pseudomonas syringae, Pseudomonas fluorescens, Xylella fastidiosa, Burkholderia fungorum, and Xanthomonas (Table 2 and Table 4, which is published as supporting information on the PNAS web site). Also, 26 PAPI-1 ORF translated sequences are similar to predicted proteins on both the 134-kb island of the mammalian pathogen Salmonella enterica (STY4521–4680) (8), and the 130-kb island of the phytopathogen Xanthomonas axonopodis (XAC2171–2286) (12) (Fig. 2 and Tables 2 and 4). Moreover, 21 additional PAPI-1 ORFs show similarity with ORFs from only one of these pathogenicity islands, 14 with S. enterica (RL052, 72, 77–79, 81–85) and 7 with X. axonopodis (RL020, 35, 63–65, 67). This abundant array of putative pathogenicity-related genes likely plays a role in the broad host range of PA14.

Interspersed with its ORFs, PAPI-1 carries at least five pairs of direct repeats (DR), a pair of inverted repeats, and an insertion sequence also found in Pseudomonas putida (28) (Fig. 2 and Tables 3 and 5, which is published as supporting information on the PNAS web site). The 63-bp DR1 repeats, which border the entire PAPI-1 element and are part of the tRNA\(^{Lys}\) gene, include the 58-bp attR sequence. A hairpin-like structure thought necessary for DNA insertion (29) occurs downstream of the right DR1, and this sequence might correspond to the

---

**Fig. 2.** The organization of PAPI-1 (A) and PAPI-2 (B). The boxes with arrows represent individual ORFs and their corresponding transcriptional orientations. Empty boxes represent pseudogenes; triangles represent tRNA genes; and the marked vertical line represents the presumptive attR “attachment” site. The numbered lines represent size (kb), and the coincident rectangles and single or double-headed arrows on the line respectively correspond to direct repeats (DR1–5), inverted repeat (IR), and insertion sequences. ORF color and pattern, respectively, correspond to the predicted protein function and the bacterial species that it is most related to, according to the key. Pathogenesis-related ORFs are indicated by red shading. Gene clusters functions are marked and correspond to the ORFs above the notations. The shaded regions show the homology between PAPI-1 and PAPI-2.

---

**Fig. 3.** Presence of PAPI-1 in P. aeruginosa clinical isolates. The thick black line represents the PAPI-1 coordinates (kb). The arrowheads indicate the position of the DR. The black rectangles correspond to the probes used for hybridization. + denotes positive hybridization; N denotes experiment not done. All strains giving positive hybridization are shown.
actual PAPI-1 integration site, generating the attR and attL sequences. We note that DR1-like sequences also occur in *P. aeruginosa* strains C and SG17M genomic islands, and in *Xylella fastidiosa* (17).

The 662-bp DR2 repeats encode two proteins of unknown function (RL035 and RL046) and may have served as a DNA transposon-generated mutants.

<table>
<thead>
<tr>
<th>Strain name</th>
<th>% Mouse mortality</th>
<th>Growth in Arabidopsis leaf</th>
<th>Closest published homologue (organism/GenBank accession no.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA14</td>
<td>100</td>
<td>$4.9 \times 10^4$</td>
<td>Wild-type</td>
</tr>
<tr>
<td>RL003</td>
<td>41</td>
<td>$2.3 \times 10^4$</td>
<td><em>P. syringae</em> pv. tomato DC3000/AAO54371</td>
</tr>
<tr>
<td>RL003R</td>
<td>95</td>
<td>$4.4 \times 10^4$</td>
<td>The mutated RL003 locus was replaced with its wild-type copy</td>
</tr>
<tr>
<td>RL008</td>
<td>38</td>
<td>$4.1 \times 10^4$</td>
<td><em>Methanosarcina acetivorans</em> C2A/AAM05338 and <em>P. aeruginosa</em> PA01/AAG05323</td>
</tr>
<tr>
<td>RL009</td>
<td>31</td>
<td>$1.3 \times 10^4$</td>
<td><em>P. aeruginosa</em> PA01/AAG05327</td>
</tr>
<tr>
<td>RL009R</td>
<td>100</td>
<td>$4.0 \times 10^4$</td>
<td>The mutated RL009 locus was replaced with its wild-type copy</td>
</tr>
<tr>
<td>RL016</td>
<td>100</td>
<td>$2.8 \times 10^4$</td>
<td><em>P. syringae</em> pv. tomato DC3000/AAO54383</td>
</tr>
<tr>
<td>RL020</td>
<td>50</td>
<td>$3.4 \times 10^4$</td>
<td>Protein-disulfide isomerase, <em>P. aeruginosa</em> PA01/AAG04371</td>
</tr>
<tr>
<td>RL022</td>
<td>88</td>
<td>$3.3 \times 10^4$</td>
<td><em>P. syringae</em> pv. tomato DC3000/AAO54394</td>
</tr>
<tr>
<td>RL029</td>
<td>38</td>
<td>$9.4 \times 10^4$</td>
<td><em>P. aeruginosa</em> C/AAN62148</td>
</tr>
<tr>
<td>RL033</td>
<td>25</td>
<td>$4.9 \times 10^4$</td>
<td>No significant similarity</td>
</tr>
<tr>
<td>RL036</td>
<td>44</td>
<td>$1.9 \times 10^5$</td>
<td>Two-component sensor <em>P. aeruginosa</em> PA14/AAM15532</td>
</tr>
<tr>
<td>RL037</td>
<td>43</td>
<td>$1.2 \times 10^5$</td>
<td>Two-component regulator pvR, <em>P. aeruginosa</em> PA14 AAM15533</td>
</tr>
<tr>
<td>RL038</td>
<td>31</td>
<td>$4.4 \times 10^4$</td>
<td>Two-component sensor rscC, <em>S. typhimurium</em> LT2/AAL21172</td>
</tr>
<tr>
<td>RL039</td>
<td>31</td>
<td>$2.7 \times 10^4$</td>
<td>Two-component regulator rscB, <em>E. coli</em> O157:H7/EDL933/AAG57352</td>
</tr>
<tr>
<td>RL043</td>
<td>75</td>
<td>$1.7 \times 10^4$</td>
<td>Probable pilus assembly chaperone cupA2, <em>P. aeruginosa</em> PA01/AAG05517</td>
</tr>
<tr>
<td>RL054</td>
<td>63</td>
<td>NT</td>
<td><em>P. aeruginosa</em> PA01/AAG05610</td>
</tr>
<tr>
<td>RL062</td>
<td>78</td>
<td>NT</td>
<td>No significant similarity</td>
</tr>
<tr>
<td>RL065</td>
<td>63</td>
<td>$4.5 \times 10^4$</td>
<td><em>X. axonopodis</em> pv. citri 306/AAM37094</td>
</tr>
<tr>
<td>RL068</td>
<td>56</td>
<td>$2.6 \times 10^4$</td>
<td>No significant similarity</td>
</tr>
<tr>
<td>RL069</td>
<td>67</td>
<td>$2.7 \times 10^4$</td>
<td>No significant similarity</td>
</tr>
<tr>
<td>RL092</td>
<td>38</td>
<td>$3.9 \times 10^4$</td>
<td>Topoisomerase I Topa, <em>Xyuella fastidiosa</em> 95c (plasmid pXF51)/AAF85572</td>
</tr>
<tr>
<td>RL095</td>
<td>50</td>
<td>$5.3 \times 10^4$</td>
<td>Single-stranded DNA binding protein Ssb, <em>P. aeruginosa</em> C/AAN62318</td>
</tr>
<tr>
<td>RL101</td>
<td>38</td>
<td>$1.8 \times 10^4$</td>
<td><em>Pseudomonas</em> sp. B13/CAD60668</td>
</tr>
<tr>
<td>RL112</td>
<td>38</td>
<td>$1.6 \times 10^4$</td>
<td>No significant similarity</td>
</tr>
<tr>
<td>RS056</td>
<td>100</td>
<td>$1.8 \times 10^4$</td>
<td><em>P. aeruginosa</em> PA01/AAG04369</td>
</tr>
</tbody>
</table>

NT, not tested.

*Strain designations correspond to the mutated ORFs (see Fig. 2).

Mice were infected with $5 \times 10^5$ bacterial cells as described, and animal mortality (7) was scored for up to 10 days. Eight to 16 mice were used per experiment.

To assess plant pathogenicity, *Arabidopsis* leaves were inoculated with $3.3 \times 10^5$ bacterial cells and assayed 4 days after infection for bacterial CFU/cm². Four different leaves were sampled. All experiments were performed twice.

Statistical significance for mortality data and bacterial growth in *Arabidopsis* leaves was determined by the t test. Differences between groups were considered statistically significant at $P \leq 0.05$ and are shown in bold.

In-frame deletion mutants.

The 248-bp DR3 repeats prescribe a 2.5-kb region of 46.4% G+C, indicating its foreign origin (Fig. 2 and Table 3). This region contains the RL087–8 genes, which are homologues of PA0984–5 that encode a bacteriocin, pyocin S5, and its immunity protein (37). We note that the RL087–8 genes are nearly identical homologues of the pyocin S5 genes found on the PAO1 island (Table 4) corresponding to PAPI-2. A pilus biogenesis system (RL77–86) is located just upstream of the left DR3 (Fig. 2). This system resembles type IV group B pili clusters found in other pathogenic bacteria, including the enteropathogenic *E. coli* bundle-forming pilus, the *S. enterica* CT18 type IVB pilus, and the *Vibrio cholerae* toxin-coregulated pilus (8, 38–40).

Interestingly, the type IV pilus biogenesis machinery is highly homologous to the type II secretion pathways (Table 6, which is published as supporting information on the PNAS web site).

Both DR4 and DR5 consist of two consecutive direct repeats. These repeats are adjacent to the RL092, RL095, RL102, RL109–11, and RL114 ORFs, which are related to plasmid-encoded replication and recombination functions, suggesting that portions of PAPI-1 might be plasmid-derived. In contrast, only two PAPI-1 ORFs (RL103 and RL110) are phase-related (Table 2). Interestingly, the chromosomal integrase RL002 and the chromosome-partitioning protein Soj RL115 genes are located at the ends of the island, similar to the *P. aeruginosa* clone C islands, suggesting that this island may have had an intermediate circular form that once integrated into tRNA sequences (19).

Finally, Fig. 2 and Table 2 demonstrate that genomic “shuffling” has also contributed to PAPI-1 organization, as the RL001, RL020, RL087, and RL088 ORFs are closely related to the PA0977–87 genes, which are located on a PAO1 genomic
island. RL054–56 and RL113 share homology with the PA01 genes PA2221–8, which also occur in a region having atypical G+C content (26).

Characterization of the PAPI-2 Pathogenicity Island. Sequencing of pH44 and pG68, which carry a second copy of RL003, reveals a 10,722-bp region, designated PAPI-2, located near the phnAB genes (Figs. 1B and 2B), a hypervariable region of the P. aeruginosa genome (14, 15, 18). Fig. 2B illustrates the organization of the 15 PAPI-2 predicted ORFs, seven of which correspond to hypothetical proteins of unknown function (Table 7, which is published as supporting information on the PNAS web site). Unlike PA01 and PAPI-1, PAPI-2 exhibits features of a genomic island, with a G+C content of 56.4%. It contains multiple predicted mobility functions, including one integrase gene, four transposase genes, and one related pseudogene; and at its left border has an almost complete IS222 element, plus a portion of ISPpu14, a putative transposase gene (Fig. 2B and Table 5).

Half of PAPI-2 is homologous to PA0977–0987, an 8.9-kb PAO1 genomic island (19), which encodes 11 predicted ORFs (Fig. 2B and Table 8, which is published as supporting information on the PNAS web site). Unlike PA0977–0987, PAPI-2 is not associated with an attR site, although it is located at the same position in the P. aeruginosa core genome (Fig. 1B). Furthermore, these two islands share upstream and downstream sequences, and six ORFs (Figs. 1B and 2B and Table 7). Although the PAO1 island, unlike PAPI-2, does not contain the entire RS02 integrase gene, it does have an intact tRNAlys (attL) at its left border and a corresponding 22-bp direct repeat at its right border (attR) (Fig. 1B). The RS03 predicted product shares homology with the N terminus of RL003, the product of the 33A9 locus (Table 7). Interestingly, the 2.5-kb left end of PAPI-2 is identical to the 2.5-kb left end of PAPI-1, from the tRNAlys gene to RL003 and RS03, respectively (Fig. 2). Finally, the PAO1 pyocin genes PA0984–85 are replaced in PAPI-2 by the cytoxin exoU gene and its chaperone spcU (RS14–15) (Table 8). exoU encodes a type III effector that plays an important role in pathogenesis (41). Its presence on PAPI-2 defines this block as a pathogenicity island.

PAPI Island ORFs Encode Pathogenicity-Related Functions. We generated and analyzed 23 mutant strains (Table 1), including 11 in-frame deletion mutants and 12 TnphoA transposon insertion mutants, to assess whether the PAPI-1 and PAPI-2 ORFs that encode hypothetical/unknown functions promote P. aeruginosa pathogenesis. None of the mutants are defective for growth in liquid culture or for the extracellular production of pyocyanin and pyoverdine and for protease activity. Because some of the known PAPI-1 genes are involved in adhesion and/or motility, we also evaluated the mutants for colony morphology, in vitro adhesion, and swimming, twitching, and swarming motilities. All of the mutants behave like the wild-type parent, indicating that these activities do not depend on the mutated gene under our experimental conditions (data not shown).

Virulence was assessed in plants and animals by using the Arabidopsis leaf infiltration and the mouse thermal injury models (21). Table 1 shows that 19 of the 23 mutants exhibit attenuated virulence phenotype in at least one of the hosts, with 11 attenuated in both. We note that caution must be taken when assigning a virulent phenotype to a particular mutated gene, as some of the mutations in our pool of TnphoA insertion mutants (Table 1) may be polar and thus also affect loss-of-function for downstream genes. Indeed an example of this affect is observed with two different mutations in the RL003 locus, the TnphoA generated mutant 33A9 (21) and the in-frame deletion mutant RL003, which are weakly and moderately virulent, respectively.

Importantly, 15 of the mutants that exhibit attenuated virulence phenotype correspond to recently discovered genes and one to a known gene (prrR) that has not been previously shown to be involved in virulence. Of the mutated ORFs, RL016 and RL022 occur within a large region (RL012–30) found in several other phytopathogenic bacteria, and RL036–37 and RL038–39 encode two-component regulatory systems, suggesting that pathogenicity activities regulated by these systems are evolutionarily conserved. To test whether two of the mutations that we generated are indeed responsible for the Arabidopsis and mice attenuated-in-virulence phenotypes, we replaced the mutated loci in the in-frame mutant strains RL003 and RL009 with the corresponding wild-type loci to generate the strains RL003R and RL009R. As predicted, these strains exhibit restored wild-type virulence in plants and mice (Table 1).

Presence of PAPI-1 in Other P. aeruginosa Clinical Isolates. We used 11 hybridization probes spanning PAPI-1 to assess the occurrence of this island in 14 P. aeruginosa pathogenic strains, 12 of which were isolated from CF patients (Fig. 3) (15, 16). Whereas CF1, CF3–5, CF27–28, CF30, and CF32 do not hybridize with any of the probes used, PAO37, CF2, and CF6 hybridize with all of them, suggesting that these strains carry the entire 108-kb PAPI-1 island. In contrast, other strains hybridize to a subset of probes. CF26 appears to carry the complete island, except for the region found between the DR2 sequences, whereas CF29 only carries its leftward half. Both PAK and PAO1 harbor only a small segment of PAPI-1, with PAK carrying its left end, and PAO1 carrying a 1.7-kb region that encodes the pyocin S5 and immunity genes, which occupy different chromosomal locations in PA01 versus PA14.

Discussion

Different P. aeruginosa strains have the remarkable ability to inhabit diverse environments and infect a range of organisms, from amoeba to humans (1, 2). This environmental and pathogenic promiscuity is due in part to the large and genetically diverse P. aeruginosa genome. Sequencing of the PAO1 reference strain reveals that it encodes an abundant repertoire of specialized genes that promote success in various environments (26). We propose that the pathogenic promiscuity of certain P. aeruginosa strains is further brought about via factors carried by pathogenicity islands. Indeed, our results for PAPI-1 and PAPI-2 provide observations and implications for pathogenicity island modular structure and evolution, relatedness to other bacterial species, and contribution to PA14 pathogenic promiscuity and the generation of pathogenic variants.

PAPI-1 has a highly mosaic structure. It harbors blocks of ORFs related to virulence functions in other human and phytopathogenic bacteria, and ORFs similar to genes found in Archaea and phages, illustrating its diverse foreign origin. The PAPI-1 border regions exemplify this mosaicism. Whereas the right border contains ORFs unrelated to any GenBank sequences, the left border carries ORFs found in Archaea species and in other P. aeruginosa strains. Interestingly, one of these ORFs, RL008, encodes a putative helicase fused to sequences homologous to a PAO1 gene that encodes an unknown function (42). By mutation analysis, this hybrid ORF encodes a mammalian virulence factor, and thus might represent a previously undescribed pathogenic function generated via gene fusion.

The highly modular organization of PAPI-1 demonstrates that it was generated by multiple recombination events, as it carries several unrelated genes and gene clusters. Indeed, a large portion of PAPI-1 is similar to ORF clusters found in the genomes of the phytopathogen X. axonopodis pv. citri (12), and the human pathogen S. enterica ser. Typhi (8). This region is interrupted by unrelated ORFs located between repeat sequences, suggesting that a fragment homologous to the X. axonopodis and S. enterica gene blocks may have been acquired by P. aeruginosa as a complete DNA fragment, and later
interrupted by the insertion of unrelated fragments. Interestingly, one of these secondary regions, RL036–39, contains two pairs of two-component regulatory systems, which we show affect plant and mammalian pathogenesis. Acquisition of regulatory systems and virulence genes from other microorganisms may have contributed to the evolution of P. aeruginosa pathogenic variants to thrive in diverse environments. For instance, the PAPI-1 group B type IV pili genes are related to virulence systems and virulence genes from other bacterial species is not surprising, because P. aeruginosa inhabits soil and water environments, and is associated with several hosts. In both islands carry genes that allow this pathogen to thrive on different strains in these species as well. The presence of virulence factors, such as P. aeruginosa virulence genes have spread to other bacterial species, to generate virulent strains in these species as well.

PAPI-1 and PAPI-2 mutational analyses demonstrate that both islands carry genes that allow this pathogen to thrive on evolutionary diverse hosts, including plants and mammals. Indeed, of 23 mutated ORFs, 19 encode functions necessary for Arabidopsis or mice virulence, with 11 required for full “wild-type” virulence in both hosts. Although the majority of these genes encode products of unknown function, their presence in P. aeruginosa clinical isolates, including those from CF patients, suggests that they direct conserved functions important for fitness and survival. Future characterization of these pathogenicity factors should provide insights into broad host pathogenic and defense mechanisms. Completion of the PAI4 genome sequence might also lead to the identification of additional PAPI blocks and virulence genes.

Perhaps our most interesting results concern the identification of virulence factors and the sequence of PAPI-1. Remarkably, 70% of its DNA sequence is unique, and 75 of its 115 predicted ORFs are unrelated to any previously identified proteins or functional domains. Significantly, many of these ORFs act in pathogenesis. They likely represent a previously undescribed repertoire of virulence activities functioning beyond the large set of virulence and regulatory genes already present on the P. aeruginosa core genome, apparently to provide promiscuous pathogenic variants with an enhanced bag of tricks with which to infect more host species.

We thank S. Lory for providing the CF strains, S. Stachel for comments and editing, and S. Gopal for helpful discussion on the statistical analysis. This work was supported in part by a grant from Aventis S.A. (to L.G.R.). E.D. is supported by a postdoctoral fellowship from Canadian Institutes of Health Research. R.L.B. was partially supported by a fellowship from the Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil. N.T.L., D.L., and J.U. were supported in part by National Heart, Lung, and Blood Institute Grant U01 HL66678 (to F.M. Ausubel).