Significant differences in type IV pilin allele distribution among *Pseudomonas aeruginosa* isolates from cystic fibrosis (CF) versus non-CF patients

Julianne V. Kus,1,3 Elizabeth Tullis,2 Dennis G. Cvitkovitch3 and Lori L. Burrows1,3

1Centre for Infection and Biomaterials Research, Hospital for Sick Children Research Institute and Department of Surgery, University of Toronto, 7142A Elm Wing, 555 University Avenue, Toronto, ON, Canada M5G 1X8
2Adult Cystic Fibrosis Clinic, St Michael’s Hospital, Toronto, ON, Canada
3Department of Microbiology, Faculty of Dentistry, University of Toronto, Toronto, ON, Canada

Type IV pili (TFP) are important colonization factors of the opportunistic pathogen *Pseudomonas aeruginosa*, involved in biofilm formation and attachment to host cells. This study undertook a comprehensive analysis of TFP alleles in more than 290 environmental, clinical, rectal and cystic fibrosis (CF) isolates of *P. aeruginosa*. Based on the results, a new system of nomenclature is proposed, in which *P. aeruginosa* TFP are divided into five distinct phylogenetic groups. Each pilin allele is stringently associated with characteristic, distinct accessory genes that allow the identification of the allele by specific PCR. The invariant association of the pilin and accessory genes implies horizontal transfer of the entire locus. Analysis of pilin allele distribution among isolates from various sources revealed a striking bias in the prevalence of isolates with group I pilin genes from CF compared with non-CF human sources (*P*<0.0001), suggesting this particular pilin type, which can be post-translationally modified by glycosylation via the action of TfpO (PilO), may confer a colonization or persistence advantage in the CF host. This allele was also predominant in paediatric CF isolates (29 of 43; 67±4 %), showing that this bias is apparent early in colonization. Group I pilins were also the most common type found in environmental isolates tested. To the authors’ knowledge, this is the first example of a *P. aeruginosa* virulence factor allele that is strongly associated with CF isolates.

INTRODUCTION

*Pseudomonas aeruginosa* remains among the top opportunistic and nosocomial pathogens as well as the major cause of morbidity and mortality among cystic fibrosis (CF) patients. It elaborates a wide variety of virulence factors including adhesins and toxins (Lyczak *et al*., 2000; Rahme *et al*., 2000; Stanislavsky & Lam, 1997). One of the most significant adhesins, polar type IV pili (TFP), is important in bacteria–host cell interactions (Hahn, 1997), in formation of biofilms (O’Toole & Kolter, 1998) and for twitching motility, a unique mechanism of surface propulsion (Mattick, 2002; Semmler *et al*., 1999). Bacteria lacking TFP have been shown to be significantly less able to bind to eukaryotic cells (Farinha *et al*., 1994; Paranchych *et al*., 1986) and to be unable to form microcolonies on artificial surfaces, necessary for initiating formation of a biofilm (O’Toole & Kolter, 1998). TFP are produced by many Gram-negative bacteria in addition to *Pseudomonas*, including *Neisseria*, *Moraxella*, *Dichelobacter*, *Eikenella* and *Myxococcus* (Strom & Lory, 1993). The pilin structural subunits (products of the pilA gene) are produced as prepilins which are cleaved near the N terminus by a specific prepilin peptidase (PilD/XcpA in *P. aeruginosa*) and N-methylated prior to assembly into the mature pilus fibre (Lu *et al*., 1997; Strom *et al*., 1993).

Previous small studies of *P. aeruginosa* pilin diversity (Castric, 1995; Castric & Deal, 1994; Spangenberg *et al*., 1995) have demonstrated that the N-terminal amino acid

**Abbreviations:** CF, cystic fibrosis; DSL, disulfide-bonded loop; TFP, type IV pili.
region of mature pilins is highly conserved. This region of the pilin is thought to mediate interaction between subunits, permitting self-assembly of the helical quaternary structure of the pilus fibre (Keizer et al., 2001). In contrast, the primary amino acid sequence of the remainder of the pilin protein is less conserved, particularly at the C terminus which contains a disulfide-bonded loop (DSL) structure (Castric & Deal, 1994; Spangenberg et al., 1995). Despite this sequence diversity, however, P. aeruginosa pili are thought to interact with the eukaryotic glycolipid receptor asialoGM3 (aGM3) (Bryan et al., 1998; Saiman & Prince, 1993) via the DSL, which is proposed to be exposed only at the pilus tip (Doig et al., 1988; Irvin et al., 1989). Unlike members of the Neisseriaceae, which can generate pilin diversity through recombination of the expressed pilin gene locus (pilE) with additional silent pilin gene cassettes (pilS) present in the genome (Gibbs et al., 1989), P. aeruginosa possesses only a single pilA structural gene (Stover et al., 2000). Therefore, pilin diversity in P. aeruginosa is more likely to arise through random mutations within a strain or by recombination with horizontally acquired pilin genes from other strains or species.

Castric & Deal (1994) divided P. aeruginosa pilin sequences into two related groups based on DNA sequence analysis of pilin genes from nine strains. Group I pilin genes are highly conserved, having a higher G+C content (approx. 51 mol%) and are longer than group II genes (approx. 48 mol%). Strains with group II pilin genes include the common laboratory strains PAO1, PA103 and PAK (Johnson et al., 1986; Pasloske et al., 1985; Sastry et al., 1985). Further analysis by Castric (1995) of the chromosome adjacent to the pilin structural gene showed that group II pilin genes were immediately upstream of a tRNA<sup>Thr</sup> gene, while strains with group I pilin genes contained an additional open reading frame (ORF), pilO, between pilA and tRNA<sup>Thr</sup>.

Another limited survey of 14 P. aeruginosa pilin gene sequences (Spangenberg et al., 1995) identified, in addition to group I and II pilins, a single unusual pilin sequence from a CF isolate, G7, with only partial amino acid identity (51%) to the N-terminal 85 aa of group I pilins. The G7 pilA gene had a G+C content of 54.8 mol% and the predicted amino acid sequence of the mature pilin was larger (173 aa) than group I pilins. A recent representational difference analysis survey comparing the highly virulent strain PA14 with PAO1 showed that the pilA gene of PA14 was identical to that of G7 (Choi et al., 2002).

In addition to variations in pilin amino acid sequence, at least one group I pilin, from strain 1244 (Castric, 1995), has been shown to be post-translationally modified by O-glycosylation on the C-terminal serine residue (Comer et al., 2002). The glycosyl moiety was recently identified as the O-antigen unit of the lipopolysaccharide of 1244, which belongs to serotype O7 (Castric et al., 2001). The carboxy terminus of the modified pilin was strongly antigenic, eliciting antibodies that cross-reacted with the O antigen (Comer et al., 2002). A pilO knockout strain continued to produce pili and to twitch, but its pilin had altered migration on polyacrylamide and isoelectric focusing gels and lacked glycosylation by specific staining. PilO appears to be the only protein required for pilin glycosylation, but the mechanism of PilO function and the biological significance of pilin glycosylation in P. aeruginosa are currently unknown.

The importance of TFP in the biology and virulence of P. aeruginosa has made them attractive targets for vaccine development (Hertle et al., 2001; Sheth et al., 1995). Saiman et al. (1989) raised cross-reactive monoclonal antibodies that recognized some pilin variants and reduced binding to eukaryotic cells. However, to design a successful P. aeruginosa vaccine based on TFP, it is important to be aware of the scope of P. aeruginosa pilin diversity and the prevalence of accessory genes such as pilO whose products are capable of post-translationally modifying the pilin subunit. In this study, we examined the pilin alleles of a large panel of environmental, rectal, clinical and CF isolates, and identified novel pilin and accessory genes. These data provide the basis for a new system of nomenclature for P. aeruginosa pilins. We show that the distribution of pilin alleles amongst CF and non-CF human isolates is sharply skewed, with the majority of CF isolates (from both paediatric and adult populations) belonging to pilin group I. This is the first example of a P. aeruginosa virulence factor allele that is strongly associated with isolates from CF patients.

**METHODS**

**Bacterial strains and media.** Patient-identity-blinded clinical and rectal isolates of P. aeruginosa (Table I, supplementary data) were prospectively collected from the microbiology laboratory at Mount Sinai Hospital in Toronto, ON. Mount Sinai is a large tertiary hospital which processes microbiology samples from several Greater Toronto Area hospitals and long-term care facilities. Rectal colonizer strains were isolated from routine rectal surveillance swabs (for resistant Enterobacteriaceae) plated on MacConkey agar supplemented with 2 μg cephalothin ml<sup>−1</sup>, to which P. aeruginosa is resistant. Clinical isolates were obtained from various inpatient specimens submitted from both sterile and non-sterile sites. CF isolates were obtained with ethics approval from St Michael’s Hospital, Toronto, ON, the Hospital for Sick Children, or as gifts from Dr David Speert, University of British Columbia, and Dr John Govan, University of Edinburgh. All CF strains were typed by PFGE to eliminate clonal isolates. Environmental isolates were gifts from Dr David Speert, Dr Stephen Lory of Harvard University, Dr Eric Deziel and Dr Richard Villenum of the INRS-Institut Armand-Frappier-Microbiologie et Biotechnologie, Dr Alan Hauser of University of Edinburgh. All CF strains were typed by PFGE to eliminate clonal isolates. Environmental isolates were gifts from Dr David Speert, University of British Columbia, and Dr Richard Villenum of the INRS-Institut Armand-Frappier-Microbiologie et Biotechnologie, Dr Alan Hauser of Northwestern University, and from Don Warburton of Health Canada. All strains were maintained as glycerol stocks at −70 °C after one passage on Luria–Bertani (LB) or Pseudomonas Isolation agar. For cloning experiments, transformed Escherichia coli TOP10 (Invitrogen) were grown on LB agar supplemented with 50 μg ampicillin ml<sup>−1</sup>.

**Twitching assay.** Each strain was tested upon receipt for twitching motility by stabbing it with a sterile toothpick to the bottom of a 3 mm thick 1% LB agar plate (Semmler et al., 1999). After
incubating overnight at 37 °C in a humidified chamber, the presence (or absence) of a hazy zone radiating from the point of inoculation at the agar–plate interface, indicating twitching motility, was recorded. To improve visualization of twitching zones, agar was removed from the plate and the twitching zones were stained with crystal violet for 5 min then rinsed with tap water to remove unbound dye.

PCR, cloning and DNA sequencing. PCR primers designed for this study and used for amplification of pilA and adjacent sequences from *P. aeruginosa* were: pilA (5′-ATG AAA GCT CAA AAA GGC TTT ACC TTG AT); pilB (5′-TCC AGC AGC ATC TTG TTG AGC AA); pilB2 (5′-TGT TCA GCT CGC GAT AAG G); and tRNA<sub>Thr</sub> (5′-CGA ATG AGC TGC TCT ACC GAC AGA GCT). Primers were synthesized by ATGC, Toronto, ON. Chromosomal DNA was prepared using Instagene (Bio-Rad) following the manufacturer’s instructions. Chromosomal DNA was isolated from mucoid strains using a modified phenol/chloroform extraction procedure (Goldberg & Ohman, 1984). The PCR mix consisted of 50 μl containing 1 μl template DNA, 3 μl 10 mM dNTPs, 50 pmol each primer, 5 μl 10 × PCR buffer, 10 μl Q solution (Qiagen) and 0·5 μl Hot Star<sup>C</sup> Taq (Qiagen). PCR (using a Perkin Elmer 2400 thermocycler) consisted of a 15 min denaturation at 95 °C, followed by 30 cycles of 30 s at 95 °C, 45 s at 60 °C and 2 min at 72 °C, with a final extension of 7 min at 72 °C, ending at 4 °C. Strains that were PCR-negative after the first round of amplification underwent a second PCR amplification with a reduced annealing temperature of 55 °C and an increased extension time of 2·5 min.

Amplification products of interest were either sequenced directly or cloned into pCR2.1 TOPO (Invitrogen) following the manufacturer’s instructions, transformed into chemically competent *E. coli* TOP10 cells and plated on LB agar plates containing 50 μg ampicillin ml<sup>−1</sup> (L + A) and 2 % X-Gal (Bioshop, Mississauga, ON). White colonies were screened for plasmid inserts using a rapid technique (Eckhardt, 1978). Recombinant clones were grown overnight on L + A plates and plasmids isolated using the Qiagen plasmid mini-prep kit as indicated by the manufacturer. The DNA sequence of cloned amplicons was obtained using the dye-deoxy terminator method by ACGT, Toronto, ON. Inserts were sequenced bidirectionally using M13 forward and reverse primers and subsequently with internal primers based on the first-round sequence.

Identification of novel accessory genes. For identification of strains containing tfpO<sub>b</sub>, tfpO<sub>b</sub>-positive and tfpZ accessory genes down-stream of pilA, specific PCR assays were performed. Primers tfpO up (5′-CGT ACT ATT CTA TTA TTG CTA G) and tfpO down (5′-CAA AGG ATG GCC TAC GAA), tfpO2 up (5′-CGT ACT CTG TTT TTC TCC TTC), tfpO2 down (5′-GCA TCT CTC GCC CAC AAC AGC), tfpY up (5′-AAA TAG CAG GAG GAT GCC G) and tfpY down (5′-CTC CCC CAC ACT ATC CGC GTA C), and tfpZ up (5′-ATA AAA CGG TGG CAG TTG G) and tfpZ down (5′-AAA GCG AAG ACG GGA A) were used. Strain 1244 (gift of Dr Peter Castric) was used as a positive control for *tfpZ* <sup>up</sup> and *tfpO* <sup>down</sup> (which amplifies *tfpO* genes). PCR-positive strains were considered to contain novel accessory DNA, which was then cloned and sequenced.

RFLP typing of *tfpO<sub>b</sub>* strains. Strains that were negative by specific PCR using primers tfpO up and down (which amplifies *tfpO<sub>b</sub>* but not *tfpO<sub>a</sub>* but positive using primers tfpO2 up and down (above) were recorded as *tfpO<sub>b</sub>-positive*. To further distinguish variations between *tfpO<sub>a</sub>-positive* strains, their pilB<sup>+</sup>RNA<sup>TM</sup> gene product was cloned into pCR2.1 and transformed into TOP10 chemically competent *E. coli*. Plasmids were isolated using QIAprep Spin Miniprep Kit following the manufacturer’s instructions (Qiagen) and 2 μl plasmid DNA was digested with EcoRI, *TaqI* or *Rsal* for 1 h and separated on a 1 % agarose gel. Samples that yielded dissimilar restriction-fragment length patterns were sequenced.

**PFGE.** The relatedness of strains was assessed by PFGE as described previously (Lightfoot & Lam, 1993; Romling & Tümmler, 2000). Briefly, cells from overnight cultures grown on Mueller–Hinton plates were washed and resuspended in SE buffer (25 mM EDTA pH 7·4, 75 mM NaCl) to a density of 8 McFarland and embedded in an equal volume of 1·6 % low-melting-point agarose (Bio-Rad). Embedded cells were incubated at 37 °C for 2 h with gentle shaking in lysis buffer (10 mM Tris/HCl pH 7·2, 50 mM NaCl, 100 mM EDTA pH 8·0, 0·2 % deoxycholate, 1 % Sarkosyl, 2 mg ml<sup>−1</sup> lysozyme ml<sup>−1</sup>). After a 15 min wash at 37 °C in wash buffer (20 mM Tris/HCl pH 8·0, 50 mM EDTA pH 8·0), embedded cells were incubated overnight at 50 °C in protease K solution (100 mM EDTA pH 8·0, 0·2 % deoxycholate, 1 % Sarkosyl, 50 μl protease K ml<sup>−1</sup>). Embedded cells were then washed 3 × 30 min in wash buffer (20 mM Tris/HCl pH 8·0, 50 mM EDTA pH 8·0). Agarose slices 2–3 mm thick were incubated overnight with *SpeI* in 1 × restriction enzyme buffer (as supplied by the manufacturer; New England Biolabs) at 37 °C. DNA fragments were separated in a 1 % PFGE-grade agarose (Bio-Rad) gel with 0·5 × Tris borate/EDTA running buffer supplemented with 50 μM thiourea (Romling & Tümmler, 2000) using a Bio-Rad CHEF II apparatus with the following programme: initial time 5 s; final time 35 s; run time 20 h; ratio 1; voltage 200 V; temperature 12 °C. The gel was post-stained with 2 μg ml<sup>−1</sup> ethidium bromide in double-distilled H<sub>2</sub>O for 30 min and de-stained for 1 h in double-distilled H<sub>2</sub>O. The bands were visualized with UV light and photographed. Banding patterns were interpreted based on the classification system outlined by Tenover *et al.* (1997). Isolates are considered subtypes (differing by one or two genetic events) when they differ by more than one band and fewer than five bands; samples are considered different strains when they differ by more than six bands (Arbeit, 1995; Tenover *et al.*, 1997).

**Analysis of pilin proteins by SDS-PAGE.** Pilin proteins were isolated using the methods of Castric (1995) with modifications. Overnight lawns of bacteria on Davis minimal agar plates were resuspended in 2 ml sterile PBS. Pili were sheared by vigorous vortexing for 5 min and cells removed by centrifugation. Supernatant proteins were precipitated by the addition of 1 M MgCl<sub>2</sub> to a final concentration of 0·1 M and incubation at 4 °C overnight. Precipitated proteins were harvested by centrifugation and resuspended in 2 % SDS sample buffer. Pilin preparations were boiled for 5 min, separated on 15% SDS-PAGE gels with a pre-stained Benchmark Protein Ladder (Invitrogen), and visualized by staining with Coomassie brilliant blue.

**Phylogenetic analysis.** Amino acid sequences of PilA variants from representative strains from each TFP group and the closest non-*P. aeruginosa* homologues (GenBank, [http://www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)) were aligned in CLUSTAL<sub>X</sub> (Thompson *et al.*, 1997) using default parameters. Phylogenetic trees were generated using PHYLIP 3.5 using default parameters and the Jones–Taylor–Thornton substitution matrix for amino acids (Felsenstein, 1989; Jones *et al.*, 1992). One thousand bootstrap replicates were performed to generate the final consensus tree which was visualized in TREEVIEW v. 1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). For comparison, and with similar results, trees were also generated with MEGA2 (Kumar *et al.*, 2001).

**Statistical analysis.** The statistical significance of differences in pilin gene distribution between sources was calculated using the chi-square test with Yates’ correction for continuity (http://www.unc.edu/~preacher/chisq/chisq.htm). Distributions were considered significantly different at *P* < 0·05.
Accession numbers. Novel sequences from each of pilin groups I, II, IV and V described in this study have been deposited in GenBank and have been assigned the following accession numbers: strain Pa302025, AAM62144 (PilAI) and AAM62145 (TfpO); strain Pa5658, AAM62146 (PilAI) and AAM62147 (TfpO); strain Pa131533, AMMS2057 (PilAI) and AMMS2058 (TfpO); strain Pa5235, AAM44066 (PilAI); strain Pa5196, AAM52059 (PilAIV), AAM52060 (TfpW) and AAM52061 (TfpX); strain Pa110594, AAM52055 (PilAV) and AAM52056 (TfpZ). Sequences of group III pilins from strains G7 and PA14 are available at accession numbers AAC41410 and AAL12242, respectively.

RESULTS

Twitching motility in CF versus non-CF strains

Macroscopic twitching motility was observed in 23 of 24 (95.8%) environmental isolates (Table I, supplementary data). Over 71% (113 of 159) of non-CF rectal and clinical isolates exhibited twitching in the agar stab assay. Loss of twitching motility in CF isolates from chronically colonized patients, as well as other phenotypic alterations such as loss of swimming motility and increased alginate production, has been reported previously (Garrett et al., 1994; Mahenthiralingam et al., 1994; Penketh et al., 1983). We saw a similar phenomenon in this study: while 35 of 43 (81.4%) paediatric CF isolates demonstrated twitching, only 36 of 66 (54.5%) adult CF isolates were positive for twitching motility.

PCR amplification analysis of pilA and flanking sequences

To amplify both pilA and any downstream sequences, an upstream primer located in the divergently transcribed, conserved pilB gene was used in conjunction with a downstream primer corresponding to the tRNA\textsuperscript{Thr} gene located 3’ to pilA in PAO1 (Fig. 1). All strains tested produced a single PCR product ranging in size from approximately 1.4 kb to greater than 4 kb. Based on product size from agarose gel analysis, 97 of 159 non-CF human isolates (61.0%), 17 of 24 environmental isolates (70.8%), and 89 of 109 CF strains (81.7%) contained DNA between pilA and tRNA\textsuperscript{Thr} (Table I, supplementary data). The remaining strains generated PCR products predicted by size to contain pilA alone, without accessory DNA. Selected products from a range of sizes were cloned into pCR2.1 and sequenced, or sequenced directly.

Group II pilins

Strains predicted to contain pilA alone were assigned to group II, as previously established by Castric & Deal (1994) (Fig. 1). The predicted amino acid sequences of PilA from two randomly selected group II strains (Pa260611, Pa270176) were similar to P. aeruginosa strain K122-4 (Paslocke et al., 1988a). Both strains had an identical R43A substitution compared with K122-4. Strains Pa281298, Pa270958, Pa100443 and Pa100683 had pilins most closely related to that of strain CD, a CF isolate identified by Paslocke et al. (1988b). All group II PilA sequences have a 12 aa C-terminal DSL region related to that of common laboratory strains PAO1 and PAK (Fig. 2). Pilins in this group have been designated PilAI\textsubscript{II}.

Unlike the other group II strains sequenced, strain Pa5235 possessed a novel pilin gene with a G+C content of 51 mol%, encoding a 154 aa prepilin with a 12 aa predicted DSL (Fig. 2). The Pa5235 pilin was 55% identical to FimZ (minor pilin precursor, GenBank accession no. YQBDZD) of Dichelobacter nodosus serogroup D, the aetiological agent of ovine footrot, whose TFP are important for both colonization and virulence factors (Kennan et al., 2001). The pilin of P. aeruginosa strain KB7 (Campbell et al., 1995; Wong et al., 1995) appears to be closely related to that of Pa5235, based on a short partial sequence corresponding to the DSL region available in the databases for comparison (GenBank accession no. Q53391). Analysis of the distribution of pilAI\textsubscript{II} in CF versus non-CF human strains (clinical and rectal isolates) showed that this pilin allele was significantly less likely (P = 0.0005) to be present in CF strains (18.3% vs 37.7%) (Table 1). This allele was present in 29.2% of environmental strains (Table 1).

Group III pilins

Of strains predicted to contain both pilA and additional DNA by amplification product size, several (Pa4494, Pa5024, Pa5112, Pa5122, Pa5223 and Pa141123) encoded pilins closely related to PilA of strain G7, a CF isolate previously found by Spangenberg et al. (1995) to have an unusually large pilin sequence, and strain PA14, a human clinical strain (Choi et al., 2002) (Table I, supplementary data). These strains were assigned to group III, and the
pilins designated PilAIII. The PilAIII proteins from these strains were 95–100% identical to the G7 pilin (Table I, Fig. 2). Immediately downstream of the pilin gene was an ORF that we named \( \text{tfpY} \) (type four pilin gene Y), which corresponds to the previously reported gene called ORF1 (Spangenberg et al., 1995). The protein encoded by \( \text{tfpY} \) is essentially identical in all strains carrying PilAIII (not shown) and is most homologous to FimB, a pilin accessory protein of unknown function from \( \text{D. nodosus} \) (Kennan et al., 2001). PCR with \( \text{tfpY} \)-specific primers showed that 41 of 159 human non-CF isolates (25–8%) and three of 24 environmental strains (12–5%) contained \( \text{tfpY} \), while only nine of 109 CF strains (8–2%) contained this gene (Table 1).

Table 1. Distribution of pilin alleles among isolates from various sources

<table>
<thead>
<tr>
<th>Accessory gene(s)</th>
<th>Accessory gene(s)</th>
<th>Accessory gene(s)</th>
<th>Accessory gene(s)</th>
<th>Accessory gene(s)</th>
<th>Accessory gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of strains ((n = 292))</td>
<td>137 (46–9%)</td>
<td>89 (30–5%)</td>
<td>53 (18–2%)</td>
<td>1 (0–3%)</td>
<td>12 (4–1%)</td>
</tr>
<tr>
<td>Clinical ((n = 95))</td>
<td>29 (30–5%)</td>
<td>36 (37–9%)</td>
<td>26 (27–4%)</td>
<td>0 (0–0%)</td>
<td>4 (4–2%)</td>
</tr>
<tr>
<td>Rectal ((n = 64))</td>
<td>18 (28–1%)</td>
<td>26 (40–6%)</td>
<td>15 (23–4%)</td>
<td>1 (1–6%)</td>
<td>4 (6–3%)</td>
</tr>
<tr>
<td>Clinical + rectal ((n = 159))</td>
<td>47 (29–6%)</td>
<td>62 (39–0%)</td>
<td>41 (25–8%)</td>
<td>1 (0–6%)</td>
<td>8 (5–0%)</td>
</tr>
<tr>
<td>Environmental ((n = 24))</td>
<td>14 (58–3%)</td>
<td>7 (29–2%)</td>
<td>3 (12–5%)</td>
<td>0 (0–0%)</td>
<td>0 (0–0%)</td>
</tr>
<tr>
<td>CF adult ((n = 66))*</td>
<td>47 (71–2%)</td>
<td>13 (19–7%)</td>
<td>4 (6–1%)</td>
<td>0 (0–0%)</td>
<td>2 (3–0%)</td>
</tr>
<tr>
<td>CF paediatric ((n = 43))†</td>
<td>29 (67–4%)</td>
<td>7 (16–3%)</td>
<td>5 (11–6%)</td>
<td>0 (0–0%)</td>
<td>2 (4–6%)</td>
</tr>
<tr>
<td>CF total ((n = 109))</td>
<td>76 (69–7%)</td>
<td>20 (18–3%)</td>
<td>9 (8–2%)</td>
<td>0 (0–0%)</td>
<td>4 (3–7%)</td>
</tr>
<tr>
<td>Clinical vs rectal P value</td>
<td>0–882</td>
<td>0–856</td>
<td>0–710</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Clinical + rectal vs environmental P value</td>
<td>0–011‡</td>
<td>0–484</td>
<td>0–244</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Clinical + rectal vs CF total P value</td>
<td>&lt;0–0001</td>
<td>0–0005</td>
<td>0–0005</td>
<td>0–470</td>
<td>ND</td>
</tr>
<tr>
<td>CF total vs environmental P value</td>
<td>0–401</td>
<td>0–361</td>
<td>0–793</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Adult CF vs paediatric CF P value</td>
<td>0–838</td>
<td>0–843</td>
<td>0–499</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

ND, Not done; numbers in these groups were too small for meaningful statistics.

*Combined adult CF strains from Toronto and Vancouver; there were no significant differences between these two groups.
†Combined paediatric CF strains from Toronto and Edinburgh; there were no significant differences between these two groups.
‡Numbers in bold are statistically significant.
Group IV pilin

Group IV currently contains a single isolate, Pa5196. The pilin gene of strain Pa5196 was novel, with a G+C content of 54.8 mol%, encoding a 155 aa prepilin protein with a 23 aa DSL (Fig. 2). The closest homologue of the Pa5196 pilin (40% identical, 55% similar) was PilE, a Neisseria meningitidis fimbrial protein precursor (GenBank accession no. S55496). Immediately downstream of pilA in Pa5196 were two novel ORFs. The first, tfpW, encodes a large hydrophobic hypothetical transmembrane protein, while the second, tfpX, encodes a putative pilin accessory protein most similar to TfpZ (below; 33% identical, 58% similar) and to PilB and FimB of Eikenella corrodens and D. nodosus, respectively. E. corrodens is a member of the HACEK group of bacteria (Haemophilus, Actinobacillus, Cardiobacterium, Eikenella, Kingella) which cause various human infections including infectious endocarditis.

Group V pilins

Strains Pa271457, Pa081061 and Pa110594 had novel pilA genes with a G+C content of 52.3–52.5 mol%. The mature pilins are one residue shorter than that of G7, with a predicted DSL of 29 aa (Fig. 2). These pilins were assigned to group V. A database search using BLASTP (Altschul et al., 1997) showed the prepilins were 53–54% identical (67–68% similar) to the G7 prepilin, and the next closest homologues were TFP-like proteins from the plant pathogen Ralstonia solanacearum (GenBank accession no. NP_518679). Adjacent to the pilA gene in each of these strains was a novel accessory gene tfpZ, which encodes a putative pilin accessory protein that is most homologous (31% identity, 53% similarity) to the PilB protein of E. corrodens strain VA1 (Villar et al., 1999). PCR with tfpZ-specific primers showed that eight of 15 non-CF human samples (5.3%) and four of 109 CF strains (3.7%) contained tfpZ DNA. These distributions are not significantly different, but the numbers in each group are low. The group V CF isolates were distributed equally among Toronto (one patient), Vancouver (one patient) and Edinburgh (two patients). This rare allele was not detected in the environmental samples tested.

Because this pilin type had not been reported previously, and we identified a number of strains with this unusual allele, strain typing by PFGE was performed to determine if the strains were clonal. PFGE of SpeI-digested chromosomal DNA showed that two of 13 group V strains were identical (Pa110594, Pa290177) and two others were closely related to these (Pa281457, Pa55177) (Fig. 3). Epidemiological analysis showed that identical strains Pa110594 and Pa290177 were collected from endotracheal tubes from different individuals in the same hospital, but the samples were collected over a month apart, and patients were in separate rooms. The closely related strains (Pa281457, Pa5517) were from two other area hospitals. The remaining tfpZ-bearing strains, including the four CF isolates, were unrelated.

Fig. 3. PFGE typing of group V strains. The relatedness of strains bearing group V pilins was assessed using PFGE of SpeI-digested chromosomal DNA, shown as a negative image of the ethidium bromide-stained gel. Strains Pa110594 and Pa290177 are identical (bold and underlined) and Pa281457 and Pa5517 (underlined) are closely related (Tenover et al., 1997). The remaining strains are considered unrelated.

Group I pilin genes and their predominance in CF isolates

Forty-seven of 159 (29.6%) human non-CF isolates, 14 of 24 (58.3%) environmental strains and 76 of 109 (69.7%) CF strains yielded PCR products of approximately 3 kb, the predicted size for pilA plus the glycosylation gene pilO. Of note, the P. aeruginosa PAO1 genome sequence (Stover et al., 2000) contains a completely unrelated gene also named ‘pilO’ (ORF PA5042), which is not involved in pilin glycosylation. Therefore, to reduce future confusion, and to harmonize the nomenclature with that of the other pilA accessory genes described above, we have renamed the glycosylation gene pilO (adjacent to pilA) as tfpO.

PCR with primers based on the published sequence of tfpO yielded amplification products for approximately half these strains, 28 of 159 human non-CF strains (17.6%), seven of 24 environmental strains (29.2%) and 45 of 109 CF strains (41.3%). The pilA1 and tfpO genes from the PCR-positive strains were highly homologous to those of the prototype strain 1244 (not shown). The original 3 kb PCR products from the remaining strains were digested with EcoRI, TaqI and Rsal to distinguish RFLPs. Three different RFLP patterns were obtained (not shown) and a representative product from each pattern was sequenced.

All three strains encoded PilA1 proteins with 70–74% amino acid identity to PilA1 from tfpO-containing strain 1244. These strains contained tfpO homologues encoding proteins with 87–92% identity with the 1244 TfpO (PilO)
protein. Sequence differences occurred mainly in the 5’ region of the tfpO gene, proximal to pilA. Based on the differences in the pilin and accessory genes, the group I pilin-containing strains could be divided into subgroups: TfpOa (Ia) strains, including the original strain 1244, and TfpOb (Ib) strains [Tables I (supplementary data) and 1]. All group Ia and Ib pilins sequenced in this study contained the terminal serine residue recently shown to be the site of glycosylation in strain 1244 (Comer et al., 2002).

**Analysis of novel pilin proteins**

The potential role of TfpOa and other novel accessory proteins in pilin modification was examined by separation of sheared pilin preparations from representative strains on SDS-PAGE. In strain 1244, glycosylation by TfpOa of the C-terminal serine residue of PilAI with a lipopolysaccharide O-antigen unit was previously shown to slow pilin migration on SDS-PAGE (Comer et al., 2002). Pilins from strains PAK (group II), 1244 (group I) and a tfpOa::Gmr insertion mutant of strain 1244 were used as controls (Fig. 4a). TfpOa strain Pa131533 pilin migrated more slowly than its predicted mass of 15·6 kDa, similar to the glycosylated 1244 pilin (Fig. 4a). Group I pilin proteins from other strains also showed reduced migration (not shown). These results show that strains carrying either version of TfpO are capable of producing modified pilins. Pilins from strains carrying tfpY or tfpZ migrated more rapidly than their predicted sizes of 17·4 and 17·7 kDa, respectively (Fig. 4a), suggesting that they may be processed differently from other pilins, generating a shorter form, or post-translationally modified. These possibilities are currently under investigation.

Interestingly, the pilin from group IV strain Pa5196, which contains tfpW–tfpX downstream of pilAIV, migrated more slowly than predicted by its molecular mass (Fig. 4b), similar to the glycosylated form of group I pilin. The pilin from Pa5196 does not contain the terminal serine residue that is glycosylated in group I pilins, suggesting that post-translational modification, if present, is likely to be at a position different from that of PilAI. The pilin of this strain is related to those of Neisseria (see below), which are modified on internal serine residues (Forest et al., 1999; Marceau & Nassif, 1999). Based on protein structure and hydropathy analyses (Fig. 5), TfpW is predicted to have multiple transmembrane domains, similar to TfpO (Castric, 1995), although there is no primary amino acid sequence homology between the proteins. We are in the process of generating Pa5196 mutants lacking TfpW to demonstrate whether this protein is involved in modification of the pilin.

**Phylogenetic analysis of PilA proteins from P. aeruginosa confirms that there are five distinct groups**

Representative amino acid sequences from each pilin group (I–V) and homologous proteins with the highest degree of similarity (based on BLASTP analysis) from species other than P. aeruginosa were aligned using CLUSTAL_X (Thompson et al., 1997). Phylogenetic trees were generated with PHYLIP 3.5 using the neighbour-joining method (Fig. 6). The pilins clustered unequivocally into the same groups to which they were previously assigned based on the length of the C-terminal DSL and the presence (or absence) and identity of characteristic accessory genes. High bootstrap confidence values confirm the clades within the tree are statistically significant. Use of only the highly conserved N-terminal regions of the proteins generated similar trees (not shown), demonstrating that the clades are not reflective only of sequence differences in the divergent C termini of the proteins.

Of interest, Pa5196 PilAIV groups more closely with E. corrodens PilA1, Neisseria gonorrhoeae PilE and N. meningitidis PilE than other P. aeruginosa pilins within this data set. PilAII from the anomalous group II strain Pa5325 groups more closely with D. nodosus FimZ than other P. aeruginosa PilAII proteins, but its relationship to P. aeruginosa group II pilins is still statistically significant (bootstrap confidence value of 78%).

**DISCUSSION**

P. aeruginosa is an opportunistic pathogen that lives in a broad range of environments and infects hosts across the evolutionary spectrum. Common virulence factors have
been shown to be required for pathogenesis in various hosts ranging from amoebae, yeast and nematodes through plants and mammals (Cosson et al., 2002; D’Argenio et al., 2001; Hogan & Kolter, 2002; Pukatzki et al., 2002; Yorgey et al., 2001). Also, no obvious differences have been detected with respect to pathogenicity between clinical and environmental strains (Alonso et al., 1999), suggesting any strain is equally capable of colonization or infection given the appropriate circumstances. Because P. aeruginosa virulence factors, including TFP, have been well studied, we were surprised when our initial analysis of pilin diversity in non-CF P. aeruginosa clinical isolates revealed previously unrecognized pilin types and accessory genes. To explain this observation, we hypothesized that CF isolates, upon which many P. aeruginosa researchers have traditionally focused, may carry a restricted range of pilin types, leading to an underestimation of the diversity of pilins in this species. We then analysed the pilin types of a large set of CF strains to test this supposition.

Comparison of the distribution of pilin alleles between CF and non-CF human isolates supports our hypothesis. Group I pilins, which can be modified by glycosylation through the activity of TfpO, were identified in 69·7% of genetically distinct CF isolates but only 29·6% of non-CF human isolates, a highly significant difference (P<0·0001; Table 1). This distribution pattern was apparent regardless of whether CF strains were collected in Toronto, Vancouver or Edinburgh CF clinics (P=0·298), showing the high prevalence of group I TFP in the CF population is not due to a regional bias. Interestingly, 29 of 43 paediatric CF isolates (67·4%) belong to pilin group I (Table 1), similar to adult CF isolates (71·2%; P=0·838), demonstrating that the bias in pilin type is already evident in patients early in the progression of CF disease. Correspondingly, group II and III pilins were more likely to be present in non-CF human strains compared with CF isolates (P=0·0005 and P=0·0005, respectively). The remaining types are currently too rare for meaningful analyses.

Whether strains carrying group I pilins are better at colonizing the unique CF lung environment than strains with other pilin types, or have associated characteristics that make them more likely to persist in the CF lung, remains to be determined experimentally. It is intriguing that the percentage of group I pilins within the environmental strains tested for this study was similar to that within CF isolates (58·3% vs 69·7%, P=0·401). The environment has been suggested to be the source of P. aeruginosa in CF patients (Speert et al., 2002), and the similarity in pilin distribution supports this theory. In contrast, other human isolates, whether they are rectal colonizing strains or clinical isolates, appear to have approximately equal distributions of strains within pilin groups I, II and III. This remarkably parallel distribution pattern may imply that opportunistic, non-CF P. aeruginosa Fig. 5. Kyte–Doolittle hydropathy plots of TfpO and TfpW proteins. The TfpOa and TfpOb proteins show almost identical hydropathy plots with some variation at the N and C termini. Although TfpW has no sequence similarity to TfpO, its hydropathy plot has similar predicted transmembrane domains. The plots were generated in GENE RUNNER 3.00 (Hastings Software) using the predicted amino acid sequences. The scales are equivalent for comparisons and positions of the amino acid residues are shown on the x-axis.

J. V. Kus and others
infections, which are often associated with use of medical devices, may arise from the patient's own transient flora. We showed previously (Gardam et al., 2002) that even severely immunocompromised transplant patients are most frequently infected by their own rectal flora.

It is interesting that strains capable of expressing the type of pilin that can be glycosylated are most prevalent in CF patients. The role of pilin glycosylation in host–bacterium interactions is currently unknown. Castric (1995) showed that inactivation of tfpO (pilO) had no apparent functional consequences with respect to pilus expression and twitching motility, but did not determine the effect of the mutation on adherence to host cells. The position of the glycosyl moiety at the pilin's C terminus, immediately adjacent to the putative receptor-binding domain, was suggested to play a role either in modulating cell–cell binding or in protecting the adhesive portion of the pilus from proteolytic degradation (Comer et al., 2002). Further investigation is necessary to determine whether pilin glycosylation confers a colonization advantage in the CF lung.

While all strains tested possessed the pilin structural gene, some strains lacked twitching motility by the agar stab assay. Biosynthesis, function and regulation of TFP remain complex and poorly defined processes involving more than 50 different genes, including some of unknown function (Jacobs et al., 2003; Mattick, 2002). Therefore, strains that lack macroscopic twitching motility may still produce adhesive pili but be unable to twitch due to defects in pilus motor function or chemotaxis, or may lack pili altogether due to mutations in a variety of genes. In two cases (strains Pa5226 and Pa5289), we found that the pilin locus was disrupted with different insertion sequence elements (Table I, supplementary data). In general, strains isolated from patients at a single time point may represent non-twitching derivatives of a colonizing strain that originally had twitching motility. Loss of twitching motility has previously been reported in serial isolates of strains from chronically colonized CF patients (Mahenthiralingam et al., 1994). This observation is supported by differences in twitching motility seen in our paediatric versus adult CF isolates.

---

**Fig. 6.** Phylogenetic relationships between PilA and related proteins. A neighbour-joining tree was produced in PHYLIP 3.5 using the Jones–Taylor–Thornton matrix for amino acid substitutions and visualized using MEGA2.1. The pilin groups that were originally generated based on presence of accessory gene(s) and length of the C-terminal DSL sequence are strongly supported by this analysis. Bootstrap replicates are indicated on branches as percentages, and indicate confidence for the relationship. GenBank accession numbers are indicated and strains from this study are underlined. Pa, *P. aeruginosa* (bolded); Mb, *Mycobacterium bovis*; Dn, *D. nodosus*; Nm, *N. meningitidis*; Ng, *N. gonorrhoeae*; Ec, *E. corrodens*; Rs, *R. solanacearum*.

---

<table>
<thead>
<tr>
<th>Group Ia</th>
<th>Group Ib</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pa SBI-N PilA&lt;sub&gt;a&lt;/sub&gt;, AAK68046</td>
<td></td>
</tr>
<tr>
<td>Pa5226 PilA&lt;sub&gt;a&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>Pa131533 PilA&lt;sub&gt;a&lt;/sub&gt;, AAM52057</td>
<td></td>
</tr>
<tr>
<td>Pa5659 PilA&lt;sub&gt;a&lt;/sub&gt;, AAM62146</td>
<td></td>
</tr>
<tr>
<td>Pa 1244 PilA&lt;sub&gt;a&lt;/sub&gt;, P18774</td>
<td></td>
</tr>
<tr>
<td>Pa302025 PilA&lt;sub&gt;a&lt;/sub&gt;, AAM62144</td>
<td></td>
</tr>
<tr>
<td>Mb &quot;pilin subunit&quot; AAS53562</td>
<td></td>
</tr>
<tr>
<td>Mb &quot;pro-pilin&quot; AAM53562</td>
<td></td>
</tr>
<tr>
<td>Mb &quot;pro-pilin&quot; AAS5365</td>
<td></td>
</tr>
<tr>
<td>Dn Fin2 YQBB2DZ</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pa5235 PilA&lt;sub&gt;a&lt;/sub&gt;, AAM44066</td>
</tr>
<tr>
<td>Pa PK PilA&lt;sub&gt;a&lt;/sub&gt;, P02973</td>
</tr>
<tr>
<td>Pa PAK PilA&lt;sub&gt;a&lt;/sub&gt;, AAG07913</td>
</tr>
<tr>
<td>Pa281298 PilA&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Nm PilE S5496</td>
</tr>
<tr>
<td>Ng PilE CAA8931</td>
</tr>
<tr>
<td>Ec PilA1 AAC28468</td>
</tr>
<tr>
<td>Pa5196 PilA&lt;sub&gt;a&lt;/sub&gt;, AAM52059</td>
</tr>
<tr>
<td>Rs PilA1 NP518678</td>
</tr>
<tr>
<td>Rs PilA NP518679</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pa811061 PilA&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Pa100594 PilA&lt;sub&gt;a&lt;/sub&gt;, AAM52055</td>
</tr>
<tr>
<td>Pa281457 PilA&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pa5024 PilA&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Pa141123 PilA&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Pa G7 PilA&lt;sub&gt;a&lt;/sub&gt;, AAC41410</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group III</th>
</tr>
</thead>
</table>

---

**http://mic.sgmjournals.org** 1323
(81.4 % vs 54.5 %; P = 0.008), even though both populations are colonized predominantly with group I strains.

Variation at the pilA locus in *P. aeruginosa* is interesting from the perspective of bacterial evolution. Since *P. aeruginosa* contains only a single pilin locus, it must either undergo mutations within the existing pilA gene or acquire foreign pilA genes through horizontal genetic transfer to generate novel pilin types. Both scenarios were evident in this study. We identified five distinct groups of pilins in *P. aeruginosa*, including some with homology to pilins from other mammalian (*D. nodosus* and *E. corrodens*) (Klenman et al., 2001; Villar et al., 2001) or plant (*R. solanacearum*, *Xanthomonas campestris*) (da Silva et al., 2002; Liu et al., 2001) pathogens, suggesting possible horizontal transfer from organisms sharing common environments.

There was also evidence of sequence variation within pilin groups, ranging from single to multiple amino acid changes (Fig. 2). Notably, each pilin type is stringently associated with specific accessory gene(s), i.e. pilAIII is always found with tfpY and never with the group V gene tfpZ. This invariant relationship implies that the pilin and its accessory gene(s) were acquired together, and allows the diagnostic use of accessory-gene-specific PCR to determine the pilin type of the strain.

The presence immediately adjacent to the pilA locus of a tRNA<sup>Thr</sup> gene suggests that bacteriophage-mediated transduction may be involved in generation of pilin diversity in *P. aeruginosa*, since tRNA genes represent preferred sites for bacteriophage integration (Reiter et al., 1989). Recent DNA sequence analysis of the highly infectious *P. aeruginosa* bacteriophage phiKZ showed it carries a tRNA<sup>Thr</sup> gene in its genome (Mesyanzhinov et al., 2002). Despite linkage of TFP to genetic competence in other species, *P. aeruginosa* has not been found to be competent for natural transformation, making transduction a more likely mechanism of pilin variation in this species.

Among the most striking and significant amino acid sequence variations we noted in this study were differences in the pilin C-terminal DSL region (Fig. 2) that has been implicated as the binding domain for the eukaryotic glycolipid asialoGM1 (aGM1). Previous studies of receptor binding using peptide mimics of the C-terminal DSLs of group II pilins from strains PAK, PAO1 and KB7 and the group I pilin from strain P1 showed that all peptides bound to aGM1 despite differences in amino acid composition (Campbell et al., 1997). In contrast, a recent study (Schroeder et al., 2001) using whole, piliated cells showed that a number of clinical isolates (of undefined pilin type), as well as laboratory strains including PAO1 and PAK, did not demonstrate differences in binding to host cells in the presence or absence of exogenously added aGM1 receptor. Therefore, the effect on host cell adhesion of the major differences in the length and amino acid sequence of the putative receptor binding domain in these novel pilins, and the effect of their presentation as peptides compared to intact pilin or on whole bacterial cells, require further studies. Similarly, the possibility that there may be other potential host receptor(s) for these novel pilins requires investigation.

In conclusion, our analyses of the type IV pilin gene locus in *P. aeruginosa* have led to the identification of novel pilins and accessory genes, expanding the known diversity of type IV pilins within this species and providing the basis for a new system of pilin nomenclature. This information will be useful for the design of comprehensive pilin-based vaccines. Our intriguing finding that the preponderance of CF isolates have group I pilin genes is, to our knowledge, the first example of a *P. aeruginosa* virulence factor allele that is strongly linked with CF, and will be the focus of future research.

**ACKNOWLEDGEMENTS**

We thank the many generous individuals listed in Methods for providing strains for this study. We also thank Dr Dave Gutman, John Stavrinides, Dr J.-M. Moncalvo, and the Bioinformatics groups at HSC for helpful discussions about phylogenetics. Winston Wong and Raynah Fernandes provided excellent technical assistance. This work was supported by grant MOP-49577 from the Canadian Institutes of Health Research to L. L. B. J.-V. K. is a Harron Scholar and a CIHR Strategic Training Fellow in Cell Signalling in Mucosal Inflammation and Pain (TP-53877).

**REFERENCES**


Alternative sigma factor AlgT (AlgU).


Is surveillance for multidrug-resistant enterobacteriaceae an effective infection control strategy in Drosophila as a model host for Pseudomonas aeruginosa infection. J Bacteriol 183, 1466–1471.


targeting the C-terminal region of the pilin structural protein.


