Use of suppression subtractive hybridization to examine the accessory genome of the Liverpool cystic fibrosis epidemic strain of *Pseudomonas aeruginosa*

Catherine H. M. Smart, Martin J. Walshaw, C. Anthony Hart and Craig Winstanley

1Division of Medical Microbiology and Genitourinary Medicine, University of Liverpool, Liverpool L69 3GA, UK
2Regional Adult Cystic Fibrosis Unit, Cardiothoracic Centre, Liverpool L14 3PE, UK

The Liverpool epidemic strain (LES) of *Pseudomonas aeruginosa* has been highly successful at colonizing cystic fibrosis (CF) patients throughout the UK, has replaced previously established strains in CF patients, has caused infections of non-CF parents of CF patients, and can cause greater morbidity in CF than other strains of *P. aeruginosa*. Using suppression subtractive hybridization (SSH) to identify strain-specific sequences, a diagnostic test for the LES based on PCR amplification of SSH sequence PS21 had previously been developed. In this study, the SSH sequence database of LES was substantially increased, using both extension of previous sequences and new rounds of subtraction. Of 92 SSH sequences identified as present in the LES but absent from strain PAO1, 25 were assessed for prevalence amongst a strain panel consisting mainly of LES and non-LES CF isolates. Preliminary analysis of genome sequence data indicated that all SSH sequences that were LES specific or found only rarely in other strains of *P. aeruginosa* were present on one of three contigs. All of the SSH sequences screened were either unstable amongst LES isolates or were not completely LES specific. Rare false positives were found with the PS21 test. The authors suggest that a second PCR assay designed to detect SSH sequence LESF9 can be used to confirm the identity of the most prevalent CF epidemic lineage in the UK.

INTRODUCTION

Cystic fibrosis (CF) is the most common life-threatening inherited disease amongst Caucasians. Although patient management and infection control policies have increased life expectancy to approximately 31 years (www.cftrust.org.uk), most patients eventually succumb to morbidity and mortality due primarily to chronic bacterial infections of the lung. The most common and important CF pathogen is *Pseudomonas aeruginosa*, which establishes chronic, long-term pulmonary infections associated with a biofilm lifestyle and the production of alginate, leading to a mucoid phenotype. Once established, *P. aeruginosa* infections cannot be eradicated, and CF patients suffer decline in lung function associated with periodic exacerbations. Until a decade ago, it was generally accepted that CF patients acquired *P. aeruginosa* infections individually from the environment, and thus carried strains unrelated to those of other patients. In 1996, Cheng et al. (1996) reported the spread of a drug-resistant strain of *P. aeruginosa* (named the Liverpool epidemic strain, LES) amongst patients in a children’s CF centre in Liverpool, UK. Subsequently other *P. aeruginosa* CF ‘epidemic’ strains have been reported in Australia (Armstrong et al., 2003; O’Carroll et al., 2004), Manchester, UK (Jones et al., 2001), and the Midlands, UK (Scott & Pitt, 2004; Chambers et al., 2005), raising serious questions about whether segregation strategies might be required to counter transmission. Recent analysis of post-2000 patient samples has identified the LES in 79 % of 80 *P. aeruginosa*-colonized CF patients in the Liverpool adult CF centre, confirming the spread and longevity of LES infections (Panagea et al., 2003). In addition to its transmissibility, the LES has been shown to replace previously established strains of *P. aeruginosa* (superinfection; McCallum et al., 2001) and has infected both non-CF parents of a CF patient, causing significant morbidity (McCallum et al., 2002). Furthermore, there is evidence for greater morbidity amongst CF patients colonized with the LES compared to those carrying non-epidemic strains of *P. aeruginosa* (Al

**Abbreviations:** CF, cystic fibrosis; FHA, filamentous haemagglutinin; LES, Liverpool epidemic strain; SSH, suppression subtractive hybridization.
Aloul et al., 2004). Recently, we have demonstrated that some isolates of the LES express an unusual ‘hypervirulence’ phenotype characterized by early expression and overexpression of quorum-sensing-regulated virulence genes (Salunkhe et al., 2005).

A survey of 31 CF centres in England and Wales, in which over 1200 isolates of P. aeruginosa were analysed, has identified the LES as the most common clone, being present in 48% of CF centres and accounting for 11% of the isolates (Scott & Pitt, 2004). The strain has also been identified amongst CF isolates in Scotland (Edenborough et al., 2004). The antibiotic susceptibilities of LES isolates vary widely (Panagea et al., 2003; Salunkhe et al., 2005), and the ‘gold-standard’ typing method, macrorestriction analysis using PFGE, is time-consuming and requires specialist equipment. A PCR-based diagnostic test for the LES (Parsons et al., 2002) has been evaluated in studies with a large number of CF isolates and found to be 100% in concordance with identification using PFGE (Panagea et al., 2003; Scott & Pitt, 2004). Furthermore, the test has successfully been applied directly to patient sputum samples (Panagea et al., 2003). The PCR assay is used to amplify one of a number of sequences (named PS21) identified previously using suppression subtractive hybridization (SSH), a technique that enables the identification of sequences present in one strain but absent from another (Parsons et al., 2002; Winstanley, 2002). However, more recently, we and others have identified a small number of non-LES isolates that are positive using the diagnostic test (Lewis et al., 2005; F. W. Scott and T. L. Pitt, unpublished data).

P. aeruginosa contains a core genome that possesses a highly conserved backbone comprising the vast majority of the genome and including most of the recognized virulence factors (Ernst et al., 2003; Wolfgang et al., 2003), and an accessory genome (or mobilome) that includes mobile genetic elements and genomic islands (Larbig et al., 2002; Spencer et al., 2003). The accessory genome of the LES is likely to contain not only strain-specific sequences that could be targeted for identification, but also novel genomic islands that may contribute to the extra abilities possessed by this strain. A genome sequence project to identify all such elements was initiated in collaboration with the Sanger Centre, Hinxton, UK (http://www.sanger.ac.uk/Projects/P.aeruginosa/).

In this study, we report the further characterization of PS21 and two other previously identified subtracted sequences, and report the results of further SSH to identify a much larger number of sequences present in the accessory genome of LES. By using preliminary data from the genome sequence project we further group many of these sequences into islands or gene clusters. Finally, we combine the data to report the development of improved PCR assays to identify the most common epidemic strain of P. aeruginosa in the UK CF community.

METHODS

Bacterial strains. Isolate LES400 was obtained from a CF patient and has been used previously for SSH (Parsons et al., 2002). The reference strain (driver) for SSH experiments was strain PA01, for which the entire genome sequence is known (Stover et al., 2000). For prevalence studies, we had several categories of strain panels (Table 1). Twenty-two isolates of the LES, 12 isolates of the Midlands1 epidemic strain, and 11 isolates of the Manchester epidemic strain were taken, mainly from isolates obtained during the survey of CF isolates in England and Wales (Scott & Pitt, 2004). Strains genetically distinct from LES isolates from adult patients in Liverpool were included from our previous SSH study (Parsons et al., 2002). In addition, we obtained 43 more recent, uncharacterized, non-LES isolates (PS21 negative) from the Liverpool adult CF unit, and 32 uncharacterized non-LES CF isolates from a neighbouring adult CF unit at Clatterbridge Hospital, Merseyside, UK. We also screened representative isolates of clone C (Römling et al., 1994) and strain PA14 (Rahme et al., 1995). In addition, four PS21 false-positive (non-LES) isolates were also screened (gifts of Ty Pitt and Fiona Scott, Health Protection Agency, UK).

GenomeWalker extension of sequences. Previously identified subtracted sequences [PS21, PSA and PS54 (Parsons et al., 2002)] were extended using the PCR-based GenomeWalker kit (Clontech), following the manufacturer’s instructions.

Construction and screening of subtraction libraries. Genomic DNA was isolated from P. aeruginosa LES400 and strain PA01 as described previously (Winstanley & Hart, 2000). SSH was carried out using the CLONTECH PCR-Select Bacterial Genome Subtraction kit (Clontech) as recommended by the supplier, but with a hybridization temperature of 73°C. In the hybridizations, DNA from the LES was used as the tester and DNA from strain PA01 was used as the driver. PCR amplicons obtained following SSH were cloned into pGEM-T (Invitrogen). The subtraction library of RsaI fragments thus constructed was screened by sequencing of plasmid DNA extracted from individual clones (Lark Technologies) using M13 forward and reverse vector primers. BLAST searches at the P. aeruginosa PA01 genome project website (http://www.pseudomonas.com) were used to determine the presence or absence of sequences in the PA01 genome. Test-specific sequences were further analysed using BLASTN and BLASTX searches of the general database using the site http://www.ncbi.nlm.nih.gov. The relative positions of sequences in the LES genome were determined by BLASTN searching the preliminary data available at the site http://www.sanger.ac.uk/ cgi-bin/blast/submit/nblast/ p_aeruginosa. Subsequent analysis of downloaded sequence data was carried out using the program ARTEMIS.

PCR amplification screening of strains. Oligonucleotide primers (Sigma-Genosys) used in PCR assays are listed in Table 2 along with the annealing temperatures used. DNA for PCR amplification was prepared by making a suspension of a few colonies in 200 μl 5% Chelex-100 (Sigma) solution. After vigorous mixing, the suspension was boiled for 5–10 min. Following centrifugation, 150 μl of supernatant containing the DNA was removed and stored at −20°C. Typically, 1 μl of this DNA was used directly in 25 μl volumes containing 1-25 units Taq DNA polymerase (Promega), 1× TaqMaster (Helena Biosciences), 300 nM each primer, 1× Taq buffer, 2-5 mM MgCl2 and 100 μM nucleotides (dATP, dCTP, dGTP, dTTP). Amplifications were carried out in an Eppendorf MasterCycler thermal cycler for 30 cycles consisting of 95°C (1 min), annealing temperature (1 min) and 72°C (2 min), with an additional extension time at 72°C (10 min) following completion of the 30 cycles.

PFGE. PFGE of SphI-digested genomic DNA was carried out using 1% (w/v) agarose gels and interpreted according to the protocol of Tenover et al. (1995).

Downloaded from www.microbiologyresearch.org by IP: 54.70.40.11 On: Wed, 26 Dec 2018 22:30:42
RESULTS AND DISCUSSION

Extending PS21 and other subtracted sequences

Using the GenomeWalker system, the PS21 sequence (previously 400 bp) was extended upstream and downstream to 1256 bp. Of the extended sequence, 750 bp shared significant similarity with the C termini of \textit{merA} (mercuric reductase) genes/proteins in the database. The best predicted protein match was with a \textit{Streptomyces} sp. MerA (Table 3). All of the best matches were to proteins from Gram-positive bacteria. The PS21 diagnostic sequence itself overlapped with 50 bp of the predicted \textit{merA} gene. PS54, another sequence identified previously as unique to the LES (Parsons \textit{et al.}, 2002), was extended from 425 to 860 bp. A predicted protein derived from nucleotide positions 131–667 contained a conserved domain found in \textit{N}-acetyltransferases (Table 3).

FHA-like genes in \textit{P. aeruginosa}

Although not unique to the LES, the previously identified subtracted sequence PSA shared similarity with the C-terminal sequence of \textit{Neisseria meningitidis} FhaB, a putative adhesin with similarity to filamentous haemagglutinin (FHA) (Parsons \textit{et al.}, 2002), and was therefore potentially virulence related. Because the PSA sequence matched strain PAO1 downstream, we extended only upstream to a total of 2751 bp. Analysis of this extended region identified sequence PSA as a replacement sequence for the C terminus of the strain PAO1 ORF PA2462, encoding an FHA-like putative adhesion-related/haemagglutinin protein. FHA is a major virulence factor involved in the adherence of \textit{Bordetella} species to epithelial cell surfaces (Locht \textit{et al.}, 1993). In conducting a web-based search of the strain PAO1 genome sequence database, Croft \textit{et al.} (2000) identified six FHA-like genes. These have been assigned ORF numbers PA0041, PA0690, PA2462, PA4082, PA4541 and PA4625 at the genome database web site (http://www.pseudomonas.com). There is a striking similarity between PA0041 and PA2462, which share an identical sequence over most of their length, suggesting that this pair arose as a result of gene duplication (Croft \textit{et al.}, 2000). We carried out database searches using the C-terminal 100 amino acids from ORFPA0041 and ORF PA2462 from the genome-sequenced strains PAO1 and PA14 (available from http://www.ncbi.nlm.nih.gov) and found a different best match for each of these sequences and the LES PA2462 sequence, indicating that C-terminal polymorphisms in these proteins are not a rare phenomenon. Furthermore, there is a second variable region of ORF PA2462 further upstream (associated with PS9).

It has been demonstrated that the insertion of a transposon into the FHA-like genes PA0041, PA2462 and PA4625 results in reduced virulence in a rat model of chronic respiratory infection (Potvin \textit{et al.}, 2003), suggesting a role for each of these genes in pathogenicity. By extending SSH sequence PSA, we found that the LES equivalent of PA2462 contained an alternative C terminus as well as an alternative variable region further upstream. By screening data now available

### Table 1. Strains used in this study

<table>
<thead>
<tr>
<th>Isolates</th>
<th>Genotype</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LES109, LES400, LES397, LES424, LES427, LESG21/23, LES416, LES417, LES430, LES431, LESB58, LESH252</td>
<td>LES</td>
<td>Known LES isolates, Liverpool CF centres</td>
<td>Parsons et al. (2002);</td>
</tr>
<tr>
<td>9245, 8916, 9057, 9069, 9190, 9231, 9608, 9750, 9946, 10066, 10112, 10397</td>
<td>Midlands1 epidemic strain</td>
<td>CF isolates, various CF centres in England and Wales</td>
<td>Salunkhe et al. (2005);</td>
</tr>
<tr>
<td>73, 126, 127, 43, 103, 139, 166, 393, 176, 214, 229, 267, 289, 291, 313, 438</td>
<td>Non-LES, genetically different (PFGE typing)</td>
<td>CF isolates, Adult CF Unit, Liverpool</td>
<td>Parsons et al. (2002)</td>
</tr>
<tr>
<td>43 isolates</td>
<td>Non-LES (PS21 negative)</td>
<td>CF isolates, Adult CF Unit, Liverpool</td>
<td></td>
</tr>
<tr>
<td>32 isolates</td>
<td>Non-LES (PS21 negative)</td>
<td>CF isolates, Adult CF Unit, Clatterbridge Hospital</td>
<td></td>
</tr>
<tr>
<td>PAO1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA14</td>
<td></td>
<td></td>
<td>Rahme \textit{et al.} (1995)</td>
</tr>
<tr>
<td>Clone C</td>
<td></td>
<td></td>
<td>Römling \textit{et al.} (1994)</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence (5’–3’)</td>
<td>Amplicon size (bp)</td>
<td>Target</td>
</tr>
<tr>
<td>----------</td>
<td>------------------</td>
<td>--------------------</td>
<td>---------</td>
</tr>
<tr>
<td>LESA3F</td>
<td>TCAGCCCATAGAAATTGC</td>
<td>375</td>
<td>LESA3</td>
</tr>
<tr>
<td>LESA3R</td>
<td>TCAAATCCGCTCAGCAAGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESA11F</td>
<td>CGGAGATGAGATTGGCCA</td>
<td>376</td>
<td>LESA11</td>
</tr>
<tr>
<td>LESA11R</td>
<td>TTGCTCTGCTGCTTCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEB7F</td>
<td>CTGATCCAGCTTCTTCTGC</td>
<td>546</td>
<td>LEB7</td>
</tr>
<tr>
<td>LEB7R</td>
<td>TGTTAAACGGAAGGCTCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEB9F</td>
<td>TGAGAGGCGCATCCAGA</td>
<td>557</td>
<td>LEB9</td>
</tr>
<tr>
<td>LEB9R</td>
<td>CCGGTAATGTCCTCTTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LED11F</td>
<td>AAGAGGCTCTGCAAGTCC</td>
<td>622</td>
<td>LED11</td>
</tr>
<tr>
<td>LED11R</td>
<td>TAAATCCGCTGCTATCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEE4F</td>
<td>ACAAGAAGAAGACAGGAC</td>
<td>367</td>
<td>LEE4</td>
</tr>
<tr>
<td>LEE4R</td>
<td>GAGATTTTCCCCGACGACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEE6F2</td>
<td>GGAATATTCAAGCCTGAC</td>
<td>516</td>
<td>LEE6</td>
</tr>
<tr>
<td>LEE6R2</td>
<td>CGCAGACAGAATCCACAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEF8F</td>
<td>AAGAGGCTCTGCAAGTCC</td>
<td>286</td>
<td>LEF8</td>
</tr>
<tr>
<td>LEF8R</td>
<td>CCAACAGATTTCCCTTCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEF9F</td>
<td>AACACTTGCTCCTATCGC</td>
<td>431</td>
<td>LEF9</td>
</tr>
<tr>
<td>LEF9R</td>
<td>CACGATATCCGCAAGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESG7F</td>
<td>AAAATCCGGAAGATGATGC</td>
<td>748</td>
<td>LESG7</td>
</tr>
<tr>
<td>LESG7R</td>
<td>CTTTCCAGATCCGCCGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESG9F</td>
<td>AGTTCAGTTCTCTCAGCA</td>
<td>383</td>
<td>LESG9</td>
</tr>
<tr>
<td>LESG9R</td>
<td>AGGAGGCGCAGATAGATCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESG11F</td>
<td>TGCGCGTGAAGAGGTCTCC</td>
<td>429</td>
<td>LESG11</td>
</tr>
<tr>
<td>LESG11R</td>
<td>CCTCAAAACTGAGTTCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESH4F</td>
<td>AACTGCAATAAGCCTCTCC</td>
<td>339</td>
<td>LESH4</td>
</tr>
<tr>
<td>LESH4R</td>
<td>CGACGGACAGCCTTTGAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LES6F</td>
<td>CGTTTGGCCAGACTGAGA</td>
<td>554</td>
<td>LES6</td>
</tr>
<tr>
<td>LES6R</td>
<td>AAGTTTCCAGGCAATTGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LES1F</td>
<td>TCGGCCGTAATGTCTCTA</td>
<td>410</td>
<td>LES1</td>
</tr>
<tr>
<td>LES1R</td>
<td>TGAGACCAGCAGTGGAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LES4F</td>
<td>GCGCACAATGACGATAGAG</td>
<td>635</td>
<td>LES4</td>
</tr>
<tr>
<td>LES4R</td>
<td>TCTACGGGTCTCGAGTCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LES8F</td>
<td>TCACTCGACGGAAATGCG</td>
<td>396</td>
<td>LES8</td>
</tr>
<tr>
<td>LES8R</td>
<td>CTGCTCTCTGCTCGAGAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESA9F</td>
<td>AAGGATCGAGAGGCGAAG</td>
<td>348</td>
<td>LESA9</td>
</tr>
<tr>
<td>LESA9R</td>
<td>TGGCGCTATGATGCGATTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESB3F</td>
<td>TGGCCGCGACTTTCTATGA</td>
<td>268</td>
<td>LESB3</td>
</tr>
<tr>
<td>LESB3R</td>
<td>ATTCTCGGCCTGCTCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESC6F</td>
<td>CATTAAAGCTCTCTGCTAG</td>
<td>401</td>
<td>LESC6</td>
</tr>
<tr>
<td>LESC6R</td>
<td>TAGATATTCTGCCTGCAGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESD9F</td>
<td>TCAACAGAGCTCCCAACG</td>
<td>554</td>
<td>LESD9</td>
</tr>
<tr>
<td>LESD9R</td>
<td>GATGGACAGATAATTGCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LES7F</td>
<td>TTGATGATGATGCGCAGCA</td>
<td>274</td>
<td>LESE7</td>
</tr>
<tr>
<td>LES7R</td>
<td>TCGGTCAATTGGGCGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESF1F</td>
<td>CAATCCACCGGAAACAG</td>
<td>391</td>
<td>LESF1</td>
</tr>
<tr>
<td>LESF1R</td>
<td>CCACTCGGAATTACCTCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESH5F</td>
<td>TTATGCTGCGTGTCGCA</td>
<td>441</td>
<td>LESH5</td>
</tr>
<tr>
<td>LESH5R</td>
<td>ATGCCACGATGATATC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESH12F</td>
<td>GAATGCGGCTGACCTACAG</td>
<td>785</td>
<td>LESH12</td>
</tr>
<tr>
<td>LESH12R</td>
<td>CGAGTCCAGAAACGCTTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
from the LES genome sequence project, we have been able not only to confirm this variation, but also to identify downstream of the PA2462 ORF a region of repeated FHA-like sequences, including an identical truncated C terminus equivalent to the PAO1 C terminus of ORF PA2462. In reporting the genome sequence of \textit{N. meningitidis} serogroup A strain Z2491, Parkhill et al. (2000) refer to repeat mediated rearrangement of the 3’ ends of a gene encoding a surface-exposed FHA-like protein. The sequence downstream of this gene contained complex repeated sequences also found within the gene. Sequencing of another \textit{N. meningitidis} strain (Z4259, serogroup C) revealed a similar repeat structure, but with additional DNA inserted into the 3’ sequence of NMA0688 at one of the repeats. This suggested that the 3’ end of the gene can be altered by recombination with alternative 3’ ends. Furthermore Klee et al. (2000) have reported that the C terminus of the FHA-like protein identified as similar to the PSA sequence (flaB of \textit{N. meningitidis}) is variable between strains. The authors hypothesize that repeated sequences could provide for variation in the C-terminal portion of the Fhb protein. They also report evidence that the flaB gene is present in two copies in at least some strains of \textit{N. meningitidis}.

The PA2462 peptide sequence shares significant similarity to \textit{Pseudomonas putida} KT2440 ORFs PP0168 and PP1449 (Nelson et al., 2002). PP1449 (1508 amino acids) is the better match, and is also identified when the KT2440 genome is interrogated with PA0041 (E value of approximately \(1 \times 10^{-26}\)). PP0168 is the largest gene (8682 amino acids) in the KT2440 genome and has the most unusual sequence (Weinel et al., 2002). The protein is rich in asparagine, valine and threonine, but contains no cysteine and is low in arginine. It contains two large threonine-rich repeats: the N-terminal repeat (nine units, each 100 amino acids long) and the C-terminal repeat (29 units, each 219 amino acids long). Although it has much in common with the \textit{P. aeruginosa} genome, the only virulence-related genes that the genome of the avirulent \textit{P. putida} KT2440 contains encode putative or confirmed adhesion proteins. Both PP0168 and PP1449 share homology with adhesin genes of enteropathogens and encode adhesins that are essential for seed colonization (Espinosa-Urgel et al., 2000; Nelson et al., 2002).

Thus, FHA-like proteins such as PA2462 may contribute to adhesion and colonization of disparate hosts, whether seed or human, and contain repeat sequences in the C terminus and downstream of the ORF with a probable role in mediating recombination. The biological function of such variations remains unknown, but since there may be a role for such proteins in the colonization of CF patients, further study is certainly merited.

\textbf{LES subtracted library}

Subtraction of the genome of strain PAO1 from that of isolate LES400 produced a library of clones, 106 of which were subjected to plasmid extraction and sequencing. Following the removal of duplicates and non-Testa-specific sequences, the subtraction added 78 sequences to the 14 obtained in a previous study (Parsons et al., 2002). The combined subtracted sequences are summarized in Table 3. Fifteen of the sequences shared similarity with mobile genetic elements, including bacteriophages D3 and D3112, both of which have been reported previously in \textit{P. aeruginosa} (Cavenagh & Miller, 1986; Wang et al., 2004). A further six subtracted sequences were associated with transposition, recombination or DNA modification. Amongst the known variable genes matching subtracted sequences were several related to siderophores. We have reported previously that the LES has a type III pyoverdine receptor (Parsons et al., 2002). The SSH identified sequences related to the synthesis of both type III pyoverdine and the corresponding receptor (Smith et al., 2005). In addition, the LES carries genes for serotype O6. Four of the subtracted sequences shared similarity with serotype-related genes. Whilst four subtracted sequences indicated the presence of the known \textit{P. aeruginosa} genomic islands PAGI-1 (Liang et al., 2001) and PAGI-2 (Larbìg et al., 2002), a further five sequences shared similarity with four linked genes from \textit{Clostridium thermocellum} ATCC 27405, suggesting that the LES may contain a low-G+C-content genomic island related to a gene cluster in this Gram-positive bacterium (Table 3). The subtraction also identified variations in the pilABC cluster. Such variations between strains of \textit{P. aeruginosa} have been reported previously (Choi et al., 2002; Kus et al., 2004).

As well as the gene related to mercury resistance associated with PS21, the subtraction identified a putative gene involved in silver resistance. Three subtracted sequences shared similarity with proteins that have roles in the production of polyketides (Table 3). Two of these sequences matched a regulatory protein (PltR) involved in the synthesis of pyoluteorin, which has antifungal activities (Bender et al., 1999). BLASTX searches for 38 % of the LES subtracted sequences revealed either no significant matches or matches with hypothetical proteins of no known function.

\textbf{Preliminary searches of the LES genome using SSH sequences}

Preliminary searches of the LES genome sequence using BLASTN identified all but two of the SSH sequences. Interestingly, there was no BLASTN match in the LES genome (99-98 % theoretical coverage) for either LESC4 or LESC5, which match adjacent genes from bacteriophage D3112, suggesting that this bacteriophage may be present in isolate LES400 but missing from the genome-sequenced isolate (LESB58). Notably, many of the SSH sequences were found on common contigs containing putative genomic islands. A single contig contained, in order, the subtracted sequences LESC6, LESH12, PS54, LESF10, LESG7, LESE5, PS3, LESF1, LESB11, LESE3, LES11, LESF12, PS1, PS12, LESD8, LESC7 and LESC4 (cluster 1). Further analysis indicated that sequence LESC6 is located at one edge of a putative genomic island incorporating all of the other SSH sequences with the...
Table 3. Summary of LES SSH

SSH sequences sharing less than 90% nucleotide sequence identity with the PAO1 genome were included in the table. ID, identity; NSH, no significant hits.

<table>
<thead>
<tr>
<th>SSH sequence</th>
<th>Length (bp)</th>
<th>GC content (%)</th>
<th>Best BLASTX match/comments, with accession nos</th>
<th>ID (%)</th>
<th>Length (aa)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacteriophage/plasmid related</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LES1</td>
<td>503</td>
<td>62.4</td>
<td>Phage-related integrase (Pseudomonas fluorescens Pf-5) (YP_260895)</td>
<td>50</td>
<td>122</td>
<td>2e⁻²⁶</td>
</tr>
<tr>
<td>LES6</td>
<td>&gt; 730</td>
<td>60-7</td>
<td>Gif-2 prophage protein (Salmonella typhimurium) (NP_460007)</td>
<td>55</td>
<td>245</td>
<td>3e⁻⁶⁸</td>
</tr>
<tr>
<td>LESA2</td>
<td>764</td>
<td>62-3</td>
<td>Orf26, serotype-converting bacteriophage D3 (P. aeruginosa) (NP_061522)</td>
<td>98</td>
<td>254</td>
<td>7e⁻¹³¹</td>
</tr>
<tr>
<td>LESC11</td>
<td>664</td>
<td>59-9</td>
<td>Orf104, serotype-converting bacteriophage D3 (P. aeruginosa) (NP_061582)</td>
<td>97</td>
<td>38</td>
<td>2e⁻³¹</td>
</tr>
<tr>
<td>LESA7</td>
<td>643</td>
<td>58-8</td>
<td>Phage-related tail-fibre protein (P. aeruginosa) (BABO3228); probable bacteriophage protein (PA0620)</td>
<td>100</td>
<td>214</td>
<td>e⁻¹³⁰</td>
</tr>
<tr>
<td>LESC4</td>
<td>324</td>
<td>64-8</td>
<td>Hypothetical protein D3112p54 (Bacteriophage D3112, P. aeruginosa) (NP_938261)</td>
<td>99</td>
<td>107</td>
<td>8e⁻⁵⁶</td>
</tr>
<tr>
<td>LESC5</td>
<td>528</td>
<td>64-2</td>
<td>Hypothetical protein D3112p53 (Bacteriophage D3112, P. aeruginosa) (NP_938260)</td>
<td>93</td>
<td>102</td>
<td>2e⁻⁴⁹</td>
</tr>
<tr>
<td>LESG3</td>
<td>401</td>
<td>56-4</td>
<td>Hypothetical protein F116p69, bacteriophage F116 (P. aeruginosa) (YP_164333)</td>
<td>85</td>
<td>34</td>
<td>9e⁻¹⁰</td>
</tr>
<tr>
<td>LESC7</td>
<td>430</td>
<td>62-8</td>
<td>Hypothetical protein mlr8022 (Mesorhizobium loti) (NP_108210); may be bacteriophage related</td>
<td>52</td>
<td>50</td>
<td>9e⁻⁷</td>
</tr>
<tr>
<td>LESF6</td>
<td>719</td>
<td>59-9</td>
<td>gp1 (Bacteriophage N15) (AAC19037)</td>
<td>42</td>
<td>163</td>
<td>4e⁻²⁸</td>
</tr>
<tr>
<td>LESG5</td>
<td>751</td>
<td>57-7</td>
<td>gp1 (Bacteriophage N15) (AAC19037); matches same sequence as LESF6, but is different from F6</td>
<td>41</td>
<td>163</td>
<td>3e⁻²⁶</td>
</tr>
<tr>
<td>LESH5</td>
<td>594</td>
<td>62-8</td>
<td>Hypothetical protein NCgl1675 (Corynebacterium glutamicum) (NP_600951); probably bacteriophage related</td>
<td>35</td>
<td>104</td>
<td>1e⁻¹¹</td>
</tr>
<tr>
<td><strong>Siderophore related</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS5</td>
<td>340</td>
<td>55-7</td>
<td>FpvAIII, type III pyoverdine receptor (P. aeruginosa) (AAN62912)</td>
<td>99</td>
<td>113</td>
<td>7e⁻⁶¹</td>
</tr>
<tr>
<td>LES8</td>
<td>527</td>
<td>61-3</td>
<td>psdL (3), type III pyoverdine-related non-ribosomal peptide synthetase (P. aeruginosa) (AY765261)</td>
<td>99</td>
<td>175</td>
<td>5e⁻⁹⁵</td>
</tr>
<tr>
<td>LESH4</td>
<td>524</td>
<td>62-0</td>
<td>psdD (3), type III pyoverdine-related non-ribosomal peptide synthetase (P. aeruginosa) (AY765261)</td>
<td>92</td>
<td>174</td>
<td>6e⁻⁸⁵</td>
</tr>
<tr>
<td>LESA1</td>
<td>461</td>
<td>55-1</td>
<td>Outer membrane ferric siderophore receptor (P. putida KT2440) (NP_743022)</td>
<td>57</td>
<td>157</td>
<td>1e⁻⁴⁶</td>
</tr>
<tr>
<td>LESA3</td>
<td>&gt; 584</td>
<td>58-6</td>
<td>Outer membrane ferric siderophore receptor (P. fluorescens Pf-5) (YP_257996)</td>
<td>71</td>
<td>196</td>
<td>3e⁻⁷⁶</td>
</tr>
<tr>
<td><strong>O serotype</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESB12</td>
<td>&gt; 596</td>
<td>55-4</td>
<td>Putative glycosyl transferase WbpT (P. aeruginosa O6 serotype) (AAF23993)</td>
<td>99</td>
<td>198</td>
<td>e⁻¹⁰⁷</td>
</tr>
<tr>
<td>PS43</td>
<td>&gt; 307</td>
<td>51-8</td>
<td>WbpO, B-band O-antigen biosynthesis (P. aeruginosa O6 serotype) (AAF23997)</td>
<td>100</td>
<td>102</td>
<td>8.2e⁻⁵⁰</td>
</tr>
<tr>
<td>PS48</td>
<td>862</td>
<td>48-6</td>
<td>Wzx, O-antigen flippase/translocase (P. aeruginosa O6 serotype) (AAF24000)</td>
<td>99</td>
<td>287</td>
<td>8.9e⁻¹⁰⁰</td>
</tr>
<tr>
<td>LESD3</td>
<td>592</td>
<td>50-8</td>
<td>WbpS (P. aeruginosa O6 serotype) (AAF24002)</td>
<td>55</td>
<td>200</td>
<td>4e⁻⁵⁷</td>
</tr>
<tr>
<td><strong>Genomic islands</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS40</td>
<td>198</td>
<td>57-5</td>
<td>Hypothetical protein (P. aeruginosa) (AAK01512); genomic island PAGI-1</td>
<td>100</td>
<td>66</td>
<td>2e⁻³¹</td>
</tr>
<tr>
<td>LES7</td>
<td>&gt; 605</td>
<td>62-0</td>
<td>Hypothetical protein (P. aeruginosa) (AAK01519); genomic island PAGI-1</td>
<td>100</td>
<td>148</td>
<td>1e⁻⁸¹</td>
</tr>
<tr>
<td>LESG6</td>
<td>237</td>
<td>62-9</td>
<td>Hypothetical protein ORFC81 (P. aeruginosa clone C) (AAN62175); genomic island PAGI-2</td>
<td>100</td>
<td>78</td>
<td>1.9e⁻³⁴</td>
</tr>
<tr>
<td>LESG7</td>
<td>1219</td>
<td>52-7</td>
<td>Hypothetical protein (Providencia rettgeri) (AAM08045); Putative excisionase ORFC56 (P. aeruginosa clone C) (AAN62150); ORFC56 is in genomic island PAGI-2</td>
<td>41</td>
<td>124</td>
<td>2e⁻²¹</td>
</tr>
</tbody>
</table>
### Table 3. cont.

<table>
<thead>
<tr>
<th>SSH sequence/Linked genes (putative island)</th>
<th>Length (bp)</th>
<th>GC content (%)</th>
<th>Best BLASTX match/comments, with accession nos</th>
<th>ID (%)</th>
<th>Length (aa)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LESF9</td>
<td>513</td>
<td>58.1</td>
<td>Type I restriction-modification system methyltransferase subunit (C. thermocellum) (ZP_00504950)</td>
<td>38</td>
<td>158</td>
<td>9e-19</td>
</tr>
<tr>
<td>LESG2</td>
<td>748</td>
<td>57.4</td>
<td>Hypothetical protein (C. thermocellum) (ZP_00504954)</td>
<td>44</td>
<td>233</td>
<td>1e-47</td>
</tr>
<tr>
<td>LESD11</td>
<td>1716</td>
<td>53.3</td>
<td>Possible ATPase (C. thermocellum) (ZP_00504957)</td>
<td>71</td>
<td>298</td>
<td>1e-17</td>
</tr>
<tr>
<td>LESB4</td>
<td>419</td>
<td>50.1</td>
<td>Serine/threonine protein kinase (C. thermocellum) (ZP_00504959); N terminus</td>
<td>41</td>
<td>39</td>
<td>0e-31</td>
</tr>
<tr>
<td>LESE4</td>
<td>623</td>
<td>49.4</td>
<td>Serine/threonine protein kinase (C. thermocellum) (ZP_00504959); same protein as LESB4</td>
<td>43</td>
<td>205</td>
<td>1e-38</td>
</tr>
<tr>
<td><strong>Drug or metal resistance/antimicrobial synthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS21</td>
<td>400</td>
<td>42.8</td>
<td>NSH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended PS21</td>
<td>1256</td>
<td>51.4</td>
<td>MerA, mercuric reductase (Streptomyces sp. CHR28) (AAF64138)</td>
<td>53</td>
<td>240</td>
<td>5e-67</td>
</tr>
<tr>
<td>LEB9</td>
<td>933</td>
<td>62.0</td>
<td>Putative silver-efflux pump (Dechloromonas aromatica RCB) (ZP_00150992); membrane-fusion protein (D. aromatica RCB) (ZP_00150992)</td>
<td>81</td>
<td>237</td>
<td>1e-102</td>
</tr>
<tr>
<td>LESE2</td>
<td>362</td>
<td>62.4</td>
<td>2.4-Diacetylphloroglucinol biosynthetic protein (polyside synthase), PhlD (P. fluorescens) (AAAX31636)</td>
<td>67</td>
<td>68</td>
<td>2e-19</td>
</tr>
<tr>
<td>LESG11</td>
<td>572</td>
<td>62.6</td>
<td>Transcriptional activator PltR (Pseudomonas sp. M18) (AAQ90169); involved in regulation of pyoluteorin synthesis</td>
<td>98</td>
<td>190</td>
<td>5e-45</td>
</tr>
<tr>
<td>LESH3</td>
<td>804</td>
<td>50.3</td>
<td>Transcriptional activator PltR (Pseudomonas sp. M18) (AAQ90169); same protein as LESG11 but different region; N-terminal sequence</td>
<td>97</td>
<td>95</td>
<td>2e-44</td>
</tr>
<tr>
<td><strong>Haemagglutinin related</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSA</td>
<td>&gt;1000</td>
<td>ND</td>
<td>Hypothetical protein (PA2459)</td>
<td>98</td>
<td>156</td>
<td>3e-87</td>
</tr>
<tr>
<td>Extended PSA</td>
<td>2751</td>
<td>50.9</td>
<td>Hypothetical protein (Bacillus thuringiensis) (YP_037116)</td>
<td>44</td>
<td>74</td>
<td>5e-10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C terminus of FbaB (N. meningitidis) (NP_283490)</td>
<td>50</td>
<td>57</td>
<td>1e-9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Large exoprotein COG3210 involved in haem utilization or adhesion (P. aeruginosa PA14) (ZP_00140193)</td>
<td>85</td>
<td>364</td>
<td>1e-16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Variation at the C terminus; hypothetical protein (PA2462)</td>
<td>95</td>
<td>319</td>
<td>1e-167</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Variation at the C terminus; hypothetical protein (PA2459)</td>
<td>98</td>
<td>177</td>
<td>1e-100</td>
</tr>
<tr>
<td>PS9</td>
<td>393</td>
<td>49.6</td>
<td>Haemagglutinin-like protein (PA2462)</td>
<td>39</td>
<td>140</td>
<td>3e-19</td>
</tr>
<tr>
<td><strong>Pili related</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEB1</td>
<td>&gt;652</td>
<td>58.0</td>
<td>Pilin biogenesis protein PilC (P. aeruginosa) (AAA25733)</td>
<td>88</td>
<td>127</td>
<td>3e-58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PilB (PA4526)</td>
<td>94</td>
<td>88</td>
<td>3e-43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PilC (PA4527)</td>
<td>89</td>
<td>94</td>
<td>7e-43</td>
</tr>
<tr>
<td>LESE11</td>
<td>508</td>
<td>53.9</td>
<td>Pilin glycosylation protein PilO (P. aeruginosa) (CAA58769)</td>
<td>99</td>
<td>169</td>
<td>1e-81</td>
</tr>
<tr>
<td><strong>DNA modification/recombination/transposases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESC12</td>
<td>516</td>
<td>50.0</td>
<td>Predicted transposase (Azoarcus sp. EbN1) (YP_158450)</td>
<td>47</td>
<td>57</td>
<td>5e-8</td>
</tr>
<tr>
<td>LESC7</td>
<td>458</td>
<td>53.9</td>
<td>Hypothetical protein MCA0729 (Methylcococcus capsulatus) (YP_113236); possible helicase</td>
<td>82</td>
<td>152</td>
<td>2e-72</td>
</tr>
<tr>
<td>LESE6</td>
<td>975</td>
<td>58.6</td>
<td>RadC, putative DNA-repair protein (Vibrio cholerae) (AAAL59704)</td>
<td>73</td>
<td>52</td>
<td>1e-16</td>
</tr>
<tr>
<td>LESH2</td>
<td>&gt;548</td>
<td>59.1</td>
<td>Periplasmic molybdate-binding protein/domain (Microbulbifer degradans) (ZP_00315765)</td>
<td>47</td>
<td>59</td>
<td>1e-10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hypothetical protein BB0398 (Bordetella bronchiseptica) (NP_886947); putative transposase</td>
<td>49</td>
<td>101</td>
<td>4e-20</td>
</tr>
</tbody>
</table>
### Table 3. cont.

<table>
<thead>
<tr>
<th>SSH sequence</th>
<th>Length (bp)</th>
<th>GC content (%)</th>
<th>Best BLASTX match/comments, with accession nos</th>
<th>ID (%)</th>
<th>Length (aa)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LESA8</td>
<td>637</td>
<td>58-4</td>
<td>RadC family protein (P. putida KT2440) (NP_744641)</td>
<td>83</td>
<td>109</td>
<td>2.5e-43</td>
</tr>
<tr>
<td>LESB5</td>
<td>357</td>
<td>58-0</td>
<td>Rhs family protein (PA14) (ZP_00136740)</td>
<td>91</td>
<td>118</td>
<td>7e-58</td>
</tr>
<tr>
<td>LESA8</td>
<td>637</td>
<td>58-4</td>
<td>Hypothetical protein (PA2457); recombination hot spot</td>
<td>85</td>
<td>118</td>
<td>4e-53</td>
</tr>
<tr>
<td><strong>Metabolism/enzymes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESA9</td>
<td>648</td>
<td>43-7</td>
<td>FOG: HEAT repeat (Anabaena variabilis) (ZP_00162995)</td>
<td>58</td>
<td>72</td>
<td>4e-16</td>
</tr>
<tr>
<td>LESA10</td>
<td>419</td>
<td>60-4</td>
<td>Asparagine synthase, glutamine hydrolysing (Psychrobacter sp.) (YP_263961)</td>
<td>65</td>
<td>126</td>
<td>9e-44</td>
</tr>
<tr>
<td>LESA9</td>
<td>648</td>
<td>43-7</td>
<td>Hypothetical protein (P. fluorescens Pf-5) (YP_260898)</td>
<td>65</td>
<td>44</td>
<td>3e-9</td>
</tr>
<tr>
<td>LESA8</td>
<td>637</td>
<td>58-4</td>
<td>Putative ATPase (Nitrobacter hamburgensis X14) (ZP_00627842)</td>
<td>20</td>
<td>202</td>
<td>8e-5</td>
</tr>
<tr>
<td><strong>Acyltransferases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS54</td>
<td>425</td>
<td>50-6</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended PS54</td>
<td>860</td>
<td>49-1</td>
<td>Acyltransferase for 30S ribosomal subunit protein S18 (Yersinia pestis) (NP_671044)</td>
<td>34</td>
<td>79</td>
<td>0-019</td>
</tr>
<tr>
<td>LESH11</td>
<td>556</td>
<td>59-9</td>
<td>Predicted acyltransferase (Ralstonia metallidurans) (ZP_00593471)</td>
<td>41</td>
<td>185</td>
<td>2e-23</td>
</tr>
<tr>
<td><strong>Transport</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESA11</td>
<td>638</td>
<td>61-9</td>
<td>Inner-membrane permease, component of ABC transporter system (Pseudomonas sp. M18) (AAY59068)</td>
<td>99</td>
<td>198</td>
<td>6e-94</td>
</tr>
<tr>
<td>LESH8</td>
<td>401</td>
<td>60-1</td>
<td>Conserved hypothetical protein (Psychrobacter sp.) (ZP_00608284)</td>
<td>28</td>
<td>133</td>
<td>2e-5</td>
</tr>
<tr>
<td><strong>Hypothetical proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESA5</td>
<td>522</td>
<td>59-4</td>
<td>Hypothetical protein (PA2790)</td>
<td>99</td>
<td>114</td>
<td>1e-61</td>
</tr>
<tr>
<td>LESH8</td>
<td>&gt;693</td>
<td>62-3</td>
<td>Hypothetical protein Paer03004107 (PA14) (ZP_00140093)</td>
<td>100</td>
<td>186</td>
<td>10e-100</td>
</tr>
<tr>
<td>PS32</td>
<td>138</td>
<td>58-7</td>
<td>Hypothetical protein PP2496 (P. putida KT2440) (NP_744644)</td>
<td>82</td>
<td>45</td>
<td>9e-18</td>
</tr>
<tr>
<td>LESH11</td>
<td>528</td>
<td>60-2</td>
<td>Hypothetical protein PP2496 (P. putida KT2440) (NP_744644); matches same protein as PS32 but in different region</td>
<td>93</td>
<td>175</td>
<td>1.0e-84</td>
</tr>
<tr>
<td>LESC10</td>
<td>494</td>
<td>58-7</td>
<td>Hypothetical protein PP3887 (P. putida KT2440) (NP_746017)</td>
<td>34</td>
<td>150</td>
<td>2e-15</td>
</tr>
<tr>
<td>LESA8</td>
<td>359</td>
<td>43-5</td>
<td>Hypothetical protein Pf002002711 (P. fluorescens P60-1) (ZP_00264577)</td>
<td>40</td>
<td>109</td>
<td>2e-15</td>
</tr>
<tr>
<td>LESA9</td>
<td>463</td>
<td>61-6</td>
<td>Hypothetical protein Avin02002756 (Azotobacter vinelandii) (ZP_00089890)</td>
<td>33</td>
<td>145</td>
<td>3e-9</td>
</tr>
<tr>
<td>LESA8</td>
<td>692</td>
<td>60-6</td>
<td>Hypothetical protein (PA5517)</td>
<td>100</td>
<td>73</td>
<td>2e-35</td>
</tr>
<tr>
<td>LESA9</td>
<td>463</td>
<td>61-6</td>
<td>Hypothetical protein S4804 (Shigella flexneri) (NP_836048)</td>
<td>29</td>
<td>84</td>
<td>5e-4</td>
</tr>
<tr>
<td>LESH11</td>
<td>493</td>
<td>56-6</td>
<td>Hypothetical protein Daro03003346 (Dechloromonas aromatica RCB) (ZP_00149999)</td>
<td>89</td>
<td>91</td>
<td>1e-43</td>
</tr>
<tr>
<td>LESA8</td>
<td>692</td>
<td>60-6</td>
<td>Hypothetical protein Det0256 (Dehalococcoides ethenogenes 195) (YP_181004)</td>
<td>44</td>
<td>69</td>
<td>1e-10</td>
</tr>
<tr>
<td><strong>NSH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS1</td>
<td>338</td>
<td>42-9</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS2</td>
<td>282</td>
<td>45-7</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS3</td>
<td>250</td>
<td>51-6</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS12</td>
<td>288</td>
<td>52-1</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
exception of LES4, which lies beyond the putative island. The diagnostic probe (PS21) appears to lie in a cargo region of the previously identified island PAGI-2 (Klockgether et al., 2004; Larbig et al., 2002) that also contains LESB9, the putative silver-efflux-related sequence (cluster 2). Twelve of the SSH sequences (in order: LESH1, LESE4, LESB4, LESD11, LESG2, LESF9, LESD7, PS32, LES5, LESA8, LESE6 and LESB2) were clustered on a different contig within a second large putative island containing a number of genes related to *C. thermocellum* (cluster 3). There were three separate contigs containing multiple SSH sequences, some of which were bacteriophage related. One contained, in order, LES1, LESA9, LESC1, LESB6, LES6 and LESG5 (cluster 4); a second contained, in order, LESA11, LESH3, LESG11, LESE2, LESG3, LESA2, LESC11, LESC10, LESH5, LESA12 and LESF3 (cluster 5); and a third contained, in order, LESC3, LESE12, LESE7, LESF8, LESF6 and LESG9 (cluster 6). There were also several other contigs containing more than one SSH sequence. Not surprisingly, one of these linked the type III pyoverdine-synthesis-related sequences LESH4 and LES8.

**Distribution of subtracted sequences**

The distribution of 25 LES sequences amongst a panel of strains was assessed by PCR amplification assays using oligonucleotide primers designed from subtracted sequences (Table 4). The sequences selected included multiple representatives for each of the six major clusters identified by preliminary analysis of the genome sequence, sequence LESD9 (which shared a contig with LESB5, PS15, LESG10 and PS2), sequence LESB3 (which shared a contig with LESH1), and the type III pyoverdine-synthesis-related sequences LESH4 and LES8, and LESA3 (which shared a contig with LESA1).

The distribution of sequences from cluster 1 confirmed that the putative genomic island was only present in LES isolates, but was completely missing from three LES isolates. LES4 is beyond the putative island and has a radically different distribution from the other SSH sequences on this contig (Table 4). All of the SSH sequences, including PS54, PS3, PS1 and PS12 from our previous study (Parsons et al., 2002), were absent from LES isolates 109 (isolated in Liverpool), 8995 (isolated in Sheffield, UK) and 10333 (isolated in Nottingham, UK). The putative island is flanked by equivalents to strain PAO1 ORFs PA0831 and PA0832.

Cluster 2, including PS21 and the silver-efflux-pump-related SSH sequence LESB9, was present in all LES isolates. All four PS21 false-positive strains were also PCR positive for LESB9, confirming the linkage of these two sequences. All SSH sequences tested from cluster 3 were PCR positive for all LES isolates but PCR negative for all other genetically characterized isolates (Table 4). Further PCR assays for sequences LESE4 and LESF9 were used to screen the uncharacterized non-LES CF isolates from Liverpool and from Clatterbridge Hospital. Three isolates (CH31, CH47 and 49278) were PCR positive for LESE4, and two of these

<table>
<thead>
<tr>
<th>SSH sequence</th>
<th>Length (bp)</th>
<th>GC content (%)</th>
<th>ID</th>
<th>Length (aa)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS15</td>
<td>882</td>
<td>59.1</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LES2</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LES4</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESD10</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESB11</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESG3</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESF2</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESF3</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESG10</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESH1</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESA9</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC1</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESA3</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESA10</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC11</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC10</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESB12</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC3</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC4</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC5</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC6</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC7</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC8</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC9</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC11</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC12</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC13</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC14</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC15</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
<tr>
<td>LESC16</td>
<td>713</td>
<td>56.9</td>
<td>NSH</td>
<td>47.5</td>
<td>69.9</td>
</tr>
</tbody>
</table>
isolates, CH47 and 49278, were also PCR positive for LESF9. All other PS21-negative isolates were PCR negative for each of these sequences. PFGE typing of isolates CH31, CH47 and 49278 confirmed that these did not match the LES pulsotype (data not shown).

The PCR-based distribution profiles for the three cluster 4 sequences were similar but not identical, and indicated some instability. Two LES isolates were PCR negative for all three sequences. Two of the three cluster 5 sequences (LESA11 and LESG11) shared identical distributions and were stable amongst LES and Midlands1 isolates. The third sequence tested from cluster 5 (LESH5) was greater than 50 kb away from the other two. Only LES isolates were PCR positive for this sequence. The three cluster 6 SSH sequences shared nearly identical distributions and were detected in all LES isolates. The other sequences screened were present in both LES and Midlands1 isolates. The Manchester epidemic strain isolates were PCR positive for only one of the sequences (LESA3, matching a putative siderophore receptor) tested (Table 4).

PCR-based identification of the LES

We have reported previously that the LES strain carries approximately 95% of the ORFs present in strain PAO1 and that estimations based on PFGE indicate that the LES genome is no bigger than that of PAO1 (Salunkhe et al., 2005). In our previous study (Parsons et al., 2002), we identified 14 LES sequences subtracted from strain PAO1, including five (PS1, PS3, PS12, PS21 and PS54) that were LES specific. All of these, except PS21, were absent from one of the LES isolates screened (isolate 109). After considerably expanding the SSH database and conducting preliminary searches of the LESB58 genome sequence, it is clear that sequences PS1, PS3 and PS54, along with 13 other SSH sequences, are part of an LES-specific island that is associated with some instability. Thus, although we have never found this island in strains other than LES, any test for the LES based on this island may be prone to false negatives. In contrast, PS21, part of a merA gene located in an apparent cargo region of PAGI-2, is stable in the LES, but occurs rarely in other P. aeruginosa strains, making tests based on PS21 theoretically susceptible to rare false positives. Sequence LESF9 forms part of a putative genomic island present in all LES isolates but is, like PS21, occasionally also found in non-LES isolates. In the absence of a definitive LES-specific PCR assay that is positive for all LES isolates tested, we suggest a strategy using a combination of PCR assays for the detection of PS21 and LESF9. Since PS21 has proved to be an effective tool, especially in a setting where LES is widespread (Panagea et al., 2003; our unpublished data), we would propose the continuation of this test. For clinicians and patients alike, it is important that all LES isolates give a positive result. In order to achieve this, in the light of the instability of the LES genome, we may have to accept the possibility of rare false positives. However, if a combination of PCR assays for PS21 and LESF9 is used, then at least with our collection of isolates,
only LES isolates are PCR positive for both of these sequences. We would suggest that any isolates that are PCR positive for just one of these sequences should be further tested using a different typing method, such as PFGE or RAPD.

The LES mobiome

The SSH data indicated that bacteriophage DNA is likely to make a significant contribution to the accessory genome of the LES. D3 is a lysogenic bacteriophage with a genome of 56-4 kb (Kropinski, 2000) (accession no. AF165214) implicated previously in serotype conversion (Newton et al., 2001) and transduction (Cavenagh & Miller, 1986) in \textit{P. aeruginosa}. SSH sequences matching widely separated regions of the D3 genome indicate the possible presence of this bacteriophage genome within the LES. D3112 is a transposable Mu-like bacteriophage with a genome of 37-6 kb (Wang et al., 2004). We identified LES sequences matching neighbouring genes within the D3112 genome, but we have no further indication concerning the extent of the D3112 DNA within the genome of the LES. Interestingly, these two sequences appear to be missing from the LES isolate chosen for genome sequencing, suggesting either that the bacteriophage has been acquired by isolate LES400 or that it has been lost by isolate LESB58. In addition, the bacteriophage-related cluster 4 was absent from some LES isolates. SSH also indicated that the LES contains a putative tail-fibre protein from another bacteriophage reported elsewhere in \textit{P. aeruginosa} (Nakayama et al., 2000). The presence of several other subtracted sequences with matches sharing lower identities with bacteriophage-related proteins, many clustered on to one of three contigs, suggests that there may be additional whole (or fragments of) bacteriophage genomes within the LES. The details will become clear when the ongoing genome sequence project is further advanced.

ACKNOWLEDGEMENTS

We would like to thank Ty Pitt and Fiona Scott for providing strains. We are also grateful to Julian Parkhill and colleagues at the Pathogen Sequencing Unit of the Wellcome Trust Sanger Unit, and Bob Hancock, Fiona Brinkman and colleagues from the \textit{Pseudomonas} Genome Project, for their excellent work in the LES and PAO1 genome sequence projects. This work was funded by the UK Cystic Fibrosis Trust.

REFERENCES


