BACTERIAL PATHOGENICITY

Cloning and analysis of genomic differences unique to *Burkholderia pseudomallei* by comparison with *B. thailandensis*

NAT F. BROWN and IFOR R. BEACHAM

School of Health Science, Griffith University, Gold Coast Campus, Southport, Queensland 4217, Australia

Meliodosis is an infectious disease caused by *Burkholderia pseudomallei*. Genomic subtractive hybridisation was performed with the closely related avirulent species *B. thailandensis* to identify virulence genes of *B. pseudomallei*. The subtractive hybridisation products were highly specific for *B. pseudomallei*. Sequence analysis revealed a number of putative virulence factors, as well as apparently novel sequences of unknown function. The subtracted library contained DNA regions of relatively low G+C mol% content, which were distributed throughout the *B. pseudomallei* genome. The distribution of subtracted sequences amongst a collection of 22 *B. pseudomallei* isolates was found to be variable, with the exception of three strains which almost universally lacked the subtracted sequences. These three strains also differed in that they were highly haemolytic, indicating a possible separate virotype.

Introduction

Meliodosis is a disease of man and animals caused by infection with the gram-negative, saprophytic soil bacterium, *Burkholderia pseudomallei* [1]. Both the disease and the bacterium are found primarily in tropical regions of south-east Asia and northern Australia [2]. In these regions, meliodosis is a major cause of morbidity and mortality, and yet it remains widely under-recognised [3]. Also adding to the significance of meliodosis is the high rate of relapse (up to 23%), despite long-term maintenance therapy [3]. The disease has various presentations, including acute, chronic and recrudescent conditions affecting many organs [3]. This suggests that *B. pseudomallei* possesses a considerable armoury of virulence factors.

Corresponding to the under-recognition of meliodosis is the general lack of knowledge regarding *B. pseudomallei* virulence. Some studies have been concerned with the role of exoproducts, such as protease, lipase, phospholipase C and haemolysin, as virulence factors [4, 5]. Of these, only protease has been shown to be involved in pathogenesis, with a protease-deficient mutant having a moderately increased LD50 in an animal model [4]. The type II O-

Received 4 Jan. 2000; revised version accepted 30 March 2000.
Corresponding author: Dr I.R. Beacham (e-mail: I.beacham @mailbox.gu.edu.au).
This study aimed to identify genetic determinants of virulence in *B. pseudomallei*. The approach taken was to perform a genome comparison by subtractive hybridisation between a virulent *B. pseudomallei* isolate and an isolate of *B. thailandensis*. The products of the subtractive hybridisation experiment were used to construct a subtracted library of *B. pseudomallei*, which was then subjected to sequence analysis. The distribution of random cloned subtracted DNA fragments among a collection of *B. pseudomallei* isolates was also investigated.

**Materials and methods**

**Bacterial strains and plasmids**

*Escherichia coli* TGI [21] was used as the host strain for molecular cloning. *Burkholderia* strains used in this study are described in Table 1 [22–24]. The sequences of the variable 320-bp region of their 16S rDNA were obtained according to the method described by Brett *et al.* [18]. All bacteria were cultured on Luria-Bertani (LB) agar at 37°C, except *E. coli* transformants, which were grown on MacConkey agar containing ampicillin 100 μg/ml. PCR products were cloned into the pGEM-T Easy plasmid vector (Promega Corporation). Haemolysis was detected on agar plates containing sheep blood 5% and incubated at 37°C for 48 h.

**Molecular biology procedures**

Genomic DNA was prepared as described previously [25], except that a loopful of growth from a fresh plate was suspended in 600 μl of lysis solution (sodium dodecyl sulphate 0.5%, protease K 100 μg/ml). Miniprep of plasmid DNA were prepared by alkaline lysis and polyethylene glycol precipitation [26].

Ligations were performed in a 10-μl final volume at 4°C overnight. Electrocompetent *E. coli* was transformed in an EC100 electroprotator (EC Apparatus Corporation) according to the manufacturer's instructions.

Southern blotting was performed according to standard techniques [26], with Hybond N+ nylon membrane (Amersham Pharmacia Biotec).

**Table 1. Characteristics of Burkholderia strains**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Source</th>
<th>Variable 16S rDNA sequence*</th>
<th>Haemolysin Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. pseudomallei</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RH1</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak N. George, Royal Brisbane Hospital</td>
</tr>
<tr>
<td>THFP30</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Strong [22]</td>
</tr>
<tr>
<td>THFP385</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Strong [22]</td>
</tr>
<tr>
<td>THFP401</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak [22]</td>
</tr>
<tr>
<td>THFP478</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Strong [22]</td>
</tr>
<tr>
<td>THFP375</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak [22]</td>
</tr>
<tr>
<td>THFP379</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak [22]</td>
</tr>
<tr>
<td>THFP419</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak [22]</td>
</tr>
<tr>
<td>76</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak [22]</td>
</tr>
<tr>
<td>GF1</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak [24]</td>
</tr>
<tr>
<td>GF2</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak [24]</td>
</tr>
<tr>
<td>GF3</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak [24]</td>
</tr>
<tr>
<td>01</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie, Menzies School of Health Research</td>
</tr>
<tr>
<td>02</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td>03</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td>04</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td>05</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td>06</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td>07</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td>08</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td>09</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td>10</td>
<td>Clinical</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>Weak Dr. B. Currie</td>
</tr>
<tr>
<td><em>B. thailandensis</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E264</td>
<td>Environmental</td>
<td>TCAATGCGTAATACCCGGAATG00</td>
<td>– Dr. T. Pitt, CPHL London</td>
</tr>
<tr>
<td><em>B. cepacia</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCC17765</td>
<td>–</td>
<td>TCTTGGCTTAATACCCGGAATG00</td>
<td>– N. George</td>
</tr>
</tbody>
</table>

*The sequence shown is the variable region of the 320-bp PCR product from the 16S rDNA gene. Underlined bases indicate bases that differ from the recognised *B. pseudomallei* 16S rDNA gene sequence [1].
Subtractive hybridisation

Suppression subtractive hybridisation [27] was performed with the CLONTECH PCR-Select™ Bacterial Genome Subtraction Kit™ (CLONTECH Laboratories) according to the manufacturer’s instructions with minor modifications. Briefly, the procedure involved the purification of genomic DNA [23] from tester and driver strains (see below) and this DNA was cut to completion with AluI restriction endonuclease. The restricted tester DNA was then split into two pools, each being ligated to separate adapters. Each tester pool was then hybridised separately with a molar excess of driver DNA at 63°C for 1.5 h. This was followed by a second round of hybridisation at 63°C for 18 h in which both hybridisation pools were combined together with additional driver DNA. The second hybridisation mixture was then diluted and amplified with primers complementary to the adapter sequences.

Nucleotide sequence analysis

Nucleotide sequence data were obtained by using T7 and SP6 primers with Big Dye chemistry on an Applied Biosystems model 377 automated DNA sequencer. The sequence data were edited with SeqEd™ v1.0.3 (Applied Biosystems) and analysed with MacVector v6.0.1 (Oxford Molecular Group). BLASTx was used to search for sequence similarity [28].

Slot-blotting

Genomic DNA, 300 ng in a volume of 200 μl, from each bacterial strain was blotted onto GeneScreen Plus™ nylon membranes (Du Pont) in a PR 648 slot-blot apparatus (Hoefer Scientific Instruments) according to the manufacturer’s instructions. Probes were labelled with α32P-dCTP (Amersham Pharmacia Biotech) by nick translation and membranes were hybridised overnight at 65°C. The membranes were then washed at 65°C with SSC (20 × 0.3 mM NaCl, 0.3% sodium citrate) at concentrations appropriate for the required stringency. Images were obtained from the blots with storage phosphor screens (Kodak) and a Personal Molecular Imager FX (BioRad Laboratories) with associated Quantity One v4.0.3 software.

Pulsed-field gel electrophoresis (PFGE)

Genomic DNA from B. pseudomallei strain 08 was prepared from growth on LB agar plates according to the method described by Nair et al. [29]. This DNA was digested with SpeI (New England Biolabs) for 18 h and electrophoresed under the conditions described by Winstanley et al. [17] in a Gene Navigator CHEF apparatus (Amersham Pharmacia Biotech).

Results and discussion

Subtractive hybridisation

The two species, B. pseudomallei and B. thailandensis, are very closely related phylogenetically, yet the former displays very high virulence compared with the latter [18,20]. There appear to be few other phenotypic differences between these species. The approach chosen was to perform genomic subtractive hybridisation with B. pseudomallei strain 08 as the tester strain (strain of interest) and B. thailandensis strain E264 as the driver strain (reference strain). B. pseudomallei 08 was selected as the tester because it was isolated from a fatal human case of melioidosis. B. thailandensis E264 was chosen as the driver because it has been shown previously to be avirulent in an animal model of melioidosis [18]. Cloned products of subtractive hybridisation would largely be made up of genomic sequences unique to B. pseudomallei. It was predicted that a substantial proportion of these clones should be representative of virulence genes, as virulence is a major phenotypic difference between these species.

The PCR products from subtractive hybridisation were used to construct a subtracted genomic library in pGEM-T Easy (Promega Corporation). From the resulting library, 43 distinct subtracted DNA clones were tested for their presence in the B. pseudomallei 08 and B. thailandensis E264 genomes by Southern (results not shown) or slot-blot (see Fig. 1) hybridisation. Of these 43 clones, only six were shown to be present in both strains 08 and E264. Only two of these (pSC-A47 and pSC-B8) were considered to be simple background clones as four appeared to represent transposable genetic elements (pSC-A31, pSC-A42, pSC-B48 and pSC-B54) [30]. The latter may have been selected as background because of copy number effects or, possibly, sequence divergence between the strains resulting in poor hybridisation [30]. The relatively low number of background clones indicated that the subtractive hybridisation procedure was effective at enriching for sequences unique to B. pseudomallei 08.

Strain distribution of random clones

The distribution of selected clones among a collection of 22 strains of B. pseudomallei, and one strain each of B. thailandensis and B. cepacia, was determined by slot-blot hybridisation. The identity of these strains was first confirmed by PCR amplification and sequencing of a variable region of the 16S rRNA genes [18] and the results are shown in Table I. All environmental isolates appeared to be B. pseudomallei by this technique, and hence are potentially virulent [18]. The two strains showing sequence differences (08 and THP478) were both isolated from clinical sources and hence can be considered to be virulent.

A total of 22 subtracted DNA clones was selected for use as probes for slot-blot hybridisation. These were
Fig. 1. Slot-blot hybridisations of genomic DNA of *Burkholderia* strains with probes consisting of the subtracted DNA fragments. The strains from which the genomic DNA was derived were arranged in the blots as follows: left to right, top row, RBH, THP380, THP385, THP401, THP478, THP375, THP379, THP419, 76, GF1, GF2, GF3, and bottom row, 01, 02, 03, 04, 05, 06, 07, 09, 10, E264, ATCC 17765 and 08. Cloned subtracted fragments used to probe the blots were: a, pSC-B45; b, pSC-A15; c, pSC-A38; d, pSC-A56; e, pSC-B5; f, pSC-B18; g, pSC-B21; h, pSC-B39; i, pSC-B42; j, pSC-B63; k, pSC-B79; l, pSC-B82. Fig. 1a also represents the slot-blot hybridisations probed with pSC-A33, pSC-A35, pSC-A39, pSC-A46, pSC-A47 pSC-A71, pSC-B8, pSC-B12, pSC-B52 and pSC-B73.

selected at random to determine if there were any patterns of distribution of subtracted sequences between strains. The results of the slot-blot experiments are shown in Fig. 1. From these experiments, two broad categories of distribution were evident among *B. pseudomallei* isolates. The first included sequences that were present in all *B. pseudomallei* strains except THP380, THP385 and THP478, and the second included sequences that were present in a random sample of strains. Eleven sequences were distributed in the former manner (Fig. 1a). Invariably, strains THP380, THP385 and THP478 also lacked 11 sequences belonging to the second category (Fig. 1b–h, j–l, with the exception of pSC-B42, Fig. 1l).
was no discernible pattern of strain distribution of clones in the second category, and the number of strains that possessed these sequences varied from 1 to 18. For example, pSC-B39 was present only in strain 08, whereas pSC-B18 was present in all strains except THP375, THP380, THP385 and THP478.

A striking feature of the strain distribution of the selected subtracted clones was their absence from strains THP380, THP385 and THP478. Furthermore, THP380 and THP385 lacked the three proposed transposable elements represented in this library [30]. These observations are strongly indicative of these strains being significantly different from the other strains tested. However, being clinical isolates, they are virulent strains and hence may differ, at least in part, by possessing different virulence determinants. This is consistent with the highly haemolytic phenotype of these strains, in contrast to the other strains tested (Table 1).

Nucleotide sequence analysis of subtracted library
From the subtracted library, 150 clones were selected for sequencing with vector primers. The resulting nucleotide sequences were used to search for similarity to known sequences with the BLASTx program [28]. It became apparent that there was a significant level of sibling redundancy among the 150 clones chosen for sequencing and that, of those 150 clones, there were 84 individual clones.

In total, 38,707 bp of nucleotide sequence data were generated from the subtracted genomic library and none of the sequences corresponded to B. pseudomallei sequences in databases. The length of strain 08 cloned subtracted genomic DNA inserts ranged from 78 bp for pSC-B56 to 1023 bp for pSC-B42. All these cloned fragments were Abs1 restriction fragments as expected and, hence, were random in distribution and did not represent entire genes. An open reading frame (ORF) analysis of each clone was performed. Any internal ORF of >300 bp, or a portion of an ORF >150 bp but truncated at the 5' or 3' end, was used as a guide to the coding potential of the clones. The vast majority of clones showed coding potential according to these criteria.

The G + C mol% content of these clones was considerably lower than the recognised G + C content of 67.9 mol% for B. pseudomallei [1]. The G + C values ranged from 41.9 mol% (pSC-A32) to 62.6 mol% (pSC-A71). This disparity in G + C mol% values was not due to contaminating DNA coming through the subtractive hybridisation procedure, as 43 clones were shown to be present in B. pseudomallei 08 DNA by hybridisation experiments. The difference in G + C mol% of the subtracted clones from the rest of the genome is particularly interesting in the light of such differences in other species being indicative of loci, including pathogenicity islands (PAIs), acquired via horizontal gene transfer [31, 32]. The two known background clones, pSC-A47 and pSC-B8 (as well as clones representing transposable elements), also had low G + C mol% contents. This may be explained by A + T rich DNA being less likely to hybridise at the temperature used and, therefore, being more likely to come through the procedure as background.

Interestingly, a significant proportion of the 84 clones displayed little similarity to known genes, although there were exceptions to this. References made to clone designations in the following text refer to the corresponding B. pseudomallei 08 cloned subtracted genomic DNA fragments. Thirty-six clones had an expected value >1 for their highest BLAST hit, which indicates a low statistical significance of these hits, and a further five clones had no BLAST hits. This shows that this subtracted library is likely to represent many novel, as yet uncharacterised, functions. Furthermore, it is also likely that most clones with hits to eukaryotic proteins of known function may share little or no functional relationship; hence, such clones also probably encode uncharacterised functions. Finally, BLAST similarities with reasonable statistical significance included a number of hypothetical proteins from species as diverse as Homo sapiens, Caenorhabditis elegans, Vibrio cholerae, Methanothermus januschii, Mycobacterium tuberculosis, Bacillus subtilis and Escherichia coli and these also represent unknown functions.

The 20 remaining clones are those which were similar to prokaryotic genes encoding proteins of known function (Table 2). Four (pSC-A31, pSC-A42, pSC-B48 and pSC-B54) were similar to genes coded within insertion sequences, were present in both tester and driver strains, and have been proposed as representing transposable elements [30]. pSC-A71 is similar to Ta1, a protein involved in the synthesis of TA, a polyketide antibiotic produced by Myxococcus xanthus [33]. The slot-blot probed with pSC-A71 showed that this sequence was widely distributed among B. pseudomallei strains (Fig. 1a). Two other clones, pSC-A10 and pSC-B73, were similar to BxiE of Streptomyces lividans (accession no. 943654) and a glycophosphoryl diester phosphodiesterase of Bacillus subtilis (accession no. Z26522), respectively. The pSC-B73 sequence was apparently present in the majority of B. pseudomallei strains (Fig. 1a).

pSC-B42 was similar to a bacteriophage P2 protein [34]. It was widely distributed among B. pseudomallei strains (Fig. 1I) and may represent a lysogenic bacteriophage. pSC-B82 was similar to an integrase encoded within a shufflon [35]. This shufflon functions to invert a 490-bp DNA segment via site-specific recombination resulting in two possible pilus-tip adhesin proteins [35]. It is also noteworthy that this shufflon is located in the major PAI of Salmonella typhi, although this shufflon
Table 2. Sequence analysis of *B. pseudomallei* 08 subtracted DNA fragments

<table>
<thead>
<tr>
<th>Clone</th>
<th>Length (bp)</th>
<th>G+C mol%</th>
<th>ORFs*</th>
<th>BLAST hits</th>
<th>expect value</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSC-A10</td>
<td>235</td>
<td>54.4</td>
<td>1</td>
<td>Hypothetical (Proteus vulgaris) 7e-07</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-A17</td>
<td>730</td>
<td>52.6</td>
<td>2</td>
<td>Permease (Thermus sp.) 2e-25</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-A31</td>
<td>243</td>
<td>56.4</td>
<td>1</td>
<td>ORF1 of IS400 (R. capsulatus) 2e-15</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-A33</td>
<td>813</td>
<td>49.9</td>
<td>2</td>
<td>Hypothetical (Enterobacter agglomerans) 1e-14</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-A35</td>
<td>504</td>
<td>43.8</td>
<td>2</td>
<td>AIDA-I precursor (E. coli) 0.008</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-A39</td>
<td>732</td>
<td>52.9</td>
<td>4</td>
<td>CprX (Actinobacillus pleuropneumoniae) 3e-41</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-A42</td>
<td>435</td>
<td>52.9</td>
<td>3</td>
<td>ORF4 of IS400 (R. capsulatus) 4e-05</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-A46</td>
<td>583</td>
<td>53.3</td>
<td>4</td>
<td>BM3 (Bord. bronchisepticum) 1e-03</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-A71</td>
<td>171</td>
<td>62.6</td>
<td>2</td>
<td>Tid (Mycosarcus sambus) 8e-19</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B6</td>
<td>219</td>
<td>56.2</td>
<td>1</td>
<td>Rotamase C (E. coli) 2e-04</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B21</td>
<td>504</td>
<td>51.6</td>
<td>1</td>
<td>Lipoprotein (Bacillus subtilis) 7e-04</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B42</td>
<td>1023</td>
<td>49.7</td>
<td>3</td>
<td>Old (bacteriophage P2) 3e-06</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B48</td>
<td>572</td>
<td>49.8</td>
<td>1</td>
<td>Transposase (Bord. parapertussis) 2e-09</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B52</td>
<td>757</td>
<td>48</td>
<td>5</td>
<td>Isolose DNA synthetase (Campylobacter jejuni) 0.46</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B54</td>
<td>328</td>
<td>54.9</td>
<td>2</td>
<td>Putative transposase (Pseudomonas alcaligenes) 2e-13</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B67</td>
<td>534</td>
<td>47</td>
<td>2</td>
<td>rhsD (E. coli) 0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B73</td>
<td>560</td>
<td>50.9</td>
<td>1</td>
<td>Core protein (E. coli) 0.15</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B78</td>
<td>248</td>
<td>48</td>
<td>0</td>
<td>Hypothetical (Bacillus subtilis) 0.090</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B79</td>
<td>741</td>
<td>43.6</td>
<td>1</td>
<td>SlaA (Serratia marcescens) 0.046</td>
<td>0.001</td>
</tr>
<tr>
<td>pSC-B82</td>
<td>155</td>
<td>56.1</td>
<td>1</td>
<td>Rcl (Salmonella typhi) 2e-14</td>
<td>0.001</td>
</tr>
</tbody>
</table>

*ORFs were determined by the start stop method. Any ORF > 300 bp or any 5’ or 3’ truncated ORF > 150 bp was considered to be a possible ORF.

system is considered to alter mating specificity in conjugation [35]. The distribution of pSC-B82 is interesting in that it hybridised well to two strains and poorly to the majority of the remaining strains (Fig. 1).

Seven clones were similar to genes encoding cell wall-associated and extracellular proteins, a group of proteins which includes most virulence factors [36]. Proteins belonging to this group, which were similar to subtracted DNA sequences, include sugar and capsular polysaccharide transporters and adhesins. pSC-A17 was similar to an ORF (in an operon considered to be involved in the metabolism of sugars) which is thought to encode a permease, although this requires confirmation [37]. pSC-A39 was similar to the capsule export protein CprX of *Actinobacillus pleuropneumoniae* [38], and was distributed widely among *B. pseudomallei* strains (Fig. 1a). pSC-A33 demonstrated similarity to diffusely adherent *E. coli* adhesin, AIDA-I, and was present in the genome of most strains (Fig. 1a). AIDA-I has been shown to mediate adherence to cells in culture [39, 40], and contains a predicted β-barrel structure indicative of an autotransporter protein [41]. pSC-B21 was similar to the HMW1 adhesin of non-typable *Haemophilus influenzae* (NTHi) [42, 43]. The HMW proteins of NTHi are considered to be major virulence determinants at many stages of NTHi disease [44, 45]. The pSC-B21 sequence was present in only two strains as determined by slot-blot analysis (see Fig. 1g). pSC-B67 was similar to rhsD of *E. coli*, a protein putatively localised to the outer membrane [46]. pSC-B79 was present in most *B. pseudomallei* strains tested (Fig. 1k), and is similar to the surface layer protein (SlaA) of *Serratia marcescens*, which is exported via a type I secretion apparatus [47]. pSC-B6 showed significant similarity to rotamase C of *E. coli*. Rotamase is a protein with peptidyl-prolyl cis-trans-isomerase activity that is located in the periplasm where it may act in protein folding [48].
There were also clones which appeared to represent genes similar to those encoding proteins involved in transcriptional regulation. Two were similar to two-component regulatory systems, pSC-A35 being similar to DmsR of Rhodobacter sphaeroides [49], and pSC-B52 showing similarity to CpxR of E. coli [50]. Each of these two clones had an identical distribution among the B. pseudomallei strains tested (Fig. 1a). The CpxA/R system of E. coli is known to be a major regulatory pathway in the response to membrane stress and protein misfolding in the periplasm [51–53]. Furthermore, this system has been found to regulate virF, the master regulator of virulence genes in Shigella sonnei [53, 54]. pSC-B78 demonstrated similarity to BlaA of Streptomyces coelicolor, which has been shown to regulate β-lactamase expression [55]. pSC-A46 was widely distributed among B. pseudomallei strains (Fig. 1a) and displayed similarity to the BbH3 and BpH3 proteins of Bordetella bronchiseptica and Bord. pertussis, respectively. In Bord. pertussis, BpH3 has been shown to bind to promoter regions and to affect transcription [56].

Mapping subtractive hybridisation products

The differing G+C mol% content of the subtractive hybridisation products, relative to the remainder of the B. pseudomallei genome, suggested that they represented one or more horizontally acquired polygenic loci. Other observations supporting this contention are the fact that many products were absent from a closely related avirulent organism and the presence of sequences similar to transposases [30] and integrases [35]. Specifically, pathogenicity islands have been well documented to encode such activities which are considered as defining characteristics [31, 32]. Other elements such as bacteriophages or integron-like structures would also be consistent with the above findings.

To determine if the subtractive hybridisation products were representative of any large loci consisting of contiguous genes, macro-restriction fragments of B. pseudomallei DNA were probed with labelled PCR products derived from the subtractive hybridisation. The results indicated that most of the fragments hybridised with the probe (Fig. 2). This conclusion was confirmed in further Southern hybridisation experiments in which four randomly selected cloned subtracted fragments were each used as probes; each hybridised to a different restriction fragment (results not shown). This suggested that the subtractive hybridisation products did not represent a small number of distinct polygenic loci.

Conclusions

To summarise, of the 38 707 bp of nucleotide sequence making up the subtracted library, the great majority was specific for B. pseudomallei. This indicates the existence of considerable genomic differences between the two closely related species B. pseudomallei and B. haeundensis, in addition to those noted previously [15–20].

The variable distribution of the cloned sequences amongst the B. pseudomallei strains tested suggests significant genotypic diversity, reminiscent of that observed in B. cepacia [57]. Such diversity may have implications for the pathogenesis of melioidosis and
may, in part, explain the variable presentation of melioidosis [3]. In addition, the fact that strains THP380, THP385 and THP478 are so different both in terms of the absence of subtracted sequences and in haemolysin production is notable and merits further investigation with a larger number of haemolytic and non-haemolytic strains.

A substantial number of genes that encode unknown functions were found in fully sequenced genomes [58–61]. The majority of the subtracted library sequences in the current work appeared to represent so far unknown biological functions and some of these are likely to be involved in virulence, given the manner of their isolation.

Of the minority of clones whose potential gene products were similar to proteins of known function from prokaryotic organisms, most included cell wall-associated and extracellular proteins, regulatory proteins and proteins involved in DNA re-arrangement via transposition and site-specific recombination. Clearly such sequences represent genes with a potential role in virulence; of particular note are those with similarity to adhesins and capsule production.

It is apparent from the pulsed-field gel analysis that the subtractive hybridisation did not identify a large locus inclusive of most of the sequences from B. pseudomallei 08. Therefore, it is difficult to rationalise the fact that this study identified 38 707 bp of nucleotide sequence with a G+C content that is significantly below that which is normal for the B. pseudomallei genome [1]. Although the existence of one or more pathogenicity islands is not excluded, it would seem that many of the sequences are dispersed.

In conclusion, of the identified sequences present in the B. pseudomallei genome which were absent from B. thailandensis, a substantial proportion represented genes of unknown function, and a proportion showed similarity to known virulence genes. The latter demand further attention to fully characterise the genes that they represent in B. pseudomallei, and to identify their role in the pathogenesis of melioidosis.

We gratefully acknowledge the assistance and access to PFGE equipment provided by Mr John Bates and Ms Helen Smith at Queensland Health Scientific Services. We also thank Des Aia Lew, Bart Currie and T. Pitt and Ms Narelle George for the generous donation of strains for this work. N.F.B. was supported by an Australian Postgraduate Award. This work was partly funded by a Griffith University Research Grant.

References


