

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

Melioidosis 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

Melioidosis 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, melioidosis is a major cause of morbidity and mortality, and yet it remains widely under-recognised [3]. Also adding to the significance of melioidosis is the high rate of relapse (up to 23%), despite long-term maintenance therapy [3]. The disease has various presentations, including acute, chronic and recrudescing conditions affecting many organs [3]. This suggests that *B. pseudomallei* possesses a considerable armoury of virulence factors.

Corresponding to the under-recognition of melioidosis 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-

antigenic polysaccharide moiety of lipopolysaccharide is a demonstrated virulence factor [6], whereas a number of incompletely characterised toxins, and a type III secretion apparatus, show potential as virulence factors [7–9].

Over the past decade a number of powerful techniques aimed at identifying virulence genes have been developed. Many of these are designed to identify genes that are specifically expressed *in vivo* [10]. Subtractive hybridisation was designed originally to select for differentially expressed genes in higher eukaryotes [11], but more recently has been applied to prokaryotes. A modified approach, with genomic DNA, selects for the differences between closely related genomes. This approach is well suited to bacterial species and has been successful in the identification of virulence genes [12–14].

Recently, it has become apparent that there is a division within the species *B. pseudomallei* [15–18]. The phenotype used most commonly to differentiate between the two types of strains is the utilisation of arabinose [15, 16, 18, 19]. Brett *et al.* [18] made a clear distinction based on phenotypic characteristics, including arabinose utilisation, exoproduct production and virulence in an animal model, as well as the nucleotide sequence of a portion of the 16S rRNA genes. The avirulent strains were referred to as '*Burkholderia pseudomallei*-like' strains [18] but have now been proposed as a new species, *B. thailandensis* [20].

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 TG1 [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%, proteinase K 100 µg/ml). Minipreps 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 electroporator (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

Strain	Source	Variable 16S rDNA sequence*	Haemolysin	Reference
<i>B. pseudomallei</i>				
RBH	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	N. George, Royal Brisbane Hospital
THP380	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Strong	[22]
THP385	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Strong	[22]
THP401	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	[22]
THP478	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Strong	[22]
THP375	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	[22]
THP379	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	[22]
THP419	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	[22]
76	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	[23]
GF1	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	[24]
GF2	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	[24]
GF3	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	[24]
01	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie, Menzies School of Health Research
02	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
03	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
04	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
05	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
06	Environmental	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
07	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
08	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
09	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
10	Clinical	TCATTCTGGCTAATACCCGGAGTGG	Weak	Dr B. Currie
<i>B. thailandensis</i>				
E264	Environmental	TCATCTGGCTAATAACCCGGGTGG	–	Dr T. Pitt, CPHL London
<i>B. cepacia</i>				
ATCC17765	–	TCCTTGGCTCTAATACAGCCGGGG	–	N. George

*The sequence shown is the variable region of the 320-bp PCR product from the 16S rRNA gene. Underlined bases indicate bases that differ from the recognised *B. pseudomallei* 16S rRNA 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 on to 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 $\alpha^{32}\text{P}$ -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×: 3 M NaCl, 0.3 M 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 Biotec).

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 1. 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

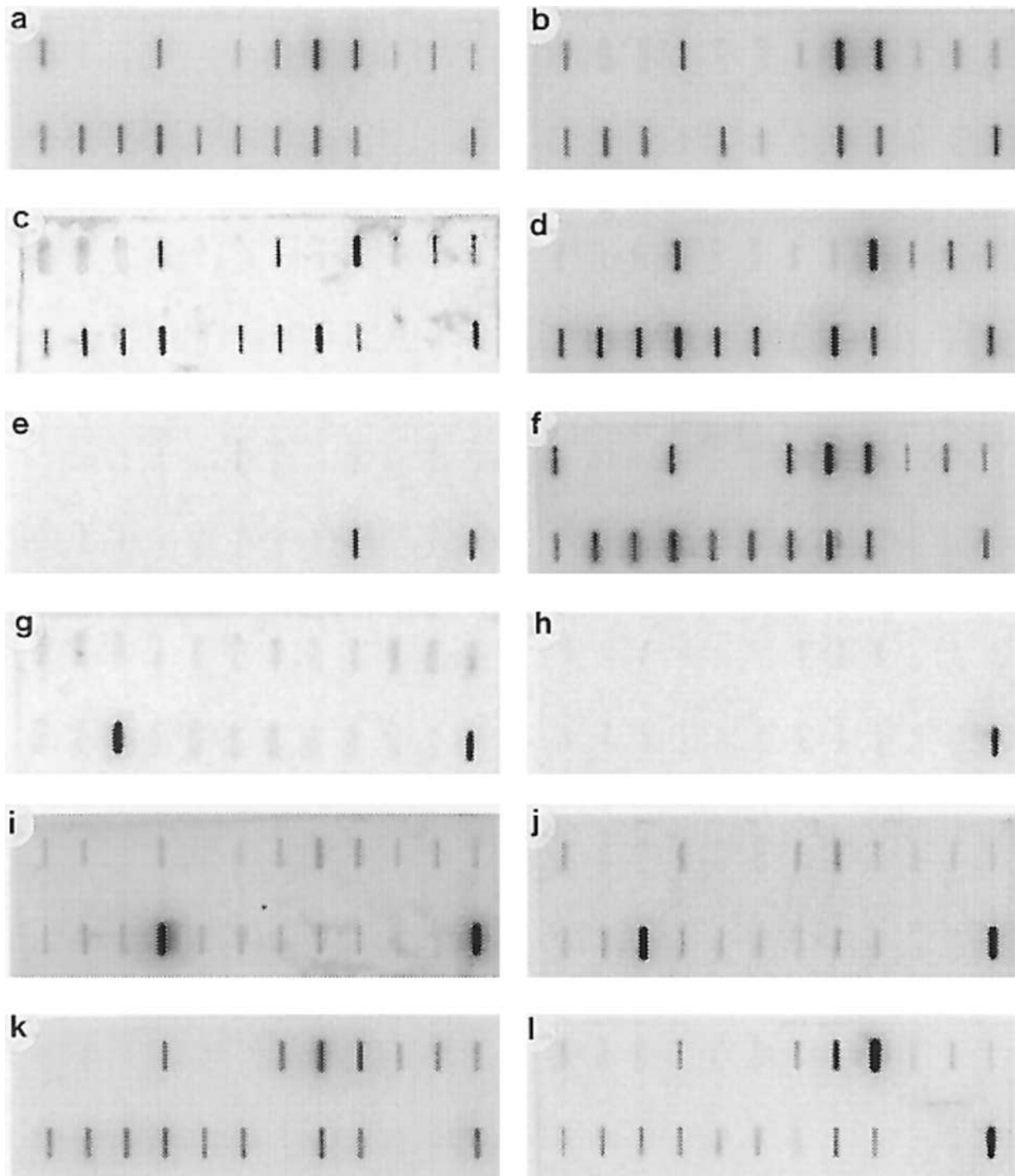


Fig. 1. Slot-blot hybridisations of genomic DNA of *Burkholderia* strains with probes consisting of the subtracted DNA fragments. The strains from which the genomic DNA was derived were arranged in the blots as follows: left to right, top row, RBH, THP380, THP385, THP401, THP478, THP375, THP379, THP419, 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. 1i). There

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 *AluI* 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*, *Arabidopsis thaliana*, *Methanococcus jannaschii*, *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. 043654) and a glycerophosphoryl 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

Clone	Length (bp)	G+C mol%	ORFs*	BLAST hits; expect value
pSC-A10	235	54.4	1	Hypothetical (<i>Pyrococcus horikoshii</i>); 7e-07 BxlE (<i>Streptomyces lividans</i>); 0.001
pSC-A17	730	52.6	2	Permease (<i>Thermus</i> sp.); 2e-25 Hypothetical (thermophilic bacterium RT8.B4); 5e-12
pSC-A31	243	56.4	1	ORF1 of IS407 (<i>B. cepacia</i>); 2e-15 Hypothetical (<i>Enterobacter agglomerans</i>); 1e-14
pSC-A33	813	49.9	2	AIDA-I precursor (<i>E. coli</i>); 0.008 No definition (<i>Caenorhabditis elegans</i>); 0.038
pSC-A35	504	43.8	2	DmsR (<i>Rhodobacter sphaeroides</i>); 5e-04 CpxR (<i>E. coli</i>); 2e-06
pSC-A39	732	52.9	4	CpxB (<i>Actinobacillus pleuropneumoniae</i>); 3e-41 ctrC (<i>Neisseria meningitidis</i>); 1e-40
pSC-A42	435	52.9	3	ORF4 of IS407 (<i>B. cepacia</i>); 4e-05 ORFA (<i>Xanthomonas campestris</i>); 0.010
pSC-A46	583	53.3	4	BbH3 (<i>Bord. bronchiseptica</i>); 1e-05 BpH3 (<i>Bord. pertussis</i>); 2e-05
pSC-A71	171	62.6	2	Tal (<i>Myxococcus xanthus</i>); 8e-19 Gramicidin S synthetase II (<i>Bacillus brevis</i>); 1e-18
pSC-B6	219	56.2	1	Rotamase C (<i>E. coli</i>); 2e-04 Lipoprotein (<i>Bacillus subtilis</i>); 7e-04
pSC-B21	504	51.6	1	HMW1 (<i>Haemophilus influenzae</i>); 2e-18 HxuA precursor (<i>Haemophilus influenzae</i>); 1e-15
pSC-B42	1023	49.7	3	Old (<i>bacteriophage P2</i>); 3e-06 Hypothetical (<i>Methanococcus jannashchii</i>); 6e-05
pSC-B48	572	49.8	1	Transposase (<i>Bord. parapertussis</i>); 2e-09 Isoleucyl-tRNA synthetase (<i>Campylobacter jejuni</i>); 0.46
pSC-B52	757	48	5	CpxR (<i>E. coli</i>); 2e-06 OmpR (<i>Proteus vulgaris</i>); 9e-05
pSC-B54	328	54.9	2	Transposase (<i>Alcaligenes eutrophus</i>); 6e-14 Putative transposase (<i>Pseudomonas alcaligenes</i>); 2e-13
pSC-B67	534	47	2	rhsD (<i>E. coli</i>); 0.013 Core protein (<i>E. coli</i>); 0.15
pSC-B73	560	50.9	1	Hypothetical (<i>Mycobacterium tuberculosis</i>); 5e-32 Glycerophoryl diester phosphodiesterase (<i>Bacillus subtilis</i>); 2e-11
pSC-B78	248	48	0	Hypothetical (<i>Bacillus subtilis</i>); 0.090 BlaA (<i>Streptomyces cacaoi</i>); 0.12
pSC-B79	741	43.6	1	SlaA (<i>Serratia marcescens</i>); 0.046 Fibre protein (<i>Mastadenovirus</i>); 0.06
pSC-B82	155	56.1	1	Rci (<i>Salmonella typhi</i>); 2e-14 rci (<i>E. coli</i>); 4e-14

*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. 11).

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 CpxB 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 cacaoi*, 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+Cmol% 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

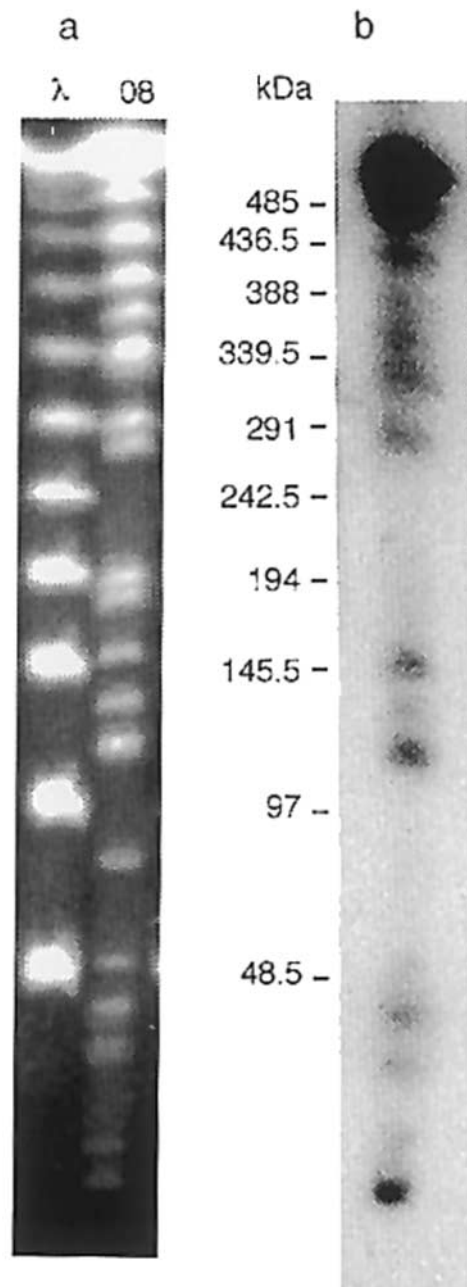


Fig. 2. Southern hybridisation of *SpeI* macro-restriction fragments of *B. pseudomallei* 08 genomic DNA probed with subtractive hybridisation-derived PCR products. (a) PFGE of *B. pseudomallei* 08 genomic *SpeI* fragments showing the λ concatemer mol. wt marker in the left lane and the 08 restriction fragments in the right lane. (b) Image of Southern hybridisation indicating many hybridising *SpeI* restriction fragments.

existence of considerable genomic differences between the two closely related species *B. pseudomallei* and *B. thailandensis*, 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 mol% 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 Drs Ala 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

1. Yabuuchi E, Kosako Y, Oyaizu H *et al.* Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. *Microbiol Immunol* 1992; **36**: 1251–1275.
2. Dance DAB. Melioidosis: the tip of the iceberg? *Clin Microbiol Rev* 1991; **4**: 52–60.
3. Leelarasamee A. *Burkholderia pseudomallei*: the unbeatable foe? *Southeast Asian J Trop Med Public Health* 1998; **29**: 410–415.
4. Sexton MM, Jones AL, Chaowagul W, Woods DE. Purification and characterization of a protease from *Pseudomonas pseudomallei*. *Can J Microbiol* 1994; **40**: 903–910.
5. Ashdown LR, Koehler JM. Production of hemolysin and other extracellular enzymes by clinical isolates of *Pseudomonas pseudomallei*. *J Clin Microbiol* 1990; **28**: 2331–2334.
6. DeShazer D, Brett PJ, Woods DE. The type II O-antigenic polysaccharide moiety of *Burkholderia pseudomallei* lipopolysaccharide is required for serum resistance and virulence. *Mol Microbiol* 1998; **30**: 1081–1100.
7. Haase A, Janzen J, Barrett S, Currie B. Toxin production by *Burkholderia pseudomallei* strains and correlation with severity of melioidosis. *J Med Microbiol* 1997; **46**: 557–563.
8. Häussler S, Nimtz M, Domke T, Wray V, Steinmetz I. Purification and characterization of a cytotoxic exolipid of *Burkholderia pseudomallei*. *Infect Immun* 1998; **66**: 1588–1593.
9. Winstanley C, Hales BA, Hart CA. Evidence for the presence in *Burkholderia pseudomallei* of a type III secretion system-associated gene cluster. *J Med Microbiol* 1999; **48**: 649–656.
10. Handfield M, Levesque RC. Strategies for isolation of *in vivo* expressed genes from bacteria. *FEMS Microbiol Rev* 1999; **23**: 69–91.
11. Sargent TD, Dawid IB. Differential gene expression in the gastrula of *Xenopus laevis*. *Science* 1983; **222**: 135–139.
12. Tinsley CR, Nassif X. Analysis of the genetic differences between *Neisseria meningitidis* and *Neisseria gonorrhoeae*: two closely related bacteria expressing two different pathogenicities. *Proc Natl Acad Sci USA* 1996; **93**: 11109–11114.
13. Townsend KM, Dawkins HJS, Zeng BJ, Watson MW, Papadimitriou JM. Cloning of a unique sequence specific to isolates of type B:2 *Pasteurella multocida*. *Res Vet Sci* 1996; **61**: 199–205.
14. Morrow BJ, Graham JE, Curtiss R. Genomic subtractive hybridization and selective capture of transcribed sequences identify a novel *Salmonella typhimurium* fimbrial operon and putative transcriptional regulator that are absent from the *Salmonella typhi* genome. *Infect Immun* 1999; **67**: 5106–5116.
15. Wuthiekanun V, Smith MD, Dance DAB, Walsh AL, Pitt TL, White NJ. Biochemical characteristics of clinical and environmental isolates of *Burkholderia pseudomallei*. *J Med Microbiol* 1996; **45**: 408–412.
16. Trakulsomboon S, Dance DAB, Smith MD, White NJ, Pitt TL. Ribotype differences between clinical and environmental isolates of *Burkholderia pseudomallei*. *J Med Microbiol* 1997; **46**: 565–570.
17. Winstanley C, Hales BA, Corkill JE, Gallagher MJ, Hart CA. Flagellin gene variation between clinical and environmental isolates of *Burkholderia pseudomallei* contrasts with the invariance among clinical isolates. *J Med Microbiol* 1998; **47**: 689–694.
18. Brett PJ, Deshazer D, Woods DE. Characterization of *Burkholderia pseudomallei* and *Burkholderia pseudomallei*-like strains. *Epidemiol Infect* 1997; **118**: 137–148.
19. Smith MD, Angus BJ, Wuthiekanun V, White NJ. Arabinose assimilation defines a nonvirulent biotype of *Burkholderia pseudomallei*. *Infect Immun* 1997; **65**: 4319–4321.
20. Brett PJ, DeShazer D, Woods DE. *Burkholderia thailandensis* sp. nov., a *Burkholderia pseudomallei*-like species. *Int J Syst Bacteriol* 1998; **48**: 317–320.
21. Brown TA (ed). Molecular biology labfax. Oxford, BIOS Scientific Publishers. 1991.
22. Lew AE, Desmarchelier PM. Molecular typing of *Pseudomonas pseudomallei*: restriction fragment length polymorphisms of rRNA genes. *J Clin Microbiol* 1993; **31**: 533–539.
23. Lew AE, Currie BJ, Withnall KM, Desmarchelier PM. The 1990–1991 outbreak of melioidosis in the Northern Territory of Australia: molecular epidemiology of *Burkholderia pseudomallei* isolates. *Southeast Asian J Trop Med Public Health* 1995; **26**: 466–471.
24. Brook MD, Currie B, Desmarchelier PM. Isolation and identification of *Burkholderia pseudomallei* from soil using selective culture techniques and the polymerase chain reaction. *J Appl Microbiol* 1997; **82**: 589–596.

25. Ausubel FW, Brent R, Kingston RE, Moore DD, Smith JA. Current protocols in molecular biology. New York, Wiley Interscience. 1987.
26. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press. 1989.
27. Diatchenko L, Lau Y-FC, Campbell AP *et al.* Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 1996; **93**: 6025–6030.
28. Altschul SF, Madden TL, Schäffer AA *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389–3402.
29. Nair S, Karim R, Cardosa MJ, Ismail G, Pang T. Convenient and versatile DNA extraction using agarose plugs for ribotyping of problematic bacterial species. *J Microbiol Methods* 1999; **38**: 63–67.
30. Brown NF, Lew AE, Beacham IR. Identification of new transposable genetic elements in *Burkholderia pseudomallei* using subtractive hybridisation. *FEMS Microbiol Lett* 2000; **183**: 73–79.
31. Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 1997; **23**: 1089–1097.
32. Groisman EA, Ochman H. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* 1996; **87**: 791–794.
33. Paitan Y, Alon G, Orr E, Ron EZ, Rosenberg E. The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. *J Mol Biol* 1999; **286**: 465–474.
34. Haggard-Ljungquist E, Barreiro V, Calendar R, Kurnit DM, Cheng H. The P2 phage *old* gene: sequence, transcription and translational control. *Gene* 1989; **85**: 25–33.
35. Zhang X-L, Morris C, Hackett J. Molecular cloning, nucleotide sequence, and function of a site-specific recombinase encoded in the major 'pathogenicity island' of *Salmonella typhi*. *Gene* 1997; **202**: 139–146.
36. Finlay BB, Falkow S. Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev* 1997; **61**: 136–169.
37. Vian A, Carrascosa AV, Garcia JL, Cortés E. Structure of the β -galactosidase gene from *Thermus* sp. strain T2: expression in *Escherichia coli* and purification in a single step of an active fusion protein. *Appl Environ Microbiol* 1998; **64**: 2187–2191.
38. Ward CK, Inzana TJ. Identification and characterization of a DNA region involved in the export of capsular polysaccharide by *Actinobacillus pleuropneumoniae* serotype 5a. *Infect Immun* 1997; **65**: 2491–2496.
39. Benz I, Schmidt MA. Diffuse adherence of enteropathogenic *Escherichia coli* strains. *Res Microbiol* 1990; **141**: 785–786.
40. Benz I, Schmidt MA. Isolation and serologic characterization of AIDA-I, the adhesin mediating the diffuse adherence phenotype of the diarrhea-associated *Escherichia coli* strain 2787 (O126:H27). *Infect Immun* 1992; **60**: 13–18.
41. Suhr M, Benz I, Schmidt MA. Processing of the AIDA-I precursor: removal of AIDA_c and evidence for the outer membrane anchoring as a β -barrel structure. *Mol Microbiol* 1996; **22**: 31–42.
42. Barenkamp SJ, Leininger E. Cloning, expression, and DNA sequence analysis of genes encoding nontypeable *Haemophilus influenzae* high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of *Bordetella pertussis*. *Infect Immun* 1992; **60**: 1302–1313.
43. St. Geme JW, Falkow S, Barenkamp SJ. High-molecular-weight proteins of nontypeable *Haemophilus influenzae* mediate attachment to human epithelial cells. *Proc Natl Acad Sci USA* 1993; **90**: 2875–2879.
44. Krasan GP, Cutter D, Block SL, St. Geme JW. Adhesion expression in matched nasopharyngeal and middle ear isolates of nontypeable *Haemophilus influenzae* from children with acute otitis media. *Infect Immun* 1999; **67**: 449–454.
45. Rao VK, Krasan GP, Hendrixson DR, Dawid S, St. Geme JW. Molecular determinants of the pathogenesis of disease due to nontypeable *Haemophilus influenzae*. *FEMS Microbiol Rev* 1999; **23**: 99–129.
46. Hill CW, Sandt CH, Vlazey DA. *Rhs* elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. *Mol Microbiol* 1994; **12**: 865–871.
47. Kawai E, Akatsuka H, Idei A, Shibatani T, Omori K. *Serratia marcescens* S-layer protein is secreted extracellularly via an ATP-binding cassette exporter, the Lip system. *Mol Microbiol* 1998; **27**: 941–952.
48. Liu J, Walsh CT. Peptidyl-prolyl cis-trans-isomerase from *Escherichia coli*: a periplasmic homolog of cyclophilin that is not inhibited by cyclosporin A. *Proc Natl Acad Sci USA* 1990; **87**: 4028–4032.
49. Ujiye T, Yamamoto I, Satoh T. The *dmsR* gene encoding a dimethyl sulfoxide-responsive regulator for expression of *dmsCBA* (dimethyl sulfoxide respiration genes) in *Rhodobacter sphaeroides* f. sp. denitrificans. *Biochim Biophys Acta* 1997; **1353**: 84–92.
50. Dong J, Iuchi S, Kwan HS, Lu Z, Lin EC. The deduced amino-acid sequence of the cloned *cpxR* gene suggests the protein is the cognate regulator for the membrane sensor, CpxA, in a two-component signal transduction system of *Escherichia coli*. *Gene* 1993; **136**: 227–230.
51. Missiakas D, Raina S. Signal transduction pathways in response to protein misfolding in the extracytoplasmic compartments of *E. coli*: role of two new phosphoprotein phosphatases PrpA and PrpB. *EMBO J* 1997; **16**: 1670–1685.
52. Dartigalongue C, Raina S. A new heat-shock gene, *ppiD*, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. *EMBO J* 1998; **17**: 3968–3980.
53. Ravio TL, Silhavy TJ. The σ^E and Cpx regulatory pathways: overlapping but distinct envelope stress responses. *Curr Opin Microbiol* 1999; **2**: 159–165.
54. Nakayama S-I, Watanabe H. Identification of *cpxR* as a positive regulator essential for expression of the *Shigella sonnei virF* gene. *J Bacteriol* 1998; **180**: 3522–3528.
55. Urabe H, Ogawara H. Nucleotide sequence and transcriptional analysis of activator-regulator proteins for beta-lactamase in *Streptomyces cacaoi*. *J Bacteriol* 1992; **174**: 2834–2842.
56. Goyard S, Bertin P. Characterization of BpH3, an H-NS-like protein in *Bordetella pertussis*. *Mol Microbiol* 1997; **24**: 815–823.
57. Lessie TG, Hendrickson W, Manning BD, Devereux R. Genomic complexity and plasticity of *Burkholderia cepacia*. *FEMS Microbiol Lett* 1996; **144**: 117–128.
58. Fleishmann RD, Adams MD, White O *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; **269**: 496–512.
59. Koonin EV, Mushegian AR, Rudd KE. Sequencing and analysis of bacterial genomes. *Curr Biol* 1996; **6**: 404–416.
60. Hinton JCD. The *Escherichia coli* genome sequence: the end of an era or the start of the FUN? *Mol Microbiol* 1997; **26**: 417–422.
61. Choudhary M, Mackenzie C, Nereng K, Sodergren E, Weinstock GM, Kaplan S. Low-resolution sequencing of *Rhodobacter sphaeroides* 2.4.1^T: chromosome II is a true chromosome. *Microbiology* 1997; **143**: 3085–3099.