J. Med. Microbiol. — Vol. 51 (2002), 374–384

ISSN 0022-2615

MICROBIAL PATHOGENICITY

Distribution of type III secretion gene clusters in *Burkholderia pseudomallei*, *B. thailandensis* and *B. mallei*

LUCILLE RAINBOW, C. ANTHONY HART and CRAIG WINSTANLEY

Department of Medical Microbiology and Genitourinary Medicine, University of Liverpool, Duncan Building, Daulby Street, Liverpool L69 3GA, UK

***Burkholderia pseudomallei*, the causative agent of melioidosis, carries a cluster of genes closely related in organisation to the type III secretion (TTS) system gene clusters of the plant pathogens *Ralstonia solanacearum* and *Xanthomonas* spp. The TTS gene cluster (TTS1) is present only in *B. pseudomallei* and not in avirulent *B. thailandensis*. Adjacent to the gene cluster encoding putative secreton structural proteins lie a number of open reading frames (ORFs) encoding putative proteins with little or no homology to known proteins, with the exception of one predicted protein with homology to *Pseudomonas syringae* HrpK. In both *R. solanacearum* and *Xanthomonas* spp., genes in this location encode secreted effector proteins. RT-PCR analysis indicated that TTS genes, including two of these ORFs, are expressed in broth at 37°C. Analysis of genome sequence data identified a second cluster of TTS genes (TTS2) present in both *B. pseudomallei* and *B. mallei* (99% identity). However, *B. mallei* appears to lack the TTS1 gene cluster. PCR assays indicated that TTS2 was also present in *B. thailandensis*. TTS1 and TTS2 are similar in gene organisation, but nucleotide sequences are sufficiently divergent to suggest that the two TTS systems may have different roles.**

Introduction

Burkholderia pseudomallei is the causative agent of melioidosis, a disease that can manifest itself as an acute, subacute or chronic infection [1, 2]. The acute form of the disease accounts for a significant proportion of often fatal, community-acquired septicaemia in areas of south-east Asia and Australia. *B. pseudomallei* also causes a subacute febrile illness leading to systemic abscess formation involving various organs. The chronic form of the disease is evident from discovery at autopsy or by subsequent activation causing other forms of the disease. It is known that *B. pseudomallei* can enter a dormant state following initial infection or recovery from clinical disease [3], and that dormant cells can be triggered, leading to relapse or the onset of acute melioidosis [4]. The mechanisms underlying these different clinical manifestations and the triggers causing the more serious forms of the disease are poorly understood.

Received 5 Sept. 2001; revised version accepted 7 Dec. 2001.
Corresponding author: Dr C. Winstanley (e-mail: C. Winstanley@liverpool.ac.uk).

It has been suggested that environmental *B. pseudomallei* isolates can be separated on the basis of a number of factors into two biotypes, best defined by the ability to assimilate arabinose. There is evidence that *B. thailandensis* (Ara^+ strains) [5] are avirulent whereas Ara^- strains constitute melioidosis-causing 'true' *B. pseudomallei* [6]. It is possible to discriminate between the two biotypes by molecular approaches such as *fliC* variation [7, 8] or multiplex PCR assays [9]. Furthermore, differences in genomic macrorestriction patterns between the two biotypes have been reported [10]. As the mechanisms of *B. pseudomallei* virulence are poorly understood, there is considerable interest in the identification of the specific factors determining the greater virulence of *B. pseudomallei*. Indeed, subtractive hybridisation has been used to identify some virulence genes present in *B. pseudomallei* but absent from *B. thailandensis* [11].

Differences in virulence between closely related bacterial pathogens are often attributable to the presence of pathogenicity islands (PIs) in virulent strains. Type III secretion (TTS) systems have been implicated in the pathogenicity of several gram-negative bacterial pathogens, including intracellular bacteria such as *Shigella*

spp. [12, 13]. The genes encoding TTS systems and secreted effector proteins are associated with PIs and, therefore, are prime candidates for involvement in the greater virulence of some strains over other related strains.

A previous study identified, in *B. pseudomallei*, a cluster of putative genes homologous to those encoding HpaP, HrcQ (HrpQ), HrcR (HrpT), HrcS (HrpU) and HrpV, all TTS-associated proteins in the plant pathogen *Ralstonia solanacearum* [14]. A further study, based on the detection of a single locus, reported a link between the presence of putative TTS system genes and the Ara^- phenotype, the more virulent biotype of *B. pseudomallei* [15]. There was one exception to the link, a *B. thailandensis* strain (E27) which was positive by both PCR and dot-blot assay.

This study reports the nucleotide sequence of the entire putative structural region of the *B. pseudomallei* TTS1 gene cluster, identifies ORFs that may encode potential effectors and provides evidence that TTS genes are expressed. The study also presents more comprehensive evidence of the link between TTS1 genes and the Ara^- biotype and reports evidence for a second TTS gene cluster (TTS2) common to *B. pseudomallei* and *B. mallei*, the causative agent of glanders.

Materials and methods

Bacterial strains

The strains used in this study have been described previously [15] and are listed in Table 1. *B. pseudomallei* and *B. thailandensis* strains were maintained on nutrient agar.

Nucleotide sequence of TTS gene PI

Cosmid clones c503-23, c503-5 and c503-V1 were previously identified as containing a TTS gene cluster [14], for which we propose the name TTS1. Fragments generated by *Hin*dIII digestion were subcloned from these cosmid clones into *Hin*dIII-digested pUC19 (Helena Biosciences). These and other subclones were sequenced by primer-walking with vector and internal oligonucleotide primer sequences.

Computer analyses

Nucleotide sequence alignments, G + C mol% values, determination of amino acid composition, predicted protein mass and alignments of predicted proteins with other related proteins (retrieved from EMBL, GenBank, PIR or SwissProt [16]) were performed with BESTFIT, GAP, PILEUP, PRETTY, COMPOSITION, PEPTIDESORT and FASTA from the GCG sequence analysis software package (Genetics Computer Group, University of Wisconsin). Hydrophobicity profiles and predictions of membrane-spanning regions were obtained with the PEPTIDESTRUCTURE and PEPPLOT programmes, employing the hydropathy measure of Kyte and Doolittle [17] and the measure of helical hydrophobic moment [18]. The programme MOTIFS was used to identify motifs within the predicted protein sequences. BLAST searches were conducted *via* the site http://www.ncbi.nlm.nih.gov/blast/blast.cgi [19]. Genome sequence data for *B. pseudomallei*, used to complete *orf13*, and for TTS2, were obtained from the Sanger Centre website http://www.sanger.ac.uk/Projects/B_pseudomallei/. These sequence data were produced by the *B. pseudomallei* Sequencing Group at the Sanger Centre and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/bps/. Preliminary sequence data for the genome sequence of *B. mallei* were obtained from the Institute for Genomic Research website at http://www.tigr.org. Searches of these data were conducted *via* the BLASTX or TBLASTN facilities provided at these websites.

Dot-blot and PCR assays

Genomic DNA extraction from *B. pseudomallei* and dot-blot hybridisation of genomic DNA (*c*. 0.25 μg) were performed as described previously [15]. Digoxigenin-11-2′-dUTP (DIG) (Roche)-labelled probes were made either by PCR amplification with c503-5 or c503-23 DNA as template in the presence of 60 μM DIG, as described previously [15], or by labelling amplicons

Table 1. Strains used in this study and results of TTS1 gene PCR/dot-blot assays

Strains	*orf11* PCR	*orf7* PCR	*orf7* Dot-blot	*bpscJ* PCR	*bpscJ* Dot-blot	*orf4/orf5* Dot-blot	*orf1-bpscQ* gap Dot-blot
B. pseudomallei E503, E504, E505, E506, 204 (E955), 576 (E957), E25 (E958), E8 (E960)	+	+	+	+	+	+	+
B. thailandensis E27 (E956)	–	–	–	+/–	–	–	–
E82 (E959), E32, E111, E125, E132, E135, E202, E216, E251, E253, E254, E255, E260	–	–	–	–	–	–	–
Neisseria meningitidis C311	ND	ND	–	ND	–	–	–

ND, not done.

directly with the DIG-High Prime labelling system (Roche) and following the manufacturer's instructions.

Oligonucleotide primers (Genosys) used in PCR or RT-PCR assays and for the labelling of probes are listed in Table 2 along with the annealing temperatures used. Typically, genomic DNA (1 μl) was used directly in 25-μl volumes containing 2 units of Dynazyme (Flowgen), 200 nM of each primer, 1× Dynazyme buffer and 100 μM nucleotides (dATP, dCTP, dGTP, dTTP). Amplifications were performed 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. At the end of the amplification, 5-μl samples were subjected to electrophoresis on a standard agarose 1.0% w/v gel to confirm the presence of an amplified product.

RNA extraction and RT-PCR

B. pseudomallei E503 (also known as strain 10-705) and *B. thailandensis* E27 were cultured in Luria broth at 37°C with and without the addition of 10 mM nitrilotriacetic acid (NTA) until they reached exponential growth (A_{600} of *c*. 0.5). Cells were harvested by centrifugation and RNA was extracted with an RNeasy kit (Qiagen). Total RNA was treated with DNAase following the manufacturer's instructions (Life Technologies) and reverse transcribed into cDNA with *c*. 1 μg of total RNA and 200 ng of random hexamers (Life Technologies) in the presence of RNaseOUT ribonuclease inhibitor (2 U/μl; Life Technologies). PCR amplification was performed as described earlier. The flagellin gene (*fliC*) was amplified with the primers BC6E and BCR14 as described previously [7].

RT-PCR products visualised after agarose 1% w/v gel electrophoresis were often faint. To ensure that there was convincing discrimination between positive and negative PCR assays, gels were Southern blotted by standard procedures and probed with equivalent labelled PCR products generated from genomic DNA. After purification through Microspin S-400 HR columns (Amersham Pharmacia Biotech), PCR products were labelled with DIG with a DIG-High Prime labelling system (Roche).

Accession number

The *B. pseudomallei* TTS1 system gene cluster nucleotide sequence reported in this paper can be found under the GenBank accession no. AF074878.

Results

Map of the TTS genes

A map of the strain E503 region sequenced from three overlapping cosmid clones is presented in Fig. 1. It has been possible to extend further to complete *orf13* with additional sequence data produced by the *B. pseudomallei* Sequencing Group at the Sanger Centre (obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/bps). The region spanning *bpscC* to *orf3* bears a close resemblance in gene organisation to the TTS gene clusters encoding the Hrp secretons of the plant pathogens *R. solanacearum* [20, 21] and *Xanthomonas* spp. (*X. campestris* pv. *vesicatoria* or *X. oryzae* pv. *oryzae*) [22–24].

Protein sequence alignments and motifs

Table 3 indicates the sequence similarities of the equivalent predicted proteins from *B. pseudomallei* and *R. solanacearum.* In most cases there was good alignment between the predicted proteins of the two organisms. BLASTP searches indicated that most of

Table 2. Oligonucleotide primers used in this study

Target DNA (amplicon size)	Oligonucleotide primer sets for PCR	Annealing temperature (°C)
orf11	PM122: 5′-ATCGCCAAATGCCGGGTTTC	55
(706 bp)	PM123: 5′-GTGCATCCATTCATCAAAG	
orf10	PM121: 5′-AGGCCCGTTACCCGATTG-3′	60
(219 bp)	PM127: 5′-CGACAGGACTGGAGAAAG-3′	
orf7	PM73: 5′-GACGTACGAACCGAACAA-3′	60
(1059 bp)	PM80: 5′-AAGAAGTACGCGGCATTC-3′	
bpscJ	PM98: 5′-TGACGGTTTGGCTGTTTG-3′	55
(481 bp)	PM99: 5′-GATGGACTGCAAAGAATG-3′	
orf4/orf5	PM84: 5′-ATCCTGTCCGAGGTTTGC-3′	50
(646 bp)	PM95: 5′-ACATTCACTTCACTCAGG-3′	
*orf1-bpscQ*gap*	PMXBA3: 5′-TGCATCTAGAATTCGGTTCGACGTTCCC-3′	65
(490 bp)	PMBAM: 5′-ATGTGGATCCGATCGACAGTTCAATCACCG-3′	
orf13	SEPCF: 5′-TGAGAGCAAACCGTAATA-3′	48
(722 bp)	SEPCR: 5′-AAGGCTCATCCAGAACGTTC-3′	
bpscU2 (TTS2)	TTS2UF: 5′-TGGCCGAAGAGAAAACCGAA-3′	63
(488 bp)	TTS2UR2: 5′-AAACAGCGAGGTCATCATC-3′	
sopE homologue	TTS3F: 5′-ATGACTTACAACCCGAGA-3′	55
(TTS3) (608 bp)	TTS3R: 5′-CTGAAGCTCGCGATACTC-3′	

*PMXBA3 and PMBAM contain terminal restriction sites that are not relevant to this study.

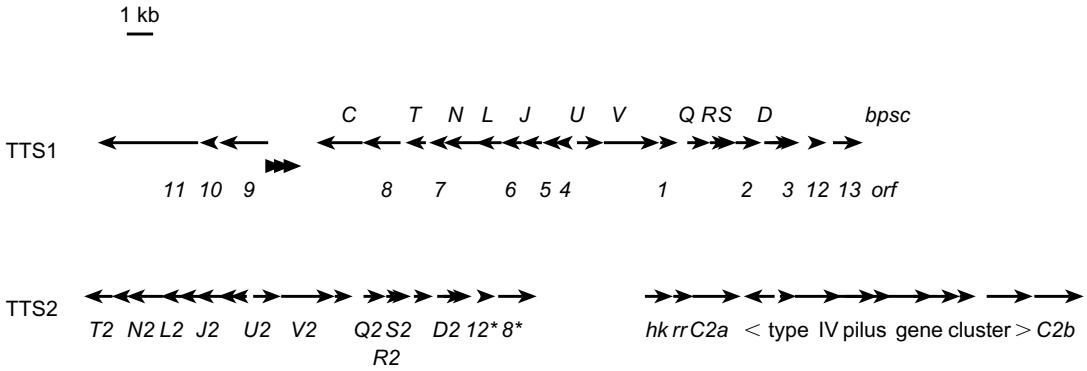

Fig. 1. Map of the TTS1 and TTS2 gene cluster. The location and lengths of predicted genes are shown. Arrows indicate the direction of transcription. The location of an imperfectly repeated sequence in TTS1 is also indicated. TTS2 genes have been labelled according to comparison with equivalent TTS1 genes. *12** and *8** indicate ORFs with homology to TTS1 *orf12* and *orf8* respectively. Genes with homology to histidine kinases (*hk*) and response regulators (*rr*) from known two-component regulatory systems are also indicated. Only genes with predicted protein sequences showing significant homology to proteins in the database are shown. The gap between *orf8** and *hk* is to scale but may contain other ORFs. The organisation of TTS2 genes is based on data obtained from the unfinished *B. pseudomallei* and *B. mallei* genome sequence projects.

Table 3. Properties of predicted TTS1 proteins and comparison with equivalent *R. solanacearum* and TTS2 proteins

Protein*	*R. solanacearum* % similarity/% identity†	TTS2 % similarity/% identity	*Yersinia* homologue	Predicted location/properties
BpscC (HrpA/HrcC)	56.2/45.8	66.4/54.1 (C2b) 39.2/29.6 (C2a)	YscC	Outer-membrane secretin
Orf8 (HrpB)	47.7/36.1	50.7/42.1	None	Transcriptional activator
BpscT (HrpC/HrcT)	50.6/41.4	56.5/48.5	YscT	Inner membrane
Orf7 (HrpD)	41.4/29.9	40.7/33.5	None	Unknown
BpscN (HrpE/HrcN)	71.5/63.5	79.9/72.2	YscN	ATPase
BpscL (HrpF)	41.0/32.1	57.3/46.4	YscL	Unknown
Orf6 (HrpH)‡	37.0/33.0	40.9/35.6	None	Unknown
BpscJ (HrpI/HrcJ)§	61.5/53.2	64.5/59.7	YscJ	Lipoprotein
Orf5 (HrpJ)	35.5/25.8	40.2/31.5	None	Unknown
Orf4 (HrpK)	41.0/30.8	42.9/35.4	None	Unknown
BpscU (HrpN/HrcU)	59.8/49.3	59.5/47.3	YscU	Inner membrane
BpscV (HrpO/HrcV)	69.3/61.1	77.6/71.2	YscV (LcrD)	Inner membrane
Orf1 (HpaP)	37.6/29.4	43.7/39.2	None	Unknown
BpscQ (HrpQ/HrcQ)	40.4/29.8	47.5/42.0	YscQ	Translocase
BpscR (HrpT/HrcR)	74.4/66.0	82.0/73.7	YscR	Inner membrane
BpscS (HrpU/HrcS)	64.0/55.8	70.1/57.5	YscS	Inner membrane
Orf2 (HrpV)	33.8/26.8	35.3/29.9	None	Unknown
BpscD (HrpW)	40.9/32.7	49.8/43.8	YscD	Inner membrane
Orf3 (HrpX)	36.4/29.9	47.6/37.8	None	Unknown

*Protein designations shown in parentheses relate to *R. solanacearum* Hrp proteins. Both the original and currently accepted designations for these proteins are given.
†% similarity/% identity values were calculated by GAP/BESTFIT with the following parameters: gap creation penalty 8, gap extension penalty 2.
‡Orf6 and HrpH do not align well. The figures presented were calculated by BESTFIT over a sequence of 113 amino acids.
§BpscJ data were calculated for the whole predicted peptide sequence not for the protein predicted after cleavage of the lipoprotein signal sequence.

the *B. pseudomallei* predicted TTS1 proteins matched best with homologues from either *R. solanacearum* or *Xanthomonas* spp. The exceptions to this were Orf2, Orf3, Orf6 and Orf7, for which the BLAST searches identified no significant similarity. Similarity/identity values were particularly high for the 11 proteins belonging to families whose members are known to be conserved between TTS systems (BpscC, BpscD, BpscN, BpscL, BpscJ, BpscQ, BpscR, BpscS, BpscT, BpscU and BpscV). There was also strong homology between Orf8 and *R. solanacearum* HrpB, a transcriptional activator responsible for regulating TTS genes in *R. solanacearum*. Further analysis with the NCBI Conserved Domain Database at http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml indicated a highly significant match with bacterial regulatory helix-turn-helix proteins of the AraC family in the C-terminal domain (Orf8 394–483). Members of the AraC-XylS family are basic. Whereas HrpB has a pI value of 8.56, Orf8 is only weakly basic, with a pI of 7.06. Even amongst those proteins exhibiting lower similarity/identity values there are similarities in predicted pI and mol. wt values (data not shown). The least convincing match between the two TTS systems was found with Orf6, which does not align well with HrpH, although the predicted pI values are similar.

The predicted properties of BpscQ, BpscR, BpscS, Orf1 and Orf2 have been discussed previously [14]. BpscN had one of the highest sequence identities with its *R. solanacearum* homologue (HrcN) and contains an ATP-binding 'A' consensus sequence (P loop; 182-APAGVGKS) and an ATP synthase $\alpha\beta$ signature (362-PAIDVLGSLS) typical of TTS-associated ATPases. BpscJ has a prokaryotic membrane lipoprotein lipid attachment site (23-VLALSVLLAGC) similar to those found in the equivalent *R. solanacearum* (HrcJ) and *Yersinia* (YscJ) proteins. As was the case with their *R. solanacearum* homologues [20], hydropathy analysis predicted that Orf7 and BpscL were highly hydrophilic.

Analysis with the programme MOTIFS indicated that the predicted protein BpscC contains a protein D signature common to type II and type III secretion outer-membrane secretin proteins (536-GQSLLIAGY STDKRANGVAGVPWLSKIPLLGALF). BpscC also contained a putative signature characteristic of bacterial regulatory proteins of the LysR family. Proteins of the LysR family possess a potential 'helix-turn-helix' DNA-binding motif in their N-terminal section. The amino acid sequence pattern used to detect these proteins is a consensus sequence derived from the complete helix-turn-helix motif and the next 10 residues. In BpscC, the consensus sequence was not detected in the N-terminal region and therefore is probably not significant.

It has been predicted that the N-terminal half of secretins faces the periplasm and this region is not well conserved [25]. The C-terminal region (β domain) is more conserved and is predicted to contain several transmembrane β-strands whose likely location is the outer membrane. BpscC aligns strongly with homologues in the conserved C-terminal region and contains the consensus motif (V,I)PXL(S,G)XIPXXGXLF common to proteins of this family [26]. BpscV aligns particularly strongly with the N-terminal region of *R. solanacearum* HrcV. In HrcV and homologues this region of the protein is predicted to be hydrophobic, with membrane-associated helices [27]. The less conserved C-terminal region is hydrophilic. BpscT aligns with the *R. solanacearum* secreton inner-membrane protein HrcT. Proteins of this family are highly hydrophobic, containing a number of membrane-associated helices.

Downstream of *bpscC* lies a region containing a 264-bp imperfect triple direct repeat sequence that marks the boundary between the gene cluster with homology to *R. solanacearum* and *Xanthomonas* spp. and several putative ORFs (Table 4) likely to include genes encoding secreted proteins if the *B. pseudomallei* TTS gene cluster continues its resemblance to that of the plant pathogens. It is possible to identify an ORF within this repeated region. If genuine, this ORF, transcribed in the same direction as *bpscC* and *orf9*, would encode a glycine-rich 461 residue peptide with predicted mol. wt and pI values of 45764 Da and 4.16, respectively, and no significant homology to known proteins. As the repeated sequences are in-frame with each other, any such peptide would contain regions of repeated amino acid sequences.

Orf9 is predicted to be a largely hydrophilic protein with a C-terminal hydrophobic α-helical domain. BLAST searches identified homology with *P. syringae* HrpK (32% identity/46% positives between residues over 338 amino acid residues (94–418); accession no. AF232004), a hydrophilic protein for which a function has yet to be ascribed. However, *P. syringae* HrpK does not contain a C-terminal hydrophobic domain. BLAST searches indicated that Orf10 shared best homology with acetyltransferases. The highest match was 35% identity and 51% positives over 65 amino acid residues (Orf10 73–132) against a ribosomal-protein-alanine N-acetyltransferase of *Bacillus halodurans* (accession no. AP001509). BLAST searches failed to identify any significant homology between Orf11 and known proteins. MOTIFS indicated that Orf11 contains a putative actinin-type actin-binding domain signature ((E,Q)x2(A,T,V)(F,Y)x2WxN). In most proteins identified as possessing this motif, including α-actinin, the signature resides within the N-terminal region. An exception to this rule is fimbrin, where the signature is located in duplicate in the C-terminal region. In Orf11, the signature is located between residues 551 and 560. Orf11 also contains a putative ATP/GTP-binding site – P-loop; (A,G)x4GK(S,T) – typical of kinases, ADP-ribosylating factors and a number of proteins involved in secretion.

Downstream of Orf3 lie other potential genes. The predicted protein of Orf12 shows strong homology to HpaB (accession no. BAB07868), a protein of the

Table 4. Predicted properties of putative proteins encoded by ORFs downstream of *bpscC* and *orf3*

ORF	Length	Mol. wt (Da)	pI	Properties/BLASTP matches
9	602 (641)	61412 (67678)	4.33 (4.98)	Homology with *Pseudomonas syringae* HrpK (properties indicated in parentheses) (338 residues; $E = 1 \times e^{-29}$)
10	161	17852	7.10	Some homology with acetyltransferases
11	1287	136278	9.34	Actinin-type actin-binding domain signature; glycine rich; ATP/GTP-binding site motif A (P-loop)
12	151	16847	4.18	Homology with *Xanthomonas oryzae* pv. *oryzae* HpaB (151 residues; $E = 1 \times e^{-36}$)
13	371	40079	10.40	Homology with *Serratia entomophila* SepC (327 residues; $E = 2 \times e^{-27}$)

Xanthomonas spp. TTS system. These proteins are small and acidic, properties suggesting that they may function as chaperones [12]. Further downstream lies Orf13, which has homology to SepC, a potential insect toxin of *Serratia entomophila* [28].

Dot-blot and PCR assays

Table 1 summarises the data obtained by dot-blot and PCR assays to determine the presence of different TTS genes in various strains of *B. pseudomallei* and *B. thailandensis*. Five TTS1 loci were targeted by dot-blot and PCR assays. Only dot-blot data are presented for two of the loci because PCR amplification with the primers PM84/PM95 (for *orf4*/*orf5*) and PMXBA3/PMBAM (for the gap between *orf1* and *bpscQ*) with the various genomic DNAs proved to be difficult. However, in both cases it was possible to produce amplicons for DIG-labelling to make probes for dot-blot analysis. Fig. 2 shows examples of PCR assay results. In all cases the *B. pseudomallei* strains tested positive for the presence of TTS-related DNA whereas *B. thailandensis* strains were negative. Strain E27, which had previously proved positive for *orf2*/*bpscD* by both PCR and dot-blot assays [15], was negative for the loci tested in this study. Occasionally, a faint amplicon of the correct size was visible when primers PM98/PM99 (for *bpscJ*) were used to amplify strain E27 genomic DNA. However, dot-blot analysis suggested that this gene is not present in strain E27, or is significantly different from the *bpscJ* genes present in *B. pseudomallei* strains.

RT-PCR

The conditions required for expression of TTS1 genes in *B. pseudomallei* were not known. In *Pseudomonas aeruginosa*, the addition of NTA to growth media induces TTS genes [29]. The initial approach in the present study involved culturing the bacteria at 37°C in the presence or absence of NTA. Two primer combinations were used to screen for gene expression from putative genes either side of the repeat sequence region of the gene cluster. The three genes targeted were *bpscJ*, *orf10* and *orf11*. The *fliC* gene was targeted as a positive control. In both strains E503 and E27, grown with and without NTA, the *fliC* gene could be amplified from cDNA. All of the controls where reverse-transcriptase was not used were negative, indicating that there was no residual genomic DNA in the cDNA samples (Fig. 3). RT-PCR amplicons for *bpscJ*, *orf10* and *orf11* were obtained with strain E503 samples only. This occurred with and without NTA, indicating that NTA is not required for the expression of these genes in *B. pseudomallei*.

TTS2 and comparisons with B. mallei

Comparisons of *B. pseudomallei* E503 TTS1 sequences with those available through the *B. pseudomallei* genome sequence project indicated that there are sequence variations (typically 1–2%). Such comparisons also suggested that there was a second TTS gene cluster in this bacterium. A cluster of genes homologous to those encoding BpscU, BpscV, Orf1, BpscQ, BpscR, BpscS, Orf2, BpscD, Orf3 and Orf4, could be identified on the same 9315-bp sequence contig. A comparison of this cluster (TTS2) with *B. mallei* genome sequence project data indicated the presence of an equivalent gene cluster (99% sequence identity), and further suggested that genes encoding Orf5, BpscJ, Orf6, BpscL and BpscN were also present in the cluster. Further analysis of another *B. pseudomallei* sequence

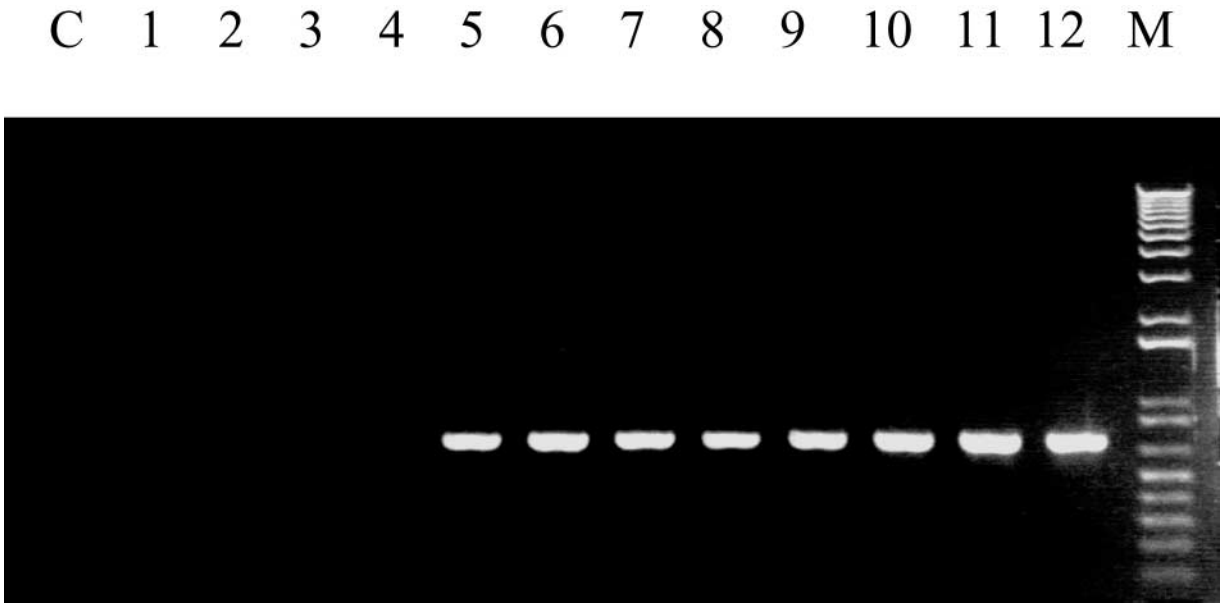

Fig. 2. PCR assays for the presence of TTS1 genes. Examples of results of PCR assays for the presence of *orf11* are shown. The strains tested were: **1**, E27; **2**, E125; **3**, E111; **4**, E32; **5**, E506; **6**, E505; **7**, E504; **8**, E503; **9**, E8; **10**, E25; **11**, 576; **12**, 204. Lanes **1**–**4**, results for *B. thailandensis* strains; **5**–**12**, results for *B. pseudomallei* strains. **C**, reagent control; **M**, 1-kb Plus DNA ladder size marker (Life Technologies, Renfrewshire).

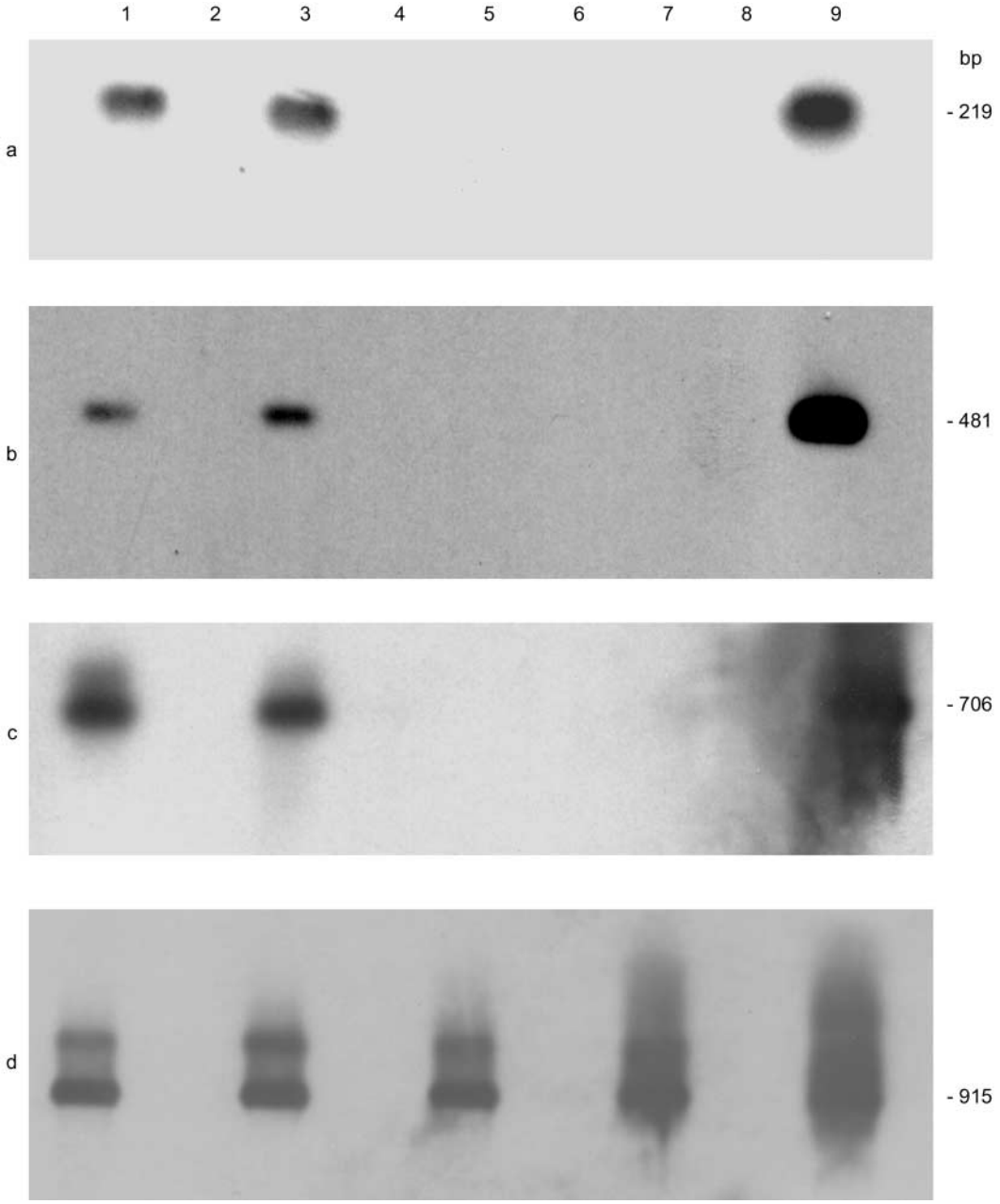

Fig. 3. RT-PCR results for detection of expression of *orf10* (**a**), *bpscJ* (**b**), *orf11* (**c**) and *fliC* (**d**) are shown. The figure comprises Southern blot hybridisation data from four gels. In each case the lanes refer to the following: **1**, strain E503 grown at 37°C in the absence of NTA; **3**, strain E503 grown at 37°C in the presence of NTA; **5**, strain E27 grown at 37°C in the absence of NTA; **7**, strain E27 grown at 37°C in the presence of NTA; **2**, **4**, **6**, **8**, controls lacking reverse-transcriptase for equivalent reactions to preceding lanes; **9**, genomic DNA controls.

contig, overlapping with the *B. mallei* sequence, suggests that the TTS2 gene cluster resembles TTS1 in gene organisation from *bpscT*-*orf12*, but that an equivalent gene to *orf8* lies downstream of the TTS2 *orf12* and that the equivalent gene to *bpscC* is several kilobases further downstream. Indeed, there are two possible *bpscC* equivalents (*bpscC2a* and *bpscC2b*) separated by a cluster of genes with homology to genes involved in type IV pilus production (Fig. 1). PCR assays were used to screen three *B. pseudomallei* and three *B. thailandensis* strains, and indicated that the TTS2 gene *bpscU2* is present in both *B. pseudomallei* and *B. thailandensis* (data not shown). There was no evidence for the presence of the TTS1 gene cluster in *B. mallei* from *orf11* to *orf12* (at the time the search was conducted the data were estimated to represent at least three-fold coverage of the genome). However, a BLASTX search with Orf13 indicated its presence in *B. mallei* (99% sequence identity). Further sequence analysis confirmed that *orf13* and DNA directly downstream are present in *B. mallei* but upstream of *orf13* the sequence similarity ends (between *orf12* and *orf13*). A PCR assay for *orf13* was conducted on a selection of *B. pseudomallei* and *B. thailandensis* strains. The results indicate that unlike *B. mallei* and *B. pseudomallei*, *B. thailandensis* strains lack *orf13* (data not shown).

Discussion

The TTS1 sequenced region of the *B. pseudomallei* E503 genome reported in this paper contains numerous putative genes with homology to TTS system genes reported previously and is highly likely to encode such a system. The overall G + C content (63 mol%) is similar to the value reported previously for the region spanning *orf1–orf2* [14] and is lower than the value estimated for this organism by the *B. pseudomallei* sequencing group (65 mol%; ftp://ftp.sanger.ac.uk/pub/pathogens/bps). However, the difference in G + C content is marginal compared with some reported for TTS PIs in other bacteria [12]. Although the genes encoding TTS systems are generally clustered together, the order of genes varies between bacteria. Whereas proteins involved in the formation of the secretion structure, or secreton, tend to be conserved throughout TTS systems, other proteins are apparent only in related groups of bacteria, or are unique to a particular TTS system. Often genes are conserved in relative location between bacteria whilst exhibiting low sequence homologies [12, 13]. The organisation of secreton genes in the *B. pseudomallei* TTS1 cluster strongly resembles that seen in the plant pathogens *R. solanacearum* and *Xanthomonas* spp. The only significant difference stems from the fact that in *Xanthomonas* spp., *hrpXv*, the equivalent of the genes encoding the putative transcriptional regulators HrpB (in *R. solanacearum*) and Orf8 (*B. pseudomallei*), is not located between the genes encoding BpscC and BpscT or homologues, but lies in an unlinked position [30].

The regions downstream of *bpscC* and its equivalents differ significantly between *B. pseudomallei* TTS1, *R. solanacearum* and *Xanthomonas* spp. In *R. solanacearum* and *Xanthomonas* spp., this domain marks the boundary between genes encoding proteins with largely structural roles in the formation of a secreton and those that encode the secreted products. Therefore, it would seem reasonable to postulate that ORFs in *B. pseudomallei* lying in an equivalent location may encode potential type III-secreted proteins or accessory proteins associated with secretion. In *B. pseudomallei*, the boundary is marked by the presence of a repeated sequence, which may be indicative of a recombination event that has linked the structural genes of the TTS system with genes encoding effector or other proteins specific to *B. pseudomallei*.

PCR and dot-blot assays in this study confirmed the association between TTS genes and an Ara^- phenotype reported previously [15]. The earlier observations that one *B. thailandensis* strain (E27) was carrying some TTS1 system-related DNA were not further substantiated. The results obtained previously for the *orf2*/*bpscD* region, and the reproducible presence of a faint band following PCR amplification with the PM98/PM99 primer set (for *bpscJ*) cannot simply be explained by the presence of TTS2 in *B. thailandensis* because none of the primers match the TTS2 sequence sufficiently well. There may be DNA in *B. thailandensis* strain E27 that is related to *B. pseudomallei* TTS1 DNA. If this is the case, the equivalent genes in strain E27 are sufficiently different from those in *B. pseudomallei* to ensure that detection is not possible by dot-blot hybridisation in most cases. The alternative explanation is that strain E27 contains DNA equivalent to some but not all the *B. pseudomallei* TTS1 gene cluster. The RT-PCR assays used in this study did not lead to the detection of any expression of TTS1 genes in E27. The problems associated with two of the PCR assays in this study may well be related to minor sequence variations between strains, causing primer mismatching.

In many cases, although TTS is triggered *in situ* by cell contact, growth conditions can induce expression of TTS genes. Secretion of ExoS and other proteins by *P. aeruginosa* can be induced artificially by growth in the presence of the chelator NTA [29]. The present observations indicate that expression of the *B. pseudomallei* TTS system is detectable during growth in Luria broth at 37°C. The addition of NTA to growth media was not required for TTS gene expression. The data presented here do not include quantification of gene expression, and there may be growth conditions that lead to higher levels of expression. However, it is clear that there was some expression of TTS genes in nutrient-rich media at 37°C. There are several examples of artificial induction of TTS, including the TTS systems of *Yersinia* spp. (induced by growth in media containing low levels of calcium ions [31]), *Shigella* spp. (in which chemical compounds, including Congo red, Evans blue and direct orange, induce secretion of Ipa proteins [32]) and *Salmonella enterica* serovar Typhimurium (in which SPI2 genes can be induced by Mg^{2+} deprivation and phosphate starvation [33]). However, temperature induction of TTS without the requirement for additional media supplements is not unprecedented. In *Bordetella bronchiseptica*, TTS is under the control of the BvgAS regulon, and can be induced in response to growth at 37°C (Bvg^+). In contrast, growth at or below 26°C, or in the presence of high concentrations of nicotinic acid or $MgSO_4$ (Bvg^-), excludes the production of almost all of the known virulence factors in *Bordetella* spp., including TTS [34].

There is considerable homology between the predicted Orf8 protein and *R. solanacearum* HrpB, the positive regulator responsible for regulating expression of TTS operons [20, 21], and a member of the AraC family. Putative binding sites for such regulatory proteins are difficult to identify from sequence data. A consensus sequence (TTTTaGYcTtTat, in which nucleotides conserved in ⩾60% of sequences are shown in capital letters) has been reported for VirF binding sites in *Yersinia* spp. [35]. The present study found a sequence conforming to this consensus upstream of the *orf8*

gene. The presence of a putative VirF-like binding site suggests that, as in *R. solanacearum*, the *B. pseudomallei* TTS gene cluster may encode a transcriptional activator that is autoregulated. In *Xanthomonas* spp. HrpXv, the equivalent AraC-like positive regulatory protein, is not autoregulated [30].

In *R. solanacearum* the *hrp* genes are expressed in minimal media but repressed in nutrient-rich media [36], which suggests that the regulatory system differs considerably from that of *B. pseudomallei* TTS1. The fact that the latter system is induced in nutrient-rich media at 37°C may be indicative of a system designed for activity in mammals rather than in plants. A number of candidate ORFs for secreted proteins have been identified by relative location compared to the TTS PIs of *R. solanacearum* and *Xanthomonas* spp. RT-PCR indicated that *orf10* and *orf11* are expressed, suggesting that this region does contain functional, expressed genes. There are putative ribosome-binding sites (AGGA) upstream of *orf9* and *orf10*. The end of *orf9* is separated from the beginning of *orf10* by only 24 bp. Therefore, it seems likely that *orf9*, *orf10* and *orf11* form an operon. Although there is a potential ORF within the repeat region, and there are potential ribosome-binding sequences upstream of this ORF (5′-GGAGGCGGAA); without evidence to the contrary it seems unlikely that this does encode a functional protein. Any such protein would contain the same sequence repeated imperfectly three times with short intervening gaps.

Only one of the predicted protein sequences derived from these ORFs was significantly homologous to proteins identified in other TTS systems. Orf9 had homology to the HrpK protein of *P. syringae*, a protein of unknown function. This homology did not include the C-terminal region, which suggests that there may be functional differences between these proteins. Although Orf10 shared some homology with a number of bacterial acetyltransferases, *E* values were much higher than the level generally considered as indicating significance ($E < 1 \times e^{-5}$). Orf11 is a large protein, which shares no significant homology to any proteins in the database, but contains a number of potential motifs. There is evidence that *B. pseudomallei* is a facultative intracellular pathogen, capable of multiplying within various cell types [37]. It has been demonstrated that, after internalisation, the organism can escape from membrane-bound phagosomes into the cytoplasm [37–39]. More recently it has been shown that *B. pseudomallei* can induce cell fusion leading to the formation of multinucleated giant cells and actin-associated membrane protrusion in phagocytic and non-phagocytic cell lines, potentially enabling the pathogen to spread between cells [40]. It is tempting to speculate that type III-secreted proteins may be implicated in such interactions with actin. However, neither ActA, SipA nor SipC contain the actinin-type actin-binding domain signature sought by the programme MOTIFS, and the significance of the signature identified in Orf11 is difficult to assess without evidence of activity. A number of protein families have the ATP/GTP-binding site motif identified in Orf11. Amongst the most interesting of these are members of the ADP-ribosylation factors family and kinases.

Downstream from *orf11*, following a gap of 310 bp, there is DNA which when translated in the opposite orientation to *orf11* gives a partial predicted protein sequence with strong homology (50% identity/65% similarity over 251 amino acids; BLASTP *E* value of $3 \times e^{-58}$) to a transposase of *R. metallidurans* (AF236817). Residual transposon or IS sequences are a feature of PIs [12] and the presence of such a sequence strongly suggests that the TTS-associated gene cluster ends at this point, and that a putative transposase may have played a role in the acquisition of this gene cluster by *B. pseudomallei* E503. Interestingly, the high levels of sequence identity between strain E503 TTS1 and the equivalent region of the strain chosen for genome sequencing, K96243, ends between *orf11* and the putative transposase gene. Indeed, BLAST searches failed to identify the strain E503 transposase gene in K96243 or *B. mallei*, although there are sequences with homology to other transposases downstream of *orf11* in K96243 (transposase of *Tn*1721/*Tn*501, $E = 3 \times e^{-22}$, over 117 amino acids). However, the K96243 sequence contains a stop codon within the matching region, suggesting either that the genome data is erroneous or that a mutation has occurred that would render the gene inactive. This observation does suggest that there may be DNA sequence differences between strains at the edge of the TTS1 gene cluster, and that there are variations in transposon or IS sequences between strains of *B. pseudomallei*.

With TTS1 predicted proteins as probes in silico, a second TTS gene cluster (TTS2) was identified in both *B. pseudomallei* and *B. mallei*. Analysis indicates that *B. pseudomallei* and *B. mallei* both carry TTS2 genes whereas only *B. pseudomallei* carries the TTS1 gene cluster. In addition, a PCR assay for detection of *bpscU2* indicated that at least part of TTS2 is present in *B. thailandensis*. Recently, it has become apparent that there is a third TTS system gene cluster (tentatively named TTS3). TTS3 includes genes with homology to some of the *Salmonella* Inv/Spa secreted virulence factors [41] and is present in *B. pseudomallei* and *B. mallei*. A PCR assay to screen for the putative gene encoding a protein with homology to *Salmonella* SopE showed evidence to suggest that TTS3 DNA is also present in *B. thailandensis* (data not shown). Thus, only the TTS1 system defines a major difference between *B. pseudomallei* and both *B. thailandensis* and *B. mallei*. One boundary marking the variation between *B. mallei* and *B. pseudomallei* lies between *orf12* and *orf13* in TTS1. At this stage, it is not clear whether *orf13* should be considered part of the TTS1 gene cluster. Orf13 is homologous to the C-terminal region

of the putative insect toxin SepC. Hurst *et al.* [28] demonstrated homology between SepC and other insect toxins. However, those regions common to insect toxins were not the same as the region of best homology with Orf13. Whether Orf13 is involved in TTS or not, it may be a gene of interest because of its presence in *B. mallei* and *B. pseudomallei* but absence from *B. thailandensis*. The TTS2 genes are organised in a similar fashion to the TTS1, *R. solanacearum* and *Xanthomonas* spp. TTS gene clusters suggesting common ancestry for all these systems. Although the *B. pseudomallei* TTS1 and TTS2 genes are most similar to each other, there has been considerable divergence suggesting that the two TTS systems may have different functions. Although the actual role of TTS during infection with *B. pseudomallei* has yet to be resolved, the likelihood is that, as in all other cases, TTS makes an important contribution to the pathogenicity of this bacterium.

This work was funded by the award of a grant to C.W. and C.A.H. from the Wellcome Trust (reference 58935). Whole genome sequencing of *B. mallei* and *B. pseudomallei* was accomplished with support from the National Institute of Health, USA and Beowulf Genomics, respectively.

References

1. Dance DA. Melioidosis: the tip of the iceberg? *Clin Microbiol Rev* 1991; **4**: 52–60.
2. Leelarasamee A, Bovornkitti S. Melioidosis: review and update. *Rev Infect Dis* 1989; **11**: 413–425.
3. Chaowagul W, Suputtamongkol Y, Dance DAB, Rajchanuvong A, Pattara-arechachai J, White NJ. Relapse in melioidosis: incidence and risk factors. *J Infect Dis* 1993; **168**: 1181–1185.
4. Currie BJ, Fisher DA, Anstey NM, Jacups SP. Melioidosis: acute and chronic disease, relapse and re-activation. *Trans R Soc Trop Med Hyg* 2000; **94**: 301–304.
5. Brett PJ, DeShazer D, Woods DE. *Burkholderia thailandensis sp. nov.*, a *Burkholderia pseudomallei*-like species. *Int J Syst Bacteriol* 1998; **48**: 317–320.
6. Smith MD, Angus BJ, Wuthiekanun V, White NJ. Arabinose assimilation defines a nonvirulent biotype of *Burkholderia pseudomallei*. *Infect Immun* 1997; **65**: 4319–4321.
7. 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.
8. Wajanarogana S, Sonthayanon P, Wuthiekanun V, Panyim S, Simpson AJH, Tungpradabkul S. Stable marker on flagellin gene sequences related to arabinose non-assimilating pathogenic *Burkholderia pseudomallei*. *Microbiol Immunol* 1999; **43**: 995–1001.
9. Dharakul T, Tassaneetrithep B, Trakulsomboon S, Songsivilai S. Phylogenetic analysis of Ara^+ and Ara^- *Burkholderia pseudomallei* isolates and development of a multiplex PCR procedure for rapid discrimination between the two biotypes. *J Clin Microbiol* 1999; **37**: 1906–1912.
10. Chaiyaroj SC, Kotrnon K, Koonpaew S, Anantagool N, White NJ, Sirisinha S. Differences in genomic macrorestriction patterns of arabinose-positive (*Burkholderia thailandensis*) and arabinose-negative *Burkholderia pseudomallei*. *Microbiol Immunol* 1999; **43**: 625–630.
11. Reckseidler SL, Deshazer D, Sokol PA, Woods DE. Detection of bacterial virulence genes by subtractive hybridization: identification of capsular polysaccharide of *Burkholderia pseudomallei* as a major virulence determinant. *Infect Immun* 2001; **69**: 34–44.
12. Hueck CJ. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* 1998; **62**: 379–433.
13. Winstanley C, Hart CA. Type III secretion systems and pathogenicity islands. *J Med Microbiol* 2001; **50**: 116–126.
14. 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.
15. Winstanley C, Hart CA. Presence of type III secretion genes in *Burkholderia pseudomallei* correlates with Ara^- phenotypes. *J Clin Microbiol* 2000; **38**: 883–885.
16. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 1988; **85**: 2444–2448.
17. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982; **157**: 105–132.
18. Eisenberg D, Weiss RM, Terwilliger TC. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci USA* 1984; **81**: 140–144.
19. 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.
20. Van Gijsegem F, Gough C, Zischek C *et al*. The *hrp* gene locus of *Pseudomonas solanacearum*, which controls the production of a type III secretion system, encodes eight proteins related to components of the bacterial flagellar biogenesis complex. *Mol Microbiol* 1995; **15**: 1095–1114.
21. Guéneron M, Timmers ACJ, Boucher C, Arlat M. Two novel proteins, PopB, which has functional nuclear localization signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the *Hrp*-secretion apparatus of *Ralstonia solanacearum*. *Mol Microbiol* 2000; **36**: 261–277.
22. Fenselau S, Bonas U. Sequence and expression analysis of the *hrpB* pathogenicity operon of *Xanthomonas campestris pv. vesicatoria* which encodes eight proteins with similarity to components of the Hrp, Ysc, Spa, and Fli secretion systems. *Mol Plant Microbe Interact* 1995; **8**: 845–854.
23. Huguet E, Hahn K, Wengelnik K, Bonas U. *hpaA* mutants of *Xanthomonas campestris pv. vesicatoria* are affected in pathogenicity but retain the ability to induce host-specific hypersensitive reaction. *Mol Microbiol* 1998; **29**: 1379–1390.
24. Zhu W, Magbanua MM, White FF. Identification of two novel *hrp*-associated genes in the *hrp* gene cluster of *Xanthomonas oryzae pv. oryzae*. *J Bacteriol* 2000; **182**: 1844–1853.
25. Guilvout I, Hardie KR, Sauvonnet N, Pugsley AP. Genetic dissection of the outer membrane secretin PulD: are there distinct domains for multimerization and secretion specificity? *J Bacteriol* 1999; **181**: 7212–7220.
26. Genin S, Boucher CA. A superfamily of proteins involved in different secretion pathways in gram-negative bacteria: modular structure and specificity of the N-terminal domain. *Mol Gen Genet* 1994; **243**: 112–118.
27. Gough CL, Genin S, Lopes V, Boucher CA. Homology between the HrpO protein of *Pseudomonas solanacearum* and bacterial proteins implicated in a signal peptide-independent secretion mechanism. *Mol Gen Genet* 1993; **239**: 378–392.
28. Hurst MRH, Glare TR, Jackson TA, Ronson CW. Plasmid-located pathogenicity determinants of *Serratia entomophila*, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of *Photorhabdus luminescens*. *J Bacteriol* 2000; **182**: 5127–5138.
29. Yahr TL, Mende-Mueller LM, Friese MB, Frank DW. Identification of type III secreted products of the *Pseudomonas aeruginosa* exoenzyme S regulon. *J Bacteriol* 1997; **179**: 7165–7168.
30. Wengelnik K, Bonas U. HrpXv, an AraC-type regulator, activates expression of five of the six loci in the *hrp* cluster of *Xanthomonas campestris pv. vesicatoria. J Bacteriol* 1996; **178**: 3462–3469.
31. Cornelis GR, Boland A, Boyd AP *et al*. The virulence plasmid of *Yersinia*, an antihost genome. *Microbiol Mol Biol Rev* 1998; **62**: 1315–1352.
32. Bahrani FK, Sansonetti PJ, Parsot C. Secretion of Ipa proteins by *Shigella flexneri*: inducer molecules and kinetics of activation. *Infect Immun* 1997; **65**: 4005–4010.
33. Deiwick J, Nikolaus T, Erdogan S, Hensel M. Environmental regulation of *Salmonella* pathogenicity island 2 gene expression. *Mol Microbiol* 1999; **31**: 1759–1773.
34. Yuk MH, Harvill ET, Miller JF. The BvgAS virulence control

system regulates type III secretion in *Bordetella bronchiseptica*. *Mol Microbiol* 1998; **28**: 945–959.
35. Wattiau P, Cornelis GR. Identification of DNA sequences recognized by VirF, the transcriptional activator of the *Yersinia yop* regulon. *J Bacteriol* 1994; **176**: 3878–3884.
36. Arlat M, Gough CL, Zischek C, Barberis PA, Trigalet A, Boucher CA. Transcriptional organization and expression of the large *hrp* gene cluster of *Pseudomonas solanacearum*. *Mol Plant Microbe Interact* 1992; **5**: 187–193.
37. Jones AL, Beveridge TJ, Woods DE. Intracellular survival of *Burkholderia pseudomallei*. *Infect Immun* 1996; **64**: 782–790.
38. Pruksachartvuthi S, Aswapokee N, Thankerngpol K. Survival of *Pseudomonas pseudomallei* in human phagocytes. *J Med Microbiol* 1990; **31**: 109–114.
39. Harley VS, Dance DAB, Drasar BS, Tovey G. Effects of *Burkholderia pseudomallei* and other *Burkholderia* species on eukaryotic cells in tissue culture. *Microbios* 1998; **96**: 71–93.
40. Kespichayawattana W, Rattanachetkul S, Wanun T, Utaisincharoen P, Sirisinha S. *Burkholderia pseudomallei* induces cell fusion and actin-associated membrane protrusion: a possible mechanism for cell-to-cell spreading. *Infect Immun* 2000; **68**: 5377–5384.
41. Attree O, Attree I. A second type III secretion system in *Burkholderia pseudomallei*: who is the real culprit? *Microbiol* 2001; **147**: 3197–3199.