MOLECULAR IDENTIFICATION AND EPIDEMIOLOGY

Flagellin gene variation between clinical and environmental isolates of *Burkholderia pseudomallei* contrasts with the invariance among clinical isolates

C. WINSTANLEY, B. A. HALES, J. E. CORKILL*, M. J. GALLAGHER* and C. A. HART*

Department of Biomedical Sciences, University of Bradford, Bradford, West Yorkshire BD7 1DP and *Department of Medical Microbiology, Royal Liverpool University Hospital, PO Box 147, Liverpool L69 3BX

The flagellin gene sequence from a clinical isolate of *Burkholderia pseudomallei* was used to design oligonucleotide primers for PCR/RFLP analysis of flagellin gene variation among clinical and environmental isolates of *B. pseudomallei*. Genes from four clinical and six environmental isolates were amplified and compared by RFLP. The clinical isolates were indistinguishable, but variation was detected among some of the environmental isolates. Sequence analysis of flagellin gene amplified products demonstrated high levels of conservation amongst the flagellin genes of clinical isolates (>99% similarity), compared to the variation observed between the clinical isolates and one of the environmental isolates (<90% similarity). Genomic comparisons with pulsed-field gel electrophoresis (PFGE) revealed differences between the relationships inferred by flagellin genotyping and PFGE, suggesting that a combination of molecular methods may be useful for the subtyping of *B. pseudomallei* strains.

Introduction

*Burkholderia pseudomallei* is the causative agent of melioidosis, an often fatal infection endemic in areas of South East Asia and Australia. It is isolated frequently from soil in regions where the disease occurs and this may be the reservoir for infection in man and other animals. There has been considerable interest in developing adequate typing schemes for *B. pseudomallei*. Several studies have used ribotyping of isolates from northern Australia [1], Western Australia [2] and Thailand [3, 4]. Ribotype groups can be resolved further by random amplified polymorphic DNA [5] or DNA macrorestriction analysis [6]. Molecular approaches to the diagnosis of melioidosis have been developed by targeting regions of the 16S rRNA [7] or 23S rRNA [8] genes with labelled probes or polymerase chain reaction (PCR) amplification.

Bacterial flagellin genes have provided highly variable targets for the development of typing methods in several bacterial species [9], including *Campylobacter* spp. [10], *Salmonella* spp. [11], *Helicobacter pylori* [12] and *Pseudomonas aeruginosa* [13]. In a preliminary study, flagellin genes from *B. cepacia* were cloned and sequenced [14], and a high degree of homology was noted between the *B. cepacia* flagellin sequence and a published *B. pseudomallei* flagellin sequence [15]. This homology means that *B. cepacia* flagellin gene sequences can be used as probes to identify *B. pseudomallei* flagellin genes in gene libraries. This paper reports the cloning and sequencing of the flagellin gene from three clinical and one environmental isolate of *B. pseudomallei*, and the development of a PCR-based approach to the study of flagellin gene variation in *B. pseudomallei*.

Materials and methods

*B. pseudomallei* strains

E505 was a clinical isolate from a patient in Preston, Lancashire, who may have acquired the strain during a visit to Goa. Clinical isolates E503, E504, E506 were from melioidosis patients in Malaysia. Environmental isolates E955–E960 were from Thailand and were kindly provided by Dr T. Pitt, Central Public Health Laboratory, 61 Colindale Avenue, London.
Cloning and sequencing of *B. pseudomallei* flagellin genes

The flagellin gene from the *B. cepacia* strain E243 was cloned and sequenced [14]. A cloned PCR product, generated with the sense primer BC1 ('5'-GTIGCICARCARAAYCTAAYGG-3') and the antisense primer BCR2 ('5'-CCNACSGTCTGCCCCCTCTG-3') and comprising c. 450 bp of the *b. cepacia* flagellin gene, was cloned into the pTAG vector with the LigATor cloning kit (R&D Systems Europe Ltd). This 450-bp cloned fragment was labelled with digoxigenin-11-2'-dUTP (DIG; Boehringer Mannheim) by PCR amplification in a 50-μl reaction containing 1 unit of Dynazyme (Flowgen), 1× Dynazyme buffer, 100 μM nucleotides (dATP, dCTP, dGTP, dTTP), 60 μM DIG and 200 nM of each of two vector-designed primers, LIGA1 ('5'-ACCGGTTACGTATCGGAT-3' and LIGA2 ('5'-CCTAGGCTCGAGAAGCTT-3'). Amplifications were performed in a MiniCycler (Genetic Research Instrumentation) for 30 cycles consisting of 95°C (1 min), 50°C (1 min) and 72°C (1.5 min). A gene library of *B. pseudomallei* E503 genomic DNA, isolated as described previously [16], was constructed with the SuperCos 1 cosmid vector (Stratagene) and the conditions recommended by the supplier. Flagellin gene-containing clones were identified by hybridisation with the DIG-labelled probe, with anti-DIG-AP Fab fragments and the chemiluminescent substrate CDP-Star as directed in the manufacturer's instruction guide (Boehringer). DNA from a flagellin gene-containing cosmid clone was purified with a Qiagen Midi-preparation Kit. A 4-kb SstI-fragment, containing the flagellin gene, was subcloned into pUC19 and both strands were sequenced by the University of Liverpool DNA Sequencing Service with a number of internal oligonucleotide primers. Flagellin gene amplified products generated from strains E505 and E506 with primers BC6 (N-terminal: '5'-AACAGGCCTGGACGCACGGATC3-3') and BCR12 (C-terminal: '5'-ACAG/TGTCCCGGTTTCCGCTTGG-3') and from strain E956 with primers BC6E (N-terminal: '5'-ACCAACAGCCTGCACGGATC3-3') and BCR14 (C-terminal: '5'-TTATTGCAGAGCTTCCGAC-3') were also cloned with the LigATor kit and sequenced. The sequences obtained for strains E505, E506 and E956 represent most but not all of the flagellin genes. These sequences have been deposited in GenBank and given the accession numbers AF030239 (E503), AF030240 (E505) and AF030241 (E956).

**PCR/RFLP analysis of *B. pseudomallei* flagellin genes**

Cell suspensions of *B. pseudomallei* isolates grown on brain heart infusion (BHI) agar were boiled for 5 min to generate crude DNA preparations. Boiled suspension (2.5 μl) was used directly in an amplification mixture consisting of 0.5 units of Dynazyme, 1× Dynazyme buffer, 100 μM nucleotides (dATP, dCTP, dGTP, dTTP) and 200 nM of the flagellin gene-specific primers BC6E (sense primer for N-terminal region) and BCR14 (antisense primer of C-terminal region), in a total volume of 25 μl. Amplifications were performed in a MiniCycler for 30 cycles consisting of 95°C (1 min), 60°C (1 min) and 72°C (1.5 min).

Samples (5 μl) of amplified product were digested with the restriction endonucleases *Hae*III, *Msp*I, *Mbo*I, *Rsa*I and *Cfo*I under the conditions recommended by the supplier (Life Technologies). These digests and PCR size marker (R&D Systems; fragment sizes: 50, 150, 300, 500, 750, 1000, 1500, 2000 bp) were then subjected to electrophoresis on MetaPhor agarose 3% w/v gels (Flowgen).

**Pulsed-field gel electrophoresis**

Bacterial cells were grown overnight at 37°C on blood agar plates, harvested by scraping from plates, washed with and suspended in buffer (10 mM Tris, 20 mM NaCl, 50 mM EDTA, pH 7.2) and mixed with an equal volume of pulsed-field certified agarose (Bio-Rad) 2.0% w/v containing 4 μl of lysozyme (25 mg/ml). The mixture was dispensed into 100-μl insert-moulds and allowed to solidify on ice. The plugs were incubated in 1 ml of lysozyme buffer (10 mM Tris, 50 mM NaCl, sodium deoxycholate 0.2% w/v, sodium lauryl sarcosine 0.5% w/v, pH 7.2) containing 40 μl of lysozyme (25 mg/ml) for 1 h at 37°C. After washing in wash buffer (20 mM Tris, 50 mM EDTA, pH 8.0) the plugs were placed in 1 ml of proteinase K buffer (100 mM EDTA, sodium deoxycholate 0.2% w/v, sodium lauryl sarcosine 1.0% w/v, pH 8.0) containing 1 mg of proteinase K and incubated overnight at 50°C. The plugs were then washed in 1 ml of wash buffer containing 20 μl of 100 mM phenyl methyl sulphonyl fluoride (PMSF) for 30 min. After four further washes each of 30 min in wash buffer, the plugs were treated with *Spe*I (BioRad) 2 units/block in 300 μl of restriction enzyme buffer and incubated at 37°C for 18 h. Restriction fragments were separated by pulsed-field gel electrophoresis (PFGE) in a CHEF DRII system (Bio-Rad) through pulsed-field certified agarose 1% w/v in 0.5× TBE (1× TBE is 0.9 M Tris-HCl, 0.9 M boric acid, 1.0 mM EDTA). Electrophoresis was performed for 20 h in 0.5× TBE at 14°C with initial and final pulse times of 5 s and 35 s respectively. Lambda phage concatamers were used as DNA size standards.

**Cable pilus typing by PCR**

Highly transmissible pathogenic *B. cepacia* (ET12 lineage) characteristically have long intertwined cable pili [17]. PCR amplification of cable pilin genes from *B. cepacia* ET12 and *B. pseudomallei* chromosomal DNA was performed in standard PCR amplification mixtures with sense ('5'-CCAAAGGACTAACCCA-3') and anti-sense ('5'-ACCGGATGTCCTCACACA-3') primers designed to produce an amplicon of 676 bp. An
initial denaturing step of 94°C for 3 min was followed by 30 cycles of 94°C (1 min), 55°C (1 min) and 72°C (1 min).

Computer analyses

Nucleotide sequence alignments, determination of amino-acid composition and alignments of predicted flagellin proteins with each other or with other flagellins (retrieved from EMBL, GenBank or SwissProt) were performed with GAP, PILEUP and FASTA from the GCG sequence analysis software package (Genetics Computer Group, University of Wisconsin) [18].

Results

Flagellin gene sequence comparisons

The complete flagellin gene sequence obtained from strain E503 was identical in length to the published *B. pseudomallei* flagellin gene sequence (1167 bp) [15] and differed in only two positions (798 and 975) both of which were synonymous substitutions, indicating that the peptide sequences were invariable. Flagellin gene and protein variation is known to be far greater in the central region than in the conserved N- and C-terminal domains [9]. To assess flagellin sequence variability in *B. pseudomallei*, sequence information from PCR-amplified products representing the flagellin gene central domain of three other strains was obtained. The flagellin gene nucleotide sequences for strains E505 and E506 were 100% identical over 766 bp. The available sequence data for E506/E505 flagellin genes indicated only one base difference when compared with E503 flagellin gene (position 798 on the complete gene sequence) and one base difference when compared with the published *B. pseudomallei* flagellin gene sequence [15] (position 975).

Flagellin genotyping by PCR/RFLP

PCR amplification was possible with all 10 clinical and environmental isolates of *B. pseudomallei* tested. Restriction digest analysis revealed no detectable variation between the four clinical isolates and the environmental isolates E955, E957 and E958. Environmental isolates E956 and E959 gave different flagellin gene PCR/RFLP patterns from the other isolates with *HaeIII*, *MspI*, *MboI*, *RsaI*, *HinIII*, *MspI*, *MboI*, *RsaI* enzymes. Restriction digest analysis revealed no detectable enzymes *HaeIII*, *MboI*, *RsaI* enzymes. Restriction digest analysis revealed no detectable enzymes.

sequence from another clinical isolate of *B. pseudomallei* [15]. Over the region shown, the E956 flagellin gene sequence shared an 89.5–89.6% identity with the flagellin genes of the clinical isolates. A 15-bp deletion was identified in the E956 flagellin gene sequence (position 449–463; Fig. 2).

PFGE genotyping

PFGE of macrorestricted chromosomal DNA from the four clinical isolates revealed two distinct patterns. Strains E503 and E505 exhibited identical restriction patterns differing in more than four band positions from the pattern found with strains E504 and E506, which were indistinguishable from each other. Of the environmental isolates, E959 and E960 produced unique patterns which were different from each other and all the other isolates. Environmental isolates E955, E956, E957 and E958 produced a pattern differing from that of clinical isolates E503 and E505 by fewer than three bands. Thus they can be considered to be similar isolates of the same strain.

Cable pilus typing by PCR

PCR amplification with *B. cepacia* cable pilin primers produced no detectable amplicons from any of the *B. pseudomallei* isolates used in this study, whereas the control *B. cepacia* (ET12) strain produced the expected amplicon of 676 bp.

Discussion

By sequence alignment, it was possible to predict that the location of the small number of variable nucleotides in the flagellin gene sequences of the clinical isolates would lead to substitutions that were synonymous. Thus, despite the geographical variation in the sources of the isolates, there was no evidence of any
Fig. 2. Alignment of *B. pseudomallei* flagellin gene sequences. PM indicates the previously published sequence of Deshazer et al. [15]. Only bases differing from this sequence are shown for the flagellin genes of E503, E505 and E956. Identical residues are indicated by (.), deletions are indicated by (-).

flagellin protein divergence amongst the clinical isolates of *B. pseudomallei*.

The flagellin gene sequence invariance observed among the clinical isolates of *B. pseudomallei* has been observed in other pathogenic bacteria [19]. The results support the observations of Brett et al. [20], who isolated flagellin proteins from four strains of *B. pseudomallei* and found them all to have a similar mass of c. 43.4 kDa. Flagellin gene sequence invariance, despite the geographical variation in isolate source (Malaysia and Preston/Goa) and the variability often observed in flagellin gene sequences [9], is indicative of a clonal population structure. However,
flagellin genotype variation was found between the clinical isolates and some environmental isolates, although three of the six environmental isolates were indistinguishable from clinical isolates by flagellin gene PCR/RFLP and one of the environmental isolates exhibited variation with only one enzyme.

Macrolestription analysis by PFGE confirmed that clinical isolates E503 and E505, and the environmental isolates E955, E957 and E958 were indistinguishable, suggesting a high degree of clonality in *B. pseudomallei* populations. However, there were differences between the relationships inferred by flagellin genotyping and PFGE. The results suggest that not only can PFGE be used to discriminate between strains of the same flagellin genotype, but flagellin genotyping can be used to distinguish between some strains exhibiting the same PFGE restriction pattern, suggesting that a combination of molecular typing methods may be useful for better discrimination.

The inability to amplify cable pilin genes from *B. pseudomallei* strains suggests either that cable pili are not present or, if present, the cable pilin gene in *B. pseudomallei* is sufficiently different from the *B. cepacia* cable pilin gene to prevent primer annealing.

The pathogenic abilities of the environmental isolates used in this study are unknown, whereas all the clinical isolates were associated with disease. Brett et al. [21] examined a number of clinical and environmental isolates for morphological and biochemical characteristics and for virulence. Having also determined the 16S rRNA sequences, the authors suggested that the highly virulent *B. pseudomallei* strains were
the true B. pseudomallei strains and that strains of low virulence could be distinguished from the true strains by significant differences in exoenzyme production, hamster virulence and 16S rRNA gene sequences. A study of many more isolates of B. pseudomallei will be required to ascertain whether flagellin genotyping can be applied to differentiate between high and low virulence strains of B. pseudomallei. However, the method reported here should prove useful for the identification of different clonal variants of the organism, and may provide a way of discriminating between true B. pseudomallei strains and B. pseudomallei-like strains.

This study was supported by a grant to C.W. from the Wellcome Trust (044249/PMG/VW). This work benefited from the use of the SEQNET facility, Daresbury, UK.

References