Comparative proteomic analysis of *Clostridium difficile* isolates of varying virulence


1Leeds Institute for Biomedical and Clinical Sciences, University of Leeds, Leeds LS1 3EX, UK
2Public Health England, Centre for Infections, London NW9 5EQ, UK
3College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh EH16 4SB, UK
4Veterinary Medicines Directorate, Addlestone, Surrey KT15 3NB, UK

The soluble proteome of three *Clostridium difficile* strains of varying pathogenic potential, designated B-1, Tra 5/5 and 027 SM, were compared using differential in-gel electrophoresis in which the proteins of each strain were labelled with CyDyes. This enabled visual inspection of the 2D profiles of strains and identification of differentially expressed proteins using image analysis software. Unlabelled protein reference maps of the predominant proteins were then generated for each strain using 2D gel electrophoresis followed by protein sequencing of each spot using a Reflectron matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer. Increased coverage of the proteome was achieved using 1D gel electrophoresis in a bottom-up approach using LC-MS/MS of 1 cm gel slices. A total of 888 different proteins were detected by comparative analysis of isolates grown in parallel for 64 h on blood agar plates. Of these, only 38% were shared between all isolates. One hundred and ten proteins were identified as showing ≥2-fold difference in expression between strains. Differential expression was shown in a number of potential virulence and colonization factors. Toxin B was detected in the more virulent strains B-1 and 027 SM, but not in the lower virulent strain Tra 5/5, despite all strains possessing an intact pathogenicity locus. The S-layer protein (Cwp2) was identified in strains 027 SM and Tra 5/5 but not strain B-1, and differences in the post-translational modification of SlpA were noted for strain B-1. The variant S-layer profile of strain B-1 was confirmed by genomic comparison, which showed a 58 kb insertion in the S-layer operon of strain B-1. Differential post-translational modification events were also noted in flagellar proteins, thought to be due to differential glycosylation. This study highlights genomic and proteomic variation of different *Clostridium difficile* strains and suggests a number of factors may play a role in mediating the varying virulence of these different strains.

INTRODUCTION

*Clostridium difficile* is usually a nosocomially acquired pathogen, and the main cause of antibiotic-associated diarrhoea. Symptoms range from mild, self-limiting diarrhoea, through to severe pseudomembranous colitis (PMC) (Larson et al., 1978; Kelly & LaMont, 1998). The emergence of hyper-virulent PCR ribotype 027 strains in hospitals worldwide has been linked to the epidemic spread of infection (McDonald et al., 2005), causing significant burdens on healthcare institutions. Systematic reviews have estimated that the incremental costs associated with *Clostridium difficile* infection (CDI) range from £4577 to £8843 per patient across Europe (Wiegand et al., 2012) and US$4846 to US$8570 in the USA (Ghantoji et al., 2010).

Although several putative virulence factors have been identified, two major toxins produced by many *Clostridium difficile* strains are accepted as the major virulence factors. However, the exact role of each toxin (Lyras et al., 2009; Kuehne et al., 2010), mechanism of action (Chumbler et al., 2012) and potential differences in toxin regulation and amounts produced in vivo (Warny et al., 2005; Freeman et al., 2007; Dupuy et al., 2008; Denève et al., 2009; Vohra & Poxton, 2011) remain unclear. In addition, the large
variation in the toxinotypes of pathogenic *Clostridium difficile* strains (Rupnik, 2008) indicates that toxin production is only one aspect of pathogenicity and associated virulence.

The identity and roles of the other *Clostridium difficile* pathogenicity factors are less well understood. Adhesion may be a critical early step in host colonization, and this adhesion is likely to be mediated by cell surface proteins. A number of cell surface proteins have been identified, although the roles of individual proteins in co-ordination of adhesion of the bacterial cell to the gut wall, and the mechanism of this process, have not been elucidated. These proteins include adhesins, S-layer proteins, cell wall proteins and the flagella proteins as well as a number of S-layer protein paralogues (Wright *et al.*, 2005).

The genomes of different *Clostridium difficile* strains from different isolates spanning the last four decades have been sequenced in recent years. Among these are strain 630 (Sebaihia *et al.*, 2006), which has more recently been updated (Monot *et al.*, 2011), and two 027 isolates, a recent, epidemic isolate, R20291, and a ‘historic’ non-epidemic isolate, CD196 (Stabler *et al.*, 2009). These have respectively enabled genomic comparison studies of multiple isolates (Stabler *et al.*, 2006; He *et al.*, 2010; Dingle *et al.*, 2013), confirming the huge genomic variation of this pathogen. Epidemic PCR ribotype 027 strains have been shown to cluster into a tight clade (Stabler *et al.*, 2009). However, disease-associated isolates belong to multiple lineages, indicating that certain genetic elements may underlie virulence, and as a consequence of the highly dynamic nature of the genome, that these elements may be transferable by horizontal gene transfer and recombination (He *et al.*, 2010; Dingle *et al.*, 2013).

Although previously published extensive genomic comparisons have highlighted differences between strains of varying mosaic genome composition, they provide only limited insight into the factors causing differences in virulence, as the strains sequenced had not been well characterized at the biological level. However, two biologically well characterized strains have now been sequenced by the Department of Bioanalysis and Horizon Technologies (Public Health England, Colindale), the genome sequences of multiple isolates (Stabler *et al.*, 2006; He *et al.*, 2010; Dingle *et al.*, 2013), confirming the huge genomic variation of this pathogen. Epidemic PCR ribotype 027 strains have been shown to cluster into a tight clade (Stabler *et al.*, 2009). However, disease-associated isolates belong to multiple lineages, indicating that certain genetic elements may underlie virulence, and as a consequence of the highly dynamic nature of the genome, that these elements may be transferable by horizontal gene transfer and recombination (He *et al.*, 2010; Dingle *et al.*, 2013).

There is a paucity of proteomic data for *Clostridium difficile*. Studies have looked at cell surface proteins (Wright *et al.*, 2005), spore proteins (Lawley *et al.*, 2009) and the insoluble proteome (Jain *et al.*, 2010) of the 630 reference strain. Changes over time (Janvilisri *et al.*, 2012) and in response to antimicrobial peptides (McQuade *et al.*, 2012) have been reported for single strains. A comparative analysis of the *Clostridium difficile* secretome (Boetzkes *et al.*, 2012) has recently been reported; however, only one comparative analysis of proteomic differences between historic and recently emerged *Clostridium difficile* isolates has so far been described (Chen *et al.*, 2013), and this does not include a low-virulence strain. Crucially, detailed biological characterization data are rarely available in conjunction with genomic and proteomic analysis.

The purpose of this study was to combine biological characterization, proteomic analysis and genomic analysis of three strains of varying virulence in order to gain an insight into the factors which may play a role in variation in virulence. The detection of proteins associated with previously described biological characteristics, and the presence or absence of corresponding genes and genetic elements were investigated in detail.

**METHODS**

**Bacterial strains and growth conditions.** Characteristics of the strains are shown in Table 1.

Three *Clostridium difficile* isolates with varying biological and virulence properties were used in this study. The first strain, B-1 (PCR ribotype 005), was isolated from a patient with pseudomembranous colitis in 1978 (Borriello & Barclay, 1983). Comparative analysis of *Clostridium difficile* strains in the hamster model (Borriello *et al.*, 1987) has shown that it is highly virulent.

The second strain, Tra 5/5 (PCR ribotype 001), was isolated in 1982 from the faeces of a healthy infant (Larson *et al.*, 1982). It is of lower virulence (demonstrated by number of animals killed after inoculation) than strain B-1 (Borriello *et al.*, 1987).

The third strain, designated 027 SM (donated by M. Wilcox, Leeds), was isolated from the Stoke-Mandeville outbreak, associated with the emergence of the ‘hypervirulent’ ribotype 027 (Anonymous, 2005; Smith, 2005). This strain has been less extensively biologically characterized, but much recent work has focused on strains of this epidemic PCR ribotype.

All strains were grown on Columbia blood agar (CBA, Oxoid) at 37 °C. Anaerobic conditions were maintained using anaerobic sachets (Oxoid). Cells were harvested for protein extraction after 64 h, when optimum reproducible protein profiles were observed. Growth curves were determined in order to establish a timescale for optimum protein synthesis. Starter cultures in fastidious anaerobic broth (FAB, Oxoid) were prepared and used to inoculate 25 ml FAB to an optical density of 0.01–0.03 at 600 nm. Absorbance was read every 4 h to produce a growth curve. The three strains had comparable growth rates.

**Proteomic analysis**

**Preparation of whole-cell protein extract.** The growth from CBA plates was harvested after 64 h. Cells were resuspended in lysis buffer (8 M urea, 2% CHAPS, 40 mM Tris base) containing PMSF protease inhibitor (0.5–1 mM), or 50 mM phosphate buffer containing 0.5 mM PMSF for cell lysis and protein extraction. Glass beads were added to the cell suspension, and homogenization was performed on the FastPrep System (MP Biomedicals). Samples were shaken for two runs of 60 s at 4 m s⁻¹, and incubated on ice before centrifugation for 30 min at 21 000 g °C, to remove cell debris. The protein concentration of the lysates was determined using Bradford’s
Table 1. Known biological characteristics of strains B-1, Tra 5/5 and 027 SM, with presence/absence of related genes and corresponding proteins

Data were taken from Borriello & Barclay (1985), except for aBorriello et al. (1987), bWarny et al. (2005), cBoriello et al. (1988), dTasteyre et al. (2000); eGharbia, unpublished data, fStabler et al. (2009). Corresponding protein detection and presence of corresponding genes are indicated for each characteristic.

<table>
<thead>
<tr>
<th>Ribotype</th>
<th>Strain B-1</th>
<th>Strain Tra 5/5</th>
<th>Strain 027 SM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
<td>Protein</td>
<td>Phenotype 005</td>
<td>Protein 001</td>
</tr>
<tr>
<td>Toxin production</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Toxin A (TcdA)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Toxin B (maximum titre) (TcdB)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Surface proteins and mucosal adherence</td>
<td>+</td>
<td>Unknown</td>
<td>+</td>
</tr>
<tr>
<td>SLP (CD2793)</td>
<td>+</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>HMW SLP</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>LMW SLP</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cwp2 (CD2791)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cwp84 (CD2787)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cwp66 (CD2789)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Fimbriae</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Pilin (CD3513)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Flagella</td>
<td>+</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>FliC (CD0239)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Flagella hook</td>
<td>+</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>(CD0255)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Motility</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Chemotaxis protein (CD0538)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Chemotaxis protein (CD0256)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

*Identified in broth culture, data not shown.
**1D gel electrophoresis.** 1D gel electrophoresis was carried out using the NuPAGE gel system (Invitrogen). NuPAGE LDS sample buffer (4 μl) and NuPAGE sample reducing agent (1.6 μl) was added to 5 μg protein sample and made up to 16 μl with distilled water. Each sample was then loaded onto NuPAGE 10 % Bistris gels alongside 1 μg Protein Molecular Weight Standards (P-6649; Molecular Probes). Gels were run in NuPAGE MES SDS running buffer for 35 min, 200 V, 120 mA, 25 W. Gels were stained with SYPRO Ruby Protein Gel stain (Bio-Rad) overnight in the dark and were destained in 7 % (v/v) acetic acid, 10 % (v/v) methanol for 30 min before scanning by the Ettan DIGE (differential in-gel electrophoresis) imager (GE Healthcare). Gels were post-stained with Sigma Brilliant Blue G-colloidal (Sigma) according to the manufacturer’s protocol.

**MALDI-TOF mass spectrometry analysis.** Peptide solution (0.7 μl) was co-crystallized on the target plate with 0.7 μl matrix solution [10 mg ml⁻¹ χ-cyano-4-hydroxycinnamic acid in 49.5 % (v/v) acetonitrile, 49.5 % (v/v) ethanol, 0.001 % (v/v) TFA] (TFA). Peptide masses were measured on a MALDI-TOF Reflexion (Waters) equipped with a 337 nm nitrogen laser. Analysis was performed in positive ion mode. Peptide masses were collected over an m/z range of 800–3000 Da, and 1.8 μg renin ml⁻¹ in 0.1 % (v/v) TFA was used to lock mass the instrument. Raw data files were then compared with protein sequence data from the National Center for Biotechnology Information (NCBInr) database, and a non-redundant database comprising strains 027 SM, B-1,Tra 5/5 and 630, using the online Mascot Wizard software (http://www.matrixscience.com/wizard.html). One missed cleavage per peptide was allowed, and a mass tolerance of 150 p.p.m. was used. Carbamidomethylation of cysteine was specified as a fixed modification, and oxidation of methionine as a variable modification. A Mascot score greater than 61 was significant (P<0.05), but protein identifications with scores lower than 61 were considered positive if they also showed a minimum of six matched peptides or a sequence coverage of over 25 % as described previously (Encheva et al., 2009).

**2D gel electrophoresis.** After being treated with the GE 2D clean up kit (GE Healthcare) according to the manufacturer’s protocol, 50–100 μg protein was diluted to 100 μl with rehydration buffer [7 M urea, 2 M thiourea, 2 % (w/v) CHAPS, 0.5 % (v/v) IPG buffer, 0.002 % (v/v) bromophenol blue, 2.8 mg DTT ml⁻¹] and loaded onto hydrated Immobiline dry strips (18 cm, pH 4–7) for isoelectric focussing. These were run overnight before equilibration with DTT [6 M urea, 75 mM Tris/HCl pH 8.8, 29.3 % (v/v) glycerol, 2 % (w/v) SDS, 0.002 % (v/v) bromophenol blue, 65 mM DTT] for 15 min followed by iodoacetamide [6 M urea, 75 mM Tris/HCl pH 8.8, 29.3 % (v/v) glycerol, 2 % (w/v) SDS, 0.002 % (v/v) bromophenol blue, 135 mM iodoacetamide] for 15 min. The second dimension was run on 10 % acrylamide gels (22 × 24 cm × 1 mm) for 6 h. Gels were stained, scanned and post-stained as described above. A number of different pH ranges were trialled, but as almost all the visible spots were located in the pH 4–7 range, this range was used for the reference maps.

**2D fluorescence difference gel electrophoresis.** Protein samples were cleaned using the GE 2D clean up kit according to the manufacturer’s protocol. The pH was adjusted to pH 8.5 before minimal labelling with 400 pmol CyDye per 50 μg protein (CyDye DIGE Fluor). The samples were incubated in the dark for 30 min before the reaction was stopped with 10 mM lysine. The CyDye labelled protein samples were diluted with an equal volume of rehydration buffer [7 M urea, 2 M thiourea, 2 % (w/v) CHAPS, 0.5 % (v/v) IPG buffer, 0.002 % (v/v) bromophenol blue, 2.8 mg DTT ml⁻¹] before being pooled, and the pooled samples then run in the first and second dimensions as described above. Two biological replicates for each of the three strains were used, giving six samples in total. In order to obtain relative quantification of proteins, an internal standard of a mix of all six samples was run on each gel, alongside two individual samples. Gels were run for each strain against the two other strains, and repeated with the biological replicates to give a total of four profiles for each strain. Each biological replicate was labelled with both Cy3 and Cy5 during the experiment. Spot detection, spot alignment, and statistical analysis of the relative protein concentrations between strains was carried out using the Progenesis SameSpots software (Non-linear Dynamics). A mean normalized volume was produced for each protein spot across all gels to give a measure of relative abundance. The fold difference is the difference between the highest and lowest mean normalized volumes. Protein spots showing differential expression were matched to reference maps, and identified by MALDI-TOF MS as outlined below.

**In-gel digestion.** Spots from 2D gels or bands from 1D gels were excised and destained in 25 mM ammonium bicarbonate, in 50 % (v/v) methanol, until no blue colour remained. Gel plugs were dehydrated with two 10 min incubations in 100 % acetonitrile, and air-dried for 10 min, followed by reduction in 10 mM DTT (30 min at 60 °C) and alkylation in 55 mM iodoacetic acid (IAA), (45 min in the dark). Trypsin digestion was carried out at 37 °C overnight in 25 mM ammonium bicarbonate solution containing trypsin (10 ng μl⁻¹, 10 μl). Peptides were extracted in 0.05 % (v/v) trifluoroacetic acid (TFA), 50 % (v/v) acetonitrile (10 μl) for direct spotting onto MALDI plates for MALDI-TOF analysis, or 0.1 % (v/v) TFA for LC-MS analysis.

**LC-MS/MS mass spectrometry analysis.** Protein extracts from two biological replicates of each strain were separated by 1D PAGE (NuPAGE 12 % Bistris gel 1.0 mm, 12 wells; Invitrogen) and the peptide mixtures from tryptic digestion of gel bands from 1D PAGE were separated and analysed using an Ultimate 3000 Dionex nano capillary HPLC system (Dionex) coupled with a Thermo LTQ Orbitrap mass spectrometer (Thermo Scientific).

Trypsin-digested peptides (5 μl) were cleaned and desalted on a reversed phase trap column (PepMap C18, 300 μm internal diameter × 5 mm, 3 μm, 100 Å, Dionex), separated on a nano C18 analytical column (PepMapC18, 75 μm internal diameter × 15 cm, 3 μm, 100 Å, Dionex), and analysed by tandem mass spectrometry (MS/MS) on the LTQ Orbitrap mass spectrometer.

Settings for the LC-MS/MS were as follows: peptide mixtures were initially trapped on the reversed-phase trap column using isocratic of 100 % solvent A (2 % acetonitrile, 0.1 % formic acid in water) at a flow rate of 25 μl min⁻¹. Subsequent separation was performed on the analytical C18 nano column using a 45 min gradient of 5 to 45 % solvent B (90 % acetonitrile, 0.1 % formic acid in water) versus solvent A, then to 90 % B for an additional 5 min with a flow rate of 300 nl min⁻¹. The mass spectrometer was operated in a data-dependent mode to switch automatically between MS and MS/MS acquisition. The full survey scan (m/z 440–2000) was acquired in the Orbitrap with a resolution of 60000 at m/z 400, which was followed by six MS/MS scans in which the most abundant peptide precursor ions detected in the preceding survey scan were dynamically selected and subjected for collision-induced dissociation (CID) to generate MS/MS spectra, which were later searched against appropriate databases to assign identities to proteins.

The electrospray voltage was set at 1.5 kV; the ion source capillary voltage and temperature were set at 35 V and 200 °C, respectively. Tube lens was 105 V and the normalized collision energy was 35 % for MS/MS. Mass accuracy is a key parameter of mass spectrometric performance. In order to achieve enhanced mass accuracy, a background polydimethylcyclosiloxane ion with m/z=445.120052 was used as ‘lock mass’. Real-time recalibration on the lock mass by
corrections of mass shift removed mass error associated with calibration of the mass scale. A mass accuracy of 1–5 p.p.m. was routinely maintained during data acquisition.

The raw data files were matched against a theoretical trypic digest of the resulting proteins of a non-redundant Clostridium difficile database containing genomic data from strains 027 SM, B-1 and Tra 5/5 and the 630 reference strain using Bioworks 3.2 software (Thermo Fisher Scientific), with a tolerance of two missed cleavages. A peptide mass tolerance of 0.05 Da for the parent ion, and 0.50 Da for the fragment ions was specified, allowing for deamination and methylation of asparagine, hydroxylation of phenylalanine and proline, oxidation of methionine, acetylation of lysine, and carbamidomethylation and carboxymethylation of cysteine. The Sequest peptide threshold value was set at deltaCn scores of greater than 0, and XCorr scores of greater than 1.9, 2.2, 3.5 and 3.5 for singly, double, triply and quadruply charged peptides. The 12 srf files for each protein extract were merged and identification verified using Scaffold (Proteome software). A minimum of two peptides was required for protein identification, although identifications based on only one peptide were considered valid if the scaffold identification probability score was ≥95 %, and the same peptide had been detected in other samples. Any additional redundancy was removed by ensuring that where protein sequences differed slightly between strains, and so the same peptides matched to more than one protein in the database, only one protein was recorded.

**Genomic analysis**

**Genome sequencing.** The three strains were sequenced by Roche, using the Roche FLX platform and following the standard Roche sequencing protocols for 2 kb Pair-End and shotgun sequencing (Roche technical manual). The raw reads were deposited in EMBL-ENA, accession numbers ERP002519 (027 SM), ERP002520 (B-1) and ERP002521 (Tra 5/5). For annotations, the raw reads were assembled using the analysis package Roche Newbler v2.5; default settings were used.

**Genome comparison.** The complete genomes of each strain were compared using the MAUVE progressive algorithm (version 2.3.1) (default setting used), available from Genome Evolution Laboratory (http://gel.ahabs.wisc.edu/mauve/) (Darling et al., 2010).

**BLAST searching of protein sequences against the unfinished genomes of each strain.** The sequence of each identified protein was searched against each unfinished genome using the pBLAST protein BLAST search, with tab delimited outputs and an e-value threshold of 10^{-5}. Default settings were used for all other parameters. The percentage identification and e-values for each protein against each genome were recorded, and a percentage identification greater than 75 % was taken to indicate the presence of the corresponding gene in a strain. Protein sequences were also BLAST searched against the 630 reference strain genome.

**Cell cytotoxicity assay.** Strains were grown on Biafra’s CCEYL agar for 48 h before inoculation of single colonies into BHI broth. After anaerobic incubation at 37 °C for 48 h, 1 ml aliquots of culture were centrifuged at 16000 g and the supernatant removed. The supernatant was serially diluted in PBS to 10^{-7}. Twenty microlitres of each dilution was added to vero cell culture monolayers prepared in 96-well microtitre trays. Cells were then incubated at 37 °C in air with 5 % CO_{2} and examined at 24 and 48 h under an inverted microscope. A positive reaction was indicated by cell rounding. Specific action of Clostridium difficile cytotoxin was confirmed by parallel neutralization with Clostridium sordellii antitoxin. Toxin titres were expressed as relative units.

**RESULTS**

Genomic analysis of strains 027 SM, B-1 and Tra 5/5 by the Department of Bioanalysis and Horizon Technologies (Public Health England, Colindale) has shown a comparable number of ORFs in strains 027 SM (3896) and Tra 5/5 (3840) to those identified in the 630 strain, but a larger number of ORFs (4061) in strain B-1. Extra-chromosomal data have been generated for three plasmids in strain B-1, and one plasmid in strain Tra 5/5.

Strains were grown on CBA for 64 h prior to protein extraction. The growth rates of the three strains were shown to be comparable, and the majority of cells are expected to be in stationary phase at this point. Extracts were analysed by both 2D gel electrophoresis followed by MALDI-TOF MS and 1D gel electrophoresis followed by LC-MS/MS.

A total of 888 different proteins were detected by the combined proteomic methodologies. Protein identifications with MASCOT identification scores, percentage peptide coverage and number of matched peptides are shown in Tables S1 and S2 (available in the online Supplementary Material). A distribution of the types of proteins identified by 2D gel electrophoresis MALDI-TOF and 1D gel electrophoresis LC-MS/MS is shown in Table 2. Of the proteins identified, only 38 % were shared between all isolates, the majority of these being metabolic enzymes. Many of the proteins identified were detected in only one of the three strains, with 136 proteins (15 %) detected only in strain 027 SM, 162 (18 %) only in strain B-1 and 49 (6 %) only in strain Tra5/5. (Files S1, S2 and S3). In addition, proteomic methodology affected protein identification. One hundred and fifty proteins (17 % of proteins) were identified in strains B-1 and 027 SM, but not strain Tra 5/5 (of lower virulence in the hamster model) (Files S1, S2 and S3). Notably, these included toxin B, flagellar-related proteins, and four surface proteins. 1D gel electrophoresis followed by LC-MS/MS identified a far greater number of proteins than 2D gel electrophoresis followed by MALDI-TOF MS. However, a subset of nine proteins were identified by MALDI-TOF only.

DIGE analysis showed differential expression of proteins between the three strains (Figs 1, 2 and 3). Five different proteins were identified as significantly upregulated in strain 027 SM: Cwp2, SlpA, the ABC transporter substrate binding protein CD0873, GlyA and FliC. Ten different proteins were identified as upregulated in strain B-1: SlpA, pyruvate flavodoxin oxireductase CD2682, two ABC transporter substrate binding proteins (CD0855 and CD2672), two glycine reductase complex components (CD2349 and CD2354), AdhE, a carbon monoxide dehydrogenase (CD0174), ThiH and FliC. Eight different proteins were identified as upregulated in strain Tra 5/5: SlpA, pyruvate flavodoxin oxireductase (CD2682), a putative isomerase (CD0840), carbon monoxide dehydrogenase (CD0174), leucyl-tRNA synthetase (CD2521), Cwp2, acetyl CoA dehydrogenase (CD0399) and a

http://jmm.sgmjournals.org
The identification of many of these proteins as differentially expressed is likely to be due to differential migration within the gel, as discussed in more detail for individual proteins below.

The protein expression profiles, and presence or absence of corresponding genes were compared with existing biological characterization data (Table 1). Proteomic and genomic data relating to specific biological characteristics are outlined in detail below.

### Toxin production

Using LC-MS/MS, toxin A was detected in all three strains, whereas toxin B was detected in strains 027 SM and B-1 but not strain Tra 5/5 (File S2). Using Vero cell cytotoxicity assay, a mean toxin titre of 5 relative units (RU) was detected for strain B-1 and strain 027 SM, compared with 3.5 RU for strain Tra 5/5 (data not shown). Neither toxin cyclopropane-fatty-acyl-phospholipid synthase (CD0177).

![Fig. 1. A DIGE gel image of strain 027 SM (Cy5, red) compared to strain B-1 (Cy3, green). The Cy2 channel (blue) is the internal standard containing a mix of protein extracts from all strains. Yellow box (1) highlights an S-layer protein (CD2791, Cwp2) with a 43.7-fold greater concentration in strain 027 SM, and a normalized volume of 5.4 compared with 0.1 for strain B-1 and 0.5 for strain Tra 5/5. Yellow box (2) highlights an S-layer protein (CD2793, SlpA LMW protein) with a 19.1-fold greater concentration in strain 027 SM, and a normalized volume of 3.5 compared with 0.2 for strain B-1 and 0.2 for strain Tra 5/5. Yellow box (3) highlights the differences in the modified flagellin proteins of the B-1 and 027 SM strains, where the series of Cy3 and Cy5 labelled proteins are clearly visible as separate spots which have migrated differently within the gel. Yellow box (4) highlights the SlpA HMW protein, which shows differential migration between the strains.](image-url)
was detected on 2D reference maps or DIGE gels, probably due to the large size of these proteins.

The genomic data confirm that all three strains possess an intact pathogenicity locus (data not shown). Single nucleotide polymorphism (SNP) variations between strains are evident, primarily in the *tcdA* and *tcdB* genes. The PaLoc sequence of strain 027 SM is almost identical to that of the previously sequenced R20291 strain (data not shown), with the same 18 base deletion in the *tcdC* gene observed in strain 027 SM and R20291, but absent in the other strains.

**Surface proteins**

During this analysis, 20 cell surface or S-layer proteins were identified, only eight of which were detected in all three strains (Tables S1 and S2). These included SlpA, Cwp66 and Cwp84. Both SecA translocases (CD0143 and CD2792) were also detected in all strains during this analysis.

The S-layer locus of the genomes of 630, 027 SM, B-1 and Tra 5/5 were compared in detail (Fig. 4) using the online progressive MAUVE algorithm (Darling et al., 2010). The majority of ORFs were present in all strains, with the exception of *cwp2* and the adjacent small hypothetical ORF (CD2790), encoding a hypothetical protein of 235 aa. These ORFs were present in strains 027 SM, Tra 5/5 and 630, but appear to have been deleted from this locus in strain B-1. The Cwp2 protein was detected in strains 027 SM and Tra 5/5, showing highly conserved sequences by BLAST analysis, but Cwp2 was not detected in strain B-1 (Figs 1–3, box 1), confirming the genetic indication that it lacks the *cwp2* gene. Interestingly, the S-layer genetic locus in strain B-1 includes an insertion of approximately 58 kb between the ORFs of Cwp66 and Cwp84 (Fig. 4), which is not present in strains 027 SM and Tra 5/5.

SlpA protein was identified in all three strains, and showed variation in predicted pI and molecular mass values (predicted pI values of 4.83, 4.76 and 4.79, and predicted masses of 80 428, 64 703 and 79 922 Da for strains 027 SM, B-1 and Tra 5/5, respectively) (Tables S1 and S2). This predicted variation is confirmed by differential migration of this protein during 2D DIGE (Figs 1–3, box 4).

SlpA was present as two separate spots on the reference maps of strain 027 SM and Tra 5/5, corresponding to the
Fig. 4. MAUVE analysis comparing the organization of the SlpA operon for strains 027 SM, B-1, Tra5/5 and the 630 reference strain. Yellow bars represent areas of genetic homology between the strains, with the height of the bar representing the level of homology, and ORFs are designated by bars with arrows below. The SlpA gene is shown in red in all strains, and highlighted in green in strain 630. The gene encoding Cwp2 is shown in orange. Hypothetical or putative ORFs are shown in grey. Arrows above the genes show the direction of transcription. Blue arrows indicate the corresponding protein was detected in this analysis. CSP notates genes encoding cell surface proteins (Slp homologues). In strain B-1 this genetic locus shows considerable differences from the other three genomes, with an insertion of approximately 58 kb. This insertion contains 50 ORFs, the majority of them putative and uncharacterized.
high molecular mass (HMW) and low molecular mass (LMW) mature proteins. However, only one SlpA protein subunit was observed on the B-1 reference map (Figs 1–3, boxes 2 and 4).

Four putative cell surface proteins, Cwp5, Cwp6, Cwp22 and Cwp25 (CD2713, CD0844, CD2786, CD2184) were detected only in strains 027 SM and B-1. The remaining five cell surface proteins were detected in one strain only. Cwp10 (CD2796) only in strain B-1, Cwp24 (CD2193) only in Tra 5/5 and three proteins, Cwp19, Cwp12 and Cwp11 (CD3567, CD2794, CD2795), only in strain 027 SM.

The majority of these surface proteins detected showed >97% sequence identity between all strains by BLAST analysis (Table 3). Greater variation was observed for Cwp19 which showed 95.7% sequence identity to strains 027 SM and Tra 5/5, but only 65.8% identity to strain B-1. SlpA and Cwp66 showed greater sequence variability between strains (Table 3). The Cwp66 in silico proteins sequences for strain 027 SM and Tra 5/5 appear more similar to the 630 reference strain (~79% sequence identity) than that of strain B-1 (~59% sequence identity).

The major pilin protein (CD3050) was detected in strain Tra 5/5 but not B-1 or 027 SM, indicating that protein production may correlate with microscopic observations of fimbriae presence (Table 1). However no other fimbriae proteins were detected. The gene encoding pilin was present in the genomes of all three strains, despite the detection of the protein in strain Tra 5/5 only.

**Flagella and motility**

Flagellin (CD0239) was identified in all three strains with a high number of unique peptides and good sequence coverage. The flagellar hook protein (CD0255) was identified in strain B-1 and strain 027 SM, but not detected in strain Tra 5/5 (Tables S1 and S2). The corresponding genes to both the flagellin protein and the flagella hook protein were found to be present in the genomes of all three strains by BLAST analysis (Table 1), indicating that the lack of detection of the flagella hook protein in strain Tra 5/5 is not due to lack of the corresponding gene. Multiple distinct spots following two-dimensional gel electrophoresis were identified as flagellin. The pattern and location of these spots is suggestive of post-translational modifications. Glycosylation of FliC in *Clostridium difficile* has been described (Twine et al., 2009), so it is likely this modification is a glycosylation event. DIGE analysis showed that migration of these protein spots is different for each strain (Figs 1–3, box 3), indicating that the post-translational events may differ among strains. A putative flagellar glycan biosynthesis locus was compared in detail using the progressive MAUVE algorithm, and found to be conserved between the three strains (data not shown).

Two chemotaxis proteins were detected during this analysis: CD0538, which was detected in strains B-1 and 027 SM, but not in strain Tra 5/5, and MotA (CD0256), which was detected in strain B-1 only (Tables S1 and S2). The corresponding genes were both found to be absent in strain Tra 5/5 by BLAST analysis (Table 1), and motA was absent in strain 027 SM, corresponding to the protein expression profile. Other chemotaxis proteins, such as CheW and CheA, were not detected.

**DISCUSSION**

The choice of two of the three *Clostridium difficile* strains used in this study (B-1 and Tra 5/5) was based on previous characterization work indicating that the strains exhibit varying biological and virulence properties (Table 1). The third strain was chosen as an example of an epidemic ribotype 027 strain. The genome of *Clostridium difficile* has undergone extensive characterization in recent years, with a number of complete genomes available (Sebaihia et al., 2006; Stabler et al., 2009; Monot et al., 2011), and full genome comparisons undertaken of a large number of strains (Stabler et al., 2006; He et al., 2010; Dingle et al., 2013). The initial sequencing of the three biologically well-characterized strains used in this study provides preliminary genomic information. These genomic data have been supplemented by the characterization of the proteome of whole-cell protein extracts from each of the three strains.

**Table 3. BLAST analysis of surface proteins**

Homologies (%) of in silico protein sequences of strains 027 SM, B-1 and Tra 5/5 and the 630 reference strain for those surface proteins showing sequence variation. The SlpA S-layer protein, showing the greatest variation, is shown in bold.

<table>
<thead>
<tr>
<th>630 protein ID*</th>
<th>Strain B-1</th>
<th>ID (%)</th>
<th>Strain 027 SM</th>
<th>ID (%)</th>
<th>Strain Tra5/5</th>
<th>ID (%)</th>
<th>Detected in strains:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cwp66 CD2789</td>
<td>BASYS03070</td>
<td>58.9</td>
<td>BASYS03265</td>
<td>79.5</td>
<td>BASYS02894</td>
<td>79.2</td>
<td>027 SM, B-1 Tra5/5</td>
</tr>
<tr>
<td>Cwp8 CD2799</td>
<td>BASYS03079</td>
<td>99.9</td>
<td>BASYS03276</td>
<td>99.5</td>
<td>BASYS02904</td>
<td>99.4</td>
<td>027 SM, Tra5/5</td>
</tr>
<tr>
<td>Cwp2 CD2791</td>
<td>BASYS03079</td>
<td>36.5</td>
<td>BASYS03267</td>
<td>97.9</td>
<td>BASYS02896</td>
<td>99.5</td>
<td>027 SM, Tra5/5</td>
</tr>
<tr>
<td>SlpA CD2793</td>
<td>BASYS03072</td>
<td>52.9</td>
<td>BASYS03269</td>
<td>55.4</td>
<td>BASYS02898</td>
<td>55.7</td>
<td>027 SM, B-1, Tra5/5</td>
</tr>
<tr>
<td>Cwp12 CD2794</td>
<td>BASYS03073</td>
<td>65.8</td>
<td>BASYS03271</td>
<td>95.7</td>
<td>BASYS02899</td>
<td>95.7</td>
<td>027 SM</td>
</tr>
</tbody>
</table>

*Sebaihia et al. (2006).
In order to achieve the best possible coverage of the proteomes of the three strains, a number of different methodologies were used. 2D reference mapping was used to provide a visual comparison of the protein profiles of the three strains, and DIGE analysis was used to provide relative quantification of protein expression between them. In addition, 1D gel electrophoresis coupled with LC-MS/MS, a more sensitive protein detection platform, was used to increase the protein coverage by improved detection of low abundance proteins. In total, 16.5%, 16.7% and 11.6% of the predicted ORFs from strains 027 SM, B-1 and Tra 5/5, respectively, have been confirmed in this study by protein identification, a similar coverage to other reported *Clostridium difficile* proteomic studies (Janvilisri et al., 2012; Chen et al., 2013).

Three aspects of previously reported biological characterization are discussed below with respect to the genomic and proteomic information reported here, to give some insight into factors potentially mediating the varying virulence of these strains.

**Toxin production**

The two toxins, toxin A and toxin B, remain the best characterized and understood virulence factors for *Clostridium difficile*, and toxin production is required for symptomatic disease. Comparative analysis of *Clostridium difficile* strains in the hamster model of disease (Borriello et al., 1987) has shown that strain B-1 is highly virulent, with challenged animals dead or moribund within 48 h. Extracellular toxin A levels were over 195 ng ml$^{-1}$ in Bactotryptose broth, while the titre of toxin B in the same batches exceeded 1/500. Strain Tra 5/5 is of lower virulence (demonstrated by number of animals dying after challenge with *Clostridium difficile*), with a significantly lower *in vitro* toxin A concentration than for strain B-1 (~25 ng ml$^{-1}$), but the same maximum cytotoxic titre (toxin B) as strain B-1 (Borriello et al., 1987). We have confirmed by Vero cell cytotoxicity assay that strain B-1 produces >10-fold greater levels of toxin *in vitro* than strain Tra 5/5; with a mean toxin titre of 5 RU compared with 3.5 RU. However, this methodology does not distinguish between activity of toxin A and toxin B.

The virulence of strain 027 SM has not been tested in the hamster model, but many studies have characterized epidemic, PCR ribotype 027 outbreak strains, and the sequenced R20291 strain (Stabler et al., 2009) was isolated during the same outbreak as strain 027 SM. It has been reported (Warny et al., 2005) that PCR ribotype 027 strains produce higher levels of toxin compared with other ribotypes (toxin A, 848 µg l$^{-1}$, toxin B 180 µg l$^{-1}$), although this has been disputed, with PCR ribotype 027 strains shown not to produce greater levels of toxin, but to produce toxin for a prolonged period of time within an *in vitro* gut model (Freeman et al., 2007). We found strain 027 SM to produce a toxin titre of 5 RU *in vitro*, a comparable level to strain B-1, and greater than strain Tra 5/5.

Strains Tra 5/5 and B-1 have previously been shown to produce comparable amounts of toxin B *in vitro* (Borriello et al., 1987), contrasting with the lack of toxin B detected here during protein analysis. This discrepancy may be due to differences in growth conditions of cultures used for protein extraction here (64 h of growth on blood agar) and by Borriello *et al.* (broth culture), or may be due to sensitivity and specificity differences in toxin detection methodologies. The relative roles played by toxin A and toxin B in virulence of *Clostridium difficile* are disputed, and toxin B gene knockout studies have been contradictory (Lyras et al., 2009; Kuehne et al., 2010), as has been reviewed in detail (Carter et al., 2012). However, toxin B has been shown to disrupt epithelial integrity in the host cells (Hecht et al., 1992) and function as a potent enterotoxin in its own right (Savidge et al., 2003). Furthermore, toxin B-positive A-negative strains have been shown to produce fluid accumulation in rabbit ligated small bowel loops and disease in the hamster animal model (Borriello et al., 1992), and be responsible for nosocomial outbreaks of *Clostridium difficile*-associated diarrhoea (Alfa et al., 2000). The apparent lower levels of toxin B produced by strain Tra 5/5 may contribute to the reported lower virulence of this strain, although the molecular reasons for reduced toxin production remain unclear, as all three strains possess an intact pathogenicity locus (PaLoc), suggesting that all strains possess the potential to produce both toxins A and B.

The 18 base deletion in *tcdC* identified in strain 027 SM has been widely reported in ribotype 027 strains (Barbut et al., 2007) and, as *tcdC* has been reported to be a negative regulator of toxin production, it has been suggested that this deletion may lead to increased toxin production by ribotype 027 strains (Carter et al., 2011). However, the role of *tcdC* as a negative regulator of toxin production has recently been disputed (Curtman et al., 2012). The work presented here supports the hypothesis that the mutation of *tcdC* is not an important factor causing increases in toxin production, as strain B-1 does not contain this deletion, yet produces comparable levels of toxin *in vitro*. Strains B-1 and Tra 5/5 were genetically very similar in the PaLoc region, yet showed ~10-fold difference in toxin levels produced, suggesting that toxin production may be additionally controlled by factors outside of the PaLoc region.

**Surface proteins and mucosal adherence**

Adhesion is a critical early step in host colonization, and a number of cell surface proteins co-ordinating adhesion of the bacterial cell to the gut wall have been identified, although the full mechanism has not been elucidated. Vegetative *Clostridium difficile* cells have been shown to adhere to Caco-2 cells (Cerquetti et al., 2002), and a number of *Clostridium difficile* proteins have been described as adhesins, including Cwp66 (Waligora et al., 2001), surface layer (S-layer) proteins (Calabi et al., 2002a), cell wall proteins (Wright et al., 2005), flagella (Tasteye et al., 2001)
and a number of S-layer protein paralogues (Wright et al., 2005). In addition, adherence of Clostridium difficile spores to Caco-2 cells has recently been investigated (Paredes-Sabja & Sarker, 2012), although the adhesins responsible have not been identified. The role adhesion plays in virulence is still not fully understood. Borriello et al. (1988) demonstrated that different Clostridium difficile strains of varying virulence exhibit corresponding variation in their ability to colonize and adhere to the gastrointestinal tract in the hamster model, and showed that strain B-1 has a high level of adherence. Waligora et al. (1999) found no difference in the adherence of toxigenic and nontoxigenic strains, and suggested that earlier observed variation was in mucosal association, rather than true cell adherence (Waligora et al., 1999).

S-layers have been shown to be essential for virulence in some bacterial pathogens, including Aeromonas salmonicida and Campylobacter fetus (Sára & Sleytr, 2000). Most Clostridium difficile strains express a HMW protein and a LMW protein, which are derived from a single gene, slpA (Calabi et al., 2001), and cleaved post-translationally. These two S-layer proteins play a role in the binding of Clostridium difficile to host cells (Calabi et al., 2002a), and induce inflammatory and regulatory cytokines in host cells (Ausielo et al., 2006), suggesting roles in adherence and interaction with the host. S-layer proteins exhibit a high degree of variation between different Clostridium difficile strains in the LMW region, but show greater conservation in the HMW region, which shows cross-immunoreactivity (Calabi et al., 2001; McCoubrey & Paxton, 2001).

The Clostridium difficile genome contains a number of slpA homologues, a nomenclature for which was proposed by Fagan et al. (2011). Calabi et al. (2001) identified 29 ORFs showing homology to slpA. Many of these ORFs were clustered near slpA, in a locus containing 17 ORFs, 11 of which encode proteins containing domains homologous to the postulated cell wall anchoring domain of SlpA (Karjalainen et al., 2001). The structure and organization of this locus suggests that these genes could be co-transcribed and form an operon (Savariau-Lacomme et al., 2003). Here we report that the organization of this operon is very similar in strains 027 SM and Tra 5/5, but strain B-1 shows genetic reorganization in this operon, with an insertion of approximately 58 kb. This insertion has replaced the Cwp2 ORF, but does not seem to block the transcription of downstream ORFs, with expression being confirmed by protein detection in many cases (SecA2 and Cwp84 were detected in all strains). A similar insertion into this genetic locus in a number of strains has recently been described by Dingle et al. (2013), which has been shown to compose a putative glycosylation locus. It is possible that strains containing this insertion have altered SlpA expression profiles, as indicated by the markedly different S-layer protein expression profile of strain B-1 described here, and if the putative glycosylation locus is functional and expressed, may also have altered glycosylation capacities. The slpA genes of different ribotype 027 strains have been reported to be highly conserved, and to show homology to those of ribotype 001 strains (Spigaglia et al., 2011). This is supported here, where the strain 027 SM and the strain Tra 5/5 (PCR ribotype 001) show similar S-layer profiles, both genetically and by protein detection and post-translational modification, whereas strain B-1 (PCR ribotype 005) differs. The majority of detected surface proteins showed high homology (>97%) between all strains by BLAST analysis. However, greater variation was observed for Cwp19, SlpA and Cwp66, all of which exhibited greater similarity between strains 027 SM and Tra 5/5 than to strain B-1. Differences in S-layer expression were also noted in strain B-1 compared with strains Tra 5/5 and 027 SM. Strain B-1 does not express Cwp2, or the LMW SlpA protein. The absence of a LMW protein indicates that post-translational modification steps may be different within this strain, or that the LMW SLP is unstable or rapidly degraded. A similar variation in the SLP protein has been seen by Calabi and Fairweather who identified a strain, designated 167, which showed an unusual S-layer 1D profile (Calabi & Fairweather, 2002b). A predominant protein of 39 kDa representing the HMW SLP was detected, but only minimal amounts of the predicted, 20 kDa LMW protein were detected. The uncleaved pre-protein of strain 167 had a predicted molecular mass of 62 312 Da, lower than that of other isolates, which is similar to the predicted mass for strain B-1 (64 703 Da). The authors proposed that the lack of LMW SLP could be due to rapid degradation of this subunit, and showed that this did not affect the pathogenicity or cellular viability of the strain. A different S-layer lattice system may be present in this, and the B-1 strain, which may be associated with the 58 kb genetic insertion.

Other well characterized components of this genetic locus include Cwp66 (CD2789) and Cwp84 (CD2787), both of which were detected in all three strains here. Cwp66 has been shown to act as an adhesin in heat-shocked cells (Waligora et al., 2001) and is thought to undergo similar cleavage events to SlpA. Cwp84 is located immediately downstream of Cwp66 and encodes a putative cysteine protease that mediates the post-translational cleavage SlpA (Kirby et al., 2009; Dang et al., 2010). Cleavage of SlpA may not be required for virulence, as Cwp84 genetic knockout mutant strains, expressing only the uncleaved SlpA precursor, were still capable of causing disease in the hamster model (Kirby et al., 2009). However, the finding that trypsin can also cleave SlpA suggests that trypsin-mediated cleavage of SlpA may occur in the absence of Cwp84 in vivo, thus SlpA processing may still be a key step in the disease causing mechanism. The slpA locus also contains the secA2 translocase gene (detected in all strains here), encoding an essential component of the general secretory pathway, the route by which SlpA is exported (Fagan & Fairweather, 2011).

The S-layer profiles of these three strains do not obviously correspond with virulence, as the two highly virulent strains, B-1 and 027 SM show marked differences in S-layer
expression profiles, whereas the lower virulence Tra 5/5 strain shows high homology with the 027 SM strain. These results correlate with many published analyses (Spigaglia et al., 2011; Dingle et al., 2013), and suggest that the relationship between expression of S-layer proteins and virulence is highly complex.

Strain B-1 shows no evidence of the presence of fimbriae, thought to play a role in adhesion, whereas strain Tra 5/5 does (Borriello et al., 1988). The pilin protein (CD3050), a component of fimbriae, was detected in strain Tra 5/5 but not B-1 or 027 SM, although the corresponding gene was present in all strains, confirming the microscopic observations of fimbriae presence. This further indicates that the relationship between strain adherence and virulence may not be a simple one.

Flagella and motility

Flagella have been shown to contribute to bacterial virulence in a number of ways (Ramos et al., 2004). They confer motility, allowing access of bacteria to mucosal tissues, and as with other cell surface structures, they can act as adhesins. Crude preparations of Clostridium difficile flagella have been shown to associate with caecal mucus of mice, and aflagellate strains of Clostridium difficile are reported to associate with caecal tissue at a rate that is 10-fold slower than that of flagellate strains (Tasteyre et al., 2001). In addition, it has been suggested that flagella can play a role in sensing environmental conditions such as ‘wetness’ (Wang et al., 2003) and hence may control expression of genes and virulence factors under certain conditions (Boin et al., 2004).

Strain B-1 has shown evidence of flagella and motility (Tasteyre et al., 2000), and epidemic 027 ribotype strains have been shown to be motile (Stabler et al., 2009), although the presence of flagella was not confirmed microscopically. Strain Tra 5/5 had not previously been tested for motility, but was found to be negative (S. E. Gharbia, data not shown). Previous comparative genomic assays have indicated that flagella production and motility may not be a common feature for Clostridium difficile isolates, and that flagella gene loci are not necessarily conserved among virulent strains (Stabler et al., 2006), although they appear to be conserved for the three strains investigated here. It has been reported by reverse transcriptase (RT)-PCR that flagellin is expressed in both flagellate and aflagellate strains (Tasteyre et al., 2000), so it is possible that strain Tra 5/5 is aflagellate, despite the detection of flagellin. This would correlate with the observed lack of motility of this strain, and the lack of detection of the flagellar hook protein. Dingle et al. have shown disruption of flagellin genes leads to lack of flagella expression and motility, but does not affect adherence to Caco-2 cells, or virulence in the hamster model (as determined by % survival) (Dingle et al., 2011), suggesting that if Tra 5/5 is aflagellate, this would not explain an apparent reduced virulence. However, recent work by Aubry et al. indicates that some flagellar mutants exhibit reduced toxin production, indicating that the flagellar regulon modulates toxin production (Aubry et al., 2012). This raises the interesting possibility that lack of flagella detection in strain Tra 5/5 may be linked to the lack of toxin detection.

The pattern and location of flagellin protein in two-dimensional gel electrophoresis is suggestive of post-translational modifications, and shows variation between strains. Similar changes in flagellum protein migration have been seen in Campylobacter jejuni, with proteins from a putative flagellar glycosyltransferase knockout mutant migrating at a less acidic pI within the gel than those with an active glycosyltransferase (van Alphen et al., 2008). Glycosylation of flagella has been shown to be important in many bacterial species, affecting motility, adhesion and pathogenesis (Logan, 2006). Differences in the flagellar-associated glycosyltransferases have been shown between epidemic 027 ribotypes and non-epidemic strains (Stabler et al., 2006), and Twine et al. (2009) have confirmed that flagellin from Clostridium difficile is linked to glycans, which show divergence between different strains. If, as seems probable, the difference in flagellin migration is due to differential glycosylation, this could affect structure and function of the flagella, altering the role played in motility and adhesion.

Twine et al. (2009) reported variation in a putative flagellar glycan biosynthesis locus between the structural fliC and flgB genes. They identified one putative flagellar glycosyltransferase downstream of fliC in strain 630, but three in other clinical isolates. This locus was examined for strains 027 SM, B-1 and Tra 5/5 by MAUVE and compared to 630 (data not shown). As reported by Twine et al. (2009) strain 630 contains one putative glycosyltransferase, however strains 027 SM, B-1 and Tra 5/5 contain the three glycosyltransferases reported in other isolates. This genetic locus appears conserved between strains 027 SM, B-1 and Tra 5/5, suggesting that the postulated difference in flagellar glycosylation is not due to genomic variation in this region.

Another component to bacterial motility is chemotaxis. It has previously been shown that gut mucus of animals and humans acts as a chemotactrant for Clostridium difficile (Boriello and Bhatt, 1995), making chemotaxin proteins potential virulence factors for this organism, as have been shown for other organisms such as Vibrio cholera (Boin et al., 2004). Chemotaxis of the strains investigated here have not been investigated, although Borriello and Barclay have noted evidence of chemotaxis by strain B-1 (unpublished data). Two chemotaxis-related genes and the corresponding proteins were detected in strain B-1, the flagellar motor rotation protein MotA (CD0256) and the chemotaxis receptor gene (CD0538). Strain 027 SM lacks the motA gene (and corresponding protein). The lack of detection of both motA and the chemotaxis receptor gene (CD0538), and corresponding proteins, in strain Tra 5/5, may also help explain the reduced virulence of this strain.
CONCLUSION
The combination of genomic, proteomic and biochemical data allows in depth investigation of potential virulence factors, which may help to elucidate the molecular reason for the varying virulence of Clostridium difficile strains. For the few potential virulence factors discussed here, a number of strain-to-strain differences have been identified (e.g. difference in toxin and flagellar expression) which could contribute to the varying virulence of these three isolates. Further analysis may highlight additional strain-to-strain variations of interest, as well as providing useful comparison data with previously sequenced strains. A detailed comparison of the insert identified here in strain B-1 with the inserts described by Dingle et al. (2013) would provide a valuable analysis into evolutionary changes over time. However, some strain to strain variations discussed here (e.g. similarity of strain B-1 and strain Tra 5/5 PaLoc, and similarity of strain 027 SM and Tra 5/5 S/A expression profile) do not correlate with the observed virulence of these three strains, further confirming that the pathogenicity of this organism is not a simple process, and that many potential virulence factors, although not required to cause disease, may contribute to disease severity.

The differences highlighted here merit further investigation, and give indications of potential targets for gene knock out studies. In addition, this study has highlighted the importance of different proteomic methodologies. The utilization of both MALDI-TOF and LC-MS/MS in this study has increased the coverage of the Clostridium difficile proteome, demonstrating the importance of using a combination of types of analysis in order to gain a more complete understanding of the full proteome of this organism, and how it varies between different strains. This could be an important consideration for proteomic characterization of other microorganisms.

ACKNOWLEDGEMENTS
This work was funded by Public Health England (formerly Health Protection Agency).

REFERENCE


Clostridium difficile has a highly mobile, mosaic genome. Nat Genet 38, 779–786.


