Plasticity of fimbrial genotype and serotype within populations of *Bordetella pertussis*: analysis by paired flow cytometry and genome sequencing

Thomas E. Vaughan,1 Catherine B. Pratt,1 Katie Sealey,2,3 Andrew Preston,2 Norman K. Fry3 and Andrew R. Gorringe1

1Public Health England, Porton Down, Salisbury SP4 0JG, UK
2Department of Biology & Biochemistry, University of Clifton, Clifton Down, Bath BA2 7AY, UK
3Respiratory and Vaccine Preventable Bacteria Reference Unit, Public Health England – Microbiology Reference Services, Colindale, 61 Colindale Avenue, London NW9 5EQ, UK

The fimbriae of *Bordetella pertussis* are required for colonization of the human respiratory tract. Two serologically distinct fimbrial subunits, Fim2 and Fim3, considered important vaccine components for many years, are included in the Sanofi Pasteur 5-component acellular pertussis vaccine, and the World Health Organization recommends the inclusion of strains expressing both fimbrial serotypes in whole-cell pertussis vaccines. Each of the fimbrial major subunit genes, *fim2*, *fim3*, and *fimX*, has a promoter poly(C) tract upstream of its −10 box. Such monotonic DNA elements are susceptible to changes in length via slipped-strand mispairing in vitro and in vivo, which potentially causes on/off switching of genes at every cell division. Here, we have described intra-culture variability in poly(C) tract lengths and the resulting fimbrial phenotypes in 22 recent UK *B. pertussis* isolates. Owing to the highly plastic nature of fimbrial promoters, we used the same cultures for both genome sequencing and flow cytometry. Individual cultures of *B. pertussis* contained multiple fimbrial serotypes and multiple different fimbrial promoter poly(C) tract lengths, which supports earlier serological evidence that *B. pertussis* expresses both serotypes during infection.

**INTRODUCTION**

*Bordetella pertussis* causes whooping cough, a continuing public health problem despite high vaccine coverage. Rates of pertussis disease have increased in many countries over recent years, thought to be due to a combination of strain evolution (Mooi et al., 2013) and faster waning immunity provided by acellular pertussis (aP) vaccines, compared to previously used whole-cell pertussis vaccines (wP) (Tartof et al., 2013). In addition, evidence from a baboon model of pertussis indicates that acellular vaccines are less effective than whole-cell vaccines at preventing acquisition and transmission of the organism (Warfel et al., 2014). To facilitate pathogenesis of pertussis disease, *B. pertussis* adheres to the mucosal surface in the upper respiratory tract via a number of adhesins, including filamentous haemagglutinin, pertactin and fimbriae (Mattoo & Cherry, 2005). *B. pertussis* expresses two serologically distinct fimbriae composed of either Fim2 or Fim3 subunits (22.5 and 22.0 kDa, respectively). Subtypes of Fim2 (*fim2-1* and *fim2-2*) and Fim3 (*fim3-1*, *fim3-2* and *fim3-3*) have also been described (Packard et al., 2004; Tsang et al., 2004). Both Fim2 and Fim3 subunits are assembled in a coil to produce long filaments on the surface of the bacteria (Heck et al., 1996). Fimbriae (co-purified Fim2/3) have been determined as protective antigens in mouse models of pertussis (Heininger et al., 1998; Robinson et al., 1989a), and are components of an acellular vaccine shown to have high efficacy in phase III clinical trials (Gustafsson et al., 1996; Olin, 1997). Household contact studies performed during these trials showed a correlation between antibody responses to fimbriae and protection against disease (Cherry et al., 1998; Storsaeter et al., 1998).

Expression of many *B. pertussis* virulence-related genes, including *fim2* and *fim3*, is regulated by the BvgA/S two-component system (Decker et al., 2012). Transcription of *fim3* is proposed to occur when a dimer of phosphorylated BvgA is bound to the −35 region of the *fim3* promoter at the same time as region 4 of the RNA polymerase σ80 subunit and C-terminal domain of an RNA polymerase z subunit (Decker et al., 2011). Poly(C) tracts and homopolymeric tracts are generally under-represented in the *B. pertussis* chromosome (Coenye & Vandamme, 2005), and the possibility of rapid antigenic variation via slipped-strand mispairing at these sites

**Abbreviation:** wP, whole-cell pertussis vaccine.

One supplementary figure and two supplementary tables are available with the online version of this paper.
has received limited research attention. However, each *fim* gene possesses a poly(C) tract between its −35 and −10 promoter regions. Transcription-permissive lengths for the fimbrial poly(C) tracts have been determined (Willems *et al.*, 1990; Chen *et al.*, 2010), but the frequency of slipped-strand mispairing has not been previously described. The fimbrial poly(C) tracts of *B. pertussis* have a notable counterpart in the porA promoter of *Neisseria meningitidis*, where altered transcription levels caused by changes in the length of a poly(G) tract between the −35 and −10 regions facilitate evasion of bactericidal anti-PorA antibodies (Tauseef *et al.*, 2013).

In this study, analysis of whole genome sequencing reads from a panel of 96 *B. pertussis* strains isolated in the UK between 1920 and 2012 revealed multiple subpopulations with differing fimbrial poly(C) tract lengths in *B. pertussis* cultures. In addition, we preserved aliquots from 22 of the cultures used for genome sequencing and analysed Fim2 and Fim3 expression in individual bacteria using two-colour flow cytometry with labelled anti-Fim2 and anti-Fim3 monoclonal antibodies. Our findings demonstrate that bacteria expressing Fim2 and/or Fim3 are present in cultures previously shown by standard slide agglutination methods to express only one fimbrial serotype. Accordingly, we conclude that expression of Fim2 and Fim3 by *B. pertussis* is more flexible than generally acknowledged. Moreover, this flexibility may help to explain the rapid shifts in the predominant serotype observed in many countries.

**METHODS**

**Bacteria.** Ninety-six *B. pertussis* isolates (UK001 to UK06, UK008 to UK011, UK014 to UK018, UK020 to UK100) were obtained from the Respiratory and Vaccine Preventable Bacteria Reference Unit (RVPRBU), Public Health England (PHE) Colindale. The majority of these isolates (*n*=87) were submitted by hospital laboratories as part of the enhanced surveillance of pertussis (Van Buynder *et al.*, 1999). Additional isolates were previously obtained from the Wellcome Bacterial Collection held by the National Collection of Type Cultures (NCTC) (*n*=4), the NCTC (*n*=2) and the former Centre for Applied Microbiology and Research, now PHE Porton Down (*n*=3) (Fry *et al.*, 2001).

Following culture for 48 to 72 h on blood charcoal agar (PHE Media Services), *B. pertussis* isolates were serotyped using a slide agglutination assay with rabbit polyclonal antisera against antigens 1 (89/596), 2 (89/598), and 3 (89/600) (National Institute for Biological Standards and Controls). Antigen 1 (agglutininogen 1) is expressed by all isolates of *B. pertussis*. Thus, Fim−, Fim2, Fim3 and Fim2,3 (Hallander *et al.*, 2005) are used instead of the historical notations serotype 1, serotype 1,2, serotype 1,3 and serotype 1,2,3 (Robinson *et al.*, 1989b), as described previously (Litt *et al.*, 2009). The terms Fim− and Fim2−Fim3− are used interchangeably in the text. Cultures for this study were grown on charcoal agar for 72 h at 37 °C. For flow cytometry analysis, bacteria were resuspended in PBS containing 1% formaldehyde and OD600 adjusted to 2.0.

**Genome sequencing.** Genome sequencing of UK isolates UK001 to UK100 was performed at the Sanger Institute. Detailed analysis of these sequences will be presented in a future publication (manuscript in preparation). Briefly, paired-end libraries were prepared from genomic DNA, amplified with a Kapa HiFi library amplification kit (Kapa Biosystems), and sequenced as 150 bp paired-end reads using a MiSeq desktop sequencer (Illumina). The unassembled short reads are available in fastq.gz format from the Sanger Institute ftp site. Locations of the files are presented in Table S1 (available in the online Supplementary Material). Additional genome sequence files (non-UK isolates) are listed in Table S2.

**Extraction of poly(C) tract spanning reads from fastq.gz files.** Each fastq.gz file was read with *kseq_test*, a minimal example program for the kseq.h fastq parser library (Li, 2011; Li, 2012). Output from *kseq_test* was piped into a script written in AWK (Aho *et al.*, 1988), within which the sequence field of each Illumina read was put through two successive filters.

The first filter compared each read to the set of all plausible fimbrial poly(C) tracts (6–20 base pairs) flanked by a minimal *B. pertussis* Tohama I-derived context (Parkhill *et al.*, 2003). The search patterns used in this first filter are shown as regular expressions in Table 1. The fim2 regular expressions contain a [CT] character list corresponding to deoxythymidine that punctuates the poly(C) tract of *B. pertussis* Tohama. Three regular expressions are necessary to prevent fim3 and fimX sequences from being misidentified as each other. Reverse-complement patterns, not shown in Table 1, were also used, generating a total of 308 patterns upon evaluation of regular expression. Non-matching reads (including all reads ending within poly(C) tracts) were discarded, and exact matches tested in the second filter.

The second filter took the segments of each read outside the initial match, joined them and aligned the resulting chimaera to the equivalent joined 30 bp segments from *B. pertussis* Tohama 1. For this alignment, EMBoss needle (Rice *et al.*, 2000) was used via the AWK getline function. Local alignment was forced with respect to query and global alignment with respect to reference (by invoking needle with an endweight, a zero end/open penalty and high gap open/extend penalties). Reads were given a positive plausibility score if this second filter resulted in a plausible alignment (aligned join-sites, ≥66.7% identity across ≥10 bp where flanking sequence was available on both sides of the initial match, ≥80% identity across ≥20 bp where flanking sequence was available on only one side of the initial match) and otherwise given a negative score. Each fastq record was reassembled into a one-line tab-delimited format and augmented

<table>
<thead>
<tr>
<th>Table 1. Patterns used in the first of two sequential text filters to identify genome sequence reads spanning the promoter poly(C) tracts of fim2, fim3 and fimX</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene</strong></td>
</tr>
<tr>
<td>fim2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>fim3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>fimX</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The patterns are printed in 1D regular expression syntax (Thompson & Ritchie, 1971). The reverse complement of each pattern was also used.
with gene name (fim2, fim3 or fimX), isolate name, fastq.gz file name and plausibility score.

The kseq_test preprocessing step and awk/needle double filter were combined as a single Bourne shell script, and summary counts extracted from the tab-delimited output using a second shell script (supplementary materials: MeasurePolyC.sh and FastplusToTable.sh). The awk script embedded in MeasurePolyC.sh induced a memory leak in GNU awk, but not in awk or nawk. Arithmetic means and standard deviations (Table 3 and supplementary material, CoverageStatsUKstrains.awk) were calculated using the Welford–Knuth algorithm (Welford, 1962; Knuth, 1997).

Detection of Fim2 and Fim3 using flow cytometry. A 200 μl aliquot of each suspension (OD600 0.1) of plate-grown bacteria was added to a well of a 96-well microtitre plate. The plate was centrifuged at 3060 g for 5 min and supernatant removed. Fim2- (06/124) and Fim3-specific (06/128) monoclonal antibodies (obtained from the National Institute of Biological Standards and Control) bound to Alexa Fluor 488 and Alexa Fluor 647, respectively, using a Zenon® Alexa Fluor® Mouse IgG1 Labelling kit (Life Technologies). A 100 μl aliquot of each antibody at a dilution of 1 in 5000 was added to each test well, and pellets resuspended. The plate was incubated at 25 °C with shaking for 30 min. Next, the plate was centrifuged as above and washed, and pellets resuspended in 200 μl blocking buffer (PBS, 1 % BSA). Samples were analysed using a Beckman Coulter CyanADP flow cytometer, and Summit® software used to calculate the median fluorescence of each fluorophore for each sample. Unstained bacterial cells were used to verify the absence of fluorophore-independent counts, which are minimized by the use of filter-sterilized buffers and sheath fluid and by the collection of 10,000 events per sample.

A gate was drawn to include the bacteria plus conjugate control (no antibody), resulting in median fluorescence >20 being considered positive for Fim expression. To ascertain the percentage of bacterial cells expressing Fim2 or Fim3, Alexa Fluor 488 fluorescence was plotted against Alexa Fluor 647 fluorescence for each sample. Gates were drawn to enclose the negative control sample and used to demarcate between a positive and negative result. This gate was subsequently applied to all test samples, allowing determination of the percentage of events or individual bacteria that were not fluorescent, singly positive for Alexa Fluor 488, singly positive for Alexa Fluor 647 or doubly positive for both Alexa Fluor 488 and Alexa Fluor 647, thereby establishing the percentage of individual bacteria expressing Fim2, Fim3 or both Fim2 and Fim3 within a population.

RESULTS

Multiple fimbrial promoter variants within cultures

From the 96 unassembled UK *B. pertussis* genome sequences, 2648 reads were matched exactly to one of the fim2 regular expressions (Table 1) or one of the reverse complements. After utilizing a second filter to eliminate false positives and uncertain matches, 2580 fim2 poly(C) tract spanning reads were retained. The twice-filtered reads contained fim2 poly(C) tracts ranging in length from 6 to 20 bp. Multiple fim2 poly(C) tract lengths were found within samples (Table 2), with a mean intra-isolate range of 4.2 bp (standard deviation, 2.3 bp). The deoxystyminidine residue that punctuates the fim2 poly(C) tract of *B. pertussis* Tohama was also found in UK001, UK002 and UK004 (58 %, 17 % and 23 % of reads, respectively), and counted as an internal deoxyctydine for the purpose of poly(C) tract length enumeration. Our numbering convention differs from that of Chen et al. (2010), who examined only T-punctuated fim2 poly(C) tracts and whose enumeration excluded deoxystyminidine on the 5’ side of the punctuating deoxythymidine.

Using the equivalent double filter for fim3, 1914 reads were matched exactly by the first filter, and 1660 fim3 poly(C) tract spanning reads retained after the second filter. The twice-filtered reads contained fim3 poly(C) tracts ranging in length from 6 to 20 bp. As with fim2, multiple poly(C) tract lengths were found within samples (Table 2), with a mean intra-isolate range of 5.2 bp (standard deviation, 3.0 bp).

For fimX, 4463 reads were matched by the first filter and 4381 fimX poly(C) tract spanning reads retained after the second filter. The fimX poly(C) tracts were less variable than those of the other two fimbrial genes (mean intra-isolate range of 1.1 bp, standard deviation of 0.4). In 77 of the 96 genome-sequenced UK isolates, fimX poly(C) tracts did not exceed 7 bp in length. Isolate UK071 contained a fimX poly(C) tract of 9 bp, the longest among the UK isolates.

Read lengths for UK025 to UK100 were all 150 bp (average of 1.26 million reads per isolate; standard deviation, 0.26 million), whereas those for other available *B. pertussis* Illumina sequences were 54 bp (average of 5.40 million reads per isolate for UK strains; standard deviation, 1.66 million). For fim2 and fimX, poly(C) tract spanning reads were more numerous in the genomes sequenced at 54 bp per read whereas for fim3, poly(C) tract spanning reads were more numerous in genomes sequenced at 150 bp per read (Table 3). Since a 54 bp read is less likely than a 150 bp read to span a given sequence, the higher tally of fim2 and fimX poly(C) tracts in the 54 bp genome sequences can be attributed to higher coverage of those genomes (71.4-fold versus 46.1-fold for the 150 bp genome sequences). For fim3, the higher tally in the 150 bp genome sequences can be attributed to the broader size range of poly(C) tracts (rangefim3 5.2 bp versus rangefim2 4.2 bp), with the longer part of this range more likely to be encompassed by longer reads.

After normalization to genome coverage, the yields of fim2 and fim3 poly(C) tract spanning reads were higher from the 150 bp than the 54 bp genomes, but the yield for fimX stayed higher for the 54 bp genomes (Table 3, column 8).

Presence of multiple serotypes within cultures

An aliquot of culture prepared for genome sequencing from strains UK025 to UK046 was preserved by the addition of formaldehyde to 1 %. Individual bacteria in preserved suspensions were analysed for expression of Fim2 and Fim3 using anti-Fim2 and anti-Fim3 monoclonal antibodies and two-colour flow cytometry. The flow cytometry scatter plots obtained and poly(C) tract lengths determined for each isolate are shown in Fig. 1. In each of the 22 isolates examined
Table 2. Promoter poly(C) tract lengths and surface expression of Fim2 and Fim3 in *B. pertussis* isolates UK025 to UK046

ENA identifiers are European Nucleotide Archive sample names. Serotype by agglutination used the nomenclature of Hallander *et al.* (2005). Each isolate was cultured twice, once for serotyping by agglutination and once for the other two analyses, with cultures being split between flow cytometry and whole genome sequencing. Values (apart from fluorescence) for the other 74 UK isolates are presented in Table S3.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Year of isolation</th>
<th>Serotype by agglutination</th>
<th>ENA identifier</th>
<th>No. of reads</th>
<th>Median poly(C) tract length</th>
<th>Range of poly(C) tract lengths</th>
<th>Median fluorescence</th>
<th>No. of reads</th>
<th>Median poly(C) tract length</th>
<th>Range of poly(C) tract lengths</th>
<th>Median fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK025</td>
<td>2008</td>
<td>Fim3</td>
<td>ERS176852</td>
<td>33</td>
<td>10</td>
<td>3 (9 to 11)</td>
<td>4.6</td>
<td>8</td>
<td>14</td>
<td>6 (13 to 18)</td>
<td>320.2</td>
</tr>
<tr>
<td>UK026</td>
<td>2008</td>
<td>Fim3</td>
<td>ERS176853</td>
<td>28</td>
<td>10</td>
<td>2 (10 to 11)</td>
<td>4.3</td>
<td>12</td>
<td>14</td>
<td>3 (13 to 15)</td>
<td>251.4</td>
</tr>
<tr>
<td>UK027</td>
<td>2008</td>
<td>Fim3</td>
<td>ERS176854</td>
<td>24</td>
<td>10</td>
<td>3 (9 to 11)</td>
<td>4.6</td>
<td>15</td>
<td>14</td>
<td>4 (12 to 15)</td>
<td>353.3</td>
</tr>
<tr>
<td>UK028</td>
<td>2009</td>
<td>Fim3</td>
<td>ERS176855</td>
<td>27</td>
<td>10</td>
<td>3 (9 to 11)</td>
<td>4.6</td>
<td>10</td>
<td>14</td>
<td>4 (12 to 15)</td>
<td>360.7</td>
</tr>
<tr>
<td>UK029</td>
<td>2009</td>
<td>Fim2</td>
<td>ERS176856</td>
<td>6</td>
<td>16</td>
<td>4 (14 to 17)</td>
<td>708.8</td>
<td>20</td>
<td>13</td>
<td>5 (11 to 15)</td>
<td>12.8</td>
</tr>
<tr>
<td>UK030</td>
<td>2011</td>
<td>Fim3</td>
<td>ERS176857</td>
<td>4</td>
<td>16</td>
<td>3 (15 to 17)</td>
<td>703.3</td>
<td>26</td>
<td>10</td>
<td>3 (9 to 11)</td>
<td>8.1</td>
</tr>
<tr>
<td>UK031</td>
<td>2011</td>
<td>Fim2</td>
<td>ERS176858</td>
<td>9</td>
<td>15</td>
<td>4 (13 to 16)</td>
<td>666.5</td>
<td>24</td>
<td>10</td>
<td>3 (9 to 11)</td>
<td>8.1</td>
</tr>
<tr>
<td>UK032</td>
<td>2011</td>
<td>Fim3</td>
<td>ERS176859</td>
<td>39</td>
<td>11</td>
<td>3 (10 to 12)</td>
<td>4.6</td>
<td>22</td>
<td>14</td>
<td>8 (12 to 19)</td>
<td>209.8</td>
</tr>
<tr>
<td>UK033</td>
<td>2011</td>
<td>Fim3</td>
<td>ERS176860</td>
<td>30</td>
<td>10</td>
<td>2 (9 to 10)</td>
<td>4.6</td>
<td>5</td>
<td>14</td>
<td>2 (14 to 15)</td>
<td>246.7</td>
</tr>
<tr>
<td>UK034</td>
<td>2011</td>
<td>Fim2</td>
<td>ERS176861</td>
<td>6</td>
<td>15</td>
<td>4 (13 to 16)</td>
<td>608.7</td>
<td>20</td>
<td>11</td>
<td>3 (10 to 12)</td>
<td>9.3</td>
</tr>
<tr>
<td>UK035</td>
<td>2012</td>
<td>Fim2</td>
<td>ERS176862</td>
<td>11</td>
<td>15</td>
<td>7 (11 to 17)</td>
<td>361</td>
<td>21</td>
<td>12</td>
<td>3 (11 to 13)</td>
<td>12</td>
</tr>
<tr>
<td>UK036</td>
<td>2012</td>
<td>Fim3</td>
<td>ERS176863</td>
<td>31</td>
<td>10</td>
<td>3 (9 to 11)</td>
<td>4.1</td>
<td>6</td>
<td>14</td>
<td>2 (13 to 14)</td>
<td>666.5</td>
</tr>
<tr>
<td>UK037</td>
<td>2012</td>
<td>Fim2</td>
<td>ERS176864</td>
<td>12</td>
<td>14</td>
<td>6 (11 to 16)</td>
<td>255.8</td>
<td>38</td>
<td>10</td>
<td>9 (9 to 11)</td>
<td>9.9</td>
</tr>
<tr>
<td>UK038</td>
<td>2012</td>
<td>Fim3</td>
<td>ERS176865</td>
<td>30</td>
<td>10</td>
<td>3 (9 to 11)</td>
<td>4.9</td>
<td>17</td>
<td>14</td>
<td>8 (12 to 19)</td>
<td>394.6</td>
</tr>
<tr>
<td>UK039</td>
<td>2012</td>
<td>Fim2</td>
<td>ERS176866</td>
<td>28</td>
<td>10</td>
<td>2 (9 to 10)</td>
<td>4.6</td>
<td>11</td>
<td>14</td>
<td>9 (12 to 20)</td>
<td>341.5</td>
</tr>
<tr>
<td>UK040</td>
<td>2012</td>
<td>Fim2</td>
<td>ERS176867</td>
<td>11</td>
<td>15</td>
<td>7 (13 to 19)</td>
<td>270</td>
<td>37</td>
<td>12</td>
<td>8 (11 to 18)</td>
<td>9.9</td>
</tr>
<tr>
<td>UK041</td>
<td>2012</td>
<td>Fim2</td>
<td>ERS176868</td>
<td>5</td>
<td>15</td>
<td>5 (13 to 17)</td>
<td>270.2</td>
<td>27</td>
<td>12</td>
<td>5 (11 to 15)</td>
<td>9.6</td>
</tr>
<tr>
<td>UK042</td>
<td>2012</td>
<td>Fim3</td>
<td>ERS176869</td>
<td>28</td>
<td>10</td>
<td>2 (10 to 11)</td>
<td>4.7</td>
<td>6</td>
<td>14</td>
<td>3 (12 to 14)</td>
<td>178.3</td>
</tr>
<tr>
<td>UK043</td>
<td>2012</td>
<td>Fim3</td>
<td>ERS176870</td>
<td>11</td>
<td>15</td>
<td>8 (9 to 16)</td>
<td>4.6</td>
<td>11</td>
<td>14</td>
<td>3 (13 to 15)</td>
<td>15.3</td>
</tr>
<tr>
<td>UK044</td>
<td>2012</td>
<td>Fim3</td>
<td>ERS176871</td>
<td>36</td>
<td>10</td>
<td>3 (9 to 11)</td>
<td>4.5</td>
<td>6</td>
<td>13.5</td>
<td>8 (11 to 18)</td>
<td>291</td>
</tr>
<tr>
<td>UK045</td>
<td>2012</td>
<td>Fim3</td>
<td>ERS176872</td>
<td>35</td>
<td>10</td>
<td>3 (10 to 12)</td>
<td>4.2</td>
<td>5</td>
<td>14</td>
<td>3 (13 to 15)</td>
<td>312.8</td>
</tr>
<tr>
<td>UK046</td>
<td>2012</td>
<td>Fim3</td>
<td>ERS176873</td>
<td>43</td>
<td>10</td>
<td>4 (9 to 12)</td>
<td>4</td>
<td>9</td>
<td>14</td>
<td>6 (12 to 17)</td>
<td>191.9</td>
</tr>
</tbody>
</table>
using two-colour flow cytometry, *B. pertussis* cells of all four possible fimbrial serotypes (Fim2⁺Fim3⁺, Fim2⁺Fim3⁻, Fim2⁻Fim3⁻ and Fim2⁻Fim3⁺) were present (Fig. 1, scatterplots; Fig. 2). One fimbrial serotype was dominant (>55% of the population) in 21 out of the 22 isolates (Fig. 2). In isolate UK043 with bvg⁻ colony morphology, Fim2⁻Fim3⁻ cells were dominant (89% of the population). FimX surface expression was not investigated owing to the lack of available antibodies.

Gates were set so that the bacteria plus conjugate control was 99% enclosed inside the Fim2⁻Fim3⁺ quadrant. Although this may misidentify some events as Fim2⁻Fim3⁻ (for instance the lower portion of the Fim3⁺ cloud of UK042 in Fig. 1), it ensures that false positives are avoided. The diagonal streaking observed in the Fim2⁺Fim3⁺ quadrant of some scatterplots (Fig. 1) is attributed to clumping of bacteria, with single bacteria towards the lower right and clumped bacteria towards the upper right. This analysis used a gate that excluded most clumped bacteria, but when a larger gate was used to include more clumped bacteria, the fraction of Fim2⁻ or Fim3⁻ bacteria remained very similar (data not plotted). Hence, the results presented here are not dependent on the chosen flow cytometry analysis gate.

In all 22 isolates, fim2 and fim3 promoter poly(C) tracts were present in the Illumina sequencing data as multiple variants. The broadest range for fim2 occurred in UK086, with poly(C) tract lengths ranging from 7 to 17 bp. The broadest range for fim3 occurred in UK057, with poly(C) tract lengths ranging from 7 to 19 bp.

### Relationship between median fimbrial poly(C) tract length and serotype

Among the 21 bvg⁺ isolates studied using paired genome sequencing and two-colour flow cytometry, Fim2 median fluorescence was high in eight phenotypically bvg⁻ isolates with median fim2 poly(C) tracts lengths of 12 to 16 bp (Table 2). In the other 13 bvg⁻ isolates tested in this paired analysis, Fim2 median fluorescence was at least 13-fold lower, and median poly(C) tract lengths were either 10 or 11 bp. Median fim2 poly(C) tract lengths greater than 16 bp were not observed in the 22 isolates tested using flow cytometry, but observed in four of the other 74 UK isolates.

Fim3 median fluorescence was high in 13 phenotypically bvg⁺ isolates with median fim3 poly(C) tract lengths of 13 or 14 bp. In the other 8 bvg⁻ isolates tested, Fim3 median fluorescence was at least 11-fold lower, and median fim3 poly(C) tract lengths were in the range of 10 to 13 bp. Median fim3 poly(C) tract lengths longer than 14 bp were not seen in the 22 isolates tested using flow cytometry, but observed in four of the other 74 UK isolates.

In UK043, fim2 and fim3 median poly(C) tract lengths were 15 and 14 bp, respectively, but median fluorescence was low for both Fim2 and Fim3. Over 70% of cells were identified as Fim2⁻Fim3⁻ (Fig. 2). This isolate displayed bvg⁻ colony morphology, and decoupling of surface expression
from poly(C) tract length was therefore expected. However, the isolate was serotyped as Fim3⁺ via agglutination.

Among the other 74 sequenced UK isolates (not analysed by flow cytometry), UK015 provided the sole example of a homogeneous fim2 promoter (38 Illumina reads spanning the fim2 poly(C) tract, all with poly(C) tract lengths of 9 nt). The range of fim2 poly(C) tract lengths in UK010 was not determined, since only one spanning read was
available. There were no unambiguous examples of homogeneous \textit{fim3} promoters (only one spanning read was available for UK017).

For Fim2 and Fim3, the relationships between promoter \textit{poly(C)} tract lengths and expression are therefore as follows: a \textit{bvg} + population is Fim2 + if the median \textit{poly(C)} tract length is \(>11\) bp (upper bound not established, but at least 16 bp), and a \textit{bvg} + population is Fim3 + if the median \textit{poly(C)} tract length is \(>13\) bp (upper bound not established, but at least 14 bp).
Optimal poly(C) tract lengths for *B. pertussis* fimbrial promoter activity have been proposed from observations with luciferase reporter gene constructs (Chen et al., 2010). In 19 of the 21 *bvg*+ isolates described here, surface expression of Fim2 or Fim3 correlated with the majority of *fim2* or *fim3* promoters having poly(C) tracts unambiguously in the proposed active range (Fig. 1, shaded in bar charts). However, this correlation was not observed for two
isolates. In UK037, the majority of \textit{fim2} poly(C) tracts were outside the proposed optimal range, but the majority of cells in this culture were Fim2\textsuperscript{+}. Similarly, in UK044, the majority of \textit{fim3} poly(C) tracts were outside the proposed optimal range, while the majority of cells were Fim3\textsuperscript{+}.
Relationship between poly(C) tract length and serotype determined via agglutination

Serotypes determined via agglutination were not uniformly in agreement with poly(C) tract length distributions. Serotypes of 13 among the 96 genome-sequenced isolates (including UK043) determined via agglutination conflicted with the rules proposed above connecting median poly(C) tract length to expression (Table 4), and if upper bounds for ‘active’ median poly(C) tract lengths could be established, additional conflicting examples would probably be apparent. For 59 of the 96 isolates, serotypes by agglutination were in agreement with median poly(C) tract lengths.

Serotyping by agglutination was consistent with the flow cytometric measurements of Fim2 and Fim3 for 20 of the 22 isolates (Table 2). The exceptions were UK030 (Fim2− Fim3+ with agglutination but 88 % Fim2− Fim3− with flow cytometry) and the bvg− isolate UK043. Four of the 96 genome-sequenced UK isolates were serotyped as Fim2,3, and one as Fim−, but these five were not included in flow cytometric analysis (Table S3: UK001, UK003, UK006, UK010 and UK068).

Subpopulations with the potential to express FimX

In 87 of the 96 UK isolates, all fimX poly(C) tracts were 8 bp or shorter, suggesting that these strains are FimX−.

The longest fimX poly(C) tract found in a UK isolate was 9 bp (UK024), which is shorter than the proposed optimum length for expression (Chen et al., 2010). Since no reagents are available to detect FimX protein, the possible existence of FimX+ subpopulations can be tested only by scanning genome sequences for fimX poly(C) tracts of plausible transcription-permissive lengths. A further 275 unassembled B. pertussis genome sequences from non-UK isolates were therefore filtered through the same script used for the 96 UK isolates. In 195 non-UK isolates, the longest observed fimX poly(C) tract length was 7 bp, and the longest fimX poly(C) tract lengths observed in the whole dataset were 10 bp (found in two isolates), which is shorter than the proposed optimum length for expression.

DISCUSSION

Fimbriae are required by B. pertussis for colonization of the human respiratory tract. We have shown using two-colour flow cytometry that individual cultures contain cells of more than one fimbrial serotype. In the same cultures, we detected multiple variants of promoters for each serologically distinct fimbrial major subunit. This study therefore provides cellular and genomic corroboration of the suggestion by Heikkinen et al. (2008) and recent serological evidence that
Fig. 2. Relative abundance of four fimbrial serotype permutations in individual cultures of 22 UK isolates of *B. pertussis* analysed by two-colour flow cytometry.
Table 4. *B. pertussis* isolates with serotypes determined by agglutination conflicting with those predicted from fim2 and fim3 median poly(C) tract length

Observed, by agglutination; Predicted, from median poly(C) tract lengths. Serotype nomenclature is as described by Hallander *et al.* (2005). Discrepancies in fimbrial serotypes between those ‘observed’ and ‘predicted’ are presented in bold type.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Serotype</th>
<th>Median poly(C) tract length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td>UK001</td>
<td>Fim−</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK003</td>
<td>Fim2,3</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK005</td>
<td>Fim3</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK010</td>
<td>Fim2,3</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK011</td>
<td>Fim3</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK015</td>
<td>Fim3</td>
<td>Fim−</td>
</tr>
<tr>
<td>UK030</td>
<td>Fim3</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK043</td>
<td>Fim3</td>
<td>Fim2,3</td>
</tr>
<tr>
<td>UK051</td>
<td>Fim3</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK061</td>
<td>Fim3</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK064</td>
<td>Fim3</td>
<td>Fim2,3</td>
</tr>
<tr>
<td>UK072</td>
<td>Fim3</td>
<td>Fim2</td>
</tr>
<tr>
<td>UK081</td>
<td>Fim3</td>
<td>Fim2</td>
</tr>
</tbody>
</table>

*B. pertussis* expresses both Fim2 and Fim3 during infection, irrespective of the predominant serotype of case isolates (Hallander *et al.*, 2014).

The intra-culture serotype heterogeneity described here reflects the intrinsic susceptibility of repetitive DNA elements (in this case, promoter poly(C) tracts) to slipped-strand mispairing during DNA replication. The steric feasibility of bulge defects in the base-pairing of nucleic acids was demonstrated more than 50 years ago, and the formation of such defects by slipped-strand mispairing was postulated immediately as a mechanism underlying point mutations in DNA regions of low complexity (Fresco & Alberts, 1960). Insertions and deletions within runs of short (1 to 10 bp) repeated elements have subsequently been reported in human genetic disorders and antigenic switching by pathogens (van der Woude & Bäumler, 2004). Sporadic Huntingdon’s disease arises from expansion of a run of sequential CAG glutamine codons in the HTT gene (Myers *et al.*, 1993), and similar CAG expansions are evident in several other neurodegenerative disorders. Analogous in-frame trinucleotide repeats (AGT or TCT) occur in the MgPa adhesion operon of *Mycoplasma genitalium*, and variations in the number of repeat units have been observed within and between isolates (Ma *et al.*, 2012). The opacity genes of *Neisseria gonorrhoeae* are subject to reversible frameshifts at a rate of $10^{-4}$–$10^{-5}$ per cell/generation, resulting from changes in the number of copies of a CTCTT repeat in the signal peptide-encoding sequence (Murphy *et al.*, 1989). Variation in the number of copies of a TAAA repeat in the nadA promoter of *Neisseria meningitidis* MC58 leads to colonies with differential NadA expression appearing on the same culture plate (Martin *et al.*, 2003).

The simplest possible repeat elements, homopolymeric tracts, account for many reported instances of slipped-strand mispairing in bacteria. There are 5684 homopolymeric tracts longer than 4 bp in the *B. pertussis* Tohama I genome sequence (Parkhill *et al.*, 2003), among which 78 % are 5 bp in length. The 6 bp homopolymeric tracts account for 15 % of the total, and frequency decreases with increasing length. In a survey of expected versus observed homopolymeric tract frequencies in 139 sequenced bacterial genomes, *B. pertussis* was ranked twelfth lowest with an expected/observed ratio of 0.23 (Coenye & Vandamme, 2005), and most reports of homopolymeric tract-associated antigenic variation are from bacteria with higher expected/observed ratios (for instance, 2.1 in *Neisseria meningitidis* and 3.5 in *Helicobacter pylori*). It is therefore striking that the earliest report of *B. pertussis* slipped-strand mispairing was in the bvgS gene where the impact is magnified by pleiotropy. Stibitz *et al.* (1989) found that bvg− strains arise by frameshift expansion of a C6 tract at bvgS codons 1103–1104, and the frequency of the bvg+- switch varied between two strains.

Transcription from the fimbrial promoters requires that their poly(C) tracts are longer (Chen *et al.*, 2010) than the footprints of some DNA polymerases (Beese *et al.*, 1993; Eom *et al.*, 1996), so the heterogeneity of fimbrial promoters reported here is perhaps unsurprising. Fazekas *et al.* (2010) were able to improve the fidelity of PCR amplification across homopolymeric tracts using PCR enzymes with extended DNA-binding footprints. Compared to a 68 °C PCR extension, DNA replication in *B. pertussis* should be less vulnerable to slipped-strand mispairing because the 3’ end of the nascent strand is less likely to dissociate at physiological temperatures. Nevertheless, the imperfect fidelity of C6 replication in bvgS, along with the apparent selection against homopolymeric tracts throughout the genome, suggests that *B. pertussis* is unable to replicate homopolymeric tracts with high fidelity if longer than 5 bp. Fimbrial poly(C) tracts of
B. pertussis, being several nucleotides longer, are directly analogous to the poly(C) tract of opc (Sarkari et al., 1994) and the poly(G) tract of porA (van der Ende et al., 1995; Tauseef et al., 2013), both in N. meningitidis. In each of these examples, expression of a surface antigen is modulated by frequent and apparently stochastic changes in the spacing between the ~35 and ~10 promoter regions.

In an organism that uses slipped-strand mispairing to induce antigenic variation, cell division must, at least some of the time, result in daughter cells inheriting surface antigens that they are unable to express anew. The correlation between fimbrial poly(C) tract length distribution and surface expression is therefore likely to be imperfect, especially if antigenic variation is rapid. This is exemplified in the present study by the two isolates with poly(C) tract lengths predominantly outside the transcription-permissive range estimated by Chen et al. (2010); though an alternative explanation might be that estimation of transcription-permissive lengths is compromised by slipped-strand mispairing in the reporter gene constructs used. In this study, we obtained statistical estimates of transcription-permissive fimbrial poly(C) tract lengths, as distinct from the previous reporter gene-based estimates, but for some isolates poly(C) tract length did not agree with serotype determined by agglutination assay (Table 4). However, the objective of the present study was not to determine precise limits for promoter activity, and the range of poly(C) tract lengths in individual cultures was sufficiently broad to encompass the multiple serotypes seen in the same cultures (Fig. 1).

The major impediment to estimating in vivo poly(C) tract diversity lies in the potential of laboratory DNA polymerases to add to the diversity in vitro. In our hands, B. pertussis fimbrial poly(C) tracts could be sequenced successfully using dideoxy chemistry if amplified in Escherichia coli as plasmid inserts, implying correct amplification by the E. coli DNA polymerase. However, when PCR-amplified fimbrial promoters are sequenced without cloning, multiple superimposed traces appear downstream of the poly(C) tract, signifying multiple poly(C) tract lengths, some of which may represent slipped-stranded mispairing in the PCR tube rather than in the organism. In the present study, we sought to avoid PCR-generated diversity using high coverage sequence data. The 10-cycle amplification in the Kapa HiFi library amplification protocol amounts to 10 000-fold less amplification than takes place in a 30-cycle PCR reaction, and therefore, the number of in vitro-generated poly(C) tract lengths is likely to be minimal. In addition, only a few of the poly(C) tract length distributions in Fig. 1 resemble the bell-shaped distribution expected for pure PCR-generated diversity.

The distribution of B. pertussis fimbrial serotypes, determined via agglutination, varies between countries and over time, and may be influenced by differing vaccination policies (reviewed by Poolman & Hallander, 2007). Our data indicate that an agglutination assay highlights the serotype of the largest subpopulation in the sample under assay. Thus, the fact that temporal and geographical trends can be discerned at all suggests that B. pertussis populations are stably diverse, with selective pressure being required to displace a dominant serotype. Detection limits for current serotyping assays are not known, but could be determined with blended samples following separate flow cytometric analyses. The proportion of UK isolates with serotype Fim2+ (via agglutination) dropped significantly from 47% (n=110) in 1988 to 2001 to 3% (n=184) in 2002 to 2004, and 2% (n=166) in 2005 to 2008, after the introduction of aP vaccines (a preschool booster in 2001 and a replacement for the previous wP administered at two, three and four months in 2004) (Litt et al., 2009), but increased in 2011–2012 (Alexander et al., 2012) to levels comparable with the pre-vaccine and whole-cell vaccine eras. It is possible that population immunity to the predominant serotype builds up until a tipping point is reached and expression of the alternative serotype is favoured. However, there is no clear pattern in the median poly(C) tract lengths over these periods, partly because of a sampling bias in favour of 2012 isolates (Fig. S1).

FimX is 73% similar to Fim2 and 70% similar to Fim3, but has been detected only with mass spectrometry and in one B. pertussis isolate (Tefon et al., 2011). No laboratory antibodies are available for more rapid detection, and the capacity of FimX to substitute for Fim2 and Fim3 is unknown. Most of the fimX poly(C) tracts counted in the present study were 6 to 8 bp long, which is probably too short to allow transcriptional initiation. The emergence of a FimX+ strain from one with exclusively short fimX poly(C) tracts might conceivably start with a low-probability DNA replication error, followed by increasingly probable recurrences as the poly(C) tract grew further beyond the size of the DNA polymerase footprint in successive generations. The analogous phenomenon in human CAG expansion disorders, where successive generations are at increased risk as their CAG repeats expand, is termed genetic anticipation (Walker, 2007). The 10 bp fimX poly(C) tracts found in two strains are comparable in length to some of the fim2 and fim3 poly(C) tracts (Fig. 1), but robust investigation of FimX genetic anticipation requires more data than were available in the genome sequences (Table 3, column 4), and would be better accomplished by generating specific antibodies for use in flow cytometry.

Thus, in this study we have characterized the distribution of poly(C) tract lengths in promoters of fim2 and fim3 and shown that cultures contain multiple subpopulations with differing poly(C) lengths. For a subset of strains, we determined expression of Fim2 and Fim3 in each bacterium using two-colour flow cytometry and showed that while a predominant serotype exists in each culture, multiple combinations of Fim2 and Fim3 expression occur in a single culture. The optimal poly(C) tract lengths for expression of Fim2 and Fim3 were greater than 11 or 13 bp respectively for most strains, but for 13 out of 96 strains, the observed serotype conflicted with that predicted based on median poly(C) tract length (Table 4). This finding
indicates that expression of Fim2 and Fim3 is more fluid than previously thought and confirms serological evidence that *B. pertussis* expresses both Fim2 and Fim3 in *vivo*. It is therefore important for *B. pertussis* vaccines that rely on protection mediated by fimbriae to elicit antibodies against both fimbrial antigens.

**ACKNOWLEDGEMENTS**

Work at PHE, Porton Down has been funded by the National Institute of Health Research Centre for Health Protection Research. K.S. is funded by a PhD Studentship from PHE.

**REFERENCES**


Edited by: P. Langford