K19 capsular polysaccharide of Acinetobacter baumannii is produced via a Wzy polymerase encoded in a small genomic island rather than the KL19 capsule gene cluster

Johanna J. Kenyon,† Mikhail M. Shneider,† Sofya N. Senchenkova,† Alexander S. Shashkov,† Maria N. Sinigina,† Sergey Y. Malanin,† Anastasiya V. Popova,† Konstantin A. Miroshnikov,† Ruth M. Hall† and Yury A. Knirel†

†These authors contributed equally to this work.

Abbreviations: CPS, capsular polysaccharide; GalNAc, N-acetylgalactosamine; GalNAcA, N-acetylgalactosaminuronic acid; GlcNAc, N-acetylglucosamine; HMBC, heteronuclear multiple bond correlation; HSCQ, heteronuclear single-quantum coherence; KL, K locus; QuiN4N, 2,4-diamino-2,4,6-trideoxy-D-glucose; ROESY, rotating-frame nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; UDP, uridine diphosphate; UndP, undecaprenyl phosphate; ST, sequence type.

The GenBank/EMBL/DDBJ accession number for the sequences of the KL19 gene clusters and the wzy-containing genomic island for RBH2 and 28 are KU165787 and KU215659, respectively.

One supplementary table is available with the online Supplementary Material.
INTRODUCTION

Long-chain capsular polysaccharide (CPS) protects bacterial cells from external stresses, and is likely an important contributor to the ability of Acinetobacter baumannii to survive in the hospital environment for long periods (Russo et al., 2010; Geisinger & Isberg, 2015; Roca et al., 2012). In A. baumannii, the CPS is produced by polymerization of an oligosaccharide (K unit) catalysed by a Wzy polymerase. Normally, the wzy gene is located within the CPS biosynthesis gene cluster, which maps at the chromosomal K locus (KL) and is responsible for the synthesis, assembly and export of the capsule (Kenyon & Hall, 2013; Hu et al., 2013).

Many different gene clusters have been found at the KL in A. baumannii (Kenyon & Hall, 2013; Hu et al., 2013; Shashkov et al., 2015a, b; Senchenkova et al., 2014, 2015a; Arbatisky et al., 2015; Kenyon et al., 2014, 2015a, b, c, d; Hamidian et al., 2014; Schultz et al., 2016; Holt et al., 2016). In most cases, they share a common genetic organization, with genes for the Wzx translocase and Wzy polymerase in a central CPS-specific portion flanked by a block of capsule export genes on one side and a block of genes for the synthesis of common sugars on the other. In one case, wzy was found outside this region immediately downstream of wzc in the KL8 gene cluster (Kenyon & Hall, 2013).

Among >90 different gene clusters that we have found at the KL in A. baumannii genomes to date (J. J. Kenyon & R. M. Hall, unpublished data), two closely related types, KL19 and KL39, that lack a wzy gene in or near the KL were identified. In this work, we examined capsule production and structure in isolates carrying the KL19 gene cluster.

METHODS

Bacterial strains. A. baumannii strain RBH2, an extensively antibiotic-resistant isolate, was recovered in 1999 from a patient at the Royal Brisbane Hospital in Brisbane, Australia. RBH2 corresponds to Q46 or Q47 (Runnegar et al., 2010) and is recorded as resistant to gentamicin, tobramycin, amikacin, ciprofloxacin and meropenem. A. baumannii strain 28 was isolated in 2002 from a burn patient at the L. D. Hernandez Research Institute of Emergency Medicine, Saint Petersburg, Russian Federation. Strain RBH2 was grown in a 10 ml overnight culture, which was diluted 1:100 and grown to mid log phase before harvesting the CPS. Strain 28 was cultivated in 2x TY medium (Sigma) overnight; cells were harvested by centrifugation (10 000 g, 20 min), washed with distilled water, suspended in aqueous 70% acetone, precipitated, and dried in air.

Isolation and SDS-PAGE visualization of CPS. Polysaccharides were extracted from isolates RBH2 (Kenyon & Reeves, 2013) and 28 (Shashkov et al., 2015a) as described previously. Samples were loaded into a polyacrylamide gel (16% separating, 4% stacking) and subjected to SDS-PAGE followed by Alcian blue staining as described elsewhere (De Castro et al., 2010).

Chemical analyses and O-deacylation. A CPS sample (1 mg) was hydrolysed with 2 M CF₂CO₂H (120 °C, 2 h). Monosaccharides were analysed by GLC of the alditol acetates (Sawardecker et al., 1965) or acetylated (S)-2-octyl glycosides (Leontin et al., 1993) on a Maestro (Agilent 7820) chromatograph (Interlab) equipped with an HP-5 column (0.32 mm x 30 m) using a temperature programme of 160 °C (1 min) to 290 °C at 7 °C min⁻¹.

A CPS sample was treated with 12.5% aqueous ammonia (37 °C, 2 h), and O-deacylated CPS was isolated by size-exclusion chromatography on a column (56 x 2.6 cm) of Sephadex G-50 Superfine (Amersham Biosciences) using 0.05 M pyridinium acetate buffer pH 4.5 as eluent and monitoring with a Knauer differential refractometer.

NMR spectroscopy. NMR spectra were recorded at 60 °C on a Bruker Avance II 600 spectrometer using a 5 mm broadband inverse probe head for solutions in 99.95% D₂O after deuterium exchange by freeze–drying sample solutions in 99.9% D₂O. Sodium 3-(trimethylsilyl)propionate-2,2,3,3-δ4 (δ0 0, δ1 ~ 1.6) was used as internal reference for calibration. Two-dimensional NMR spectra were obtained using standard Bruker software, and the Bruker TopSpin 2.1 program was used to acquire and process the NMR data. The two-dimensional 1H/1H total correlation spectroscopy (TOCSY) and rotating frame Overhauser effect spectroscopy (ROESY) spectra were recorded with 60 ms duration of MLEV-17 spin-lock and 150 ms mixing time, respectively. The 1H/13C heteronuclear multiple-bond correlation (HMBC) spectrum was recorded with a 60 ms delay for evolution of long-range spin couplings.

Bioinformatics and sequence analysis. The sequence type (ST) in the Pasteur and Oxford multilocus sequence typing (MLST) schemes was determined from the genome sequence at http://pubmlst.org/abau- mannii/. Gene clusters at the KL were characterized by comparison with known A. baumannii gene clusters using the annotation system described previously (Kenyon & Hall, 2013). Functions of proteins were predicted using BLAST and Pfam bioinformatics tools (Finn et al., 2014; Altschul et al., 1990). BLAST was also used to identify Wzy candidates by searching for known Wzy domains (PF04932, PF14296, PF14897 and PF13425), and the genomic region was also further analysed using the PHAST annotation program (Zhou et al., 2011). GenBank and WGS accession numbers of sequences analysed in this study are listed in Table S1 (available in the online Supplementary Material). To identify the additional genomic segment, a portion of sequence including the candidate wzy gene and ~5 kb of sequence on either side was aligned with the complete genome sequence of A. baumannii A1 (GenBank accession number CP010781.1) (Holt et al., 2015).

RESULTS

Isolates carrying the KL19 capsule biosynthesis gene cluster

The KL19 gene cluster (Fig. 1) was originally identified in the genome of A. baumannii RBH2 (GenBank accession number KU165787), an extensively antibiotic-resistant...
isolates were found to belong to GenBank accession number KU215659), which was isolated in 1999 from a patient at the Royal Brisbane Hospital in Brisbane, Australia. This sequence shares 99.9 % identity (100 % coverage) with the sequence of the K19 capsule in the genome of A. baumannii strain 28 (GenBank accession number KU215639), which was isolated in 2002 from a burn patient in the Russian Federation. Both isolates were found to belong to ST111 in the Pasteur MLST scheme, and ST1134 in the Oxford scheme.

The central region of the KL19 gene cluster (Fig. 1) contains genes previously identified as required for the synthesis of UDP-2-acetamido-2-deoxy-D-galactopyranosuronic acid (UDP-D-GalpNAcA) and UDP-2-acetamido-4-acylamino-2,4,6-trideoxy-D-glucopyranose (UDP-D-QuipNAc4NR) (Kenyon & Hall, 2013). There is an itrA1 gene encoding an initiating transferase and two genes for glycosyltransferases, Gtr41 and Gtr2, indicating a trisaccharide K unit. A wza gene for the repeat unit translocase is present but a gne1 gene was not found. Therefore, it could be expected that a long-chain K19 polysaccharide may not be produced.

Fig. 1. KL1, KL19 and KL39 capsule biosynthesis gene clusters. Genes are depicted by arrows that indicate direction of transcription, with gene names shown above. The figure is drawn to scale from GenBank accession numbers CP010781.1 (88154–90337) for KL1, KU165787.1 for KL19, and AMHP01000019.1 (55887–79680) for KL39, respectively. A scale bar and colour scheme key are shown below, and functions of proteins encoded in modules are shown above. Dark grey shading between gene clusters indicates nucleotide sequence identity >90 %. Light grey shading is for nucleotide sequence identity 70–90 %.

K19 capsule is a high molecular mass polysaccharide

Extracellular polysaccharide was extracted from isolates RBH2 and 28, and analysed. Polysaccharide from RBH2 was visualized using SDS-PAGE with Alcian blue stain, which is specific for negatively charged CPS. The RBH2 extract contained a high molecular mass capsule comparable in size to those of other A. baumannii strains known to carry different K gene clusters (Fig. 2), but the quantity of CPS was significantly less (Fig. 2). In Sephadex G-50 size-exclusion chromatography, CPS of each strain was eluted immediately after the void volume of the column, indicating a molecular mass greater than 30 kDa. These data indicate that a long-chain polysaccharide is produced by isolates carrying KL19, suggesting that a Wzy polymerase must be encoded elsewhere in the genome.

Polymerase Wzy and acetyltransferase Atr are encoded in a small genomic island (GI)

The genome sequences of isolates RBH2 and 28 were searched for genes encoding potential Wzy candidates with known polymerase domains. The PF04932 (Wzy_C) domain associated with Wzy polymerases was found in a protein (GenPept accession number ALV86836.1) encoded by an ORF of 1125 bp. This gene was located ~3 kb downstream of the chromosomal cpn60 gene, which encodes a GroEL chaperonin protein (GenPept accession number ALV86838.1; Fig. 3). It could not be found in most A. baumannii genomes. Alignment with the complete genome of A. baumannii A1 (GenBank accession number CP010781.1) (Holt et al., 2015) revealed a 6094 bp insertion in the chromosomes of both RBH2 and 28, and a 9 bp direct repeat of the immediate 3’-end of cpn60.

The GI includes three other ORFs (Fig. 4). The gene adjacent to the putative wzy predicts an acetylttransferase designated Atr25 (GenPept accession number EKL59360.1).
acetylated (S)-2-octyl glycosides showed that GalN has the D configuration.

The $^{13}$C NMR (Fig. 5) and $^1$H NMR spectra of the CPS were complicated by non-stoichiometric O-acetylation [there were signals for an O-acetyl group at $\delta$H 2.07, $\delta$C 21.5 (CH$_3$) and 174.3 (CO)]. After O-deacetylation of the CPS with aqueous ammonia, the spectra looked typical of a polysaccharide having a trisaccharide K unit. The $^{13}$C NMR spectrum of the O-deacetylated CPS showed signals for three anemic carbons at $\delta$ 98.3–102.7, one group each of CH$_3$-C, HOCH$_2$-C and HO$_2$C-C (C-6) of 6-deoxyhexose, hexose and hexuronic acid residues at $\delta$ 17.7, 61.7 and 174.9, respectively, four nitrogen-bearing carbons at $\delta$ 49.2–58.1, other sugar-ring carbons at $\delta$ 68.1–79.3, and four N-acetyl groups at $\delta$ 23.3–23.9 (CH$_3$) and 174.9–176.1 (CO) (Table 1). The $^1$H NMR spectrum contained signals for three anemic protons at $\delta$ 4.60–5.24, one CH$_3$-C group (H-6) of a 6-deoxyhexose at $\delta$ 1.17, other sugar protons at 3.50–4.37, and four N-acetyl groups at $\delta$ 1.94–2.09.

The NMR spectra were assigned using a set of shift-correlated two-dimensional NMR experiments and spin-systems for three sugar residues, A, B and C, were identified (Table 1). Unit C showed correlations of H-1 to H-2,3,4,5 and of H-6 (CH$_3$) at $\delta$ 1.17 to H-5,4,3 in the correlation (COSY) and TOCSY spectra, which are characteristic of the glucos configuration. In contrast, only H-1 to H-2,3,4 correlations were observed for units A and B, which,

**Structure of the K19 CPS of A. baumannii 28**

Sugar analysis of the isolated K19 CPS of A. baumannii 28 by GLC of the acetylated alditols revealed 2-amino-2-deoxygalactose (galactosamine, GalN). GLC analysis of the

belonging to Pfam PF01757 (Acyl_transf_3) and sharing 51% identity with Atr1 (GenPept accession number AJF80035.1) encoded by KL1. Atr1 is responsible for 6-O-acetylation of the D-GalpNac residue in 50–75% of K1 units (Russo et al., 2013). However, in RBH2 the atr25 gene is interrupted by an ISAbai insertion (GI variant 1 in Fig. 4). Functions of orf1 and orf2 in the GI remain unknown. The predicted Orf1 is 50% identical to a protein (GenPept accession number ABH32277.1) annotated as Orf66 (Hu et al., 2013), which is encoded by a gene between mviN and fkbB, just outside the KL35 gene cluster in A. baumannii LUH5535 (also known as PSgc6). Orf1 also shares 51% identity with a hypothetical protein (GenPept accession number AJF81225.1) encoded elsewhere in the genomes of A. baumannii isolates A1, AB0057, AYE, 307–0294 and others belonging to Global Clone 1 (GC1; equivalent to CC1 in the Pasteur MLST scheme). Homologues of known or unknown function were not found for the protein encoded by orf2.

**Fig. 2.** Capsules produced by isolates with different gene clusters at the KL. SDS-PAGE followed by Alcian blue staining was used to visualize capsule extracts. Isolate names are shown above, and the KL type for each strain is indicated below. Separating and stacking phases of polyacrylamide gel are indicated on the left.

**Fig. 3.** Location of the wzy-atr25 genomic insertion. The figure is drawn to scale using AB0057 as a representative genome (GenBank accession number NC_011586.1). A scale bar is shown below, and numbers inside the circle indicate relative Mb positions. The location of the insertion in RBH2/28 is shown. Loci that direct the synthesis of polysaccharides or antibiotic resistance are shown. Small black squares indicate locations of genes that are included in the Pasteur MLST scheme.

28
therefore, are galacto configured. That unit A is a hexose derivative followed from correlations of H-4 with H-5 and H-6a,6b in the ROESY spectrum. Correlations between H-4 and H-5 in the ROESY spectrum and between H-5 and C-6 (\HO_2C) at \(\delta\) 4.05/174.9 in the \(^{1}\text{H},^{13}\text{C}\) HMBC spectrum enabled identification of unit B as a hexuronic acid derivative. Correlations of protons at the nitrogen-linked carbons with the corresponding carbons at \(\delta\) 49.2–58.1 in the \(^{1}\text{H},^{13}\text{C}\) heteronuclear single-quantum coherence (HSQC) spectrum demonstrated that units A and B are 2-amino sugars and unit C is a 2,4-diamino sugar. These data together indicated that the K19 unit consists of one residue each of GalNAC (A), GalNAC (B) and QuiNAc4NAc (C).

A comparison of the \(^{13}\text{C}\) NMR chemical shifts of the CPS with published data (Knirel et al., 1986; Lipkind et al., 1988) showed that units A and B are \(\alpha\)-linked, whereas unit C is \(\beta\)-linked. In the ROESY spectrum, there were A H-1/B H-4, B H-1/C H-3 and C H-1/A H-3 correlations, which defined the sequence and linkage pattern in the K unit. The sites of glycosylation were confirmed by significant downfield displacements of the \(^{13}\text{C}\) NMR signals for the linkage carbons, namely, C-3 of units A and C and C-4 of unit B, as compared with their positions in the non-substituted monosaccharides (Knirel et al., 1986; Lipkind et al., 1988). Analysis of the displacements of the signals for the linkage and neighbouring carbons using known regularities (Knirel et al., 1987; Shashkov et al., 1988) showed that all constituent monosaccharides have the same absolute configuration, namely the \(\mathbf{D}\) configuration.

A comparison of the NMR spectra of the O-deacetylated and initial CPS revealed marked shifts of approximately one-quarter of the \(^{13}\text{C}\) NMR signals for C-5 and C-6 of unit A from \(\delta\) 71.5 and 61.7 to \(\delta\) 69.5 and 64.8 and \(^{1}\text{H}\) NMR signals for H-6a,6b from \(\delta\) 3.66 to \(\delta\) 4.17 and 4.22, respectively. These displacements are indicative of O-acetylation of the GalNAC residue at position 6 in ~25% K units.

Therefore, the K19 CPS of strain 28 has the structure shown in Fig. 6. The structure of the K19 unit, including the O-acetylation pattern, is similar to that of the K1 unit established earlier (Russo et al., 2013; Fig. 6).

### Analysis of the K19 CPS of *A. baumannii* RBH2

The isolated K19 CPS of *A. baumannii* RBH2 was analysed by NMR spectroscopy as described above for the CPS of stain 28. A comparison of the \(^{13}\text{C}\) NMR spectra (Fig. 5) and two-dimensional \(^{1}\text{H},^{13}\text{C}\) HSQC spectra (not shown) indicated that the CPS from strain RBH2 has the same structure as the O-deacetylated CPS from strain 28. Therefore, the CPSs of *A. baumannii* RBH2 and 28 are identical except that the former lacks O-acetylation. This finding correlates with interruption of the atr25 gene in the GI of strain RBH2 by an \text{ISAb}a1 insertion (Fig. 4) and further confirms that At25 is responsible for 6-O-acetylation of the GalpNAc residue in the K19 unit.

### Assignment of the linkage formed by the Wzy polymerase

The ItrA1 (GenPept accession number ALV86826.1) encoded by KL19 is 97% identical to ItrA1 (GenPept accession number CAM88568.1), the initiating transferase (also known as WeeH) encoded by KL1 in *A. baumannii* isolate AYE, which has been shown to link \text{d-QuipNAc4NR} [R = acetyl or (S)-3-hydroxybutanoyl] to the lipid carrier in the inner membrane (Morrison & Imperiali, 2013). Hence, \text{d-QuipNAc4NAc} is the first sugar of the K19 unit (Fig. 6), and the \(\beta\)-\text{d-QuipNAc4NAc-(1→3)-d-GalpNAc} linkage in K19 is formed by WzyGlc, encoded by the 6.1 kb GI.
Assignment of glycosyltransferases to specific sugar linkages

The KL19 gene cluster is similar to the KL1 gene cluster (Fig. 1) previously described for *A. baumannii* GC1 isolates AB307-0294 and AYE (Kenyon & Hall, 2013), A1 (Holt *et al.*, 2015) and LUH5542 (Hu *et al.*, 2013), and the structure of K1 is known (Russo *et al.*, 2013). This allows the bond formed by each glycosyltransferase to be inferred. The KL1 (GenBank accession number CP010781.1, bases 87 963–110 544) and KL19 gene clusters share >90% identity across most of their length. However, in KL19, the wzx<sub>K19</sub> and gtr41 genes replace four genes, wzx<sub>K1</sub>, gtr1, wzy

**Fig. 5.** <sup>13</sup>C NMR spectra of CPS (a) and O-deacetylated CPS (b) from *A. baumannii* 28, and CPS from *A. baumannii* RBH2 (c). Arabic numerals refer to carbons in sugar residues denoted by letters as indicated in Table 1. Signals for carbons A5 and A6 of 6-O-acetylated GlcNAc are annotated in italics.
Table 1. $^1$H and $^{13}$C NMR chemical shifts ($\delta$, ppm) of the O-deacetylated CPS from A. baumannii 28

<table>
<thead>
<tr>
<th>Residue</th>
<th>C-1</th>
<th>C-2</th>
<th>C-3</th>
<th>C-4</th>
<th>C-5</th>
<th>C-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$3-$\alpha$-$D$-GlcpNAc-(1$\rightarrow$4) $\rightarrow$4-$\alpha$-$D$-GalpNAcA-(1$\rightarrow$3)-$\beta$-$D$-QuipNAc4NR-(1$\rightarrow$) A</td>
<td>100.0</td>
<td>49.2</td>
<td>77.7</td>
<td>69.4</td>
<td>71.5</td>
<td>61.7</td>
</tr>
<tr>
<td>B</td>
<td>4.91</td>
<td>4.29</td>
<td>4.00</td>
<td>4.23</td>
<td>4.28</td>
<td>3.66, 3.66</td>
</tr>
<tr>
<td>C</td>
<td>5.24</td>
<td>4.20</td>
<td>3.89</td>
<td>4.37</td>
<td>3.40</td>
<td>4.05</td>
</tr>
</tbody>
</table>

Chemical shifts for NAc groups are: $\delta_H$ 1.94–2.09; $\delta_C$ 23.3–23.9 (CH$_3$) and 174.9–176.1 (CO). $^1$H NMR chemical shifts are italicized.

and atr1 (Fig. 1). Both gene clusters contain gtr2 in the central region, and the one internal linkage, $\alpha$-$D$-GalpNAcA-(1$\rightarrow$3)-$\beta$-$D$-QuipNAc4NR, that is shared by K1 and K19 would be formed by Gtr2 (GenPept accession number ALV86825.1). Hence, Gtr41 (GenPept accession number ALV86824.1) should generate the remaining $\alpha$-$D$-GalpNAcA-(1$\rightarrow$4)-$D$-GalpNAcA linkage in K19, and Gtr1 (GenPept accession number AJF80033.1) should catalyse formation of the $\alpha$-$D$-GlcpNAc-(1$\rightarrow$4)-$D$-GalpNAcA linkage in K1.

**Relationship between KL1 and KL19**

K19 includes the gne1 gene, which catalyses the conversion of D-GlcpNAc to $D$-GalpNAc and is present in most KL gene clusters but missing from KL1. This difference explains the absence of D-GalpNAc from K1. Additionally, K19 includes a different acetyl-/acyl transferase gene (qhbC), for which the last 362 bp of 651 bp shares 84% identity with the last 362 bp of qhbA (660 bp) from KL1. QhbA (also known as WeeI) modifies UDP-$D$-QuipNAc4N with either an acetyl or (S)-3-hydroxybutanoyl group in the ratio 1:1 (Morrison & Imperiali, 2013). The K19 unit contains $D$-QuipNAc4N (Fig. 6) and it is likely that QhbC (GenPept accession number ALV86827.1) is only able to add an acetyl group to UDP-$D$-QuipNAc4N. In addition, Wzy$_{KL}$ is not significantly related to the Wzy$_{KL1}$ protein encoded by K1. This is expected granted that the two proteins have different substrate specificities.

---

**Fig. 6.** Structures of the K19 strain 28 (this work), K19* strain RBH2 (this work), and K1 strain AB307-0294 (Russo et al., 2013) CPS units, and a predicted structure of the K39 CPS unit based on genetic analysis. Sugar names are abbreviated as shown in the text, and brackets indicate the ends of the K units. Grey shading shows differences between structures. Glycosyltransferases and Wzy polymerases are shown as predicted linkage each is predicted to form.
Other isolates that carry KL19 and the 6.1 kb GI

The KL19 gene cluster was detected in a further ten draft genome sequences available in the WGS database, many of which were from isolates recovered between 2006 and 2013 in the USA and Chile (strain information and WGS accession numbers are provided in Table 2). The wzy-atr25 GI was found in all of these. However, in two isolates, the GI is interrupted by an IS\textsubscript{Aba125} (GI variant 2 in Fig. 4; Tables 2 and 3). Overall, the isolates carrying KL19 and the wzy-atr25 GI belong to a range of different STs (Table 2), and

Table 2. GenBank and WGS accession numbers of isolates carrying KL19

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sequence type*</th>
<th>ST allelic profile†</th>
<th>Location</th>
<th>Year</th>
<th>Source</th>
<th>GenBank/WGS accession number of KL19</th>
<th>Insertion variant</th>
<th>GenBank/WGS accession number of GI</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBH2 (F2)</td>
<td>ST111</td>
<td>3,3,2,2,4,8,12</td>
<td>Brisbane, Australia</td>
<td>1999</td>
<td>–</td>
<td>KU165787</td>
<td>1</td>
<td>KU165787</td>
</tr>
<tr>
<td>28</td>
<td>ST111</td>
<td>3,3,2,2,4,8,12</td>
<td>Russian Federation</td>
<td>2002</td>
<td>Burn wound</td>
<td>KU216569</td>
<td>–</td>
<td>KU216569</td>
</tr>
<tr>
<td>MAR002</td>
<td>ST271</td>
<td>3,3,2,2,29,4,14</td>
<td>–</td>
<td>–</td>
<td>Wound</td>
<td>JRHB01000001.1</td>
<td>(88 089 to 111 886)</td>
<td>JRHB01000001.1</td>
</tr>
<tr>
<td>AB\textsubscript{1536-8}</td>
<td>ST413</td>
<td>1,3,2,2,5,8,12</td>
<td>USA</td>
<td>2006</td>
<td>–</td>
<td>AMHA01000170.1</td>
<td>(11 792 to 35 589)</td>
<td>AMHA01000138.1</td>
</tr>
<tr>
<td>AB\textsubscript{1594-8}</td>
<td>ST417</td>
<td>1,2,2,11,1,5</td>
<td>USA</td>
<td>2006</td>
<td>Sputum</td>
<td>AMHD01000187.1</td>
<td>(11 821 to 35 617)</td>
<td>AMHD01000026.1</td>
</tr>
<tr>
<td>AB\textsubscript{TG2018}†</td>
<td>ST417</td>
<td>1,2,2,11,1,5</td>
<td>USA</td>
<td>2012</td>
<td>Sputum</td>
<td>AMIE01000139.1</td>
<td>(1 to 19 054)</td>
<td>AMIE01000071.1</td>
</tr>
<tr>
<td>AB\textsubscript{909-01-7}‡</td>
<td>ST417</td>
<td>1,2,2,11,1,5</td>
<td>USA</td>
<td>2012</td>
<td>Blood</td>
<td>AMHY01000102.1</td>
<td>(1 to 19 055)</td>
<td>AMHY01000086.1</td>
</tr>
<tr>
<td>AB\textsubscript{2009-04-01-7}</td>
<td>ST417</td>
<td>1,2,2,11,1,5</td>
<td>USA</td>
<td>2012</td>
<td>Wound</td>
<td>AMHS01000014.1</td>
<td>(31 602 to 55 397)</td>
<td>AMHS01000024.1</td>
</tr>
<tr>
<td>AB\textsubscript{2008-23-01-01-7}‡</td>
<td>ST417</td>
<td>1,2,2,11,1,5</td>
<td>USA</td>
<td>2012</td>
<td>Clinical environment</td>
<td>AMHQ01000044.1</td>
<td>(21 935 to 40 989)</td>
<td>AMHQ01000076.1</td>
</tr>
<tr>
<td>TG27399</td>
<td>ST417</td>
<td>1,2,2,11,1,5</td>
<td>USA</td>
<td>2013</td>
<td>Sputum</td>
<td>ASGM01000057.1</td>
<td>(31 620 to 55 416)</td>
<td>ASGM01000040.1</td>
</tr>
<tr>
<td>TG27395</td>
<td>ST417</td>
<td>1,2,2,11,1,5</td>
<td>USA</td>
<td>2013</td>
<td>Sputum</td>
<td>ASGL01000017.1</td>
<td>(45 812 to 69 607)</td>
<td>ASGL010000115.1</td>
</tr>
<tr>
<td>Ab5</td>
<td>–‡</td>
<td>3,2,2,2,4,8</td>
<td>Chile</td>
<td>2007</td>
<td>Wound</td>
<td>LANH01000021.1</td>
<td>(108 552 to 132 349)</td>
<td>LANH01000010.1</td>
</tr>
</tbody>
</table>

*ST in Pasteur MLST scheme.
†Allele order: cpn60, fuaA, gltA, pyrG, recA, rplB, rpoB.
‡Sequence is found in multiple contigs.
§fusA allele not detected.

Table 3. Description of GI variants

<table>
<thead>
<tr>
<th>GI variant</th>
<th>Interruption*</th>
<th>Position in atr25 (933 bp)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IS\textsubscript{Aba1} (&lt;)</td>
<td>829</td>
</tr>
<tr>
<td>2</td>
<td>IS\textsubscript{Aba125} (&lt;)</td>
<td>903</td>
</tr>
<tr>
<td>3</td>
<td>IS\textsubscript{Aba1} (&gt;)</td>
<td>823</td>
</tr>
</tbody>
</table>

*Symbols < and > indicate direction of transcription as shown in Fig. 4.
†Number of nucleotides from the start codon.
were recovered from a wide variety of clinical specimens and geographical locations over a period of ten years.

Other isolates carrying the 6.1 kb GI
The genetic content of KL19 is nearly identical to that of the KL39 gene cluster (Fig. 1) that we have found in the genomes of several A. baumannii isolates (see Table S1 for WGS accession numbers). However, KL39 contains the qhbA gene for acetyl-β-hydroxybutyranoyl-transferase in place of the qhbC gene for acetyltransferase (Fig. 1). All genomes carrying KL39 included a variant form of the GI containing a ISAb1 insertion at a different position in the atr25 gene (Table 3; GI variant 3 in Fig. 4).

As noted previously (Holt et al., 2016), two further isolates, Naval-83 and AB058 (Table S1), which belong to GC1, were also found to carry the uninterrupted form of the GI. The genome of Naval-83 includes a variant of the KL1 gene cluster that contains two ISaBa125 elements interrupting the wzy and atr1 genes. The available KL1 gene cluster sequence in AB058 was split into many contigs, and only an incomplete copy of the wzy gene could be found. To confirm that an IS interrupts wzy in AB058, the gene cluster will require re-sequencing.

DISCUSSION

In all A. baumannii gene clusters described to date, wzy and wzy have been found within or, less commonly, immediately adjacent to the KL (Kenyon & Hall, 2013; Hu et al., 2013; Shashkov et al., 2015a, b, c, d; Senchenkova et al., 2014, 2015a, b; Arbatsky et al., 2015; Kenyon et al., 2014, 2015a, b, c, d; Hamidian et al., 2014; Schultz et al., 2016; Holt et al., 2016). In this paper, we describe two gene clusters, KL19 and KL39, which lack a wzy gene, and report the novel finding of a small GI containing a wzy gene located elsewhere in the genome. It is not clear how the GI integrated into the chromosome. Though a 9bp direct repeat borders the GI, it was only found in the one location in all of the isolates that carry it, and inverted repeats or an integrase (int) gene could not be found, suggesting that it may not be mobile.

The 6.1 kb GI was found in all KL19 isolates examined, and interestingly these belong to different STs (Table 2) and were recovered from a wide variety of sources and countries. This indicates that the KL19 gene cluster and the 6.1 kb GI have been circulating globally for some time; RBH2 was isolated in 1999. The GI was also found in isolates with other gene clusters at the KL that either lack a wzy gene (KL39) or contain an interrupted wzy (KL1c variant). Several studies have now shown that the GI is a hotspot for recombination (Kenyon & Hall, 2013; Holt et al., 2016; Snitkin et al., 2011), and it is possible that the GI was acquired into a strain before the genetic content at the KL was lost or interrupted. The presence of an alternate wzy elsewhere in the genome may have relieved the pressure of maintaining the wzy gene in the gene cluster present at the KL.

Several of the genomes examined contained an IS element within the atr25 gene of the GI (Fig. 4), and it is expected that the isolates carrying insertion variants are likely to produce long-chain CPSs that are not modified by O-acetylation, as was found for RBH2. The atr25 and wzy genes overlap, and hence are translationally coupled, and the insertion of an IS in atr25 could reduce the expression of the downstream wzy. However, the ISAb1 element includes a strong outward-facing promoter (Hamidian & Hall, 2013), which is directed toward wzy in the RBH2 GI (variant 1 in Fig. 4). Though wzy mutants can be constructed in the laboratory, natural A. baumannii isolates with wzy missing from the KL that have been found to date are compensated by a second wzy elsewhere (Kenyon & Hall, 2013). This is likely to be due to the fact that capsule is critical for virulence.

The structures of the CPSs of A. baumannii 28 and RBH2 established in this work are closely related to the known structure of the K1 capsule (Fig. 6) reported in a number of strains belonging to the KL1 group, including AB307-0294 (Russo et al., 2013), and AYE and LUH5542 (our unpublished data). K19 differs in the presence of d-GlcNAc, likely due to the absence of gene1 from KL1, the substituent at N-4 of QuipNAc4N-acetyl versus either (S)-3-hydroxybutyranoyl or acetyl, and the degree of O-acetylation on the GlcNAc/GalNAc residue (~25% versus 50–75%). In addition, the linkages between the K units made by WzyK1 and WzyGI are different (1→4 versus 1→3, respectively).

As more structures of capsules from isolates with related gene clusters are determined, functions of enzymes involved in CPS biosynthesis will be able to be predicted with increasing certainty, forming the foundations for future biochemical studies.

ACKNOWLEDGEMENTS

We thank Mohammad Hamidian (University of Sydney, Australia) for assistance with sequence typing, as well as Derek J. Pickard (Wellcome Trust Sanger Institute, UK) and Kathryn E. Holt (University of Melbourne, Australia) for the sequencing and assembly of the RBH2 genome. The research was performed using the equipment of the Interdisciplinary Center for Collective Use of Kazan Federal University, supported by the Education of the Russian Federation (ID RFMEFI59414X0003).

REFERENCES


Edited by: D. W. Hood and M. Whiteley