The kleboxymycin biosynthetic gene cluster is encoded by several species belonging to the *Klebsiella oxytoca* complex

Preetha Shibu[1]#†, Frazer McCuaig[2]#, Anne L. McCartney[3], Magdalena Kujawska[4], Lindsay J. Hall[4,5], Lesley Hoyles[2,6]*

[1]Life Sciences, University of Westminster, United Kingdom

[2]Department of Biosciences, Nottingham Trent University, United Kingdom

[3]Department of Food and Nutritional Sciences, University of Reading, United Kingdom

[4]Gut Microbes & Health, Quadram Institute Bioscience, Norwich Research Park, Norwich, United Kingdom

[5]Chair of Intestinal Microbiome, ZIEL – Institute for Food & Health, Technical University of Munich, Freising, Germany

[6]Department of Metabolism, Digestion and Reproduction, Imperial College London, United Kingdom

#These authors made an equal contribution to the work; shared authorship.

***Corresponding author:** Lesley Hoyles, lesley.hoyles@ntu.ac.uk

†**Present address:** Berkshire and Surrey Pathology Services, Frimley Health NHS Trust, Wexham Park Hospital, Slough, United Kingdom.

Running title: Distribution of the kleboxymycin BGC in *Klebsiella*

Abbreviations: AAHC, antibiotic-associated haemorrhagic colitis; AMR, antimicrobial resistance; BGC, biosynthetic gene cluster; CARD, Comprehensive Antibiotic Resistance Database; ESBL, extended spectrum β-lactamase; KPC, *K. pneumoniae* carbapenemase; MAG, metagenome-assembled genome; MSA, multiple-sequence alignment; NEC, necrotizing enterocolitis; PAI, pathogenicity island; PBD, pyrrolobenzodiazepine; TM, tilimycin; TV, tillivaline; VFDB, Virulence Factors of Pathogenic Bacteria Database.

Keywords: tilimycin, tillivaline, *Klebsiella michiganensis*, antibiotic-associated haemorrhagic colitis, necrotizing enterocolitis

Data statement: Supplementary data associated with this article are available from figshare.

Data summary: Draft genome sequences for PS_Koxy1, PS_Koxy2 and PS_Koxy4 have been deposited with links to BioProject accession number PRJNA562720 in the NCBI BioProject database.

IMPACT STATEMENT

Extended analyses of the genomes of *Klebsiella* spp. have revealed the kleboxymycin biosynthetic gene cluster (BGC) is restricted to species of the *Klebsiella oxytoca* complex (*K. oxytoca*, *K. michiganensis*, *K. pasteurii* and *K. grimontii*). Species- and/or gene-specific differences in the cluster's sequences may be relevant to virulence of *K. oxytoca* and related species. The finding of the kleboxymycin BGC in the preterm infant gut microbiota may have implications for disease presentation in a subset of neonates.

ABSTRACT

As part of ongoing studies with clinically relevant *Klebsiella* spp., we characterized the genomes of three clinical GES-5-positive strains originally identified as *Klebsiella oxytoca*. Average nucleotide identity and phylogenetic analyses showed the strains to be *Klebsiella michiganensis.* In addition to encoding GES-5, the strains encoded SHV-66, a β-lactamase not previously identified in *K. michiganensis*. The strains further encoded a range of virulence factors and the kleboxymycin biosynthetic gene cluster (BGC), previously only found in *K. oxytoca* strains and one strain of *Klebsiella grimontii*. This BGC, associated with antibiotic-associated haemorrhagic colitis, has not previously been reported in *K. michiganensis*, and this finding led us to carry out a wide-ranging study to determine the prevalence of this BGC in *Klebsiella* spp. Of 7,170 publicly available *Klebsiella* genome sequences screened, 88 encoded the kleboxymycin BGC. All BGC-positive strains belonged to the *K. oxytoca* complex, with strains of four (*K. oxytoca*, *K. pasteurii*, *K. grimontii*, *K. michiganensis*) of the six species of the complex found to encode the complete BGC. In addition to being found in *K. grimontii* strains isolated from preterm infants, the BGC was found in *K. oxytoca* and *K. michiganensis* metagenome-assembled genomes recovered from neonates. Detection of the kleboxymycin BGC across the *K. oxytoca* complex may be of clinical relevance and this cluster should be included in databases characterizing virulence factors, in addition to those characterizing BGCs.

INTRODUCTION

Recent work has demonstrated genomic characterization of *Klebsiella oxytoca* strains is inadequate, with large numbers of genomes deposited in public databases erroneously assigned to *K. oxytoca* instead of *Klebsiella michiganensis* or *Klebsiella grimontii* (1–4). Consequently, *K. michiganensis* and *K. grimontii* are clinically relevant but under-reported in the literature (1,5). Three additional novel species (*Klebsiella huaxiensis*, *Klebsiella spallanzanii*, *Klebsiella pasteurii*) belonging to the *K. oxytoca* complex have been proposed recently (6,7).

Little is known about the antibiotic-resistance and virulence genes encoded by *K. oxytoca* and related species. In the course of ongoing *Klebsiella*–phage work, with three GES-5-positive clinical strains (8) of *K. michiganensis*, we sought to determine whether widely recognized virulence factors such as enterobactin, yersiniabactin and salmochelin are encoded in the strains' genomes, and the kleboxymycin biosynthetic gene cluster (BGC), as this was until recently a little-studied BGC implicated in non-*Clostridioides difficile* antibiotic-associated haemorrhagic colitis (AAHC) (9–12). AAHC is caused by the overgrowth of cytotoxin-producing *K. oxytoca* secondary to use of antibiotics such as penicillin or amoxicillin, resulting in the presence of diffuse mucosal oedema and haemorrhagic erosions (13,14). This type of colitis is distinct from the more common form of antibiotic-associated diarrhoea caused by toxin-producing *Clostridiodes difficile*, which usually gives rise to watery diarrhoea resulting in mild to moderate disease.

Our findings led us to investigate the distribution of the kleboxymycin BGC in a range of *Klebsiella* and related species. Here, we report the detailed genotypic characterization of our GES-5-positive strains, and on the distribution of the kleboxymycin BGC in *Klebsiella* spp.

METHODS

Phenotypic characterization of isolates. Strains PS_Koxy1, PS_Koxy2 and PS_Koxy4 had been recovered from a throat swab, urine and rectal swab, respectively. The strains were from the study of Eades *et al.* (8), but no other information on the strains was available to us. The isolates were obtained from Imperial College Healthcare NHS Trust Tissue Bank. Once in the laboratory, API 20E strips

(bioMerieux) were used to confirm the identities of the isolates as *K. oxytoca*, following manufacturer's instructions.

Antimicrobial susceptibility testing. The three isolates were screened on Mueller–Hinton agar for possible carbapenemase production following UK national guidelines (15). This involved testing all presumptive isolates against 18 antimicrobials (amikacin, amoxycillin, augmentin, aztreonam, cefotaxime, cefoxitin, ceftazidime, cefuroxime, ciprofloxacin, colistin, ertapenem, gentamicin, meropenem, tazocin, temocillin, tigecycline, tobramycin, trimethoprim) following the EUCAST disc diffusion method. Control organisms used to monitor test performance of the antimicrobials were *Escherichia coli* ATCC 25922 and *Staphylococcus aureus* ATCC 29213. Zone sizes were read using calibrated callipers and interpreted as sensitive or resistant by referring to the EUCAST breakpoint guidelines (16).

Carbapenemase confirmatory tests were performed on all three isolates as they were found to be resistant to the indicator carbapenem ertapenem. The isolates were further tested with ertapenem E-strips (Launch) to ascertain the Minimum Inhibitory Concentration (MIC). After incubation for 24 h at 30–35 °C the MIC gradients were read against the EUCAST breakpoint guidelines and those showing MICs >0.12 µg/mL were considered resistant to ertapenem (16).

Extraction of DNA. Isolates were plated onto MacConkey agar no. 3 (Oxoid) and incubated aerobically and overnight at 37 °C. On three separate passages, a single colony of each isolate was picked and streaked out. After the third passage, DNA was extracted from a loopful of cells using the Gentra PureGene Qiagen DNA extraction kit (Qiagen). DNA quality was assessed by agarose gel electrophoresis, and quantified using the Nanodrop instrument.

Whole-genome sequencing, assembly and annotation. Extracted DNA was frozen at -20 °C and sent to the Quadram Institute Bioscience, Norwich for library preparation and sequencing. Samples were run on an Illumina Nextseq500 instrument using a Mid Output Flowcell [NSQ® 500 Mid Output KT v2 (300 CYS)] following Illumina's recommended denaturation and loading procedures,

which included a 1% PhiX spike-in (PhiX Control v3). Data were uploaded to Basespace, where the raw data were converted to eight fastq files for each sample (four for R1, four for R2), which were subsequently concatenated to produce one R1 and one R2 file per strain. Sequence data were quality checked using fastqc v0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/): no adaptor or over-represented sequences were present. Data were trimmed using trimmomatic 0.39 (SLIDINGWINDOW:5:20 MINLEN:50) (17), and paired reads retained. Genomes were assembled using the trimmed paired reads with SPAdes v3.13.0 (default settings) (18). Gene predictions and annotations were completed using Prokka v1.14.5 (default settings) (19). The data have been deposited with links to BioProject accession number PRJNA562720 in the NCBI BioProject database.

Analyses of draft genome sequences. Completeness and contamination of the three genomes was assessed using CheckM v1.0.18 (20). Average nucleotide identity (ANI) of genomes with their closest relatives and type strains of species was assessed using FastANI (21). PhyloPhlAn 0.99 (22) was used to determine phylogenetic placements of strains within species. The three draft genomes were uploaded to the *Klebsiella oxytoca* MLST website (https://pubmlst.org/koxytoca/) hosted by the University of Oxford (23) to determine allele number against previously defined house-keeping genes (*rpoB*, *gapA*, *mdh*, *pgi*, *phoE*, *infB* and *tonB*). Kleborate (24,25) and Kaptive (http://kaptive.holtlab.net) were used to attempt to identify capsular (K) and O antigen types (24–26). Presence of antibiotic-resistance genes within strains was determined by BLASTP analysis of amino acid sequences of predicted genes within genomes against the Comprehensive Antibiotic Resistance Database (CARD) database v3.0.7 (downloaded 27 July 2019; protein homolog dataset) (27); only strict and perfect matches with respect to CARD database coverage and bit-score cut-off recommendations are reported, to reduce the potential for reporting false-positive results. Virulence genes were identified by BLASTP of genome amino acid sequences against the Virulence Factors of Pathogenic Bacteria Database (VFDB; ‘core dataset’ downloaded 27 July 2019) (28); results are reported for ≥70 % identity and 90 % query coverage (1).

Characterization of the kleboxymycin BGC in genomes. The annotated reference sequence of the kleboxymycin BGC was downloaded from GenBank (accession number MF401554 (9)) and used as a BLASTP database for searches with the protein sequences encoded within the genomes of PS_Koxy1, PS_Koxy2 and PS_Koxy4. Initially, Geneious Prime v2019.2.1 was used to identify regions of the three draft genomes encoding the complete BGC, and to align them to MF401554.

All annotated non-redundant *Klebsiella* genome assemblies available in the NCBI Genome database on 2 September 2019 (*n* = 7,170; **Supplementary Table 1**) were downloaded (19). The protein sequences of the annotated assemblies were searched for the kleboxymycin BGC using the reference sequence and BLASTP v2.9.0+, and the resulting hits were filtered based on >70 % identity and >70 % coverage to identify isolates potentially carrying genes from the BGC. *K. grimontii* (*n*=3) and *K. michiganensis* (*n*=2) and *K. oxytoca*-related metagenome-assembled genomes (MAGs) (*n*=25) from (1) were also subject to BLASTP searches. Genomes that encoded the full BGC (i.e. all 12 BGC genes on a contiguous stretch of DNA) were identified from the BLAST results. The protein sequences encoded in the BGC were extracted from the annotated assemblies using samtools v1.9 faidx (29) and concatenated into a single sequence (the sequence data are available as supplementary material from figshare). These concatenated sequences were used to produce a multiple-sequence alignment (MSA) in Clustal Omega v1.2.4, along with the BGC sequences of the three *K. michiganensis* clinical isolates, the reference sequence (9), a recently described *K. grimontii* sequence (30) and a homologous sequence found in *Pectobacterium brasiliense* BZA12 (to be used as an outgroup in later phylogenetic analyses; identified as encoding the complete kleboxymycin BGC through NCBI BLASTP). Phylogenetic analyses were carried out on the MSA using the R package Phangorn v2.5.5 (31), producing a maximum-likelihood tree, which was visualised and rooted (on *P. brasiliense* BZA12) using the Interactive Tree of Life (iTOL v5.5) (32). To examine variation at the individual protein level, further within-species MSAs were produced for each of the 12 protein sequences in the BGC. Each of these alignments was used as the basis for a consensus sequence, produced using EMBOSS Cons v6.6.0.0, representing each of the four species carrying the BGC. An MSA and per cent identity matrix were then generated for each protein between the consensus

sequences of *K. oxytoca*, *K. grimontii*, *K. michiganensis* and *K. pasteurii*, along with the reference sequence (9).

The species affiliations of the genomes encoding the full kleboxymycin BGC were determined using FastANI v1.2 (21) against genomes of type strains of the *K. oxytoca* and *K. pneumoniae* complexes (7,33) and *K. aerogenes* ATCC 13048^{T} (assembly accession number GCA_003417445), with PhyloPhlAn 0.99 used to conduct a phylogenetic analysis to confirm species affiliations.

RESULTS

Antimicrobial susceptibility profiles of the three clinical isolates

PS_Koxy1, PS_Koxy2 and PS_Koxy4 were confirmed to be isolates of *K. oxytoca* by phenotypic testing (API 20E profile 5245773, 97.8 % identity). The strains were all found to be resistant to amoxicillin (zone diameter <14 mm), augmentin (<19 mm), aztreonam (<21 mm), cefotaxime (<17 mm), cefoxitin (<19 mm), ceftazidime (<19 mm), cefuroxime (<18 mm), ciprofloxacin (<19 mm), ertapenem (<22 mm), gentamicin (<14 mm), tazocin (<17 mm), temocillin (<19 mm), tobramycin (<14 mm) and trimethoprim (<14 mm), and sensitive to amikacin (>18 mm), colistin (MIC <2 μg/mL), meropenem (>22 mm) and tigecycline (>18 mm). Ertapenem resistance was confirmed by E-test (had an MIC > 0.12 μg/mL), as ertapenem and meropenem are used as an indicator antibiotic for the detection of carbapenemase.

Genotypic characterization of the three clinical isolates

We generated draft genome sequence data for PS_Koxy1, PS_Koxy2 and PS_Koxy4 to accurately identify the strains and provide us with genomic data that would be useful in our future phage studies (**Table 1**).

The ANI values for the three genomes compared with genomes of *K. oxytoca* and related species were determined. PS_Koxy1, PS_Koxy2 and PS_Koxy4 all shared 98.81 %, 98.71 % and 98.71 % ANI, respectively, with the type strain of *K. michiganensis* (W14^{T}, GCA_901556995), and 99.98 to 100.00 % ANI with each other. Based on current recommendations, ANI of 95–96 % and

above with the genome of the type strain is indicative of species affiliation (34). Inclusion of the genomes with those previously confirmed to belong to *K. oxytoca*, *K. michiganensis* and *K. grimontii* (1) along with recently proposed species *K. huaxiensis*, *K. spallanzanii* and *K. pasteurii* (6,7,33) in a phylogenetic analysis confirmed the affiliation of strains PS_Koxy1, PS_Koxy2 and PS_Koxy4 with *K. michiganensis* (**Supplementary Figure 1**). The strains were ST138, in accord with Eades *et al.* (8). No capsule or O antigen types could be assigned to the strains using Kaptive, but all three strains were best matched with KL68 [PS_Koxy1, 17/18 genes matched (*cpsACP* missing); PS_Koxy2 and PS_Koxy4, 16/18 genes matched (*cpsACP* and KL68_18 missing)] and O1v1 [4/7 genes (*wzm*, *wzt*, *glf*, *wbbO*) matched in all strains].

Antimicrobial resistance (AMR) gene profiles of the three *K. michiganensis* clinical isolates

The protein sequences encoded by the strains' genomes were compared against the CARD database (27). Presence of the β-lactamase with carbapenemase activity GES-5 (100 % identity, bit-score 591 – perfect CARD match) was confirmed, so too was that of the extended spectrum β-lactamase (ESBL) CTX-M-15 (100 % identity, bit-score 593 – perfect CARD match) (**Figure 1a**) (8). PS_Koxy1, PS_Koxy2 and PS_Koxy4 also encoded SHV-66 (99.65 % identity, bit-score 580 – strict CARD match). In addition to the β-lactamases GES-5, CTX-M-15 and SHV-66, PS_Koxy1, PS_Koxy2 and PS_Koxy4 encoded several other antibiotic-resistance genes (**Figure 1a**), some reported as rare (e.g. *acrB*, *acrD*, *emrB*, *mdtB*, *mdtC*) in *K. oxytoca* genomes by CARD while others were common (e.g. *baeR*, *fosA5*, CRP, *marA*) (**Table 2**). Furthermore, PS_Koxy1 encoded AAC(6')-Ib7; PS_Koxy1 and PS_Koxy2 encoded *tet*(*A*); PS_Koxy2 and PS_Koxy4 encoded QnrB1 (**Figure 1a**).

Identification of virulence factors encoded in the strains' genomes using VFDB

The three strains had identical virulence factor profiles (**Figure 1b**), encoding the plasminogen activating omptin Pla, the Mg^{2+} transport proteins MgtBC, Hsp60, autoinducer-2 (LuxS), type I fimbriae, type 3 fimbriae, type 6 secretion system I, *Escherichia coli* common pilus and

enterobactin. They also encoded numerous proteins associated with capsule, regulation of capsule synthesis (RcsAB) and LPS, with several of the latter sharing identity with *Haemophilus* endotoxins (RfaD, GalU, LpxC, GmhA/LpcA, KdsA). All six proteins required for allantoin utilization were encoded in the strains' genomes.

Detection of the complete kleboxymycin BGC in clinical isolates

It has long been known that *K. oxytoca* gut colonization is linked with AAHC (14). Schneditz *et al.* (10) showed tillivaline (TV), a pyrrolobenzodiazepine (PBD) derivative produced by *K. oxytoca*, is one of the enterotoxins responsible for causing AAHC. This toxic product is encoded by the heterologous expression of the kleboxymycin [also known as tilimycin (TM) (12)] BGC comprising 12 genes (9). Protein sequences of the reference sequence (9) were used to create a BLASTP database against which the proteins encoded in the genomes of PS_Koxy1, PS_Koxy2 and PS_Koxy4 were compared. The genomes of PS_Koxy1, PS_Koxy2 and PS_Koxy4 encoded a complete kleboxymycin BGC (**Supplementary Figure 2**). All genes in each of the genomes shared >99 % identity and >99 % query coverage with the genes of the reference sequence (12): *mfsX*, 99.76 % identity; *uvrX*, 99.87 %; *hmoX*, 99.80 %; *adsX*, 99.85 %; *icmX*, 99.52 %; *dhbX*, 99.62 %; *aroX*, 99.74 %; *npsA*, 99.80 %; *thdA*, 98.68 %; *npsB*, 99.93 %; *npsC*, 98.47 %; *marR*, 99.39.

Our strains were *K. michiganensis* ST138, so we downloaded and assembled (from BioProject PRJEB30858) available sequence data from *K. oxytoca* ST138 strains described recently (35) and determined whether they were in fact *K. michiganensis* and encoded the kleboxymycin BGC. All strains were confirmed to be *K. michiganensis* on the basis of ANI analysis, and encoded the complete kleboxymycin BGC (**Supplementary Figure 3**).

Schneditz *et al.* (10) reported *npsA*/*npsB* were functionally conserved in six sequenced strains of *K. oxytoca* (**Table 3**), based on a BLASTP analysis. Full details of the analysis are unavailable, with only a brief mention of presence being determined based on BLASTP sequence identities >90 % with no indication of sequence coverage. All the genomes included in the study of Schneditz *et al.* (10) were compared with those of the type strain of *K. oxytoca* and related species to confirm their species affiliations (**Table 3**). While some strains were *K. oxytoca*, others belonged to *K.*

michiganensis, *K. pasteurii*, *K. grimontii* and *Raoultella ornithinolytica*. Using thresholds of 70 % identity and 70 % query coverage in our BLASTP analyses to reduce the potential for detecting false positives, we reanalysed the genomes included in the study of Schneditz *et al.* (10). Our results agreed with those of (10) for all genomes, except we detected *npsA*/*npsB* (and all other genes encoded in the kleboxymycin BGC) in *K. grimontii* SA2. *K. oxytoca* 10–5243, *K. pasteurii* 10–5250, *K. oxytoca* 11492-1, *K. oxytoca* 10–5248 and *K. grimontii* M5a1 also encoded the whole kleboxymycin BGC. All genes in all matches shared greater than 90 % identity across greater than 99 % query coverage. *K. michiganensis* 10–5242, E718 and KCTC 1686 did not encode homologues associated with the kleboxymycin BGC. *K. oxytoca* 10-5245 encoded almost-complete homologues of four genes [EHS96696.1 (*marA*) 98.79 % identity, 99.39 % coverage; EHS96697.1 (*npsC*) 95.38 % identity, 99.23 % coverage; (EHS96698.1 (*mfsX*) 96.68 % identity, 99.87 % coverage; EHS96699.1 (*uvrX*) 94.88 % identity, 99.76 % coverage] in contig JH603137.1.

Detection of the kleboxymycin BGC in the faecal microbiota of preterm infants

Our previous work had highlighted the preterm infant gut microbiota harbours a range of species belonging to the *K. oxytoca* complex (1). BLASTP searches of the two *K. michiganensis* (P049A W, GCA_008120305; P095L Y, GCA_008120085) and three *K. grimontii* (P038I, GCA_008120465; P043G P, GCA_008120425; P079F P, GCA_008120915) strains we previously characterized showed all three *K. grimontii* strains encoded the kleboxymycin BGC (**Supplementary Figure 4**). All BGC genes in their genomes shared >98 % identity and >99 % query coverage with the genes of the reference sequence (9): *mfsX*, 100 % identity; *uvrX*, 99.60–99.73 %; *hmoX*, 99.80 %; *adsX*, 99.69 %; *icmX*, 100 %; *dhbX*, 100 %; *aroX*, 99.94–99.74 %; *npsA*, 99.21–99.41 %; *thdA*, 98.68 %; *npsB*, 99.31–99.38 %; *npsC*, 99.23–100 %; *marR*, 100 %. The BGC was also detected in 8/25 of the preterm-associated *K. oxytoca* complex MAGs (3 *K. oxytoca*, 5 *K. michiganensis*) we described previously (1). An MSA of the preterm-associated genomes' BGC against the reference sequence (9) suggested species-specific clustering of the sequences (**Supplementary Figure 4**).

Prevalence of the kleboxymycin BGC in *Klebsiella* spp.

Given the work detailed above had detected the kleboxymycin BGC in several different but closely related *Klebsiella* species and in a range of clinical and gut-associated isolates, and Hubbard *et al.* (30) recently detected the BGC in a strain of *K. grimontii*, we chose to increase the scope of our analysis to include 7,170 publicly available assembled *Klebsiella* genomes (including our three clinical strains, and five isolates from preterm infants (1)) (**Supplementary Table 1**).

While NCBI annotations for genomes are improving, we have noted issues with identities attributed to *Klebsiella* genomes (1). Consequently, the identity of all genomes included in this work was first confirmed by ANI analysis (**Supplementary Table 1**). Of the 7,170 *Klebsiella* spp. genomes screened, the majority (*n*=6,245) were *K. pneumoniae*, followed by *K. variicola* subsp. *variicola* (*n*=241), *K. quasipneumoniae* subsp. *similipneumoniae* (*n*=184), *K. aerogenes* (*n*=168), *K. quasipneumoniae* subsp. *quasipneumoniae* (*n*=120), *K. michiganensis* (*n*=79), *K. oxytoca* (*n*=66), *K. grimontii* (*n*=24), *K. variicola* subsp. *tropica* (*n*=19), '*K. quasivariicola*' (*n*=11), *K. pasteurii* (n=6), *K. huaxiensis* (*n*=5), *K. africana* (*n*=1) and *K. spallanzanii* (*n*=1). One-hundred and ten (1.5 %) had one or more matches with the 12 genes encoded within the kleboxymycin BGC reference sequence, with all except two genomes (both *K. pneumoniae*) belonging to species of the *K. oxytoca* complex (**Supplementary Table 2**). Ninety-six genomes – all belonging to the *K. oxytoca* complex – encoded at least 12 genes belonging to the BGC (**Supplementary Table 2**), and were examined further.

One genome (GCA_002856195) encoding 12 BGC genes was found to encode two stretches of the same protein with the other cluster-associated genes non-contiguous, while one (GCA_004005605) encoded 13 BGC genes (one gene duplicated) in a non-contiguous arrangement. Fifty-five out of 66 (83.3 %) *K. oxytoca* genomes encoded the entire kleboxymycin BGC, as did 19/24 (79.2 %) *K. grimontii*, 9/79 (11.4 %) *K. michiganensis* and 5/6 (83.3 %) *K. pasteurii* genomes (**Figure 2a**). Phylogenetic analysis (**Figure 2b**) confirmed ANI data (**Supplementary Table 2**) that showed all genomes belonged to species of the *K. oxytoca* complex. The 88 genomes confirmed to encode the complete kleboxymycin BGC included the type strains of *K. grimontii* and *K. pasteurii*. The BGC cluster sequences grouped according to species, and the reference sequence (9) clustered with *K. grimontii* sequences and was closely related to the type strain of that species (**Figure 2c**).

Species-specific consensus sequences were generated for all genes within the kleboxymycin BGC and are available as supplementary material from figshare. Similarity values for each gene within the BGC consensus sequences across the four species are available in **Supplementary Table 3**.

DISCUSSION

Genotypic and phenotypic characteristics of the three clinical *K. michiganensis* strains

The three clinical strains characterized herein had previously been included in a study of outbreak strains encoding GES-5 and CTX-M-15 (8), the first report of GES-5-positive clinical isolates of *K. oxytoca* ST138 in the UK. Subsequently, it has been shown that the GES-5 gene in ST138 strains is encoded on an IncQ group plasmid (35). The whole-genome sequence data reported on previously (8) were not available to us, nor were any clinical details on the strains beyond their site of isolation. API 20E (this study), MALDI-TOF and limited sequence analysis (8) had shown the strains to be *K. oxytoca*. Our previous work with isolates recovered from preterm infants had shown that API 20E testing on its own was insufficient to accurately identify *K. oxytoca* strains (1). The strains described by Eades *et al.* (8) were characterized before the availability of MALDI-TOF databases capable of splitting species of the *K. oxytoca* complex (MALDI-TOF was only able to identify as *K. oxytoca* but did not have sufficient resolution to identify individual species within the complex) (7). As we are using PS_Koxy1, PS_Koxy2, PS_Koxy4 in ongoing phage work, we generated draft genome sequences for the strains, to accurately identify them and facilitate detailed host–phage studies in the future.

ANI and phylogenetic analyses confirmed all three strains belonged to the species *K. michiganensis*, not *K. oxytoca* (**Supplementary Figure 1**). In addition to the AMR genes GES-5 (β-lactamase with carbapenemase activity) and CTX-M-15 (an ESBL responsible for resistance to cephalosporins) reported previously (8), the strains encoded SHV-66, an ESBL not previously reported in *K. oxytoca* and related species (**Figure 1a**, **Table 2**). SHV-66 has previously only been reported in a minority of β-lactamase-producing *K. pneumoniae* in Guangzhou, China (36). In this

study, SHV-66 (99.65 % identity, bit-score 580 – strict CARD match) was also found in *K. michiganensis* strains E718 (37), GY84G39 (unpublished), K1439 (unpublished) and 2880STDY5682598 (5) (accession numbers GCA_000276705, GCA_001038305, GCA_002265195 and GCA_900083915, respectively), included in the phylogenetic analysis shown in **Supplementary Figure 1**. Moradigaravand *et al.* (5) noted in their study that 2880STDY5682598 encoded a bla_{SHV} gene, but did not document its type nor indicate its novelty.

PS_Koxy1, PS_Koxy2 and PS_Koxy4 were resistant to the β-lactam antibiotics amoxicillin and aztreonam, augmentin and tazocin (both containing a β-lactam antibiotic and β-lactamase inhibitor), the cephalosporins cefotaxime, cefoxitin, ceftazidime and cefuroxime, the fluoroquinolone ciprofloxacin, the carbapenem ertapenem, the aminoglycosides gentamicin and tobramycin, the carboxypenicillin temocillin, and the antifolate antibacterial trimethoprim. They were sensitive to amikacin (aminoglycoside), colistin (polymyxin), meropenem (carbapenem) and tigecycline (glycylcycline). bla_{OXA-1} is frequently associated with $bla_{CTX-M-15}$, making isolates resistant to β-lactam–β-lactamase inhibitor combinations (38): all three strains were resistant to augmentin (amoxicillin/clavulanic potassium) and tazocin (piperacillin/tazobactam), with neither OXA-1 nor CTX-M-15 considered common in *K. oxytoca* genomes (**Figure 1b**, **Table 2**). The strains carried a range of plasmid-encoded enzymes [AAC(6')-Ib7, *aadA*, APH(3'')-Ib, APH(6)-Id, AAC(3)-IIe] conferring resistance to gentamicin and tobramycin, though they remained sensitive to amikacin.

Clinical significance of *K. michiganensis*

K. michiganensis was originally proposed to describe an isolate closely related to *K. oxytoca* recovered from a toothbrush holder (39). The bacterium is now recognized as an emerging pathogen, due to improved genomic characterization of clinical isolates that would have previously been described as *K. oxytoca* based on simple phenotypic tests (1–4). Reports specific to *K. michiganensis* have only recently begun to appear in the literature, with the main focus being carriage of AMR genes.

A carbapenemase-producing *K. michiganensis* isolate was recovered from the stool sample of an immunocompromised patient hospitalized with acute diarrhoea in Zhejiang, China (2). On analysis of genomic data, the isolate was found to harbour the carbapenemase-encoding genes bla_{KPC-2}, bla_{NDM-}

$_2$ and *bla*$_{NDM-5}$. This study was the first to isolate *K. michiganensis* carrying these genes in China. In 2019, Seiffert *et al.* (4) reported on a bloodstream infection caused by *K. michiganensis* in an immunocompromised patient in St Gall, Switzerland. Susceptibility testing revealed the isolate to be resistant to first- to fourth-generation cephalosporins and the carbapenem ertapenem, and sensitive only to amikacin and colistin. Genomic analyses showed the strain to encode genes conferring resistance to β-lactamases (*bla*$_{TEM-1B,}$ *bla*$_{OXY-1-1}$), aminoglycosides [*aac(3)-IId*, *aph(3')-Ia*, *aadA5*, *strA*, *strB*], macrolides [*mph(A)*], sulphonamides (*su11*, *su12*), tetracycline [*tet(B)*] and trimethoprim (*dfrA17*). *K. pneumoniae* carbapenemase (KPC)3 gene was encoded on plasmid IncFIIK2-FIB. This was the first study to identify *bla*$_{KPC}$ in *K. michiganensis* in Europe. *bla*$_{KPC-3}$- and *bla*$_{KPC-2}$-encoding isolates of *K. michiganensis* recovered from clinical samples have also been reported in the US (40).

During a study of multidrug-resistant *Enterobacteriaceae* recovered from inpatients in 10 private hospitals in Durban, South Africa, a strain of *K. michiganensis* belonging to ST170 and encoding *bla*$_{NDM-1}$ was isolated (3). Plasmid-associated fluoroquinolone resistance genes [*aac(6')lb-cr*, *oqxB*, *oqxA*, QnrS1] have also been reported for *K. michiganensis* (ST170) isolated in Durban (41). In addition, a *bla*$_{OXA-181}$- and *bla*$_{NDM-1}$-encoding strain of *K. michiganensis* isolated from a cancer patient in KwaZulu-Natal Province, South Africa has been reported (42).

ESBL-producing *K. michiganensis* was recently found to be responsible for an outbreak in a neonatal ward in South East Queensland, Australia (43). The bacterium was introduced to the ward due to it contaminating detergent bottles used for washing milk expression equipment. During our recent work characterizing *Klebsiella* spp. recovered from preterm infants (1), we showed two of the 109 infants harboured *K. michiganensis* in their gut microbiota. No *K. oxytoca* strains were recovered, and three of the infants harboured *K. grimontii* in their faeces. Phenotypic testing showed the *K. michiganensis* isolates had intermediate resistance to both benzylpenicillin (encoded FosA5) and meropenem (encoded OXY-1-2, specific to *K. michiganensis*). They were able to persist in macrophages, suggesting *K. michiganensis* has immune evasion abilities similar to *K. pneumoniae*. Genomic characterization showed the strains encoded the acquired antimicrobial resistance genes *emrB* and *emrR*, and a range of core antimicrobial resistance genes (*acrB*, *acrD*, CRP, *marA*, *mdtB*, *mdtC*, *baeR*, FosA5, *pmrF*, *oqxA*, *oqxB*, *msbA*, KpnE, KpnF, KpnG, *Escherichia ampH* β-lactamase).

Enterobactin and yersiniabactin were also encoded by both isolates studied, with *in vitro* assays confirming the two strains were able to produce siderophores. Other virulence factors encoded by the strains included plasminogen activator, *E. coli common* pilus, autoinducer-2, type 3 fimbriae, RcsAB, OmpA, Hsp60 and the AcrAB efflux pump. We also recovered 21 *K. michiganensis* MAGs from the preterm infant gut microbiota compared with four *K. oxytoca* MAGs. Similar to the isolates we studied, all the MAGs encoded antimicrobial resistance genes (*emrR*, *bacA*, *acrB*, *acrD*, CRP, *marA*, *mdtB*, *pmrF*, *msbA*, KpnG, *Escherichia ampH* β-lactamase) and virulence factors (including enterobactin Hsp60, *E. coli* common pilus, autoinducer-2, type 3 fimbriae, RcsAB and AcrAB efflux pump) (1).

As in the above-mentioned studies, all three clinical *K. michiganensis* strains characterized in this study encoded a range of AMR genes [including *emrB*, *acrB*, *marA*, *sul1*, *sul2*, *acrD*, *emrR*, CRP, *mdtBC*, *baeR*, TEM-1, SHV-66, OXA-1, CTX-M-15, GES-5, OXA1-4, *aadA*, *aph(3'')-Ib7*, *aph(6)-Id*, *dfrA14*, *bacA*, FosA5, *pmrF*, *oqxAB*, *msbA*, KPnEFG, *E. coli* β-lactamase and *aac(3)-Ile*], further demonstrating this bacterium contributes to the clinical resistome (**Figure 1a**, **Table 2**). As with the isolates and MAGs described in our infant study (1), the three clinical strains of *K. michiganensis* encoded the virulence factors plasminogen activator, Hsp60, autoinducer-2, type 3 fimbriae, *E. coli* common pilus, RcsAB and enterobactin. All three strains encoded the allantoinase gene cluster associated with liver infection. We previously only detected this gene cluster in 5/21 of the MAGs described in our studies of preterm infants (1). The contribution of *K. michiganensis* to liver disease remains to be determined.

Of the 54 genomes deposited in GenBank as *K. michiganensis*, only 47 were identified as this bacterium using ANI analysis (**Supplementary Table 1**): three were *K. grimontii*, three were *K. pasteurii* and one was *K. spallanzanii* (confirmed by phylogenetic analysis, data not shown). Of the 79 *K. michiganensis* included in this study, 32 had been listed in GenBank as *K. oxytoca* (*n*=28), *K. pneumoniae* (*n*=1) or *Klebsiella* sp. (*n*=3) (**Supplementary Table 1**). While we only comment on the *K. michiganensis* genomes here, misidentified genomes were observed across all species of *Klebsiella* deposited in GenBank, with the exception of *K. aerogenes*. Therefore, there is still a need to carefully check identities of genomes deposited in public repositories prior to using them in any work. It is also

clear that *K. michiganensis* may be more clinically relevant than previously thought, given that it represented almost half of the 181 *K. oxytoca* complex genome assemblies available publicly and included in this study (**Figure 2a**, **Supplementary Table 4**).

A recent report shows multi-drug-resistant *K. michiganensis* is of veterinary relevance (44). Two *K. michiganensis* isolates (both encoding five plasmids) recovered from lavage samples from diseased horses in Austria were resistant to cefotaxime, gentamicin, tobramycin, tetracycline, doxycycline, chloramphenicol and trimethoprim/sulfamethoxazole. The strains encoded: $bla_{\text{OXY-4-1}}$ and $bla_{\text{CTX-M-1}}$; *aac(3)-IId*, *aadA5*, *aph(3'')-Ib*, *aph(3')-Ia* and *aph(6)-Id* (active against aminoglycosides); *tet(B)* (active against tetracyclines); *catA1* (active against chloramphenicol); and *sul1*, *sul2* and *dfrA17* (active against trimethoprim/sulfamethoxazole). Interestingly, CTX-M-1 was detected on transconjugants/transformants in mating experiments. While there is clear overlap in the AMR genes encoded by animal- and human-associated isolates, how similar they are to one another at the genomic level is unknown.

Distribution of the kleboxymycin BGC in *Klebsiella* spp.

As relatively little is known about the virulence factors of *K. oxytoca* and related species, and the VFDB is limited with respect to the number of *Klebsiella* spp. on which it reports information, we wanted to see whether our strains encoded the kleboxymycin BGC responsible for generating microbiome-associated metabolites known to directly contribute to AAHC (9,10). The cytotoxic nature of a heat-stable, non-proteinaceous component of spent media from *K. oxytoca* strains isolated from patients with AAHC was first reported in 1990 (45). With respect to *K. oxytoca* being a causative agent of AAHC, the bacterium has fulfilled Koch's postulates (13). While a commensal of the gut microbiota of some individuals, it has been suggested that cytotoxic *K. oxytoca* is a transient member of the gut microbiota (45).

TV is a PBD produced by *K. oxytoca* and is a causative agent of AAHC (10). The TV biosynthesis genes are encoded on a non-ribosomal peptide synthase operon and include *npsA*, *thdA* and *npsB*. The genes *aroX* and *aroB* are also essential for TV production (11). *npsA*, *thdA*, *npsB* and *aroX* are located on a pathogenicity island (PAI). In clinical isolates, the PAI was present in 100 % of

toxin-producing isolates, but only 13 % of non-toxin-producing isolates (10). AAHC is characterized by disruption of epithelial barrier function resulting from apoptosis of epithelial cells lining the colon. TV exerts its apoptotic effect by binding to tubulin and stabilising microtubules, leading to mitotic arrest (12).

A second PBD generated by the same pathway as TV has been identified (11). TM [also called kleboxymycin (9)] has stronger cytotoxic properties than TV, having a PBD motif with a hydroxyl group at the C11 position, while TV has an indole ring. When deprived of indole by the inactivation of the indole-producing tryptophanase gene *tnaA*, *K. oxytoca* produces TM but not TV. TV production is restored with the addition of indole, as indole spontaneously reacts with TM to produce TV. Limited interconversion between TM and TV may also occur spontaneously *in vivo* (9). TM is a genotoxin and triggers apoptosis by interacting with DNA, which leads to the activation of damage repair mechanisms, causing DNA strand breakage (12). DNA interaction is prevented in the case of TV by its indole ring, and both the molecular targets and apoptotic mechanisms of TM and TV are distinct. The kleboxymycin BGC is not native to *K. oxytoca*, nor the wider *Enterobacteriaceae*. Instead, the BGC is thought to have been acquired via horizontal gene transfer from *Xenorhabdus* spp., which in turn acquired the BGC from bacteria of the phylum *Actinobacteria* (9).

In the current study, we found the kleboxymycin BGC in our *K. michiganensis* isolates and that it was common among four species of the *K. oxytoca* complex, with *K. oxytoca* and *K. grimontii* strains making the largest contribution and the type strains of *K. grimontii* and *K. pasteurii* encoding the BGC (**Figure 2**). Prior to this study, sequences from two *K. oxytoca* strains (MH43-1, GenBank accession number MF401554 (9); AHC-6, GenBank accession number HG425356 (10)) were available for the kleboxymycin BGC. Draft genome sequences do not appear to be publicly available for either of these strains. However, our analysis of the kleboxymycin BGC across the *K. oxytoca* complex has shown that MH43-1 is a strain of *K. grimontii* (**Figure 2c**). Hubbard *et al*. (30) recently reported on a strain of *K. grimontii* that encoded the BGC, based on antiSMASH analysis. Comparison of the AHC-6 sequence with that of MH43-1 and other sequences included in this study shows AHC-6 is a strain of *K. oxytoca* (99.0–99.55 % nucleotide similarity with the BGCs encoded

by the three *K. oxytoca* MAG sequences included in **Supplementary Figure 4**). It is likely that as more genomes of *K. oxytoca* complex species are deposited in public databases, the range of species encoding the kleboxymycin BGC will increase.

All three of our *K. michiganensis* strains encoded the kleboxymycin BGC (**Figure 2**, **Supplementary Figure 2**), as did strains of *K. grimontii* we previously isolated from preterm infants and *K. oxytoca* and *K. michiganensis* MAGs recovered from publicly available shotgun metagenomic data (**Supplementary Figure 3**). Whether the BGCs encoded in our clinical and infant-associated strains are functional will be the subject of future studies. The discovery of the kleboxymycin BGC in strains and MAGs recovered from preterm infants is of particular concern. Gut colonization is linked with AAHC, with disease caused by the overgrowth of cytotoxin-producing strains secondary to use of antibiotics (14). AAHC presents as diffuse mucosal oedema and haemorrhagic erosions (14), and patients pass bloody diarrhoea (46). The gut microbiota of preterm infants is shaped by the large quantity of antibiotics these infants are given immediately after birth to cover possible early onset infection, with 'blooms' of bacteria preceding onset of infection (1). Blood in the stool is frequently associated with necrotizing enterocolitis (NEC) in preterm infants, which shares similar pathological hallmarks to AAHC – i.e. intestinal necrosis. Notably, NEC is difficult to diagnose in the early stages and is often associated with sudden serious deterioration in infant health, with treatment options limited due to emerging multi-drug-resistant bacteria associated with disease. Previous studies have linked *Klebsiella* spp. to preterm NEC (supported by corresponding clinical observations), with bacterial overgrowth in the intestine linked to pathological inflammatory cascades, facilitated by a 'leaky' epithelial barrier and LPS–TL4 activation. Recent work has demonstrated *K. oxytoca* isolates recovered from infants with NEC can produce kleboxymycin (TM) and TV (47). Taken together with the results from our study, we suggest specific virulence factors – i.e. kleboxymycin-related metabolites encoded by atypical *Klebsiella* spp. – may also play a role in NEC, and this warrants further study.

Attempts have been made to link specific subtypes of *K. oxytoca* to AAHC (48). Cytotoxic effects were limited to *K. oxytoca*, with faecal (and to a lesser extent skin) isolates of *K. oxytoca* most commonly associated with cytotoxicity (48). No genetic relationship was associated with cytotoxic

strains based on pulsed-field gel electrophoresis, and 31/97 strains exhibited evidence of cytotoxin production (i.e. reduced viability of Hep2 cells). Joainig *et al.* (48) isolated genetically distinct cytotoxin-positive and -negative strains from one AAHC patient, leading them to suggest that, when detected in faeces, *K. oxytoca* should be considered an opportunistic pathogen able to produce disease upon antibiotic treatment. They also found that, in patients with acute or chronic diarrhoeal diseases, more than half of the isolates recovered were cytotoxin-positive. Given that *K. oxytoca*-related species are not routinely screened for in such samples, it is possible that kleboxymycin-producing isolates may make a greater contribution to diarrhoeal diseases than currently recognized, especially in patients suffering from non-*C. difficile*-associated disease. We have shown that there are species-specific differences in the kleboxymycin BGC (**Figure 2c**). These differences may have implications for virulence of strains and warrant further study. It is hoped that the identification of an increased range of strains (including type strains) encoding the kleboxymycin BGC will facilitate such studies.

ACKNOWLEDGEMENTS

Imperial College Healthcare NHS Trust Tissue Bank is thanked for providing access to clinical strains. Imperial Health Charity is thanked for contributing to registration fees for the Professional Doctorate studies of PS. This work used the computing resources of CLIMB (49), the UK MEDical BIOinformatics partnership (UK Med-Bio, supported by Medical Research Council grant number MR/L01632X/1) and the Nottingham Trent University Hamilton High Performance Computing Cluster. The research placement of FM was funded by Nottingham Trent University. LJH is funded by a Wellcome Trust Investigator Award (no. 100/974/C/13/Z); a BBSRC Norwich Research Park Bioscience Doctoral Training grant no. BB/M011216/1 (supervisor LJH, student MK); an Institute Strategic Programme Gut Microbes and Health grant no. BB/R012490/1 and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356; and an Institute Strategic Programme Gut Health and Food Safety grant no. BB/J004529/1 to LJH. PS did all phenotypic characterization work. PS initial bioinformatics analyses (assembly, annotation, CARD, initial BGC work), while FM and LH undertook the large-scale analyses and infant-associated work; ALM and MK prepared and pre-processed DNA for sequencing; LJH, ALM and LH supervised the work. We

would like to thank Dave Baker and the QIB core sequencing team for WGS library preparation and sequencing. All authors contributed to interpretation and analyses of data, and writing of the manuscript.

REFERENCES

1. Chen Y, Brook TC, Soe CZ, O'Neill I, Alcon-Giner C, Leelastwattanagul O, et al. Preterm infants harbour diverse *Klebsiella* populations, including atypical species that encode and produce an array of antimicrobial resistance- and virulence-associated factors. Microb Genom. 2020 May 21;761924.
2. Zheng B, Xu H, Yu X, Lv T, Jiang X, Cheng H, et al. Identification and genomic characterization of a KPC-2-, NDM-1- and NDM-5-producing *Klebsiella michiganensis* isolate. J Antimicrob Chemother. 2018 01;73(2):536–8.
3. Pedersen T, Sekyere JO, Govinden U, Moodley K, Sivertsen A, Samuelsen Ø, et al. Spread of plasmid-encoded NDM-1 and GES-5 carbapenemases among extensively drug-resistant and pandrug-resistant clinical *Enterobacteriaceae* in Durban, South Africa. Antimicrob Agents Chemother. 2018;62(5).
4. Seiffert SN, Wüthrich D, Gerth Y, Egli A, Kohler P, Nolte O. First clinical case of KPC-3-producing *Klebsiella michiganensis* in Europe. New Microbes New Infect. 2019 May;29:100516.
5. Moradigaravand D, Martin V, Peacock SJ, Parkhill J. Population structure of multidrug-resistant *Klebsiella oxytoca* within hospitals across the United Kingdom and Ireland identifies sharing of virulence and resistance genes with *K. pneumoniae*. Genome Biol Evol. 2017 Mar 1;9(3):574–84.
6. Hu Y, Wei L, Feng Y, Xie Y, Zong Z. *Klebsiella huaxiensis* sp. nov., recovered from human urine. Int J Syst Evol Microbiol. 2019 Feb;69(2):333–6.
7. Merla C, Rodrigues C, Passet V, Corbella M, Thorpe HA, Kallonen TVS, et al. Description of *Klebsiella spallanzanii* sp. nov. and of *Klebsiella pasteurii* sp. nov. Front Microbiol. 2019;10:2360.

8. Eades C, Davies F, Donaldson H, Hopkins K, Hill R, Otter J, et al. GES-5 carpabapenemase-producing *Klebsiella oxytoca* causing clinical infection in a UK haematopoetic stem cell transplantation unit. 26th European Congress of Clinical Microbiology and Infectious Diseases, Amsterdam, The Netherlands; 2016.

9. Tse H, Gu Q, Sze K-H, Chu IK, Kao RY-T, Lee K-C, et al. A tricyclic pyrrolobenzodiazepine produced by *Klebsiella oxytoca* is associated with cytotoxicity in antibiotic-associated hemorrhagic colitis. J Biol Chem. 2017 24;292(47):19503–20.

10. Schneditz G, Rentner J, Roier S, Pletz J, Herzog KAT, Bücker R, et al. Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis. Proc Natl Acad Sci U S A. 2014 Sep 9;111(36):13181–6.

11. Dornisch E, Pletz J, Glabonjat RA, Martin F, Lembacher-Fadum C, Neger M, et al. Biosynthesis of the Enterotoxic pyrrolobenzodiazepine natural product tilivalline. Angew Chem Int Ed Engl. 2017 13;56(46):14753–7.

12. Unterhauser K, Pöltl L, Schneditz G, Kienesberger S, Glabonjat RA, Kitsera M, et al. *Klebsiella oxytoca* enterotoxins tilimycin and tilivalline have distinct host DNA-damaging and microtubule-stabilizing activities. Proc Natl Acad Sci U S A. 2019 26;116(9):3774–83.

13. Högenauer C, Langner C, Beubler E, Lippe IT, Schicho R, Gorkiewicz G, et al. *Klebsiella oxytoca* as a causative organism of antibiotic-associated hemorrhagic colitis. N Engl J Med. 2006 Dec 7;355(23):2418–26.

14. Beaugerie L, Metz M, Barbut F, Bellaiche G, Bouhnik Y, Raskine L, et al. *Klebsiella oxytoca* as an agent of antibiotic-associated hemorrhagic colitis. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2003 Sep;1(5):370–6.

15. Public Health England. UK Standards for Microbiology Investigations P 8: Laboratory detection and reporting of bacteria with carbapenem-hydrolysing β-lactamases (carbapenemases) [Internet]. 2013. Available from: https://webarchive.nationalarchives.gov.uk/20140712120634/http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1317138518829

16. EUCAST. Clinical breakpoints – bacteria v 5.0 [Internet]. 2015. Available from: http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_5.0_Breakpoint_Table_01.pdf
17. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014 Aug 1;30(15):2114–20.
18. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol J Comput Mol Cell Biol. 2012 May;19(5):455–77.
19. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl. 2014 Jul 15;30(14):2068–9.
20. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015 Jul;25(7):1043–55.
21. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018 30;9(1):5114.
22. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun. 2013;4:2304.
23. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124.
24. Lam M, Wyres KL, Duchêne S, Wick RR, Judd LM, Gan Y-H, et al. Population genomics of hypervirulent *Klebsiella pneumoniae* clonal-group 23 reveals early emergence and rapid global dissemination. Nat Commun [Internet]. 2018 Jul 13 [cited 2018 Aug 14];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6045662/
25. Wick RR, Heinz E, Holt KE, Wyres KL. Kaptive Web: User-Friendly capsule and lipopolysaccharide serotype prediction for *Klebsiella* genomes. J Clin Microbiol. 2018 Jun;56(6).

26. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, et al. Identification of *Klebsiella* capsule synthesis loci from whole genome data. Microb Genomics. 2016 Dec;2(12):e000102.

27. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 2017 Jan 4;45(D1):D566–73.

28. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019 Jan 8;47(D1):D687–92.

29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009 Aug 15;25(16):2078–9.

30. Hubbard ATM, Newire E, Botelho J, Reiné J, Wright E, Murphy EA, et al. Isolation of an antimicrobial-resistant, biofilm-forming, *Klebsiella grimontii* isolate from a reusable water bottle. MicrobiologyOpen. 2020 Mar 3;e1023.

31. Schliep KP. phangorn: phylogenetic analysis in R. Bioinforma Oxf Engl. 2011 Feb 15;27(4):592–3.

32. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019 Jul 2;47(W1):W256–9.

33. Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A, Brisse S. Description of *Klebsiella africanensis* sp. nov., *Klebsiella variicola* subsp. *tropicalensis* subsp. nov. and *Klebsiella variicola* subsp. *variicola* subsp. nov. Res Microbiol. 2019 May;170(3):165–70.

34. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol. 2018 Jan;68(1):461–6.

35. Ellington MJ, Davies F, Jauneikaite E, Hopkins KL, Turton JF, Adams G, et al. A multi-species cluster of GES-5 carbapenemase producing *Enterobacterales* linked by a geographically disseminated plasmid. Clin Infect Dis Off Publ Infect Dis Soc Am. 2019 Nov 20;

36. Zuo B, Liu Z-H, Wang H-P, Yang Y-M, Chen J-L, Ye H-F. [Genotype of TEM- and SHV-type beta-lactamase producing *Klebsiella pneumoniae* in Guangzhou area]. Zhonghua Yi Xue Za Zhi. 2006 Nov 7;86(41):2928–32.

37. Liao T-L, Lin A-C, Chen E, Huang T-W, Liu Y-M, Chang Y-H, et al. Complete genome sequence of *Klebsiella oxytoca* E718, a New Delhi metallo-β-lactamase-1-producing nosocomial strain. J Bacteriol. 2012 Oct;194(19):5454.

38. Sugumar M, Kumar KM, Manoharan A, Anbarasu A, Ramaiah S. Detection of OXA-1 β-lactamase gene of *Klebsiella pneumoniae* from blood stream infections (BSI) by conventional PCR and in-silico analysis to understand the mechanism of OXA mediated resistance. PloS One. 2014;9(3):e91800.

39. Saha R, Farrance CE, Verghese B, Hong S, Donofrio RS. *Klebsiella michiganensis* sp. nov., a new bacterium isolated from a tooth brush holder. Curr Microbiol. 2013 Jan;66(1):72–8.

40. Hazen TH, Mettus R, McElheny CL, Bowler SL, Nagaraj S, Doi Y, et al. Diversity among blaKPC-containing plasmids in *Escherichia coli* and other bacterial species isolated from the same patients. Sci Rep. 2018 06;8(1):10291.

41. Osei Sekyere J, Amoako DG. Genomic and phenotypic characterisation of fluoroquinolone resistance mechanisms in *Enterobacteriaceae* in Durban, South Africa. PloS One. 2017;12(6):e0178888.

42. Founou RC, Founou LL, Allam M, Ismail A, Essack SY. Genomic characterisation of *Klebsiella michiganensis* co-producing OXA-181 and NDM-1 carbapenemases isolated from a cancer patient in uMgungundlovu District, KwaZulu-Natal Province, South Africa. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2018 13;109(1):7–8.

43. Chapman P, Forde BM, Roberts LW, Bergh H, Vesey D, Jennison AV, et al. Genomic investigation reveals contaminated detergent as the source of an extended-spectrum-β-lactamase-producing *Klebsiella michiganensis* outbreak in a neonatal unit. J Clin Microbiol. 2020 Apr 23;58(5).

44. Loncaric I, Cabal Rosel A, Szostak MP, Licka T, Allerberger F, Ruppitsch W, et al. Broad-spectrum cephalosporin-resistant *Klebsiella* spp. Isolated from diseased horses in Austria. Anim Open Access J MDPI. 2020 Feb 20;10(2).

45. Higaki M, Chida T, Takano H, Nakaya R. Cytotoxic component(s) of *Klebsiella oxytoca* on HEp-2 cells. Microbiol Immunol. 1990;34(2):147–51.

46. Youn Y, Lee SW, Cho H-H, Park S, Chung H-S, Seo JW. Antibiotics-associated hemorrhagic colitis caused by *Klebsiella oxytoca*: two case reports. Pediatr Gastroenterol Hepatol Nutr. 2018 Apr;21(2):141–6.

47. Paveglio S, Ledala N, Rezaul K, Lin Q, Zhou Y, Provatas AA, et al. Cytotoxin-producing *Klebsiella oxytoca* in the preterm gut and its association with necrotizing enterocolitis. Emerg Microbes Infect. 2020 Dec;9(1):1321–9.

48. Joainig MM, Gorkiewicz G, Leitner E, Weberhofer P, Zollner-Schwetz I, Lippe I, et al. Cytotoxic effects of *Klebsiella oxytoca* strains isolated from patients with antibiotic-associated hemorrhagic colitis or other diseases caused by infections and from healthy subjects. J Clin Microbiol. 2010 Mar;48(3):817–24.

49. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, Poplawski R, et al. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. Microb Genomics. 2016;2(9):e000086.

Table 1. Sequencing summary statistics for the three clinical isolates characterized in this study (all 60× coverage)

Isolate	Size (bp)	No. contigs	N50	CDS	Completeness, contamination*
PS_Koxy1	6,271,778	135	170,962	5,948	100 %, 0.30 %
PS_Koxy2	6,275,379	111	172,981	5,946	100 %, 0.30 %
PS_Koxy4	6,294,880	112	146,343	5,978	100 %, 0.30 %

*Determined using CheckM v1.0.18.

Table 2. Summary of information for CARD genes found in *K. michiganensis* PS_Koxy1, PS_Koxy2 and PS_Koxy4

CARD data from analyses of 257 *K. oxytoca* complex genomes (https://card.mcmaster.ca/prevalence). CARD Prevalence 3.0.7 is based on sequence data acquired from NCBI on 7 May 2020.

ARO: accession	Name	Definition	Prevalence (%) in *K. oxytoca**
3000074	*emrB*	Translocase in the *emrB*-TolC efflux protein in *E. coli*. It recognises substrates including carbonyl cyanide *m*-chlorophenylhydrazone, nalidixic acid and thioloactomycin.	0.78
3000165	*tet*(*A*)	Tetracycline efflux pump found in many species of Gram-negative bacteria.	5.84
3000216	*acrB*	Protein subunit of AcrA-AcrB-TolC multidrug efflux complex. AcrB functions as a heterotrimer which forms the inner membrane component and is primarily responsible for substrate recognition and energy transduction by acting as a drug/proton antiporter.	0.78
3000263	*marA*	In the presence of antibiotic stress, *E. coli* overexpresses the global activator protein MarA, which besides inducing MDR efflux pump AcrAB, also down-regulates synthesis of the porin OmpF.	98.05
3000410	*sul1*	Sulfonamide resistant dihydropteroate synthase of Gram-negative bacteria. It is linked to other resistance genes of class 1 integrons.	13.23
3000412	*sul2*	Sulfonamide resistant dihydropteroate synthase of Gram-negative bacteria, usually found on small plasmids.	2.72
3000491	*acrD*	Aminoglycoside efflux pump expressed in *E. coli*. Its expression can be induced by indole, and is regulated by *baeRS* and *cpxAR*.	0.78
3000516	*emrR*	EmrR is a negative regulator for the EmrAB-TolC multidrug efflux pump in *E. coli*. Mutations lead to EmrAB-TolC overexpression.	99.61
3000518	CRP	CRP is a global regulator that represses MdtEF multidrug efflux pump expression.	100
3000793	*mdtB*	MdtB is a transporter that forms a heteromultimer complex with MdtC to form a multidrug transporter. MdtBC is part of the MdtABC-TolC efflux complex.	0.39
3000794	*mdtC*	MdtC is a transporter that forms a heteromultimer complex with MdtB to form a multidrug transporter. MdtBC is part of the MdtABC-TolC efflux complex. In the absence of MdtB, MdtC can form a homomultimer complex that results in a functioning efflux complex with a narrower drug specificity.	0.39
3000828	*baeR*	BaeR is a response regulator that promotes the expression of MdtABC and AcrD efflux complexes.	99.22
3000873	TEM-1	TEM-1 is a broad-spectrum beta-lactamase found in many Gram-negative bacteria. Confers resistance to penicillins and first generation cephalosphorins.	5.45
3001121	SHV-66	SHV-66 is an extended-spectrum beta-lactamase found in *K.*	ND*

ARO: accession	Name	Definition	Prevalence (%) in *K. oxytoca**
		pneumoniae.	
3001396	OXA-1	OXA-1 is a beta-lactamase found in *E. coli*.	1.95
3001878	CTX-M-15	CTX-M-15 is a beta-lactamase found in the family *Enterobacteriaceae*.	0.78
3002334	GES-5	GES-5 is a beta-lactamase found in the family *Enterobacteriaceae*.	ND
3002392	OXY-1-4	OXY-1-4 is a beta-lactamase found in *K. oxytoca*.	3.11
3002578	AAC(6')-Ib7	AAC(6')-Ib7 is a plasmid-encoded aminoglycoside acetyltransferase in *Enterobacter cloacae* and *Citrobacter freundii*.	ND
3002601	*aadA*	ANT(3")-Ia is an aminoglycoside nucleotidyltransferase gene encoded by plasmids, transposons, integrons in *Enterobacteriaceae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Vibrio cholerae*.	8.95
3002639	APH(3'')-Ib	APH(3")-Ib is an aminoglycoside phosphotransferase encoded by plasmids, transposons, integrative conjugative elements and chromosomes in *Enterobacteriaceae* and *Pseudomonas* spp.	4.67
3002660	APH(6)-Id	APH(6)-Id is an aminoglycoside phosphotransferase encoded by plasmids, integrative conjugative elements and chromosomal genomic islands in *K. pneumoniae*, *Salmonella* spp., *E. coli*, *Shigella flexneri*, *Providencia alcalifaciens*, *Pseudomonas* spp., *V. cholerae*, *Edwardsiella tarda*, *Pasteurella multocida* and *Aeromonas bestiarum.*	5.45
3002714	QnrB1	Plasmid-mediated quinolone resistance protein found in *K. pneumoniae*.	0.39
3002859	*dfrA14*	Integron-encoded dihydrofolate reductase found in *E. coli.*	5.45
3002986	*bacA*	BacA recycles undecaprenyl pyrophosphate during cell wall biosynthesis, which confers resistance to bacitracin.	0.39
3003209	*fosA5*	Fosfomycin resistance gene isolated from clinical strain of *E. coli* E265. It is susceptible to amikacin, tetracycline and imipenem, and resistant to sulphonamide, cephalosporins, gentamicin, ciprofloxacin, chloramphenicol and streptomycin.	93.39
3003578	*pmrF*	Required for the synthesis and transfer of 4-amino-4-deoxy-L-arabinose to Lipid A, which allows Gram-negative bacteria to resist the antimicrobial activity of cationic antimicrobial peptides and antibiotics such as polymyxin.	0.39
3003922	*oqxA*	RND efflux pump conferring resistance to fluoroquinolone.	91.83
3003923	*oqxB*	RND efflux pump conferring resistance to fluoroquinolone.	1.56
3003950	*msbA*	Multidrug resistance transporter homolog from *E. coli* and belongs to a superfamily of transporters that contain an adenosine triphosphate (ATP) binding cassette (ABC) which is also called a nucleotide-binding domain (NBD). MsbA is a member of the MDR-ABC transporter group by sequence homology. MsbA transports lipid A, a major component of the bacterial outer cell membrane, and is the only bacterial ABC transporter that is essential for cell viability.	98.83
3004580	*K. pneumoniae* KpnE	KpnE subunit of KpnEF resembles EbrAB from *E. coli*. Mutation in KpnEF resulted in increased susceptibility to cefepime, ceftriaxon, colistin, erythromycin, rifampin, tetracycline, and streptomycin as well as enhanced sensitivity toward sodium dodecyl sulfate,	98.83

ARO: accession	Name	Definition	Prevalence (%) in *K. oxytoca**
		deoxycholate, dyes, benzalkonium chloride, chlorhexidine and triclosan.	
3004583	*K. pneumoniae* KpnF	KpnF subunit of KpnEF resembles EbrAB from *E. coli*. Mutation in KpnEF resulted in increased susceptibility to cefepime, ceftriaxon, colistin, erythromycin, rifampin, tetracycline, and streptomycin as well as enhanced sensitivity toward sodium dodecyl sulfate, deoxycholate, dyes, benzalkonium chloride, chlorhexidine and triclosan.	99.22
3004588	*K. pneumoniae* KpnG	KpnG consists of ~390 residues and resembles EmrA of *E. coli*. Disruption of the pump components KpnG-KpnH significantly decrease resistance to azithromycin, ceftazidime, ciprofloxacin, ertapenem, erythromycin, gentamicin, imipenem, ticarcillin, norfloxacin, polymyxin-B, piperacillin, spectinomycin, tobramycin and streptomycin.	99.22
3004612	*E. coli ampH*	AmpH is a class C ampC-like beta-lactamase and penicillin-binding protein identified in *E. coli*.	99.22
3004621	AAC(3)-IIe	Plasmid-encoded aminoglycoside acetyltransferase in *E. coli*.	1.95

*ND, no data.

Table 3. Genomes included in analyses conducted by Schneditz *et al.* (10) with corrected species affiliations (originally reported as *K. oxytoca*)

Assembly accession	Strain	Species	ANI with shown genome*	*npsA*/*npsB*
GCA_000240325.1	KCTC 1686	*K. michiganensis*	98.69 %, GCA_901556995.1	–
GCA_000247835.1	10–5242	*K. michiganensis*	97.59 %, GCA_901556995.1	–
GCA_000247855.1	10–5243	*K. oxytoca*	99.31 %, GCA_900977765.1	+
GCA_000247875.1	10–5245	*K. oxytoca*	99.13 %, GCA_900977765.1	–
GCA_000247895.1	10–5246	*Raoultella ornithinolytica*	99.21 %, GCA_001598295.1	–
GCA_000247915.1	10–5250	*K. pasteurii*	99.29 %, GCA_901563825.1	+
GCA_000252915.3	11492-1	*K. oxytoca*	99.15 %, GCA_900977765.1	+
GCA_000276705.2	E718	*K. michiganensis*	98.37 %, GCA_901556995.1	–
GCA_000427015.1	SA2	*K. grimontii*	99.33 %, GCA_900200035.1	+
GCA_001078235.1	10–5248	*K. oxytoca*	99.25 %, GCA_900977765.1	+
GCA_001633115.1	M5a1	*K. grimontii*	99.40 %, GCA_900200035.1	+

*GCA_901556995.1 = *K. michiganensis* W14^{T}; GCA_900200035.1, *K. grimontii* 06D021^{T}; GCA_900977765.1 = *K. oxytoca* ATCC 13182^{T}; GCA_001598295.1 = *R. ornithinolytica* NBRC 105727^{T}; *K. pasteurii* SB3355^{T} GCA_901563825.1.

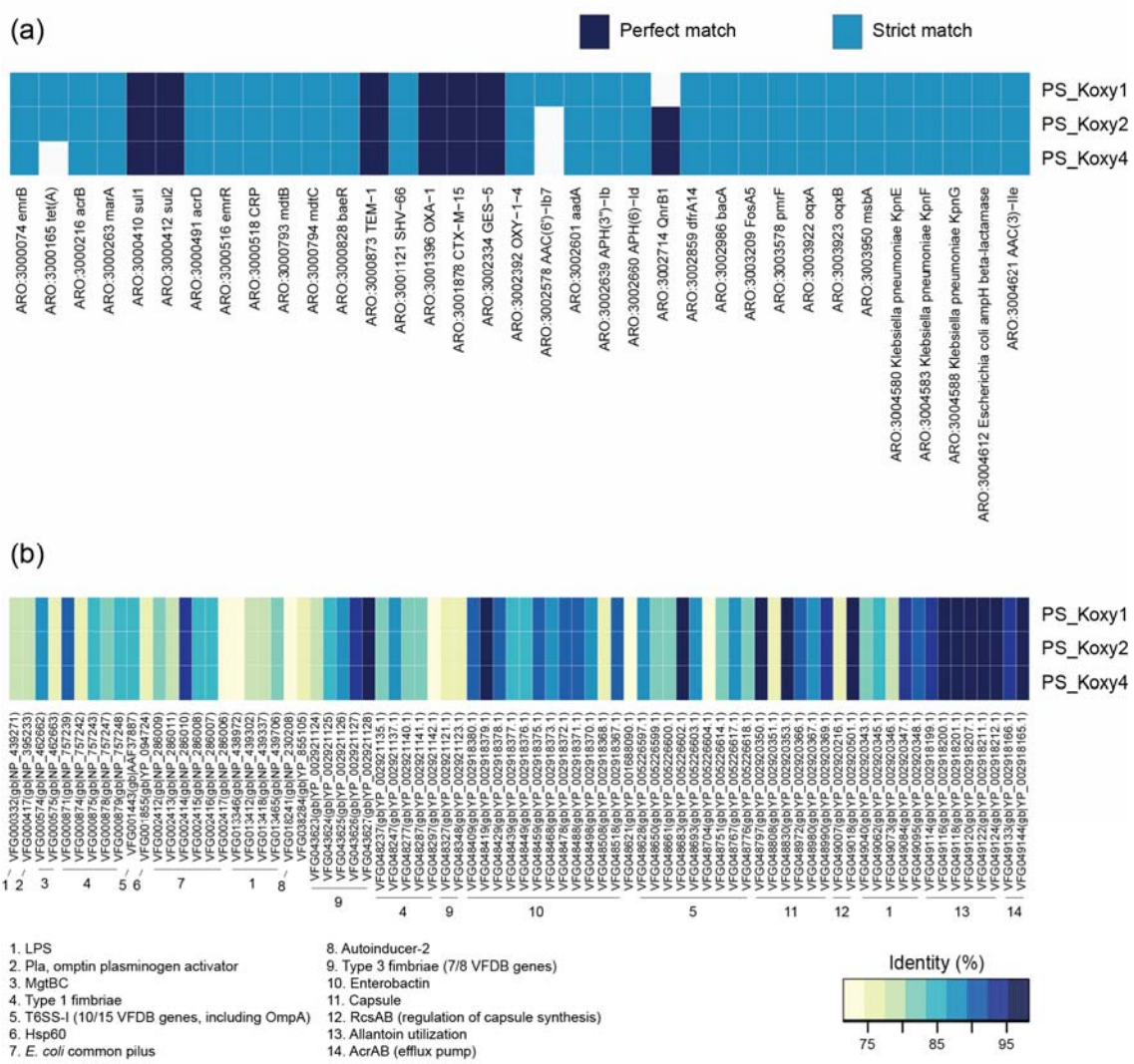


Figure 1. Antimicrobial resistance (a) and virulence factor (b) genes detected in the genomes of the GES-5-positive *K. michiganensis* strains. (a) Protein sequences encoded within the strains' genomes were compared against sequences in CARD (27). Strict CARD match, not identical but the bit-score of the matched sequence is greater than the curated BLASTP bit-score cut-off; perfect CARD match, 100 % identical to the reference sequence along its entire length. (b) Protein sequences encoded within the strains' genomes were compared against sequences in VFDB (28). Only BLASTP results for proteins sharing >70 % identity and 90 % query coverage with VFDB protein sequences are shown.

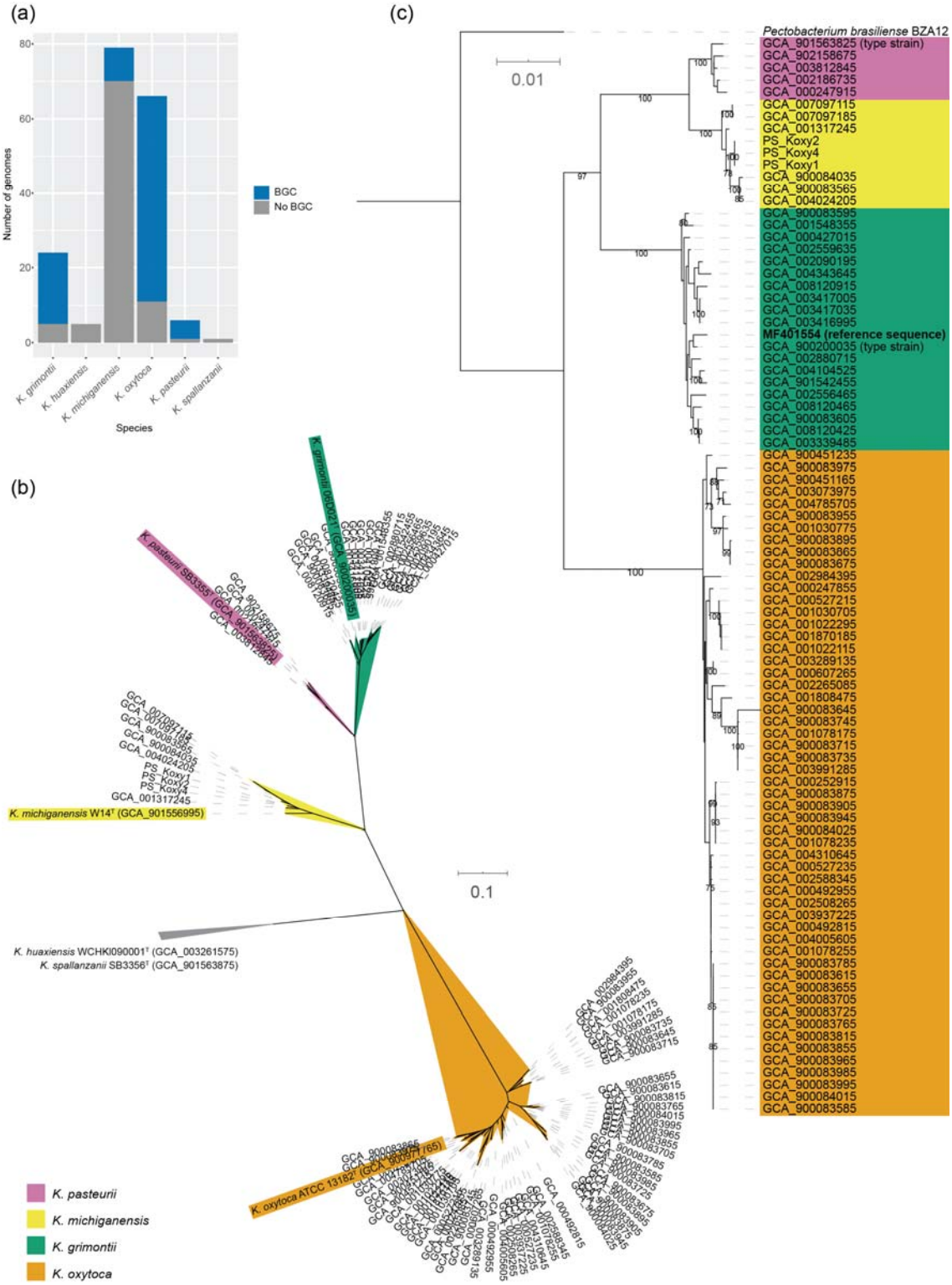
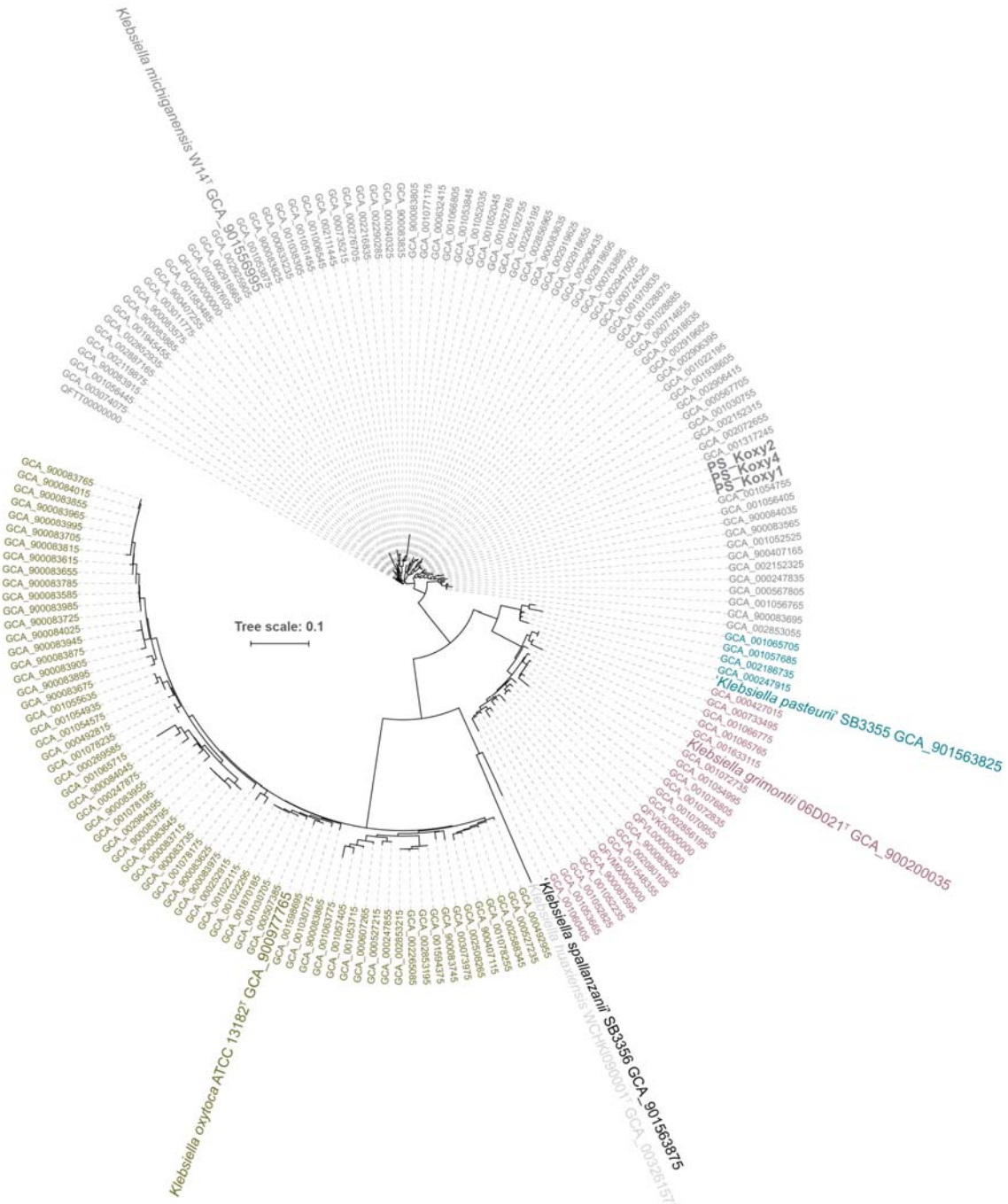


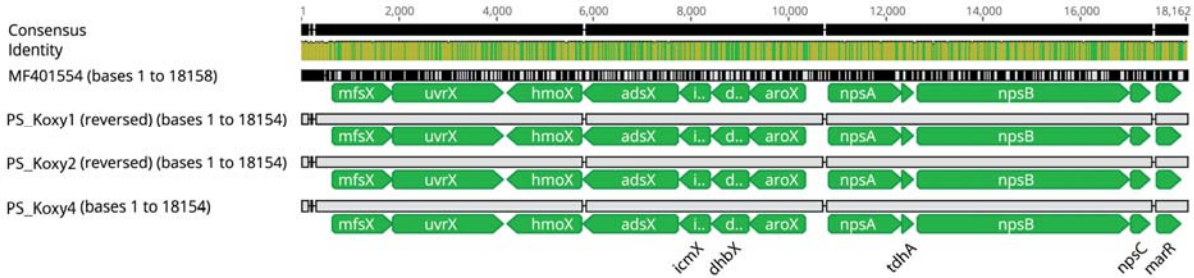
Figure 2. Distribution of the kleboxymycin BGC in *Klebsiella* spp. genomes. (a) Distribution of the *K. oxytoca* complex genomes encoding the entire kleboxymycin BGC. (b) Unrooted maximum-likelihood tree [generated using PhyloPhlAn (22) and 380 protein-encoding sequences from each

genome] confirming species affiliations of the 88 genomes within the *K. oxytoca* complex (7) encoding the kleboxymycin BGC. Type strains are shown with coloured backgrounds corresponding to the legend. The clade associated with *K. huaxiensis* and *K. spallanzanii* has been collapsed because of space constraints. (c) Maximum-likelihood tree generated with the concatenated protein sequences for the kleboxymycin BGC of the 88 genomes found to encode all 12 genes of the BGC plus the reference sequence (9). The tree was rooted using the kleboxymycin-encoding BGC of *Pectobacterium brasiliense* BZA12. Values at nodes, bootstrap values expressed as a percentage of 100 replicates. (b, c) Scale bar, average number of substitutions per position.



Supplementary Figure 1. Phylogenetic tree showing the relationship of the three GES-5-positive clinical isolates (PS_Koxy1, PS_Koxy2, PS_Koxy4) and the 167 strains included in our previous study (1), plus type strains of species of the *K. oxytoca* complex (7).

(a)

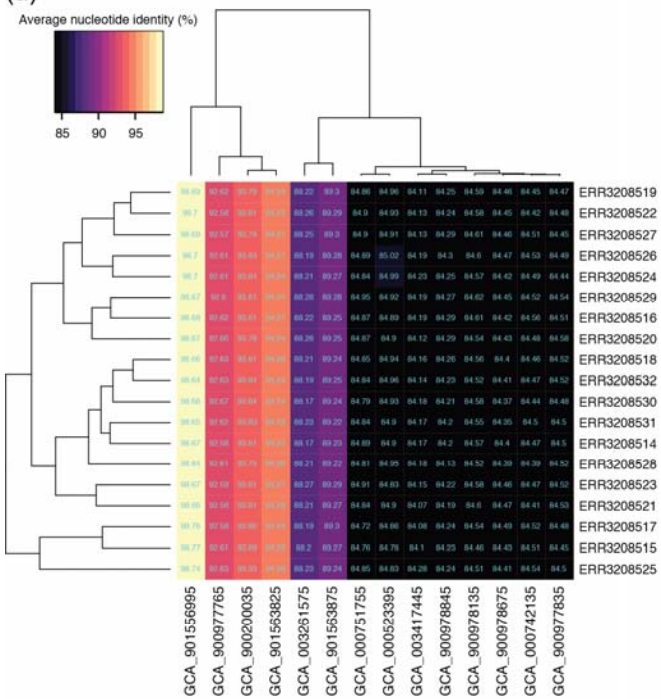


(b)

	mfsX	*uvrX*	*hmoX*	*adsX*	*icmX*	*dhbX*	*aroX*	*npsA*	*thdA*	*npsB*	*npsC*	*marR*
PS_Koxy1 ORF	PSKoxy1_03008	PSKoxy1_03007	PSKoxy1_03006	PSKoxy1_03005	PSKoxy1_03004	PSKoxy1_03003	PSKoxy1_03002	PSKoxy1_03001	PSKoxy1_03000	PSKoxy1_02999	PSKoxy1_02998	PSKoxy1_02997
PS_Koxy2 ORF	PSKoxy2_03101	PSKoxy2_03100	PSKoxy2_03099	PSKoxy2_03098	PSKoxy2_03097	PSKoxy2_03096	PSKoxy2_03095	PSKoxy2_03094	PSKoxy2_03093	PSKoxy2_03092	PSKoxy2_03091	PSKoxy2_03090
PS_Koxy4 ORF	PSKoxy4_04877	PSKoxy4_04878	PSKoxy4_04879	PSKoxy4_04880	PSKoxy4_04881	PSKoxy4_04882	PSKoxy4_04883	PSKoxy4_04884	PSKoxy4_04885	PSKoxy4_04886	PSKoxy4_04887	PSKoxy4_04888

Supplementary Figure 2. Alignment of the kleboxymycin BGCs from the three clinical *K. michiganensis* strains with the complete cluster of *K. oxytoca* MH43-1 [GenBank accession number MF401554 (9)]. (a) The image (alignment view) was generated via the progressiveMauve algorithm plugin of Geneious Prime v2019.2.1 (default settings, full alignment), with gene names for the three clinical isolates assigned manually. (b) Genes corresponding to Prokka-generated annotations. Consensus identity is the mean pairwise nucleotide identity over all pairs in the column: green, 100 % identity; greeny-brown, at least 30 % and under 100 % identity; red, below 30 % identity.

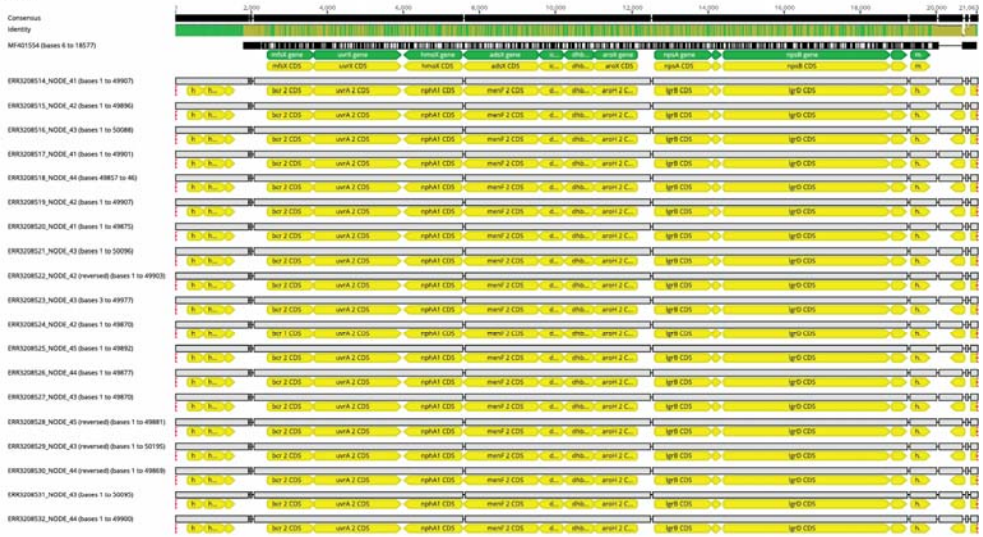
(a)



(b)

SRA accession	Size (bp)	No. contigs	N50	CDS
ERR3208514	6,302,966	335	114,717	5,880
ERR3208515	6,262,953	328	118,337	5,837
ERR3208516	6,287,539	341	124,824	5,862
ERR3208517	6,450,015	326	124,100	6,110
ERR3208518	6,317,731	331	118,589	5,901
ERR3208519	6,323,285	334	124,824	5,900
ERR3208520	6,323,567	341	114,717	5,898
ERR3208521	6,325,218	336	122,308	5,900
ERR3208522	6,329,222	336	124,824	5,911
ERR3208523	6,326,723	364	124,861	5,898
ERR3208524	6,315,651	362	103,560	5,896
ERR3208525	6,241,566	312	106,994	5,818
ERR3208526	6,329,241	362	114,717	5,904
ERR3208527	6,330,129	361	114,398	5,904
ERR3208528	6,322,073	380	114,717	5,899
ERR3208529	6,346,312	543	105,659	5,896
ERR3208530	6,309,195	402	114,398	5,887
ERR3208531	6,330,172	381	120,681	5,907
ERR3208532	6,319,562	337	124,861	5,900

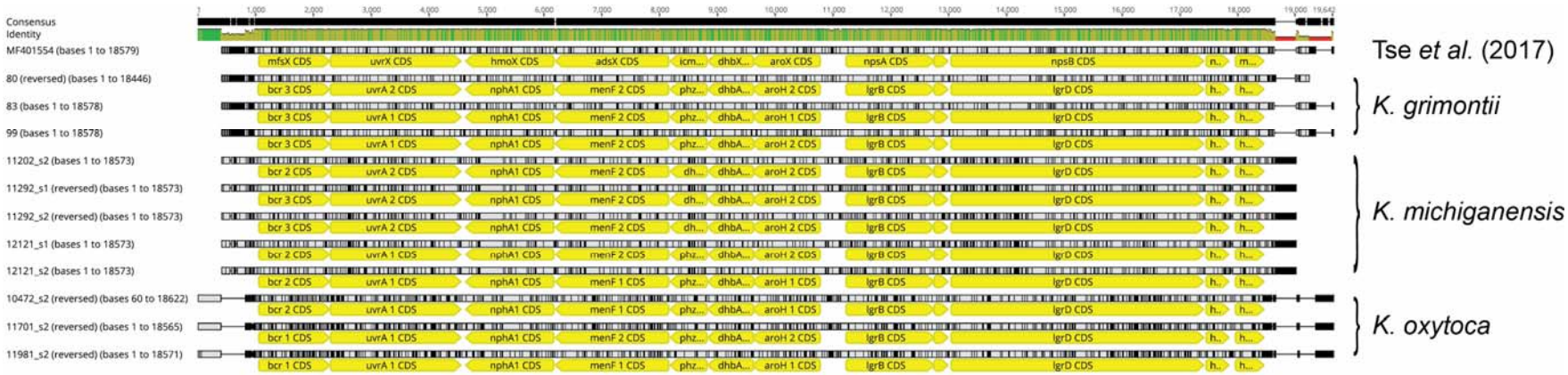
(c)



740

Supplementary Figure 3. Species identification of ST138 isolates described by Ellington et al. (35) and detection of the kleboxymycin BGC within their genomes. (a) Bidirectional clustered heatmap showing that all ST138 isolates described recently are *K. michiganensis*, not *K. oxytoca*, sharing 98.64–98.77 % ANI with GCA_901556995 (*K. michiganensis* W14^{T}). (b) Assembly statistics for the genomes assembled in this study with the Sequence Read Archive (SRA) accession numbers associated with the raw data. (c) The kleboxymycin BGC is present in all 19 of the *K. michiganensis* ST138 isolates of Ellington et al. (35). The image (alignment view) was generated via the progressiveMauve algorithm plugin of Geneious Prime v2019.2.1 (default settings, full alignment). Prokka-assigned gene annotations have been left for the *de novo*-assembled genomes. Consensus identity is the mean pairwise nucleotide identity over all pairs in the column: green, 100 % identity; greeny-brown, at least 30 % and under 100 % identity; red, below 30 % identity.

ERR3208533 was also assembled but was found to be *K. oxytoca* and lacking the kleboxymycin BGC (data not shown, but assembly available from figshare).



Supplementary Figure 4. Detection of the kleboxymycin BGC in strains and MAGs recovered from the faecal microbiota of preterm infants. The strains and MAGs were characterized by us previously (1). The image (alignment view) was generated via the progressiveMauve algorithm plugin of Geneious Prime v2019.2.1 (default settings, full alignment). Prokka-assigned gene annotations have been left for the infant-associated strains and MAGs. Two pairs of the *K. michiganensis* MAGs came from the same infants (11292, 12121) sampled at two different time points. Consensus identity is the mean pairwise nucleotide identity over all pairs in the column: green, 100 % identity; greeny-brown, at least 30 % and under 100 % identity; red, below 30 % identity.