SCIENTIFIC REPORTS

OPEN

Received: 5 October 2017
Accepted: 14 December 2017
Published online: 10 January 2018

Transfer of a *bla*$_{CTX-M-1}$-carrying plasmid between different *Escherichia coli* strains within the human gut explored by whole genome sequencing analyses

Per Kristian Knudsen[1,2], Karianne Wiger Gammelsrud[2,3], Kristian Alfsnes[4], Martin Steinbakk[5], Tore G. Abrahamsen[1,2], Fredrik Müller[2,3] & Jon Bohlin[6]

Horizontal transfer of antibiotic resistance determinants contributes to dissemination of antibiotic resistance. Such transfer of resistance genes within the human gut has been documented in some *in vivo* studies. The present study investigated seven *bla*$_{CTX-M-1}$-carrying *Escherichia coli* isolates from three consecutive faecal samples collected from one cystic fibrosis patient in a nine-months period, by analysing whole genome sequencing data. The analyses showed that the seven *E. coli* isolates represented three genetically diverse strains. All isolates contained *bla*$_{CTX-M-1}$-carrying IncI1 plasmids that shared a common 101 kb backbone differing by only four SNPs. The plasmids harboured by the three different *E. coli* strains varied within limited regions suggestive of recombination events, according to the phylogenetic topology of the genomes of the isolates harbouring them. The findings strongly suggest that horizontal transfer of a *bla*$_{CTX-M-1}$-carrying plasmid had occurred within the patient´s gut. The study illustrates the within-host diversity of faecally carried resistant *E. coli* isolates and highlights the value of collecting multiple bacterial colonies from longitudinally collected samples to assess faecal carriage of resistant enterobacteria. The clustering of the plasmids with the corresponding *E. coli* strains carrying them indicates that the plasmids appear to have adapted to their respective *E. coli* hosts.

During the last 10 to 15 years a global epidemic of extended-spectrum beta-lactamase (ESBL)-producing enterobacteria has emerged[1]. The CTX-M enzymes constitute by far the most common group of ESBLs[1–3]. The *bla*$_{CTX-M}$ genes encoding the CTX-M enzymes are probably originally mobilized from the chromosomes of various species of the *Kluyvera* genus to plasmids well adapted particularly to *Escherichia coli*[4]. Spread of CTX-M enzymes has to some extent occurred through dissemination of successful, virulent clones, e.g. *E. coli* ST131. According to Woerther *et al.* a huge diversity of *bla*$_{CTX-M}$-carrying *E. coli* strains is present in the community with each carrier mostly harbouring his/her own strain that may disseminate in the immediate surroundings[1]. Previous studies of faecal carriage of ESBL-producing *E. coli* have demonstrated dissemination between household members[5,6]. Molecular characterization of the bacterial isolates revealed clonally related *E. coli* strains carrying the same ESBL genotype, but also genetically different strains carrying the same ESBL genotype[5]. Other recent studies have

[1]Department of Paediatric Medicine, Division of Paediatric and Adolescent Medicine, Oslo University Hospital, PB 4956 Nydalen, 0424, Oslo, Norway. [2]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, PB 1171 Blindern, 0318, Oslo, Norway. [3]Department of Microbiology, Division of Laboratory Medicine, Oslo University Hospital, PB 4956 Nydalen, 0424, Oslo, Norway. [4]Department of Molecular Biology, Domain of Infection Control and Environmental Health, Norwegian Institute of Public Health, PB 4404 Nydalen, 0403, Oslo, Norway. [5]Department of Antibiotic Resistance and Infection Prevention, Domain of Infection Control and Environmental Health, Norwegian Institute of Public Health, PB 4404 Nydalen, 0403, Oslo, Norway. [6]Department of Methodology Research and Analysis, Domain of Infection Control and Environmental Health, Norwegian Institute of Public Health, PB 4404 Nydalen, 0403, Oslo, Norway. Correspondence and requests for materials should be addressed to P.K.K. (email: p.k.knudsen@medisin.uio.no)

	Faecal samples		
	Sample 1 (time 0)	Sample 2 (after 6 months)	Sample 3 (after 9 months)
E. coli phenotypes isolated	433-tz (SXT resistant)	432-AT128 (SXT resistant)	431-Ts-o (SXT resistant)
	433-at (SXT and TET resistant)	432-CT-a-32 (SXT and TET resistant)	431-Ts-lys (SXT and TET resistant)
			431-Tz (No additional resistance)

Table 1. Seven extended-spectrum beta-lactamase (ESBL)-producing *E. coli* isolates from three faecal samples collected from one patient at three different time points. Additional phenotypic resistance detected in each isolate is shown in the parentheses. SXT, trimethoprim-sulfamethoxazole; TET, tetracycline.

shown extensive within-host diversity in faecally carried ESBL-producing *E. coli* isolates with identical bla_{CTX-M} variants detected in genetically different *E. coli* isolates from the same individual[7,8]. This is suggestive of transfer of resistance genes between different strains of *E. coli*. The gut of humans and animals is the main reservoir of *E. coli* and other enterobacteria and may constitute a suitable environment for horizontal gene transfer (HGT)[9]. The large number of microorganisms, more than 10^{13} bacterial cells residing in the gut[10], and probably also biofilm formation contribute to high bacterial density and direct contact between bacterial cells, thus facilitating HGT within the gastrointestinal tract[9,11]. A common mechanism of HGT is by conjugation with plasmids acting as vectors. Numerous studies have documented conjugational transfer of resistance genes between bacteria *in vitro* and also in some animal models[12–14]. However, not many *in vivo* studies have actually proven HGT of resistance determinants within the human gut[15–18].

Whole genome sequencing (WGS) of bacteria offers vast opportunities for detailed characterisation of bacterial isolates, including detection and classification of plasmids[19,20]. However, assembly of plasmids from WGS data generated by high-throughput sequencing technologies that produce short reads (≈100–300 bp) might be challenging[20]. Not many studies have so far used *in silico* plasmid extraction from bacterial WGS data to study *in vivo* HGT in the human gut[21,22].

In a study of faecal carriage of resistant enterobacteria in children[23] we identified the plasmid-carried $bla_{CTX-M-1}$ in seven different *E. coli* isolates from three consecutive faecal samples from one patient. The seven *E. coli* isolates represented three different phenotypes according to their trimethoprim-sulfamethoxazole and tetracycline susceptibility patterns. The aim of this study was to assess the genetic relationship between these seven *E. coli* isolates and to further characterise their $bla_{CTX-M-1}$-carrying plasmids by reconstructing them *in silico* from WGS data. Specifically, we wanted to explore whether $bla_{CTX-M-1}$ was located on a common plasmid in all the isolates.

Methods

Ethics. The study was approved by the Regional Committee for Medical and Health Research Ethics – South East ("REK sør-øst")(reference number 581–06–03092) and the study was conducted in accordance with the regulations from the committee and the principles of the Declaration of Helsinki. Written, informed consent was obtained from the participating child´s parents.

Collection and characteristics of the *E. coli* isolates and patient information. In a study of the prevalence of resistant intestinal *Enterobacteriaceae* in children[23], one of the participants, a child with cystic fibrosis, submitted three faecal samples with time intervals of six months and three months, respectively. The faecal samples were analysed with a Direct MIC-gradient Strip Method to detect resistant *Enterobacteriaceae*[24]. Two phenotypically different ESBL-producing *E. coli* isolates were detected both in sample one and sample two: one of the ESBL-producing *E. coli* isolates was resistant to trimethoprim-sulfamethoxazole, whereas the other was resistant to both trimethoprim-sulfamethoxazole and tetracycline. In sample three, a third ESBL-producing *E. coli* phenotype, susceptible to all non-beta-lactam antibiotics, was detected in addition to the two other phenotypes. For all of the three samples, one of each ESBL-producing *E. coli* phenotype was selected for WGS, in total seven isolates as shown in Table 1.

The patient had been diagnosed with cystic fibrosis at Oslo University Hospital three years prior to the first faecal sample. Her airways were chronically colonized with *Haemophilus influenzae* and *Staphylococcus aureus*. In addition, *Stenotrophomonas maltophilia* and *Achromobacter xylosoxidans*, but not *Pseudomonas aeruginosa*, had been cultured sporadically. At CF diagnosis she was treated with ceftazidime, tobramycin and cloxacillin intravenously for 14 days. After this she had not been treated with any third-generation cephalosprins or carbapenems, but she had received repeated oral antibiotic courses, mainly trimethoprim-sulfamethoxazole, but also cephalexin, cloxacillin and amoxicillin. She also received several courses with these antibiotics during the study period.

Whole genome sequencing (WGS). Genomic DNA was extracted using MagNa Pure 96 (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. DNA concentrations were measured using a Qubit fluorometer (Thermo Fisher Scientific, MA, USA) to determine DNA input. Isolates 431-Ts-o, 431-Ts-lys, 431-Tz, 433-tz, 433-at: Genomic libraries were prepared using KAPA HyperPlus Library Preparation Kit (Kapa Biosystems, MA, USA) and WGS was performed on an Illumina MiSeq platform using v2 reagent kits generating 2 × 250 bp paired-end reads (Illumina, San Diego, CA, USA) at the Norwegian Institute of Public Health, Oslo, Norway. Isolates 432-AT128 and 432-CT-a-32: Genomic libraries were prepared and WGS was performed on an Illumina HiSeq 2500 platform generating 2 × 125 bp paired-end reads at BGI Tech Solutions Co., Limited, Hong Kong.

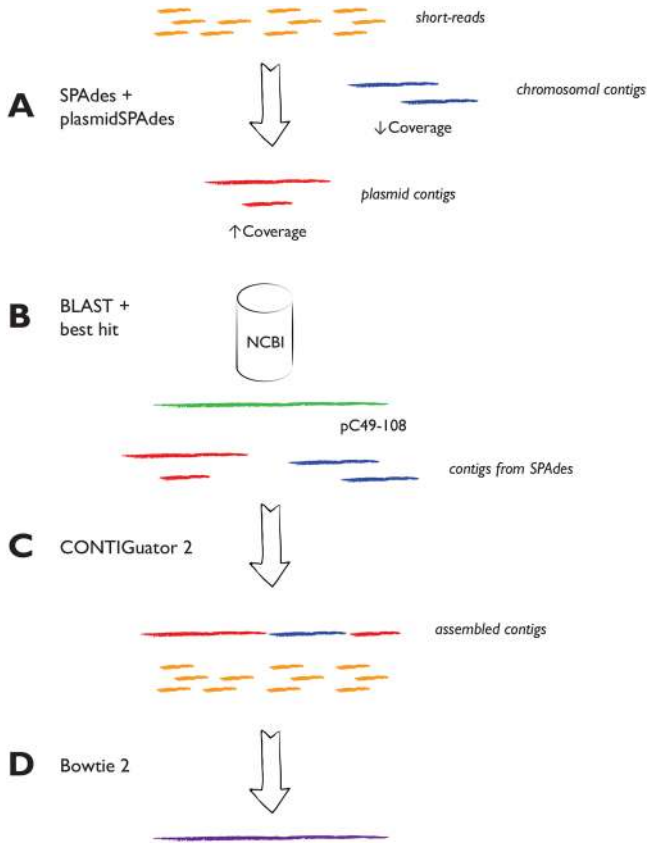

Figure 1. The plasmid assembly method. Short reads from each isolate are de novo assembled using SPAdes, and putative plasmid contigs are assembled and extracted by plasmidSPAdes (**A**). The putative plasmid contigs are subsequently queried against the NCBI plasmid database and if suitable matches are detected, the best hit is used as a reference sequence (**B**). The assembled plasmid sequences obtained from mapping to the reference plasmid is then used as templates for which contigs from the corresponding *de novo* assembled *E. coli* genomic sequences are sorted against, using CONTIGuator2 (**C**). The contigs that have a 99% nucleotide identity as well as consisting of more than 10% of reference plasmid DNA are retained as putative plasmid contigs while those falling short of the criteria are excluded. Finally, the newly assembled plasmid contigs are used as a reference to which all the original short reads from each *E. coli* isolate are mapped using BowTie2 (**D**).

Assembly of genomes and $bla_{CTX\text{-}M\text{-}1}$-carrying plasmids. The sequenced reads were trimmed for adapter sequences using trimmomatic[25] and subsequently quality corrected using BayesHammer[26]. Whole genome *de novo* assembly was carried out using SPAdes v 3.9.0[27], and PlasmidSPAdes[28] was used to assemble and extract putative plasmid sequences from the sequenced reads (Fig. 1A). PlasmidSPAdes was run on all isolates, and the predicted plasmid contigs from the isolate having the largest contigs (isolate 431-Ts-lys), *i.e.* the largest contiguous predicted plasmid contigs also containing $bla_{CTX\text{-}M\text{-}1}$, were BLASTed[29] against the NCBI databases restricted to bacterial plasmids (Fig. 1B). The plasmid with the best hit (pC49–108, Accession number KJ484638) in terms of bit score and size was used as a reference for a subsequent plasmid assembly. This assembly was performed as follows: first, the plasmid-sequence of pC49–108 was used as a reference for the assembly of the putative plasmid sequences from each of the seven isolates. Second, each of the resulting assembled plasmid sequences were further used as templates of which contigs from the corresponding *de novo* assembled *E. coli* genomic sequences were sorted against, using CONTIGuator v 2.7[30] (Fig. 1C). Contigs were assessed using the Artemis Comparison Tools[31] included in the CONTIGuator package. The contigs with a 99% nucleotide identity as well as they consisted of more than 10% of reference plasmid DNA were retained as putative plasmid contigs while those falling short of the criteria were excluded. This procedure was carried out to fill in genetic regions not found in the reference plasmid. The contig-assembled plasmid was in turn used as a reference to which reads from each *E. coli* isolate was mapped using both BowTie2 v 2.1.0[32] and SAMtools v 1.3.1[33] (Fig. 1D). Assembly details of the *E. coli* genomes can be found in Supplementary Table S1.

Detection of resistance genes and typing of the *E. coli* isolates. The assembled *E. coli* genomes were submitted to the web-based ResFinder service v 2.1 (Center for Genomic Epidemiology, DTU, Denmark) to identify acquired resistance genes[34]. Hits were accepted for matches with ≥99% nucleotide identity between the resistance gene in the database and the corresponding sequence in the genome, and the length of the query sequence covering ≥95% of the length of the gene in the database. Multilocus sequence typing (MLST) and serotyping of

the *E. coli* isolates were performed *in silico* from the assembled genomes by MLST v 1.8 and SerotypeFinder v 1.1 (Center for Genomic Epidemiology, DTU, Denmark), respectively, using default settings[35,36].

Phylogenetic analyses of the *E. coli* genomes. After *de novo* assembly, phylogenetic analyses were performed on the single nucleotide polymorphisms (SNPs) from the complete *E. coli* genome sequences with the putative plasmid sequences removed. The plasmid sequences were excluded by mapping all reads from each *E. coli* genome to the *E. coli* K-12 strain MG1655 chromosome (Accession number NC_000913.3) with BowTie 2 v 2.1.0[32]. Further, SNP calling was carried out using parSNP v 1.2 from the Harvesttools suite[37] by extracting the core genome, containing both coding and non-coding regions, and removal of putative recombinant regions with Gubbins v 2.2[38]. The phylogenetic analysis was performed using the RAxML maximum likelihood estimation program v 8.2.4[39]. The tree was bootstrapped 500 times and the nucleotide substitution matrix (GTR) was found with PhyML v 3.0[40] and the R-package 'ape'[41] (https://www.R-project.org/). The tree was obtained using FigTree v 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Characterization and comparison of the plasmids. All putative plasmid sequences within the genome of each isolate were also identified with PlasmidFinder v 1.3 (Center for Genomic Epidemiology, DTU, Denmark)[19] using the *Enterobacteriaceae* database with the detection threshold set to > 95% sequence identity. Characterization of the assembled $bla_{\text{CTX-M-1}}$-containing plasmids by incompatibility (Inc) group and plasmid MLST (pMLST) was performed with PlasmidFinder v 1.3 and pMLST v 1.4 (Center for Genomic Epidemiology, DTU, Denmark), respectively[19]. SNP calling of the assembled $bla_{\text{CTX-M-1}}$-containing plasmids was performed with BowTie2 v 2.1.0[32] and SAMtools[33], but the low number of SNPs (four) did not allow for a reliable phylogenetic analysis and was therefore omitted. The core backbone of the plasmids was determined with both Roary v 3.8.0[42] (default settings) and ParSNP (fraction of coding/non-coding DNA shared amongst all plasmids). The gene presence-absence heat map of the plasmids was created with R v 3.3.3 from the Roary output and based on complete linkage hierarchical clustering with Euclidean distance. Comparison of all the assembled $bla_{\text{CTX-M-1}}$-containing plasmids was performed and visualized using MAUVE[43]. All the plasmids were annotated using Prokka v 1.11[44]. Annotation of the plasmids was additionally improved by BLASTing of "hypothetical proteins" to NCBI databases.

Data availability. The *E. coli* genome short reads that were generated and analysed in this work are deposited to the European Nucleotide Archive (ENA), and the accession numbers of the seven genomes are shown in Supplementary Table S1. The fasta-files containing the seven assembled plasmid sequences are also available in the Supplementary Information.

Results

The *E. coli* isolates. Analyses of the assembled *E. coli* genomes with respect to acquired resistance genes by ResFinder (Center for Genetic Epidemiology, DTU) detected $bla_{\text{CTX-M-1}}$ in all seven *E. coli* isolates. The SNP-based phylogenetic tree presented in Fig. 2 shows that the seven isolates represent three different strains; strain 1: isolates 433-tz, 432-AT128, 431-Ts-o; strain 2: isolates 433-at, 432-CT-a-32, 431-Ts-lys; strain 3: isolate 431-Tz. The results of *in silico* typing (MLST and serotyping) of each isolate are shown in Table 2.

The *E. coli* plasmids. *In silico* analyses of the genomes from the seven *E. coli* isolates by PlasmidFinder detected seven different plasmid replicons in the three isolates representing strain 1: IncI1, IncFII (pRSB107), IncQ1, IncFIA, IncFIB (AP001918), Col8282, ColpVC (Table 2). This indicates that these three isolates contain several (up to seven) different plasmids. In addition, another IncFII plasmid replicon was found in one of the strain 1 isolates (isolate 433-tz). The three isolates representing strain 2 contained three plasmid replicons: IncI1, IncFIA, Col (MG828), whereas isolate 431-Tz (strain 3) contained one IncI1 replicon only (Table 2).

All the assembled $bla_{\text{CTX-M}}$ carrying plasmids from the seven *E. coli* isolates were IncI1 plasmids representing sequence type ST3. A Mauve comparison indicates that all the assembled plasmid sequences were almost identical, but with some differences among them located in delimited variable regions (Fig. 3).

All seven IncI1 plasmids shared a common 101 kb nucleotide sequence, which constitutes a core backbone of the complete 115 kb plasmid sequence (88% coverage). A corresponding SNP analysis based on the plasmid sequences (including the reference pC49–108) resulted in only four SNP differences within the backbone. Genes associated with conjugation (e.g. *traC* and *pil* genes) were detected within this common backbone. A list of annotated genes found in each plasmid is shown in Supplementary Table S2.

Comparison of the annotated plasmids with respect to the variable regions (nucleotide sequences not shared by all the plasmids) showed that these regions contained different genes, including genes encoding transposases and integrases. All three plasmids from the isolates belonging to strain 2 (the most complete plasmids) contained genes encoding TnpA transposase, transposase for Tn21 transposon, TnpR resolvase, and an integrase/recombinase (Supplementary Table S2). Corresponding genes were not found in the four plasmids from the other two *E. coli* strains (strain 1 and 3) that contained larger deletion blocks (Fig. 3).

Due to the close nucleotide similarity between the plasmids we created a gene presence- absence heat map of all plasmids in the present study, including the pC49–108 reference sequence, consisting of genes not included in the core backbone. From Fig. 4 it can be seen that the plasmids from *E. coli* strains 1 and 2 cluster according to their corresponding strain, while the plasmid from strain 3 clusters closely with the GenBank reference plasmid.

A further interrogation of the plasmids available in public repositories that resemble those in the present study the most, indicates that pC49–108, and to a lesser degree pH2291–112 (Accession number KJ484629), pC59–112 (Accession number KJ484637), pC60–108 (Accession number KJ484635) are the closest. These are all IncI1/ST3 plasmids harbouring $bla_{\text{CTX-M-1}}$ detected in *E. coli* isolates from humans and chicken[45].

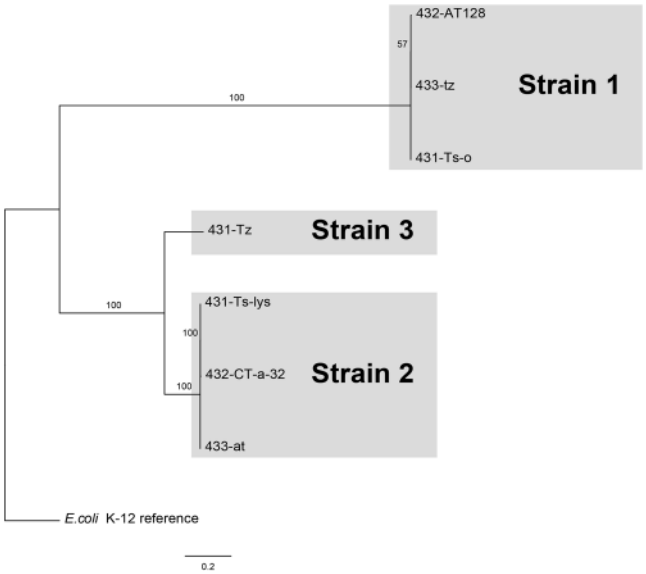

Figure 2. Maximum Likelihood based phylogenetic analysis of core genome SNPs from seven $bla_{CTX\text{-}M\text{-}1}$-containing *E. coli* isolates, with the reference *E. coli* K-12 strain MG1655 (Accession number NC_000913.3) included as out-group. Branch numbers designate bootstrap support.

Isolate-ID	Sample time	Phenotypic resistance	Strain	Sequence type (MLST)	Serotype	Resistance genes*	Plasmid replicons
433-tz	Study start	3.generation cephalosporins, SXT	1	ST1640	O124:H25	$bla_{CTX\text{-}M\text{-}1}$, *dfrA14*, *dfrA17*, *sul2*	IncI1, IncFII (pRSB107), IncQ1, IncFIA, IncFIB (AP001918), Col8282, ColpVC, IncFII
432-AT128	6 months	3.generation cephalosporins, SXT	1	ST1640	O124:H25	$bla_{CTX\text{-}M\text{-}1}$, *dfrA14*, *dfrA17*, *sul2*	IncI1, IncFII (pRSB107), IncQ1, IncFIA, IncFIB (AP001918), Col8282, ColpVC
431-Ts-o	9 months	3.generation cephalosporins, SXT	1	ST1640	O124:H25	$bla_{CTX\text{-}M\text{-}1}$, *dfrA14*, *dfrA17*, *sul2*	IncI1, IncFII (pRSB107), IncQ1, IncFIA, IncFIB (AP001918), Col8282, ColpVC
433-at	Study start	3.generation cephalosporins, SXT, TET	2	ST6331	O82:H21	$bla_{CTX\text{-}M\text{-}1}$, *dfrA17*, *sul1*, *sul2*, *tetA*	IncI1, IncFIA, Col (MG828)
432-CT-a-32	6 months	3.generation cephalosporins, SXT, TET	2	ST6331	O82:H21	$bla_{CTX\text{-}M\text{-}1}$, *dfrA17*, *sul1*, *sul2*, *tetA*	IncI1, IncFIA, Col (MG828)
431-Ts-lys	9 months	3.generation cephalosporins, SXT, TET	2	ST6331	O82:H21	$bla_{CTX\text{-}M\text{-}1}$, *dfrA17*, *sul1*, *sul2*, *tetA*	IncI1, IncFIA, Col (MG828)
431-tz	9 months	3.generation cephalosporins	3	ST2144	O166:H49	$bla_{CTX\text{-}M\text{-}1}$	IncI1

Table 2. Characteristics of seven *E. coli* isolates from faecal samples collected from one cystic fibrosis patient at three different time points. SXT, trimethoprim-sulfamethoxazole; TET, tetracycline. *Resistance genes detected by ResFinder (Centre for Genomic Epidemiology, DTU, Denmark) that explain the observed phenotypic resistance to 3. generation cephalosporins, trimethoprim-sulfamethoxazole and tetracycline.

ResFinder detected no other resistance genes in any of the $bla_{CTX\text{-}M\text{-}1}$ containing IncI1 plasmids from the seven isolates. This is in accordance with the phenotypic resistance profile of the *E. coli* isolate 431-Tz (susceptible to all the non-beta-lactam antibiotics tested), which harboured no other plasmid replicons than IncI1 (Table 2). *dfrA* genes in combination with *sul* genes were detected by ResFinder in all the six isolates representing strain 1 and strain 2 that were resistant to trimethoprim-sulfamethoxazole. *tet* genes were detected in the three tetracycline resistant strain 2 isolates (Table 2). These genes are probably located on some of the other plasmids detected in these strains (Table 2).

Discussion

In this study we explored seven ESBL-producing *E. coli* isolates and their corresponding $bla_{CTX\text{-}M\text{-}1}$-carrying plasmids isolated from faecal samples from one cystic fibrosis patient at three different time points. The assessment of the isolates and the plasmids was exclusively based on WGS analyses. Assembly of plasmids from WGS short reads is challenging[20], and complete reconstruction is often not possible due to the presence of repeat sequences,

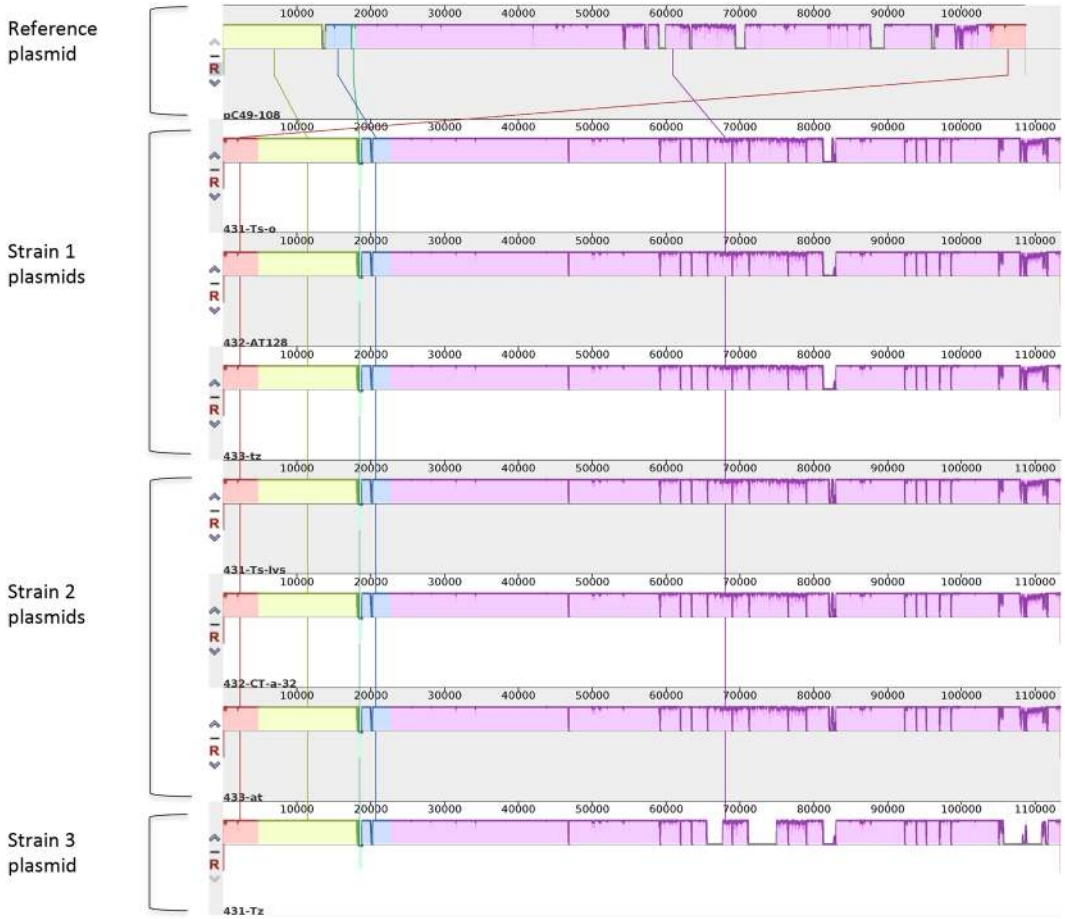

Figure 3. A Mauve comparison of seven $bla_{CTX\text{-}M\text{-}1}$-containing IncI1/ST3 plasmids extracted *in silico* from seven *E. coli* isolates representing three different strains, and a reference plasmid (pC49–108, accession no. KJ484638). The height of the coloured segments indicates the degree of similarity between the plasmids where low height/missing region designates low similarity or absence of sequence. The label of each plasmid is located below the corresponding plasmid sequence.

especially in large plasmids (> 50 kbp)[46]. Here we aimed to reconstruct plasmid sequences by combining *de novo* assembly and reference-based read mapping relying on the successful detection of a reference plasmid (pC49–108) in the NCBI databases with a high degree of similarity with our plasmid sequences.

All seven ESBL-producing *E. coli* isolates carried $bla_{CTX\text{-}M\text{-}1}$, but the isolates represented three different sequence types and serotypes. The phylogenetic analyses (Fig. 2) confirmed that this was in accordance with three genetically diverse groups of isolates that we designated strain 1, 2 and 3 (Table 2).

All seven $bla_{CTX\text{-}M\text{-}1}$-carrying plasmids belonged to the same Inc group (IncI1) and represented the same pMLST (ST3). They shared a common 101 kb nucleotide core backbone containing only four SNP differences. This high degree of similarity between the plasmid backbones, also visualized by the Mauve comparison in Fig. 3, strongly suggests that they originate from a common plasmid that may have been transferred between the different *E. coli* strains. The environment within the human gut is favourable for horizontal gene transfer[9,47], and the transfer of the $bla_{CTX\text{-}M\text{-}1}$-containing plasmid is likely to have happened within the patient's gut. It is well known that plasmids belonging to the IncI1 group often carry ESBL-coding genes and that they have the capacity of conjugational transfer[48]. Genes associated with conjugation (e.g. *traC* and *pil* genes, shown in Supplementary Table S2) were detected within the common backbone of all our plasmids and support the hypothesis that the plasmids probably have been transferred between different *E. coli* strains within the patient´s gut. The diversity of CTX-M-1 producing *E. coli* isolates found in our study is similar to findings in recent studies that have demonstrated great within-host diversity in faecally carried ESBL-producing *E. coli* isolates[7,8]. Stoesser *et al.* sequenced the genomes of 16 *E. coli* colonies from each of eight faecal samples from different individuals in Cambodia[7]. In two of the eight individuals, the same $bla_{CTX\text{-}M}$ variant occurred in different *E. coli* clones, and/or different $bla_{CTX\text{-}M}$ variants occurred in the same clone. Similar findings were demonstrated by Jørgensen *et al.* in a longitudinal study of faecal carriage of ESBL-producing *E. coli* and *Klebsiella pneumoniae*[8]; ESBL production commonly occurred in diverse strains within the same host, and the same $bla_{CTX\text{-}M}$ variant was detected in genetically different *E. coli* isolates from the same faecal sample. These findings suggest horizontal transfer of $bla_{CTX\text{-}M}$ within the gut, but none of these studies assessed the mobile genetic elements (e.g. plasmids) that may have been involved in

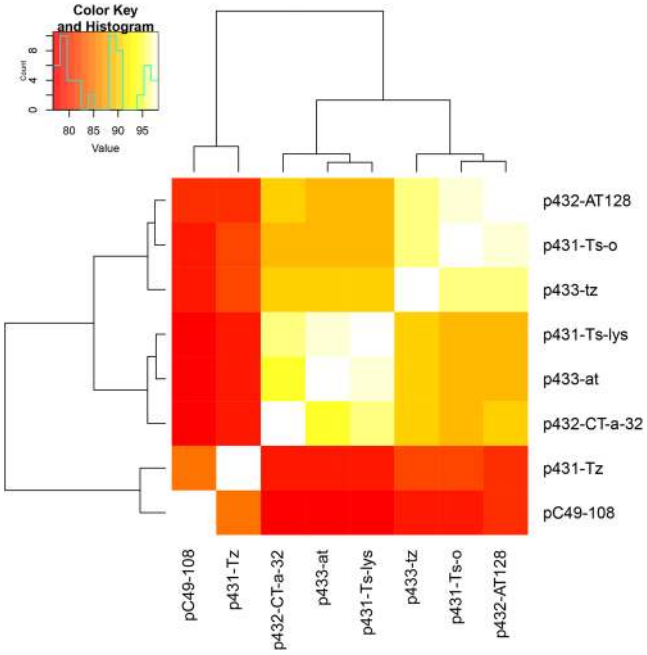

Figure 4. Gene presence-absence heat map of seven $bla_{CTX\text{-}M\text{-}1}$-carrying Incl1/ST3 plasmids extracted from seven *E. coli* isolates and a reference plasmid (pC49–108, accession number KJ484638) obtained from the NCBI database. The key legend top left designates % of the genes common to the plasmids, exempting core backbone genes present in all.

such transmission. Conlan *et al.* studied transmission of plasmids by analysing WGS data obtained from several sequencing platforms. They detected bla_{KPC}-carrying pKpQIL plasmids with a common backbone in several *K. pneumoniae* isolates representing two different STs and in an *E. coli* isolate from faecal samples collected from the same patient at different time points[21]. Horizontal transfer of resistance genes may be facilitated by antibiotic exposure[49], and the extensive antibiotic exposure in our patient may have promoted transfer of $bla_{CTX\text{-}M}$ between different *E. coli* strains. Such antibiotic treatment may also select resistant (e.g. ESBL-producing) strains of enterobacteria and concomitantly supress dominant populations of anaerobic species in the gut, thus promoting cell-to-cell contact between enterobacteria that enhances HGT. Stecher *et al.* showed that gut inflammation could boost HGT between pathogenic and commensal *Enterobacteriaceae*[50]. This may also be of relevance in patients with CF since this disease is associated with intestinal inflammation[51–53].

The $bla_{CTX\text{-}M\text{-}1}$-carrying plasmids contained variable regions with genes encoding transposases and integrases (Supplementary Table S2). Notably, the most complete plasmids, harboured by the strain 2 isolates, contained several genes associated with recombination and transposable elements that were not detected in the plasmids from strains 1 and 3. The plasmids from *E. coli* strains 1 and 3 contained deletion blocks (Fig. 3) that may indicate that recombination events have happened in the plasmids within the different hosts. The phylogenetic SNP analysis of the *E. coli* genomes (Fig. 2), without the putative plasmid sequences, indicates that the plasmids differ according to the phylogenetic topology of the genomes of the isolates harbouring them. The gene presence-absence heat map of the plasmids (Fig. 4) also show that the plasmids cluster according to the three different *E. coli* strains, suggesting that the plasmids may have adapted within their respective hosts. The differences between the groups of clustered plasmids shown in Fig. 4 are most likely due to deleted and/or acquired accessory genes.

The closer clustering of the plasmids discussed above, as compared to the pC49–108 reference, and the large shared backbone of all plasmids considered, may suggest what best can be described as an adaptive ability. It has been demonstrated that plasmids co-evolve with their hosts[54,55], which may result in the loss of genes conferring fitness cost. In this respect it is interesting to note that the strain represented by the single 431-Tz isolate, that contained the plasmid with largest deletion blocks, was isolated at a later time point than the isolates representing the other strains. It is therefore tempting to speculate that the plasmid was introduced to one of these strains and subsequently transferred to the other strains where co-evolution with the respective host in each strain has resulted in deletion events in the plasmids. The heavy antibiotic exposure to the patient may also have promoted the adaptive events in the plasmids within the *E. coli* hosts, as described by Porse *et al.*[54]. In an experimental study they demonstrated adaptive evolution by deletion of a costly region from a multidrug resistant plasmid in *E. coli* during antibiotic exposure, and that the increased plasmid stability was maintained also after the antibiotic selection pressure was removed.

One limitation of the present study is the use of short read sequencing data to reconstruct large plasmids (>100 kb). Several investigators have demonstrated that it may be challenging to obtain correct assemblies of large plasmids from short reads[46,56,57]. The occurrence of multiple repetitive nucleotide sequences (e.g. transposons and IS-sequences) and/or recombination events, both known to be common in plasmids, are demanding obstacles to overcome. For instance, if resistance genes are flanked by repetitive sequences (e.g. transposon elements) that

are longer than the read lengths obtained by short-read based sequencing, it can be difficult to determine the correct genetic context of the resistance genes[57]. Therefore, long-read based sequencing methodology (e.g. Single Molecule Real-time Sequencing (SMRT), Pacific Biosciences or Nanopore sequencing, Oxford Nanopore) may hold great promise in obtaining properly assembled and closed plasmids[20]. Without these methods available, and only relying on the short read sequencing approach used in the present study, there may be inconsistencies regarding the plasmid's variable regions as well as reverse complemented regions, and the extracted plasmids cannot be considered as fully closed. However, all the plasmids in our study cluster according to the *E. coli* strain to which they belong (Fig. 4), although the cluster analysis is based only on genes from the plasmid's variable regions. There may be a possibility that the plasmids extracted by our method are divided between multiple plasmids, which may appear as one contig from the assembled short read data, or that the plasmid backbone has been integrated into different plasmid segments. However, such segments would most likely have been detected during the contig extraction step (Fig. 1C) by inconsistencies between the plasmids of the multiple isolates. We acknowledge that this could be difficult to verify in studies including only single isolates. The consistency of the finding of $bla_{\text{CTX-M-1}}$-carrying Incl1 plasmids sharing a common backbone in two different *E. coli* strains (strain 1 and 2) isolated from three consecutive faecal samples during a nine-months period, strongly supports that a common plasmid is present in the different strains. As Incl1 plasmids are known to have conjugational capacity[48], it is not unlikely that the plasmid has been horizontally transferred between the two strains and even to or from a third strain. However, this cannot be proven due to the methodological limitations described. While the availability of long read sequencing technology is increasing, short read sequencing platforms are still far more commonly available. Thus, extraction of plasmids from short-read data, and methods facilitating this, will probably still be used for some time to come.

In summary, we reconstructed $bla_{\text{CTX-M-1}}$ containing Incl1 plasmids *in silico* from *E. coli* whole genome sequenced short reads by combining *de novo* assembly and reference-based read mapping relying of the successful detection of a reference plasmid in the NCBI databases with a high degree of similarity with our plasmid sequences. Seven $bla_{\text{CTX-M-1}}$-carrying plasmids extracted from isolates that represented three genetically diverse *E. coli* strains isolated from faecal samples from one individual shared a common backbone. The findings strongly suggest horizontal transfer of the $bla_{\text{CTX-M-1}}$-carrying plasmid between different *E. coli* strains within the patient´s gut, although our results need to be interpreted with some caution due to the limitations associated with plasmid assembly from short-read sequencing data. The differences between the plasmids, indicative of recombination events, corresponded to the *E. coli* strain carrying them, indicating that the plasmids appear to have adapted to their respective *E. coli* hosts.

The study illustrates the within-host diversity of faecally carried resistant *E. coli* isolates and highlights the value of collecting several bacterial colonies from longitudinally collected samples to assess faecal carriage of resistant enterobacteria.

References

1. Woerther, P. L., Burdet, C., Chachaty, E. & Andremont, A. Trends in human fecal carriage of extended-spectrum beta-lactamases in the community: toward the globalization of CTX-M. *Clin. Microbiol. Rev.* **26**, 744–758, https://doi.org/10.1128/cmr.00023-13 (2013).
2. Geser, N., Stephan, R., Korczak, B. M., Beutin, L. & Hachler, H. Molecular identification of extended-spectrum-beta-lactamase genes from Enterobacteriaceae isolated from healthy human carriers in Switzerland. *Antimicrob. Agents Chemother.* **56**, 1609–1612, https://doi.org/10.1128/aac.05539-11 (2012).
3. Ruppe, E. *et al.* Clinical predictive values of extended-spectrum beta-lactamase carriage in patients admitted to medical wards. *Eur. J. Clin. Microbiol. Infect. Dis.* **31**, 319–325, https://doi.org/10.1007/s10096-011-1313-z (2012).
4. Canton, R., Gonzalez-Alba, J. M. & Galan, J. C. CTX-M Enzymes: Origin and Diffusion. *Frontiers in microbiology* **3**, 110, https://doi.org/10.3389/fmicb.2012.00110 (2012).
5. Rodriguez-Bano, J., Lopez-Cerero, L., Navarro, M. D., Diaz de Alba, P. & Pascual, A. Faecal carriage of extended-spectrum beta-lactamase-producing Escherichia coli: prevalence, risk factors and molecular epidemiology. *The Journal of antimicrobial chemotherapy* **62**, 1142–1149, https://doi.org/10.1093/jac/dkn293 (2008).
6. Valverde, A. *et al.* High rate of intestinal colonization with extended-spectrum-beta-lactamase-producing organisms in household contacts of infected community patients. *J. Clin. Microbiol.* **46**, 2796–2799, https://doi.org/10.1128/jcm.01008-08 (2008).
7. Stoesser, N. *et al.* Extensive Within-Host Diversity in Fecally Carried Extended-Spectrum-Beta-Lactamase-Producing Escherichia coli Isolates: Implications for Transmission Analyses. *J. Clin. Microbiol.* **53**, 2122–2131, https://doi.org/10.1128/jcm.00378-15 (2015).
8. Jorgensen, S. B. *et al.* Fecal carriage of extended spectrum beta-lactamase producing Escherichia coli and Klebsiella pneumoniae after urinary tract infection - A three year prospective cohort study. *PLoS One* **12**, e0173510, https://doi.org/10.1371/journal.pone.0173510 (2017).
9. Huddleston, J. R. Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infection and drug resistance* **7**, 167–176, https://doi.org/10.2147/idr.s48820 (2014).
10. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **14**, e1002533, https://doi.org/10.1371/journal.pbio.1002533 (2016).
11. Sorensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. & Wuertz, S. Studying plasmid horizontal transfer *in situ*: a critical review. *Nature reviews. Microbiology* **3**, 700–710, https://doi.org/10.1038/nrmicro1232 (2005).
12. Duval-Iflah, Y., Raibaud, P., Tancrede, C. & Rousseau, M. R-plasmic transfer from Serratia liquefaciens to Escherichia coli *in vitro* and *in vivo* in the digestive tract of gnotobiotic mice associated with human fecal flora. *Infect. Immun.* **28**, 981–990 (1980).
13. Schjorring, S., Struve, C. & Krogfelt, K. A. Transfer of antimicrobial resistance plasmids from Klebsiella pneumoniae to Escherichia coli in the mouse intestine. *The Journal of antimicrobial chemotherapy* **62**, 1086–1093, https://doi.org/10.1093/jac/dkn323 (2008).
14. Faure, S., Perrin-Guyomard, A., Delmas, J. M., Chatre, P. & Laurentie, M. Transfer of plasmid-mediated CTX-M-9 from Salmonella enterica serotype Virchow to Enterobacteriaceae in human flora-associated rats treated with cefixime. *Antimicrob. Agents Chemother.* **54**, 164–169, https://doi.org/10.1128/aac.00310-09 (2010).
15. Gottig, S., Gruber, T. M., Stecher, B., Wichelhaus, T. A. & Kempf, V. A. *In vivo* horizontal gene transfer of the carbapenemase OXA-48 during a nosocomial outbreak. *Clin. Infect. Dis.* **60**, 1808–1815, https://doi.org/10.1093/cid/civ191 (2015).

16. Lester, C. H., Frimodt-Moller, N., Sorensen, T. L., Monnet, D. L. & Hammerum, A. M. *In vivo* transfer of the vanA resistance gene from an Enterococcus faecium isolate of animal origin to an E. faecium isolate of human origin in the intestines of human volunteers. *Antimicrob. Agents Chemother.* **50**, 596–599, https://doi.org/10.1128/aac.50.2.596-599.2006 (2006).
17. Trobos, M., Lester, C. H., Olsen, J. E., Frimodt-Moller, N. & Hammerum, A. M. Natural transfer of sulphonamide and ampicillin resistance between Escherichia coli residing in the human intestine. *The Journal of antimicrobial chemotherapy* **63**, 80–86, https://doi.org/10.1093/jac/dkn437 (2009).
18. Karami, N. *et al*. Transfer of an ampicillin resistance gene between two Escherichia coli strains in the bowel microbiota of an infant treated with antibiotics. *The Journal of antimicrobial chemotherapy* **60**, 1142–1145, https://doi.org/10.1093/jac/dkm327 (2007).
19. Carattoli, A. *et al*. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903, https://doi.org/10.1128/aac.02412-14 (2014).
20. Orlek, A. *et al*. Plasmid Classification in an Era of Whole-Genome Sequencing: Application in Studies of Antibiotic ResistanceEpidemiology. *Frontiers in microbiology* **8**, 182, https://doi.org/10.3389/fmicb.2017.00182 (2017).
21. Conlan, S. *et al*. Plasmid Dynamics in KPC-Positive Klebsiella pneumoniae during Long-Term Patient Colonization. *mBio* **7**, https://doi.org/10.1128/mBio.00742-16 (2016).
22. Willemsen, I. *et al*. Retrospective identification of a previously undetected clinical case of OXA-48-producing K. pneumoniae and E. coli: the importance of adequate detection guidelines. *Infection* **44**, 107–110, https://doi.org/10.1007/s15010-015-0805-7 (2016).
23. Knudsen, P. K. *et al*. Impact of extensive antibiotic treatment on faecal carriage of antibiotic-resistant enterobacteria in children in a low resistance prevalence setting. *PLoS One* **12**, e0187618, https://doi.org/10.1371/journal.pone.0187618 (2017).
24. Gammelsrud, K. W., Knudsen, P. K. & Hoiby, E. A. A novel Direct MIC-gradient Strip Method to screen for antibiotic-resistant faecal Enterobacteriaceae. *J. Microbiol. Methods* **129**, 94–97, https://doi.org/10.1016/j.mimet.2016.08.007 (2016).
25. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, https://doi.org/10.1093/bioinformatics/btu170 (2014).
26. Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14**(Suppl 1), S7, https://doi.org/10.1186/1471-2164-14-s1-s7 (2013).
27. Bankevich, A. *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477, https://doi.org/10.1089/cmb.2012.0021 (2012).
28. Antipov, D. *et al*. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* **32**, 3380–3387, https://doi.org/10.1093/bioinformatics/btw493 (2016).
29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410, https://doi.org/10.1016/s0022-2836(05)80360-2 (1990).
30. Galardini, M., Mengoni, A. & Bazzicalupo, M. Mapping contigs using CONTIGuator. *Methods Mol. Biol.* **1231**, 163–176, https://doi.org/10.1007/978-1-4939-1720-4_11 (2015).
31. Carver, T. J. *et al*. ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422–3423, https://doi.org/10.1093/bioinformatics/bti553 (2005).
32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357–359, https://doi.org/10.1038/nmeth.1923 (2012).
33. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, https://doi.org/10.1093/bioinformatics/btp352 (2009).
34. Zankari, E. *et al*. Identification of acquired antimicrobial resistance genes. *The Journal of antimicrobial chemotherapy* **67**, 2640–2644, https://doi.org/10.1093/jac/dks261 (2012).
35. Larsen, M. V. *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* **50**, 1355–1361, https://doi.org/10.1128/jcm.06094-11 (2012).
36. Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M. & Scheutz, F. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J. Clin. Microbiol.* **53**, 2410–2426, https://doi.org/10.1128/jcm.00008-15 (2015).
37. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 524, https://doi.org/10.1186/preaccept-2573980311437212 (2014).
38. Croucher, N. J. *et al*. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15, https://doi.org/10.1093/nar/gku1196 (2015).
39. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, https://doi.org/10.1093/bioinformatics/btu033 (2014).
40. Guindon, S. *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321, https://doi.org/10.1093/sysbio/syq010 (2010).
41. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
42. Page, A. J. *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693, https://doi.org/10.1093/bioinformatics/btv421 (2015).
43. Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403, https://doi.org/10.1101/gr.2289704 (2004).
44. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069, https://doi.org/10.1093/bioinformatics/btu153 (2014).
45. Wang, J. *et al*. Nucleotide sequences of 16 transmissible plasmids identified in nine multidrug-resistant Escherichia coli isolates expressing an ESBL phenotype isolated from food-producing animals and healthy humans. *The Journal of antimicrobial chemotherapy* **69**, 2658–2668, https://doi.org/10.1093/jac/dku206 (2014).
46. Arredondo-Alonso, S., van Schaik, W., Willems, R. J. & Schürch, A. C. On the (Im)possibility to Reconstruct Plasmids from Whole Genome Short-Read SequencingData. *bioRxiv*, 1-18, https://doi.org/10.1101/086744 (2016).
47. van Schaik, W. The human gut resistome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140087, https://doi.org/10.1098/rstb.2014.0087 (2015).
48. Carattoli, A. Resistance plasmid families in Enterobacteriaceae. *Antimicrob. Agents Chemother.* **53**, 2227–2238, https://doi.org/10.1128/aac.01707-08 (2009).
49. Beaber, J. W., Hochhut, B. & Waldor, M. K. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427**, 72–74, https://doi.org/10.1038/nature02241 (2004).
50. Stecher, B. *et al*. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc. Natl. Acad. Sci. USA* **109**, 1269–1274, https://doi.org/10.1073/pnas.1113246109 (2012).
51. Smyth, R. L., Croft, N. M., O'Hea, U., Marshall, T. G. & Ferguson, A. Intestinal inflammation in cystic fibrosis. *Arch. Dis. Child.* **82**, 394–399 (2000).
52. Bruzzese, E. *et al*. Intestinal inflammation is a frequent feature of cystic fibrosis and is reduced by probiotic administration. *Aliment. Pharmacol. Ther.* **20**, 813–819, https://doi.org/10.1111/j.1365-2036.2004.02174.x (2004).
53. Munck, A. Cystic fibrosis: evidence for gut inflammation. *Int. J. Biochem. Cell Biol.* **52**, 180–183, https://doi.org/10.1016/j.biocel.2014.02.005 (2014).
54. Porse, A., Schonning, K., Munck, C. & Sommer, M. O. Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts. *Mol. Biol. Evol.* **33**, 2860–2873, https://doi.org/10.1093/molbev/msw163 (2016).

55. Norberg, P., Bergstrom, M., Jethava, V., Dubhashi, D. & Hermansson, M. The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination. *Nature communications* **2**, 268, https://doi.org/10.1038/ncomms1267 (2011).
56. Conlan, S. *et al*. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Sci. Transl. Med.* **6**, 254ra126, https://doi.org/10.1126/scitranslmed.3009845 (2014).
57. Sheppard, A. E. *et al*. Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC. *Antimicrob. Agents Chemother.* **60**, 3767–3778, https://doi.org/10.1128/aac.00464-16 (2016).

Acknowledgements

This work was supported by grants from the Southern and Eastern Norway Regional Health Authority (Helse Sør-Øst) (grant 2014071), Barnestiftelsen, Oslo University Hospital and the Norwegian Surveillance System for Antimicrobial Drug Resistance (NORM).

Author Contributions

P.K.K. and J.B. designed the work and wrote the manuscript. P.K.K. performed the bacteriological work and WGS analyses. J.B. developed the method for plasmid assembly, performed WGS analyses and created the figures. K.W.G. designed the original study and collected the clinical samples. K.A. performed the whole genome sequencing of the *E. coli* isolates and contributed to the creation of the figures. All authors (P.K.K., K.W.G., K.A., M.S., T.G.A., F.M. and J.B.) contributed to the interpretation of the results and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-18659-2.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.