High rates of human fecal carriage of *mcr-1*-positive multi-drug resistant *Enterobacteriaceae* isolates emerge in China in association with successful plasmid families

Lan-Lan Zhong[1,2*], Hang TT Phan[3*], Cong Shen[1,2], Karina Doris-Vihta[3], Anna E Sheppard[3], Xi Huang[2,4], Kun-Jiao Zeng[1,2], Hong-Yu Li[5], Xue-Fei Zhang[1,2], Sandip Patil[1,2], Derrick W Crook[3], A Sarah Walker[3], Yong Xing[1,2], Jia-lin Lin[1,2], Lian-Qiang Feng[1,2], Yohei Doi[6], Yong Xia[7], Nicole Stoesser[3†#], Guo-Bao Tian[1,2†#]

Author affiliations:

[1] Department of Immunology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

[2] Key Laboratory of Tropical Diseases Control (Sun Yat-sen University), Ministry of Education, Guangzhou, China

[3] Modernising Medical Microbiology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

[4] Program of Immunology, Affiliated Guangzhou Women and Children's Medical Center, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

[5] Department of Clinical Laboratory, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China

[6] University of Pittsburgh Medical Centre, Pittsburgh, Pennsylvania, USA

[7] Department of Clinical Laboratory Medicine, The Third Affiliated Hospital of Guangzhou Medical University, Guangdong, China

* These authors contributed equally to this article

† These senior authors contributed equally to this article

Running head: Human fecal carriage of *mcr-1* in China

Keywords: *mcr-1*, genomic epidemiology, China, ESBL

Address correspondence to:

Nicole Stoesser, nicole.stoesser@ndm.ox.ac.uk

Mailing address: Department of Microbiology (Research), John Radcliffe Hospital Level 7, Headley Way, Headington, OX3 9DU, UK.

Phone: +44 (0)1865 220856

Fax: +44 (0)1865

Alternate corresponding author:

Dr Guo-Bao Tian, tiangb@mail.sysu.edu.cn

Mailing address: Department of Immunology, Zhongshan School of Medicine, Sun Yat-sen University, 74 Zhongshan 2nd Road, Guangzhou 510080, China.

Fax/Phone: +86 (0)20 87335387

ARTICLE SUMMARY

mcr-1 renders colistin ineffective, which is of relevance to the management of drug-resistant *Enterobacteriaceae* infections. We highlight rapid increases in human gastrointestinal *mcr-1* carriage prevalence (2011-2016, Guangzhou), using bacterial genomics to characterize the genetic diversity facilitating *mcr-1* spread amongst humans/animals.

ABSTRACT

Objectives

mcr-1-mediated colistin resistance in *Enterobacteriaceae* is concerning, as colistin is used in treating multidrug-resistant *Enterobacteriaceae* infections. Rates of human *mcr-1* gastrointestinal carriage have historically been low. We identified trends in human fecal *mcr-1*-positivity rates and colonization with *mcr-1*-positive+third-generation cephalosporin-resistant (3GC-R) *Enterobacteriaceae* in Guangzhou, China, and investigated the genetic contexts of *mcr-1* in a subset of *mcr-1*-positive+3GC-R strains.

Methods

Fecal samples were collected from in-patients and out-patients submitting specimens to three hospitals (2011-2016). *mcr-1* carriage trends were assessed using iterative sequential regression. A subset of *mcr-1*-positive isolates was sequenced (whole genome sequencing [WGS], Illumina), and genetic contexts (flanking regions, plasmids) of *mcr-1* characterized.

Results

Of 8,022 fecal samples collected, 497 (6.2%) were *mcr-1*-positive, and 182 (2.3%) harbored *mcr-1*-positive+3GC-R *Enterobacteriaceae*. We observed marked increases in *mcr-1* (0% [Apr/2011] to 31% [Mar/2016]) and more recent (since January 2014; 0% [Apr/2011] to 15% [Mar/2016]) increases in human colonization with *mcr-1*-positive+3GC-R *Enterobacteriaceae* ($p<0.001$). *mcr-1*-positive+3GC-R isolates were commonly multi-drug resistant.

WGS of *mcr-1-positive*+3GC-R isolates (70 *Escherichia coli*, 3 *Klebsiella pneumoniae*) demonstrated bacterial strain diversity (48 *E. coli* sequence types); *mcr-1* in association with common plasmid backbones (IncI, IncHI2/HI2A, IncX4) and sometimes in multiple plasmids; frequent *mcr-1* chromosomal integration; and high mobility of the *mcr-1*-associated insertion sequence IS*Apl1*. Sequence similarity with published *mcr-1* plasmid sequences was consistent with spread amongst animal/human reservoirs.

Conclusions

The high prevalence of *mcr-1* in multidrug-resistant *E. coli* colonizing humans is a clinical threat; diverse genetic mechanisms (strains/plasmids/insertion sequences) have contributed to the dissemination of *mcr-1*, and will facilitate its persistence.

MAIN TEXT

INTRODUCTION

Colistin is one of the antibiotics of last resort for managing multidrug-resistant Gram-negative infections. Colistin resistance has historically largely been due to cell wall modifications, utilization of efflux pumps, and capsule formation(1). Transmissible, *mcr-1*-mediated colistin resistance was recently identified in *Escherichia coli* and *Klebsiella pneumoniae* isolates from hospitalized humans, animals (pigs) and raw meat (pigs and chicken) in China(2), with higher rates in animal samples (~19% versus ~1% in humans).

Subsequently, *mcr-1*-harboring strains have been identified in humans, animals and raw meat sampled globally (e.g.(3-13). These strains have predominantly been *E. coli*(14) or *Salmonella* spp.(3, 5), with up to 20% carriage prevalence in swine and poultry(6, 11), and *mcr*-1-positive isolates from chickens as early as the 1980s in China and 2007 in

France(6, 15). Prevalence in humans remains low(2, 4, 7, 10, 16), and mostly restricted to hospitalized patients(17, 18). However *mcr-1* can be carried in the healthy human gut(18, 19).

The association of *mcr-1* with other broad-spectrum resistance mechanisms, such as extended-spectrum β-lactamases (ESBLs) and/or carbapenemases(20-25), could represent a major clinical problem. The identification of *mcr-1* in multiple plasmid types, including IncI2(2, 25), IncHI2(22), IncX4(12, 25), IncP(26) and IncF(23) plasmids, is consistent with multiple *mcr-1* mobilization events, potentially facilitating the association of *mcr-1* with other resistance mechanisms, thereby creating multidrug-resistant bacterial hosts.
In this study, we investigate human fecal carriage of *mcr-1* and of *mcr-1* in third-generation cephalosporin-resistant (3GC-R)-*Enterobacteriaceae* in Guangzhou, China, over five years. Given that colistin resistance is of particular clinical relevance in the context of multi-drug resistance, we used whole-genome sequencing (WGS) to characterize a subset of 3GC-R isolates to identify relevant genetic structures, using 3GC resistance as a marker for wider multi-drug resistance.

METHODS

All in-patients and out-patients submitting any clinical specimens during the study timeframe to the hospital microbiology laboratories for diagnostic purposes were asked to participate in the study by means of an invitation included with the diagnostic test report returned to the patient. Samples came from three hospitals in Guangzhou, Guangdong province, serving a population of ~15 million over ~10,000 km^2. Recruitment and sample

collection occurred continuously (except January 2012, February 2013 and February 2014; holiday months, staff shortages); samples were not de-duplicated by patient. Ethical approval was given by Sun Yat-Sen University; individual consent for the use of fecal samples was obtained from patients.

Fecal samples were collected into sterile fecal specimen containers and plated onto Columbia blood agar (CBA) within 2 hours of collection. A cotton swab was used to inoculate the agar with specimen (plate incubated for 18-24 hours, 37°C). Subsequently, up to 10 colonies of *Enterobacteriaceae* were sub-cultured to MacConkey agar+cefotaxime (2 mg/L), and species identification confirmed by 16S rDNA sequencing (see web-only Supplementary methods). All cefotaxime-resistant *Enterobacteriaceae* isolates were stored (lysogeny broth (LB)+30% glycerol, -80°C). Sweeps of cultured growth from the original CBA plates were also similarly stored. All frozen sweeps of cultured growth from feces (n=8,022) and individual cefotaxime-resistant *Enterobacteriaceae* isolates (n=20,332) were subsequently re-cultured and screened for *mcr-1* by PCR (see web-only Supplementary methods). Cefotaxime-resistant isolates were screened for $bla_{\text{CTX-M}}$, and alleles determined by sequencing (Supplementary methods). Species identification for *mcr-1* positive isolates was performed by API20E. Minimum inhibitory concentrations (MICs) were determined for all *mcr-1*-positive isolates by agar dilution (EUCAST breakpoints, version 6.0; Clinical and Laboratory Standards Institute, document M100-S25) (Fig.1).

Every cefotaxime-resistant, *mcr-1*-positive *Enterobacteriaceae* isolate of distinct species and MIC profile to May 2015 (n=45), and a random subset from June 2015-March 2016 (n=44/142 [31%]) were selected for WGS. DNA was extracted using the cetyl-trimethyl-ammonium bromide (CTAB)/chloroform method (Supplementary methods).

DNA extracts were sequenced on the Illumina HiSeq 4000 platform at the Beijing Genomics Institute, using both paired-end (150bp reads, ~350bp insert) and mate-pair (50bp reads, ~6kb insert) approaches (n=69 isolates) or paired-end reads only (n=20 isolates; Supplementary Table S1). Libraries were prepared using standardized protocols incorporating fragmentation by ultra-sonication, end repair, adaptor ligation, and PCR amplification (Supplementary methods).

Preliminary species identification for isolates was derived from WGS using Kraken(27); read-data were then mapped to species-specific references(28). Hybrid *de novo* assemblies of paired-end and mate-pair data, or paired-end data alone, were generated using SPAdes(29) version 3.6 with the "--careful" option, a set of automatically determined k-mer values (21, 33, 55, 77), and by removing contigs <500bp or with k-mer coverage <1. *In silico* MLST was determined using BLASTn and publicly available databases (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli, http://bigsdb.web.pasteur.fr/klebsiella/klebsiella.html). Resistance genes were identified from *de novo* assemblies using a curated database of resistance genes(30) and BLASTn/mapping-based identification (scripts available at: https://github.com/hangphan/resistType). Contigs were annotated using PROKKA(31).

Contigs containing *mcr-1* were defined as "plasmid" if they contained annotations consistent with plasmid-associated loci and no obvious chromosomal loci, or "chromosomal" if >75% of annotations (excluding hypothetical proteins) were consistent with a chromosomal location.

The integrity of assemblies containing *mcr-1* in chromosomal locations was assessed using REAPR(32). We excluded any sequences with assembly sizes ≥6.5Mb and/or mixed calls from *in silico* species identification/MLST.

The phylogeny was reconstructed using IQtree(33). Branch lengths were corrected for recombination using ClonalFrameML(34)(Supplementary methods). The phylogeny was represented in the interactive tree of life viewer (iTOL v3, http://itol.embl.de). Insertion sequences (IS) were downloaded from ISFinder (https://www-is.biotoul.fr); sequence assemblies were queried against this database with BLASTn (requiring >95% sequence identity over >90% of the reference sequence length).

Circularization of *mcr-1*-harboring plasmid contigs was confirmed using Bandage(35). For single *mcr-1*-harboring plasmid contigs which were not circularizable, we also used Bandage to visualize the sequencing assembly graph generated by SPAdes and manually resolved the most likely *mcr-1* plasmid structures based on node (contig) linkage and contig coverage (Supplementary methods).

Iterative sequential regression (ISR) in R was used to characterize trends in fecal *mcr-1* positivity (Supplementary methods).

Sequence data have been deposited in NCBI (BioProject: PRJNA354216; Supplementary Table S1).

RESULTS

Trends in fecal *mcr-1* prevalence, and fecal *mcr-1*/cefotaxime-resistant *Enterobacteriaceae* prevalence

Sweeps of cultured growth from 497/8,022 fecal samples (6.2%) were *mcr-1* PCR-positive, and 182 (2.3%) fecal samples harbored *mcr-1*-positive+cefotaxime-resistant *Enterobacteriaceae* (Fig.2). The proportion of both *mcr-1*-positive and *mcr-1*-positive+cefotaxime-resistant samples increased significantly over time ($p<0.001$). For *mcr-1*-positive+cefotaxime-resistant samples, this was driven specifically by increases after January 2014 ($p<0.001$, Fig.2; 95% CI for estimated date of trend change: 01/April/2013-01/Nov/2014). There was no evidence of a change in fecal sampling rates over time (negative binomial regression; incidence rate ratio [IRR] 1.02 (95% CI 0.96-1.08), p-value = 0.6).

From fecal samples harboring *mcr-1*-postive+cefotaxime-resistant *Enterobacteriaceae*, 187 distinct isolates from 182 fecal samples from 179 individuals were identified (*E. coli*, n=173; *K. pneumoniae*, n=13; *Enterobacter cloacae*, n=1). Of these, 23 isolates from 22 individuals had ertapenem MICs of ≥0.5mg/L (earliest isolate in 2013). 144/179 (80%)

individuals were hospital in-patients at sampling (median stay 14 days [range: 3-258 days; IQR: 7-21 days]).

Whole genome sequencing of *mcr-1*+cefotaxime-resistant *Enterobacteriaceae*

Species and strain-level diversity

Of 89/187 isolates selected for WGS (see methods), 11/45 consecutive isolates pre-May 2015 and 5/44 randomly selected isolates post-June 2015 failed quality control (Table S1), leaving 73 sequences for analysis (70 *E. coli*, 3 *K. pneumoniae*). For *E. coli*, these represented 48 sequence types (STs) (Fig.3), ST156 being the most common (n=5), with most isolates (n=33) representing singleton STs. Other global disease-causing lineages were also identified, including ST131, ST155 and ST405(36), and for *K. pneumoniae*, ST15 and ST307. Four pairs of isolates were separated by 0 SNVs (SYSU0077/SYSU0078; SYSU0002/SYSU0009; SYSU0014/SYSU0015; SYSU0025/SYSU0026), and one pair by 3 SNVs (SYSU0041/SYSU0049), representing likely direct/indirect transmissions between patients. All others were >1126 SNVs apart.

mcr-1 and insertion sequence diversity

A novel *mcr-1* allele was identified (G3A; loss of the first methionine [n=2; SYSU0052, SYSU0010]); these isolates remained colistin-resistant. Another isolate had *mcr-1* disrupted by an IS*1294* element, previously described downstream of *mcr-1* (n=1; SYSU0039)(8).

Insertion sequences may contribute to plasmid-plasmid and plasmid-chromosome rearrangements and gene mobilization. We found 99 different IS types in the 73 sequenced isolates, with median 15 (range: 7-29) IS types per isolate.

Antimicrobial susceptibility profiles and mechanisms of multi-drug resistance in mcr-1-positive+cefotaxime-resistant Enterobacteriaceae

Rates of cross-class resistance in *mcr-1*-positive+cefotaxime-resistant isolates were high (Table 1); only carbapenems, nitrofurantoin and tigecycline demonstrated susceptibility rates ≥80% (similar rates in subset of sequenced isolates [Table 1]). Amongst the sequenced isolates, 58/73 (79%) harbored $bla_{\text{CTX-M}}$, including: (i) group 9-like alleles: 14(n=22), 27(4), 65(7), 110(1), 130(1); (ii) group 1-like alleles: 1-like(n=1), 3 (n=1), 11(1), 55/55-like(21), 136(1); and (iii) hybrid alleles: 64(3), 123(2). Thirteen (18%) sequenced isolates had two $bla_{\text{CTX-M}}$ variants, and in three (carrying $bla_{\text{CTX-M-14,27/55,55/123}}$), $bla_{\text{CTX-M}}$ was located on the same contig as *mcr-1*. For eleven (15%) isolates, no cefotaxime resistance mechanism could be identified (Table S1). Multiple other resistance genes were identified; three isolates harbored bla_{CMY} and one bla_{DHA} (Table S1). In carbapenem non-susceptible isolates that were sequenced (6/23), no genes encoding carbapenemases were identified; carbapenem non-susceptibility in these isolates was attributed to the presence of ESBLs ($bla_{\text{CTX-M}}$, $bla_{\text{OXA-10}}$) + porin gene mutations.

Genetic (plasmid/chromosomal) contexts of mcr-1

The genetic context of *mcr-1* was diverse. In 62/73 (85%) isolates *mcr-1* was located on plasmid-associated contigs, in 4 (5%) on chromosomal contigs, in 2 (5%) likely on chromosomal contigs, and in 5 (7%) the location was unclear due to assembly/annotation limitations. *mcr-1* copy number per sequenced isolate was estimated at 0.27-3.54 (median 0.97); copy number values >1 were likely due to its presence either on multi-copy plasmids; as multiple copies within the same plasmid (e.g. SYSU0077 [IncH]; SYSU0093 [IncI2]); or on different plasmids within the same isolate (e.g. SYSU0072 [IncH and IncX4]; SYSU0220 [novel plasmid described below]).

mcr-1 on IncI plasmids

In 27/62 (44%) isolates where *mcr-1* was plasmid-associated, it was co-located with an IncI2 replicon (n=16). Fourteen sequences represented circularizable, complete plasmid structures; this included the earliest sequenced *mcr-1* plasmid, to our knowledge (SYSU0011, May 2011).

In these plasmids, the backbone and *mcr-1* location were largely preserved; this was also true of the six non-circularizable IncI contigs harbouring *mcr-1* (Fig.4). In relation to *mcr-1*, we observed variable presence of IS*Apl1*, but when present, it was always located upstream of *mcr-1* between *mcr-1* and *nikB*, most consistent with ancestral IS*Apl1*-mediated acquisition of *mcr-1* into an IncI plasmid backbone, and subsequent loss of IS*Apl1*. These plasmids were also undergoing significant evolution by mutation/recombination, rearrangement, indels, and acquisition/loss of smaller mobile genetic elements (MGEs), such as IS*Ecp1*+$bla_{\text{CTX-M-55}}$ (SYSU0060, SYSU0045; Fig.4).

The *mcr-1*-IncI2 plasmids were also genetically highly similar to three previously sequenced plasmids (Fig.4): SZ02 (Accession number: KU761326, blood culture, August 2015, Suzhou); the *bla*$_{\text{CTX-M-55}}$-harboring pA31-12 (Accession number: KX034083.1, chicken, August 2012, Guangzhou), and pHNSHP45 (Accession number: KP347127.1, pig, July 2013, Shanghai), consistent with the dissemination of this plasmid in human cases/carriers, pigs and chickens across China.

We also observed highly genetically related IncI plasmid backbones in two different *E. coli* host strains (ST156, 354; SYSU0060, SYSU0062) isolated from the same patient/same day, potentially consistent with within-host transfer, as well as highly genetically related plasmids within different *E. coli* strains and humans across periods of time consistent with direct transmission/acquisition from common sources (SYSU0019, SYSU0007).

mcr-1 on IncHI2/HI2A plasmids

In 10/62 (34%) isolates where *mcr-1* was plasmid-associated, it co-localized with either IncHI2/HI2A (n=7), or was plausibly on an IncHI2 plasmid (n=12). As these plasmids are large and more likely to include repeats, we were unable to fully reconstruct them. Five contigs were short (<3,000bp); the others are represented in Fig.5, alongside two closed reference *mcr-1*-IncHI2 plasmids, pSA26 (Accession number: KU743384, blood culture, Saudi Arabia), and pHNSHP45-2 (Accession number: KU341381, pig feces, Shanghai). These demonstrate that a homogenous backbone sequence ranging from

~38kbp to ~224kbp (of evaluable sequence) has been circulating within the study population between 2012-early 2016, likely derived from an ancestral plasmid similar to the reference plasmids (Fig.5).

We observed apparently frequent IS/transposon-associated indel events in *mcr-1*-IncHI/HI2A plasmids, including of IS*Apl1*, which was either at the 5' end of *mcr-1* (SYSU0003), flanking it on both sides (SYSU0014), or absent (SYSU0026). As *mcr-1* was located in the same wider genetic context in 14/17 study plasmid contigs, this likely represents a single IS*Apl1*-*mcr-1*-IS*Apl1* acquisition and subsequent loss of IS*Apl1* elements(37), similar to that seen in *mcr-1*-IncI plasmids in this study. Alternatively it could represent a genetic "hotspot" facilitating multiple IS*Apl1*-*mcr-1*-IS*Apl1* insertion events.

We also observed a *mcr-1* duplication event within otherwise identical plasmid sequences in two isolates taken 7 days apart (SYSU0077, SYSU0078); and inversions of IS*Apl1*-*mcr-1* within the backbone (SYSU0002, SYSU0009, SYSU0055), all consistent with high rates of plasticity involving IS*Apl1*-*mcr-1* and IncHI2/HI2A structures.

mcr-1 on IncX4 plasmids

In 15/62 (24%) isolates where *mcr-1* was plasmid-associated, it co-localized (n=10) or was plausibly associated with (n=5) an IncX4 replicon; seven of these represented circularizable plasmid sequences (Fig.6). IncX4-*mcr-1* plasmids had a highly conserved, syntenic backbone, with limited nucleotide and indel variation over 2.5 years. They were

also highly similar to two reference IncX4-*mcr-1* plasmids, SZ04 (peritoneal fluid, Suzhou) and pAF48 (urine, Johannesburg, South Africa), consistent with national and international spread. IS*Apl1* was absent in all of these plasmids, but IS*26*-like elements were located upstream of *mcr-1*, suggesting the mechanism of mobilization to IncX4 is different from that in IncI and IncHI2/HI2A plasmids.

Other genetic contexts of mcr-1

In 8/73 (11%) sequenced isolates no associated plasmid replicon could be identified. In one isolate (SYSU0220) we identified a new *mcr-1* harboring plasmid (48,944bp, Fig.S1) similar to pHNFP671 (accession number: KP324830.1, IncP, pig feces, isolated prior to 2014).

Chromosomal *mcr-1* integration was confirmed in four *E. coli* STs (457, 1114, 1684, 2936), and plausibly in two others (101, 2705), consistent with at least 4 independent chromosomal integration events in diverse strains (5% sequenced *E. coli* isolates), suggesting capacity for non-lineage-specific vertical dissemination of *mcr-1*.

DISCUSSION

Widespread transmissible colistin resistance is of major concern in the context of multidrug-resistance, as colistin is commonly used to treat infections caused by MDR-*Enterobacteriaceae*, despite its drawbacks(39, 40). In this, the largest study of human fecal carriage and WGS of *mcr-1* isolates to our knowledge, we observed alarming sequential increases in carriage of *mcr-1*, and of *mcr-1*-positive+cefotaxime-resistant

Enterobacteriaceae fecal carriage (predominantly *E. coli*) over five years. We also found significant diversity and genetic plasticity of MGEs harboring *mcr-1* that may explain some of these dramatic increases.

IncI plasmids are narrow host-range plasmids(41), commonly isolated from *E. coli* and *Salmonella* spp., and implicated in the spread of the ESBL gene $bla_{\text{CTX-M}}$ amongst *Enterobacteriaceae*, particularly in China(42, 43). CTX-M-55 has been found in *E. coli* in animals in China, and CTX-M-55/55-like variants were seen with *mcr-1* in IncI2 plasmids in two different *E. coli* strains (ST156, ST117) in this study. Previously, it has been postulated that an IncI2 plasmid backbone acquired a 3,080bp IS*Ecp1*-$bla_{\text{CTX-M-15}}$ complex from an IncA/C plasmid, with a mutation resulting in conversion to $bla_{\text{CTX-M-55}}$ within this structure(44). The same signature sequence was observed in our isolates, as well as another recently sequenced IncI2 plasmid harboring $bla_{\text{CTX-M-55}}$ from a chicken *E. coli* isolate in Guangzhou (August 2012)(45), consistent with the spread of this plasmid across humans and animals. $bla_{\text{CTX-M-64}}$ was also observed on an IncI *mcr-1* plasmid in this study (SYSU0115, ST155); this allele is thought to have arisen from recombination between $bla_{\text{CTX-M-14}}$ and $bla_{\text{CTX-M-15}}$ on IncI plasmids in food animals in China(44).

IncHI2/HI2A plasmids are typically large (>250kb)(46), multidrug-resistant plasmids that have been associated with a range of antibiotic and metal resistance genes in *Salmonella* spp. and *E. coli* isolated from humans and food-producing animals(47). Similar to the IncI plasmids, our genetic analyses suggest an initial *mcr-1* acquisition event, and subsequent loss of IS*Apl1* units and rearrangements of the plasmid backbone.

In at least four sequenced isolates, *mcr-1* had been integrated into the chromosome, a phenomenon observed in at least two other studies (*E. coli* ST156, Beijing (48); *E. coli* ST410, Germany (49)). Our data suggest two additional *mcr-1* chromosomal integrations - one in association with an integrative element promoting IS activity, and one in association with an IS*Ec23*/IS*30*-like/IS*1294* element, which has played a role in the mobilization of $bla_{\text{CMY-2}}$ in *Salmonella* spp. and *E. coli*(50). The multifarious means by which chromosomal integration of *mcr-1* appears to occur is of concern, as this may facilitate more stable inheritance of this gene.

The variable extent of genetic diversity observed within the plasmid families here (IncHI2/HI2A and IncI > IncX4) may reflect different evolutionary rates for these plasmid families, or may be consistent with earlier acquisition of IS*Apl1-mcr-1* composite transposons by IncHI2/HI2A and IncI2, most likely within poultry and pig farms in which regular colistin exposure represents a major selection pressure(2). The acquisition of *mcr-1* by IncX4 plasmids appears to be unrelated to the presence of IS*Apl1*, and possibly involves IS*26/26*-like structures.

This study has several limitations. Firstly, although we observed significant increases in fecal carriage of *mcr-1*-harboring isolates, we did not assess possible risk factors associated with increasing incidence. We did not de-duplicate samples by patient for the whole dataset; however, analysis of participant study identifiers for the *mcr-1*/3GC-resistant isolates suggests that replicate sample submission was minimal (~2%), and

would not have explained the incidence trends observed. We only investigated the culturable component of feces for overall *mcr-1* prevalence, as DNA was extracted from sweeps of cultured feces, and we may therefore have underestimated true *mcr-1* colonization by not performing PCR direct on DNA extracts from whole feces. Our WGS strategy only targeted culturable, cefotaxime- and colistin-resistant isolates from feces, as we were predominantly interested in investigating the genomic epidemiology of multi-drug resistant *Enterobacteriaceae*. The diversity of strains and MGEs harboring *mcr-1* may therefore be even greater. We did not screen for *mcr-2*, which has also been shown to confer colistin resistance, nor did we screen potential animal or non-human reservoirs. Finally, due to resource limitations we were unable to sequence all *mcr-1*-positive isolates, or to re-sequence those that failed. We were also unable to undertake any long-read sequencing, increasingly the method of choice in assembling plasmids.

Despite these limitations, we have demonstrated that human fecal carriage of *mcr-1* positive *E. coli* has increased dramatically in Guangzhou over the last two years, reaching similar proportions (20-30%) to animals (pigs, chickens) over the preceding 3-4 years in the same region of China(2). Our genetic analyses suggest the rapid emergence of several major plasmid vectors of *mcr-1* within numerous multidrug-resistant *E. coli* strains carried by humans, and highlight the significant degree of plasticity in these plasmid vectors harboring *mcr-1* over short periods of time.

FUNDING

This work was supported by: the National Natural Science Foundation of China (grant numbers 81722030, 81471988); Science and Technology Planning Project of Guangdong (grant number 2016A020219002); The 111 Project (grant numbers B13037, B12003); and Fundamental Research Funds for the Central Universities (grant number 16ykzd09).

Additional funding support was provided through a University of Oxford/National Institutes of Health Research (NIHR) Academic Clinical Lectureship to NS and the NIHR Oxford Biomedical Research Centre. DWC is a NIHR senior investigator.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

ACKNOWLEDGEMENTS

We would like to acknowledge the work of the Modernizing Medical Microbiology (MMM) sequencing pipeline team (Nicholas Sanderson) and informatics team (Ian Szwajca, Trien Do, Dona Foster) - collectively represented as the Modernizing Medical Microbiology Informatics Group.

REFERENCES

1. **Olaitan AO, Morand S, Rolain JM.** 2014. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. Front Microbiol **5:**643.

2. **Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu LF, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu JH, Shen J.** 2016. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis **16:**161-168.

3. **Quesada A, Ugarte-Ruiz M, Iglesias MR, Porrero MC, Martinez R, Florez-Cuadrado D, Campos MJ, Garcia M, Piriz S, Saez JL, Dominguez L.** 2016. Detection of plasmid mediated colistin resistance (MCR-1) in Escherichia coli and Salmonella enterica isolated from poultry and swine in Spain. Res Vet Sci **105:**134-135.

4. **Prim N, Rivera A, Rodriguez-Navarro J, Espanol M, Turbau M, Coll P, Mirelis B.** 2016. Detection of mcr-1 colistin resistance gene in polyclonal Escherichia coli isolates in Barcelona, Spain, 2012 to 2015. Euro Surveill **21**.

5. **Webb HE, Granier SA, Marault M, Millemann Y, den Bakker HC, Nightingale KK, Bugarel M, Ison SA, Scott HM, Loneragan GH.** 2016. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis **16:**144-145.

6. **Perrin-Guyomard A, Bruneau M, Houee P, Deleurme K, Legrandois P, Poirier C, Soumet C, Sanders P.** 2016. Prevalence of mcr-1 in commensal Escherichia coli from French livestock, 2007 to 2014. Euro Surveill **21**.

7. **Kluytmans-van den Bergh MF, Huizinga P, Bonten MJ, Bos M, De Bruyne K, Friedrich AW, Rossen JW, Savelkoul PH, Kluytmans JA.** 2016. Presence

of mcr-1-positive *Enterobacteriaceae* in retail chicken meat but not in humans in the Netherlands since 2009. Euro Surveill **21**.

8. **Stoesser N, Mathers AJ, Moore CE, Day NP, Crook DW.** 2016. Colistin resistance gene mcr-1 and pHNSHP45 plasmid in human isolates of Escherichia coli and Klebsiella pneumoniae. Lancet Infect Dis **16:**285-286.

9. **Olaitan AO, Chabou S, Okdah L, Morand S, Rolain JM.** 2016. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis **16:**147.

10. **Kuo SC, Huang WC, Wang HY, Shiau YR, Cheng MF, Lauderdale TL.** 2016. Colistin resistance gene mcr-1 in Escherichia coli isolates from humans and retail meats, Taiwan. J Antimicrob Chemother doi:10.1093/jac/dkw122.

11. **Nguyen NT, Nguyen HM, Nguyen CV, Nguyen TV, Nguyen MT, Thai HQ, Ho MH, Thwaites G, Ngo HT, Baker S, Carrique-Mas J.** 2016. The use of colistin and other critical antimicrobials on pig and chicken farms in southern Vietnam and their association with resistance in commensal Escherichia coli. Appl Environ Microbiol doi:10.1128/AEM.00337-16.

12. **Mulvey MR, Mataseje LF, Robertson J, Nash JH, Boerlin P, Toye B, Irwin R, Melano RG.** 2016. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis **16:**289-290.

13. **Grami R, Mansour W, Mehri W, Bouallegue O, Boujaafar N, Madec JY, Haenni M.** 2016. Impact of food animal trade on the spread of mcr-1-mediated colistin resistance, Tunisia, July 2015. Euro Surveill **21**.

14. **Falgenhauer L, Waezsada SE, Yao Y, Imirzalioglu C, Kasbohrer A, Roesler U, Michael GB, Schwarz S, Werner G, Kreienbrock L, Chakraborty T,**

consortium R. 2016. Colistin resistance gene mcr-1 in extended-spectrum beta-lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany. Lancet Infect Dis **16:**282-283.

15. **Shen Z, Wang Y, Shen Y, Shen J, Wu C.** 2016. Early emergence of mcr-1 in Escherichia coli from food-producing animals. Lancet Infect Dis **16:**293.
16. **Quan J, Li X, Chen Y, Jiang Y, Zhou Z, Zhang H, Sun L, Ruan Z, Feng Y, Akova M, Yu Y.** 2017. Prevalence of mcr-1 in Escherichia coli and Klebsiella pneumoniae recovered from bloodstream infections in China: a multicentre longitudinal study. Lancet Infect Dis doi:10.1016/S1473-3099(16)30528-X.
17. **Zhang R, Huang Y, Chan EW, Zhou H, Chen S.** 2016. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis **16:**291-292.
18. **Wang Y, Tian GB, Zhang R, Shen Y, Tyrrell JM, Huang X, Zhou H, Lei L, Li HY, Doi Y, Fang Y, Ren H, Zhong LL, Shen Z, Zeng KJ, Wang S, Liu JH, Wu C, Walsh TR, Shen J.** 2017. Prevalence, risk factors, outcomes, and molecular epidemiology of mcr-1-positive *Enterobacteriaceae* in patients and healthy adults from China: an epidemiological and clinical study. Lancet Infect Dis doi:10.1016/S1473-3099(16)30527-8.
19. **Arcilla MS, van Hattem JM, Matamoros S, Melles DC, Penders J, de Jong MD, Schultsz C, consortium C.** 2016. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis **16:**147-149.
20. **Skov RL, Monnet DL.** 2016. Plasmid-mediated colistin resistance (mcr-1 gene): three months later, the story unfolds. Euro Surveill **21**.

21. **Haenni M, Poirel L, Kieffer N, Chatre P, Saras E, Metayer V, Dumoulin R, Nordmann P, Madec JY.** 2016. Co-occurrence of extended spectrum beta lactamase and MCR-1 encoding genes on plasmids. Lancet Infect Dis **16:**281-282.

22. **Zhi C, Lv L, Yu LF, Doi Y, Liu JH.** 2016. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis **16:**292-293.

23. **Zeng KJ, Doi Y, Patil S, Huang X, Tian GB.** 2016. Emergence of the Plasmid-Mediated mcr-1 Gene in Colistin-Resistant Enterobacter aerogenes and Enterobacter cloacae. Antimicrob Agents Chemother **60:**3862-3863.

24. **Du H, Chen L, Tang YW, Kreiswirth BN.** 2016. Emergence of the mcr-1 colistin resistance gene in carbapenem-resistant *Enterobacteriaceae*. Lancet Infect Dis **16:**287-288.

25. **Li A, Yang Y, Miao M, Chavda KD, Mediavilla JR, Xie X, Feng P, Tang YW, Kreiswirth BN, Chen L, Du H.** 2016. Complete sequences of mcr-1-harboring plasmids from extended spectrum beta-lactamase (ESBL)- and carbapenemase-producing *Enterobacteriaceae* (CPE). Antimicrob Agents Chemother doi:10.1128/AAC.00550-16.

26. **Malhotra-Kumar S, Xavier BB, Das AJ, Lammens C, Butaye P, Goossens H.** 2016. Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. Lancet Infect Dis **16:**283-284.

27. **Wood DE, Salzberg SL.** 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol **15:**R46.

28. **Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW, Kasarskis A, Batty EM, Kos V, Wilson DJ, Phetsouvanh R, Wyllie D, Sokurenko E, Manges AR, Johnson TJ, Price LB, Peto TE, Johnson JR, Didelot X, Walker AS, Crook DW, Modernizing Medical Microbiology Informatics G.** 2016. Evolutionary History of the Global Emergence of the Escherichia coli Epidemic Clone ST131. MBio **7**.

29. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455-477.

30. **Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, Johnson JR, Walker AS, Peto TE, Crook DW.** 2013. Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. J Antimicrob Chemother **68:**2234-2244.

31. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068-2069.

32. **Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD.** 2013. REAPR: a universal tool for genome assembly evaluation. Genome Biol **14:**R47.

33. **Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ.** 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **32:**268-274.

34. **Didelot X, Wilson DJ.** 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol **11:**e1004041.

35. **Wick RR, Schultz MB, Zobel J, Holt KE.** 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics **31:**3350-3352.

36. **Woodford N, Turton JF, Livermore DM.** 2011. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. FEMS Microbiol Rev **35:**736-755.

37. **Snesrud E, He S, Chandler M, Dekker JP, Hickman AB, McGann P, Dyda F.** 2016. A Model for Transposition of the Colistin Resistance Gene mcr-1 by ISApl1. Antimicrob Agents Chemother doi:10.1128/AAC.01457-16.

38. **Kusumoto M, Ooka T, Nishiya Y, Ogura Y, Saito T, Sekine Y, Iwata T, Akiba M, Hayashi T.** 2011. Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. Nat Commun **2:**152.

39. **Landersdorfer CB, Nation RL.** 2015. Colistin: how should it be dosed for the critically ill? Semin Respir Crit Care Med **36:**126-135.

40. **Humphries RM.** 2015. Susceptibility testing of the polymyxins: where are we now? Pharmacotherapy **35:**22-27.

41. **Carattoli A.** 2009. Resistance plasmid families in *Enterobacteriaceae*. Antimicrob Agents Chemother **53:**2227-2238.

42. **Wong MH, Liu L, Yan M, Chan EW, Chen S.** 2015. Dissemination of IncI2 Plasmids That Harbor the blaCTX-M Element among Clinical Salmonella Isolates. Antimicrob Agents Chemother **59:**5026-5028.

43. **Lv L, Partridge SR, He L, Zeng Z, He D, Ye J, Liu JH.** 2013. Genetic characterization of IncI2 plasmids carrying blaCTX-M-55 spreading in both pets and food animals in China. Antimicrob Agents Chemother **57:**2824-2827.

44. **Liu L, He D, Lv L, Liu W, Chen X, Zeng Z, Partridge SR, Liu JH.** 2015. blaCTX-M-1/9/1 Hybrid Genes May Have Been Generated from blaCTX-M-15 on an IncI2 Plasmid. Antimicrob Agents Chemother **59:**4464-4470.

45. **Sun J, Li XP, Yang RS, Fang LX, Huo W, Li SM, Jiang P, Liao XP, Liu YH.** 2016. Complete Nucleotide Sequence of IncI2 Plasmid Co-harboring blaCTX-M-55 and mcr-1. Antimicrob Agents Chemother doi:10.1128/AAC.00774-16.

46. **Garcia-Fernandez A, Carattoli A.** 2010. Plasmid double locus sequence typing for IncHI2 plasmids, a subtyping scheme for the characterization of IncHI2 plasmids carrying extended-spectrum beta-lactamase and quinolone resistance genes. J Antimicrob Chemother **65:**1155-1161.

47. **Fang L, Li X, Li L, Li S, Liao X, Sun J, Liu Y.** 2016. Co-spread of metal and antibiotic resistance within ST3-IncHI2 plasmids from E. coli isolates of food-producing animals. Sci Rep **6:**25312.

48. **Yu H, Qu F, Shan B, Huang B, Jia W, Chen C, Li A, Miao M, Zhang X, Bao C, Xu Y, Chavda KD, Tang YW, Kreiswirth BN, Du H, Chen L.** 2016. Detection of mcr-1 colistin resistance gene in carbapenem-resistant *Enterobacteriaceae* (CRE) from different hospitals in China. Antimicrob Agents Chemother doi:10.1128/AAC.00440-16.

49. **Falgenhauer L, Waezsada SE, Gwozdzinski K, Ghosh H, Doijad S, Bunk B, Sproer C, Imirzalioglu C, Seifert H, Irrgang A, Fischer J, Guerra B, Kasbohrer A, Overmann J, Goesmann A, Chakraborty T.** 2016. Chromosomal Locations of mcr-1 and blaCTX-M-15 in Fluoroquinolone-Resistant Escherichia coli ST410. Emerg Infect Dis **22**.

50. **Tagg KA, Iredell JR, Partridge SR.** 2014. Complete sequencing of IncI1 sequence type 2 plasmid pJIE512b indicates mobilization of blaCMY-2 from an IncA/C plasmid. Antimicrob Agents Chemother **58:**4949-4952.

TABLES

Table 1. Antimicrobial susceptibility profiles of *mcr-1*-positive *Enterobacteriaceae* isolates obtained following selective screening culture of feces on agar supplemented with cefotaxime (2 mg/L).

	Number of isolates resistant (%)														
	Ampicillin	Amoxicillin-clavulanate	Cefotaxime	Ceftazidime	Cefepime	Colistin	Gentamicin	Amikacin	Ertapenem	Imipenem	Meropenem	Fosfomycin	Nitrofurantoin	Ciprofloxacin	Tigecycline
All isolates (n=187)	183 (98)	176 (94)	165 (88)	131 (70)	139 (74)	179 (96)	130 (70)	47 (25)	23 (12)	4 (2)	16 (9)	101 (54)	19 (10)	152 (81)	24 (13)
Sequenced isolates (n=73)	73 (100)	59 (81)	67[a] (92)	60 (82)	57 (78)	71 (97)	54 (74)	20 (27)	6 (8)	0 (0)	0 (0)	43 (59)	5 (7)	63 (86)	9 (12)

[a] six isolates had MICs of ≤1 (susceptible) when tested following selective screening culture. Of these, one harbored *bla*$_{\text{CTX-M-55}}$ and one *bla*$_{\text{CTX-M-65}}$; for the other four, no discernible third generation cephalosporin resistance mechanism was identified following re-culture and DNA extraction for WGS.

FIGURE LEGENDS

Figure 1. Flow diagram summarizing sampling/laboratory/sequencing workflows.

Figure 2. Monthly proportions of *mcr-1*-positive human fecal samples in Guangzhou, China, 2011-2016 for: Panel A, fecal samples harboring *mcr-1*-positive isolates, and Panel B, fecal samples harboring *mcr-1*-positive/cefotaxime-resistant isolates. Black line represents estimated prevalence by iterative sequential regression (ISR), with gaps representing months with missing data. Blue lines at the base of the graph represent 95% confidence intervals around the breakpoints estimated by the ISR model. Panel C, raw counts for each category and sampling denominator (total number of fecal samples), by calendar year (or partial year as specified).

Figure 3. Phylogeny of sequenced *Escherichia coli* study isolates, plus available reference sequences from NCBI (n=11), and associated sequence types (ST; "NF" denotes "not found"); plasmid incompatibility groups; insertion sequences within 5kb of *mcr-1* on either the same contig or contigs associated through the assembly graph; and antimicrobial resistance genes present within the isolates (presence represented by respective coloured shapes).

Figure 4. Alignment of *mcr-1* IncI2 study plasmids (complete plasmids, denoted by * if circularized using SPAdes assembly only, and by ** if circularized using SPAdes+Bandage); incompletely resolved plasmid contigs; and three reference plasmid sequences (sequence labels in bold). Dates of isolation are shown on the right side. Two

plasmid sequences were obtained from isolates cultured from the same patient (sequence labels in blue). Pink/blue cross-links between aligned sequences demonstrate regions of sequence homology (BLASTn matches of ≥500bp length, >95% sequence identity); dark regions within these cross-links demonstrate sequence variation. Tick marks represent 10kb of sequence. Colored arrows demonstrate ORFs of particular relevance; additional coding sequence annotations are shown above reference plasmid SZ02.

Figure 5. Alignment of *mcr-1* IncHI2/HI2A plasmid contigs and two reference plasmid sequences (sequence labels in bold followed by dates of host strain isolation). Pink/blue cross-links between aligned sequences demonstrate regions of sequence homology (BLASTn matches of ≥500bp length, >95% sequence identity); dark regions within these cross-links demonstrate sequence variation. Tick marks represent 10kb of sequence. Colored arrows demonstrate ORFs of particular relevance.

Figure 6. Alignment of *mcr-1* IncX4 study plasmids (complete plasmids, denoted by * if circularized using SPAdes assembly only, and by ** if circularized using SPAdes+Bandage); incompletely resolved plasmid contigs; and three reference plasmid sequences (sequence labels in bold). Dates of isolation are shown on the right side. Pink/blue cross-links between aligned sequences demonstrate regions of sequence homology (BLASTn matches of ≥500bp length, >95% sequence identity); dark regions within these cross-links demonstrate sequence variation. Tick marks represent 10kb of sequence. Colored arrows demonstrate ORFs of particular relevance. Three contigs were excluded as they were small (2.0-8.75kb).