

ARTICLE

DOI: 10.1038/s41467-018-03205-z

OPEN

The global distribution and spread of the mobilized colistin resistance gene *mcr-1*

Ruobing Wang¹, Lucy van Dorp¹, Liam P. Shaw¹, Phelim Bradley³, Qi Wang¹, Xiaojuan Wang¹, Longyang Jin¹, Qing Zhang⁴, Yuqing Liu⁴, Adrien Rieux⁵, Thamarai Dorai-Schneiders⁶, Lucy Anne Weinert⁷, Zamin Iqbal¹, Xavier Didelot⁹, Hui Wang¹ & Francois Balloux^{1,2}

Colistin represents one of the few available drugs for treating infections caused by carbapenem-resistant *Enterobacteriaceae*. As such, the recent plasmid-mediated spread of the colistin resistance gene *mcr-1* poses a significant public health threat, requiring global monitoring and surveillance. Here, we characterize the global distribution of *mcr-1* using a data set of 457 *mcr-1*-positive sequenced isolates. We find *mcr-1* in various plasmid types but identify an immediate background common to all *mcr-1* sequences. Our analyses establish that all *mcr-1* elements in circulation descend from the same initial mobilization of *mcr-1* by an ISAp1 transposon in the mid 2000s (2002–2008; 95% highest posterior density), followed by a marked demographic expansion, which led to its current global distribution. Our results provide the first systematic phylogenetic analysis of the origin and spread of *mcr-1*, and emphasize the importance of understanding the movement of antibiotic resistance genes across multiple levels of genomic organization.

¹Department of Clinical Laboratory, Peking University People's Hospital, Beijing 100044, China. ²UCL Genetics Institute, University College London, Gower Street, London WC1E 6BT, UK. ³Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ⁴Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Sciences, Shandong Province, Jinan 250100, China. ⁵UMR PVBMT, CIRAD, 97410 St Pierre, Reunion, France. ⁶Division of Infection and Pathway Medicine, 49 Little France Crescent, Edinburgh EH16 4SB, UK. ⁷Department of Veterinary Medicine, Cambridge CB3 0ES, UK. ⁸European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge CB10 1SD, UK. ⁹Department of Infectious Disease Epidemiology, Imperial College London, Norfolk Place 21, London W2 1PG, UK. These authors contributed equally: Ruobing Wang, Lucy van Dorp, Liam P. Shaw. Correspondence and requests for materials should be addressed to H.W. (email: wanghui@pkuph.edu.cn) or to F.B. (email: f.balloux@ucl.ac.uk)

Colistin was largely abandoned as a treatment for bacterial infections in the 1970s owing to its toxicity and low renal clearance, but has been reintroduced in recent years as an antibiotic of 'last resort' against multi-drug-resistant infections¹. It is therefore alarming that the prevalence of resistance to colistin has become a significant concern, following the identification of plasmid-mediated colistin resistance conferred by the *mcr-1* gene in late 2015². Colistin resistance is emblematic of the growing problems of antimicrobial resistance worldwide.

Up until 2015, resistance to colistin had only been linked to mutational and regulatory changes mediated by chromosomal genes^{3,4}. The mobilized colistin gene *mcr-1* was first described in a plasmid carried by an *Escherichia coli* isolated in China in April 2011². The presence of colistin resistance on mobile genetic elements poses a significant public health risk, as these can spread rapidly by horizontal transfer, and may entail a lower fitness cost⁵. At the time of writing, *mcr-1* has been identified in numerous countries across five continents. Significantly, *mcr-1* has also been observed on plasmids containing other antimicrobial resistance genes such as carbapenemases^{6–8} and extended-spectrum β-lactamases^{9–11}.

The *mcr-1* element has been characterized in a variety of genomic backgrounds, consistent with the gene being mobilized by a transposon^{12–16}. To date, *mcr-1* has been observed on a wide variety of plasmid types, including IncI2, IncHI2, and IncX4¹⁷. Intensive screening efforts for *mcr-1* have found it to be highly prevalent in a number of environmental settings, including the Haihe River in China¹⁸, recreational water at public urban beaches in Brazil¹⁹, and fecal samples from otherwise healthy individuals^{20,21}. Although both Brazil and China have now banned the use of colistin in agriculture, the evidence that *mcr-1* can spread within hospital environments even in the absence of colistin use²² as well as in the community²¹ raises the possibility that the spread of *mcr-1* will not be contained by these bans.

The global distribution of *mcr-1* over at least five continents is well documented, but little is known about its origin, acquisition, emergence, and spread. In this study, we aim to shed light on these fundamental issues using whole-genome sequencing (WGS) data from 110 novel *mcr-1*-positive isolates from China in conjunction with an extensive collection of publicly available sequence data sourced from the NCBI repository as well as the Short Read Archive (SRA).

Our data and analyses support an initial single mobilization event of *mcr-1* by an *ISAp1-mcr-1-orf-ISAp1* transposon around 2006. The transposon was immobilized on several plasmid backgrounds following the loss of the flanking *ISAp1* elements, and spread through plasmid transfer. The current distribution of *mcr-1* points to a possible origin in Chinese livestock. Our results illustrate the complex dynamics of antibiotic resistance genes across multiple embedded genetic levels (transposons, plasmids, bacterial lineages and bacterial species), previously described as a nested 'Russian doll' model of genetic mobility²³.

Results

Data set. We compiled a global data set of 457 *mcr-1*-positive isolates (Fig. 1a), including 110 new WGS from China, of which 107 were sequenced with Illumina short reads and three with PacBio long-read technology. One hundred and ninety-five isolates were sourced from publicly available assemblies in the NCBI GenBank repository (73 completed plasmids, 1 complete chromosome, 121 assemblies). A further 152 sequences were sourced from the NCBI SRA, after being identified as *mcr-1*-positive using a k-mer index of a snapshot of the SRA as of December 2016 (see Methods). The whole data set consists of 256 short-read data sets, 6 long-read PacBio WGS, 121 draft assemblies, and 74 completed

assemblies. Accession numbers and metadata for the 457 isolates are provided in supplementary data 1.

Isolates carrying *mcr-1* were identified from 31 countries (Fig. 1a). The countries with the largest numbers of *mcr-1*-positive samples are China (212), Vietnam (58) and Germany (25). Within China, nearly half (45%) of positive isolates stem from the Shandong province (Fig. 1b). The vast majority of *mcr-1*-positive isolates belong to *E. coli* (411), but the data set also comprises *mcr-1*-positive isolates from another seven bacterial species: *Salmonella enterica* (29), *Klebsiella pneumoniae* (8), *Escherichia fergusonii* (2), *Kluyvera ascorbata* (2), *Citrobacter braakii* (2), *Cronobacter sakazakii* (1), and *Klebsiella aerogenes* (1) (Fig. 1a). The majority of isolates for which sampling dates were available (80%), were collected between 2012 and 2016, with the oldest available isolates dating back to 2008 (Fig. 1c). Isolates with metadata on the sample source (*n* = 360) came from a range of animal (*n* = 222), human (*n* = 108) and environmental (*n* = 30) hosts.

The large number of *mcr-1*-positive isolates from China, and the high incidence in the Shandong province can be largely ascribed to the inclusion of our 110 newly sequenced isolates including 49 from Shandong and to another 37 isolates from a previous large sequencing effort¹³. However, even after discounting the isolates from these two sources, China remains, together with Vietnam one of the two countries with the highest number of sequenced *mcr-1*-positive isolates.

Evolutionary model. It has been proposed that *mcr-1* is mobilized by a composite transposon formed of a ~2600 bp region containing *mcr-1* (1626 bp) and a putative open reading frame encoding a PAP2 superfamily protein (765 bp), flanked by two *ISAp1* insertion sequences¹². *ISAp1* is a member of the IS30 family of insertion sequences, which utilize a 'copy-out, paste-in' mechanism with a targeted transposition pathway requiring the formation of a synaptic complex between an inverted repeat (IR) in the transposon circle and an IR-like sequence in the target. Snesrud and colleagues¹² hypothesized that after the initial formation of such a composite transposon, these insertion sequences would have been lost over time, leading to the stabilization of *mcr-1* in a diverse range of plasmid backgrounds (Fig. 2). In the following, we sought to test this model by performing an explicit phylogenetic analysis of the region surrounding *mcr-1* using our comprehensive global data set.

Immediate genomic background of *mcr-1*. If there had been a unique formation event for the composite transposon, followed by progressive transposition and loss of insertion sequences, we would expect to be able to identify a common immediate background region for *mcr-1* in all samples. Indeed, we were able to identify and align a shared region or remnants of it in all 457 sequences surrounding *mcr-1* (see Methods), supporting a single common origin for all *mcr-1* elements sequenced to date (Fig. 3a). The majority of the sequences contained no trace of *ISAp1* (*n* = 260) indicating that the *mcr-1* transposon had been completely stabilized in their genomic background. Forty-two sequences contained indication of the presence of *ISAp1* both upstream and downstream, either in full copies (*n* = 16), a full copy upstream and a partial copy downstream (*n* = 7), a partial copy upstream and a full copy downstream (*n* = 1), or partial copies upstream and downstream (*n* = 18). Some sequences only had *ISAp1* present upstream as a complete (*n* = 55) or partial (*n* = 99) sequence, and one sequence had only a partial downstream *ISAp1* element. The downstream copy of *ISAp1* was inverted in some sequences (*n* = 3) and some sequences had full

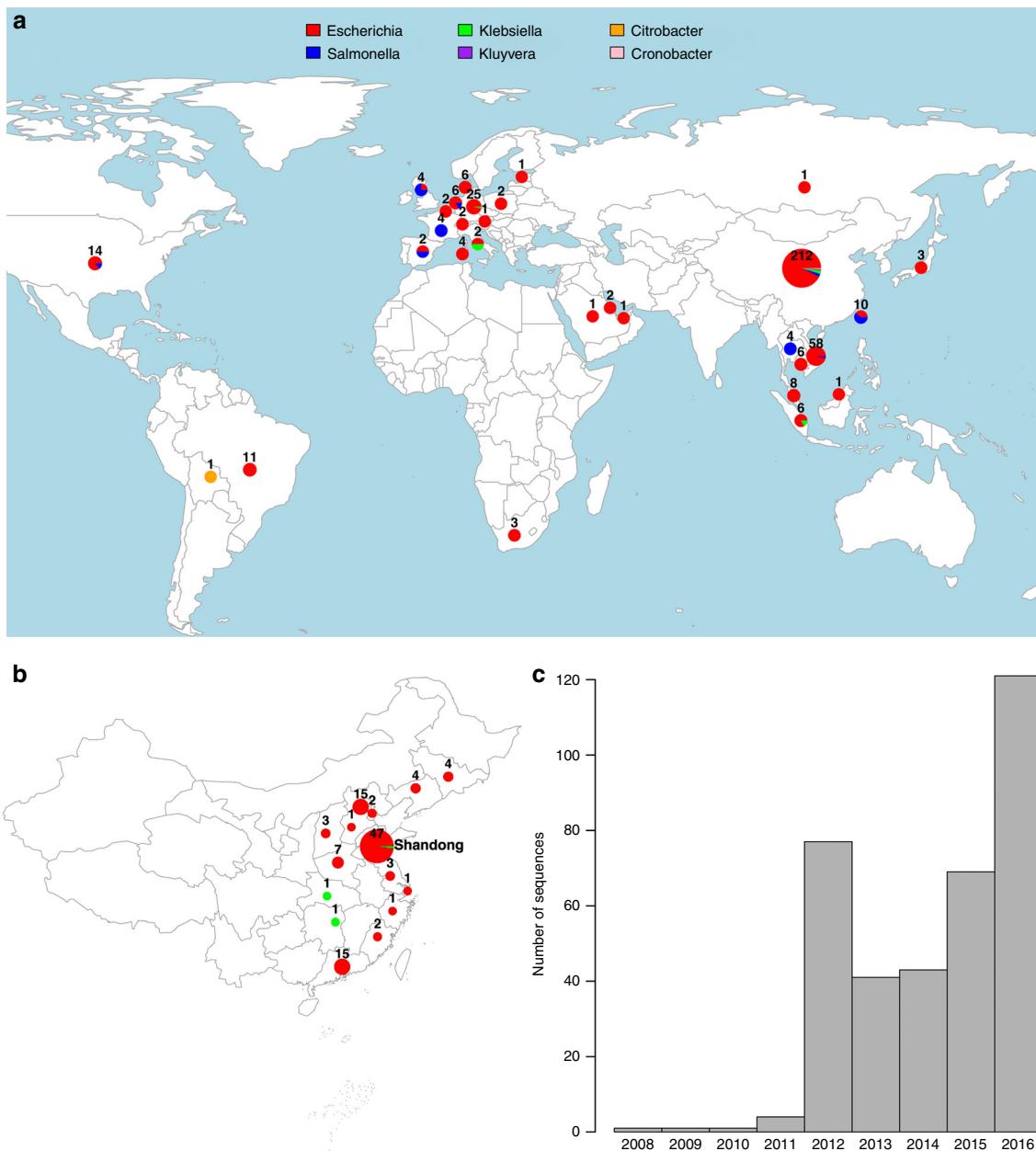


Fig. 1 Overview of the *mcr-1*-positive isolates included. **a** Global map of *mcr-1*-positive isolates included colored by genus with the number and size of pies providing the sample size per location; **b** Map of novel Chinese isolates sequenced for this study; **c** Histogram of sampling dates (years) of the isolates. Maps were created using the R package *rworldmap* using the public domain Natural Earth data set

copies of *ISAp1* present elsewhere on the same contig ($n = 7$), consistent with its high observed activity in transposition²⁴.

Further inspection of the transposon alignment revealed that the 186 bp region between the 3'-end of the upstream *ISAp1* and *mcr-1* contained IR-like sequences similar to the IRR and IRL of *ISAp1* (respectively: 93–142 bp, 23/50 identity; and 125–175 bp, 21/50 identity). The most variable positions in this 186 bp region were at 177 bp and 142 bp, approximately coinciding with the end of the alignment with the IRRs and were more variable in sequences lacking *ISAp1*, suggesting possible loss of function of the transposition pathway associated with *ISAp1* (Fig. 3d). Some of these SNPs occurred in a stretch previously identified as the promoter region for *mcr-1*²⁵, and this region showed strong signals of recombination. A small number of sequences (3%) had SNPs present in *mcr-1* itself. These tended to be at the upstream/

5'-end of the sequence, particularly in the first three positions. A subset of the sequences from Vietnam ($n = 28$) included a secondary 1.7 kb insertion downstream of *mcr-1* containing a putative transposase, indicating subsequent rearrangements involving this region after initial mobilization of the transposon (Fig. 3e).

To reconstruct the phylogenetic history of the composite *mcr-1* transposon, we created a sequence alignment for 457 sequences (Fig. 3c) after removal of recombinant regions identified with ClonalFrameML, including the region immediately upstream of *mcr-1* between positions 1212–1247 (Fig. 3d). A midpoint-rooted maximum parsimony phylogeny showed that there was a dominant sequence type with subsequent diversification, likely indicating the ancestral form of the composite transposon (Fig. 4). There was no discernible clustering of isolates by bacterial species

(Supplementary Fig. 1) or sample source (Supplementary Fig. 2), suggesting the composite transposon does not evolve differently in these different backgrounds.

A Bayesian dating approach (BEAST) was applied to infer a timed phylogeny of the maximal alignable region of the *mcr-1* carrying transposon (see Methods). Based on this 3522 site alignment we infer a common ancestor for 364 dated isolates in 2006 (Supplementary Fig. 3; 2002–2008 95% highest posterior density (HPD) with a strict clock and coalescent model) with a mutation rate around 7.51×10^{-5} substitutions per site per year (Supplementary Table 1). There was no clear overall geographic clustering in the Maximum Clade Credibility (MCC) tree (Supplementary Figure 4).

Wider genomic background of *mcr-1*. Next, we explored the wider genomic background upstream and downstream of the conserved transposon sequences. We had sufficiently long assembled contigs for 182 isolates to identify plasmid types based on co-occurrence with plasmid replicons (see Methods) and identified *mcr-1* in 13 different plasmid backgrounds. IncI2 and IncX4 were the dominant plasmid types, accounting for 51 and 38% of the isolates, respectively (Fig. 5) similar to the proportions observed by Matamoros et al.¹⁵. One isolate in our data set was definitively located on a complete chromosome. Though, we cannot rule out the presence of a few other chromosomal copies of *mcr-1* located on short contigs.

The distribution of transposons carrying one or two copies of *ISAp1* was highly heterogeneous across these plasmid types. For example, sequences with one or two copies of *ISAp1* were found on six and four types, respectively, which supports their mobility compared with those without *ISAp1*, which were found in five plasmid types. Of the contigs carrying one copy of *ISAp1*, 61% were found in IncI2 plasmids, and 50% of contigs carrying two copies of *ISAp1* belonged to IncHI2 plasmids. Conversely, the common IncX4 plasmids carried only two transposons with two copies of *ISAp1* and none with a single copy of the element.

We identified two extended plasmid backbone sequences that could be aligned. The first such alignment encompasses a shared sequence of 7161 bp between 108 plasmid backgrounds and has been previously referred to as 'Type A'¹³. These sequences contain 54 sequences co-occurring with an IncI2 replicon, with 54 of unknown plasmid type, and encompass a large fraction of the genetic diversity found in the *mcr-1* transposon, although a large proportion (9/108) belonged to the dominant sequence type (i.e., B07_1_WFA61 in Fig. 4). The second alignment is 34,761 bp long and is common to nine IncX4 plasmids and partly overlaps with a background previously defined as 'Type D'¹³.

We applied BEAST to infer a timed phylogeny for each of these alignable regions after removal of SNPs showing evidence of recombination. For the IncI2 background we infer that a common ancestor to all 108 isolates existed in 2006 (1998–2010 95% CI relaxed exponential clock model) assuming a constant population size model (Supplementary Fig. 5). For the IncX4 backgrounds we dated the common ancestor of the eight isolates to 2011 (2010–2013 95% CI relaxed exponential clock model) assuming a constant population size model (Supplementary Fig. 6). Posterior density distributions of root dating under different population models are shown in Supplementary Figs. 7–8. The difference in dating inferred for these two plasmid backgrounds and the recent date obtained for IncX4 highlight the dynamic nature of the integration of the *mcr-1* carrying transposon, even if in the IncX4 phylogeny isolates from East Asia and Europe and the Americas cluster together. The inferred mutation rates obtained for the IncI2 and IncX4 backgrounds consistently lie around $5\text{--}10 \times 10^{-5}$ substitutions per site per year (Supplementary Table 1).

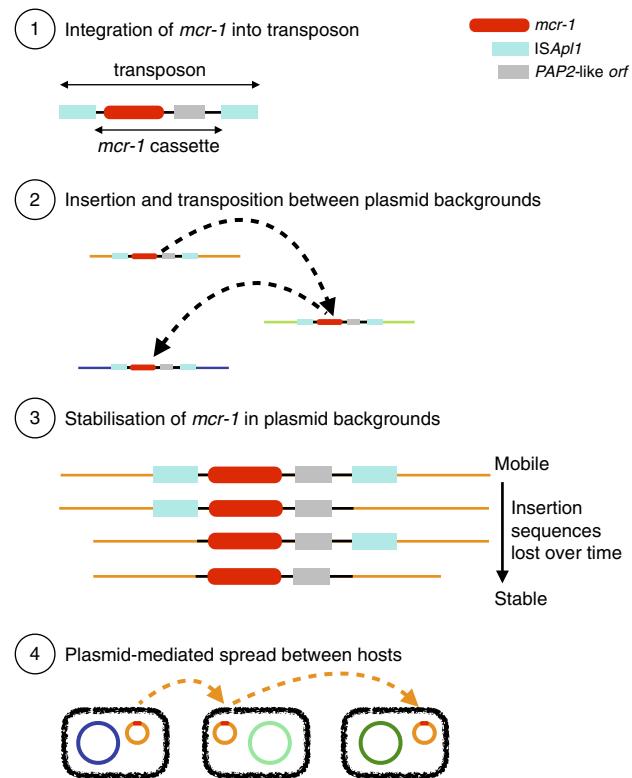


Fig. 2 Schematic representation of the evolutionary model for the steps in the spread of the *mcr-1* gene. (1) The formation of the original composite transposon, followed by (2) transposition between plasmid backgrounds and (3) stabilization via loss of *ISAp1* elements before (4) plasmid-mediated spread

Environmental distribution of the composite *mcr-1* transposon. It has been suggested that agricultural use of colistin, as has been widespread in China since the early 1980s²¹, caused the initial emergence and spread of *mcr-1*^{26,27}. According to the evolutionary model in Fig. 2, the ancestral mobilizable state is represented by the transposon carrying both its *ISAp1* elements. The transposon is thought to lose its capability for mobilization after the loss of both *ISAp1* elements, although a single copy is reportedly sufficient to keep some ability to mobilize, with the upstream copy being functionally more important¹². Comparing human ($n = 108$) and non-human ($n = 252$) isolates, there were significantly more sequences with some trace of the insertion sequence *ISAp1* both upstream and downstream in non-human isolates (32/220 vs. 5/108, χ^2 -test, $p = 0.033$). This comparison held when only comparing agricultural isolates to human isolates ($n = 213$) (28/213 vs. 5/108, χ^2 -test, $p = 0.029$). Furthermore, of the 42 isolates that had *ISAp1* fragments both upstream and downstream, the majority were from Asia ($n = 30$) with only a quarter from Europe ($n = 10$) (χ^2 test, $p = 0.12$). This result is not driven by an over representation of agricultural isolates from Asia in our data set (χ^2 test, $p = 0.38$).

Discussion

We assembled a global data set of 457 *mcr-1*-positive sequenced isolates and could show that there was a single integration event of *mcr-1* into an *ISAp1* composite transposon, followed by its subsequent spread between multiple genomic backgrounds. Our phylogenetic analyses point to a date for the insertion of *mcr-1* into the gene transposon shared across our isolates in the mid 2000s (2002–2008 95% HPD). We could identify the likely sequence of the ancestral transposon type and show the pattern of

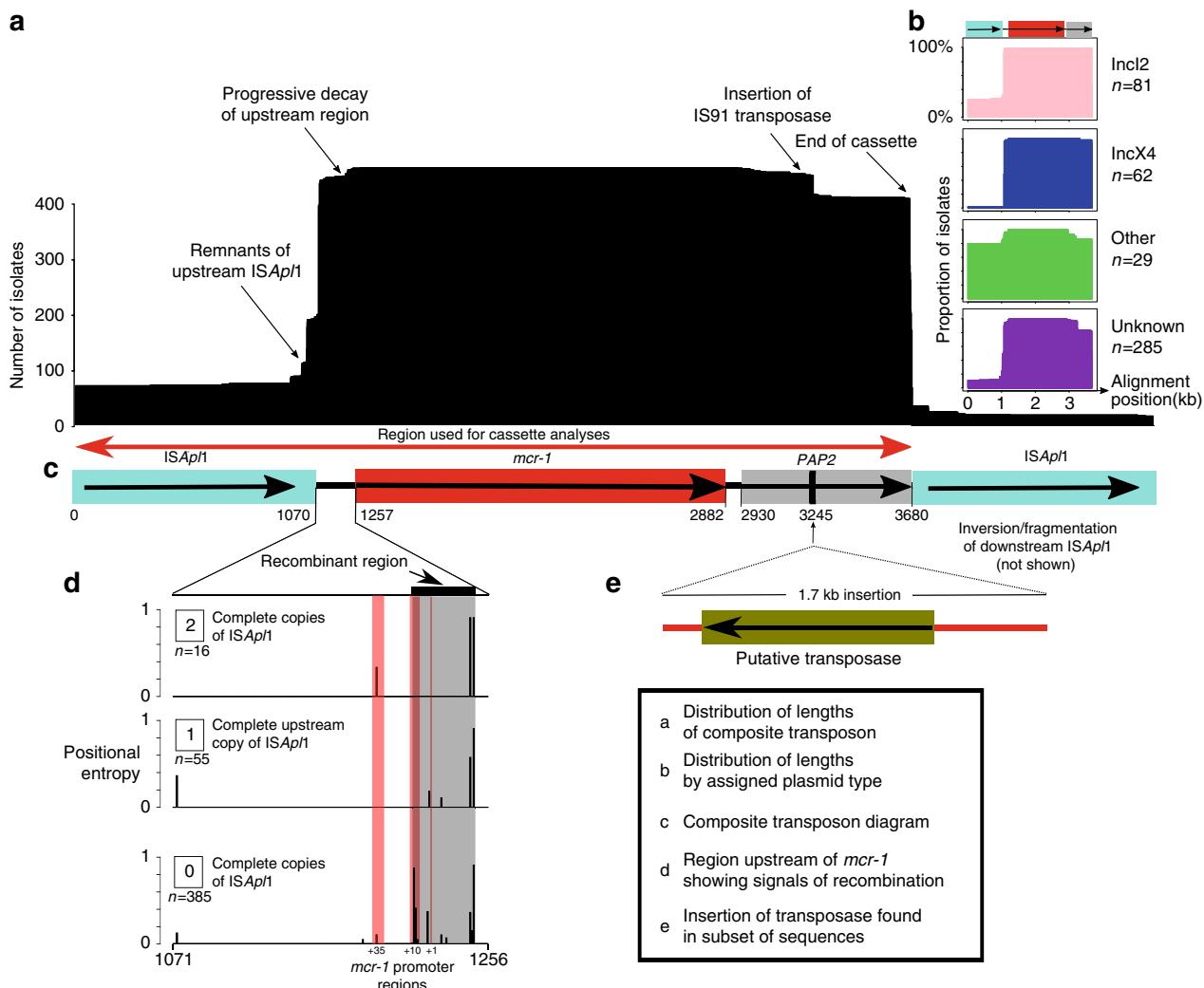


Fig. 3 The genetic element carrying *mcr-1* is a composite transposon and is alignable across our global data set. **a** Length distribution of the alignment across sequences. **b** Length distribution subset by plasmid type. **c** The composite transposon, consisting of ISAp1, *mcr-1*, a PAP2 orf, and ISAp1. The region indicated by the red arrow was used in phylogenetic analyses, after the removal of recombination. **d** The 186 bp region upstream of *mcr-1* showed strong signals of recombination (gray box) that coincided with the promoter regions of *mcr-1* (red box), and this diverse region was removed from the subsequent alignment. **e** Twenty-eight sequences from Vietnam had a 1.7 kb insertion containing a putative transposase, suggesting subsequent rearrangement after initial mobilization

diversity supports a single mobilization with subsequent diversification during global spread.

Despite the limited number of WGSs for samples before 2012, with the oldest sequence available from 2008 (Fig. 1c), our estimate is consistent with the majority of available evidence from retrospective surveillance data,²⁶ which has found the presence of *mcr-1* in samples dating back to 2005 in Europe²⁸. One retrospective study of Chinese isolates from 1970–2014 reported three *mcr-1*-positive *E. coli* dating from the 1980s²⁹, although *mcr-1* then did not reappear until 2004. This observation seems surprising in light of our results, which clearly exclude such an early spread of *mcr-1* at least on this ISAp1 transposon background.

Our estimates of the age of spread of the representative IncI2 and IncX4 plasmid backgrounds are more recent, dating to ~2008 and 2013, respectively, but are both consistent with the age of the transposon mobilization event. We did not constrain the evolutionary rates in any of our phylogenetic analyses. It is thus encouraging that the different rates are highly consistent between the *mcr-1* transposon and the two plasmid backgrounds. Although this points to high internal consistency between our

estimates, we were surprisingly unable to find any previously published estimates for the evolutionary rate of bacterial plasmids.

The current distribution and observed genetic patterns are in line with a center of origin in China. This is the place where we observe the highest proportion of isolates carrying intact or partial copies of the ISAp1 flanking elements. Transposon sequences carrying ISAp1 elements were also overrepresented in environmental and agriculture isolates, relative to those collected from humans. This pattern is in line with agricultural settings acting as the source of *mcr-1* within bacteria isolated from humans²¹. The current global distribution has been achieved through multiple translocations, and is illustrated by the interspersed geographic origins in our phylogenetic reconstructions. A likely driver for the global spread is trade, in particular food animals³⁰ and meat, although direct global movement by colonized or infected humans²⁰ is also likely to have played a role in the current distribution.

The origin of *mcr-1* prior to its mobilization remains elusive. Despite an exhaustive search of sequence repositories, including

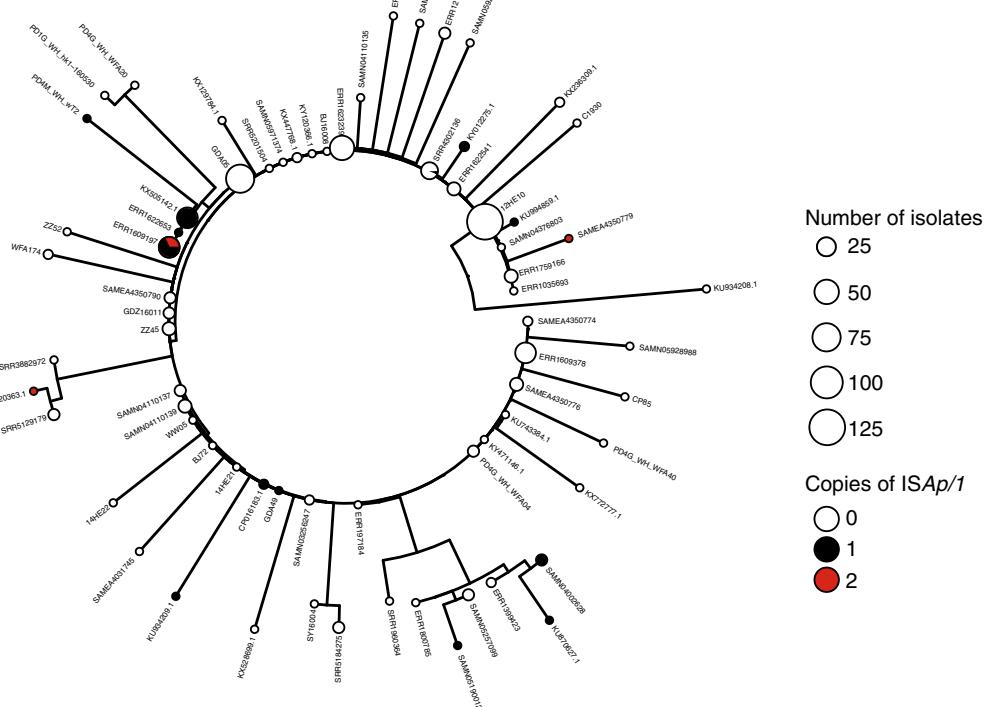


Fig. 4 Phylogeny of the *mcr-1* composite transposon indicates a dominant sequence type with subsequent diversification. Midpoint-rooted maximum parsimony phylogeny based on the 3522 bp alignment of 457 sequences (recombinant regions removed). Size of points indicates the number of identical sequences, with a representative sequence for each shown next to each tip

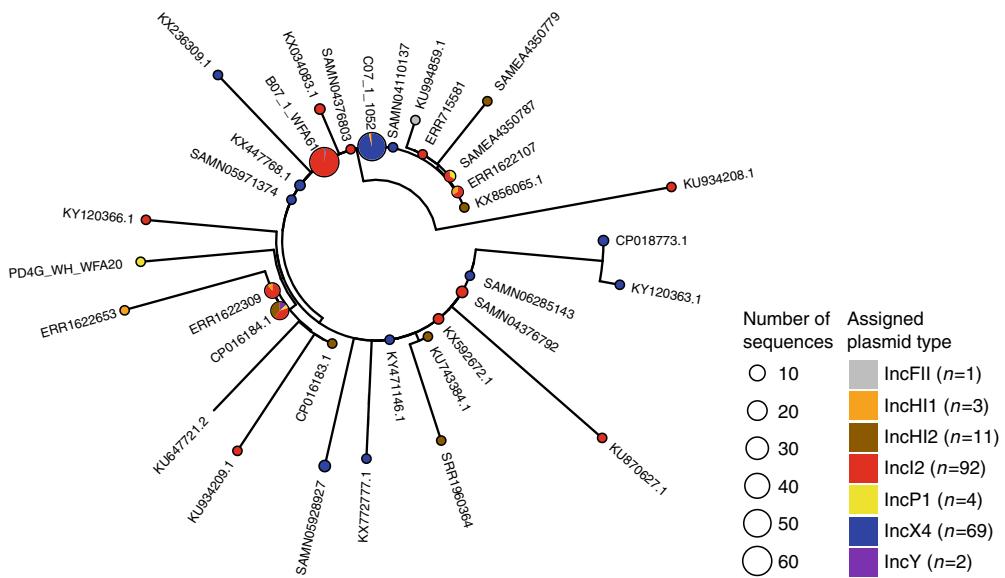


Fig. 5 The distribution of plasmid types shown on the transposon phylogeny. Maximum parsimony tree (homoplastic sites removed, midpoint rooted, as in Fig. 4) based on the composite transposon alignment for 172 sequences containing a plasmid replicon on the same contig i.e. those with an assigned plasmid type (color). Incl2 and IncX4 are the most common plasmid types. An example sequence ID is shown for each unique sequence

the short-read archive, we found not a single *mcr-1* sequence outside the *ISAp1* transposon background. *ISAp1* was first identified in the pig pathogen *Actinobacillus pleuropneumoniae*³¹ suggesting that it may also have been an ancestral host for *mcr-1*, although to our knowledge no *mcr-1*-positive *A. pleuropneumoniae* isolates have been described. The phosphoethanolamine transferase from *Paenibacillus sophorae* has

also been proposed as a possible candidate³². However, this seems most unlikely as *Paenibacilli* are Gram-positive and are thus intrinsically resistant to polymyxins³³. Moreover, although the two sequences share functional similarities, this should be interpreted as a case of possible parallel evolution rather than direct filiation³³. *Moraxella* has also been suggested as being the source of *mcr-1*³⁴, following the identification of genes in *Moraxella* with

limited homology to *mcr-1* (~60% nucleotide sequence identity). However, this sequence identity is too low for *Moraxella* to be considered as viable candidates for the origin of *mcr-1*. The search for the initial source of *mcr-1* remains open until a *mcr-1* sequence is identified outside of the *ISAp1* sequence background.

We note that there are an increasing number of mobilized genes that can confer colistin resistance, with *mcr-2* reported less than a year after *mcr-1* was initially described³⁵ and more recent descriptions of the phylogenetically distant *mcr-3*, *mcr-4*, and *mcr-5*^{36–38}. There appear to be commonalities between the mechanisms of the *mcr* genes, despite their different sequences and locations near to different insertion sequences. For example, *mcr-2* has 76.7% nucleotide identity to *mcr-1* and was found in colistin-resistant isolates that did not contain *mcr-1*, and appeared to be mobilized on an IS1595 transposon³⁵. Despite the different insertion sequences, intriguingly, this mobile element also contained a similar protein downstream of the *mcr* gene. Indeed, in *mcr-1*, -2, and -3, the *mcr* gene has a downstream open reading frame encoding, respectively, a putative PAP2 protein¹², a PAP2 membrane-associated lipid phosphatase³⁵, and a diacylglycerol kinase³⁶, all of which have transmembrane domains and are involved in the phosphatidic acid pathway^{39,40}. Although the PAP2-like orf in *mcr-1* has been shown not to be required for colistin resistance⁴¹, the presence of similar sequences downstream of other *mcr* genes implies some functional role, either in the formation of the mobile element and/or in its continued mobilization.

In summary, we assembled the largest data set to date of *mcr-1*-positive sequenced isolates through our own sequencing efforts combined with an exhaustive search of publicly available sequence databases including unassembled data sets from the SRA. Although this allowed us to obtain a truly global data set of 457 *mcr-1*-positive isolates covering 31 countries and five continents, we appreciate that the data is likely affected by complex sampling biases, with an over representation of samples from places with active surveillance and well-funded research communities. Equally, although we took advantage of the most sophisticated bioinformatics and phylogenetic tools currently available, the complex ‘Russian doll’ dynamics of the transposon, plasmids, and bacterial host limits our ability to reach strong inferences on some important aspects of the spread of *mcr-1*. Nevertheless, we believe our results highlight the potential for phylogenetic reconstruction of antimicrobial resistance elements at a global scale. We hope that future efforts relying on more sophisticated computational tools and even more extensive genetic sequence data will become part of the routine toolbox in infectious disease surveillance.

Methods

Compilation of genomic data set. We blasted for *mcr-1* in all NCBI GenBank assemblies (as of 16th March 2017, $n = 90,759$) using a 98% identity cutoff. 195 records (0.21%; 121 assemblies, 73 complete plasmids, one complete chromosome) contained at least one contig with a full-length hit to *mcr-1* (1626 bases). We only included samples with a single copy of *mcr-1*. The only isolate with multiple copies was a previously published isolate with three chromosomal copies of *mcr-1* and seven copies of *ISAp1*⁴².

We searched a snapshot of all WGS bacterial raw read data sets in the SRA (December 2016), looking for samples containing *mcr-1* by using a k-mer index ($k = 31$), which we had previously constructed⁴³, software available at: <https://github.com/phelimb/bigsi>. A total of 184 data sets were found to contain at least 70% of the 31-mers in *mcr-1*. After removing duplicates (i.e., those with a draft assembly available) we could assemble contigs with *mcr-1* for 152 of these.

Our final data set comprised 457 isolates from six genera across 31 different countries, ranging in date from 2008 to 2017. Where only a year was provided as the date of isolate collection the date was set to the midpoint of that year.

Whenever identified isolates did not comprise previously assembled genomes or complete plasmids, the raw fastq files were first inspected using FastQC and trimmed and filtered on a case-by-case basis. De novo assembly was then conducted using Plasmid SPADES 3.10.0 using the *-careful* switch and otherwise

default parameters⁴⁴. For those isolates sequenced using PacBio a different pipeline was employed. Correction, trimming, and assembly of raw reads was performed using Canu⁴⁵ and assembled reads were corrected and trimmed using the tool Circlator⁴⁶. The quality of resultant assemblies was assessed using infseq. In both cases the *mcr-1* carrying contigs in this final data set were identified using blastn v2.2.31⁴⁷.

We ran Plasmid Finder 1.3⁴⁸ with 95% identity to identify plasmid replicons on the *mcr-1* carrying contigs. In total, 182 unique contigs could be assigned a plasmid type using this method.

Novel samples from China. We selected 110 *mcr-1*-positive isolates from China for WGS from a larger survey effort of both clinical and livestock isolates. A total of 2824 non-repetitive clinical isolates, including 1637 *E. coli* and 1187 *K. pneumoniae* were collected from 15 provinces of mainland China from 2011 to 2016. Seventy-two isolates were resistant to polymyxin B, comprising 40 *E. coli* and four *K. pneumoniae* carrying *mcr-1*. Livestock samples were collected from four provinces of China in 2013 and 2016. One broiler farm of the Shandong province provided chicken anal swabs, liver, heart, and wastewater isolated in 2013. In 2016, samples including feces, wastewater, anal swabs, and internal organs of sick livestock were collected from swine farms, cattle farms, and broiler farms in four provinces (Jilin, Shandong, Henan, and Guangdong). A total of 601 *E. coli* and 126 *K. pneumoniae* were isolated, of which 167 (137 *E. coli* and 30 *K. pneumoniae*) were resistant to polymyxin B. We detected *mcr-1* in 135 *E. coli* and two *K. pneumoniae*, as well as in eight *E. coli* isolated from environmental samples, which were collected from influents and effluents of four tertiary care teaching hospitals.

All of the isolates were sent to the microbiology laboratory of Peking University People’s Hospital and confirmed by routine biochemical tests, the Vitek system (bioMérieux, Hazelwood, MO, USA) and/or MALDI-TOF (Bruker Daltonics, Bremen, Germany). The minimal inhibitory concentrations (MICs) of polymyxin B was determined using the broth dilution method. The breakpoints of polymyxin B for *Enterobacteriaceae* were interpreted with the European Committee on Antimicrobial Susceptibility Testing (EUCAST, http://www.eucast.org/clinical_breakpoints) guidelines. Colistin-resistant isolates (MIC of $\geq 2 \mu\text{g/ml}$) were screened for *mcr-1* by PCR and sequencing as described previously⁴⁹.

Identification and alignment of *mcr-1* transposon. We searched for the *mcr-1* carrying transposon region across isolates by blasting for its major components: *ISAp1* (*Actinobacillus pleuropneumoniae* reference sequence: EF407820), *mcr-1* (from *E. coli* plasmid pHNSHP45: KP347127.1), and short sequences representing the sequences immediately upstream and downstream of *mcr-1* (from KP347127.1) using blastn-short. Contiguous sequences containing *mcr-1* were aligned using Clustal Omega⁵⁰ and then manually curated and amended using jalview⁵¹, resulting in a 3679 bp alignment containing the common ~2600 bp identified by Snesrud and colleagues¹². The downstream copy of *ISAp1* was more often fragmented or inverted. Twenty-eight isolates, which were all assemblies from the same study in Vietnam had a ~1.7 kb insertion downstream of *mcr-1* (Fig. 3e) before the downstream *ISAp1* element.

Phylogenetic analyses. For constructing the transposon phylogeny, we excluded the downstream *ISAp1* and the insertion sequence observed in a small number of samples, as well as regions identified as having signals of recombination by ClonalFrameML⁵², resulting in a 3522 bp alignment. We removed two homoplastic sites (requiring >1 change on the phylogeny), before constructing a maximum parsimony neighbor-joining tree based on the Hamming distance between sequences. We calculated branch lengths using non-negative least squares with nnls.phylo in phangorn v2.2.0⁵³. Phylogenies were visualized with ggtree v1.8.1⁵⁴.

Phylogenetic dating. Given recombination can conceal the clonal phylogenetic signal we also applied ClonalFrameML⁵² to identify regions of high recombination in a subset of IncI2 and IncX4 plasmid background alignments. Where recombination hotspots were identified, they were removed from the alignment. In the IncI2 alignment, this resulted in removing 1281 positions. No regions of high recombination were detected in the IncX4 alignment. We applied root-to-tip correlations to test for a temporal signal in the data using TempEST⁵⁵. There was a significantly positive slope for all three alignments (Supplementary Fig. 9–11).

We applied BEAUTi and BEAST v2.4.7^{56,57} to estimate a timed phylogeny from an alignment of IncI2 plasmids (7161 sites, 110 isolates) and IncX4 plasmids (34,761 sites, 8 isolates). Sequences were annotated using their known sampling times expressed in years. For both plasmid alignments, the HKY substitution model was selected based on evaluation of all possible substitution models in bModelTest. BEAST analyses were then applied under both a coalescent population model (the coalescent Bayesian skyline implementation) and an exponential growth model (Coalescent Exponential population implementation). In addition, a strict clock, with a lognormal prior, and a relaxed clock (both lognormal and exponential) were tested. MCMC was run for 50,000,000 iterations sampling every 2000 steps and convergence was checked by inspecting the effective sample sizes and parameter value traces in the software Tracer v1.6.0. Analyses were repeated three times to ensure consistency between the obtained posterior distributions. Posterior trees for the best fitting model were combined in TreeAnnotator after a

10% burn-in to provide an annotated MCC tree. MCC trees were plotted using ggTree⁵⁴ for both backgrounds: IncI2 (Supplementary Fig. 12) and IncX4 (Supplementary Fig. 13). The model fit across analyses was compared using the Akaike's information criteria model through 100 bootstrap resamples as described in Baele and colleagues⁵⁸ and implemented in Tracer v1.6 (Supplementary Table 2).

Phylogenetic dating on the transposon was performed using an alignment of 364 isolates, which included only those with information on isolation date, across 3522 sites. As before BEAST analyses were applied under both a coalescent population model (coalescent Bayesian skyline implementation) and an exponential growth model (coalescent exponential population implementation). In addition, a strict clock, with a lognormal prior, and a relaxed clock (both lognormal and exponential) were tested. Analyses were run under a HKY substitution model for 600 million iterations sampling every 5000 steps. Only analyses using a strict clock model reached convergence after 600 million iterations. The resultant set of trees were thinned by sampling every 10 trees and excluding a 10% burn-in and combined using TreeAnnotator to produce a MCC tree. MCC trees were plotted using ggTree⁴⁸. As before the model fit was evaluated using AICM's implemented in Tracer v1.6.

Environmental distribution. For the purpose of testing the distribution of sequences containing some trace of ISApI1, we classed isolates into broad categories as either environmental ($n = 39$; bird, cat, dog, fly, food, penguin, reptile, vegetables), agricultural ($n = 213$; chicken, cow, pig, poultry feed, sheep, turkey), or human ($n = 108$). We did not correct for study site with subsampling as we found great diversity within sites, consistent with a recent study showing multiple diverse *mcr-1*-positive strains within a single hospital sewage sample⁵⁹.

Data access. All the data generated or analyzed in this study are available within the paper and its supplementary information files. Accession numbers and metadata for all 457 isolates is provided in supplementary data 1. The newly sequenced 110 *mcr-1*-positive genomes have been submitted to the Short Read Archive under Bioproject: PRJNA408214, Accession: SRP118547.

Received: 6 October 2017 Accepted: 29 January 2018

Published online: 21 March 2018

References

- Grégoire, N., Aranzana-Climent, V., Magréault, S., Marchand, S. & Couet, W. Clinical pharmacokinetics and pharmacodynamics of colistin. *Clin. Pharmacokinet.* **56**, 1441–1460 (2017).
- Liu, Y.-Y. et al. 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 (2016).
- Olaitan, A. O., Morand, S. & Rolain, J.-M. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. *Front. Microbiol.* **5**, 643 (2014).
- Lee, J.-Y., Choi, M.-J., Choi, H. J. & Ko, K. S. Preservation of acquired colistin resistance in Gram-negative bacteria. *Antimicrob. Agents Chemother.* **60**, 609–612 (2016).
- Carattoli, A. Plasmids and the spread of resistance. *Int. J. Med. Microbiol.* **303**, 298–304 (2013).
- Poirel, L., Kieffer, N., Liassine, N., Thanh, D. & Nordmann, P. Plasmid-mediated carbapenem and colistin resistance in a clinical isolate of *Escherichia coli*. *Lancet Infect. Dis.* **16**, 281 (2016).
- Du, H., Chen, L., Tang, Y.-W., Kreiswirth, B. N. & Rolain, J.-M. Emergence of the *mcr-1* colistin resistance gene in carbapenem-resistant Enterobacteriaceae. *Lancet Infect. Dis.* **16**, 287–288 (2016).
- Yao, X., Doi, Y., Zeng, L., Lv, L. & Liu, J.-H. Carbapenem-resistant and colistin-resistant *Escherichia coli* co-producing NDM-9 and MCR-1. *Lancet Infect. Dis.* **16**, 288–289 (2016).
- Zhang, X.-F. et al. Possible transmission of *mcr-1* -harboring *Escherichia coli* between companion animals and human. *Emerg. Infect. Dis.* **22**, 1679–1681 (2016).
- Falgenhauer, L. et al. Colistin resistance gene *mcr-1* in extended-spectrum β -lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany. *Lancet Infect. Dis.* **16**, 282–283 (2016).
- Haenni, M. et al. Co-occurrence of extended spectrum β -lactamase and MCR-1 encoding genes on plasmids. *Lancet Infect. Dis.* **16**, 281–282 (2016).
- Snesrud, E. et al. A model for transposition of the colistin resistance gene *mcr-1* by ISApI1. *Antimicrob. Agents Chemother.* **60**, 6973–6976 (2016).
- Wang, Y. et al. Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1 in Chinese poultry production. *Nat. Microbiol.* **2**, 16260 (2017).
- Li, R. et al. Genetic characterization of *mcr-1* -bearing plasmids to depict molecular mechanisms underlying dissemination of the colistin resistance determinant. *J. Antimicrob. Chemother.* **72**, 393–401 (2017).
- Matamoros, S. et al. Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the *mcr-1* gene indicates bacterial diversity but plasmid restriction. *Sci. Rep.* **7**, 15364 (2017).
- Zhou, H.-W. et al. Occurrence of plasmid- and chromosome-carried *mcr-1* in waterborne *Enterobacteriaceae* in China. *Antimicrob. Agents Chemother.* **61**, e00017–e00017 (2017).
- Matamoros, S. et al. Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the *mcr-1* gene indicates bacterial diversity but plasmid restriction. *Sci. Rep.* **7**, 15364 (2017).
- Yang, D. et al. The occurrence of the colistin resistance gene *mcr-1* in the Haihe River (China). *Int. J. Environ. Res. Public Health* **14**, 576 (2017).
- Fernandes, M. R. et al. Colistin-resistant *mcr-1* -positive *Escherichia coli* on public beaches, an infectious threat emerging in recreational waters. *Antimicrob. Agents Chemother.* **61**, e00234–17 (2017).
- von Wintersdorff, C. J. H. et al. Detection of the plasmid-mediated colistin-resistance gene *mcr-1* in faecal metagenomes of Dutch travellers. *J. Antimicrob. Chemother.* **71**, 3416–3419 (2016).
- Wang, Y. et al. 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.* **17**, 390–399 (2017).
- Tian, G.-B. et al. MCR-1-producing *Klebsiella pneumoniae* outbreak in China. *Lancet Infect. Dis.* **17**, 577 (2017).
- 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 (2016).
- Snesrud, E. et al. Analysis of serial isolates of *mcr-1*-positive *Escherichia coli* reveals a highly active ISApI1 transposon. *Antimicrob. Agents Chemother.* **61**, e00056–17 (2017).
- Poirel, L. et al. Genetic features of *mcr-1*-producing colistin-resistant *Escherichia coli* isolates in South Africa. *Antimicrob. Agents Chemother.* **60**, 4394–4397 (2016).
- Poirel, L. & Nordmann, P. Emerging plasmid-encoded colistin resistance: the animal world as the culprit? *J. Antimicrob. Chemother.* **71**, 2326–2327 (2016).
- Schwarz, S. & Johnson, A. P. Transferable resistance to colistin: a new but old threat. *J. Antimicrob. Chemother.* **71**, 2066–2070 (2016).
- Haenni, M. et al. Co-occurrence of extended spectrum β -lactamase and *mcr-1* encoding genes on plasmids. *Lancet Infect. Dis.* **16**, 281–282 (2016).
- Shen, Z., Wang, Y., Shen, Y., Shen, J. & Wu, C. Early emergence of *mcr-1* in *Escherichia coli* from food-producing animals. *Lancet Infect. Dis.* **16**, 293 (2016).
- Grami, R. et al. Impact of food animal trade on the spread of *mcr-1* -mediated colistin resistance, Tunisia, July 2015. *Eurosurveillance* **21**, 30144 (2016).
- Tegetmeyer, H. E., Jones, S. C. P., Langford, P. R. & Baltes, N. ISApI1, a novel insertion element of *Actinobacillus pleuropneumoniae*, prevents ApxIV-based serological detection of serotype 7 strain AP76. *Vet. Microbiol.* **128**, 342–353 (2008).
- Gao, R. et al. Dissemination and Mechanism for the MCR-1 Colistin Resistance. *PLOS Pathog.* **12**, e1005957 (2016).
- Di Conza, J. A., Radice, M. A. & Gutkind, G. O. *mcr-1*: rethinking the origin. *Int. J. Antimicrob. Agents* **50**, 737 (2017).
- Kieffer, N., Nordmann, P. & Poirel, L. *Moraxella* species as potential sources of *mcr*-like polymyxin resistance determinants. *Antimicrob. Agents Chemother.* **61**, e00129–17 (2017).
- Xavier, B. B. et al. Identification of a novel plasmid-mediated colistin-resistance gene, *mcr-2*, in *Escherichia coli*, Belgium, June 2016. *Eurosurveillance* **21**, 30280 (2016).
- Yin, W. et al. Novel plasmid-mediated colistin resistance gene *mcr-3* *Escherichia coli*. *MBio* **8**, e00543–17 (2017).
- Carattoli, A. et al. Novel plasmid-mediated colistin resistance *mcr-4* gene in *Salmonella* and *Escherichia coli*, Italy 2013, Spain and Belgium, 2015 to 2016. *Eurosurveillance* **22**, 30589 (2017).
- Borowiak, M. et al. Identification of a novel transposon-associated phosphoethanolamine transferase gene, *mcr-5*, conferring colistin resistance in d-tartrate fermenting *Salmonella enterica* subsp. *enterica* serovar Paratyphi B. *J. Antimicrob. Chemother.* **72**, 3317–3324 (2017).
- Athenstaedt, K. & Daum, G. Phosphatidic acid, a key intermediate in lipid metabolism. *Eur. J. Biochem.* **266**, 1–16 (1999).
- Epand, R. M., Walker, C., Epand, R. F. & Magarvey, N. A. Molecular mechanisms of membrane targeting antibiotics. *Biochim. Biophys. Acta - Biomembr.* **1858**, 980–987 (2016).
- Zurfluh, K., Kieffer, N., Poirel, L., Nordmann, P. & Stephan, R. Features of the *mcr-1* cassette related to colistin resistance. *Antimicrob. Agents Chemother.* **60**, 6438–6439 (2016).

42. Yu, C. Y. et al. Complete genome sequencing revealed novel genetic contexts of the *mcr-1* gene in *Escherichia coli* strains. *J. Antimicrob. Chemother.* **72**, 1253–1255 (2016).

43. Bradley, P., Bakker, H. den, Rocha, E., McVean, G. & Iqbal, Z. Real-time search of all bacterial and viral genomic data. *bioRxiv* 234955 <https://doi.org/10.1101/234955> (2017).

44. Antipov, D. et al. plasmidSPAdes: assembling plasmids from 555 whole-genome sequencing data. *Bioinformatics* **32**, 3380–3387 (2016).

45. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **gr.215087.116** <https://doi.org/10.1101/gr.215087.116> (2017).

46. Hunt, M. et al. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* **16**, 294 (2015).

47. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

48. Carattoli, A. et al. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).

49. Wang, X. et al. Molecular epidemiology of colistin-resistant *Enterobacteriaceae* in inpatient and avian isolates from China: high prevalence of *mcr*-negative *Klebsiella pneumoniae*. *Int. J. Antimicrob. Agents* **50**, 536–541 (2017).

50. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

51. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

52. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLOS Comput. Biol.* **11**, e1004041 (2015).

53. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).

54. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

55. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).

56. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUTi and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).

57. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).

58. Baele, G. et al. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167 (2012).

59. Zhao, F., Feng, Y., Lü, X., McNally, A. & Zong, Z. Remarkable diversity of *Escherichia coli* carrying *mcr-1* from hospital sewage with the identification of two new *mcr-1* variants. *Front. Microbiol.* **8**, 2094 (2017).

Acknowledgements

L.v.D., X.D., H.W., T.S., L.A.W., and FB acknowledge financial support from the Newton Trust UK-China NSFC initiative (grants MR/P007597/1 and 81661138006). F.B. additionally acknowledges support from the BBSRC GCRF scheme. L.P.S. was supported by a PhD scholarship from EPSRC (EP/F500351/1). P.B. is funded by Wellcome Trust on a “Genomic Medicine and Statistics DPhil” grant. A.R. was co-funded by the European

Union: European regional development fund (ERDF), by the Conseil Régional de La Réunion and by the Centre de Coopération internationale en Recherche agronomique pour le Développement (CIRAD). H.W. was supported by National Natural Science Foundation of China (81625014). L.A.W. is supported by a Dorothy Hodgkin Fellowship funded by the Royal Society (Grant Number DH140195) and a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 109385/Z/15/Z). Z.I. was funded by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 102541/A/13/Z). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

H.W. and F.B. conceived the project and designed the experiments. R.W., Q.W., X.W., L. J. Q.Z., Y.L., and H.W. collected samples. R.W., Q.W., X.W., L.J., Q.Z., and Y.L. performed microbial identification, antimicrobial susceptibility testing, screening for *mcr-1*, and DNA extraction for WGS. R.W. L.v.D., L.P.S. assembled the new sequence data. L.v.D. and L.P.S. curated the global data set and performed the computational analyses. T.D.-S. advised on functional aspects of colistin resistance. A.R., L.A.W., and X.D. helped with the phylogenetic reconstructions. P.B. and Z.I. performed the search for *mcr-1*-positive samples on the Short Read Archive. H.W. takes responsibility for the accuracy and availability of the epidemiological and raw sequence data, and F.B. for all bioinformatics and computational methods and results. L.v.D., L.P.S., and F.B. wrote the paper with contributions from X.D. and H.W. All authors read and commented on successive drafts and all approved the content of the final version.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-03205-z>.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018