Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the *mcr-1* gene indicates bacterial diversity but plasmid restriction

Jarne M. van Hattem,† Sébastien Matamoros,† Maris S. Arcilla,
Niels Willemse,
Damian C. Melles, John Penders, Trung Nguyen Vinh, Ngo Thi Hoa, COMBAT consortium, Menno D. de Jong, and Constance Schultsz

†Contributed equally

COMBAT consortium: Martin C. J. Bootsma, Perry JJ van Genderen, Abraham Goorhuis, Martin Grobusch, Nicky Molhoek, Astrid M. L. Oude Lashof, Ellen E. Stobberingh, and Henri A. Verbrugh

Scientific Reports. 2017 Nov 10;7(1):15364

ABSTRACT

To understand the dynamics behind the worldwide spread of the *mcr-1* gene, we determined the population structure of *Escherichia coli* and of mobile genetic elements (MGEs) carrying the *mcr-1* gene. After a systematic review of the literature we included 65 *E. coli* whole genome sequences (WGS), adding 6 recently sequenced travel related isolates, and 312 MLST profiles. We included 219 MGEs described in 7 Enterobacteriaceae species isolated from human, animal and environmental samples. Despite a high overall diversity, 2 lineages were observed in the *E. coli* population that may function as reservoirs of the *mcr-1* gene, the largest of which was linked to ST10, a sequence type known for its ubiquity in human faecal samples and in food samples. No genotypic clustering by geographical origin or isolation source was observed. Amongst a total of 13 plasmid incompatibility types, the IncI2, IncX4 and IncHI2 plasmids accounted for more than 90% of MGEs carrying the *mcr-1* gene. We observed significant geographical clustering with regional spread of IncHI2 plasmids in Europe and IncI2 in Asia. These findings point towards promiscuous spread of the *mcr-1* gene by efficient horizontal gene transfer dominated by a limited number of plasmid incompatibility types.

INTRODUCTION

Antimicrobial resistance (AMR) represents a growing threat to global health[1]. With barely any new antimicrobial drugs in development[2], limiting the spread of AMR is key in order to maintain current treatment options[3].

Colistin is an antibiotic of the polymyxin class, discovered in 1950 and effective against Gram-negative bacteria[4]. The emergence of multidrug-resistant Gram-negative bacteria, especially those producing carbapenemases, has reintroduced colistin as a last resort antibiotic for the treatment of severe infections[5]. In contrast to its limited use in humans, colistin is widely used in food-producing animals[6]. While colistin resistance was long thought to be caused by chromosomal mutations only[7], the emergence of plasmid-mediated resistance, conferred by the mobilized colistin resistance (*mcr-1*) gene, was recently reported[8]. This gene encodes for a protein of the phosphoethanolamine transferase enzyme family, and its expression results in the addition of a phosphoethanolamine to lipid A, the target of colistin, decreasing the interaction between colistin and the bacterial lipopolysaccharide[8]. Since its discovery in 2015 in China, this gene has been described in several bacterial species that were isolated from animals, animal food products, humans and environmental samples from around the world[9-13]. Our previous study in travellers indicated acquisition of *mcr-1* carrying bacteria by healthy individuals during travel to destinations around the world, potentially related to food exposure, as well as rapid clearance after return[14]. It has been suggested that *mcr-1* has spread from food animals to humans[8,15-17], but there is a lack of comparison of *mcr-1* carrying isolates on a global level to support this hypothesis.

We studied the global population structure as well as the geographic and host distribution of *mcr-1*-carrying *Escherichia coli*, and mobile genetic elements (MGEs), to establish the population structure and to assess whether the spread of the *mcr-1* gene is linked to clonal dissemination or transmission of MGEs from animal, human, or environmental sources within geographic regions.

RESULTS

Literature search

A systematic review of the literature on *mcr-1*, published until 1 January 2017 resulted in the inclusion of 95 articles, representing a total of 410 entries (whole genome sequences, MLST profiles, and/or plasmid types) for analysis (See detailed methods and results in Supplementary data, Supplementary Figure 1 and Supplementary Table 1).

Population structure

Whole genome sequencing (WGS)

The genomes of 65 *mcr-1*-carrying *E. coli* were analysed, including 6 genomes from *E. coli* isolated from travellers that were sequenced for the purpose of the present study. Isolates originated from Asia (n = 36; 55.4%), Europe (n = 20; 30.8%), North-America (n = 4; 6.2%), South-America (n = 4; 6.2%) and Africa (n = 1; 1.5%). 45 were of animal origin (69.2%), 19 of human origin (29.2%) and one strain (1.5%) was isolated from water (Supplementary Table 1).

The average size of the genomes (all contigs in each assembly, representing chromosomes and plasmids) of these 65 isolates was 4.9 Mbp, with a median number of genes identified of 4785 (ranging from 4266 to 7083), representing a pangenome of 23248 genes and a core genome (defined by genes present in at least 99% of the isolates) of 2216 genes. An unbiased analysis of the population structure was performed using a Bayesian approach with the BAPS software[18], based on the nucleotide alignment of the core genome sequences. It revealed the presence of 5 distinct phylogenetic clusters (Fig. 1; Supplementary Figure 2; Supplementary Table 1). The largest cluster (cluster 1) consisted of 26 isolates from 16 different STs (26/65; 40.0%) and the second cluster consisted of 24 isolates from 15 different STs (36.9%). No significant relationship between clustering (BAPS) and geographical origin or isolation source was observed (χ^2-test) (Fig. 1) except that all 5 isolates that belong to BAPS cluster 3 are from Europe. Twenty isolates showed less than 10 SNPs/Mbp difference with at least 1 other isolate and were considered clonally related (Supplementary Tables 2 and 3).

Multilocus sequence typing (MLST)

For 312 *E. coli* isolates originating from 69 studies, a MLST profile was published or could be deduced from the corresponding WGS. Of these, 206 were isolated from animals or animal products (66.0%), 101 were isolated from humans (32.4%), including the 6 travel acquired isolates, and 5 from the environment (1.6%). 141 Isolates from 25 studies (141/312; 45.2%) originated from Asia and 125 isolates from 25 studies (40.1%) from Europe, together accounting for 85.3% of all included isolates. The isolates represented 112 unique sequence types (STs) with ST10 being most common, comprising 40/312 (12.8%) isolates originating from Africa, Asia, Europe and South-America.

eBURST analysis[19] was performed on all isolates included in the study, to identify their genetic relatedness based on their MLST profiles. Three main clusters were identified, for which the predicted founders, i.e. the ST in a cluster from which all other SLVs and DLVs in the cluster have most likely diversified[19], were ST10, ST1114 and ST410. The largest cluster contained all 40 ST10 isolates and an additional 46 isolates in 21 STs that were single (SLV) or double locus variants (DLV) of ST10 (86/312; 27.6%) (Supplementary Figure 3). The predicted founder of the second largest cluster was ST1114, a SLV of ST165

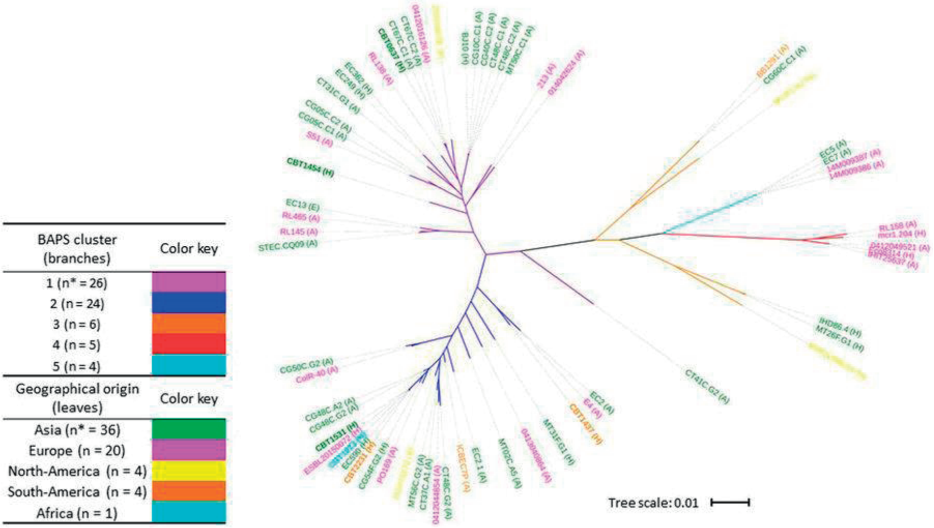


Figure 1. Maximum-likelihood tree based on concatenated core genome sequences of 65 *mcr-1*-carrying *E. coli* isolates. Branch colours indicate phylogenetic clusters as determined by BAPS. Isolates from ST10, ST165 and closely related isolates are all grouped in the BAPS cluster 2 (dark blue). Leaf (isolates identifiers) colours indicate geographical region of origin. Isolation source is indicated in brackets: A = animal or meat; H = human; E = environment. The 6 travellers' isolates that were sequenced for this study are highlighted in bold and names start with CBT. Tree scale in number of substitutions per site. *Number of isolates.

and ST100, and included 19 isolates belonging to 7 different STs (5.4%), while the third cluster was centred on ST410 and included 14 isolates from 3 different STs (4.5%).

A maximum-likelihood tree based on concatenated MLST gene sequences showed a main clade of 128 isolates (represented by blue branches in Fig. 2 and Supplementary Figure 4; bootstrap value of the main branch = 0.98), including most, but not all, isolates from the eBURST clusters of ST10 and ST1114 (Supplementary Figure 4A). All isolates from these 2 eBURST clusters for which a WGS was available were grouped in BAPS cluster 2. Similarly, all the isolates from the eBURST cluster ST410 grouped into BAPS cluster 1, along with 6 isolates from ST155. Seven isolates belonged to the globally successful extra-intestinal pathogenic *E. coli* clone ST131 (Supplementary Figure 4A).

As observed in the WGS analysis, animal isolates were interspersed with isolates from humans and the environment throughout the tree, as were isolates from different continents indicating a lack of clustering by isolation source or geographical origin (Fig. 2). Similarly, no clustering by health status of the host was observed (Supplementary Figure 4B).

Mobile genetic elements

The plasmid incompatibility group of the *mcr-1*-carrying plasmids could be determined for 217 Enterobacteriaceae isolates from 7 different species (*Escherichia sp., Salmonella* sp., *Klebsiella* sp., *Cronobacter* sp., *Enterobacter* sp., *Kluyvera* sp. and *Shigella* sp.), representing a total of 219 plasmids since 2 isolates carried 2 different plasmids (Table 1). These plasmids were described in 71 studies (1 to 33 plasmids per study, average = 3.1). In addition, the gene was integrated in the chromosome of 6 isolates. The incompatibility group could not be determined for 27 of the 65 isolates for which WGS was available. Similarly the plasmid type was not available for 182 of the 312 isolates included in the MLST analysis. A total of 14 different plasmid incompatibility groups were identified. 198/219 (90.4%) of the identified plasmids belonged to one of 3 incompatibility groups: IncX4 (77/219 plasmids, 35.2%), IncI2 (76/219 plasmids, 34.7%) and IncHI2 (45/219 plasmids, 20.5%). 50/76 IncI2 plasmids (65.8%) originated from Asia and 33/45 IncHI2 plasmids (73.3%) from Europe. IncX4 plasmids were more evenly distributed: 44/77 (57.1%) were recovered from Europe, 29 from Asia (37.7%) and 4 from other regions (5.2%). Observed proportions were significantly different from expected for IncI2 (χ^2-test, $p < 0.001$) and IncHI2 ($p < 0.001$) but not for IncX4. The distribution of these 3 plasmid types was not significantly different from expected between animal (χ^2-test, $p = 0.24$), human ($p = 0.88$) and environmental sources ($p = 0.38$). Isolates from the BAPS groups 1 and 2 carried plasmids from the 3 major types in similar proportions (Supplementary Figure 5A; Supplementary Table 1). Isolates from the eBURST clusters of ST10 carried plasmids belonging to 7 different incompatibility groups, including IncHI2, IncI2 and IncX4. No clustering of plasmid type with MLST phylogeny was observed either (Supplementary Figure 5B; Supplementary Table 1).

Figure 3 shows the alignment of the complete sequences or contigs from IncI2 (panel A), IncX4 (B) and IncHI2 (C) incompatibility group plasmids. IncHI2 plasmids had the largest size, with sequence lengths up to 267486 bp.

The ISA*pl1* transposon element situated upstream of the *mcr-1* gene was present in 7/9 (77.8%) IncHI2 plasmids, but only in 11/29 (37.9%) of IncI2 plasmids and completely missing in all of the 24 reported IncX4 plasmids (Fig. 3). In the isolates from travellers the ISA*pl1* transposon was identified in 3 out of our 6 *mcr-1*-carrying contigs including one isolate from a traveler to Asia (ST101, IncI2 incompatibility group), one to Africa (ST744, IncHI2) and one to South-America (ST744, incompatibility group not identified).

Antimicrobial resistance genes

All sequences of the *mcr-1* gene collected in the present study were 100% identical to the original sequence described by Liu *et al.*[8].

Multiple resistance genes were detected in most of the studied isolates (Supplementary Results). The florfenicol resistance gene *floR* was present in 32 (49.2%) isolates; in

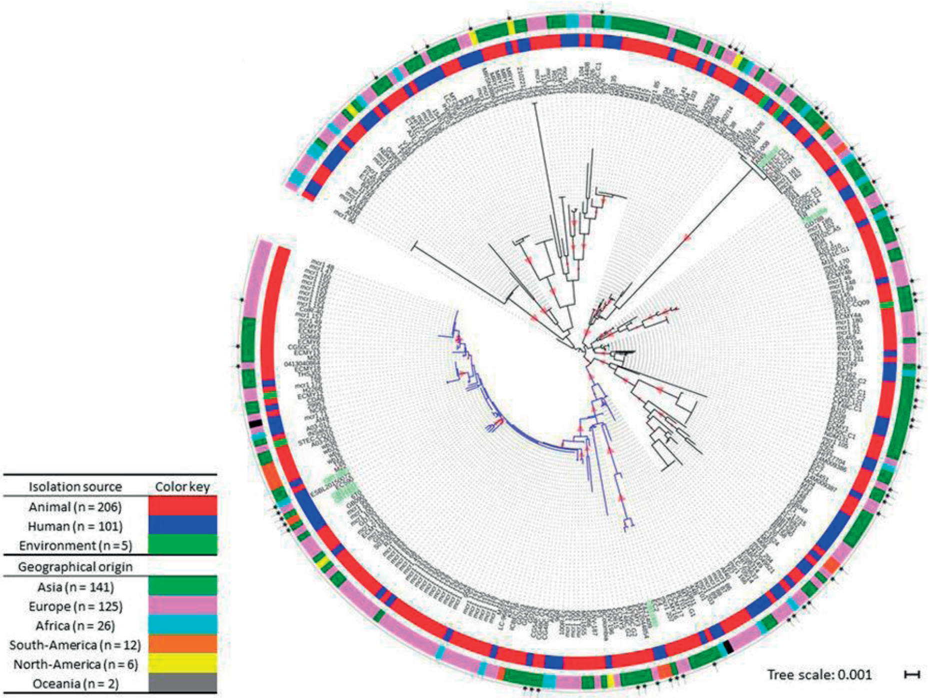


Figure 2. Phylogeny of the *mcr-1*-carrying *E. coli* isolates. Maximum-likelihood tree based on concatenated MLST gene sequences, mid-point rooted. Inner coloured circle: isolation source; outer circle: region of origin. Stars indicate the isolates from which a whole genome sequence was available. The 6 travellers' isolates that were sequenced for this study are highlighted in green. Bootstrap values between 0.9 and 1 are indicated by red triangles (size proportional to bootstrap value). The blue branches represent the main clade of 128 isolates including most isolates from ST10. Tree scale in number of substitutions per site. See Supplementary Figure 4A for additional information on the relationship between STs, eBURST clustering and WGS BAPS clustering.

22 of 45 isolates from animals (48.9%) and 10 of 19 isolates from humans (52.6%). The *baeR* and *baeS* genes, encoding novobiocin resistance, were found in 64 (98.5%) and 65 (100%) isolates, respectively.

Plasmid analysis from the WGS data showed that 4 of the 29 *mcr-1*-carrying IncI2 plasmids (13.8%) contained an additional ESBL gene. IncHI2 plasmids (n = 9) carried between 0 and 12 additional AMR genes. In particular, 4 plasmids carried CTX-M ESBL genes and 2 carried the *floR* gene. In 4 out of the 9 IncHI2 plasmids analysed in this study, the *mcr-1* gene was shown to be integrated alongside a large multi-drug resistance (MDR) gene cassette (Fig. 3C). None of the IncX4 plasmids carried additional AMR genes.

Table 1. Incompatibility types of mcr-1 carrying plasmids and distribution by geographical regions.

	IncF	IncFI	IncFIB	IncFII	IncHI1	IncHI1A/ IncHI1B/ IncFIA	IncHI2	IncI2	IncI2-IncX4	IncP	IncX3-X4	IncX4	IncY	repB (p0111)	**Total plasmids**		Chromosome
Africa	0	0	1	0	0	0	4	8	0	0	0	1	0	0	**14**	**6.4%**	0
Asia	1	3	0	0	0	2	7	50	0	1	1	29	2	0	**96**	**43.8%**	2
Europe	0	0	2	1	1	0	33	5	0	3	0	44	0	1	**90**	**41.1%**	4
North-America	1	0	0	0	0	0	1	4	0	0	0	1	0	0	**7**	**3.2%**	0
Oceania	0	0	0	0	0	0	0	2	0	0	0	0	0	0	**2**	**0.9%**	0
South-America	0	0	0	0	0	0	0	7	1	0	0	2	0	0	**10**	**4.6%**	0
Total (% of total)	**2 (0.9)**	**3 (1.4)**	**3 (1.4)**	**1 (0.5)**	**1 (0.5)**	**2 (0.9)**	**45 (20.5)**	**76 (34.7)**	**1 (0.5)**	**4 (1.8)**	**1 (0.5)**	**77 (35.2)**	**2 (0.9)**	**1 (0.5)**	**219 (100)**	**100.0%**	**6**

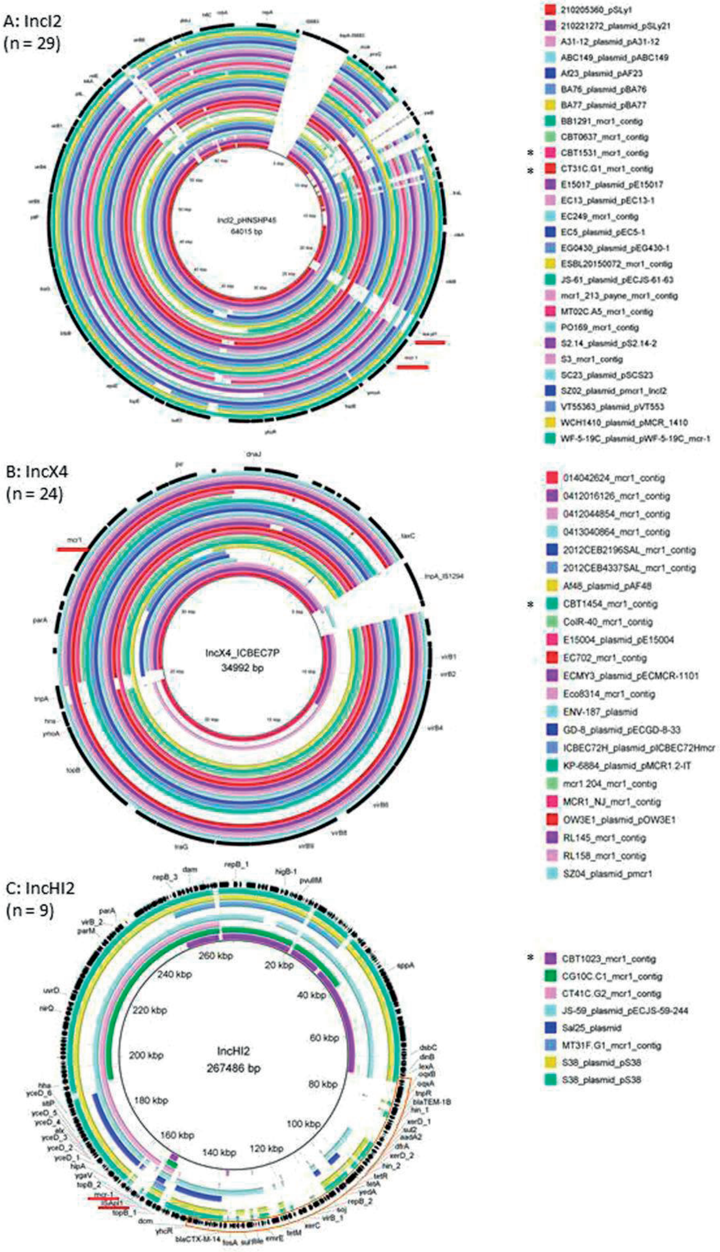


Figure 3. Alignment of *mcr-1*-containing plasmids and contigs. Panel A: IncI2 plasmids (n = 29); panel B: IncX4 (n = 24); panel C: IncHI2 (n = 9). Black outer ring: plasmid used as reference for the alignment; name and size of the reference indicated in the middle of each panel. Plasmid names followed by "_mcr1_contig" refer to assembled contigs from whole genome sequences. Other names refer to plasmid sequences deposited in online databases. The *mcr-1* gene and ISA-pl1 location are underlined in red. Plasmids indicated with an asterisk are from the 6 travellers' isolates that were sequenced for this study. Panel C: Putative MDR cassette is highlighted in orange.

DISCUSSION

Analysis of all reported WGS of *mcr-1*-carrying isolates shows that the population of *E. coli* is highly diverse, but is dominated by 2 large groups of related isolates. Most of the isolates from BAPS group 2 grouped into a MLST cluster centred on ST10. An over-representation of isolates related to ST10 and ST165 (a SLV of ST1114) in *mcr-1*-carrying *E. coli* isolates was previously reported at a smaller scale in isolates from European farm animals[20]. *E. coli* ST10 and closely related STs are frequently recovered from food and human intestinal samples and studies have shown a higher prevalence of plasmid-carried AMR genes in ST10, including CTX-M ESBL genes, compared to other STs[21-24].

The other BAPS group of interest in our study (group 1) included isolates belonging to ST155. This ST has been described as a major vector of spread of ESBL genes from animals to humans[25]. It is thus possible that zoonotic transmission leads to the spread of the *mcr-1* gene, as has been suggested in studies from China and Vietnam[16,17], notably through the 2 main phylogenetic clusters identified in this study.

Additionally, we found clonally related isolates, including some belonging to ST744, a SLV of ST10, carrying the *mcr-1* gene on different plasmid backbones and recovered from different continents (see Supplementary Results). These results point towards a worldwide dissemination *of mcr-1* driven mainly by highly promiscuous plasmids rather than the worldwide spread of one or more *mcr-1*-carrying clones. We hypothesize that several populations of *E. coli* isolates, notably those related to ST10 or ST155, acquired the *mcr-1* gene due to their intrinsic ability of acquiring AMR genes and their high prevalence in humans and food animals. These populations of commensal isolates then may play a crucial role as a reservoir for this gene, which can explain their over-representation in the present study.

In the timeframe of our literature search, 3 *E. coli* strains carrying the *mcr-2* gene were isolated from animals in Belgium. These isolates belonged to ST10 (2 isolates) and ST167 which is a SLV of ST10 and carried the gene on an IncX4-type plasmid. No WGS data was available from these isolates[26]. More data about the *mcr-2* gene is needed to assess its spread and determine if *E. coli* ST10 plays a similar role in its dissemination as it does for *mcr-1*.

More than 90% of published plasmid types carrying *mcr-1* genes belonged to either IncI2, IncX4 or IncHI2. Almost 75% of the isolates carrying an IncHI2 plasmid originated from Europe: 26 from animals and 7 from humans (Table 1 and Supplementary Table 1). In a traveller's isolate acquired in Tunisia, the *mcr-1*-carrying plasmid was identified as an IncHI2-type backbone of the ST4 pMLST subtype which co-carried a CTX-M-1 ESBL gene (Fig. 3C). This traveller reported consumption of beef, chicken and eggs during travel to Tunisia which can potentially be the source for the acquisition of the *mcr-1* positive isolate. When investigating the presence of the *mcr-1* gene in cephalosporin resistant

E. coli isolates from chicken farms in Tunisia, Grami *et al.*[27] found that all *mcr-1*-carrying plasmids from their study (n = 37) also belonged to the IncHI2-type, ST4 subtype and harboured CTX-M-1 genes. PFGE typing of the isolates harbouring this plasmid showed various bacterial genetic backgrounds. Interestingly, these chickens were all imported from France, either as adults or chicks. Other studies showed the presence of this IncHI2, CTX-M-1 and *mcr-1* combination in *Salmonella enterica* Typhimurium isolates from meat samples in Portugal from 2011[28,29] and diarrhoeic veal calves in France[30]. The IncHI2 subtype ST4 was also detected in an *E. coli* isolate from retail chicken breast in Germany[31] and the faecal sample of a veal calve from the Netherlands[32], suggesting widespread dissemination of this particular plasmid in European farm animals and possible transmission to humans.

The high prevalence of novobiocin *baeR* and *baeS* and florfenicol *floR* resistance genes[33,34] in the genomes of isolates of human and animal origin together with the fact that florfenicol and novobiocin are used almost exclusively in veterinary medicine further supports the potential role of food animals as an important reservoir of *mcr-1* containing bacteria and MGEs[15].

In contrast with the IncHI2 plasmids, 65.8% of all IncI2 plasmids recovered so far originated from Asia, with a much lower prevalence in *mcr-1* carrying Enterobacteriaceae from other regions. Taken together, these elements point toward a more regional circulation and dissemination of the *mcr-1*-carrying plasmids IncI2 and IncHI2.

We found the ISA*pl1* transposon element associated with the *mcr-1* gene, as originally described by Liu *et al.*[8] to be present in a minority of studied plasmids and contigs. However, since some of the *mcr-1*-carrying contigs were obtained by assembly of Illumina short reads from WGS data, we cannot exclude that some of these gaps are explained by an incomplete assembly of (plasmid) sequences. The ISA*pl1* transposon element is considered to be the main driver of horizontal gene transfer of the *mcr-1* gene and has been shown to be highly unstable in IncI2 plasmids[35-37]. The absence of the ISA*pl1* transposon element in *mcr-1*-carrying IncX4 plasmids as described here has recently been proposed to be essential for the maintenance of the *mcr-1* gene in this particular backbone, but the exact mechanism still requires further investigation[38].

WGS analysis provided in-depth information about the *mcr-1*-carrying *E. coli* isolates and their phylogenetic relationship, but the number of available genomes was limited. On the other hand, whilst MLST data have a lower resolution, the higher number of available profiles allowed analysis of the isolates' origin (geographical, source of isolation, diseased status of the host, etc.).

A limitation of our study is the potential for bias. The overrepresentation of isolates originating from Asia and Europe could be explained by a higher prevalence of *mcr-1* genes on these continents, but the effect of publication bias cannot be excluded. Isolates from North-America only represented 2.2% of the collection. Noteworthy, colistin, except for ophthalmic ointment, has never been marketed for use in animals in the United States[39,40].

Sampling bias should also be considered when several isolates with an identical ST are presented from a single study, as is the case for ST100 and ST752. Additionally, in the absence of a control population of *mcr-1*-negative isolates obtained from similar sources as the *mcr-1* positive isolates, results of analysis of population structures should be interpreted with caution. Because many studies screened existing collections of (resistant) isolates for colistin resistance or presence of *mcr-1*, selection bias has probably been introduced.

The findings in this study suggests that the *mcr-1* gene has locally and globally disseminated through MGEs that are mainly IncHI2, IncI2 and IncX4 plasmids and provides additional support for the hypothesis of the animal reservoir, that is driven by the use of colistin in livestock, as a source of *mcr-1* in humans. A global ban of colistin use in animals to preserve colistin for use in human medicine seems therefore justified.

MATERIAL AND METHODS

Literature search

Relevant papers that published on *mcr-1* and *mcr-2* were identified in PubMed, Web of Science, Scopus, ScienceDirect and Google Scholar using the query 'mcr-1 OR mcr1 OR mcr-2 OR mcr2 OR (mcr AND colistin)' (see Supplementary Material for full search strategies). To be able to study the associations between phylogeny, geographic distribution and isolation source we only included sequences from papers that provided sufficient metadata. As a consequence, plasmids and genomes sequences that were deposited in online databases without metadata were not included in the analysis.

Selection of isolates for whole genome sequencing

We subjected 6 *mcr-1* positive isolates that were collected as part of a prospective study (COMBAT study) aimed at studying acquisition of extended-spectrum β-lactamase (ESBL) -producing Enterobacteriaceae during travel to whole-genome sequencing[14,41]. Additionally, we included 22 whole genome sequences of isolates from Vietnamese chickens and humans that were still unpublished when performing our literature search[17].

Whole genome sequencing of *mcr-1*-positive *E. coli* isolates

Bacterial DNA was extracted from fresh pure cultures using the Qiagen DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany). Library preparation was done according to manufacturer's instruction (Illumina, San Diego, CA, USA) and sequenced using Illumina MiSeq technology with 150nt paired-end settings. Sequences have been deposited in the European Nucleotide Archive under the accession numbers SAME104030441 to SAME104030446.

Bio-informatic analysis

MLST analysis

For each *mcr-1*-carrying *E. coli* isolate for which the ST or the whole genome sequence was available, the sequences of the corresponding alleles were downloaded from the *E. coli* MLST genes repository of the University of Warwick (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/handlers/getFileData/home/cbailster/mlst/zope/Extensions/gadfly/Ecoli/DB/) and concatenated. When STs of isolates were not described in literature, the ST was determined from available whole genomes using the online service provided by the Center for Genomic Epidemiology (https://cge.cbs.dtu.dk/services/MLST/) according to the Achtman MLST scheme[42,43]. MLST clusters (STs and their single locus or double locus variants) were defined using e-burst V3 (http://eburst.mlst.net/v3/enter_data/single/)[19] and goeBURST v1.2.1[44] using only profiles from this study.

WGS and plasmid analysis

Raw sequence reads in fastq format or pre-assembled sequences in fasta format were downloaded from online databases for all available isolates (Supplementary Table 1). Additional sequences not yet deposited in online databases were requested from their respective authors. The quality of the raw sequence reads was checked using fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), quast[45] and KmerFinder 2.0 (https://cge.cbs.dtu.dk/services/KmerFinder/) (see Supplementary Methods for more details). Reads were trimmed using Trimmomatic V0.33[46]. *De-novo* genome assembly was performed with SPAdes 3.9[47] for Illumina short reads and with Canu v1.3 for PacBio long reads[48]. Contigs of less than 500 bp long were removed from the genomes to improve the overall quality of the assembly. Size of the genomes was calculated by adding the length of all remaining contigs. Identification of open reading frames (ORFs) and gene contents in the assembled genomes (*de-novo* assemblies and pre-assembled sequences) was performed using Prokka v1.11[49]. Core genome analysis was performed with Roary v3.6.8[50]. Clustering of isolates was performed using the hierBAPS module of the Bayesian Analysis of Population Structure (BAPS) software v6.0[18]. The core genome alignment output provided by Roary was used as input for BAPS with 2 levels of hierarchy and a maximum number of cluster (K) of 10. The estimated number of clusters was 5 for both levels of hierarchy.

Sequences (concatenated MLST loci or concatenated core genes) were aligned using mafft v6.864b[51]. The resulting alignment was used as input for calculation of distances and tree building using RAxML v8.1.6[52]. MLST and WGS trees were visualized using iTOL v3.3.1 (http://itol.embl.de/)[53].

Identification of plasmid incompatibility group and typing of IncHI2 plasmids were performed on assembled sequences (*de-novo* or pre-assembled) via the CGE online services PlasmidFinder v1.3 (https://cge.cbs.dtu.dk/services/PlasmidFinder/) and pMLST v1.4 (https://cge.cbs.dtu.dk/services/pMLST/)[54]. Alignment and visualization of plasmids was performed with BRIG v0.95[55]. The majority of the isolates and plasmids described in this study were sequenced using a short read technology (Illumina). This technology does not allow for a high quality assembly of the plasmids due to the high number of repeat regions present in these MGEs. Therefore no phylogenetic analysis of the *mcr-1*-carrying plasmids was conducted in this study.

Two different databases were used for identification of other antibiotic resistance genes: ResFinder (https://cge.cbs.dtu.dk/services/ResFinder/[56]) was used to detect acquired resistance genes commonly located on mobile genetic elements (MGEs) and CARD Resistance Gene Identifier (https://card.mcmaster.ca/analyze/rgi[57]) was used to detect chromosomal genes.

Statistics

The distribution of isolates and plasmids from different geographical origins and isolation sources was determined by a χ^2-test comparing the expected distribution (proportions of the total studied population) to the observed proportions using GraphPad Prism6 (La Jolla, CA, USA).

ACKNOWLEDGEMENTS

The authors would like to thank the team of the Department of Genome analysis of the AMC (in particular Dr. F. Baas and L. Koster) for their support in the sequencing bacterial genomes. Financial support for this study was provided by the European COMPARE project (http://www.compare-europe.eu/) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 643476. The VIBRE and COMBAT studies were supported by The Netherlands Organisation for Health Research and Development/The Netherlands Organisation for Scientific Research (ZonMw) under grants number 205100012 and 50–51700–98–120 respectively.

AUTHOR CONTRIBUTIONS

S.M. and J.v.H. contributed equally to this work. S.M and J.v.H. performed the systematic review, performed the experiments, interpreted the results and wrote the manuscript. N.W. contributed to bio-informatics analyses. M.A., D.M., J.P., M.d.J., the COMBAT consortium members, N.V., N.T.H. and C.S. contributed isolates and genome sequence data and associated meta data. M.d.J. and C.S. designed the study. All authors contributed to the writing of the manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

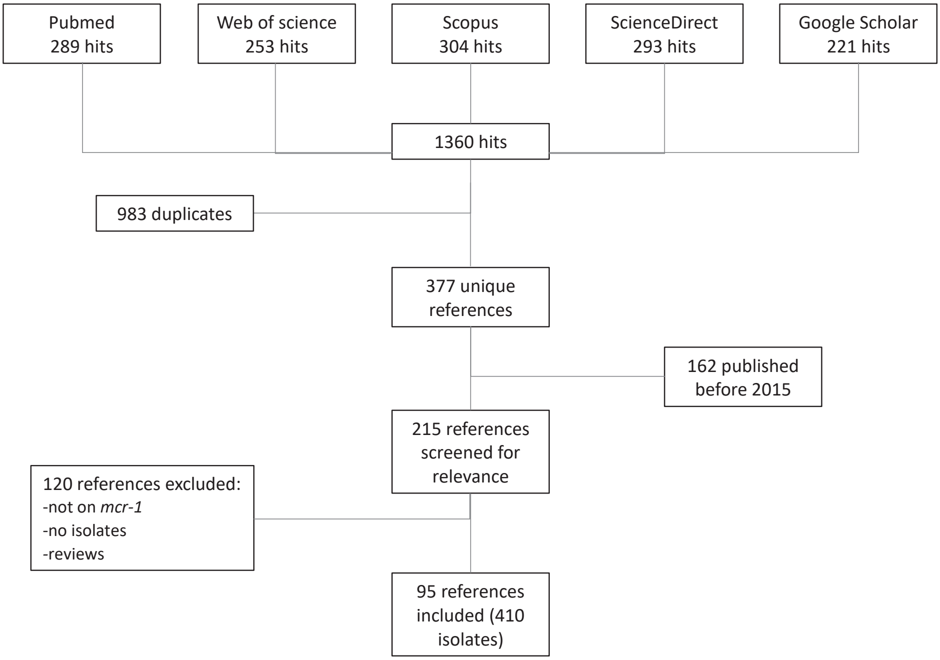
REFERENCES

1. Appelbaum PC. 2012 and beyond: Potential for the start of a second pre-antibiotic era? J. Antimicrob. Chemother. 2012;67:2062–2068.
2. Brown ED, Wright GD. Antibacterial drug discovery in the resistance era. Nature. 2016;529:336–343.
3. Laxminarayan R, et al. Antibiotic resistance-the need for global solutions. Lancet Infect. Dis. 2013;13:1057–1098.
4. Koyama Y, Kurosasa A, Tsuchiya A, Takakuta K. A new antibiotic 'colistin'produced by spore-forming soil bacteria. J Antibiot. 1950;3:457–458.
5. Lim LM, et al. Resurgence of colistin: a review of resistance, toxicity, pharmacodynamics, and dosing. Pharmacotherapy. 2010;30:1279–91.
6. Catry B, et al. Use of colistin-containing products within the European Union and European Economic Area (EU/EEA): development of resistance in animals and possible impact on human and animal health. Int. J. Antimicrob. Agents. 2015;46:297–306.
7. Olaitan AO, Morand S, Rolain J. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. Front. Microbiol. 2014;5:1–18.
8. 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*. 3099, (2015).
9. Skov, R. L. & Monnet, D. L. Plasmid-mediated colistin resistance (mcr-1gene): three months later, the story unfolds. *Eurosurveillance* 21, (2016).
10. Schwarz S, Johnson AP. Transferable resistance to colistin: a new but old threat. J. Antimicrob. Chemother. 2016;71:2066–2070.
11. Leangapichart T, et al. Acquisition of mcr-1 Plasmid-Mediated Colistin Resistance in Escherichia coli and Klebsiella pneumoniae during Hajj 2013 and 2014. Antimicrob. Agents Chemother. 2016;60:6998–6999
12. Vading M, et al. Frequent acquisition of low-virulence strains of ESBL-producing Escherichia coli in travellers. J. Antimicrob. Chemother. 2016;71:3548–3555.
13. Bernasconi OJ, et al. Travelers Can Import Colistin-Resistant Enterobacteriaceae, Including Those Possessing the Plasmid-Mediated mcr-1 Gene. Antimicrob. Agents Chemother. 2016;60:5080–5084.
14. Arcilla MS, et al. Dissemination of the *mcr-1* colistin resistance gene. Lancet Infect. Dis. 2016;16:147–149.
15. Poirel L, Nordmann P. Emerging plasmid-encoded colistin resistance: the animal world as the culprit? J. Antimicrob. Chemother. 2016;71:2326–2327.
16. Wang Y, et al. Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1 in Chinese poultry production. Nat. Microbiol. 2017;2:16260.
17. Trung NV, et al. Zoonotic Transmission of mcr-1 Colistin Resistance Gene from Small-Scale Poultry Farms, Vietnam. Emerg. Infect. Dis. J. 2017;23:529.
18. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol. Biol. Evol. 2013;30:1224–1228.
19. Feil Edward J, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 2004;186:1518–1530

20. El Garch F, et al. mcr-1 is borne by highly diverse Escherichia coli isolates since 2004 in food-producing animals in Europe. Clin. Microbiol. Infect. 2017;23:51e1–51e4
21. Manges AR, Johnson JR. Food-Borne Origins of Escherichia coli Causing Extraintestinal Infections. Clin. Infect. Dis. 2012;55:712–719.
22. Oteo J, et al. Extended-spectrum β-lactamase-producing Escherichia coli in Spain belong to a large variety of multilocus sequence typing types, including ST10 complex/A, ST23 complex/A and ST131/B2. Int. J. Antimicrob. Agents. 2009;34:173–176.
23. Fam N, et al. CTX-M-15-producing Escherichia coli clinical isolates in Cairo (Egypt), including isolates of clonal complex ST10 and clones ST131, ST73, and ST405 in both community and hospital settings. Microb. Drug Resist. 2011;17:67–73.
24. Manges AR, et al. Clonal distribution and associated characteristics of Escherichia coli clinical and surveillance isolates from a military medical center. Diagn. Microbiol. Infect. Dis. 2017;87:382–385.
25. Skurnik D, et al. Emergence of antimicrobial-resistant Escherichia coli of animal origin spreading in humans. Mol. Biol. Evol. 2016;33:898–914.
26. Xavier, B. B. *et al*. Identification of a novel plasmid-mediated colistinresistance gene, mcr-2, in Escherichia coli, Belgium, june 2016. *Eurosurveillance* 21, (2016).
27. Grami R, et al. Impact of food animal trade on the spread of mcr-1-mediated colistin resistance, tunisia, july 2015. Eurosurveillance. 2016;21:1–5.
28. Tse H, Yuen K-Y. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect. Dis. 2015;3099:3099.
29. Figueiredo R, Henriques A, Sereno R, Mendonça N, da Silva GJ. Antimicrobial Resistance and Extended-Spectrum β-Lactamases of Salmonella enterica Serotypes Isolated from Livestock and Processed Food in Portugal: An Update. Foodborne Pathog. Dis. 2014;12:110–117.
30. Haenni M, et al. Co-occurrence of extended spectrum β lactamase and MCR-1 encoding genes on plasmids. Lancet Infect. Dis. 2016;3099:3099.
31. Falgenhauer L, et al. Chromosomal Locations of mcr-1 and bla(CTX-M-15) in Fluoroquinolone-Resistant Escherichia coli ST410. Emerg. Infect. Dis. 2016;22:1689–1691.
32. Veldman K, et al. Location of colistin resistance gene mcr-1 in Enterobacteriaceae from livestock and meat. J. Antimicrob. Chemother. 2016;71:2340–2342.
33. Lilic M, Jovanovic M, Jovanovic G, Savic DJ. Identification of the CysB-regulated gene, hslJ, related to the Escherichia coli novobiocin resistance phenotype. FEMS Microbiol. Lett. 2003;224:239–246.
34. Baranova N, Nikaido H. The BaeSR Two-Component Regulatory System Activates Transcription of the yegMNOB (mdtABCD) Transporter Gene Cluster in Escherichia coli and Increases Its Resistance to Novobiocin and Deoxycholate. J. Bacteriol. 2002;184:4168–4176.
35. Stoesser N, Mathers AJ, Moore CE, Day NP, Crook DW. Colistin resistance gene mcr-1 and pHN-SHP45 plasmid in human isolates of Escherichia coli and Klebsiella pneumoniae. Lancet Infect. Dis. 2016;3099:3099.
36. 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. 2016;3099:3099.
37. Snesrud, E. *et al*. Analysis of Serial Isolates of mcr-1- positive Escherichia coli Reveals a Highly Active ISApl1 Transposon. *Antimicrob. Agents Chemother*. 10.1128/AAC.00056-17 (2017).
38. Sun J, et al. Genetic Analysis of the IncX4 Plasmids: Implications for a Unique Pattern in the mcr-1 Acquisition. Sci. Rep. 2017;7:424.

39. FDA. *2015 Summary Report On Antimicrobials Sold or Distributed for Use in Food-Producing Animals*. at https://www.fda.gov/downloads/ForIndustry/UserFees/AnimalDrugUserFeeActADUFA/UCM534243.pdf(2016).
40. European Medicines Agency. *Updated advice on the use of colistin products in animals within the European Union: development of resistance and possible impact on human and animal health*. 44 (2016).
41. Arcilla MS, et al. Import and spread of extended-spectrum beta-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. Lancet Infect. Dis. 2017;17:78–85.
42. Larsen MV, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. 2012;50:1355–1361.
43. Wirth T, et al. Sex and virulence in Escherichia coli: An evolutionary perspective. Mol. Microbiol. 2006;60:1136–1151.
44. Francisco AP, Bugalho M, Ramirez M, Carriço Ja. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics. 2009;10:152.
45. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–1075.
46. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–2120.
47. Bankevich A, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 2012;19:455–477.
48. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. bioRxiv at http://biorxiv.org/content/early/2017/01/04/071282.abstract (2017).
49. Seemann, T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014).
50. Page AJ, et al. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3693.
51. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 2008;9:286–298.
52. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–1313.
53. Letunic I, Bork P. Interactive tree of life (iTOL)v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–W245.
54. Carattoli A, et al. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrob. Agents Chemother. 2014;58:3895–3903.
55. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Geno7mics. 2011;12:402.
56. Zankari E, et al. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 2012;67:2640–2644.
57. McArthur AG, Wright GD. Bioinformatics of antimicrobial resistance in the age of molecular epidemiology. Curr. Opin. Microbiol. 2015;27:45–50.

SUPPLEMENTARY METHODS AND RESULTS



Supplementary Figure 1. Flow chart of the systematic literature review according to the PRISMA guidelines.

Systematic literature review

Relevant papers that had published on *mcr-1* and *mcr-2* were identified in Pubmed, Web of Science, Scopus, ScienceDirect and Google Scholar using the query 'mcr-1 OR mcr1 OR mcr-2 OR mcr2 OR (mcr AND colistin)'. The database was accessed on 4 January 2017.

Criteria for considering studies for this study

Eligible for inclusion were all studies publishing:

- a multilocus sequence typing (MLST) profile (sequence type; ST) of at least one *mcr-1* positive *E. coli* isolate, or
- a whole genome sequence (WGS) of any species. For *E. coli* species the WGS was used for both phylogeny and plasmid analysis. For other species the WGS was used only for plasmid analysis, or
- the incompatibility group of a plasmid carrying a *mcr-1* or *mcr-2* gene.

All studies which did not satisfy these criteria were excluded. Studies presenting data on the same isolate were included only once.

Selection of studies

Two authors (SM and JvH) independently assessed the titles and abstracts of studies identified in terms of their relevance. Full versions of articles were obtained if the initial assessment of these met the inclusion criteria.

A flow chart of the search strategy is presented in Supplementary Figure 1. A complete list of included isolates and plasmids and their references can be found in Supplementary Table 1.

The number of entries (WGS, MLST profile, plasmid sequence or plasmid type) per included study ranged from 1 to 42 (average = 4.3; median = 2).

Search strategies

Pubmed, 289 hits

mcr1[All Fields] OR
mcr-1[All Fields] OR
mcr2[All Fields] OR
mcr-2[All Fields]) OR
("mcr"[All Fields] AND ("colistin"[MeSH Terms] OR "colistin"[All Fields])

Web of Science [v.5.23], 253 hits

TOPIC: (mcr1 OR mcr-1 OR mcr2 OR mcr-2 OR (colistin AND mcr))
Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI.

Scopus, 304 hits

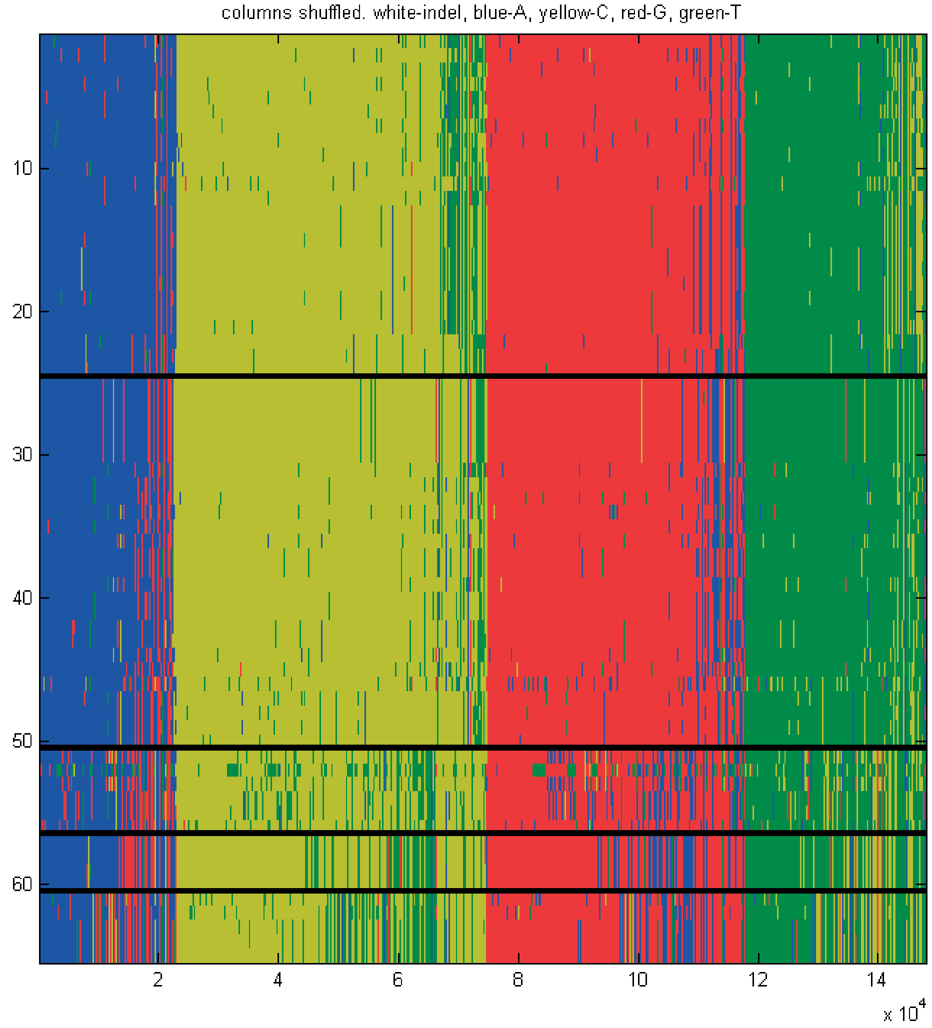
(TITLE-ABS-KEY (mcr1)) OR (TITLE-ABS-KEY (mcr-1)) OR (TITLE-ABS-KEY (mcr2)) OR (TITLE-ABS-KEY (mcr-2)) OR (TITLE-ABS-KEY (colistin AND mcr))

Science direct, 293 hits

"mcr1" OR "mcr-1" OR "mcr2" OR "mcr-2" OR (colistin AND mcr) AND LIMIT-TO(yearnav, "2017,2016,2015").

Google scolar, 221 hits

mcr1 OR mcr-1 OR mcr2 OR mcr-2 OR (colistin AND mcr)



Supplementary Figure 2. Output alignment of the BAPS clustering. A shuffled alignment of the core genome SNPs is divided by black bars to indicate five BAPS groups. Each row represents an isolate and the colors in every column represent a different nucleotide. On the left-side X-axis the number of nucleotides is indicated. Clusters were numbered from 1 to 5 according to the number of isolates they contain and cluster numbers are indicated on the right-side X-axis.
White-indel = insertion or deletion; blue-A = adenine; yellow-C = cytosine; red-G = guanine; green-T = thymine

SNP counts and clonal relationships of the isolates

SNP count was performed using kSNP v3.021[1] using a k-mer size of 19 bp and retaining only core SNPs (SNPs present in the core genome). The calculation of the genetic distance was performed using the average size of the genomes observed (4.9 Mbp). A

Hamming distance matrix was created comparing the SNP/Mbp distance between each pair of isolates in the study (Supplementary Table 3).

Among the 2080 pairs of isolates studied, 20 showed between 0 and 10 SNPs/Mbp differences and were considered closely related clones (Supplementary Table 4).

Among these, 13 pairs came from live chicken from southern Vietnam and were distributed across 5 farms (Supplementary Table 4). These results suggest regional circulation of *mcr-1* carrying clones. A pair of isolates originating from the Netherlands showed 0 SNP/Mbp difference and was isolated from the same batch of chicken meat. Another pair, originating from Malaysia, was isolated from the same chicken. These two pairs of isolates most probably represent duplicates. In only one of these isolates (EC5, from Malaysia) could the plasmid incompatibility group be identified (IncI2). Thus removing the pairs of isolates from the analysis did not modify the statistical outcome of the plasmid geographical distribution. Two closely related isolates sequenced for the present study (5.5 SNPs/Mbp difference) were recovered from travellers to China and Tunisia respectively. However, they carried the *mcr-1* gene on different plasmids (IncI2 and IncHI2 respectively). These 2 isolates also differed by less than 10 SNPs/Mbp with an isolate recovered from a patient in Denmark that carried an IncI2 plasmid. An isolate recovered from a traveller to South-America, formed a pair with an isolate from a patient in Malaysia. The plasmid type could not be identified for these 2 isolates. Finally a pair of isolates belonging to ST131 and showing 5.5 SNPs/Mbp difference between them were recovered from a patient in Spain and a chicken in Germany. These two isolates carried the *mcr-1* gene on two different plasmid types, IncX4 and IncHI2 respectively.

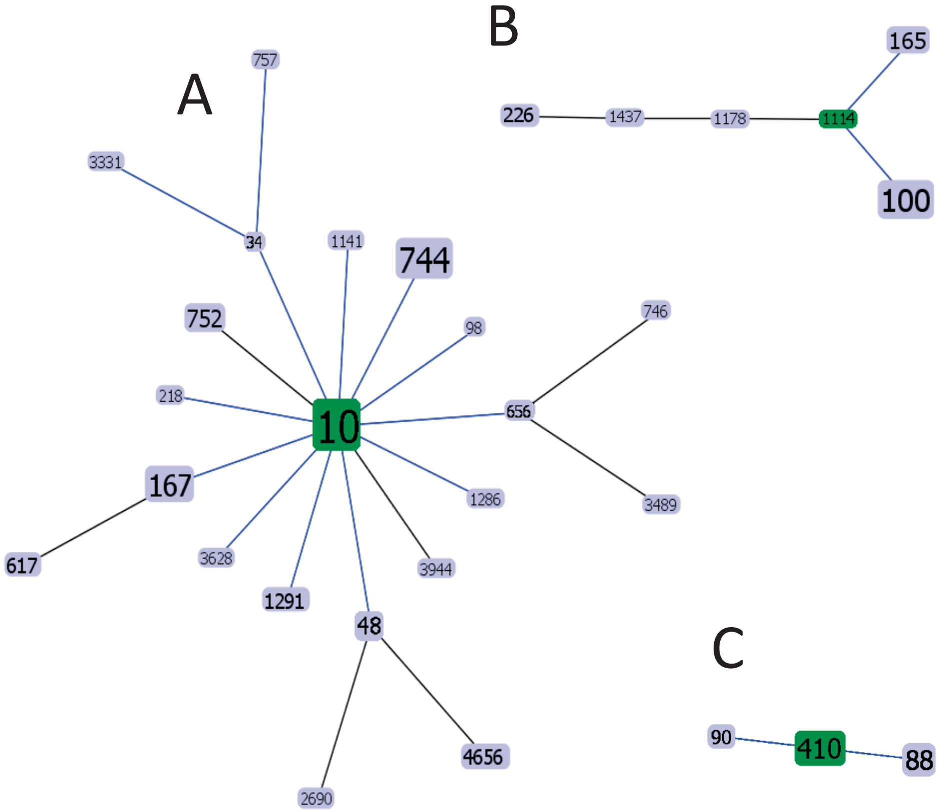
These findings implicate that, despite the spread of certain *E. coli* clones to different continents, the *mcr-1*-carrying plasmids are acquired separately in most cases.

WGS – quality check

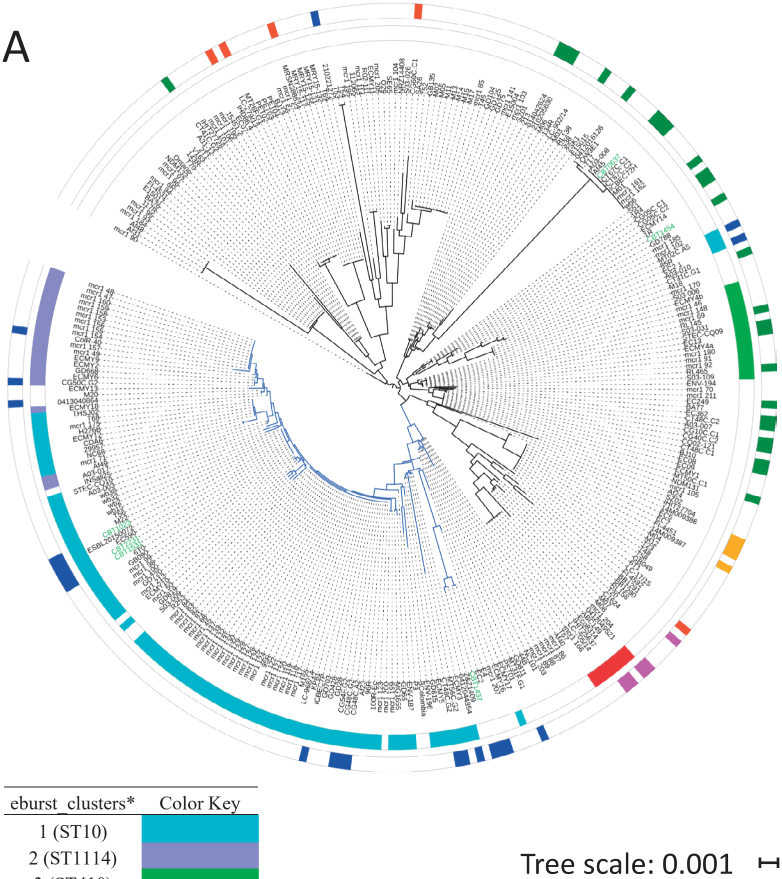
A total of 68 *mcr-1*-carrying *E. coli* genomes were analysed, including those 6 from travellers that were sequenced for the purpose of the present study. WGS datasets of 3 previously published isolates did not pass quality check and were discarded from the analysis. Two of these isolates (BB1290 [11.5 Mbp] and PO155 [8.5 Mbp]) had an unusually large genome. When tested using KmerFinder 2.0, the WGS fastq files of BB1290 and PO155 exhibited k-mer scores suggesting the presence of sequences from a different species in addition to the expected *E. coli*: *Enterobacter cloacae* and *Salmonella enterica* respectively. For isolate STEC-CQ10, only the sequences of the identified genes were available rather than the complete genome sequence or scaffolds. This resulted in a very fragmented genome (n = 6903 contigs), lack of intergenic regions and ultimately a serious decrease in quality of the population core genome determination.

N50 values, a statistic that defines assembly quality, for these isolates were 1096 for STEQ-CQ10; 14079 for PO155 and 31281 for BB1290. These values were much lower than values observed for the other genomes of the collection (N50 > 100000).

All these results taken together resulted in the exclusion of the WGS of STEQ-CQ10, BB1290 and PO155 from the phylogenetic analysis.



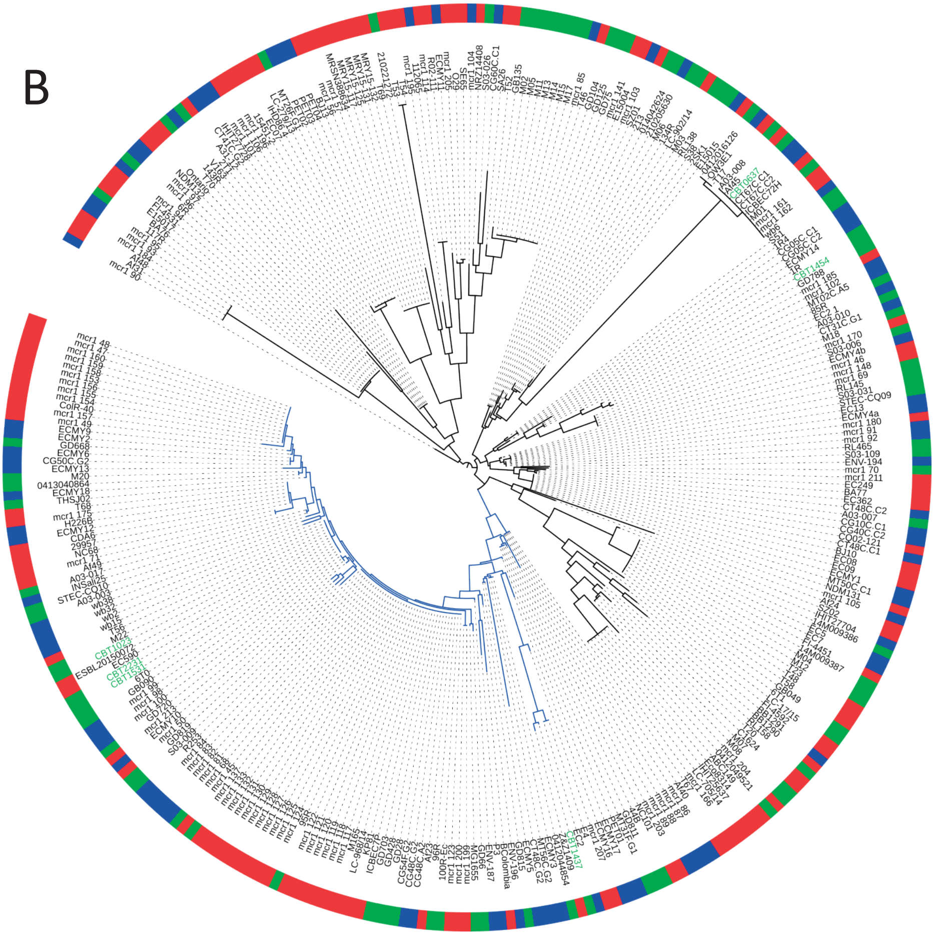
Supplementary Figure 3. Representation of the eBURST analysis showing the MLST clonal relationships of the 3 largest clusters containing 120 of the 312 included *mcr-1*-carrying *E. coli* isolates. Each rectangle represents a distinct sequence type (ST), with the number inside the rectangle referring to the ST's nomenclature according to the University of Warwick MLST database. The predicted founder ST of each group is indicated in green. Black lines indicate links drawn without recourse to tiebreak rules by goeBURST [2]; Blue lines indicate links drawn using tiebreak rule. The size of the rectangles that indicate the STs are proportional to the number of isolates they contain.
A: eBURST cluster 1 (n = 86 isolates).
B: eBURST cluster 2 (n = 20 isolates).
C: eBURST cluster 3 (n = 14 isolates).



eburst_clusters*	Color Key
1 (ST10)	🟦
2 (ST1114)	🟪
3 (ST410)	🟩
ST131	🟥
BAPS cluster	**Color key**
1	🟩
2	🟦
3	🟧
4	🟪
5	🟨

A: Colours of the inner circle indicate in which eBURST cluster (SLVs and DLVs clusters) the isolates belonged after MLST analysis; no colour indicates that the isolates were not part of one of the 3 indicated clusters. ST131 is not part of any eBURST cluster and is indicated separately. Colours of the outer circle indicate in which BAPS cluster the isolates belonged after whole genome sequence (WGS) analysis. No colour indicates that no WGS was available.

Supplementary Figure 4. Phylogeny of 312 *mcr-1*-carrying *E. coli* isolates. Maximum-likelihood trees based on concatenated MLST gene sequences, mid-point rooted.
The 6 travellers' isolates that were sequenced for this study are highlighted in green. The branch of the main clade containing most isolates from ST10 is coloured in blue. Tree scale in number of substitutions per site.
*The predicted founder ST of each cluster is indicated in brackets.

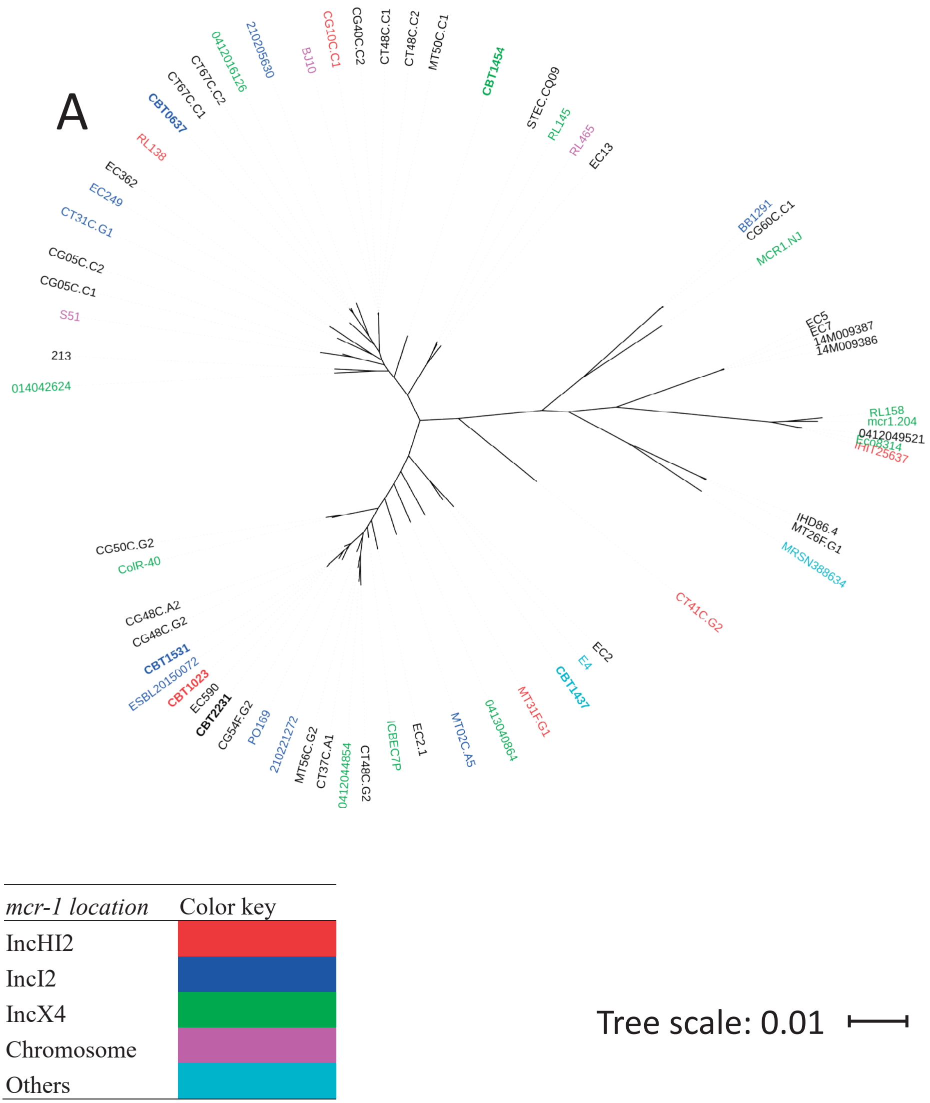


Color key	Health status
🟥	Diseased
🟩	Healthy
🟦	Unknown

B: Colours of the circle indicate the health status of the host from which the isolate was recovered, if available.

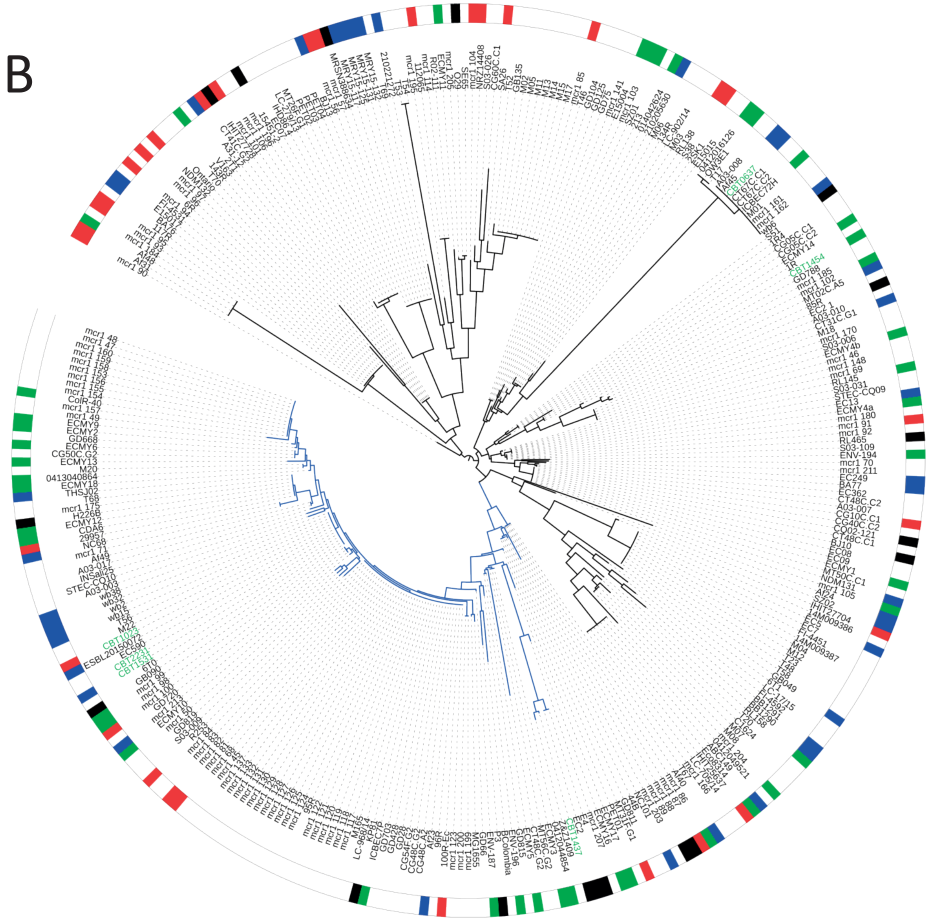
MLST analysis

Four different STs consisting of more than 2 isolates that were possibly clonally related were found in a single region: ST100 (10 isolates) and ST752 (3 isolates) in Europe; ST156 (12 isolates) and ST165 (3 isolates) in Asia. However all 3 isolates from ST752 were reported in a single study [3], as were 8 out of 10 isolates from ST100 [4].



A: Maximum-likelihood tree based on concatenated core genome sequences of the 65 *mcr-1*-carrying *E. coli* isolates. Isolate names are coloured according to the location of the *mcr-1* gene, either on an identified plasmid, chromosome of the isolate or unknown location. The 6 travellers' isolates that were sequenced for this study are highlighted in bold.

Supplementary Figure 5. Association between location of the *mcr-1* gene and WGS and MLST phylogeny of population of studied *mcr-1*-carrying *E. coli* isolates.
Only plasmids IncHI2, IncI2 and IncX4 are indicated by separate colours, all other plasmid types are indicated as "other". Trees scales in number of substitutions per site.



mcr-1 location	Colour key
IncHI2	🟥
IncI2	🟦
IncX4	🟩
Chromosome	⬛
Other	▨
Unknown	⬜

B: Maximum-likelihood tree based on concatenated MLST gene sequences, mid-point rooted. Colours indicate the location of the *mcr-1* gene. The branch of the main clade containing most isolates from ST10 is coloured in blue. The 6 travellers' isolates that were sequenced for this study are highlighted in green.

Antimicrobial resistance genes

Using the ResFinder tool, a total of 65 different unique acquired antimicrobial resistance (AMR) genes were identified in the *E. coli* WGS dataset. Many additional chromosomally encoded AMR genes were found using the CARD database. The most prevalent acquired gene (after *mcr-1*) was *tetA*, found in 52/65 (80%) of isolates followed by *sul2* (39/65; 60%) and the beta-lactamase gene TEM-1b in 38 isolates (58%). One isolate only carried just one additional acquired gene, the AmpC beta-lactamase gene *blaCMY-2*, and one isolate carried no other acquired AMR gene apart from *mcr-1*. In addition to *mcr-1*, each of the 65 isolates carried 4 or 5 genes of the *pmr* family (A, B, C, E and F variants). These genes are involved in the resistance to polymyxin, but are present in all *E. coli* strains [5]. Genetic alignments showed SNPs differences between the *pmrA* and *B* genes found in this study and both the sensitive and resistant templates previously described [5], thus preventing from drawing any conclusion relative to the involvement of these genes in the colistin resistance mechanisms. Carbapenemase encoding genes were found in the genomes of 4 isolates; NDM-5 (in isolates BJ10 and MCR1.NJ from China and USA respectively) and KPC-2 (EC362 and EC249, both from Singapore). 29/65 isolates (44.6%) carried plasmid-encoded CTX-M ESBL genes with CTX-M-55 being most prevalent in 18/29 (62.1%) of these isolates. Ten (15.4%) isolates were found to carry plasmid mediated *Qnr* quinolone resistance genes, 9 isolates carried *qnrS1* and one both *qnrS1* and *qnrS2*. In addition, 36 (55.4%) isolates displayed chromosomal mutations in both *parC* and *gyrA* genes associated with resistance to (fluoro)quinolones [6]. 34 (52.3%) isolates carried the *strA* and *strB* streptomycin resistance genes. The florfenicol resistance gene *floR* was present in 32 (49.2%) isolates; in 22 of 45 isolates from animals (48.9%) and 10 of 19 from humans (52.6%). The *baeR* and *baeS* genes are regulators of resistance mechanisms to novobiocin [7,8] and were both found in 64 (98.5%) and 65 (100%) isolates, respectively.

Supplementary data references

1. Gardner, S. N., Slezak, T. & Hall, B. G. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31, 2877–2878 (2015).
2. Francisco, A. P., Bugalho, M., Ramirez, M. & Carriço, J. a. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 10, 152 (2009).
3. Veldman, K. *et al.* Location of colistin resistance gene mcr-1 in Enterobacteriaceae from livestock and meat. *J. Antimicrob. Chemother.* 71, 2340–2342 (2016).
4. El Garch, F. *et al.* mcr-1 is borne by highly diverse Escherichia coli isolates since 2004 in food-producing animals in Europe. *Clin. Microbiol. Infect.* 23, 51e1-51e4 (2017).
5. Quesada, A. *et al.* Polymorphism of genes encoding PmrAB in colistin-resistant strains of Escherichia coli and Salmonella enterica isolated from poultry and swine. *J. Antimicrob. Chemother.* 70, 71–74 (2015).
6. Bagel, S., Hüllen, V., Wiedemann, B. & Heisig, P. Impact of gyrA and parC mutations on quinolone resistance, doubling time, and supercoiling degree of Escherichia coli. *Antimicrob. Agents Chemother.* 43, 868–875 (1999).
7. Lilic, M., Jovanovic, M., Jovanovic, G. & Savic, D. J. Identification of the CysB-regulated gene, hslJ, related to the Escherichia coli novobiocin resistance phenotype. *FEMS Microbiol. Lett.* 224, 239–246 (2003).
8. Baranova, N. & Nikaido, H. The BaeSR Two-Component Regulatory System Activates Transcription of the yegMNOB (mdtABCD) Transporter Gene Cluster in Escherichia coli and Increases Its Resistance to Novobiocin and Deoxycholate. *J. Bacteriol.* 184, 4168–4176 (2002).

SUPPLEMENTARY TABLES

Supplementary tables are too extensive to be printed and can be downloaded as an Excel file from:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5681592/bin/41598_2017_15539_MOESM2_ESM.xlsx