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

Sébastien Matamoros[1*], Jarne M. van Hattem[1*], Maris S. Arcilla[2], Niels Willemse[1], Damian C. Melles[2], John Penders[3,4], Trung Nguyen Vinh [1,5,6], Ngo Thi Hoa[6,7], COMBAT consortium, Menno D. de Jong[1], Constance Schultsz[1,5]

Department of Medical Microbiology, Academic Medical Center (AMC), Amsterdam, the Netherlands[1]; Department of Medical Microbiology and Infectious Diseases, Erasmus University Medical Center, Rotterdam, the Netherlands[2]; School for Public Health and Primary Care (Caphri), Department of Medical Microbiology, Maastricht University Medical Center (MUMC), Maastricht, the Netherlands[3]; School for Nutrition and Translational Research in Metabolism (NUTRIM), MUMC, Maastricht, the Netherlands[4]; Department of Global Health-Amsterdam Institute for Global Health and Development, AMC, Amsterdam, the Netherlands[5]; Oxford University Clinical Research Unit, Centre for Tropical Medicine, Ho Chi Minh City, Vietnam[6]; Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK[7]

*S.M. and J.M.v.H. contributed equally to this work.

*Address correspondence to Sébastien Matamoros (s.p.matamoros@amc.nl)

Running title: Phylogenetic analysis of *mcr-1*-carrying *E. coli* isolates

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.

Keywords

mcr-1; *Escherichia coli*; antimicrobial resistance; multilocus sequence typing; whole genome sequencing; plasmid; phylogeny; colistin; extended spectrum beta-lactamase (ESBL); international travel; ISA*pl1* transposon element.

41 ## Introduction

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

45 Colistin is an antibiotic of the polymyxin class, discovered in 1950 and effective against Gram-
46 negative bacteria [4]. The emergence of multidrug-resistant Gram-negative bacteria, especially those
47 producing carbapenemases, has reintroduced colistin as a last resort antibiotic for the treatment of
48 severe infections [5]. In contrast to its limited use in humans, colistin is widely used in food-producing
49 animals [6]. While colistin resistance was long thought to be caused by chromosomal mutations only [7],
50 the emergence of plasmid-mediated resistance, conferred by the mobilized colistin resistance (*mcr-*
51 *1*) gene, was recently reported [8]. This gene encodes for a protein of the phosphoethanolamine
52 transferase enzyme family, and its expression results in the addition of a phosphoethanolamine to
53 lipid A, the target of colistin, decreasing the interaction between colistin and the bacterial
54 lipopolysaccharide [8]. Since its discovery in 2015 in China, this gene has been described in several
55 bacterial species that were isolated from animals, animal food products, humans and environmental
56 samples from around the world [9–13]. Our previous study in travellers indicated acquisition of *mcr-1*
57 carrying bacteria by healthy individuals during travel to destinations around the world, potentially
58 related to food exposure, as well as rapid clearance after return [14]. It has been suggested that *mcr-1*
59 has spread from food animals to humans [8,15–17], but there is a lack of comparison of *mcr-1* carrying
60 isolates on a global level to support this hypothesis.

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

65

66

67 # Results

68 ## Literature search

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

73 ## Population structure

74 ### *Whole genome sequencing (WGS)*

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

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

92 ***Multilocus sequence typing (MLST)***

93 For 312 *E. coli* isolates originating from 69 studies, a MLST profile was published or could be deduced
94 from the corresponding WGS. Of these, 206 were isolated from animals or animal products (66.0%),
95 101 were isolated from humans (32.4%), including the 6 travel acquired isolates, and 5 from the
96 environment (1.6%). 141 Isolates from 25 studies (141/312; 45.2%) originated from Asia and 125
97 isolates from 25 studies (40.1%) from Europe together accounting for 85.3% of all included isolates.
98 The isolates represented 112 unique sequence types (STs) with ST10 being most common,
99 comprising 40/312 (12.8%) isolates originating from Africa, Asia, Europe and South-America.
100 eBURST analysis [19] was performed on all isolates included in the study, to identify their genetic
101 relatedness based on their MLST profiles. Three main clusters were identified, for which the
102 predicted founders, i.e. the ST in a cluster from which all other SLVs and DLVs in the cluster have
103 most likely diversified [19], were ST10, ST1114 and ST410. The largest cluster contained all 40 ST10
104 isolates and an additional 46 isolates in 21 STs that were single (SLV) or double locus variants (DLV) of
105 ST10 (86/312; 27.6%) (Supplementary Figure 3). The predicted founder of the second largest cluster
106 was ST1114, a SLV of ST165 and ST100, and included 19 isolates belonging to 7 different STs (5.4%),
107 while the third cluster was centred on ST410 and included 14 isolates from 3 different STs (4.5%).
108
109 A maximum-likelihood tree based on concatenated MLST gene sequences showed a main clade of
110 128 isolates (represented by blue branches in Figure 2 and Supplementary Figure 4; bootstrap value
111 of the main branch = 0.98), including most, but not all, isolates from the eBURST clusters of ST10 and
112 ST1114 (Supplementary Figure 4A). All isolates from these 2 eBURST clusters for which a WGS was
113 available were grouped in BAPS cluster 2. Similarly, all the isolates from the eBURST cluster ST410
114 grouped into BAPS cluster 1, along with 6 isolates from ST155. Seven isolates belonged to the
115 globally successful extra-intestinal pathogenic *E. coli* clone ST131 (Supplementary Figure 4A).
116 As observed in the WGS analysis, animal isolates were interspersed with isolates from humans and
117 the environment throughout the tree, as were isolates from different continents indicating a lack of

118 clustering by isolation source or geographical origin (Figure 2). Similarly, no clustering by health
119 status of the host was observed (Supplementary Figure 4B).

120 ## Mobile genetic elements

121 The plasmid incompatibility group of the *mcr-1*-carrying plasmids could be determined for 217
122 Enterobacteriaceae isolates from 7 different species (*Escherichia sp., Salmonella* sp., *Klebsiella* sp.,
123 *Cronobacter* sp., *Enterobacter* sp., *Kluyvera* sp. and *Shigella* sp.), representing a total of 219 plasmids
124 since 2 isolates carried 2 different plasmids carrying the *mcr-1* gene (Table 1). These plasmids were
125 described in 71 studies (1 to 33 plasmids per study, average = 3.1). In addition, the gene was
126 integrated in the chromosome of 6 isolates. The incompatibility group could not be determined for
127 27 of the 65 isolates for which WGS was available. Similarly the plasmid type was not available for
128 182 of the 312 isolates included in the MLST analysis. A total of 14 different plasmid incompatibility
129 groups were identified. 198/219 (90.4%) of the identified plasmids belonged to one of 3
130 incompatibility groups: IncX4 (77/219 plasmids, 35.2%), IncI2 (76/219 plasmids, 34.7%) and IncHI2
131 (45/219 plasmids, 20.5%). 50/76 IncI2 plasmids (65.8%) originated from Asia and 33/45 IncHI2
132 plasmids (73.3%) from Europe. IncX4 plasmids were more evenly distributed: 44/77 (57.1%) were
133 recovered from Europe, 29 from Asia (37.7%) and 4 from other regions (5.2%). Observed proportions
134 were significantly different from expected for IncI2 (χ^2-test, $p < 0.001$) and IncHI2 ($p < 0.001$) but not
135 for IncX4. The distribution of these 3 plasmid types was not significantly different from expected
136 between animal (χ^2-test, $p = 0.24$), human ($p = 0.88$) and environmental sources ($p = 0.38$) . Isolates
137 from the BAPS groups 1 and 2 carried plasmids from the 3 major types in similar proportions
138 (Supplementary Figure 5A; Supplementary Table 1). Isolates from the eBURST clusters of ST10 carried
139 plasmids belonging to 7 different incompatibility groups, including IncHI2, IncI2 and IncX4. No
140 clustering of plasmid type with MLST phylogeny was observed either (Supplementary Figure 5B;
141 Supplementary Table 1).

142 Figure 3 shows the alignment of the complete sequences or contigs from IncI2 (panel A), IncX4 (B)
143 and IncHI2 (C) incompatibility group plasmids. IncHI2 plasmids had the largest size, with sequence
144 lengths up to 267486 bp.
145 The ISA*pl1* transposon element situated upstream of the *mcr-1* gene was present in 7/9 (77.8%)
146 IncHI2 plasmids, but only in 11/29 (37.9%) of IncI2 plasmids and completely missing in all of the 24
147 reported IncX4 plasmids (Figure 3). In the isolates from travellers the ISA*pl1* transposon was
148 identified in 3 out of our 6 *mcr-1*-carrying contigs including one isolate from a traveler to Asia (ST101,
149 IncI2 incompatibility group), one to Africa (ST744, IncHI2) and one to South-America (ST744,
150 incompatibility group not identified).

151 ### Antimicrobial resistance genes

152 All sequences of the *mcr-1* gene collected in the present study were 100% identical to the original
153 sequence described by Liu *et al.* [8].
154 Multiple resistance genes were detected in most of the studied isolates (Supplementary Results). The
155 florfenicol resistance gene *floR* was present in 32 (49.2%) isolates; in 22 of 45 isolates from animals
156 (48.9%) and 10 of 19 isolates from humans (52.6%). The *baeR* and *baeS* genes, encoding novobiocin
157 resistance, were found in 64 (98.5%) and 65 (100%) isolates, respectively.
158 Plasmid analysis from the WGS data showed that 4 of the 29 *mcr-1*-carrying IncI2 plasmids (13.8%)
159 contained an additional ESBL gene. IncHI2 plasmids (n=9) carried between 0 and 12 additional AMR
160 genes. In particular, 4 plasmids carried CTX-M ESBL genes and 2 carried the *floR* gene. In 4 out of the
161 9 IncHI2 plasmids analysed in this study, the *mcr-1* gene was shown to be integrated alongside a
162 large multi-drug resistance (MDR) gene cassette (Figure 3C). None of the IncX4 plasmids carried
163 additional AMR genes.
164

165 ## Discussion

166 Analysis of all reported WGS of *mcr-1*-carrying isolates shows that the population of *E. coli* is highly
167 diverse, but is dominated by 2 large groups of related isolates. Most of the isolates from BAPS group

168 2 grouped into a MLST cluster centred on ST10. An overrepresentation of isolates related to ST10
169 and ST165 (a SLV of ST1114) in *mcr-1*-carrying *E. coli* isolates was previously reported at a smaller
170 scale in isolates from European farm animals [20]. *E. coli* ST10 and closely related STs are frequently
171 recovered from food and human intestinal samples and studies have shown a higher prevalence of
172 plasmid-carried AMR genes in ST10, including CTX-M ESBL genes, compared to other STs [21–24].
173 The second BAPS group of interest in our study included isolates belonging to ST155. This ST has
174 been described as a major vector of spread of ESBL genes from animals to humans [25]. It is thus
175 possible that zoonotic transmission lead to the spread of the *mcr-1* gene, as has been suggested in
176 studies from China and Vietnam [16,17], notably through the 2 main phylogenetic clusters identified in
177 this study.
178 Additionally, we found clonally related isolates, including some belonging to ST744, a SLV of ST10,
179 carrying the *mcr-1* gene on different plasmid backbones and recovered from different continents
180 (see Supplementary Results). These results point towards a worldwide dissemination *of mcr-1* driven
181 mainly by highly promiscuous plasmids rather than the worldwide spread of one or more *mcr-1*-
182 carrying clones. We hypothesize that several populations of *E. coli* isolates, notably those related to
183 ST10 or ST155, acquired the *mcr-1* gene due to their intrinsic ability of acquiring AMR genes and their
184 high prevalence in humans and food animals. These populations of commensal isolates then may
185 play a crucial role as a reservoir for this gene, which can explain their over-representation in the
186 present study.
187 In the timeframe of our literature search, 3 *E. coli* strains carrying the *mcr-2* gene were isolated from
188 animals in Belgium. These isolates belonged to ST10 (2 isolates) and ST167 which is a SLV of ST10 and
189 carried the gene on an IncX4-type plasmid. No WGS data was available from these isolates [26]. More
190 data about the *mcr-2* gene is needed to assess its spread and determine if *E. coli* ST10 plays a similar
191 role in its dissemination as it does for *mcr-1*.
192 More than 90% of published plasmid types carrying *mcr-1* genes belonged to either IncI2, IncX4 or
193 IncHI2. Almost 75% of the isolates carrying an IncHI2 plasmid originated from Europe: 26 from

194 animals and 7 from humans (Table 1 and Supplementary Table 1). In a traveller's isolate acquired in
195 Tunisia, the *mcr-1*-carrying plasmid was identified as an IncHI2-type backbone of the ST4 pMLST
196 subtype which co-carried a CTX-M-1 ESBL gene (Figure 3C). This traveller reported consumption of
197 beef, chicken and eggs during travel to Tunisia which can potentially be the source for the acquisition
198 of the *mcr-1* positive isolate. When investigating the presence of the *mcr-1* gene in cephalosporin
199 resistant *E. coli* isolates from chicken farms in Tunisia, Grami *et al.* [27] found that all 37 *mcr-1*-carrying
200 plasmids also belonged to the IncHI2-type, ST4 subtype and harboured CTX-M-1 genes. PFGE typing
201 of the isolates harbouring this plasmid showed various bacterial genetic backgrounds. Interestingly,
202 these chickens were all imported from France, either as adults or chicks. Other studies showed the
203 presence of this IncHI2, CTX-M-1 and *mcr-1* combination in *Salmonella enterica* Typhimurium isolates
204 from meat samples in Portugal from 2011 [28,29] and diarrhoeic veal calves in France [30]. The IncHI2
205 subtype ST4 was also detected in an *E. coli* isolate from retail chicken breast in Germany [31] and the
206 faecal sample of a veal calve from the Netherlands [32], suggesting widespread dissemination of this
207 particular plasmid in European farm animals and possible transmission to humans.
208 The high prevalence of novobiocin *baeR* and *baeS* and florfenicol *floR* resistance genes [33,34] in the
209 genomes of isolates of human and animal origin together with the fact that florfenicol and
210 novobiocin are used almost exclusively in veterinary medicine further supports the potential role of
211 food animals as an important reservoir of *mcr-1* containing bacteria and MGEs [15].
212 In contrast with the IncHI2 plasmids, 65.8% of all IncI2 plasmids recovered so far originated from
213 Asia, with a much lower prevalence in *mcr-1* carrying Enterobacteriaceae from other regions. Taken
214 together, these elements point toward a more regional circulation and dissemination of the *mcr-1*-
215 carrying plasmids IncI2 and IncHI2.
216 We found the ISA*pl1* transposon element associated with the *mcr-1* gene, as originally described by
217 Liu *et al.* [8] to be present in a minority of studied plasmids and contigs. However, since some of the
218 *mcr-1*-carrying contigs were obtained by assembly of Illumina short reads from WGS data, we cannot
219 exclude that some of these gaps are explained by an incomplete assembly of (plasmid) sequences.

220 The ISA*pl1* transposon element is considered to be the main driver of horizontal gene transfer of the
221 *mcr-1* gene and has been shown to be highly unstable in IncI2 plasmids [35–37]. The absence of the
222 ISA*pl1* transposon element in *mcr-1*-carrying IncX4 plasmids as described here has recently been
223 proposed to be essential for the maintenance of the *mcr-1* gene in this particular backbone, but the
224 exact mechanism still requires further investigation [38].
225 WGS analysis provided in-depth information about the *mcr-1*-carrying *E. coli* isolates and their
226 phylogenetic relationship, but the number of available genomes was limited. On the other hand,
227 whilst MLST data have a lower resolution, the higher number of available profiles allowed analysis of
228 the isolates' origin (geographical, source of isolation, diseased status of the host, etc.).
229
230 A limitation of our study is the potential for bias. The overrepresentation of isolates originating from
231 Asia and Europe could be explained by a higher prevalence of *mcr-1* genes on these continents, but
232 the effect of publication bias cannot be excluded. Isolates from North-America only represented
233 2.2% of the collection. Noteworthy, colistin, except for ophthalmic ointment, has never been
234 marketed for use in animals in the United States [39,40].
235
236 Sampling bias should also be considered when several isolates with an identical ST are presented
237 from a single study, as is the case for ST100 and ST752. Additionally, in the absence of a control
238 population of *mcr-1*-negative isolates obtained from similar sources as the *mcr-1* positive isolates,
239 results of analysis of population structures should be interpreted with caution. Because many studies
240 screened existing collections of (resistant) isolates for colistin resistance or presence of *mcr-1*,
241 selection bias has probably been introduced.
242 The findings in this study suggests that the *mcr-1* gene has locally and globally disseminated through
243 MGEs that are mainly IncHI2, IncI2 and IncX4 plasmids and provides additional support for the
244 hypothesis of the animal reservoir, that is driven by the use of colistin in livestock, as a source of *mcr-*

245 *1* in humans. A global ban of colistin use in animals to preserve colistin for use in human medicine
246 seems therefore justified.
247

248 Material and Methods

249 Selection of isolates for whole genome sequencing

250 We subjected 6 *mcr-1* positive isolates that were collected as part of a prospective study (COMBAT
251 study) aimed at studying acquisition of extended-spectrum β-lactamase (ESBL) -producing
252 Enterobacteriaceae during travel to whole-genome sequencing [14,41]. Additionally, we included 22
253 whole genome sequences of isolates from Vietnamese chickens and humans that were still
254 unpublished when performing our literature search [17].
255

256 Literature search

257 Relevant papers that published on *mcr-1* and *mcr-2* were identified in PubMed, Web of Science,
258 Scopus, ScienceDirect and Google Scholar using the query 'mcr-1 OR mcr1 OR mcr-2 OR mcr2 OR
259 (mcr AND colistin)' (see Supplementary Material for full search strategies).To be able to study the
260 associations between phylogeny, geographic distribution and isolation source we only included
261 sequences from papers that provided sufficient metadata. As a consequence, plasmids and
262 sequences that were deposited in online databases without metadata were not included in the
263 analysis.
264

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

266 Bacterial DNA was extracted from fresh pure cultures using the Qiagen DNeasy Blood and Tissue kit
267 (Qiagen, Hilden, Germany). Library preparation was done according to manufacturer's instruction
268 (Illumina, San Diego, CA, USA) and sequenced using Illumina MiSeq technology with 150nt paired-end

269 settings. Sequences have been deposited in the European Nucleotide Archive under the accession

270 numbers SAME104030441 to SAME104030446 .

271

272 Bio-informatic analysis

273 *MLST analysis*

274 For each *mcr-1*-carrying *E. coli* isolate for which the ST or the whole genome sequence was available,

275 the sequences of the corresponding alleles were downloaded from the *E. coli* MLST genes repository

276 of the University of Warwick

277 (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/handlers/getFileData/home/cbailster/mlst/zope/Extensio

278 ns/gadfly/Ecoli/DB/) and concatenated. When STs of isolates were not described in literature, the ST

279 was determined from available whole genomes using the online service provided by the Center for

280 Genomic Epidemiology (https://cge.cbs.dtu.dk/services/MLST/) according to the Achtman MLST

281 scheme [42,43]. MLST clusters (STs and their single locus or double locus variants) were defined using e-

282 burst V3 (http://eburst.mlst.net/v3/enter_data/single/) [19] and goeBURST v1.2.1 [44] using only profiles

283 from this study.

284 *WGS and plasmid analysis*

285 Raw sequence reads in fastq format or pre-assembled sequences in fasta format were downloaded

286 from online databases for all available isolates (Supplementary Table 1). Additional sequences not

287 yet deposited in online databases were requested from their respective authors. The quality of the

288 raw sequence reads was checked using fastqc

289 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), quast [45] and KmerFinder 2.0

290 (https://cge.cbs.dtu.dk/services/KmerFinder/) (see Supplementary Methods for more details). Reads

291 were trimmed using Trimmomatic V0.33 [46]. *De-novo* genome assembly was performed with SPAdes

292 3.9 [47] for Illumina short reads and with Canu v1.3 for PacBio long reads [48]. Contigs of less than 500 bp

293 long were removed from the genomes to improve the overall quality of the assembly. Size of the

294 genomes was calculated by adding the length of all remaining contigs. Identification of open reading

295 frames (ORFs) and gene contents in the assembled genomes (*de-novo* assemblies and pre-assembled
296 sequences) was performed using Prokka v1.11 [49]. Core genome analysis was performed with Roary
297 v3.6.8 [50]. Clustering of isolates was performed using the hierBAPS module of the Bayesian Analysis of
298 Population Structure (BAPS) software v6.0 [18]. The core genome alignment output provided by Roary
299 was used as input for BAPS with 2 levels of hierarchy and a maximum number of cluster (K) of 10. The
300 estimated number of clusters was 5 for both levels of hierarchy.
301 Sequences (concatenated MLST loci or concatenated core genes) were aligned using mafft v6.864b
302 [51]. The resulting alignment was used as input for calculation of distances and tree building using
303 RAxML v8.1.6 [52]. TMLST and WGS trees were visualized using iTOL v3.3.1 (http://itol.embl.de/) [53].

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

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

316 ***Statistics***

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

320

321

322 ## References

323

324 1. Appelbaum, P. C. 2012 and beyond: Potential for the start of a second pre-antibiotic era? *J.*
325 *Antimicrob. Chemother.* **67,** 2062–2068 (2012).

326 2. Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature* **529,**
327 336–343 (2016).

328 3. Laxminarayan, R. *et al.* Antibiotic resistance-the need for global solutions. *Lancet Infect. Dis.*
329 **13,** 1057–1098 (2013).

330 4. Koyama, Y., Kurosasa, A., Tsuchiya, A. & Takakuta, K. A new antibiotic 'colistin' produced by
331 spore-forming soil bacteria. *J Antibiot* **3,** 457–458 (1950).

332 5. Lim, L. M. *et al.* Resurgence of colistin: a review of resistance, toxicity, pharmacodynamics,
333 and dosing. *Pharmacotherapy* **30,** 1279–91 (2010).

334 6. Catry, B. *et al.* Use of colistin-containing products within the European Union and European
335 Economic Area (EU/EEA): development of resistance in animals and possible impact on human
336 and animal health. *Int. J. Antimicrob. Agents* **46,** 297–306 (2015).

337 7. Olaitan, A. O., Morand, S. & Rolain, J. Mechanisms of polymyxin resistance: acquired and
338 intrinsic resistance in bacteria. *Front. Microbiol.* **5,** 1–18 (2014).

339 8. Liu, Y.-Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in
340 animals and human beings in China: a microbiological and molecular biological study. *Lancet*
341 *Infect. Dis.* **3099,** (2015).

342 9. Skov, R. L. & Monnet, D. L. Plasmid-mediated colistin resistance (mcr-1 gene): three months
343 later, the story unfolds. *Eurosurveillance* **21,** (2016).

344 10. Schwarz, S. & Johnson, A. P. Transferable resistance to colistin: a new but old threat. *J.*
345 *Antimicrob. Chemother.* **71,** 2066–2070 (2016).

346 11. Leangapichart, T. *et al.* Acquisition of mcr-1 Plasmid-Mediated Colistin Resistance in
347 Escherichia coli and Klebsiella pneumoniae during Hajj 2013 and 2014. *Antimicrob. Agents*

348 *Chemother.* **60,** 6998–6999 (2016).

349 12. Vading, M. *et al.* Frequent acquisition of low-virulence strains of ESBL-producing Escherichia
350 coli in travellers. *J. Antimicrob. Chemother.* **71,** 3548–3555 (2016).

351 13. Bernasconi, O. J. *et al.* Travelers Can Import Colistin-Resistant Enterobacteriaceae, Including
352 Those Possessing the Plasmid-Mediated mcr-1 Gene. *Antimicrob. Agents Chemother.* **60,**
353 5080–5084 (2016).

354 14. Arcilla, M. S. *et al.* Dissemination of the mcr-1 colistin resistance gene. *Lancet*
355 *Infect. Dis.* **16,** 147–149 (2016).

356 15. Poirel, L. & Nordmann, P. Emerging plasmid-encoded colistin resistance: the animal world as
357 the culprit? *J. Antimicrob. Chemother.* **71,** 2326–2327 (2016).

358 16. Wang, Y. *et al.* Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1
359 in Chinese poultry production. *Nat. Microbiol.* **2,** 16260 (2017).

360 17. Trung, N. V. *et al.* Zoonotic Transmission of mcr-1 Colistin Resistance Gene from
361 Small-Scale Poultry Farms, Vietnam. *Emerg. Infect. Dis. J.* **23,** 529 (2017).

362 18. Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M. & Corander, J. Hierarchical and spatially
363 explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* **30,** 1224–1228
364 (2013).

365 19. Feil Edward J an d Li, B. C., Aanensen, D. M., Hanage, W. P. & Spratt, B. G. eBURST: inferring
366 patterns of evolutionary descent among clusters of related bacterial genotypes from
367 multilocus sequence typing data. *J. Bacteriol.* **186,** 1518–1530 (2004).

368 20. El Garch, F. *et al.* mcr-1 is borne by highly diverse Escherichia coli isolates since 2004 in food-
369 producing animals in Europe. *Clin. Microbiol. Infect.* **23,** 51e1-51e4 (2017).

370 21. Manges, A. R. & Johnson, J. R. Food-Borne Origins of Escherichia coli Causing Extraintestinal
371 Infections. *Clin. Infect. Dis.* **55,** 712–719 (2012).

372 22. Oteo, J. *et al.* Extended-spectrum β-lactamase-producing Escherichia coli in Spain belong to a
373 large variety of multilocus sequence typing types, including ST10 complex/A, ST23 complex/A

374 and ST131/B2. *Int. J. Antimicrob. Agents* **34,** 173–176 (2009).

375 23. Fam, N. *et al.* CTX-M-15-producing Escherichia coli clinical isolates in Cairo (Egypt), including
376 isolates of clonal complex ST10 and clones ST131, ST73, and ST405 in both community and
377 hospital settings. *Microb. Drug Resist.* **17,** 67–73 (2011).

378 24. Manges, A. R. *et al.* Clonal distribution and associated characteristics of Escherichia coli
379 clinical and surveillance isolates from a military medical center. *Diagn. Microbiol. Infect. Dis.*
380 **87,** 382–385 (2017).

381 25. Skurnik, D. *et al.* Emergence of antimicrobial-resistant Escherichia coli of animal origin
382 spreading in humans. *Mol. Biol. Evol.* **33,** 898–914 (2016).

383 26. Xavier, B. B. *et al.* Identification of a novel plasmid-mediated colistinresistance gene, mcr-2, in
384 Escherichia coli, Belgium, june 2016. *Eurosurveillance* **21,** (2016).

385 27. Grami, R. *et al.* Impact of food animal trade on the spread of mcr-1-mediated colistin
386 resistance, tunisia, july 2015. *Eurosurveillance* **21,** 1–5 (2016).

387 28. Tse, H. & Yuen, K.-Y. Dissemination of the mcr-1 colistin resistance gene. *Lancet Infect. Dis.*
388 **3099,** 3099 (2015).

389 29. Figueiredo, R., Henriques, A., Sereno, R., Mendonça, N. & da Silva, G. J. Antimicrobial
390 Resistance and Extended-Spectrum β-Lactamases of Salmonella enterica Serotypes Isolated
391 from Livestock and Processed Food in Portugal: An Update. *Foodborne Pathog. Dis.* **12,** 110–
392 117 (2014).

393 30. Haenni, M. *et al.* Co-occurrence of extended spectrum β lactamase and MCR-1 encoding
394 genes on plasmids. *Lancet Infect. Dis.* **3099,** 3099 (2016).

395 31. Falgenhauer, L. *et al.* Chromosomal Locations of mcr-1 and bla(CTX-M-15) in Fluoroquinolone-
396 Resistant Escherichia coli ST410. *Emerg. Infect. Dis.* **22,** 1689–1691 (2016).

397 32. Veldman, K. *et al.* Location of colistin resistance gene mcr-1 in Enterobacteriaceae from
398 livestock and meat. *J. Antimicrob. Chemother.* **71,** 2340–2342 (2016).

399 33. Lilic, M., Jovanovic, M., Jovanovic, G. & Savic, D. J. Identification of the CysB-regulated gene,

400 hslJ, related to the Escherichia coli novobiocin resistance phenotype. *FEMS Microbiol. Lett.*
401 **224,** 239–246 (2003).
402 34. Baranova, N. & Nikaido, H. The BaeSR Two-Component Regulatory System Activates
403 Transcription of the yegMNOB (mdtABCD) Transporter Gene Cluster in Escherichia coli and
404 Increases Its Resistance to Novobiocin and Deoxycholate. *J. Bacteriol.* **184,** 4168–4176 (2002).
405 35. Stoesser, N., Mathers, A. J., Moore, C. E., Day, N. P. & Crook, D. W. Colistin resistance gene
406 mcr-1 and pHNSHP45 plasmid in human isolates of Escherichia coli and Klebsiella
407 pneumoniae. *Lancet Infect. Dis.* **3099,** 3099 (2016).
408 36. Falgenhauer, L. *et al.* Colistin resistance gene mcr-1 in extended-spectrum β-lactamase-
409 producing and carbapenemase-producing Gram-negative bacteria in Germany. *Lancet Infect.*
410 *Dis.* **3099,** 3099 (2016).
411 37. Snesrud, E. *et al.* Analysis of Serial Isolates of mcr-1- positive Escherichia coli Reveals a Highly
412 Active ISApl1 Transposon. *Antimicrob. Agents Chemother.* (2017). doi:10.1128/AAC.00056-17
413 38. Sun, J. *et al.* Genetic Analysis of the IncX4 Plasmids: Implications for a Unique Pattern in the
414 mcr-1 Acquisition. *Sci. Rep.* **7,** 424 (2017).
415 39. FDA. *2015 Summary Report On Antimicrobials Sold or Distributed for Use in Food-Producing*
416 *Animals*. (2016). at
417 <https://www.fda.gov/downloads/ForIndustry/UserFees/AnimalDrugUserFeeActADUFA/UCM
418 534243.pdf>
419 40. European Medicines Agency. *Updated advice on the use of colistin products in animals within*
420 *the European Union: development of resistance and possible impact on human and animal*
421 *health.* **44,** (2016).
422 41. Arcilla, M. S. *et al.* Import and spread of extended-spectrum beta-lactamase-producing
423 Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre
424 cohort study. *Lancet Infect. Dis.* **17,** 78–85 (2017).
425 42. Larsen, M. V. *et al.* Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin.*

426 *Microbiol.* **50,** 1355–1361 (2012).

427 43. Wirth, T. *et al.* Sex and virulence in Escherichia coli: An evolutionary perspective. *Mol.*
428 *Microbiol.* **60,** 1136–1151 (2006).

429 44. Francisco, A. P., Bugalho, M., Ramirez, M. & Carriço, J. a. Global optimal eBURST analysis of
430 multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* **10,** 152 (2009).

431 45. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome
432 assemblies. *Bioinformatics* **29,** 1072–1075 (2013).

433 46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence
434 data. *Bioinformatics* **30,** 2114–2120 (2014).

435 47. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to
436 Single-Cell Sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).

437 48. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting
438 and repeat separation. *bioRxiv* (2017). at
439 <http://biorxiv.org/content/early/2017/01/04/071282.abstract>

440 49. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069
441 (2014).

442 50. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31,**
443 3691–3693 (2015).

444 51. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment
445 program. *Brief. Bioinform.* **9,** 286–298 (2008).

446 52. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
447 phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).

448 53. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and
449 annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44,** W242–W245 (2016).

450 54. Carattoli, A. *et al.* In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid
451 Multilocus Sequence Typing. *Antimicrob. Agents Chemother.* **58,** 3895–3903 (2014).

452 55. Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator
453 (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12,** 402 (2011).
454 56. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob.*
455 *Chemother.* **67,** 2640–2644 (2012).
456 57. McArthur, A. G. & Wright, G. D. Bioinformatics of antimicrobial resistance in the age of
457 molecular epidemiology. *Curr. Opin. Microbiol.* **27,** 45–50 (2015).
458
459

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

469 ## Additional information

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

476 ### Competing financial interests
477 The author(s) declare no competing financial interests.

478 ### The COMBAT consortium (in alphabetical order)
479 Martin C.J. Bootsma[8], Perry J. van Genderen[9], Abraham Goorhuis[10], Martin P. Grobusch[10], Nicky
480 Molhoek[9], Astrid M.L. Oude Lashof[3], Ellen E. Stobberingh[3], Henri A. Verbrugh[2].
481 Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, the
482 Netherlands and Department of Mathematics, Faculty of Science, Utrecht University, Utrecht, the
483 Netherlands[8]. Department of Internal Medicine, Havenziekenhuis - Institute for Tropical Diseases,
484 Rotterdam, the Netherlands[9]. Center of Tropical Medicine and Travel Medicine, Academic Medical
485 Centre (AMC), Amsterdam, the Netherlands[10].

Figure legends

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 colours (isolates identifiers) 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.

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.

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.

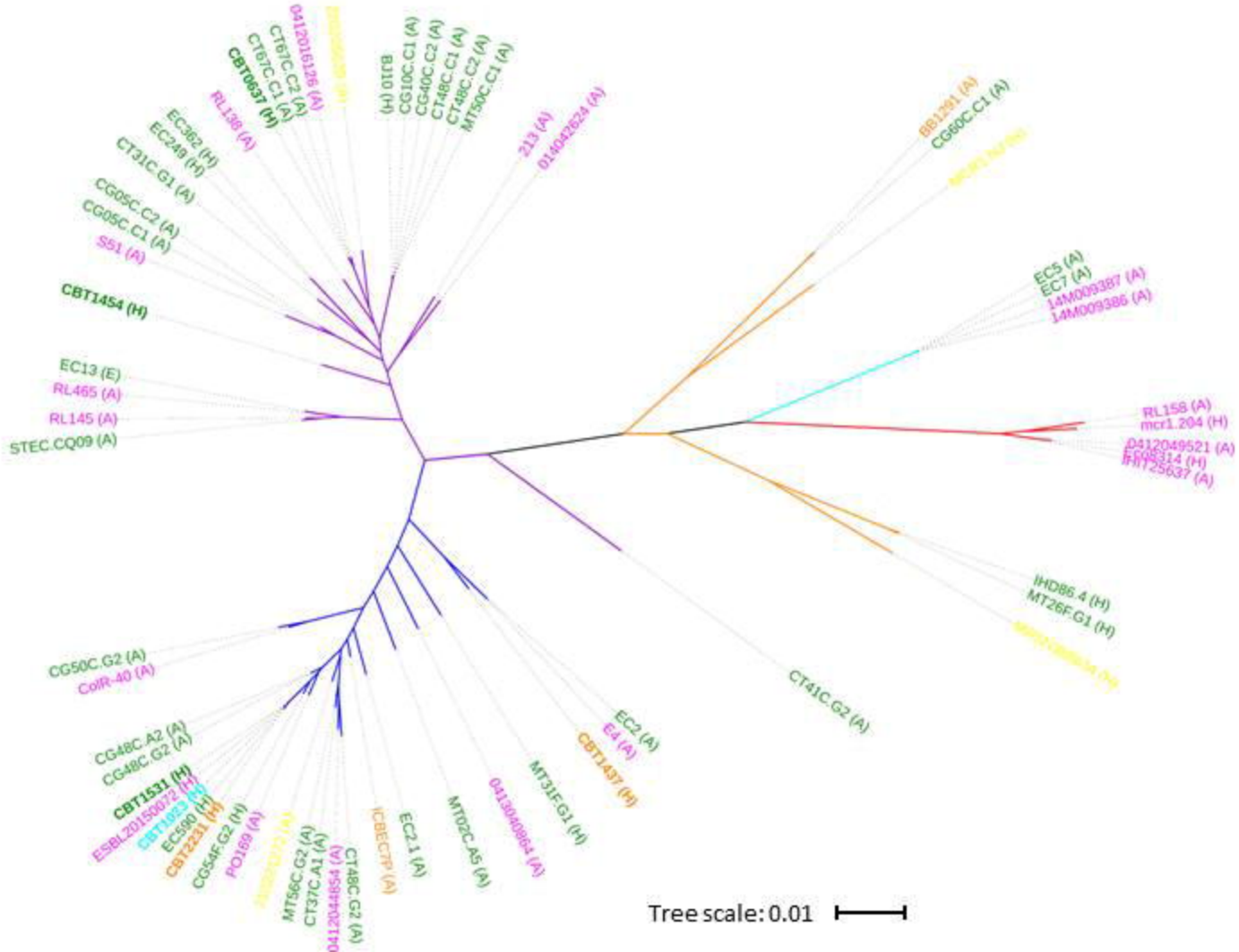
Panel C: Putative MDR cassette is highlighted in orange.

	IncF	IncFI	IncFIB	IncFII	IncHI1	IncHI1A/ IncHI1B/ IncFIA	IncHI2	IncI2	IncI2/ IncX4	IncP	IncX3/ IncX4	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**

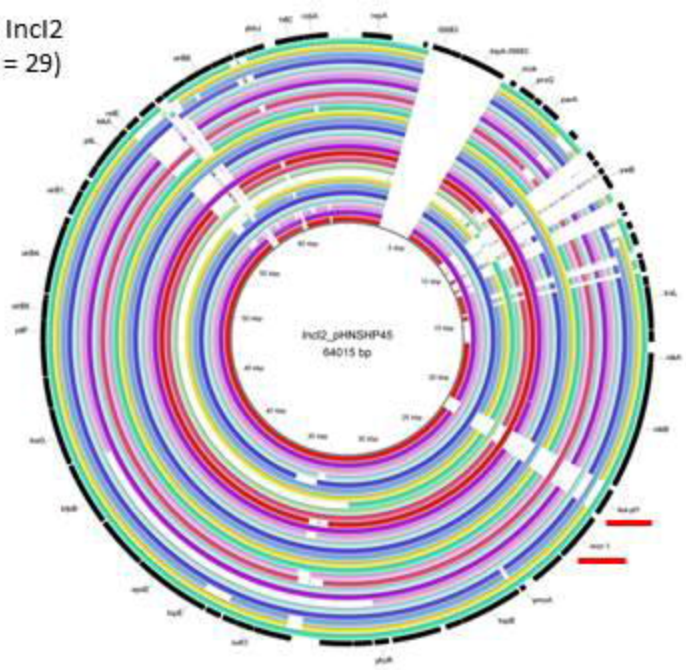
Table 1: Incompatibility types of *mcr-1* carrying plasmids and distribution by geographical regions

BAPS cluster (branches)	Color key
1 (n* = 26)	
2 (n = 24)	
3 (n = 6)	
4 (n = 5)	
5 (n = 4)	

Geographical origin (leaves)	Color key
Asia (n* = 36)	
Europe (n = 20)	
North-America (n = 4)	
South-America (n = 4)	
Africa (n = 1)	

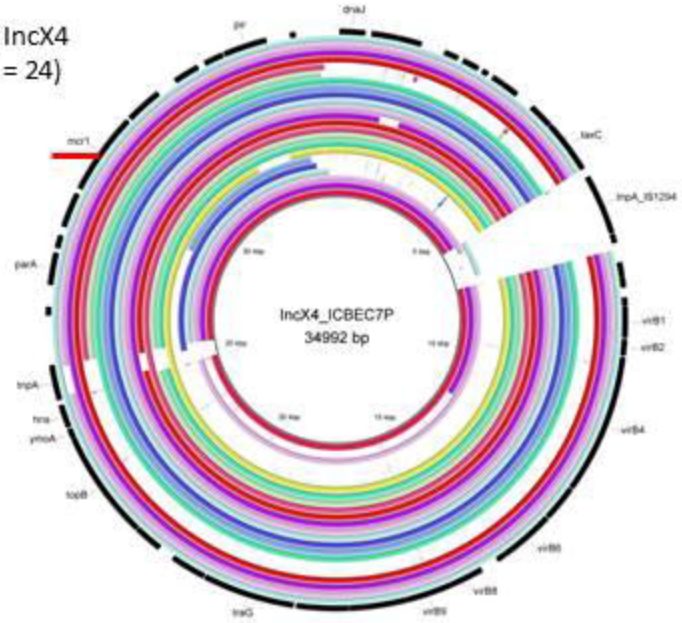


A: IncI2
(n = 29)



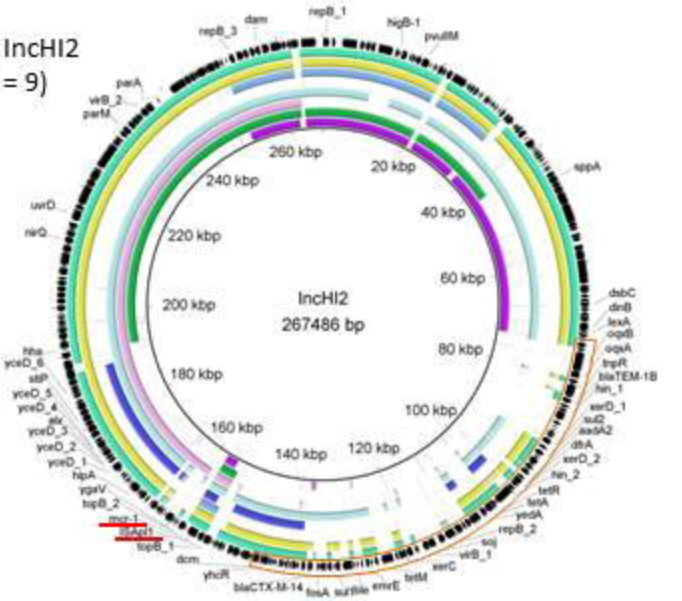
210205360_pSLy1
210221272_plasmid_pSLy21
A31-12_plasmid_pA31-12
ABC149_plasmid_pABC149
Af23_plasmid_pAF23
BA76_plasmid_pBA76
BA77_plasmid_pBA77
BB1291_mcr1_contig
CBT0637_mcr1_contig
* CBT1531_mcr1_contig
* CT31C.G1_mcr1_contig
E15017_plasmid_pE15017
EC13_plasmid_pEC13-1
EC249_mcr1_contig
EC5_plasmid_pEC5-1
EG0430_plasmid_pEG430-1
ESBL20150072_mcr1_contig
JS-61_plasmid_pECJS-61-63
mcr1_213_payne_mcr1_contig
MT02C.A5_mcr1_contig
PO169_mcr1_contig
S2.14_plasmid_pS2.14-2
S3_mcr1_contig
SC23_plasmid_pSCS23
SZ02_plasmid_pmcr1_IncI2
VT55363_plasmid_pVT553
WCH1410_plasmid_pMCR_1410
WF-5-19C_plasmid_pWF-5-19C_mcr-1

B: IncX4
(n = 24)

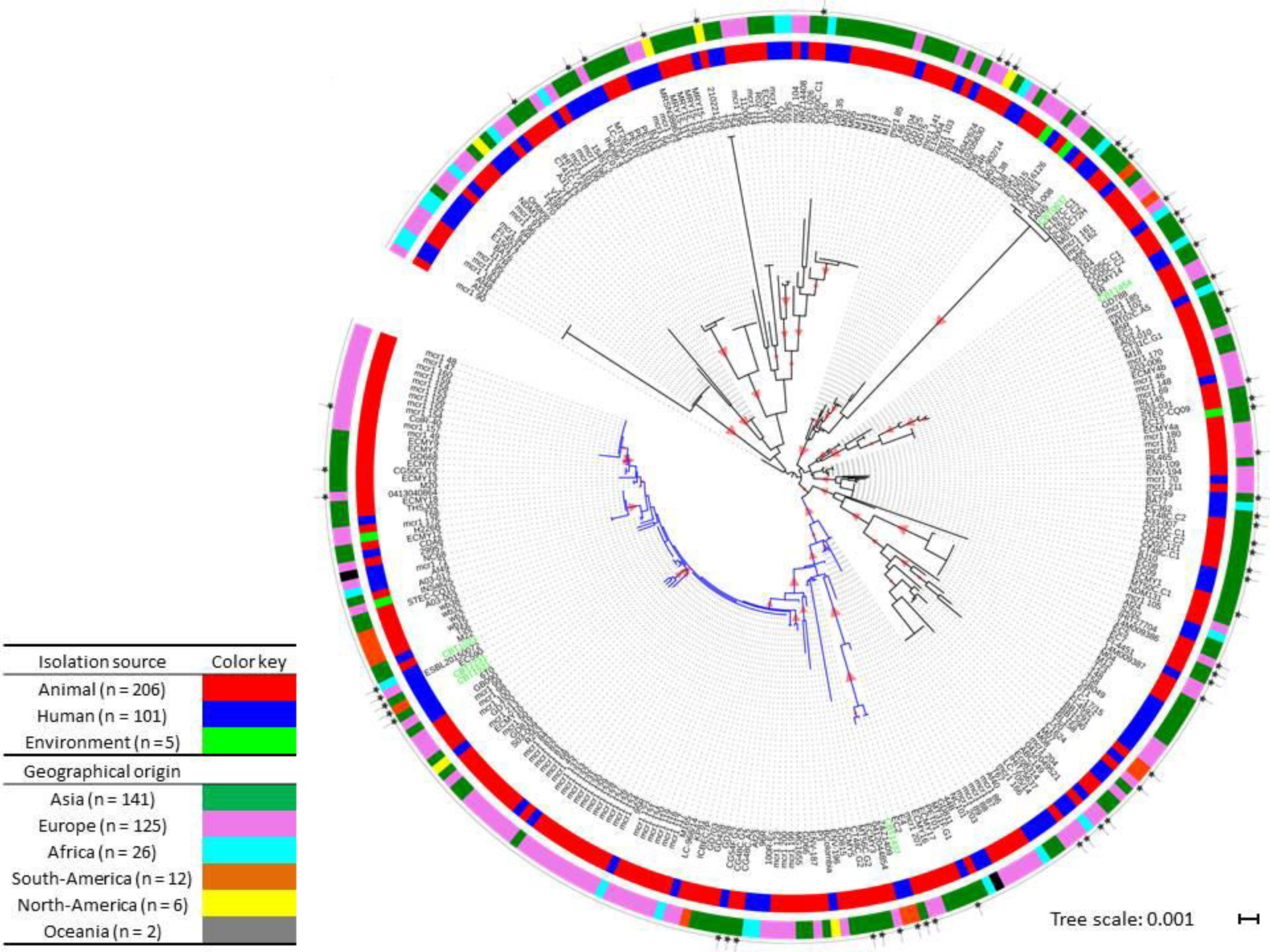


014042624_mcr1_contig
0412016126_mcr1_contig
0412044854_mcr1_contig
0413040864_mcr1_contig
2012CEB2196SAL_mcr1_contig
2012CEB4337SAL_mcr1_contig
Af48_plasmid_pAF48
* CBT1454_mcr1_contig
ColR-40_mcr1_contig
E15004_plasmid_pE15004
EC702_mcr1_contig
ECMY3_plasmid_pECMCR-1101
Eco8314_mcr1_contig
ENV-187_plasmid
GD-8_plasmid_pECGD-8-33
ICBEC72H_plasmid_pICBEC72Hmcr
KP-6884_plasmid_pMCR1.2-IT
mcr1.204_mcr1_contig
MCR1_NJ_mcr1_contig
OW3E1_plasmid_pOW3E1
RL145_mcr1_contig
RL158_mcr1_contig
SZ04_plasmid_pmcr1

C: IncHI2
(n = 9)



* CBT1023_mcr1_contig
CG10C.C1_mcr1_contig
CT41C.G2_mcr1_contig
JS-59_plasmid_pECJS-59-244
Sal25_plasmid
MT31F.G1_mcr1_contig
S38_plasmid_pS38
S38_plasmid_pS38


Isolation source
Color key
Animal (n = 206)
Human (n = 101)
Environment (n = 5)
Geographical origin
Asia (n = 141)
Europe (n = 125)
Africa (n = 26)
South-America (n = 12)
North-America (n = 6)
Oceania (n = 2)
Tree scale: 0.001