

OPEN

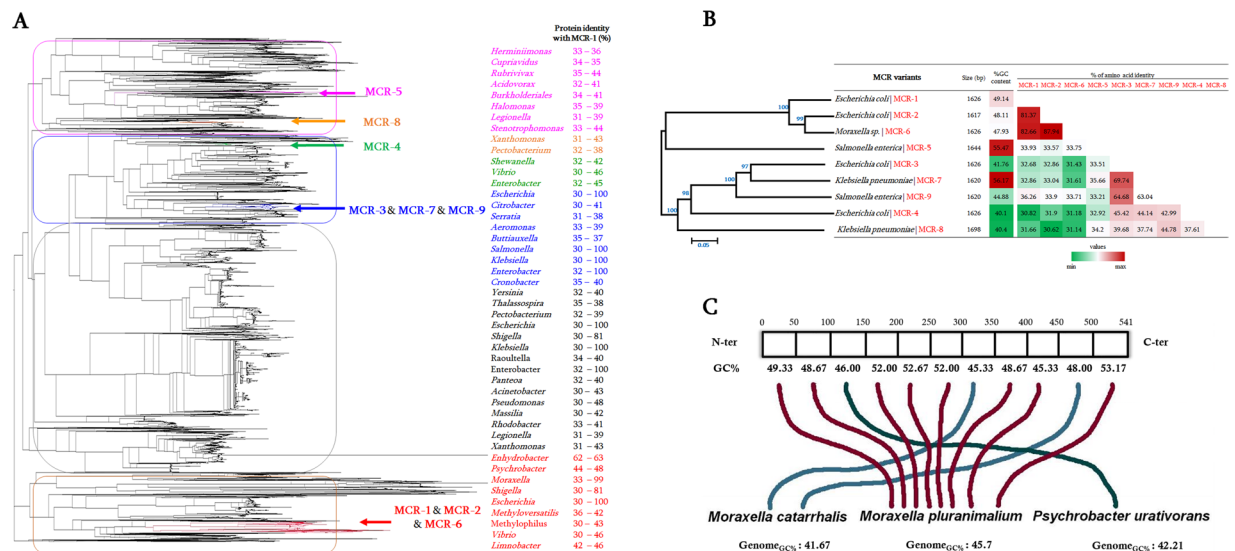
# Massive analysis of 64,628 bacterial genomes to decipher water reservoir and origin of mobile colistin resistance genes: is there another role for these enzymes?

Mariem Ben Khedher<sup>1</sup>, Sophie Alexandra Baron<sup>1,2</sup>, Toilhata Riziki<sup>1</sup>, Raymond Ruimy<sup>3</sup>, Didier Raoult<sup>1,2</sup>, Seydina M. Diene<sup>1,2\*</sup> & Jean-Marc Rolain<sup>1,2\*</sup>

Since 2015, new worrying colistin resistance mechanism, mediated by *mcr-1* gene has been reported worldwide along with eight newly described variants but their source(s) and reservoir(s) remain largely unexplored. Here, we conducted a massive bioinformatic analysis of bacterial genomes to investigate the reservoir and origin of *mcr* variants. We identified 13'658 MCR-1 homologous sequences in 494 bacterial genera. Moreover, analysis of 64'628 bacterial genomes (60 bacterial genera and 1'047 species) allows identifying a total of 6'651 significant positive hits (coverage >90% and similarity >50%) with the nine MCR variants from 39 bacterial genera and more than 1'050 species. A high number of MCR-1 was identified in *Escherichia coli* ( $n = 862$ ). Interestingly, while almost all variants were identified in bacteria from different sources (i.e. human, animal, and environment), the last variant, MCR-9, was exclusively detected in bacteria from human. Although these variants could be identified in bacteria from human and animal sources, we found plenty MCR variants in unsuspected bacteria from environmental origin, especially from water sources. The ubiquitous presence of *mcr* variants in bacteria from water likely suggests another role in the biosphere of these enzymes as an unknown defense system against natural antimicrobial peptides and/or bacteriophage predation.

The increase in bacterial resistance to antibiotics is a phenomenon that has attracted much attention in recent years. This resistance has greatly affected the effectiveness of antibiotics in both hospital and non-hospital settings. However, antibiotic resistance existed before the use of antibiotics in human medicine. Indeed, prior to the 19th century, antibiotic resistance as well as genetic sequences of resistance to antibiotics can be found<sup>1,2</sup>. This is not surprising given that antibiotic secretion is a natural phenomenon in many organisms and resistance as well<sup>3</sup> and thus other roles of 'antibiotic resistance' enzymes is expected. Nowadays, the bacterial resistance covers almost all antibiotics, including the carbapenems used for treatment against infections caused by multi-drug resistant (MDR) bacteria<sup>4,5</sup>. The increase is related to the selection pressure that favour organisms with a natural resistance (e.g. *Stenotrophomonas* sp. and thienamycin) and those that have acquired, by horizontal transfer, sequences allowing them to become resistant. Thus, this issue in human medicine is constantly highlighted by international health organizations such as the CDC (Centers for Diseases Control and Prevention) or the WHO (World Health Organization), which are calling for action to combat this phenomenon<sup>6,7</sup>. Moreover, it has been demonstrated recently that genes that were initially identified in human as Metallo- $\beta$ -lactamase (MBL) proteins fold, exhibited several other activities such as nuclease, glyoxalase or  $\beta$ -lactamase<sup>8</sup>. The same sequence may have a different predominant activity depending on the organism considered in a specific ecosystem or in the biosphere such as nuclease in eukaryotes and  $\beta$ -lactamase in bacteria. Thus, vertebrates such as humans have in their genome eighteen genes whose sequences are annotated as MBL and whose natural activity, when studied, is essentially that of nuclease/ribonuclease, glyoxalase, and metabolism-associated enzymes<sup>9</sup>. This may be a typical model

<sup>1</sup>Aix Marseille Univ., IRD, APHM, MEPHI, IHU-Mediterranee Infection, Marseille, France. <sup>2</sup>IHU-Mediterranee Infection, Marseille, France. <sup>3</sup>Centre Hospitalier Universitaire de Nice, Laboratoire de bactériologie, Nice, France. \*email: [seydina.diene@univ-amu.fr](mailto:seydina.diene@univ-amu.fr); [jean-marc.rolain@univ-amu.fr](mailto:jean-marc.rolain@univ-amu.fr)

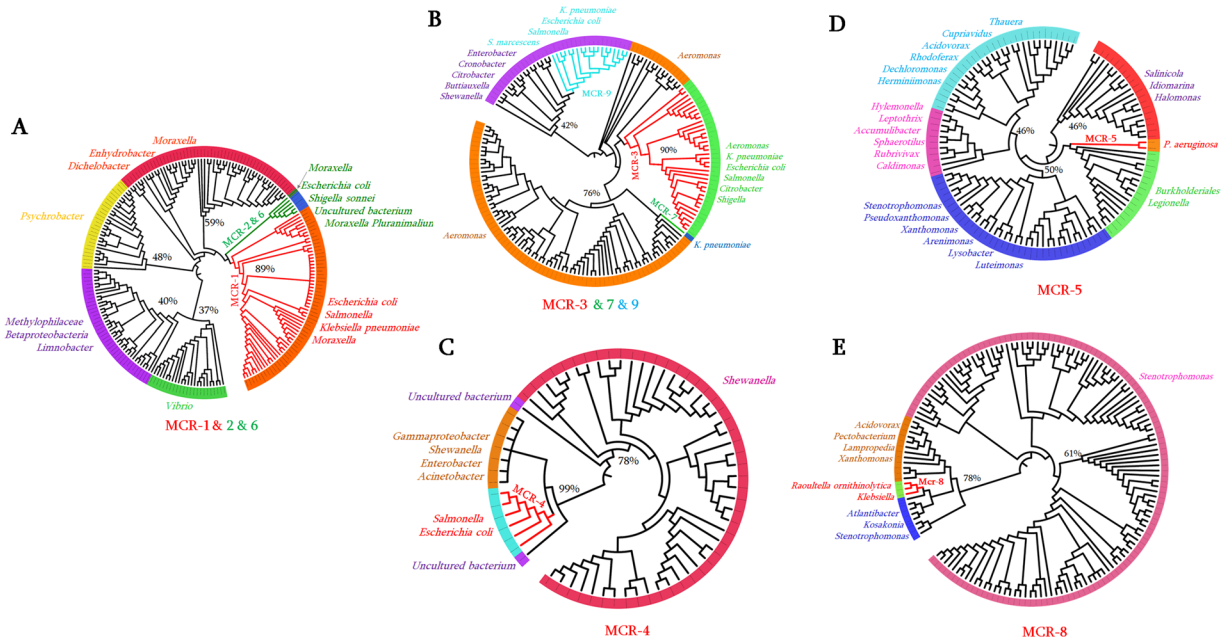


**Figure 1.** General representation of the distinctive homologous MCR sequences. (A) Phylogenetic tree inferred from PETs sequences from different bacterial species and highlighting MCR-1 to MCR-9 group using Archaeopteryx software. Protein identity of MCR-1 with all homologous sequence is indicated, lowest and highest % identity are indicated for each genus. Only representative bacterial genera are shown on this figure. (B) General features of known MCR variants including their phylogenetic relationship, nucleotide size and %CG content. Pairwise comparison of amino acid identity of MCR variants is also presented. (C) Rhizome analysis of MCR-1. This latter is here split into fragment of 50 aa and phylogenetic tree inferred for each fragment. Each presented line (blue and red) refer to the putative closest ancestor of each corresponding fragment. 8 out of the 11 fragments appeared from *M. pluranimalium* species.

of exaptation that define the use and/or evolution of one gene for different functions when useful in a specific ecosystem. Among resistance to antibiotics that have been recently described, those mediated by mobile colistin resistance (*mcr*) gene variants appear to be particularly important, since colistin has been considered as one of the reference antibiotic of the 21st century<sup>10</sup>. Indeed, until 2015, colistin resistance was extremely rare as it was not detected and was exclusively linked to chromosomal mutations in genes involved in lipid A decoration, i.e. *pmrA/B*, *phoP/Q*, *ccrA/B*, *lpxACD*, or *mgrB* genes that resulted in a modification of bacterial membranes by adding sugar (phosphoethanolamine (PETn) or 4-amino-4-deoxy-L-arabinose (L-ara4N)) to the lipid A moiety<sup>10,11</sup>. The colistin resistance mechanism also involves genes encoding phosphoethanolamine transferase (PET) and/or glycosyltransferase proteins that are essential for membrane phospholipid biosynthesis and appear extremely ubiquitous due to their presence in all areas of life including bacteria, archaea and eukaryotes (plants, arthropods)<sup>12–14</sup>. In 2015, a new transferable colistin resistance mechanism has been described, i.e. *mcr-1* gene, encoding for a PET<sup>15</sup> which nowadays, along with new variants of this gene, has been described worldwide in a wide variety of bacterial species and is believed to pose a major public health concern as it can be widely disseminated among pathogenic bacteria by horizontal transfer via transposons and recombinant plasmids<sup>16,17</sup>. In fact, these genes have been already reported in almost all common pathogens and in various sources including food, animal, human and environment (water, soil, etc.) samples<sup>13,17</sup>. Nowadays, nine *mcr* gene variants (from *mcr-1* to *mcr-9*) have been described, thus amplifying the concern and the interest granted to colistin resistance. Interestingly, most of these *mcr* genes have been discovered in bacteria from animal and/or environmental origin<sup>18–20</sup>. Indeed, the origin of *mcr-1* and *mcr-2* is believed to be from *Moraxella* species<sup>21</sup>, *mcr-3* from *Aeromonas veronii*<sup>22</sup>, *mcr-4* from *Shewanella*<sup>23</sup>, *mcr-8* recently described in *Raoultella ornithinolytica*<sup>24</sup> and *mcr-9* from *Salmonella enterica* serovar Typhimurium<sup>25</sup>. However, despite the large number of reported publications of these mobilized colistin resistance genes, the sources, origins and roles of these enzymes remain uncertain and not yet explored. Here, we conducted a massive bioinformatic analysis of 64'628 downloaded bacterial genomes to investigate, by a neutral approach of identifying sequences by similarity, the presence and the source of *mcr* gene variants. Such approach allows us to speculate that those mobile genes encoding colistin resistance were probably not only dedicated to this activity because of their ubiquitous presence. This finding demonstrates that mobile colistin resistance genes is probably an example of exaptation of this enzymatic activity from a selfish sequence that challenges the current classification by COGs (cluster of orthologous groups) that postulate an unequivocal link between phylogeny and function.

## Results

**Homologous sequences of mobile colistin resistance genes in available database.** As a starting point, the first MCR-1.1 protein described (NG\_050417) was used as query in a BlastP analysis against the NCBI database to fish out a total of 13'658 protein hits with aa identity ranged from 30% to 100% and alignment  $\geq 30\%$  (Fig. 1A). Results include all MCR-variants (from MCR-2 to MCR-9) with aa identity ranged from 30.82%



**Figure 2.** Phylogenetic subtrees highlighting the position of known MCR variants relative to the other species. (A) Phylogenetic subtree of MCR-1; MCR-2 and MCR-6; (B) Subtree of MCR-3, MCR-7, and MCR-9; (C) subtree of MCR-4; (D) subtree of MCR-5; and (E) subtree of MCR-8. The average percentage of aa identity with MCR variants are indicated.

with MCR-4.1 to 82.66% with MCR-6.1 (Fig. 1B). Interestingly, reference MCR variants exhibit almost the same sequence size (1'635-bp in average) but show a significant difference in their GC content ranged from 40.1% for MCR-4.1 to 56.17% for MCR-7.1 (Fig. 1B), likely suggesting different origins. As shown in Fig. 1A, inferred phylogenetic tree with all obtained sequences reveals a total of 494 bacterial genera (only representative genera are shown on this figure) highlighting a high diversity of PET proteins among bacteria. Moreover, among the 13'658 retrieved hits, 90,89% ( $n = 12'410$ ) exhibit aa identity less than 40% with MCR-1 and only 9.11% ( $n = 1'244$ ) had aa identity >40% with MCR-1. This finding suggests that these PET enzymes are not specifically hosted by some bacterial genera but are present in a wide range of bacterial species, demonstrating that these latter are extremely ubiquitous in microorganisms.

**Rhizome analysis of MCR-1 sequence.** To investigate the origin of MCR-1 protein, this latter was split into 11 fragments (50 aa in size) and each fragment was blasted against the NCBI database to retrieve the 100 best BlastP hits. Then, phylogenetic tree analysis was assessed for each fragment to determine its origin. As shown in Fig. 1C, 8 out of the 11 MCR-1 fragments appear to be from *Moraxella* genus, especially *Moraxella pluranimalium*.

**Subtree analysis of each MCR variants.** As seen on Fig. 1A, the phylogenetic tree analysis shows that the 9 MCR variants appear in different branches of the tree. So, subtree containing each MCR variant was retrieved for further detail analysis:

**MCR-1, 2, and 6.** These three MCR sequences share more than 80% of aa identity (Fig. 1B) and therefore appear clustered together on the inferred phylogenetic tree (Fig. 1A). Therefore, as shown in Fig. 2A, the phylogenetic subtree of MCR-1, 2 and 6 reveals that the putative progenitors of these MCR variants would be *Moraxella*, *Enhydrobacter*, *Dichelobacter*, *Psychrobacter*, *Methylophilaceae*, *Limnobacter*, and *Vibrio*. The aa identity with respect to MCR-1 ranged from the lowest 37% with sequences from *Vibrio* to the highest 59% with those from *Moraxella*, *Enhydrobacter* and *Dichelobacter*. It is interesting to note that all these bacteria originate from water, the environment and/or the soil (Supplementary Table 1).

**MCR-3, 7, and 9.** The three sequences share between 63.04 and 69.74% of aa identity despite the significant difference of their %GC content (Fig. 1B). The retrieved phylogenetic subtree containing the three variants reveals (Fig. 2B) distinct clades of sequences where the most closely related PET sequences were from *Aeromonas* species, well described as water source bacteria. MCR-3 is identified in different *Enterobacteriaceae* species and in *Aeromonas* (Fig. 2B). Interestingly, MCR-7 appears on a distinct clade from this tree and shares in average 76% aa identity with *Aeromonas* sequences. MCR-9 sequence appears to be present in *Enterobacteriaceae* species from diverse bacterial genera, including *Serratia*, *Escherichia*, *Salmonella*, and *Klebsiella* (Fig. 2B). As shown in this figure, MCR-9 shares in average 42% aa identity with sequences from *Cronobacter*, *Enterobacter*, *Shewanella* and *Buttiauxella*; these latter being described with isolation source of water, soil, environment and animals (Supplementary Table 2).

**MCR-4.** This MCR variant exhibits the lowest %GC content (40.1%GC) and shares only 44.14% of aa identity with the most closely related MCR sequence, i.e. MCR-7 (Fig. 1B). As shown in Fig. 2C, although it has been reported in *Enterobacteriaceae* such as *E. coli* and *Salmonella*, MCR-4 sequence clearly originates from *Shewanella* species with sequences sharing on average 78% of aa identity.

**MCR-5.** Homologous sequences to MCR-5 appear extremely diverse and present in various bacterial genera (Fig. 2D; Supplementary Table 1). Interestingly, as *Pseudomonas* genus where MCR-5 has been already reported<sup>26</sup>, all bacteria from this subtree (Fig. 2D) have been originating from water, soil and environment (Supplementary Table 1). The most closely related sequences to MCR-5 from this tree were identified in *Legionella* and *Burkholderiales*.

**MCR-8.** MCR-8 is highly distant to the other variants. This latter has so far only been reported in *K. pneumoniae* and *R. ornithinolytica*; and appears to be closely related to PET sequences from *Atlantibacter*, *Kosakonia*, *Xanthomonas*, *Lampromedia*, *Pectobacterium*, *Acidovorax*, and *Stenotrophomonas* (Fig. 2E). *Stenotrophomonas* exhibits a large number of homologous MCR-8 sequences, thus suggesting as the origin of this MCR variant.

Thus, based on these results, which show different bacterial origins and hosts of these MCR sequences, we have investigated the nine MCR variants in all available sequenced bacterial genomes with clinical interest and those from water and environment sources in order to identify the putative reservoirs of these enzymes.

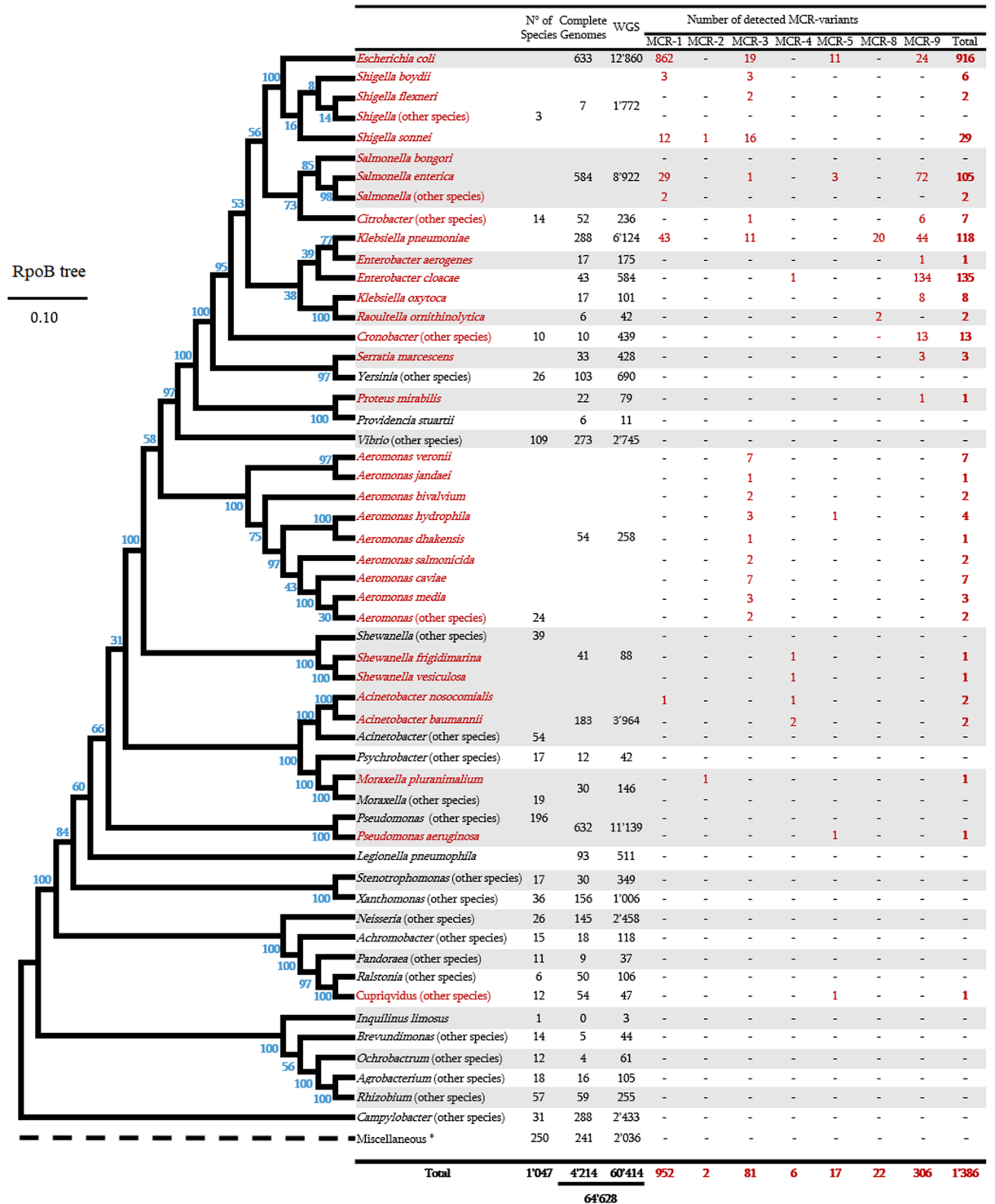
**MCR variants in sequenced bacterial genomes.** A total of 64'628 Gram-negative bacterial genomes were downloaded from the NCBI RefSeq database and include 60 bacterial genera constituted by 1'047 bacterial species (Fig. 3). As presented in this figure, a RpoB-based phylogenetic tree was inferred to establish their phylogenetic relationship. Only 4'214 out of the 64'628 genomes (6.52%) were complete genomes and the remaining genomes were a Whole-genome sequences (WGS). BlastP analysis of the nine MCR sequence variants against all downloaded data set provided significant positive results for MCR-1.1, MCR-2.1, MCR-3.1, MCR-4.1, MCR-5.1, MCR-8.1 and MCR-9.1 with aa identity values ranging from 91 to 100% and alignment  $\geq 98\%$ . A total of 1'386 MCR hits were identified in 15 out of the 60 genera analysed. As shown in Fig. 3, 952 MCR-1 hits were identified, including 862 hits in *E. coli* genomes, 43 hits in *K. pneumoniae*, 31 hits in *Salmonella* spp., 15 hits in *Shigella* spp. and one hit in *Acinetobacter baumannii*. MCR-2 was detected in only two genomes (*M. pluranimalium* and *S. sonnei*) (Fig. 3). MCR-3 was detected in diverse bacterial genera and was interestingly detected in all the *Aeromonas* species analyzed and almost in all *Shigella* species. As found in our first analysis (Fig. 2B), *Aeromonas* species appears clearly as the bacterial reservoir of MCR-3 variant (Fig. 3). Six MCR-4 hits were detected in our analysis, including 1 hit in *Enterobacter cloacae*, 2 hits in *Shewanella* spp. and 3 hits in *A. baumannii* and *A. nosocomialis*. MCR-5 was identified in only five species, including *E. coli* (n = 11 hits), *S. enterica* (n = 3 hits), *A. salmonicida* (n = 1 hit), *Cupriavidus* spp. (n = 1 hit) and *P. aeruginosa* (n = 1 hit). MCR-8 was identified in only *K. pneumoniae* (n = 20 hits) and its related species i.e. *R. ornithinolytica* (n = 2 hits), previously classified in *Klebsiella* genus. It is interesting to note that the newly described MCR-9 from *S. enterica* serotype *Typhimurium* strain isolated from a patient in Washington in 2010, was the most distributed sequence among bacterial species and was identified in 309 genomes (aa identity  $\geq 99.81\%$  and 100% of alignment) representing 10 out of the 14 MCR-positive genera (Fig. 3). Common *Enterobacteriaceae* species appear as the reservoir of this gene, especially *E. cloacae*, *S. enterica*, *K. pneumoniae*, and *E. coli*, where 134, 72, 44, and 24 hits were identified respectively. Moreover, we observed for the first time the presence of MCR-9 variant in *Serratia marcescens* and *Proteus mirabilis* genomes, those being naturally resistant to colistin.

Except MCR-1-producing *E. coli* strains that have been well described<sup>27</sup>, detailed metadata of all MCR-producer strains including strain names, aa identity with MCR variants, Bioproject number, collection dates, geographical location, isolation sources, and host organisms are given in Supplementary Table 2.

Apart from these significant occurrences (aa identity  $\geq 90\%$  and alignment  $\geq 98\%$ ) with reference MCR variants (from MCR-1 to MCR-9), we searched for new MCR candidate sequences using threshold values of  $50\% \leq$  aa identity  $\geq 90\%$  and alignment  $\geq 90\%$ . According to that, 5'265 additional hits were identified and distributed among 25 genera including among others *Escherichia*, *Aeromonas*, *Vibrio*, *Stenotrophomonas*, *Moraxella*, *Klebsiella*, *Salmonella*, *Shewanella*, *Shigella*, *Enterobacter*, *Raoultella*, *Serratia*, *Citrobacter*, *Xanthomonas*, *Acinetobacter* and *Proteus* (Supplementary Fig. 1). Interestingly, 1'244 out of the 5'265 hits exhibit aa identity between 75% and 90% with MCR variants, those being putative new MCR variants (Supplementary Fig. 1B). The *E. coli* species appears in this analysis as the main reservoir of these genes with a total of 3'090 MCR sequences identified (Supplementary Fig. 1). Surprisingly, a huge number of MCR-9-like sequences were identified with a total of 3'056 hits. Moreover, most of these bacterial genera including, *Stenotrophomonas*, *Vibrio*, *Aeromonas*, *Shewanella*, *Moraxella*, *Buttiauxella*, and *Salmonella* are bacteria of environmental sources.

As shown in Supplementary Fig. 2, all identified MCR hits from analysed genomes i.e. the 1'386 significant hits (aa identity  $\geq 90\%$ ) and the 5'265 additional hits ( $50\% \leq$  aa identity  $\geq 90\%$ ) were pooled together to infer a phylogenetic tree. We noticed that new putative MCR sequences (in blue) significantly clustered with the significant described MCR variants (in red). This suggests the presence of real MCR candidates that may soon emerge in common pathogenic bacteria. Moreover, our analyses reveal that all the analysed 64'628 genomes have constitutive PET sequences that exhibited an identity between 30% and 50% with MCR variants (data not shown).

**Genomic environment of *mcr* variant genes in bacterial genomes.** As previously described, mobilized colistin resistance genes are known to be transmitted between bacteria through mobile genetic elements (transposons and plasmids). Except *mcr-1* gene that has been well described in the literature regarding to its transmission mechanism<sup>15</sup>, we investigated, from downloaded genomes, the genetic environment of the other *mcr* genes to figure out their mode of transmissions.



**Figure 3.** All bacterial genomes analysed here and the RpoB-based phylogenetic tree of all investigated bacterial species. Bacterial species identified as MCR-producers and the number of significant MCR hits are highlighted in red. The number of species and genomes analyzed (complete and Whole genome sequences) are also indicated. \*, refers to rare bacterial species from the environment (list of the 27 genomes are given in Supplementary Table 1).

As shown in Supplementary Fig. 3A, while *mcr-2* exhibits a different genetic structure in the two genomes identified, *mcr-3* gene appears in a same sequence composed by it-self and a *dgkA* gene encoding for diacylglycerol kinase protein in four different bacterial species (Supplementary Fig. 3B) and was associated with an insertion sequence (IS) element upstream or/and downstream to *mcr-3*. Genome analysis reveals that this genetic structure of *mcr-3* in *K. pneumoniae* was identical in the 11 genomes positive for *mcr-3*. For *Aeromonas*, the

*mcr-3*-containing transposon was also identical in 26 out of the 28 genomes positive for *mcr-3*. The 21 genomes of *Shigella* also exhibited the same *mcr-3*-containing transposon. Interestingly, in the 19 *E. coli* genomes, only two of them exhibited the structure presented in this Supplementary Fig. 3B.

The *mcr-5*-containing transposon presented in Supplementary Fig. 3C appears in the same structure in the 11 *E. coli* genomes and the 3 *Salmonella* genomes. We also observed that this *mcr-5* gene is associated in this mobile genetic element with transposase and IS elements IS5/Tn3 in the different genomes. This genetic organization suggests sequence acquisition by sequence recombination via these transposases and IS elements in these genomes.

Among the 20 positive genomes of *K. pneumoniae* and 2 of *R. ornithinolytica* for *mcr-8* gene, the Tn5 transposon (Supplementary Fig. 3D) had the same structure in 13 out of the 22 genomes. This transposon located on conjugative plasmid in both species was highly similar and bordered in both sides by transposase *tnpA*/IS903 (Supplementary Fig. 3D). Interestingly, *mcr-8* gene appeared associated with *copR*, *baeS* and *dggA* genes encoding respectively for a transcriptional activator protein CopR, an integral membrane sensor signal transduction histidine kinase and a diacylglycerol kinase. Moreover, we identified two glycosyl transferases (GT) upstream and downstream of the *mcr-8* gene and two *ampC* serine- $\beta$ -lactamase genes. This genetic organization also suggests sequence acquisition of the transposon Tn903 by horizontal transfer from exogenous origin.

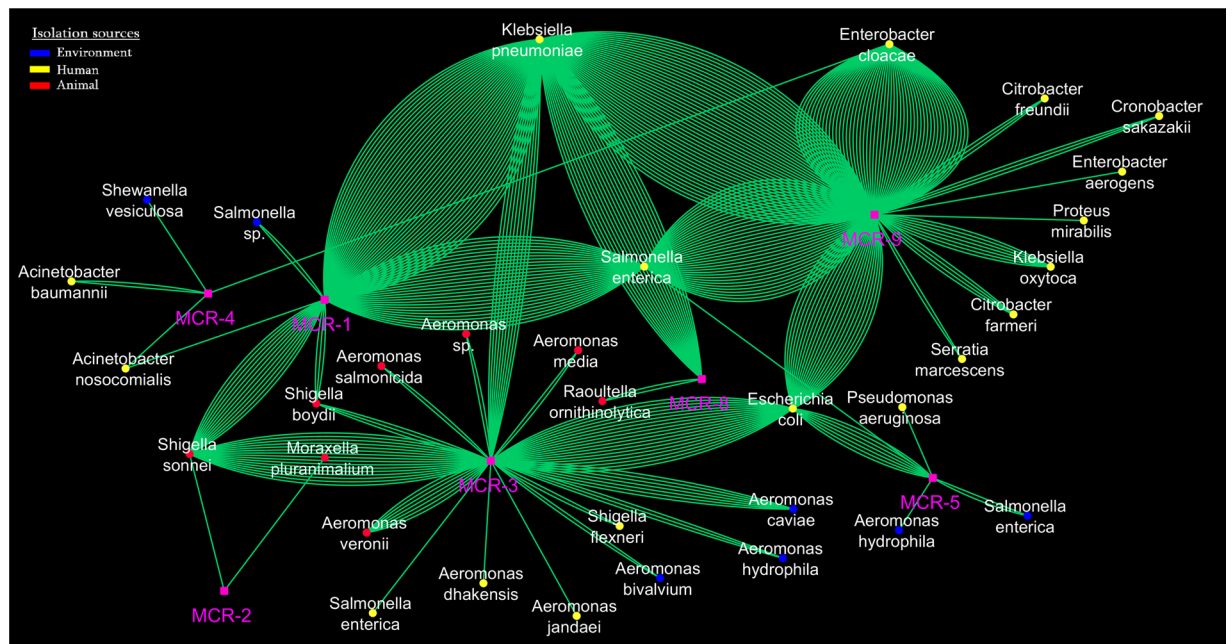
Finally, the transposon containing *mcr-9* shown in Supplementary Fig. 4 has the same structure in the nine genomes analysed, except in *P. mirabilis*. This transposon was highly similar and bordered in both sides by a transposase IS5/IS6. We identified a *wbuC* family gene encoding for cupin fold metallo-protein, which was highly conserved in this transposon. Moreover, a conserved nickel/cobalt operon (i.e. *rcnA*, *rcnR*, *pcoE* and *cusS* genes), which plays an important role in copper tolerance under anaerobic growth and nickel homeostasis in bacteria was identified in almost all analysed transposons (Supplementary Fig. 4). Some specific sequences including transcription factors (e.g. Ni(II)/Co(II)-binding transcriptional repressor RcnR, response regulator, transcriptional regulator ArdK, and Sensor kinase CusS), antimicrobial resistance genes (phosphotransferases (*aph(3'')*)-Ib, *aph(6)*-Id), copper-binding protein PcoE and zinc metalloprotease Zinc M) were identified in these transposons (Supplementary Fig. 4). This genetic organization suggests also sequence acquisition by transposition. Interestingly, *mcr-9* gene is strongly associated with *wbuC* gene encoding for Cupin fold metallo-protein located upstream in the different transposon of different bacterial species. This may suggest an essential role of this *wbuC* gene for the activity of MCR-9 enzyme.

**MCR-4 variant in *Acinetobacter* species.** So far, MCR-4 has only been described for this bacterial species in a genome of *A. baumannii*. Here, we identified three genomes of *Acinetobacter* (two *A. baumannii* and one *A. nosocomialis*) containing *mcr-4* gene variant. The two *A. baumannii* strains (AB18PR0652 and MRSN15313) were isolated respectively from feces from China in May 2018 and from a cerebrospinal fluid from Brazil (Salvado) dating from 2008, while the latter genome has not been described or analyzed in the literature (Supplementary Table 2). In both isolates, the *mcr-4* gene was located on plasmid pAB18PR065 and pAB-MCR4.3, respectively. As shown in Supplementary Fig. 5A, these two plasmids were compared to that identified in *A. nosocomialis* isolate T228 and appear very similar. Moreover, the six *mcr-4*-containing transposons identified here were compared (Supplementary Fig. 5B). As shown in this figure, these transposons were similar only between *Acinetobacter* species and were integrated into these plasmids by splitting an HNH endonuclease encoding gene (Supplementary Fig. 5B). In addition, a recombinase and two toxin/antitoxin systems were identified in *Acinetobacter* and *Shewanella* on the *mcr-4*-harboring-transposon.

**Metadata analysis of all MCR positive isolates in analysed genomes.** The presented metadata in the Supplementary Table 2 were subjected to cytoscape analysis in order to highlight links between isolates, MCR variants and isolation sources. So, as shown in Fig. 4, according to the isolation sources, almost all MCR variants were detected in bacterial species isolated from the three investigated sources (i.e. Human, Animal, and Environment). Three bacterial species including *K. pneumoniae*, *E. cloacae*, and *S. enterica*, all isolated from Human, appeared as the most predominant bacterial species harboring MCR enzymes. MCR-3 was most detected variant in different species and was also mostly detected in isolates from animals. Interestingly, only MCR-9 variant was detected in bacterial strains isolated exclusively from human and these isolates were almost all *Enterobacteriaceae* species (Fig. 4). These findings suggest that MCR-9 variant can become soon the most concern in human medicine regarding the colistin resistance.

## Discussion

Colistin resistance has become an increasing concern for human medicine since the emergence and spread of mobile colistin resistance genes<sup>15,17</sup>. The presence of these genes worldwide in all ecosystems including animals, environment (soil, water, plant) and humans raises the question of their sources, origins and roles in a global context of reservoir of antibiotic resistance genes. We show here how a large-scale data analysis approach, by studying more than 60'000 genomes, can help us to understand the bacterial reservoir and source of a gene family. We figured out that almost all mobile colistin resistance genes (from MCR-1 to MCR-9) described originated from environmental bacteria, especially from water sources. Indeed, as a proof of concept, the rhizome analysis performed reveals the origin of MCR-1 from *M. pluranimalium*, as previously suggested by another approach<sup>28</sup>. Moreover, our findings reveal that PET genes are extremely ubiquitous in bacteria since detected in 1'047 species analyzed here. Our massive genome analysis approach allowed us to identify the source and to detect the presence of these MCR-like sequences from available bacterial genomes. This was the case for MCR-3 and MCR-7 originating from *Aeromonas* spp.. Interestingly, among these MCR-3-like sequences, one of them identified from *Salmonella* and *Buttiauxella* spp. has recently been published as the new MCR-9 variant<sup>25</sup>. We identified *Shewanella* spp. as progenitor of MCR-4 sequences, that have been detected in *Acinetobacter* spp., also recently reported<sup>29</sup>. MCR-8



**Figure 4.** Network analysis of metadata of MCR variant producer isolates (excluding *E. coli mcr-1*) highlighting the link between each MCR variant and the different bacterial species and their isolation sources such as Human, Animal, and the Environment.

newly reported in *Raoultella* spp.<sup>24</sup> was also identified and appears to be present in *Stenotrophomonas* spp.. As a confirmation of this hypothesis, this latter has been recently described in *Stenotrophomonas* strain from sewage water<sup>30</sup>. Interestingly, whilst MCR-8 has only been reported in less than five studies to date, we identified 22 MCR-8 in already sequenced genomes of *Klebsiella* and *Raoultella* spp.. All these newly described MCR variants are real true witnesses of our results which predict putative 5'265 MCR-like sequences that could emerged soon in common pathogenic bacteria.

Interestingly, we observed in our study that most of the bacteria analysed are from water sources as highlighted in the Supplementary Table 1. This suggests that water environment appears to be the main reservoir and source of these MCR-like genes. Indeed, almost all MCR variants described so far have been reported from bacteria from water in different environments. MCR-1 was detected and identified directly from swine wastewater of purification facilities<sup>31</sup>, directly from water or from bacteria (e.g. *E. coli*, *E. cloacae*, and *P. mirabilis*) collected from urban surface water, rivers or lakes<sup>32,33</sup>. This gene was reported from wastewater treatment plants in China<sup>34</sup>, in South Africa<sup>35</sup>, in Italia<sup>36</sup>, in Germany<sup>37</sup> or from hospital sewage water in China<sup>38</sup>. This variant was also reported from various bacteria isolated from other water environments, such as seawater in Algeria<sup>39</sup>, from well water in rural China<sup>40</sup> or from freshwater of intensive aquaculture in China<sup>18</sup>. MCR-3 was reported from *Aeromonas* species isolated from river water<sup>32</sup>. It has recently been reported that the aquatic bacteria of the genus *Shewanella* are a reservoir/progenitor for *mcr-4* encoding gene<sup>41</sup>. MCR-5 and MCR-8 were described from the same *Stenotrophomonas* strain isolated from sewage water in China<sup>30</sup>. Overall, urban wastewater treatment plants are one of the most common reservoir and source of antibiotic resistance genes (ARGs) against various antibiotic classes ( $\beta$ -lactams, aminoglycosides, quinolones, tetracyclines, sulfonamides, macrolides)<sup>42</sup>. Of course, these ARGs, including MCR-like genes, were also reported from other sources including plants, soils<sup>43</sup> and various animal species<sup>43</sup>, most of them being in direct contact with water. Therefore, according to this, aquatic environment represents a great source and reservoir of MCR-carrier bacteria. Obviously, the global overuse of antibiotics in human medicine, animals and agriculture, has a great impact on the selection and the emergence of ARGs and multidrug resistant bacteria but this could not explain that more than 10% of bacterial genomes analyzed in our study contain a PET gene.

Indeed, our results revealed that PET genes, including *mcr*-like genes, are extremely ubiquitous in bacteria and in all areas of life<sup>12-14</sup>, thus raising the question of their real roles. It is unlikely that colistin pressure in the environment and animals, especially in water environment, may be the only cause of the existence of such an ubiquitous presence of these enzymes that could be mobilized within bacteria. This raises the question of the other roles and functions these enzymes may have in nature and in the biosphere. As a proof of concept of the existence of other roles for these enzymes, we found in our study that bacteria that are intrinsically resistant to colistin i.e. *Serratia* and *Proteus* also contain MCR variants. Because polymyxins share chemical similarities with cationic antimicrobial peptides (cAMPs) secreted by vertebrates and invertebrates, it is possible that decoration of lipid A of those bacteria with ethanolamine constitutes an immune defense against those metabolites. There is evidence that expression of PET in *Haemophilus ducreyi* is associated with polymyxin resistance as well as resistance to human defensins<sup>44</sup>. Although this hypothesis may explain another role of these enzymes against innate immunity from eukaryotic cells, this could not explain the huge prevalence of those enzymes in water environments.

Interestingly, the most common cell living in water sources are bacteriophages<sup>45,46</sup>. Predation of bacterial cells by bacteriophages is linked to the interaction between phages and O-antigens structures of Gram-negative bacteria<sup>47</sup>. One may speculate that decoration of lipid A with ethanolamine is a “camouflage” of the antigen receptor of bacteria against bacteriophages. This defense hypothesis against bacteriophage can be supported by a recent study reporting a disruption of *mcr-1* gene by a bacteriophage P1-like in an atypical enteropathogenic *E. coli* recovered from China<sup>48</sup>. Thus, we believe that PET genes in bacteria, by adding sugars to the bacterial cell membrane constitutes a very ubiquitous and old immune defense system that bacteria have developed over time, not only to resist to polymyxins, but also to resist to eukaryotic innate immune system and bacteriophage predation.

Our innovative approach, i.e. massive genome analysis, appears to be a powerful tool and a new concept that opens a new field of big data analysis and research; *mcr* gene variants are here only an example but this approach can be carried out to explore the source and the origin of any gene family. PET genes such as *mcr* genes are extremely ubiquitous in bacteria mainly from the environment, soil and especially from water environments. These latter could be mobilized by transposable elements and transmitted by horizontal transfer to bacteria as reported in the literature since the emergence of *mcr-1* in 2015. We speculate that these bacterial PET enzymes may play other major roles like defense system in the biosphere against bacteriophage in water environment but also against cAMP secreted by vertebrates (e.g. human) and invertebrates. Thus, this should be further explored in the future because this may have a great impact, not only for human health, but also as a one health perspective.

## Methods

**Looking for homologous MCR-1 sequences from the NCBI database.** Reference sequence of MCR-1 (AKF16168.1) was used as query for BlastP analysis against the NCBI database using as threshold e-value  $10e-5$ . All homologous sequence with identity  $\geq 30\%$  and alignment  $\geq 30\%$  were kept for further analysis. Protein alignments were performed using Mafft tool (<https://mafft.cbrc.jp/alignment/software/>) and phylogenetic trees were inferred using neighbor-joining method in FastTree program (<http://www.microbesonline.org/fasttree/>) and visualized with FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>), Archaeopteryx 0.9920 (<http://en.bio-soft.net/tree/ATV.html>), and iTOL, an interactive online program (<https://itol.embl.de>). A rhizome analysis<sup>49</sup> was carried out for MCR-1 to investigate putative ancestral recombination events of bacterial sequences leading to creation of chimerical gene. For this purpose, protein sequence of MCR-1 (AKF16168.1) was split into 10 fragments of 50 amino acids. Each fragment was blasted against the NCBI database using as threshold e-value  $10e-5$  to select the 100 best homologous sequences according to the aa identity  $\geq 50\%$  and coverage  $\geq 80\%$ . From these sequences, phylogenetic tree with each fragment was constructed to infer the closest ancestor.

**Looking for all MCR variants in bacterial genomes.** Whole genome sequences of the most common Gram-negative bacteria were retrieved from RefSeq data: NCBI Reference Sequence Database (<ftp://ftp.ncbi.nlm.nih.gov/genomes>). This access to the repository is made through version 0.2.7 of the NCBI genome download scripts, which allowed the download of the entire genome assembly, including all required data for each genome, genomic sequences, annotated proteins and assembly structure reports. Reference protein sequence of each MCR variant was also retrieved from the NCBI database. To the best of our knowledges, today (July 10, 2019), nine MCR variants have been described including MCR-1, MCR -2, MCR -3, MCR -4, MCR -5, MCR -6, MCR -7, MCR-8 and MCR -9 with respectively 18, 2, 29, 6, 3, 1, 1,2 and 1 subvariants. BlastP analysis was conducted to investigate the presence of the nine MCR variants (i.e. MCR-1.1, MCR-2.1, MCR-3.1, MCR-4.1, MCR-5.1, MCR6.1, MCR-7.1, MCR-8.1 and MCR-9.1) in the large downloaded data. MCR hits were considered significant on the basis of 90% aa identity and 98% of alignment. Reference sequences of MCR variants were obtained from the Pathogen Detection database (<https://www.ncbi.nlm.nih.gov/pathogens/>).

**Analysis of genetic environment of *mcr* gene variants.** Genetic environments of *mcr* genes were investigated in the selected genomes using the Easyfig version x2.2.3 program (<http://mjsull.github.io/Easyfig/>) and CGView program (<http://wishart.biology.ualberta.ca/cgview/>). An annotation using Prokka (<http://www.vicbioinformatics.com/software/prokka.shtml>) and BlastP were made for genetic environment verification.

**Metadata and network analysis.** Metadata of each bacterial genome containing significant hits of *mcr* gene variants ( $\geq 90\%$  identity,  $\geq 98\%$  alignment) including bacterial strain name, bioproject number, collection date, geographical location, isolation source and host organisms, were collected from PATRIC database (<https://www.patricbrc.org/>) and detailed in Supplementary Table 2. To investigate putative links between bacterial strain harboring *mcr* gene variants, metadata informations are subjected to networking analysis using the cytoscape software v.3.7.

## Data availability

The datasets generated and/or analysed during the current study are available in the Refseq repository, [<https://www.ncbi.nlm.nih.gov/refseq/>].

Received: 11 February 2020; Accepted: 12 March 2020;

Published online: 06 April 2020

## References

1. Dcosta, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
2. Bhullar, K. *et al.* Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS One* **7**, 1–11 (2012).
3. Palareti, G. *et al.* Comparison between different D-Dimer cutoff values to assess the individual risk of recurrent venous thromboembolism: Analysis of results obtained in the DULCIS study. *Int. J. Lab. Hematol.* **38**, 42–49 (2016).
4. Rolain, J. M. & Cornaglia, G. Carbapenemases in Enterobacteriaceae: The magnitude of a worldwide concern. *Clin. Microbiol. Infect.* **20**, 819–820 (2014).
5. Bush, K. *et al.* Tackling antibiotic resistance. *Nat. Rev. Microbiol.* **9**, 894–6 (2011).



6. Burnham, C. A. D., Leeds, J., Nordmann, P., O'Grady, J. & Patel, J. Diagnosing antimicrobial resistance. *Nat. Rev. Microbiol.* **15**, 697–703 (2017).
7. Venter, H. Reversing Resistance to Counter Antimicrobial Resistance in the World Health Organisation's Critical Priority of Most Dangerous Pathogens. *Biosci. Rep.* BSR20180474, <https://doi.org/10.1042/BSR20180474> (2019).
8. Diene, S. M. *et al.* Human metallo- $\beta$ -lactamase enzymes degrade penicillin. *Sci. Rep.* **9**, 12173 (2019).
9. Pettinati, I., Brem, J., Lee, S. Y., McHugh, P. J. & Schofield, C. J. The Chemical Biology of Human Metallo- $\beta$ -Lactamase Fold Proteins. *Trends Biochem. Sci.* **41**, 338–355 (2016).
10. Yu, Z., Qin, W., Lin, J., Fang, S. & Qiu, J. Antibacterial mechanisms of polymyxin and bacterial resistance. *Biomed Res. Int.* **2015** (2015).
11. Bakour, S. *et al.* Emergence of Colistin- and Carbapenem-Resistant *Acinetobacter baumannii* ST2 Clinical Isolate in Algeria: First Case Report. *Microb. Drug Resist.* **21**, 279–85 (2015).
12. Nishimura, Y. & Eguchi, T. Biosynthesis of archaeal membrane lipids: Digeranylgeranyl glycerophospholipid reductase of the thermoacidophilic archaeon *Thermoplasma acidophilum*. *J. Biochem.* **139**, 1073–1081 (2006).
13. Baron, S., Hadjadj, L., Rolain, J. M. & Olaitan, A. O. Molecular mechanisms of polymyxin resistance: Knowns and unknowns. *Int. J. Antimicrob. Agents* **48**, 583–591 (2016).
14. Mambelli, L. I. *et al.* Phosphoethanolamine induces caspase-independent cell death by reducing the expression of C-RAF and inhibits tumor growth in human melanoma model. *Biomed. Pharmacother.* **103**, 18–28 (2018).
15. Liu, Y.-Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *Lancet. Infect. Dis.* **16**, 161–8 (2016).
16. Erik, S. & McGann Patrick, C. M. The Birth and Demise of the ISAp11-mcr-1-ISAp11 Composite Transposon: the Vehicle for Transferable Colistin Resistance. *MBio* **9**, 1–16 (2018).
17. Hadjadj, L. *et al.* Study of mcr-1 gene-mediated colistin resistance in enterobacteriaceae isolated from humans and animals in different countries. *Genes (Basel)* **8**, 1–16 (2017).
18. Cabello, F. C. & Godfrey, H. P. Aquaculture, Exaptation, and the Origin of mcr -Positive Colistin Resistance. *Antimicrob. Agents Chemother.* **62**, 1–2 (2018).
19. Telke, A. A. & Rolain, J.-M. Functional genomics to discover antibiotic resistance genes: The paradigm of resistance to colistin mediated by ethanolamine phosphotransferase in *Shewanella* algae MARS 14. *Int. J. Antimicrob. Agents* **46**, 648–652 (2015).
20. Teale, C. *et al.* mcr-1 and mcr-2 (mcr-6.1) variant genes identified in *Moraxella* species isolated from pigs in Great Britain from 2014 to 2015. *J. Antimicrob. Chemother.* **72**, 2745–2749 (2017).
21. Kieffer, N., Nordmann, P. & Poirel, L. *Moraxella* Species as Potential Sources of MCR-Like Polymyxin Resistance Determinants. **61**, 1–7 (2017).
22. Ling, Z. *et al.* Chromosome-mediated mcr-3 variants in *aeromonas veronii* from chicken meat. *Antimicrob. Agents Chemother.* **61** (2017).
23. Huang, M. *et al.* Action and mechanism of the colistin resistance enzyme MCR-4. *Commun. Biol.* **2** (2019).
24. Wang, X. *et al.* Emergence of Colistin Resistance Gene mcr-8 and Its Variant in *Raoultella ornithinolytica*. *Front. Microbiol.* **10**, 228 (2019).
25. Carroll, L. M. *et al.* Identification of novel mobilized colistin resistance gene mcr-9 in a multidrug-resistant, colistin-susceptible *enterica* serotype typhimurium isolate. *MBio* **10**, 1–6 (2019).
26. Erik Snedrud, R., Maybank, Y. I., Kwak, A. R., Jones, M. K., Hinkle, P. M. Chromosomally Encoded mcr-5 in Colistin-Nonsusceptible *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **4–8** (2018).
27. Gao, R. *et al.* Dissemination and Mechanism for the MCR-1 Colistin Resistance. 1–19, <https://doi.org/10.1371/journal.ppat.1005957> (2016).
28. Nicolas, K., Patrice, N. & Laurent, P. *Moraxella* Species as Potential Sources of MCR-Like Polymyxin Resistance Determinants. *Antimicrob. Agents Chemother.* **61**, e00129–17 (2017).
29. Ma, F. *et al.* Identification of a novel plasmid carries mcr-4.3 in *Acinetobacter baumannii* from China. *Antimicrob. Agents Chemother.* **86** (2019).
30. Li, J. *et al.* Co-Occurrence of Colistin and Meropenem Resistance Determinants in a *Stenotrophomonas* Strain Isolated from Sewage Water. *Microb. Drug Resist.* **25**, mdr.2018.0418 (2019).
31. Yuan, Q.-B., Zhai, Y.-F., Mao, B.-Y., Schwarz, C. & Hu, N. Fates of antibiotic resistance genes in a distributed swine wastewater treatment plant. *Water Environ. Res.*, <https://doi.org/10.1002/wer.1125> (2019).
32. Tuo, H. *et al.* The Prevalence of Colistin Resistant Strains and Antibiotic Resistance Gene Profiles in Funan River, China. *Front. Microbiol.* **9**, 3094 (2018).
33. Alhaj Sulaiman, A. A. & Kassem, I. I. First report of the plasmid-borne colistin resistance gene (mcr-1) in *Proteus mirabilis* isolated from domestic and sewer waters in Syrian refugee camps. *Travel Med. Infect. Dis.* 101482, <https://doi.org/10.1016/j.tmaid.2019.101482> (2019).
34. Chen, K. *et al.* Widespread distribution of mcr-1-bearing bacteria in the ecosystem, 2015 to 2016. *Eurosurveillance* **22** (2017).
35. Igwaran, A., Iweriebor, B. & Okoh, A. Molecular Characterization and Antimicrobial Resistance Pattern of *Escherichia coli* Recovered from Wastewater Treatment Plants in Eastern Cape South Africa. *Int. J. Environ. Res. Public Health* **15**, 1237 (2018).
36. Caltagirone, M. *et al.* Occurrence of Extended Spectrum  $\beta$ -Lactamases, KPC-Type, and MCR-1.2-Producing Enterobacteriaceae from Wells, River Water, and Wastewater Treatment Plants in Oltrepò Pavese Area, Northern Italy. *Front. Microbiol.* **8**, 2232 (2017).
37. Hembach, N. *et al.* Occurrence of the mcr-1 Colistin Resistance Gene and other Clinically Relevant Antibiotic Resistance Genes in Microbial Populations at Different Municipal Wastewater Treatment Plants in Germany. **8**, 1–11 (2017).
38. Jin, L. *et al.* Emergence of mcr-1 and carbapenemase genes in hospital sewage water in Beijing, China. *J. Antimicrob. Chemother.* **73**, 84–87 (2018).
39. Drali, R. *et al.* Emergence of mcr - 1 plasmid-mediated colistin-resistant *Escherichia coli* isolates from seawater. *Sci. Total Environ.* **642**, 90–94 (2018).
40. Sun, P. *et al.* Occurrence of blaKPC-2, blaCTX-M, and mcr-1 in Enterobacteriaceae from Well Water in Rural China. *Antimicrob. Agents Chemother.* **61**, 1–5 (2017).
41. Zhang, H., Wei, W., Huang, M., Umar, Z. & Feng, Y. Definition of a Family of Nonmobile Colistin Resistance (NMCR-1) Determinants Suggests Aquatic Reservoirs for MCR-4. *Advanced Science*, <https://doi.org/10.1002/adv.201900038> (2019).
42. Pärnänen, K. M. M. *et al.* Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. *Sci. Adv.* **5**, 1–10 (2019).
43. Tyrrell, C., Burgess, C. M., Brennan, F. P. & Walsh, F. Antibiotic resistance in grass and soil. *Biochem. Soc. Trans.* **47**, 477–486 (2019).
44. Trombley, M. P. *et al.* Phosphoethanolamine transferase LptA in *haemophilus ducreyi* modifies lipid a and contributes to human defensin resistance *in vitro*. *PLoS One* **10**, 1–20 (2015).
45. Robert, A. & Edwards, F. R. Viral metagenomics. *Nat Rev Microbiol* **3**, 286–289 (2004).
46. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
47. Kulikov, E. E., Golomidova, A. K., Prokhorov, N. S., Ivanov, P. A. & Letarov, A. V. High-throughput LPS profiling as a tool for revealing of bacteriophage infection strategies. *Sci. Rep.* **9**, 1–10 (2019).
48. Bai, L. *et al.* A novel disrupted mcr-1 gene and a lysogenic phage P1-like sequence detected from a large conjugative plasmid, cultured from a human atypical enteropathogenic *Escherichia coli* (aEPEC) recovered in China. *J. Antimicrob. Chemother.* **72**, 1531–1533 (2017).
49. Merhej, V. & Raoult, D. Rhizome of life, catastrophes, sequence exchanges, gene creations, and giant viruses: how microbial genomics challenges Darwin. *Front. Cell. Infect. Microbiol.* **2**, 1–17 (2012).

## Acknowledgements

Financial support from the IHU Mediterranee-Infection, Marseille, France; CookieTrad for English corrections of the manuscript are gratefully acknowledged. This work was supported by the French Government under the “Investments for the Future” program managed by the National Agency for Research (ANR) and the IHU Mediterranee-Infection (reference 10-IAHU-03), Marseille, France.

## Author contributions

S.M.D., D.R., and J.-M.R. conceived and designed the study. M.B.K., S.M.D., S.A.B., T.R., R.R. and J.-M.R. analysed and interpreted data. M.B.K., S.M.D., J.-M.R. drafted the manuscript. S.M.D., S.A.B., R.R., D.R. and J.-M.R. made corrections and critical revisions of the manuscript. All of the authors read and approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-63167-5>.

**Correspondence** and requests for materials should be addressed to S.M.D. or J.-M.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2020