Genomics Proteomics Bioinformatics 20 (2022) 1168–1179

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

ORIGINAL RESEARCH

Genomic Shift in Population Dynamics of *mcr-1*-positive *Escherichia coli* in Human Carriage

Yingbo Shen [1,2,3,#], Rong Zhang [4,#], Dongyan Shao [1], Lu Yang [1], Jiayue Lu [4], Congcong Liu [4], Xueyang Wang [1], Junyao Jiang [1], Boxuan Wang [1], Congming Wu [1,2], Julian Parkhill [5], Yang Wang [1,2], Timothy R. Walsh [6,*], George F. Gao [3,7,*], Zhangqi Shen [1,2,*]

[1] *College of Veterinary Medicine, China Agricultural University, Beijing 100193, China*
[2] *Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510641, China*
[3] *CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China*
[4] *The Second Affiliated Hospital of Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310009, China*
[5] *Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK*
[6] *Department of Biology, University of Oxford, Oxford OX1 3SZ, UK*
[7] *Chinese Center for Disease Control and Prevention, Beijing 102206, China*

Received 29 November 2021; revised 7 November 2022; accepted 29 November 2022
Available online 5 December 2022

Handled by Feng Gao

KEYWORDS

Colistin;
mcr-1;
Escherichia coli;
Genomics;
Human

Abstract Emergence of the **colistin** resistance gene, ***mcr-1***, has attracted worldwide attention. Despite the prevalence of *mcr-1*-positive ***Escherichia coli*** (MCRPEC) strains in **human** carriage showing a significant decrease between 2016 and 2019, genetic differences in MCRPEC strains remain largely unknown. We therefore conducted a comparative **genomic** study on MCRPEC strains from fecal samples of healthy human subjects in 2016 and 2019. We identified three major differences in MCRPEC strains between these two time points. First, the insertion sequence IS*Apl1* was often deleted and the percentage of *mcr-1*-carrying IncI2 plasmids was increased in MCRPEC strains in 2019. Second, the antibiotic resistance genes (ARGs), *aac(3)-IVa* and $bla_{\text{CTX-M-1}}$, emerged and coexisted with *mcr-1* in 2019. Third, MCRPEC strains in 2019 contained more virulence genes, resulting in an increased proportion of extraintestinal pathogenic *E. coli* (ExPEC) strains (36.1%) in MCRPEC strains in 2019 compared to that in 2016 (10.5%), implying that these strains could occupy intestinal ecological niches by competing with other commensal bacteria. Our

* Corresponding authors.
E-mail: szq@cau.edu.cn (Shen Z), gaof@im.ac.cn (Gao GF), timothy.walsh@zoo.ox.ac.uk (Walsh TR).
# Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

https://doi.org/10.1016/j.gpb.2022.11.006

results suggest that despite the significant reduction in the prevalence of MCRPEC strains in humans from 2016 to 2019, MCRPEC exhibits increased resistance to other clinically important ARGs and contains more virulence genes, which may pose a potential public health threat.

Introduction

Antimicrobial resistance (AMR) has become a significant threat to human health. Due to the extensive emergence of multidrug-resistant (MDR) Gram-negative bacteria in humans since the 2000s, colistin, its nephrotoxicity and neurotoxicity notwithstanding [1], has been reintroduced clinically as a last-line and restricted antibiotic for treating severe infections caused by MDR bacteria. Paradoxically, colistin has also been used in food-producing animals in treatment, metaphylaxis, and as a feed additive for promoting growth [2]. The extensive use of colistin in animal production has led to selective pressure and enhanced the emergence and persistence of the mobile colistin resistance gene, *mcr-1* [3]. *mcr-1* has been reported in animals, animal-derived foods, environments, companion animals, wildlife, and humans in more than 60 countries across six continents [4]. Since the discovery of *mcr-1*, nine additional *mcr* genes, *mcr-2* to *mcr-10*, have been identified [5–7].

To prevent the further dissemination of *mcr* genes and mitigate the impact of *mcr-1*-positive *Escherichia coli* (MCRPEC) strains, China banned colistin as an animal feed additive in April 2017 [8]. Correspondingly, colistin resistance has significantly decreased in both animals and humans in China after this policy was implemented [9,10]. Hitherto, only one study from a single province in China has investigated the genomic epidemiology of MCRPEC with a few MCRPEC strains sequenced from humans (n = 75 in 2017 and n = 8 in 2018) after colistin ban [9]. In contrast, our previous study revealed that human *mcr-1* carriage in China decreased from 14.3% (644/4498) in 2016 to 6.3% (357/5657) in 2019 [10]. Herein, we used data from our previous studies to conduct a comprehensive comparative genomic study on MCRPEC strains collected from intestine in human subjects across 30 provinces/autonomous regions/municipalities in China in 2016 (n = 287) [11] and 17 provinces/autonomous regions in 2019 (n = 83) [10], respectively. The sequences from the former study [11] were randomly selected by criteria of 10 of each province/autonomous region/municipality (if available) to keep balanced distribution, while all recovered MCRPEC strains from the latter study [10] were sequenced due to largely decreased prevalence and most of *mcr-1*-positive samples without successful isolates. We investigated the genomic differences in MCRPEC strains at these two time points, thus providing evidence of their evolutionary adaptation and a shift in population dynamics.

Results

Strain collection

We retrieved 287 and 83 MCRPEC strains, respectively, from the fecal samples of healthy humans in 2016 [11] and 2019 [10]. All samples were derived from 30 provinces/autonomous regions/municipalities across China and from the same hospital with the same method in each year (Table S1). In 2016, the sequences were from all 30 provinces/autonomous regions/municipalities, whereas in 2019, only 83 MCRPEC strains were successfully cultured from the fecal samples in 17 provinces/autonomous regions. The median total genome size was 4.95 Mb [interquartile range (IQR): 4.8–5.11 Mb]; and the N50 of all assembled contigs was 127,102 bp (IQR: 103,000–169,765 bp) (Table S1).

Loss of IS*Apl1* in transposon Tn*6330*

mcr-1 was commonly adjacent to *pap2* and two copies of IS*Apl1* to form the floating cassette, IS*Apl1-mcr-1-pap2*-IS*Apl1*, named transposon Tn*6330*, which is likely to mediate the translocation of *mcr-1* [12]. The median size of the *mcr-1*-positive contigs in the MCRPEC strains from 2019 (60,160 bp; IQR: 32,489–63,441 bp) was approximately twice that from 2016 (32,798 bp; IQR: 24,354–60,956 bp) (P = 0.002, Mann–Whitney U test), suggesting more incomplete Tn*6330* and fewer IS*Apl1* in the *mcr-1*-positive contigs from the 2019 strains (**Figure 1**A; Table S1). We further detected these contigs with IS*Apl1*. The results showed that 49 (17.1%) MCRPEC strains contained the complete Tn*6330* in 2016, while only one MCRPEC strain contained the complete Tn*6330* in 2019. Additionally, a significantly greater proportion of MCRPEC strains were IS*Apl1*-negative in 2019 (74/83, 89.2%) than in 2016 (177/287, 61.7%; P < 0.0001) (Figure 1B; Table S2). Moreover, 59 (20.6%) and 8 (9.6%) MCRPEC strains contained only a single copy of IS*Apl1* upstream of *mcr-1* in 2016 and 2019, respectively; two MCRPEC strains harbored a single copy of IS*Apl1* downstream of *mcr-1* in 2016, and none in 2019.

Changes in *mcr-1*-positive plasmids

Overall, *mcr-1* was found in seven and six known plasmid types in MCRPEC strains collected from 2016 and 2019, respectively, with IncI2, IncX4, IncHI2, IncP, IncY, and IncFIB types in common (Figure S1A; Table S2). Among the seven known plasmid types detected, only the IncI2 type displyed a significant change, *i.e.*, higher proprotion in 2019 than in 2016 (n = 44, 53.0% in 2019 *vs.* n = 75, 26.1% in 2016; P < 0.001, χ^2 test) (Figure S1A; Table S2). We also found a group of *mcr-1*-carrying short contigs which were too short to determine their locations, termed "unknown" group. This group showed a significant decrease in MCRPEC strains in 2019 than in 2016 (n = 51, 17.8% in 2016 *vs.* n = 5, 6.0% in 2019; P = 0.014, χ^2 test) (Figure S1A; Table S2). Among the MCRPEC strains without IS*Apl1*, the predominant plasmid type changed from IncX4 (49.7%, 88/177) in 2016 to IncI2 (56.8%, 42/74) in 2019 (Figure S1B; Table S2).

Co-selection of *mcr-1*

MCRPEC strains in 2019 and in 2016 have a similar number and α-diversity of antibiotic resistance genes (ARGs) [median:

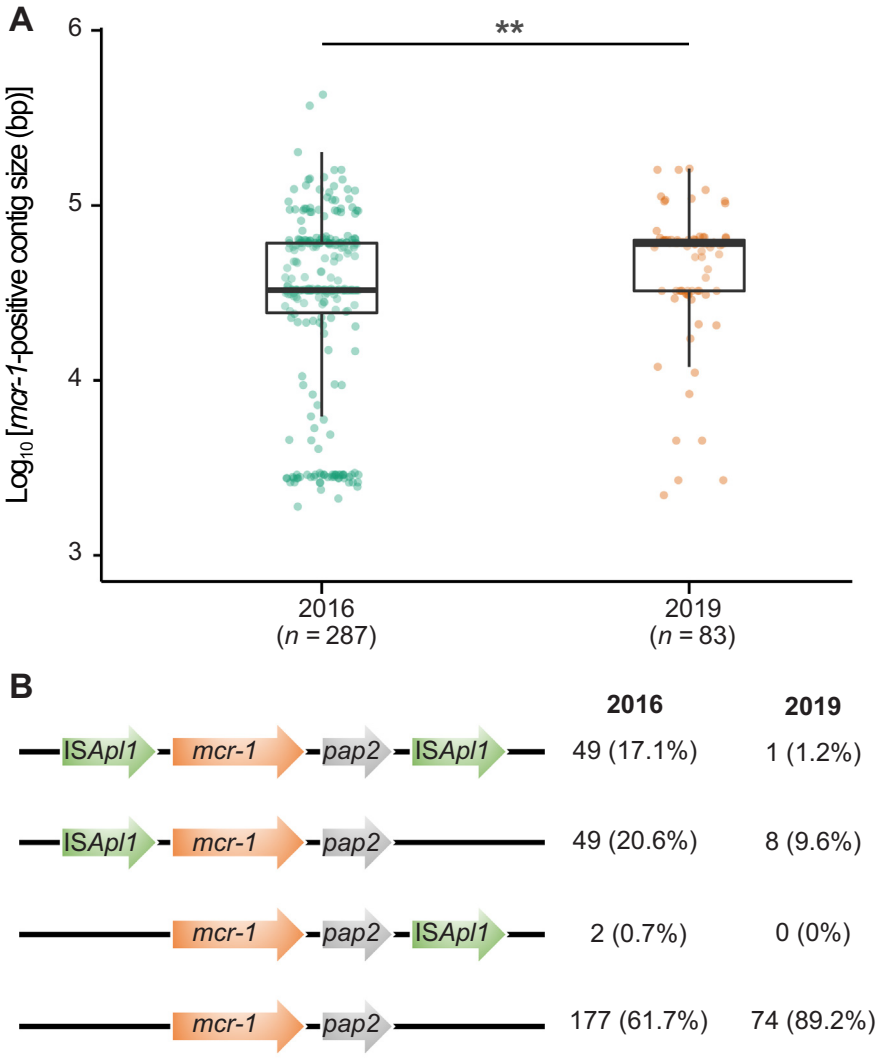

Figure 1 Distribution of transposon Tn*6330* among MCRPEC strains

A. Comparison of *mcr-1*-positive contig sizes between MCRPEC strains collected in 2016 and 2019. Data are shown as box plots; the horizontal box lines represent the first quartile, the median, and the third quartile. Whiskers denote the range of points within the first quartile − 1.5× the interquartile range and the third quartile + 1.5× the interquartile range. **, $P < 0.01$ (Mann–Whitney *U* test). **B.** Distribution of transposon Tn*6330* among MCRPEC strains collected in 2016 and 2019. MCRPEC, *mcr-1*-positive *Escherichia coli*.

16 (IQR: 13–19) in 2016 *vs.* 15 (IQR: 1–18.25) in 2019; $P = 0.355$, Mann–Whitney *U* test) (Figure S2A and B). However, the β-diversity of the ARGs in MCRPEC strains in 2019 demonstrated a different community structure than that in 2016 ($P = 0.001$; permutational multivariate analysis of variance, PERMANOVA) (Figure S2C). For ARGs, the aminoglycoside resistance genes, *aac(3)-IVa* (13.9% in 2016 *vs.* 25.3% in 2019; $P = 0.022$, χ^2 test) and *aph(4)-Ia* (10.5% in 2016 *vs.* 21.7% in 2019; $P = 0.0125$, χ^2 test), the β-lactam resistance gene, $bla_{\text{CTX-M-1}}$ (11.1% in 2016 *vs.* 36.1% in 2019; $P < 0.001$, χ^2 test), and the trimethoprim resistance gene, *dfrA5* (13.9% in 2016 *vs.* 25.3% in 2019; $P = 0.022$, χ^2 test) were significantly enriched in 2019. Conversely, the chloramphenicol resistance gene, *cmlA* (40.1% in 2016 *vs.* 20.5% in 2019; $P = 0.002$, χ^2 test), the fluoroquinolone resistance gene, *oqxAB* (45.3% in 2016 *vs.* 16.9% in 2019; $P < 0.001$, χ^2 test), and the tetracycline resistance gene, *tet*(M) (14.6% in 2016 *vs.* 4.8% in 2019, $P = 0.028$; χ^2 test) were significantly decreased in 2019 (Figure 2A; Table S3). Among these genes, $bla_{\text{CTX-M-1}}$ coexisted with *mcr-1* in three (1.0%) and seven (8.4%) *mcr-1*-positive contigs from 2016 and 2019, respectively ($P = 0.0015$, Fisher's exact test), nine of which belong to the IncI2 plasmid.

Changes in virulence genes and clusters of orthologous groups

The median number of virulence genes (VGs) in each MCRPEC strain was significantly higher in 2019 [174 (IQR: 146–209.5) in 2019 *vs.* 149 (IQR: 131–174) in 2016; $P < 0.001$, Mann–Whitney *U* test) (Figure S3A). Given that different pathogenic phenotypes (PPs) can consist of a largely discrepant number of VGs, we determined the number of PPs and found that MCRPEC strains had more PPs in 2019 than in 2016 [median: 30 (IQR: 27–34) in 2019 *vs.* 27 (IQR: 23–31) in 2016; $P < 0.001$, Mann–Whitney *U* test) (Figure S3B). The α-diversity (Shannon index) and β-diversity (Jaccard dis-

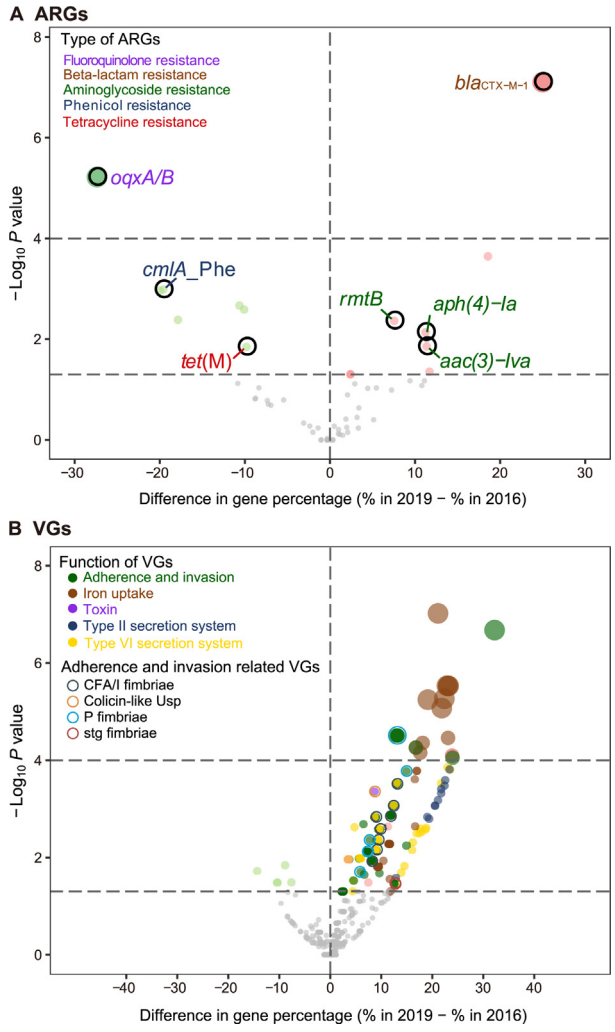

Figure 2 Volcano plot of ARGs and VGs among MCRPEC strains

A. Volcano plot of ARGs. **B.** Volcano plot of VGs. The X-axis represents difference in gene percentages between MCRPEC strains in 2016 and 2019 (2019 minus 2016). The Y-axis represents the $\log_{10}$-transformed *P* value ($-\text{Log}_{10}$ *P* value) of the percentage of each gene between two groups using the χ^2 test. Each dot represents a separate gene which is clustered in the same color as other genes with similar functions. The color would be influenced by the overlapping dot(s) with different color(s). The two horizontal dash lines denote *P* values as 0.0001 and 0.05, respectively. ARG, antibiotic resistance gene; VG, virulence gene.

tance) of the VGs in MCRPEC strains showed significantly higher abundances in 2019 than in 2016 ($P = 0.00011$, Mann–Whitney *U* test and $P = 0.001$, PERMANOVA, respectively) (Figure S3C and D). MCRPEC strains from 2019 possessed significantly higher percentages of VGs encoding adherence factors, including fimbriae (CFA/I, type I, Stg, P, and *E. coli* common pili), invasion factors (Tia and KpsT), iron-uptake factors (aerobactin siderophore, hemin uptake, iron-regulated element, iron/manganese transport, salmochelin siderophore, and yersiniabactin siderophore), and toxins (colicin-like Usp, type VI secretion system, and type II secretion system) (Figure 2B; Table S4). Furthermore, the median number of clusters of orthologous groups (COGs) of MCRPEC proteins was generally higher in 2019 than in 2016 as per the eggNOG database [13] (**Figure 3**; Table S5). Interestingly, the number of proteins ascribed to M (cell wall/membrane/envelope biogenesis), O (post-translational modification, protein turnover, chaperones), D (cell cycle control, cell division, chromosome partitioning), W (extracellular structures), K (transcription), P (inorganic ion transport and metabolism), C (energy production and conversion), F (nucleotide transport and metabolism), H (coenzyme transport and metabolism), and Q (secondary metabolites biosynthesis, transport, and catabolism) were significantly elevated in MCRPEC strains from 2019 (Figure 3; Table S5).

Genome-wide association studies

We further performed gene-based and single nucleotide polymorphism (SNP)-based genome-wide association studies (GWAS) to identify the changes in addition to the ARGs and VGs of human intestinal MCRPEC strains between 2016 and 2019. In the gene-based analysis, 126 genes showed significant differences in gene percentage ($P < 0.01$; Benjamini–Hochberg method) (**Figure 4**A; Table S6). Overall, the percentages of genes conferring heavy metal resistance and mobile elements, such as cation-efflux genes (*cusA*, *cusF*, *cusC*, *cusB*, and *cusS*), copper resistance genes (*copA*, *copB*, *copC*, and *copE*), silver resistance genes (*silE* and *silP*), and ISA*pl1*, were significantly reduced in 2019. Moreover, genes composing type IV secretion system, ferric aerobactin siderophore genes (*iutA*, *iucB*, *iucC*, and *iucD*), *L*-idonate transport genes (*idnO* and *idnT*), type I fimbriae gene (*fimA*), F1C fimbriae gene (*focC*), group II capsular polysaccharide unit gene (*kpsM*), metal chaperone gene (*yciC*), and numerous hypothetical protein-encoding genes were associated more with 2019 MCRPEC strains (Figure 4A; Table S6). The SNP-based results revealed 43 core-genome SNPs (cgSNPs) in the MCRPEC strains between 2016 and 2019 after filtering the SNPs with an adjusted *P* value $< 1 \times 10^{-4}$ (Benjamini–Hochberg method). Of these, 40 cgSNPs were synonymous mutations without amino acid changes (Figure 4B; Table S7), and three cgSNPs induced amino acid changes in three proteins: the inner membrane protein, YqjF (an amino acid in the non-coding region changed into a Met upstream of the original Met, resulting in a Met–Met concatenation), the *L*-ribulose-5-phosphate 4-epimerase, UlaF (F88L), and the periplasmic alpha-amylase, MalS (T479A) (Figure 4B; Table S7).

Open pangenome

The pangenome accumulation curves revealed an open pangenome in MCRPEC strains from two time points (Figure S4A), and the median of the total genome size was significantly larger in the 2019 strains [5.04 Mb (IQR: 4.90–5.23 Mb)] than in the 2016 strains [4.91 Mb (IQR: 4.80–5.07 Mb)] ($P < 0.001$, Mann–Whitney *U* test; Figure S4B). We observed a significant difference in the distribution of core genes between the 2016 and 2019 MCRPEC strains ($P = 0.010$, Mann–Whitney *U* test; Figure S5A; Table S2), although these two groups of strains had same median number

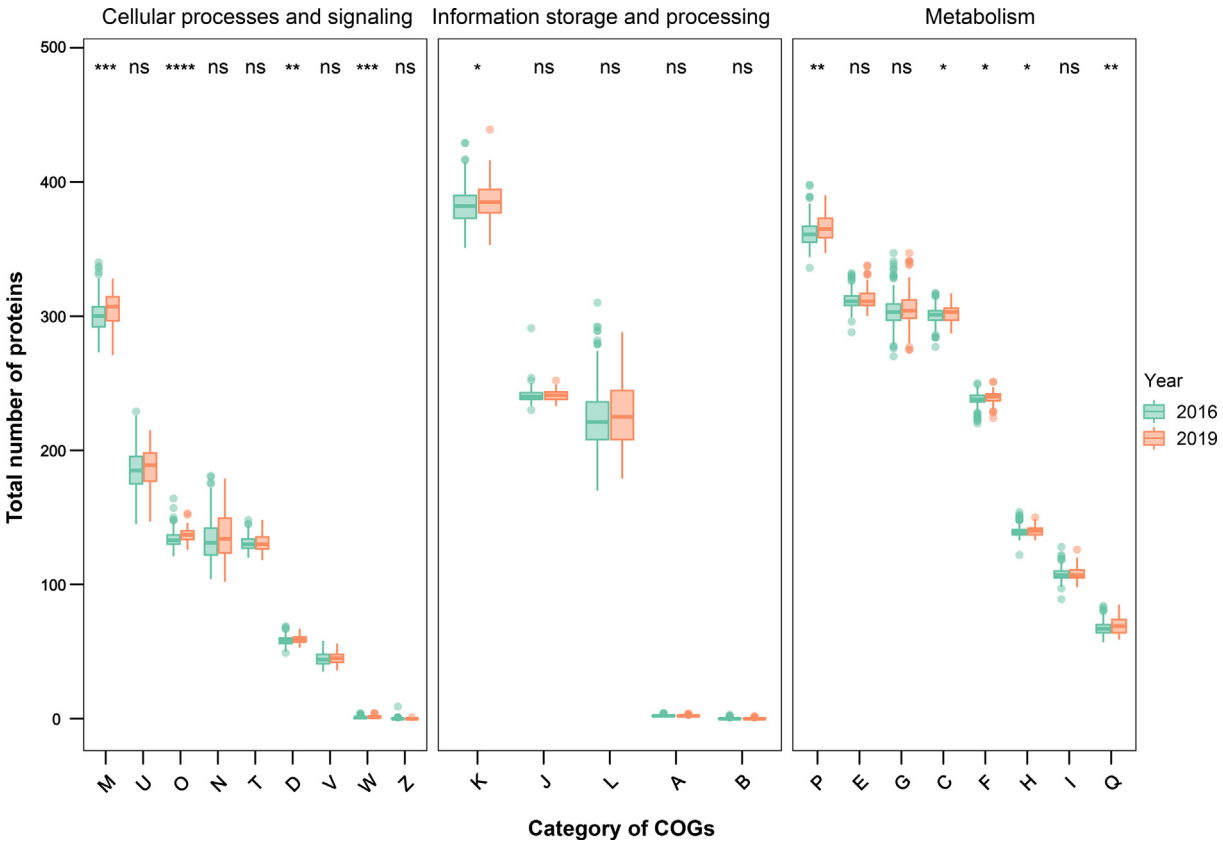

Figure 3 Differences in COGs of proteins among MCRPEC strains in 2016 and 2019

Data are shown as box plot. The horizontal box lines represent the first quartile, the median, and the third quartile. Whiskers denote the range of points within the first quartile − 1.5× the interquartile range and the third quartile + 1.5× the interquartile range. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, no significance (Wilcoxon test). COG, cluster of orthologous groups; M, cell wall/membrane/envelope biogenesis; U, intracellular trafficking, secretion, and vesicular transport; O, post-translational modification, protein turnover, chaperones; N, cell motility; T, signal transduction mechanisms; D, cell cycle control, cell division, chromosome partitioning; V, defense mechanisms; W, extracellular structures; Z, cytoskeleton; K, transcription; J, translation, ribosomal structure and biogenesis; L, replication, recombination and repair; A, RNA processing and modification; B, chromatin structure and dynamics; P, inorganic ion transport and metabolism; E, amino acid transport and metabolism; G, carbohydrate transport and metabolism; C, energy production and conversion; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; Q, secondary metabolites biosynthesis, transport, and catabolism.

(n = 2441) and IQR (2440–2442) of core genes. Moreover, the median number of accessory genes in the 2019 strains was significantly higher than that in the 2016 strains [2327 (IQR: 2214.5–2495.5) in 2019 *vs.* 2207 (IQR: 2078–2370.5) in 2016] ($P < 0.001$, Mann–Whitney U test; Figure S5B; Table S2). These results are consistent with the increased number of VGs in MCRPEC strains from 2019, which contributes to more accessory genes and a larger genome.

Population structure dynamics

Independent phylogenetic trees were established based on 203,510 and 168,226 cgSNPs derived from 287 and 83 MCRPEC strains from 2016 and 2019, respectively. The number of population lineages, using hierBAPS [14], changed from four (3.8%, 5.6%, 26.8%, and 63.8%) in 2016 to three (14.4%, 41.0%, and 44.6%) in 2019 (Figure S6; Table S8), indicating different population structures. ClermonTyping [15] revealed that phylogenetic group A was dominant in all MCRPEC strains (60.5%, 224/370) and significantly decreased in 2019 (64.5% in 2016 *vs.* 47.0% in 2019; $P = 0.006$, χ^2 test) (**Figure 5**; Table S2). A previous study has reported that most of the strains in phylogenetic groups B2, D, and F were extraintestinal pathogenic *E. coli* (ExPEC) strains, which contained more VGs than commensal *E. coli* strains in other phylogenetic groups [16]. We found that the percentages of MCRPEC strains in these three groups significantly increased in 2019 [6.6% (19/287) in 2016 *vs.* 14.5% (12/83) in 2019; $P = 0.0409$, χ^2 test], suggesting an increase in the percentage of ExPEC strains. We thus analyzed the pathotypes of all MCRPEC strains using the collected VG database [17]. We identified 30 (10.5%) ExPEC strains in the MCRPEC strains from 2016 and 30 (36.1%) ExPEC strains in the MCRPEC strains from 2019. This result revealed a significantly increased proportion of ExPEC strains in 2019 ($P < 0.001$, χ^2 test), suggesting a population shift from mainly commensal *E. coli* strains to ExPEC strains (Figure 5; Table S2). Moreover, these ExPEC strains were distributed across various provinces

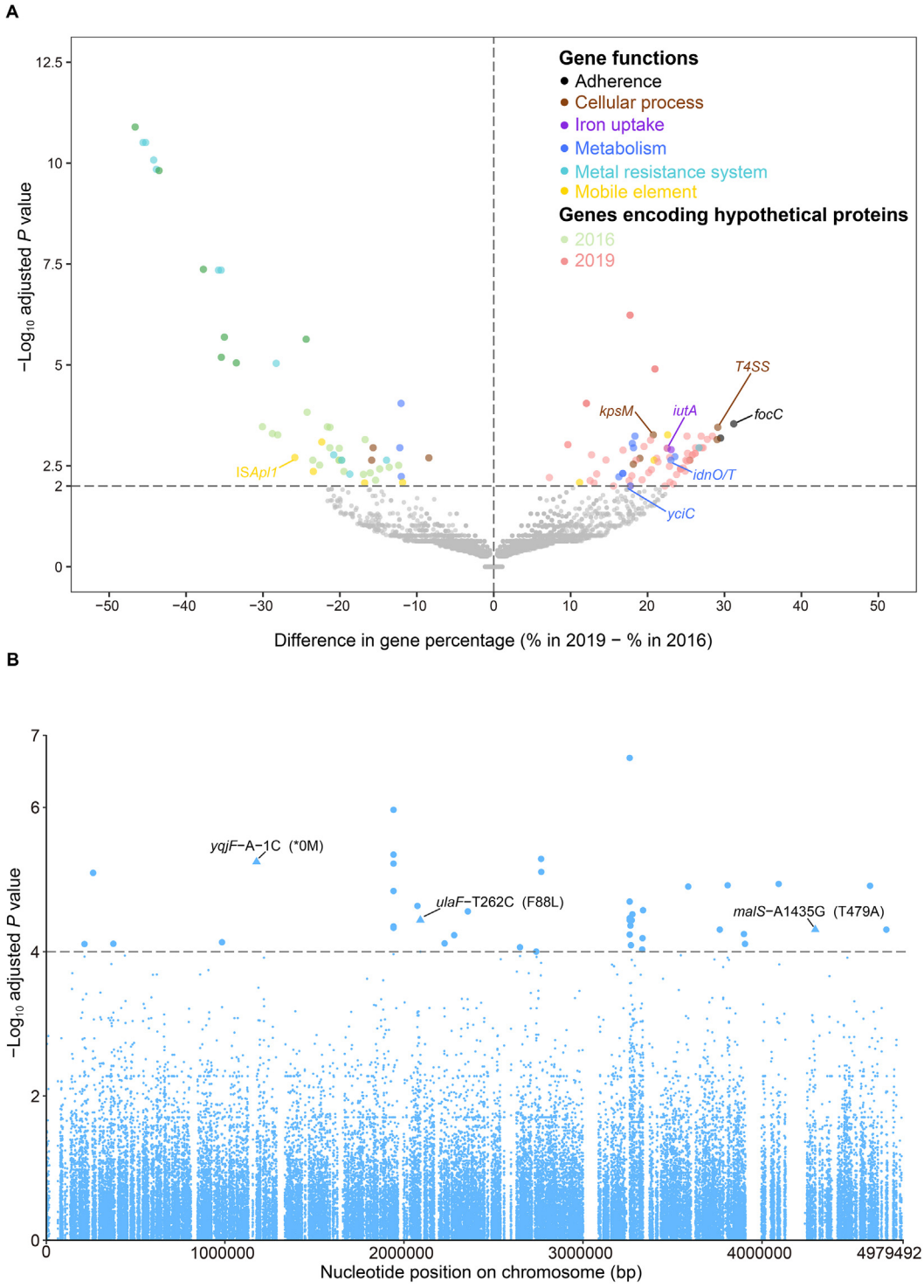

Figure 4 GWAS of MCRPEC strains between 2016 and 2019

A. Gene-based. The X-axis represents different gene percentages between 2016 and 2019 (2019 minus 2016). The Y-axis represents the $\log_{10}$-transformed adjusted *P* value ($-\text{Log}_{10}$ adjusted *P* value) of the percentage of each gene between two groups using the Benjamini–Hochberg method. Each dot represents a separate gene which is clustered in the same color as other genes with similar functions. The dash line indicates the cut-off of adjusted *P* value with 1×10^{-2}, and genes showing a significant difference in percentage are colored above this line. Genes encoding hypothetical proteins are indicated in green (2016) or pink (2019). The specific genes are labeled with text in the figure. **B.** SNP-based. Dots represent each cgSNP. The Y-axis represents the $\log_{10}$-transformed adjusted *P* value ($-\text{Log}_{10}$ adjusted *P* value) with the Benjamini–Hochberg method. The X-axis represents the nucleotide position on the chromosome of reference strain B78, which was randomly selected using snippy to generate cgSNPs. The dashed line indicates the cut-off of adjusted *P* value with 1×10^{-4} after adjustment using Benjamini–Hochberg method. The items above line are considered as cgSNPs significantly associated with MCRPEC strains in 2019, of which dots and triangles indicate with synonymous and non-synonymous mutations, respectively. The text indicates nucleotide changes and positions of the corresponding genes, and amino acid mutations are listed in bracket. The nucleotide change and position of corresponding gene were calculated from initiating codon on the DNA strain with same transcription orientation of corresponding gene. The negative number indicates the position on the upstream of the gene and the asterisk indicates the amino acid in the non-coding region. GWAS, genome-wide association study; SNP, single nucleotide polymorphism; cgSNP, core-genome SNP.

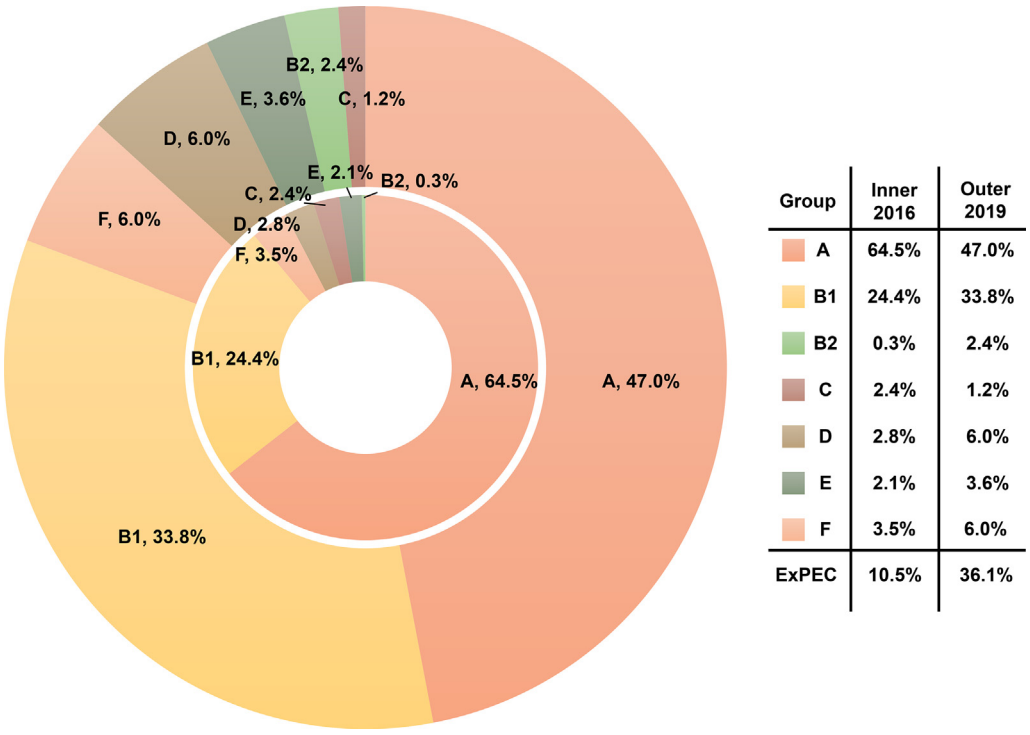

Figure 5 Phylogroup profiles of MCRPEC strains
The phylogroups are indicated with different colors. The inner and outer circles represent strains from 2016 and 2019, respectively. ExPEC, extraintestinal pathogenic *E. coli*.

without obvious geographic clustering (Table S2). In addition, 10 intestinal pathogenic *E. coli* strains were identified, including 8 (2.8%) MCRPEC strains from 2016 [6 enteropathogenic *E. coli* (EPEC) strains and 2 enterotoxigenic *E. coli* (ETEC) strains] and 2 (2.4%) MCRPEC strains from 2019 [2 diffusely adherent *E. coli* (DAEC) strains]. We further performed multilocus sequence typing (MLST) for all the MCRPEC strains, and identified 170 distinct sequence types (STs), of which, 135 and 54 were found in 2016 and 2019, respectively (Figure S7). ST10 was the most prevalent in 2016 (16.0%, 46/287) and 2019 (7.2%, 6/83), without a significant decrease ($P = 0.064$, χ^2 test). However, 116 STs detected in 2016 were absent in 2019, yet 35 new STs were uniquely present in 2019 (Table S2). Taking together, the dynamics of MCRPEC strain population depending on hierBAPS, ClermonTyping, and MLST indicated a population shift between 2016 and 2019. Additionally, MLST data showed a diversity of MCRPEC strains from two time points suggesting that the population shift is not clonally related.

Discussion

In this study, we revealed three major genomic changes in MCRPEC strains between 2016 and 2019 that might be associated with its decreasing carriage in human intestines in China. First, the loss of IS*Apl1* in Tn*6330* in MCRPEC strains in 2019 (Figure 1B), which provides a more stable genetic context for *mcr-1* on chromosome or immobile plasmids, may lead to decreased dissemination of *mcr-1*. The similar loss of IS*Apl1* in MCRPEC strains was reported in global dataset [18]; however, we provided a recent event of human MCRPEC strains and observed a higher percentage of none IS*Apl1* structure, which confirmed the evolutionary trajectory. Most incomplete Tn*6330* elements were located on the IncI2 plasmid (Figure S1), which possesses a competitive advantage over other plasmids (such as IncHI2, IncFII, and IncY) and has become the predominant vector for *mcr-1* dissemination [19,20]. A similar plasmid shift was also identified in animal- and patient-derived MCRPEC strains [9,21]. These results suggested a possible evolutionary pathway for *mcr-1*: either it was lost from fitness-cost plasmids or its surrounding IS*Apl1* elements were lost in fitness-enhanced plasmids (*e.g.*, IncI2). Second, the significant enrichment of other ARGs *aac(3)-IVa* and *aph(4)-Ia* (aminoglycosides), $bla_{\text{CTX-M-1}}$ (β-lactams), and *dfrA5* (trimethoprim) in MCRPEC strains in 2019 (Figure 2A) suggests co-selection of *mcr-1* with these antibiotics used in both animals and humans [22]. Notably, apramycin, an aminoglycoside antibiotic extensively and largely used in animals in China (198.5 tons in 2018) [23], has recently passed phase I clinical trials in human medicine (Innovative Medicines Initiative supported antibiotic passes Phase I clinical trials, 2020; https://www.imi.europa.eu/news-events/newsroom/imi-supported-antibiotic-passes-phase-i-clinical-trials). Therefore, co-selecting *mcr-1* by antibiotics used in both animals and humans poses potential public health risks. Third, both the COG and GWAS analyses revealed that 2019 MCRPEC strains possessed enriched VGs and more metabolism-associated proteins (Figures 3B, 4, and 5, Figure S3), indicating that those strains can occupy an intestinal niche over normal commensal *E. coli* strains and also have the ability of surviving in other harsh environments. For example, the Stg

and P fimbriae mediate adherence to renal and bladder epithelial cells and are commonly associated with ExPEC strains [24,25]. Additionally, more iron-uptake factors were associated with MCRPEC in 2019, which is consistent with more COG proteins potentially increasing ion transport and further indicating an enhanced ability to compete for iron within the host in host–pathogen interactions [26,27]. Subsequently, the MCRPEC strain population structure showed an increased proportion (from 10.5% to 36.1%) of ExPECs in 2019 (Figure 5). All these findings highlighted a shift in population dynamics of MCRPEC strains of human carriage, indicating a promising evolutionary trajectory of MCRPEC strains with more virulence system and higher capacity of metabolism, which may pose a public health threat.

Consequently, we believe that a "trade-off" strategy exists between resistance and virulence in *E. coli* [28]. This study reveals that MCRPEC strains possess more VGs in 2019, suggesting that these strains may break this equilibrium with high virulence and resistance. There is an ongoing debate that ExPECs with more VGs may promote survival and colonization in human intestine without causing self-limited diseases in humans, indicating that virulence may be a by-product of commensalism [29]. However, immunocompromised populations may be vulnerable to intestinal MCRPEC-associated infections, leading to treatment difficulties because clinical MCRPEC strains exhibit higher resistance levels to other antimicrobials than do clinical non-MCRPEC strains [30]. These population changes indicate a disconcerting evolutionary direction after the reduced prevalence of MCRPEC, leading to a prospering of ExPEC strains which successfully compete with commensal *E. coli* and expand owing to their high ability to occupy ecological niches.

This study has several limitations. First, we did not include MCRPEC strains from other one health sectors such as animal, food, environment, and infectious samples that would have enabled us to survey the landscape of the whole MCRPEC strain population, and sample sizes in each province were low. However, our dataset is consistent in location and origin, which minimizes the sample bias and focuses the genomic differences on human carriage. In addition, the 2019 isolates were distributed unevenly especially in Zhejiang Province (54.2% in 2019 and 3.5% in 2016); however, we had to include as many strains as possible to enlarge our dataset beyond a significant decrease in overall population of MCRPEC. Second, ExPECs are facultative pathogens living with the normal gut microbiota but cause extraintestinal infections of elderly or immunocompromised patients [31]. Additionally, all MCRPEC strains were isolated from healthy humans without obvious intestinal or other diseases; hence, the exact pathogenicity and risk to human health requires further comprehensive studies. Finally, although our comparative analysis suggests that human fecal MCRPEC strain populations have experienced a genetic shift, it is unknown as to whether the ExPECs recruited the *mcr-1*-positive plasmids via stochastic or other processes, nor do we know what molecular changes and/or other unknown drivers may have contributed to this shift.

Despite these limitations, our study describes the population dynamics of MCRPEC strains in the human gut in China. We reveal three major changes in genomic composition between these two time points: 1) stabilization of *mcr-1* by loss of IS*Apl1* and converging of *mcr-1* from fitness-cost plasmids to less burdensome plasmids, *e.g.*, IncI2; 2) association with hitchhiker genes conferring resistance to other clinically

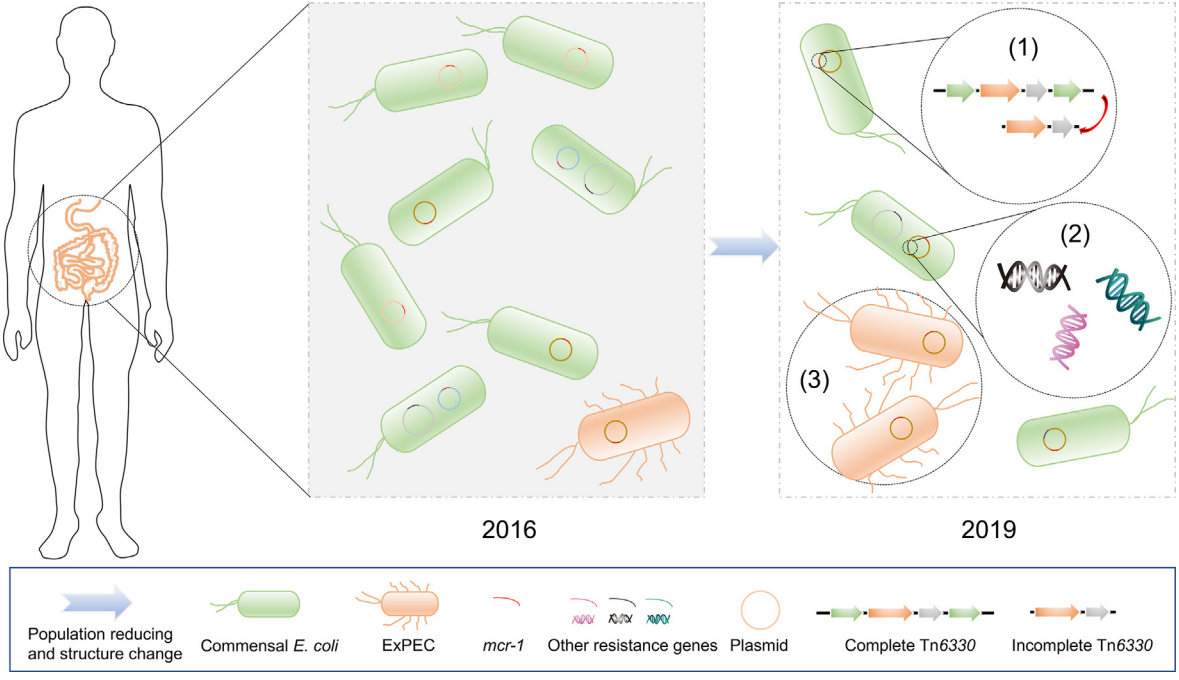

Figure 6 Schematic diagram of the MCRPEC population structure dynamics between 2016 and 2019
We found three major differences on MCRPEC population between 2016 and 2019: 1) loss of IS*Apl1* and costly plasmids for stabilizing *mcr-1* in an advantageous plasmid (IncI2), 2) hitchhiking by other resistance genes, *e.g.*, $bla_{CTX\text{-}M\text{-}1}$ and *aac(3)-IVa*, and 3) MCRPEC population shift by reducing the proportion of commensal *E. coli* strains and increasing the proportion of ExPEC strains that can occupy ecological niches and pose greater risks to human health.

important antibiotics potentially resulting in co-selection; and 3) an increase in the proportion of *mcr-1*-positive ExPEC strains that could occupy wider ecological niches and occasionally cause gastroenteritis (**Figure 6**).

The MCRPEC strain population largely decreased in 2019 from that in 2016 [9,10]; however, counterintuitively, this change additionally accompanies a "seesaw" effect on the proportion of *mcr-1* between commensal *E. coli* and ExPEC strains: the large MCRPEC strain population was mainly associated with commensal *E. coli* strains, while the reduced MCRPEC strain population was largely related with ExPEC strains. Similar to China, other countries, including Thailand, Japan, Brazil, India, Malaysia, and Argentina, have banned colistin as a growth promoter in the animal industry [2], which will most likely reduce the prevalence of MCRPEC strains; however, the change of MCRPEC strain populations in humans and animals remains uncertain. Based on the "AMR One Health" concept, continued surveillance of MCRPEC strains in the food chain, including animals, environment, food, and humans, is necessary, and antibiotics, especially apramycin and β-lactams, must be used prudently in both animals and humans.

Conclusion

Our comparative genomic analysis of the MCRPEC strains from human carriage revealed the loss of IS*Apl1 in Tn6330*, coexistence of *mcr-1* with other resistance genes, increased association of *mcr-1* with IncI2 plasmids, and more VGs contributing to greater stability and ability to cause disease. Notably, the population structure of MCRPEC strains shifted from a large proportion of commensal lineages to ExPEC strains, indicating higher risks to public health. Our study provides important genomic data on human MCRPEC strains, giving insight into the population dynamics of these significantly reduced resistant bacteria.

Materials and methods

Genomic sequencing and statistics

To compare population structures of MCRPEC strains, 370 isolates were collected in 2016 (n = 287, from a total of 5159 samples) and 2019 (n = 83, from a total of 5657 samples) from the feces of healthy people from 30 hospitals in 30 provinces/autonomous regions/municipalities across China (Figure 1; Table S1). The genomic sequences from the 2016 MCRPEC isolates were published previously [11] (BioProject: PRJNA400107). The 2019 MCRPEC isolates were retrieved from another study [10], and the genomic DNA of the 83 MCRPEC isolates from 2019 was extracted for whole-genome sequencing using the Wizard Genomic DNA Purification Kit (Catalog No. A1125, Promega, Beijing, China). The DNA libraries were prepared using the KAPA HyperPrep Kit (Catalog No. KK8505, Roche, Basel, Switzerland) following standard protocols and subsequently sequenced on the Illumina HiSeq X Ten platform with 150 bp paired-end strategy (Sinobiocore, Beijing, China). The draft genomes were assembled using the SPAdes v3.13.1 algorithm [32]. The *mcr-1*-positive contigs were extracted from assembly genomes after seeking the position of *mcr-1* using the BLAST+ algorithm [33]. Data for the whole genome and the *mcr-1*-positive contigs, including total length, N50, and GC content (%) were obtained using script compiled in Python v3.7.6. The figures corresponding to the whole-genome size and *mcr-1*-positive size were generated using ggstatsplot v0.6.1 in R v3.6.3. We extracted the 50 bp, 50 bp, and 53 bp of the 5′ end, middle, and 3′ end of IS*Apl1* respectively, and manually calculated the exact presence of IS*Apl1* after using *mcr-1*-positive contigs to map with these sequences.

Molecular characteristic profiles

All assemblies were screened for ARGs, VGs, and plasmid replicons against the Antibiotic Resistance Gene-ANNOTation (ARG-ANNOT) [34], Virulence Factors Database (VFDB) [35], and PlasmidFinder [36] databases with the BLASTN algorithm [33]. The ARG and VG matrices were defined as present or absent and evaluated using the vegan package v2.5-6 (http://CRAN.R-project.org/package=vegan) in R with Shannon index values and Jaccard distances to assess the α- and β-diversities, respectively, of the MCRPEC strains between two time points. Non-metric multidimensional scaling of the β-diversity was conducted based on the matrix of the Jaccard distances of the ARGs and VGs using the cmdscale command in R. The percentages of ARGs and VGs in two MCRPEC populations were visualized using a volcano plot in ggplot2 v3.3.2 [37] of R. The Inc type of the plasmid carrying *mcr-1* was defined as co-existence of the *mcr-1* and replicon genes in the same contig. We deduced the plasmid types using contigs that do not contain the replicon gene, to BLAST in National Center of Biotechnology Information (NCBI) database to find special skeleton genes in each plasmid. As results, the undetermined types were defined as "unknown".

COG protein analysis

All genomic assemblies were annotated using Prokka v1.11 [38] and followed by a pangenomic analysis using Roary v3.6.0 [39]. The pangenomic results were annotated to COG clusters using eggNOG-mapper v5.0 [13] with the taxonomic scope on bacteria. The abbreviative categories of the COGs are listed at https://www.sbg.bio.ic.ac.uk/~phunkee/html/old/COG_classes.html, and the COG results were plotted as boxplots using ggplot2 in R.

Pathotype analysis

The pathotype was defined as a group of strains of a single species that cause a common disease using a common set of VGs. The pathotypes are described according to infectious site as intestinal pathogenic *E. coli*, including EPEC, enterohemorrhagic *E. coli* (EHEC), ETEC, enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), and DAEC [40], and as ExPEC, including uropathogenic *E. coli* (UPEC), meningitis-associated *E. coli* (MNEC), and avian pathogenic *Escherichia coli* (APEC) [41]. We collected VGs associated with all pathotypes described in a previous study [17] as a

database, and blasted our sequences against this database to detect the pathotypes of all MCRPEC strains.

GWAS

For GWAS, we used gene-based and SNP-based strategies to assess the gene differences and mutations, respectively, associated with two groups. First, the pangenome-wide association analysis was performed using Scoary v1.6.16 [42] with the gene presence/absence dataset produced by Roary with the "--collapse" option to identify genes that are often inherited together and with a phylogeny tree generated in Roary to avoid lineage effects. We used the Benjamini–Hochberg method to control false positives, and the genes signigicantly associated with two groups were visualized using a volcano plot in R. Second, cgSNPs of all MCRPEC strains were extracted using snippy v4.6.0 (https://github.com/tseemann/snippy), and the randomly selected strain, B78, was used as a reference. The maximum-likelihood (ML) tree was generated using FastTree 2 with the "-gtr" option [43]. pyseer v1.3.6 [44] was used to analyze cgSNP differences between MCRPEC strains in 2016 and 2019. We used the cgSNPs from snippy-core as the SNP input and patristic distances from the ML phylogeny tree from FastTree to test for lineage effects. Mutations signigicantly associated with two time points were screened for synonymous and non-synonymous variants using Artemis v18.1.0 [45]. All variants were mapped against the concatenation sequence of the reference genome and visualized using a Manhattan plot with ggplot2 in R.

Phylogeny analysis

The matrix with the resulting gene content of the pangenome from all MCRPEC strains was used as input to plot the pangenomic curve using the vegan package in R [36]. The cgSNP-based phylogenetic tree was produced using the parsnp tool in Harvest suite v1.2 [46], and a strain was randomly chosen as a reference sequence with the parameter "-r!". Lineages among the MCRPEC populations were predicted by alignment with the hierBAPS algorithm [14]. The midpoint phylogenetic tree was unrooted and visualized using Interactive Tree Of Life (iTOL) v5 [47]. Phylogroups of MCRPEC strains were detected using ClermonTyping [15]. MLST was performed using SRST2 [48], and the minimum spanning tree was generated in BioNumerics v7.0 (Applied Maths, Sint-Martens-Latem, Belgium) using the BURST algorithm.

Statistical analyses

Categorical data were analyzed via χ^2 tests or Fisher's exact test. Continuous data with normal and non-normal distributions were analyzed with *t*-test and Mann–Whitney *U* test/Wilcoxon test, respectively. Shannon indexes of the ARGs and VGs between two time points were compared using *t*-test, and the significance of the β-diversity between groups was determined via PERMANOVA using the vegan package (http://CRAN.R-project.org/package = vegan) in Adonis with 999 permutations. All statistical analyses were computed in R v3.6.3.

Ethical statement

All samples were collated by The Second Affiliated Hospital of Zhejiang University School of Medicine at two time points in the same hospitals from previous studies [10,11]. Ethical permission was agreed by the Zhejiang University ethics committee under the project DETER-XDR-China (Approval No. I2019001117). Individual written consent forms were translated into Mandarin and consent was obtained for all healthy individuals.

Data availability

All new sequences used in this study were have been deposited in the BioProject database in NCBI (BioProject: PRJNA714274), which are publicly accessible at https://www.ncbi.nlm.nih.gov/, and also in the Genome Warehouse [49] at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation (GWH: GWHBJFF00000000–GWHBJIJ00000000), which are publicly accessible at https://ngdc.cncb.ac.cn/gwh/.

CRediT author statement

Yingbo Shen: Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft, Funding acquisition. **Rong Zhang:** Investigation, Resources, Data curation. **Dongyan Shao:** Software, Formal analysis, Data curation, Visualization. **Lu Yang:** Validation, Data curation, Visualization. **Jiayue Lu:** Investigation. **Congcong Liu:** Investigation. **Xueyang Wang:** Investigation. **Junyao Jiang:** Investigation. **Boxuan Wang:** Investigation. **Congming Wu:** Investigation. **Julian Parkhill:** Writing - review & editing. **Yang Wang:** Writing - review & editing, Funding acquisition. **Timothy R. Walsh:** Writing - review & editing, Funding acquisition. **George F. Gao:** Conceptualization, Supervision, Project administration, Writing - review & editing. **Zhangqi Shen:** Conceptualization, Methodology, Writing - review & editing, Funding acquisition. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported in part by the grants from the Guangdong Laboratory for Lingnan Modern Agriculture Project, China (Grant No. NT2021006), the National Natural Science Foundation of China (Grant Nos. 81861138051, 32002335, 32141001, and 31761133004), the National Postdoctoral Program for Innovative Talents, China (Grant No. BX20190359), and the UK MRC DETER-XDR-China-HUB (Grant No. MR/S013768/1).

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gpb.2022.11.006.

ORCID

ORCID 0000-0002-7259-6195 (Yingbo Shen)
ORCID 0000-0002-2174-7985 (Rong Zhang)
ORCID 0000-0003-2083-0872 (Dongyan Shao)
ORCID 0000-0002-6530-0544 (Lu Yang)
ORCID 0000-0001-5448-8476 (Jiayue Lu)
ORCID 0000-0002-1989-6711 (Congcong Liu)
ORCID 0000-0002-7745-5615 (Xueyang Wang)
ORCID 0000-0001-5737-3930 (Junyao Jiang)
ORCID 0000-0001-7276-370X (Boxuan Wang)
ORCID 0000-0002-4238-9080 (Congming Wu)
ORCID 0000-0002-7069-5958 (Julian Parkhill)
ORCID 0000-0002-5928-9377 (Yang Wang)
ORCID 0000-0003-4315-4096 (Timothy R. Walsh)
ORCID 0000-0002-3869-615X (George F. Gao)
ORCID 0000-0003-4789-7540 (Zhangqi Shen)

References

[1] Li J, Nation RL, Turnidge JD, Milne RW, Coulthard K, Rayner CR, et al. Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. Lancet Infect Dis 2006;6:589–601.
[2] Shen Y, Zhang R, Schwarz S, Wu C, Shen J, Walsh TR, et al. Farm animals and aquaculture: significant reservoirs of mobile colistin resistance genes. Environ Microbiol 2020;22:2469–84.
[3] Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, 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 2016;16:161–8.
[4] Ling Z, Yin W, Shen Z, Wang Y, Shen J, Walsh TR. Epidemiology of mobile colistin resistance genes *mcr-1* to *mcr-9*. J Antimicrob Chemother 2020;75:3087–95.
[5] Partridge SR, Di Pilato V, Doi Y, Feldgarden M, Haft DH, Klimke W, et al. Proposal for assignment of allele numbers for mobile colistin resistance (*mcr*) genes. J Antimicrob Chemother 2018;73:2625–30.
[6] Carroll LM, Gaballa A, Guldimann C, Sullivan G, Henderson LO, Wiedmann M. Identification of novel mobilized colistin resistance gene *mcr-9* in a multidrug-resistant, colistin-susceptible *Salmonella enterica* serotype typhimurium isolate. mBio 2019;10:e00853-19.
[7] Wang X, Wang Y, Zhou Y, Li J, Yin W, Wang S, et al. Emergence of a novel mobile colistin resistance gene, *mcr-8*, in NDM-producing *Klebsiella pneumoniae*. Emerg Microbes Infect 2018;7:122.
[8] Walsh TR, Wu Y. China bans colistin as a feed additive for animals. Lancet Infect Dis 2016;16:1102–3.
[9] Shen C, Zhong LL, Yang Y, Doi Y, Paterson DL, Stoesser N, et al. Dynamics of *mcr-1* prevalence and *mcr-1*-positive *Escherichia coli* after the cessation of colistin use as a feed additive for animals in China: a prospective cross-sectional and whole genome sequencing-based molecular epidemiological study. Lancet Microbe 2020;1:e34–43.
[10] Wang Y, Xu C, Zhang R, Chen Y, Shen Y, Hu F, et al. Changes in colistin resistance and *mcr-1* abundance in *Escherichia coli* of animal and human origins following the ban of colistin-positive additives in China: an epidemiological comparative study. Lancet Infect Dis 2020;20:1161–71.
[11] Shen Y, Zhou H, Xu J, Wang Y, Zhang Q, Walsh TR, et al. Anthropogenic and environmental factors associated with high incidence of *mcr-1* carriage in humans across China. Nat Microbiol 2018;3:1054–62.
[12] Li R, Xie M, Zhang J, Yang Z, Liu L, Liu X, et al. Genetic characterization of *mcr-1*-bearing plasmids to depict molecular mechanisms underlying dissemination of the colistin resistance determinant. J Antimicrob Chemother 2017;72:393–401.
[13] Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 2019;47:D309–14.
[14] Cheng L, Connor TR, Siren J, Aanensen DM, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol Biol Evol 2013;30:1224–8.
[15] Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. ClermonTyping: an easy-to-use and accurate *in silico* method for *Escherichia* genus strain phylotyping. Microb Genom 2018;4:e000192.
[16] Johnson JR, Russo TA. Molecular epidemiology of extraintestinal pathogenic *Escherichia coli*. EcoSal Plus 2018;8:ESP-0004-2017.
[17] Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nat Rev Microbiol 2004;2:123–40.
[18] Wang R, van Dorp L, Shaw LP, Bradley P, Wang Q, Wang X, et al. The global distribution and spread of the mobilized colistin resistance gene *mcr-1*. Nat Commun 2018;9:1179.
[19] Wu R, Yi LX, Yu LF, Wang J, Liu Y, Chen X, et al. Fitness advantage of *mcr-1*-bearing IncI2 and IncX4 plasmids *in vitro*. Front Microbiol 2018;9:331.
[20] Yang J, Wang HH, Lu Y, Yi LX, Deng Y, Lv L, et al. A ProQ/FinO family protein involved in plasmid copy number control favours fitness of bacteria carrying *mcr-1*-bearing IncI2 plasmids. Nucleic Acids Res 2021;49:3981–96.
[21] Jiang Y, Zhang Y, Lu J, Wang Q, Cui Y, Wang Y, et al. Clinical relevance and plasmid dynamics of *mcr-1*-positive *Escherichia coli* in China: a multicentre case-control and molecular epidemiological study. Lancet Microbe 2020;1:e24–33.
[22] van Duijkeren E, Schink AK, Roberts MC, Wang Y, Schwarz S. Mechanisms of bacterial resistance to antimicrobial agents. Microbiol Spectr 2018;6:ARBA-0019-2017.
[23] Ministry of Agriculture, Rural Affairs PRC. Report on the Use of Veterinary Antibiotics of China in 2018. Official Veterinary Bulletin 2019;21:57–9.
[24] Lymberopoulos MH, Houle S, Daigle F, Leveille S, Bree A, Moulin-Schouleur M, et al. Characterization of Stg fimbriae from an avian pathogenic *Escherichia coli* O78:K80 strain and assessment of their contribution to colonization of the chicken respiratory tract. J Bacteriol 2006;188:6449–59.
[25] Lane MC, Mobley HLT. Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. Kidney Int 2007;72:19–25.
[26] Skaar EP. The battle for iron between bacterial pathogens and their vertebrate hosts. PLoS Pathog 2010;6:e1000949.
[27] Weinberg ED. The development of awareness of iron-withholding defense. Perspect Biol Med 1993;36:215–21.
[28] Johnson JR, Goullet P, Picard B, Moseley SL, Roberts PL, Stamm WE. Association of carboxylesterase B electrophoretic pattern with presence and expression of urovirulence factor determinants and antimicrobial resistance among strains of *Escherichia coli* that cause urosepsis. Infect Immun 1991;59:2311–5.
[29] Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, et al. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. Mol Biol Evol 2007;24:2373–84.

[30] Wang Y, Tian GB, Zhang R, Shen Y, Tyrrell JM, Huang X, et al. Prevalence, risk factors, outcomes, and molecular epidemiology of *mcr-1*-positive Enterobacteriaceae in patients and healthy adults from China: an epidemiological and clinical study. Lancet Infect Dis 2017;17:390–9.
[31] Kohler CD, Dobrindt U. What defines extraintestinal pathogenic *Escherichia coli*? Int J Med Microbiol 2011;301:642–7.
[32] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77.
[33] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.
[34] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212–20.
[35] Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis - 10 years on. Nucleic Acids Res 2016;44:D694–7.
[36] Carattoli A, Zankari E, Garcia-Fernandez A, Larsen MV, Lund O, Villa L, et al. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 2014;58:3895–903.
[37] Wickham H. ggplot2: elegant graphics for data analysis. Switzerland: Springer Cham; 2009.
[38] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–9.
[39] Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015;31:3691–3.
[40] Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev 1998;11:142–201.
[41] Russo TA, Johnson JR. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. J Infect Dis 2000;181:1753–4.
[42] Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 2016;17:238.
[43] Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One 2010;5:e9490.
[44] Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. Bioinformatics 2018;34:4310–2.
[45] Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 2012;28:464–9.
[46] Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 2014;15:524.
[47] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 2019;47:W256–9.
[48] Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genome Med 2014;6:90.
[49] Chen M, Ma Y, Wu S, Zheng X, Kang H, Sang J, et al. Genome Warehouse: a public repository housing genomescale data. Genomics Proteomics Bioinformatics 2021;19:584–9.