

# Global emergence of a hypervirulent carbapenem-resistant *Escherichia coli* ST410 clone

Received: 10 May 2023

Accepted: 22 November 2023

Published online: 12 January 2024

 Check for updates

Xiaoliang Ba<sup>1,8</sup>, Yingyi Guo<sup>2,8</sup>, Robert A. Moran<sup>3</sup>, Emma L. Doughty<sup>3</sup>, Baomo Liu<sup>4</sup>, Likang Yao<sup>2</sup>, Jiahui Li<sup>2</sup>, Nanhao He<sup>2</sup>, Siqian Shen<sup>5,6</sup>, Yang Li<sup>7</sup>, Willem van Schaik<sup>3</sup>, Alan McNally<sup>3</sup>, Mark A. Holmes<sup>1,9</sup> ✉ & Chao Zhuo<sup>2,9</sup> ✉

Carbapenem-resistant *Escherichia coli* (CREC) ST410 has recently emerged as a major global health problem. Here, we report a shift in CREC prevalence in Chinese hospitals between 2017 and 2021 with ST410 becoming the most commonly isolated sequence type. Genomic analysis identifies a hypervirulent CREC ST410 clone, B5/H24RxC, which caused two separate outbreaks in a children's hospital. It may have emerged from the previously characterised B4/H24RxC in 2006 and has been isolated in ten other countries from 2015 to 2021. Compared with B4/H24RxC, B5/H24RxC lacks the *bla*<sub>OXA-181</sub>-bearing X3 plasmid, but carries a F-type plasmid containing *bla*<sub>NDM-5</sub>. Most of B5/H24RxC also carry a high pathogenicity island and a novel O-antigen gene cluster. We find that B5/H24RxC grew faster in vitro and is more virulent in vivo. The identification of this newly emerged but already globally disseminated hypervirulent CREC clone, highlights the ongoing evolution of ST410 towards increased resistance and virulence.

Antimicrobial resistance (AMR) has become one of the most serious public health problems of the 21st century, threatening the effective prevention and treatment of infections in humans and animals alike<sup>1,2</sup>. A bacterial strain is considered multidrug-resistant (MDR) if it is resistant to at least one agent in three or more antimicrobial classes<sup>3</sup>. Carbapenems are a group of last-line antibiotics for treating infections caused by MDR Gram-negative bacteria<sup>4</sup>. However, carbapenem resistance has been observed in many pathogens, including members of the Enterobacteriaceae family such as *Escherichia coli*<sup>5,6</sup>. The most frequently observed carbapenem resistance mechanism in Enterobacteriaceae is the production of carbapenemases that hydrolyse

carbapenems<sup>4,6</sup>. *Klebsiella pneumoniae* carbapenemase, New Delhi metallo- $\beta$ -lactamase (NDM), imipenemase, Verona integron-encoded metallo- $\beta$ -lactamase, and oxacillinase-48 (OXA-48) are the most commonly encountered carbapenemases in Enterobacteriaceae<sup>7-9</sup>. Among these, NDM has been reported to be highly prevalent in many parts of the world<sup>9-11</sup>.

In the past two decades, carbapenem-resistant *E. coli* (CREC) has emerged rapidly and become increasingly prevalent around the globe, causing various hard-to-treat clinical infections<sup>9,10,12</sup>. The global CREC population is made up of diverse multilocus sequence types (STs) with varied geographic distributions. A previous study reported two major

<sup>1</sup>Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom. <sup>2</sup>State Key Laboratory of Respiratory Disease, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China. <sup>3</sup>Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, United Kingdom. <sup>4</sup>Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China. <sup>5</sup>Institute of Antibiotics, Huashan Hospital, Fudan University, Shanghai, China. <sup>6</sup>Key Laboratory of Clinical Pharmacology of Antibiotics, Ministry of Health, Shanghai, China. <sup>7</sup>Department of Clinical Laboratory, Children's Hospital of Soochow University, Suzhou, Jiangsu, China. <sup>8</sup>These authors contributed equally: Xiaoliang Ba, Yingyi Guo. <sup>9</sup>These authors jointly supervised this work: Mark A. Holmes, Chao Zhuo. ✉ e-mail: [mah1@cam.ac.uk](mailto:mah1@cam.ac.uk); [chaosheep@sina.com](mailto:chaosheep@sina.com)

(>10%) STs (ST410 and ST131) and three minor (5–10%) STs (ST1248, ST167, and ST405) in a global collection of 229 CREC collected by two global surveillance programs during 2015–2017<sup>13</sup>. No CREC from China was reported in that study<sup>13</sup>, although CREC isolates are frequently isolated in Chinese healthcare settings. Two Chinese national surveillance studies carried out over a similar period (2014–2017) revealed that most CREC isolates in China belonged to ST167 or ST131, with ST410 the third most prevalent<sup>9,14</sup>.

*E. coli* ST410 is an extraintestinal pathogen associated with multidrug resistance, and has been recognised as a high-risk international clone<sup>15,16</sup>. Whole-genome sequence analysis and evolutionary reconstruction of *E. coli* ST410 revealed that ST410 is comprised of two lineages, namely lineage A with *fimH53* (A/H53) and lineage B with *fimH24* (B/H24)<sup>15</sup>. The B/H24 lineage is further divided into four sub-lineages: B1/H24, B2/H24R, B3/H24Rx and B4/H24RxC. Sub-lineage B2/H24R is characterised by fluoroquinolone resistance associated with mutations in *gyrA* and *parC*, B3/H24Rx is defined by the introduction of the extended-spectrum  $\beta$ -lactamase-encoding gene *bla*<sub>CTX-M-15</sub>, and B4/H24RxC is defined by a further introduction of the carbapenem resistance gene *bla*<sub>OXA-181</sub> carried on an IncX3 plasmid<sup>15</sup>. A more recent study has proposed a modification to this classification, where ST410-B1 is identical to B1/H24, ST410-B2 includes both B2/H24R and B3/H24Rx, and ST410-B3 is identical to B4/H24Rx<sup>17</sup>. That study characterised ST410-B2 by the introduction of fluoroquinolone resistance mutations and ST410-B3 by the acquisition of a four-amino-acid (YRIN N337N) insertion in the *ftsI*-encoded penicillin-binding protein 3 (PBP3, also called FtsI)<sup>17</sup>. The YRIN insertion in PBP3, together with YRIK and TIPY insertions, were reported to confer reduced susceptibility to a broad range of  $\beta$ -lactams such as ceftazidime, cefepime and aztreonam<sup>18,19</sup>. Chen et al. reported that nearly all (292/293) ST410-B3 isolates analysed contain the *ftsI* YRIN insertion<sup>17</sup>. Over the past decade, there have been increasing reports of serious infections and possible hospital outbreaks involving the carbapenem-resistant ST410 sub-lineage B4/H24RxC in both developed and low- and middle-income countries<sup>15,16,20–23</sup>.

In this work, we investigate the population of clinical CREC isolates from Chinese hospital patients between 2017 and 2021. The most commonly isolated CREC lineage, ST410, is analysed further. We examine the global ST410 population structure by comparing 847 publicly available ST410 genomes to the 109 ST410 genomes generated in this study. A hypervirulent MDR ST410 clone is identified and designated B5/H24RxC. Comparative genomic analyses and phenotypic assays are performed to characterise B5/H24RxC and provide insights into its emergence and evolution.

## Results

### Characteristics of CREC isolates from Chinese hospitals

A total of 388 CREC isolates were collected from hospitals across 26 Chinese provinces between 2017 and 2021 (Fig. 1a, c; Supplementary Data 1). The isolates were recovered from various clinical samples, including ascites, bile, blood, bronchoalveolar fluid, pus, wound secretion, sputum and urine. The most common clinical sample types were urine ( $n = 111$ ), sputum ( $n = 64$ ) and blood ( $n = 47$ ) (Fig. 1b), indicating possible associations between CREC and urinary tract infections (UTIs), pneumonia and bloodstream infections. All CREC isolates were resistant to at least one of the carbapenem antibiotics tested. The isolates had a median MIC for imipenem of 8 mg/L (range 0.125 to >128 mg/L), a median MIC for meropenem of 32 mg/L (<0.03 to >128 mg/L) and a median MIC for ertapenem of 32 mg/L (<0.03 to >128 mg/L) (Fig. 1f; Supplementary Data 1). It is to be noted here that analysis for ertapenem MIC was based only on the 168 isolates collected by Guangzhou Medical University, excluding isolates collected by China Antimicrobial Surveillance Network (CHINET) as ertapenem MIC was not routinely determined by CHINET. Amongst the isolates tested for tigecycline and polymyxin resistance, 40 were resistant to

tigecycline (MIC, 1–4 mg/L) and 12 were resistant to polymyxin B (MIC, 4–64 mg/L) (Supplementary Data 1).

Genomic analysis revealed that the most prevalent carbapenem resistance gene in this CREC collection was *bla*<sub>NDM-5</sub> ( $n = 282$ ), followed by *bla*<sub>NDM-1</sub> ( $n = 42$ ) (Fig. 1e; Supplementary Data 1). Interestingly, only 87.1% (338/388) of the CREC isolates possessed at least one carbapenem resistance gene, suggesting alternative carbapenem resistance mechanisms in the rest of the isolates. Eleven isolates carried the colistin resistance gene *mcr-1* (Supplementary Data 1). The CREC isolates belonged to 71 STs with 16 of them not assigned an ST. The most commonly isolated STs were ST410 ( $n = 109$ ), ST167 ( $n = 41$ ), ST131 ( $n = 12$ ) and ST617 ( $n = 12$ ) (Fig. 1d). This indicated a change in CREC population in China relative to studies conducted between 2015 and 2017<sup>9,14</sup>, where ST410 was reported to be the third most commonly isolated CREC ST behind ST131 and ST167.

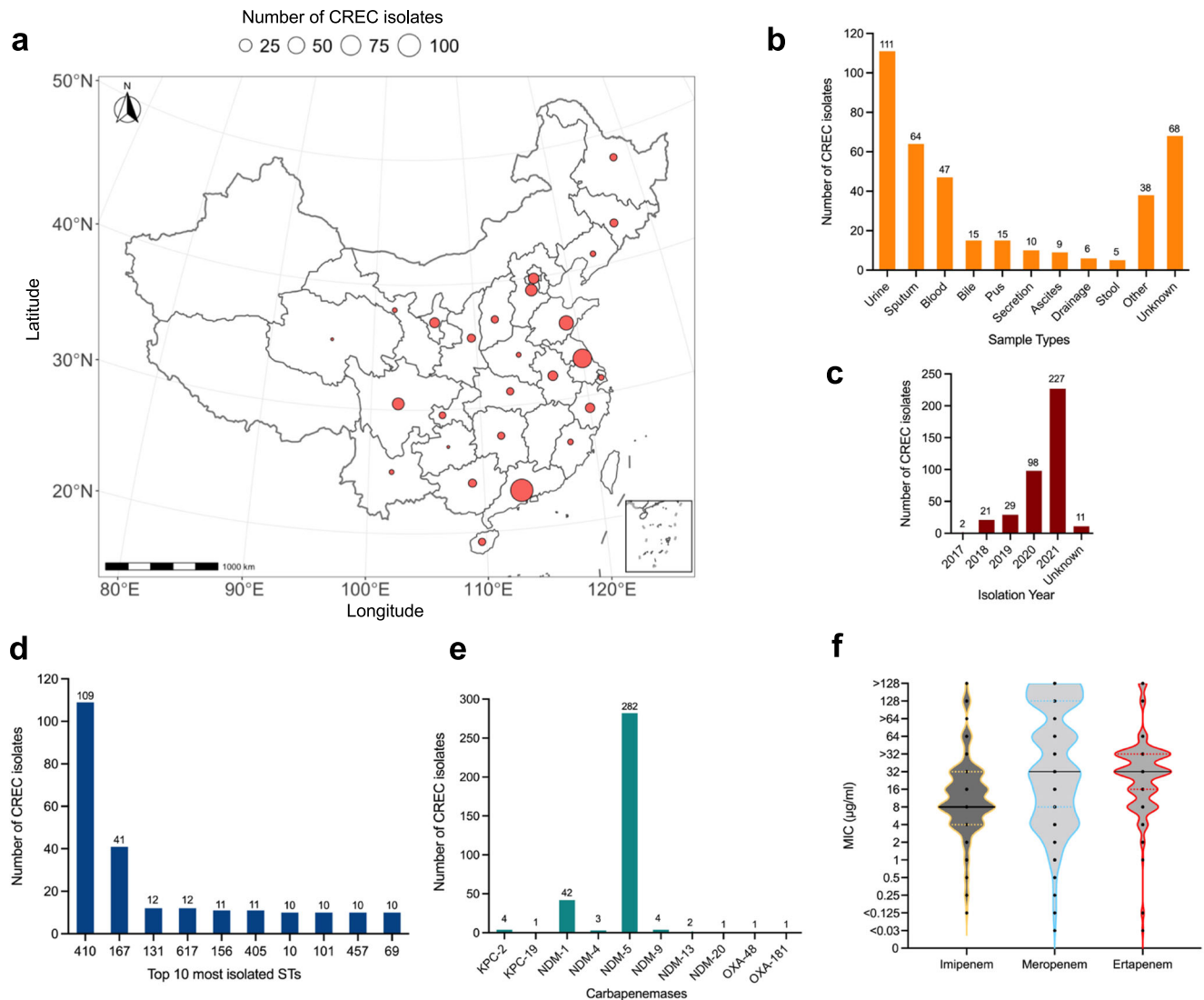
### Outbreaks of a CREC ST410 clone in a children's hospital

Of the 109 ST410 CREC, 49 were isolated from 47 inpatients admitted to a children's hospital in eastern China between 2018 and 2020. Most of the patients ( $n = 33$ , 70.2%) were under 60 days of age and 72.3% ( $n = 34$ ) of them were female (Fig. 2b). The patients were admitted to different hospital departments and suffered from various infections, including UTI ( $n = 31$ , 66.0%), pneumonia, septicaemia and bacteraemia (Supplementary Data 3). Patients were grouped into two cohorts according to their time of admission and the length of hospital stay (Fig. 2a). One cohort stayed in the hospital from March 2018 to August 2019 and the other from January 2020 to September 2020.

Four groups of *E. coli* ST410 were identified in this children's hospital, characterised by SNP analysis relative to the complete genome of isolate 19-7 (Genbank: <CP123017> to <CP123023>) which was from a patient with UTI in the same hospital (Fig. 2c). A pairwise comparison of the SNP differences across all isolates showed that isolates within the two groups differed by no more than 23 SNPs, suggesting possible outbreaks in the hospital based on a previously recommended threshold of  $\leq 25$  SNPs<sup>24</sup>. The outbreaks happened within multiple hospital departments as well as inter-departmentally as indicated by where and when the patients stayed during their treatment (Fig. 2a). Group-1 included 20 isolates mainly from between March 2018 and August 2019, with the exception of isolate 20-20 from June 2020, while group-2 only included 27 isolates from 2020 (Fig. 2a, c). The pairwise comparison also revealed that there were just 29–69 SNPs between isolates in these two groups. Singletons 18-4 and 18-10 did not belong to either group and each formed a group of their own, but 18-4 was closely related with the isolates in the two main groups as indicated by a pairwise SNP distance of 29 to 48. However, isolate 18-10 was distant from all other isolates with a pairwise SNP distance of 215–240. The dividing of the groups was further supported by a Fastbaps analysis of the phylogeny of the isolates, although 18-4 was separated from Group-1 at the second Fastbaps level (Fig. S3c, S3d). Apart from isolate 18-10, the CREC ST410 in this children's hospital were found to be distinct from previously reported ST410 clones and therefore were analysed further to define their genomic characteristics and determine their origins.

### Global population structure of *E. coli* ST410 reveals an MDR clone, B5/H24RxC

In order to analyse the CREC ST410 isolated from the children's hospital in a global context, we constructed a phylogeny of 956 *E. coli* ST410 isolates based on a core-genome SNP alignment (Fig. 3a). Fastbaps clustered this global collection into 11 BAP groups based on the SNP alignment and phylogeny. Most of the isolates were assigned to two of the main groups, BAP1 ( $n = 429$ ) and BAP2 ( $n = 434$ ), which made up 90.3% of the collection. The previously reported MDR clone B4/H24RxC, which carries the *bla*<sub>OXA-181</sub> carbapenemase gene in an X3 plasmid, was found to belong to BAP1. All of the CREC ST410 isolates



**Fig. 1 | Characteristics of CREC isolates ( $n = 388$ ) from Chinese hospitals. **a** The geographical distribution of the CREC isolates collected in this study shown on the map of P. R. China. **b** Bar chart showing the various types of clinical samples used in this study for the isolation of CREC. **c** Bar chart showing the distribution of isolation years. **d** Bar chart showing the 10 most identified STs in this CREC collection. **e** Bar chart showing identified carbapenem-resistant genes in this CREC collection.**

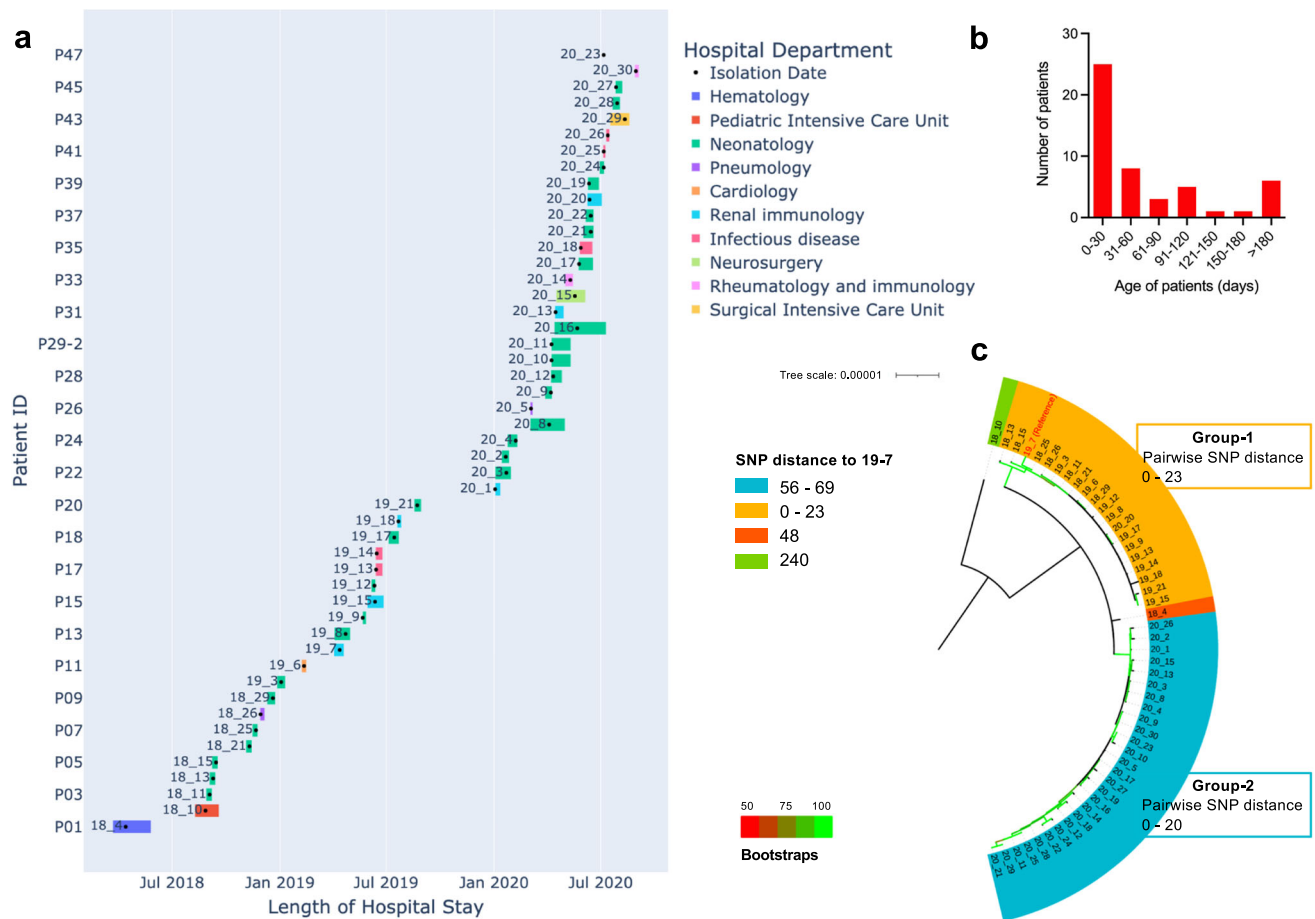
**f** Violin plot showing the distribution of MICs of imipenem, meropenem and ertapenem for the CREC isolates, the median MIC for each carbapenem antibiotics is represented with a black line. The plot for ertapenem MIC was generated based on data for isolates ( $n = 168$ ) in the Guangzhou Medical University (GMU) collection. Source data are provided as a Source Data file.

from the children's hospital outbreaks, apart from isolate 18-10, also belonged to group BAPI. However, the children's hospital isolates clustered into a subclade that was distinct from the B4/H24RxC clone. This subclade, here designated B5/H24RxC ( $n = 174$ ), rarely contained the X3 plasmid-associated *bla*<sub>OXA-181</sub> of B4/H24RxC. Instead, 97.7% (170/174) of the B5/H24RxC genomes contained a F-type plasmid (Fig. 3a, S2), carrying the *bla*<sub>NDM-5</sub> carbapenemase gene. Another obvious difference between the two clones was that most B5/H24RxC genomes lacked a small plasmid with a pColKp3-like replicon, which was present in the majority of B4/H24RxC genomes (Fig. 3a, S2).

Further comparative analyses of the two clones showed that B5/H24RxC contained a larger (two-tailed unpaired *t* test,  $p = 0.047$ ) number of acquired antimicrobial resistance genes (ARGs) and mutations that confer resistance (Fig. 3b; Supplementary Data 4). B5/H24RxC isolates were found to have more (two-tailed unpaired *t* test,  $p < 0.0001$ ) putative virulence genes than B4/H24RxC isolates (Fig. 3c). B5/H24RxC isolates had a median of 176 putative virulence genes (range 159–190) while the B4/H24RxC clone had a median of 166 (143–181). The difference was mainly caused by the presence of the

high pathogenicity island (HPI) that was originally identified in *Yersinia enterocolitica*<sup>25,26</sup>, in 77.6% (135/174) of the B5/H24RxC clone but in none of the isolates of the B4/H24RxC clone (Fig. S2). The HPI in these isolates contained all 11 virulence-associated genes (*fyuA*, *irp1*, *irp2*, *ybtA*, *ybtE*, *ybtP*, *ybtQ*, *ybtS*, *ybtT*, *ybtU*, and *ybtX*). In silico serotyping revealed that, 99.0% (384/388) of all isolates of both clones produced a H9 flagellar antigen, but the majority (95.3%, 204/214) of the B4/H24RxC isolates had an O8 lipopolysaccharide while the majority (74.7%, 130/174) of B5/H24RxC isolates had an Onovel1 (as labelled in the EcoH database by Ingle et al.<sup>27</sup>) lipopolysaccharide (Fig. 3d). BLAST analysis revealed that Onovel1 matched (96.3% coverage and 98.66% identity) a previously reported O-antigen Ogn5 (LC177549.1) identified in enterotoxigenic *E. coli* of various STs<sup>28</sup> (Fig. 3e).

The isolates of the B5/H24RxC clone in this international collection were isolated between 2015 and 2021 from humans, food and companion animals across 11 countries on 5 continents (Fig. S1; Supplementary Data 2). The majority of the isolates (89.1%, 155/174) were from countries in eastern and south-eastern Asia, such as China ( $n = 78$ ) and Thailand ( $n = 71$ ).



**Fig. 2 | Outbreaks of a ST410 lineage in a children's hospital.** **a** Gantt plot showing the length of hospital stay of the patients in the children's hospital in eastern China. Patient ID are presented on the y axis and the length of stay of each patient is represented with coloured bars. A black dot within the coloured bars indicates the time of the isolation of the isolates. **b** Bar chart showing the age

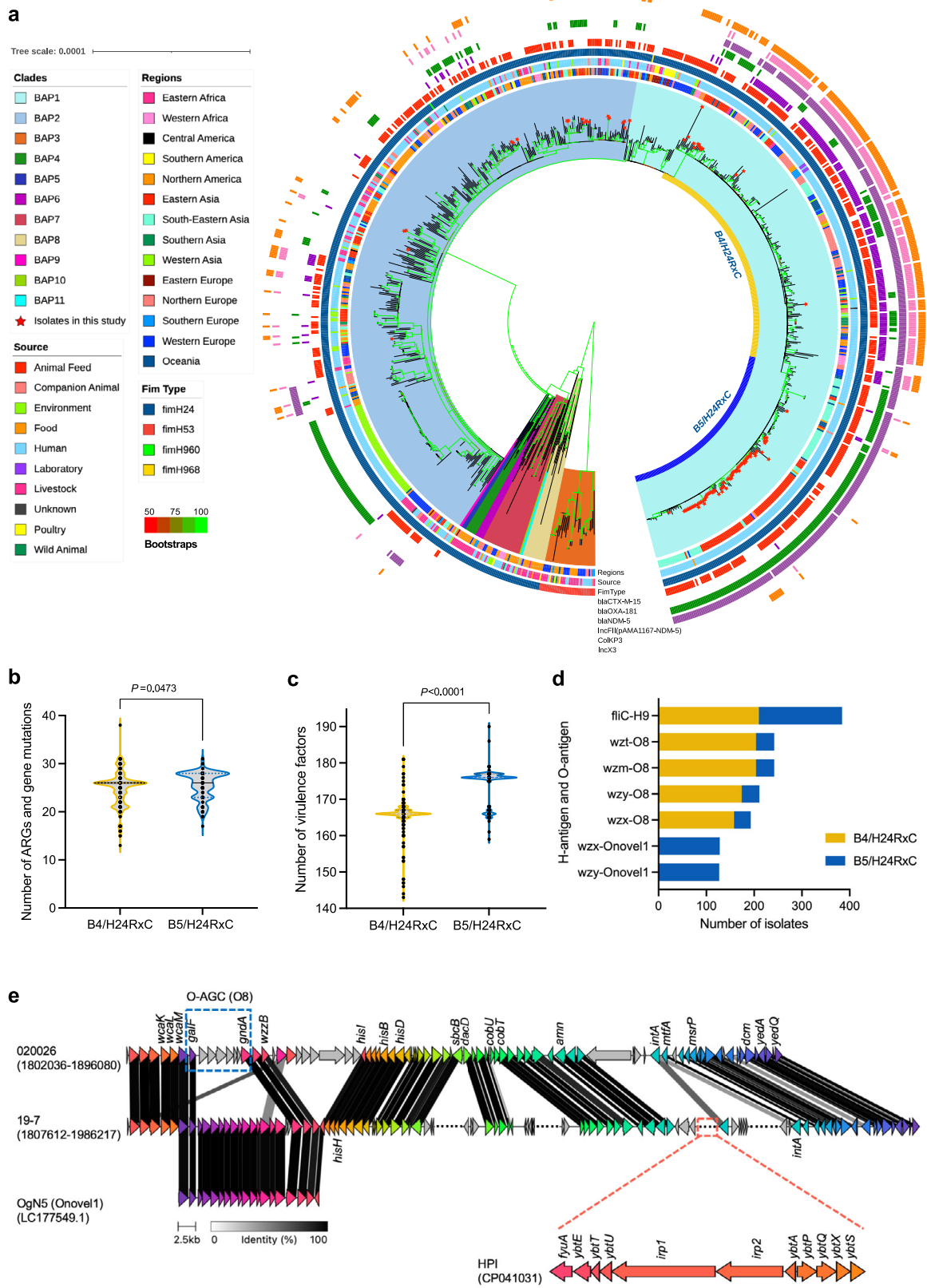
distribution of the patients. **c** Maximum-likelihood core-genome SNP phylogeny of the 49 ST410 CREC isolates in the children's hospital. Colours indicates the SNP distance to the reference genome 19-7. Bootstrap values are represented by gradient colours. Source data are provided as a Source Data file.

### Emergence of the B5/H24RxC MDR clone driven by recombination and horizontal gene transfer

A smaller scale phylogeny reconstruction was performed only on the B4/H24RxC and the B5/H24RxC MDR clones using the complete genome of a previously reported isolate O20026 (B4/H24RxC; Genbank: CP034954 to CP034958) as the reference. Gubbins identified several recombination regions associated with the B5/H24RxC MDR clone, including a 5.7-kb region that contained genes of unknown function in all 174 isolates of the clone (Fig. S5). By excluding SNPs within the recombination regions, the average pairwise distance among the 174 isolates of B5/H24RxC MDR clone was 41, ranging from 0 to 136 (Supplementary Data 5), suggesting likely ongoing global dissemination of this recently emerged clone. A total of 204 SNPs were identified on the branch separating the B5/H24RxC MDR clone from its most closely related isolate (Fig. S5), of which 171 were within the recombination regions, giving a per site *r/m* ratio (the probabilities that a given site was altered through recombination and mutation) of 5.18. This suggests that homologous recombination contributed to the evolutionary events resulting the emergence of the B5/H24RxC MDR clone. Further comparative analysis of the recombination regions in both clones also revealed the O-antigen gene cluster (O-AGC) switch from O8 in B4/H24RxC to Onovel1 in B5/H24RxC, and the HPI gene cluster in B5/H24RxC clone (Fig. 3e).

Both clones possessed a F-type plasmid containing FII-1, FIA-1, and FIB-49 replicons. The plasmid consisted of a backbone and an

antibiotic resistance region. The resistance region was bounded at one end by a copy of *ISI* and at the other end by a partial copy of *Tn5403* (Fig. 4a). Most differences between variants of this F-type plasmid were within the resistance region, while their backbones were almost identical (Fig. 4a). The conserved backbone contained three replicons and genes associated with stability (restriction-modification and toxin-antitoxin systems). Although it contained *finO* and *traX* genes, these were the only remnants of a F-like transfer region and this plasmid was not expected to be conjugative. Outside the resistance region, the backbone was only interrupted by two insertion sequences, *ISEc12* and *ISI* (Fig. 4a). *ISEc12* was flanked by a target site duplication but *ISI* was not, suggesting that it had mediated a deletion event post-insertion. Importantly, 695 bp immediately adjacent to the left end of the resistance region were absent from plasmid variants found in B5/H24RxC (Fig. 4a), and appeared to have been lost in a deletion event mediated by the *ISI* at the boundary of the resistance region. This deletion was clear evidence that the backbone variant present in B4/H24RxC was ancestral to that in B5/H24RxC. The resistance region was comprised of a series of ARG-containing translocatable element sequences interspersed with copies of *IS26* (Fig. 4b). Resistance region variants present in the B5/H24RxC clone included an *IS26*-flanked segment that contained *bla*<sub>NDM-5</sub>, *bla*<sub>TEM-1</sub>, and *sul1*, *aadA2* and *dfra12* in a class 1 integron. Although the *bla*<sub>NDM-5</sub> gene was also present in some B4/H24RxC isolates, in those it was found in variants of the *bla*<sub>OXA-181</sub> carrying X3 plasmid (Fig. S6) as previously reported<sup>16,17</sup>.



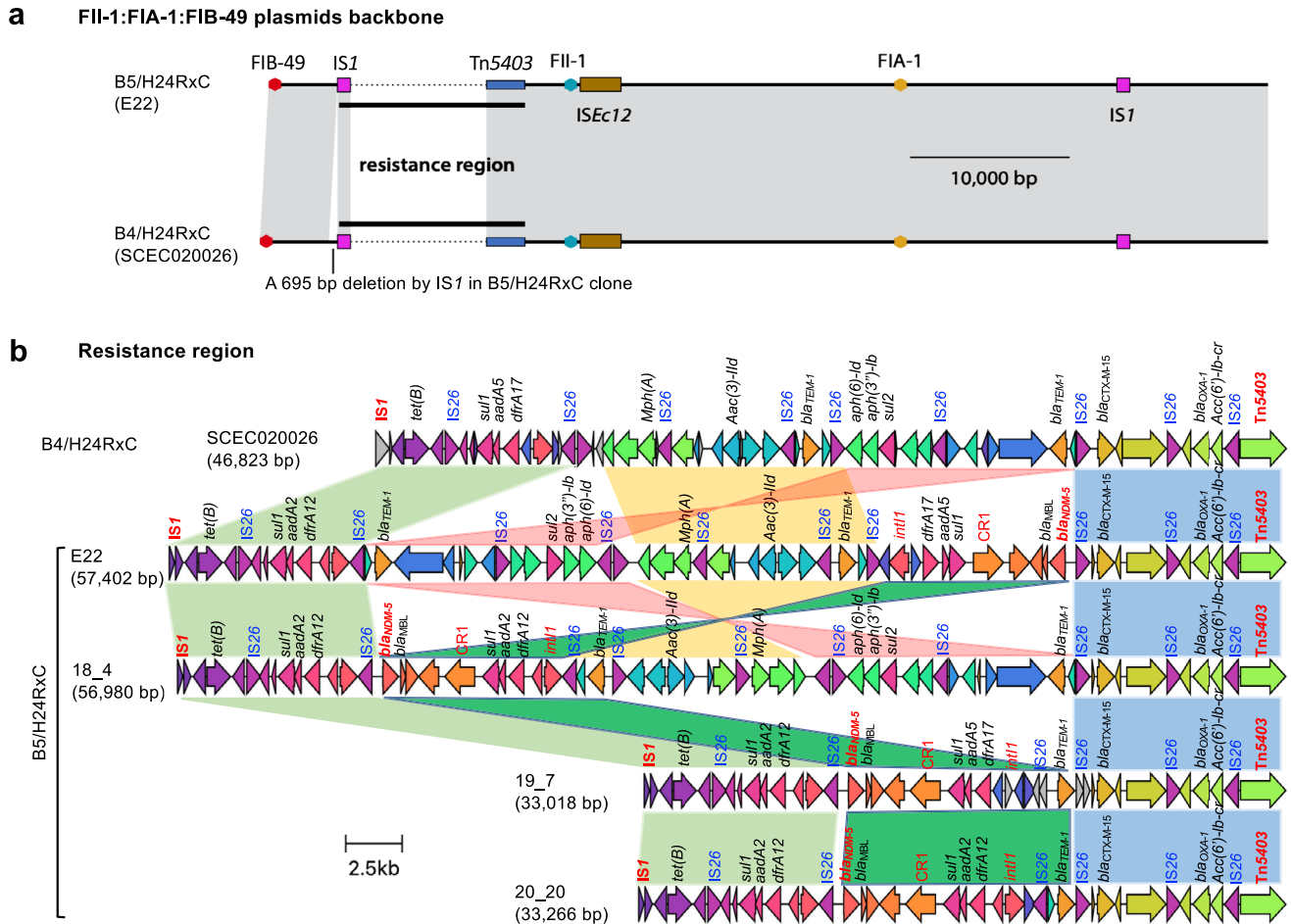
**Core-genome genes and SNPs associated with B5/H24RxC MDR clone**

Pangenome-wide association study (pan-GWAS) using Scoary for the two clones identified 199 genes that were positively or negatively associated with B5/H24RxC (Fig. 5; Supplementary Data 6). Amongst these genes, 84 were annotated as non-hypothetical, of which 44 were

over-represented in B5/H24RxC and 40 under-represented. In addition to the expected identification of accessory genes carried by different plasmids in these two clones, chromosomal accessory genes were also found to be associated with B5/H24RxC. A number of these chromosomal genes, including all 11 genes located in the HPI, were found in the majority of B5/H24RxC isolates (75.3–77.6%, 131–135/174), but in none

**Fig. 3 | Phylogeny of a global ST410 collection.** **a** Midpoint rooted maximum-likelihood phylogeny of 956 global ST410 was constructed using a core-genome SNP alignment generated by Snippy v4.6.0 with ST410 isolate YD786 (GenBank accession [CP013112.1](#)) as the reference. Branch support was performed with 1000 bootstrap replicates. Bootstrap values are represented with gradient colours. Isolates from this study are indicated with a red star. **b** Violin plot showing the distribution of total number of ARGs and mutations that confer resistance in B4/H24RxC and B5/H24RxC clones. Statistical difference between the two clones was

assessed with two-tailed unpaired Student's *t* test. **c** Violin plot showing distribution of total number of virulence factors in B4/H24RxC and B5/H24RxC clones. Statistical difference was assessed with two-tailed unpaired Student's *t* test. **d** Bar plot showing the presence of the lipopolysaccharide (O) and flagellar (H) surface antigens in B4/H24RxC and B5/H24RxC clones. **e** Comparison of the recombination regions in strain O20026 and 19-7 identified the O-antigen switch from O8 in B4/H24RxC to Onov11 (OgN5) in B5/H24RxC and the HPI gene cluster in B5/H24RxC clone. Source data are provided as a Source Data file.



**Fig. 4 | FII-1:FIA-1:FIB-49 plasmids analysis.** **a** Comparison of the backbone of the F-type plasmids in the B4/H24RxC and B5/H24RxC clones. **b** Comparison of resistance regions found in F-type plasmids. Genbank accessions for plasmids:

pCTXM15\_020026 (CP034956), pE22P1 (CP123037), p18-4P1 (CP123014), p19-7P2 (CP123019), p20-20P2 (CP123031).

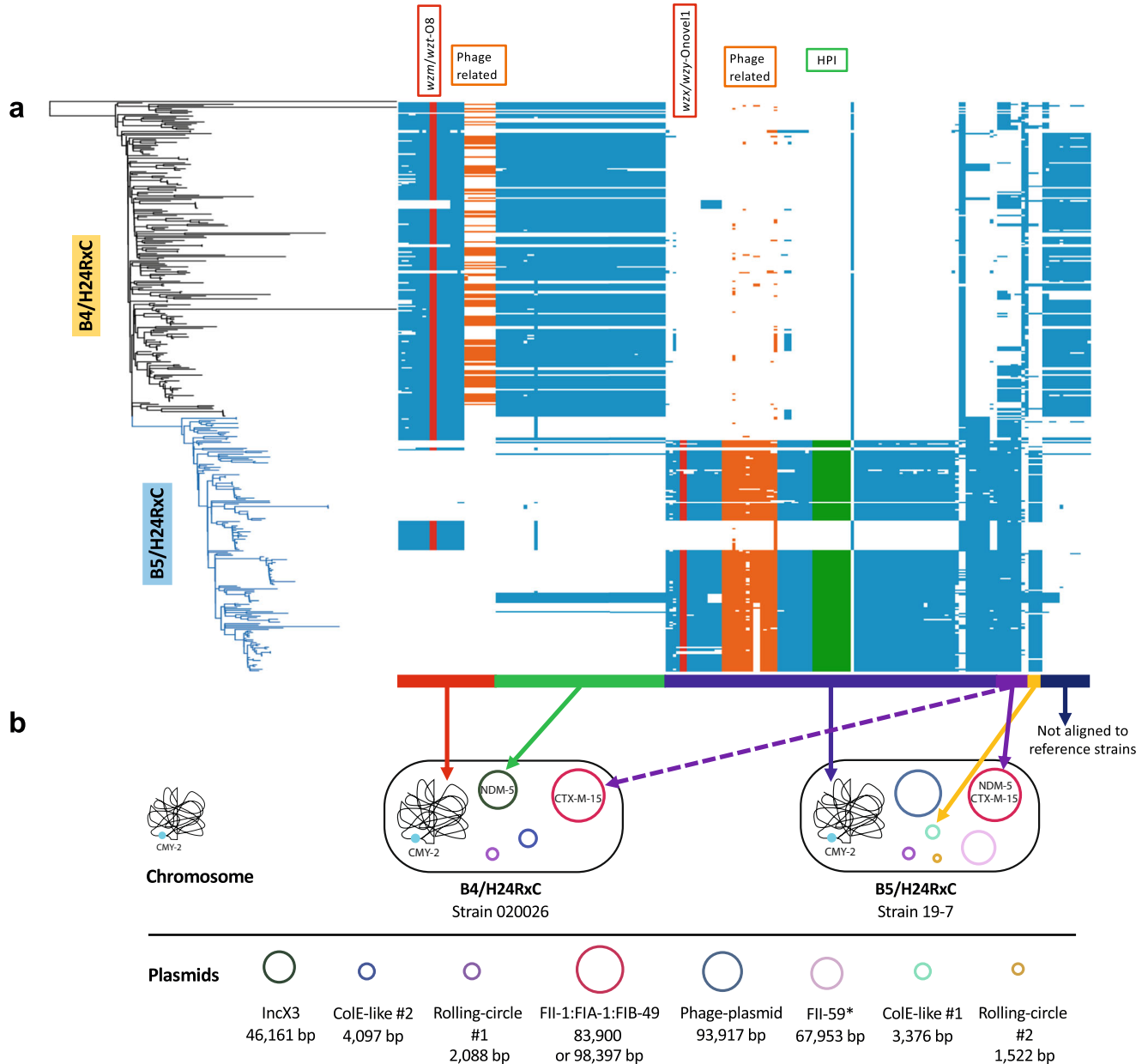
of the B4/H24RxC isolates. A prophage integrase gene (*intA*), a histone-like protein gene (*hns*) and two 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase genes (*dapH*) were also found in B5/H24RxC (75.3–77.6%, 131–135/174) but not in B4/H24RxC. Consistent with the serotyping result, O-antigen genes *wzm/wzt* O8 were associated with B4/H24RxC (204/214, 95.3%) and *wzx/wzy* Onov11 were associated with B5/H24RxC (130/174, 74.7%) (Fig. 5).

Similarly, association analysis using Scoary revealed that 423 chromosomal SNPs were positively associated with B5/H24RxC (Supplementary Data 7). Amongst the 423 SNPs, 403 were in coding sequences and 20 were in intergenic regions. Only 63 of the 403 substitutions in coding sequences were non-synonymous, present in 32 genes (Fig. S7; Supplementary Data 7), most of which were found in the recombination regions, such as genes in the histidine operon (*hisB*, *hisD* etc.), the colanic acid biosynthesis genes cluster (*wcaL* and *wcaM*) and the nitrogen assimilation transcriptional regulator gene *nac*. Premature stop codons were found in three genes which encode a PstS

family phosphate ABC transporter substrate-binding protein, the inverse autotransporter adhesin-like protein YeeJ and the cardiolipin synthase CIsB. 145 SNPs were exclusively found in B5/H24RxC, with only 12 present in all 174 isolates investigated. Gubbins revealed that amongst the B5/H24RxC specific SNPs, 129 were found to be introduced via recombination events (Fig. S7; Supplementary Data 7).

**Time of origin of the B5/H24RxC MDR clone**

The estimated time of the most recent common ancestor (TMRCA) of different phylogenetic groups was investigated with BEAST2 on the 500 Treemmer selected ST410 genomes from the 956 global collection. A mutation rate of 6.42E-7 SNPs per site per year [95% highest posterior density (HPD) intervals 5.69E-7, 7.19E-7] was estimated. The analysis estimated the age of the ST410 lineage to be approximately 205 years, with a TMRCA of around 1816 (95% HPD, 1739–1879) (Fig. 6a), close to a previous estimate of 1803<sup>15</sup>. The B4/H24RxC ancestor was estimated to have originated in 2003 (95% HPD,



**Fig. 5 | Core-genome genes associated with the B5/H24RxC MDR clone.**  
**a** Presence and absence of the genes positively and negatively associated with the B5/H24RxC clone mapped to the phylogeny. Genes are aligned and ordered against the complete genomes of the reference strains 020026 (Genbank: CP034954 to CP034958) and 19-7 (CP123017 to CP123023). Genes located in phage regions are shaded in orange, genes found in the high pathogenicity island (HPI) are in green

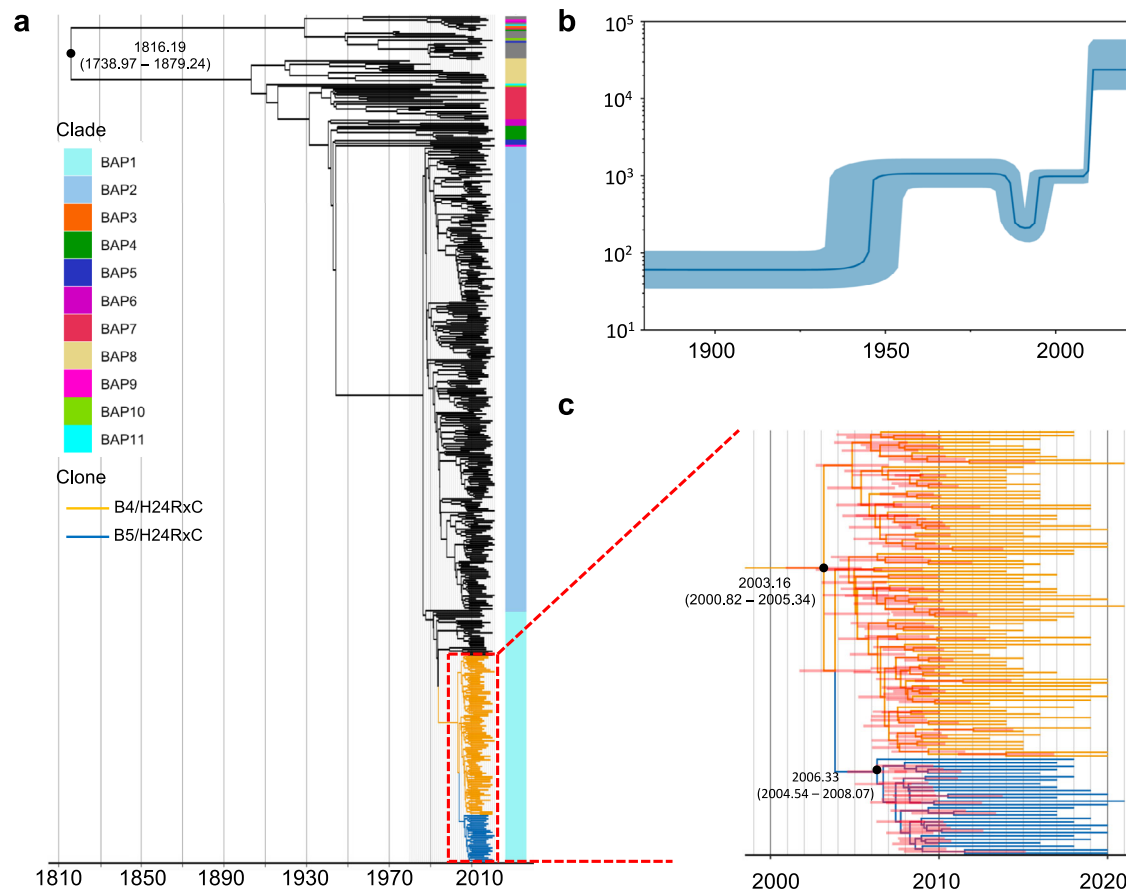
and O-group genes are in red. **b** Schematic representation of the chromosome and plasmids in the reference strains 020026 and 19-7. Colour bars and arrows indicate the location of the genes in the genomes of the reference strains. *bla*<sub>CMY-2</sub> was chromosomally integrated in B4/H24RxC as reported previously<sup>17</sup> and in B5/H24RxC. Source data are provided as a Source Data file.

2000–2005), which is identical to the estimate of 2003 from the same previous study<sup>15</sup>. The TMRCA of B5/H24RxC was estimated at around May 2006 (95% HPD, 2004–2008) (Fig. 6c).

**B5/H24RxC shows a fitness advantage and enhanced virulence**  
 In two different assays, isolates of the B5/H24RxC clone exhibited a fitness advantage compared with isolates of the B4/H24RxC clone in LB medium. Growth curve analysis showed that the isolates of B5/H24RxC grew considerably better, with a reduced doubling time of 68.5 mins in half strength LB (average of 6 isolates) compared with isolates of B4/H24RxC (83.5 mins, average of 2 isolates) (Fig. 7a, b; Table S2). In full strength LB and 1/10 strength LB, B5/H24RxC isolates also showed significantly better growth (Fig. S8; Table S2). Genomic

analysis revealed that these isolates of the B5/H24RxC clone possessed gene *fyuA* but not *yodB*, while isolates of the B4/H24RxC clone contained gene *yodB* but not *fyuA*. A qPCR-based (targeting *yodB* and *fyuA*) growth competition experiment between isolates of the two clones also confirmed that isolates of B5/H24RxC grew faster than isolates of B4/H24RxC, as demonstrated by the increasing ratio of *fyuA* to *yodB* levels in the same culture (Fig. 7c, d).

In a wax moth larvae infection model, larvae infected with isolates of B5/H24RxC had a significantly lower survival rate ( $P < 0.002$  or  $P < 0.001$ , Log-rank [Mantel–Cox] test) than those infected with isolates of B4/H24RxC (Fig. 7e; Table S3) after 72 h. The result also showed that larvae infected with isolates of B5/H24RxC had a lower survival rate than those infected with hypervirulent *Klebsiella pneumoniae*



**Fig. 6 | Coalescence-based analysis of *E. coli* ST410.** **a** A time-calibrated phylogeny was reconstructed using BEAST2.0 based on the nonrecombinant SNPs for the 500 selected *E. coli* ST410. The MDR clone B4/H24RxC and B5/H24RxC are coloured in orange and blue, respectively. **b** The Bayesian skyline plot illustrates the

predicted demographic changes of the ST410 clades. The thick solid line represents the median estimate of the effective population size, with 95% confidence interval shown in lighter blue area. **c** An enlarged phylogenetic tree showing the B4/H24RxC and B5/H24RxC clones.

strain K1088 ( $P=0.0053$  to  $P=0.0263$ , Log-rank [Mantel–Cox] test) and a similar survival rate to those infected with hypervirulent *Acinetobacter baumannii* strain AB5075 ( $P>0.1$ , Log-rank [Mantel–Cox] test). Based on this data relative to these two hypervirulent bacterial strains, we consider that B5/H24RxC is a hypervirulent *E. coli* clone. This result was expected as these B5/H24RxC isolates possessed the characterised pathogenicity island HPI, which was not present in the B4/H24RxC isolates. The HPI also contributes to iron acquisition by *E. coli* through the production of yersiniabactin<sup>29</sup>. Our iron source growth assay confirmed the enhanced ability of B5/H24RxC to utilise iron in comparison to B4/H24RxC (Fig. 7f). The growth of all tested isolates was completely inhibited in the presence of 300  $\mu$ M DIP, suggesting that they were equally resistant to iron-deprived conditions. The addition of  $\text{FeCl}_2$  restored growth of all isolates, while the addition of holo-transferrin only restored the growth of B5/H24RxC isolates. The result also showed that a higher concentration of haemoglobin and hemin was required to restore growth of the B4/H24RxC isolates than the B5/H24RxC isolates. Interestingly, growth was restored in none of the isolates with the addition of lactoferrin, which was different from a previous study where B4/H24RxC isolates tested were shown to be able to utilise iron from lactoferrin<sup>16</sup>. It suggests that not all B4/H24RxC isolates could utilise iron from lactoferrin as we had used two different isolates in this study.

Biofilm formation assays revealed that isolates of B4/H24RxC clone were poor biofilm formers as suggested previously<sup>16</sup>. The biofilm formation abilities of isolates of the B5/H24RxC clone were also categorised as weak or non-production (Fig. S9) following previously proposed interpretation criteria<sup>30,31</sup>.

### B5/H24RxC is globally disseminated

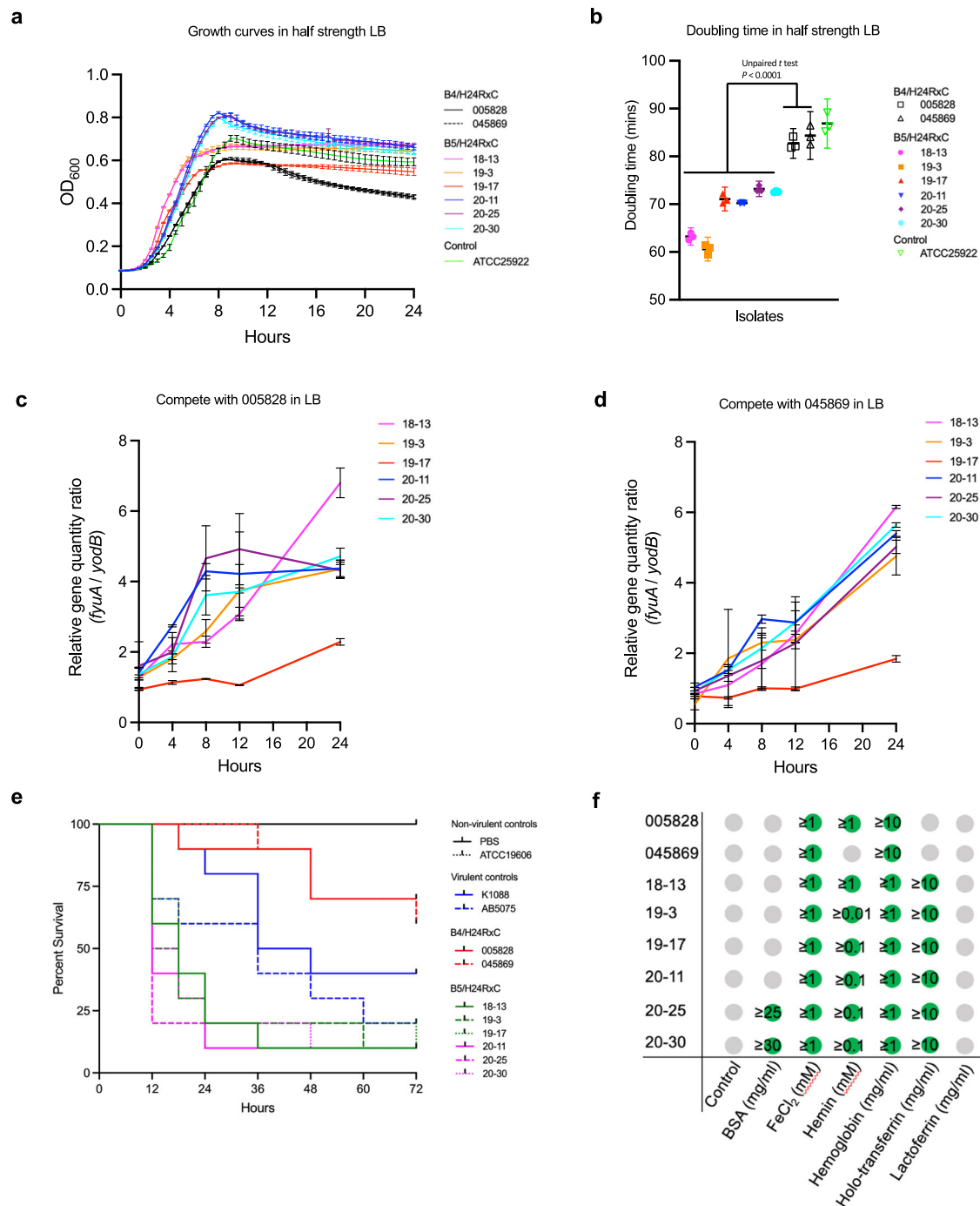
The isolates of the B5/H24RxC clone ( $n=174$ ) in this international collection ( $n=956$ ) were collected between 2015 and 2021 from humans, food and companion animals across 11 countries on 5 continents (Fig. S1; Supplementary Data 2). The majority of the isolates (89.1%, 155/174) were from countries in eastern and south-eastern Asia, such as China ( $n=78$ ) and Thailand ( $n=71$ ).

To capture a more up-to-date picture of the spread of the B5/H24RxC clone worldwide since our analysis, we screened a new collection of *E. coli* ST410 genomes ( $n=714$ , Supplementary Data 8) available from 14 Jan 2022 to 27 Sept 2023 on Enterobase. A further 84 B5/H24RxC genomes were identified from the collection (Fig. S10; Supplementary Data 8). The isolates providing these genomes were collected between 2016 and 2023 across 13 countries on 5 continents (Fig. S10), mainly from humans, with two isolates from dogs and two from environmental samples. The countries with the most isolates were Thailand ( $n=26$ ), United States ( $n=17$ ), Germany ( $n=7$ ), China ( $n=6$ ), Cambodia ( $n=6$ ), Australia ( $n=5$ ) and Norway ( $n=4$ ), suggesting the ongoing spread of this clone globally.

### Discussion

In this study, we report the emergence of a hypervirulent MDR clone of *E. coli* ST410. *E. coli* ST410 has emerged recently as a successful MDR extraintestinal pathogenic *E. coli* lineage responsible for increasing numbers of infections worldwide<sup>15,16</sup>. Our genomic surveillance of CREC in Chinese hospitals between 2017 and 2021 revealed that ST410 was the most commonly isolated CREC ST, overtaking ST167, ST131





**Fig. 7 | Phenotypic comparison of B4/H24RxC and B5/H24RxC clones. a** Growth curves in half strength LB for strains of both clones. Strain ATCC 25922 was included as a growth control. Data are shown as mean  $\pm$  SD from  $n = 3$  biological replicates. **b** Doubling time in half strength LB for isolates of both clones. Strain ATCC 25922 was included as a growth control. Data are shown as mean  $\pm$  SD from  $n = 3$  biological replicates. Statistical difference was assessed with two-tailed unpaired Student's  $t$  test. **c, d** qPCR-based competition assay for strains of both clones. The figures show the relative quantity ratio of gene *fyuA* in B5/H24RxC strains to gene *yodB* in B4/H24RxC strains 005828 and 045869. Data are shown as mean  $\pm$  SD from  $n = 3$  biological replicates. Statistical difference was assessed with two-tailed unpaired Student's  $t$  test. **e** Survival curves for wax moth larvae (*G.*

*mellonella*) infected with  $\sim 2 \times 10^6$  CFU of different isolates in the B4/H24RxC and B5/H24RxC clones. Hypervirulent *Klebsiella pneumoniae* strain K1088<sup>65</sup> and hyper-virulent *Acinetobacter baumannii* strain AB5075<sup>66</sup> were used as positive controls while *Acinetobacter baumannii* ATCC 19606 and PBS were used as negative controls. The curves represent the mean of three biological repeats. **f** The ability to utilise different iron sources by B4/H24RxC and B5/H24RxC clones. Green circles indicate growth and the numbers inside show the lowest tested concentration of the iron sources needed for the isolates to grow. Grey circles indicate no growth was observed at any concentration of the iron sources used.  $n = 2$  biological independent experiments with the same result. Source data are provided as a Source Data file.

and ST617 which were reported to be most common during surveillance carried out in 2015–2017<sup>7,14</sup>. Previous studies have identified significant clades in the ST410 lineages based on their genomic characteristics, and reported a globally widespread MDR clone, B4/

H24RxC<sup>15,16</sup>. The clone identified here, B5/H24RxC, possessed distinctive features that distinguish it from B4/H24RxC. It was similarly globally disseminated and was found to have caused two separate outbreaks in a children's hospital in eastern China.

Genomic analysis revealed that the B5/H24RxC clone was closely related to and likely emerged from B4/H24RxC. Both clones were grouped into BAPI by Fastbaps<sup>32</sup> based on a SNP alignment and a phylogeny of an international collection of ST410 genomes ( $n = 956$ ). However, the B5/H24RxC clone lacked the X3 plasmid-borne *bla*<sub>OXA-181</sub>, which is a defining feature of B4/H24RxC<sup>15</sup>. Instead, B5/H24RxC carried *bla*<sub>NDM-5</sub> in a F-type plasmid derived from an ancestral variant found in B4/H24RxC, which lacks *bla*<sub>NDM-5</sub>. However, *bla*<sub>NDM-5</sub>-bearing plasmids of the same F-type lineage, F1:A1:B49, have been described in some B4/H24RxC isolates<sup>17</sup>, suggesting either that this plasmid lineage acquired *bla*<sub>NDM-5</sub> in a B4/H24RxC host that was the progenitor to B5/H24RxC, or that the IS26-associated *bla*<sub>NDM-5</sub> was acquired on at least two occasions. Importantly, the majority of B5/H24RxC isolates investigated in this study carried the high pathogenicity island HPI and a novel O-antigen (Onovel1) which were likely introduced via recombination (Fig. S2). Based on these findings, we argue that adding a new sub-lineage (B5/H24RxC) to the Roer et al.<sup>15</sup> classification scheme provides better resolution for distinguishing ST410 *E. coli* with distinctive genotypic and phenotypic features than the modified classification scheme proposed by Chen et al.<sup>17</sup>.

Our wax larvae infection assay demonstrated that the HPI-containing B5/H24RxC was more virulent than B4/H24RxC, adding further evidence that HPI contributes to enhanced virulence in *E. coli* as previously described<sup>29,33</sup>. The HPI element is frequently found in the genomes of extraintestinal pathogenic *E. coli* (ExPEC) associated with UTIs<sup>34</sup>, especially in uropathogenic *E. coli* (UPEC) strains<sup>35,36</sup>. The HPI allows *E. coli* to obtain iron from the urinary tract, which is normally a low-iron environment. This enhances the ability of the bacteria to colonise and persist within the urinary tract and causing infections<sup>37,38</sup>. Our data from the children's hospital in eastern China also showed that a large proportion of HPI-carrying B5/H24RxC isolates were from patients with UTI, suggesting that HPI may have played a role in the emergence of this highly virulent clone. However, it should be noted that we did not explore other genetic factors, apart from the HPI, that could contribute to the enhanced virulence in the B5/H24RxC clone. Another important genomic difference that may also have contributed to the emergence and dissemination of B5/H24RxC is the O-antigen switching from O8 in B4/H24RxC to Onovel1 in B5/H24RxC. It is not clear why *E. coli* changes its O-antigen, but it is known that the ability to vary the O-antigen structure is important for bacterial adaptation to changing environments, including evasion of host immune responses<sup>39,40</sup>. Our data also indicated that B5/H24RxC had a fitness advantage over B4/H24RxC, exhibiting a higher growth rate in vitro. These traits could explain the successful global dissemination of B5/H24RxC over the last 15 years.

Additionally, non-synonymous SNPs associated with B5/H24RxC clone were identified in 32 genes, which are mostly involved in metabolism and cell wall biosynthesis, such as the genes in the histidine operon and the colanic acid biosynthesis genes cluster. It is unclear whether these substitutions in these genes contributed to the emergence of B5/H24RxC from B4/H24RxC, but they may have played a role in its increased fitness. Interestingly, the inverse autotransporter adhesin-like protein YeeJ, which was reported to promote biofilm formation<sup>41</sup>, was disrupted by a premature stop codon caused by an SNP. This may have contributed to the poor ability of B5/H24RxC to form biofilm (Fig. S9), although B4/H24RxC with an intact YeeJ is also a poor biofilm former<sup>16</sup>. It should also be noted that, in this study, non-synonymous SNPs associated with B5/H24RxC were not further investigated to validate their possible links to some phenotypic changes. Therefore, it remains unclear if YeeJ plays a role in biofilm formation in these CREC clones.

Coalescence analysis on a global selection of 500 ST410 isolates indicated a most recent common ancestor existed ~205 years ago (1816; 95% HPD, 1739–1879), similar to a previous estimate of 1803<sup>15</sup>. The B4/H24RxC clone was estimated to have originated in 2003, which

also agrees with the same previous study<sup>15</sup>. The TMRCA of the B5/H24RxC clone was estimated at around May 2008 (95% HPD, 2004–2008).

In this study, we have identified a CREC ST410 clone, B5/H24RxC, that may have greater clinical impact relative to its precursors due to the acquisition of resistance elements and the high pathogenicity island, HPI. Although the majority of B5/H24RxC were of human origin because of sampling bias, a few isolates ( $n = 3$ ) were collected from companion animals and food samples, implying that various routes of transmission may be contributing to its dissemination. It is important to note that at the time of this study, most B5/H24RxC isolates were collected from low- and middle-income countries. The multidrug-resistant and hypervirulent nature of this clone, coupled with various socioeconomic factors in these countries, makes it a challenging problem for healthcare systems already stretched to their limits. The emergence of both B4/H24RxC and B5/H24RxC MDR clones over the last two decades highlights the rapidly evolving landscape of pathogenic *E. coli*, which is being driven by continued evolution towards enhanced resistance and virulence, which in turn is being driven by recombination of key loci involved in mammalian pathogenesis and colonisation. To mitigate the impact of newly emerged and future clones, additional research is required to understand the evolutionary mechanisms involved in the emergence of the new *E. coli* clones that are of human and veterinary clinical importance.

## Methods

### Ethics and consent

This study was approved by the medical ethics committee of the First Affiliated Hospital of Guangzhou Medical University (GMU) on 21 May 2018. For the work involving the CREC isolates from the children's hospital, ethics was approved by the medical ethics committee of the Children's Hospital of Soochow University on 5 January, 2021. Individual consent was obtained from the patients' guardians by hospital staff.

### Bacterial isolates and antimicrobial susceptibility testing

An initial collection of 168 CREC isolates (GMU collection) were isolated from patients admitted to municipal hospitals in Guangzhou, Jiangsu and Beijing during 2018–2021. All isolates were identified as *E. coli* by a VITEK 2 Compact system (BioMérieux, Marcy l'Etoile, France) and confirmed to be carbapenem resistant by antimicrobial susceptibility testing. Minimum inhibitory concentrations (MICs) of meropenem, imipenem and ertapenem were determined by broth microdilution method according to the Clinical Laboratory and Standards Institute (CLSI) M100 (31st edition). The MICs of other antibiotics, including polymyxin B and tigecycline, were also determined by broth microdilution method according to European Committee on Antimicrobial Susceptibility Testing (EUCAST). For polymyxin B and tigecycline, the breakpoints defined by the EUCAST were used. All antibiotics were purchased from Macklin, Shanghai, China. *E. coli* ATCC 25922 and *P. aeruginosa* ATCC 27853 were used for standardisation.

A further 220 CREC collected by the China Antimicrobial Surveillance Network (CHINET) from Chinese hospitals in 26 provinces during 2017–2021 were also included in this study. Whole-genome sequencing data and relevant metadata of the isolates, including MICs for meropenem, imipenem, polymyxin B and tigecycline, were provided by CHINET. Metadata for isolates from both collections are presented in Supplementary Data 1.

### Whole-genome sequencing and publicly available sequences

Whole-genome sequencing was performed for all CREC isolates ( $n = 168$ ) for the initial collection. Briefly, single colonies from an overnight agar plate were cultured in 4 ml of LB broth at 37 °C for 16 h, and genomic DNA was extracted using a Bacterial DNA Kit D3350

(Omega BioTek, USA). Sequencing was conducted by Novogene (Beijing, China) using an Illumina Novaseq 6000 platform (Illumina, San Diego, CA, USA). Five isolates were also subjected to whole-genome sequencing using the long-read MinION Sequencer (Nanopore; Oxford, UK) by Novogene (Beijing, China). These five isolates (E22, 18-4, 20-16, 19-7, and 20-20; Table S1) belong to the B5/H24RxC clone and they were selected based on their genomic differences, representing the range of diversity as suggested by their positions in the phylogeny.

For the analysis of a resistant clone of *E. coli* ST410 identified in this study, Enterobase (<http://enterobase.warwick.ac.uk/>) was searched for *E. coli* ST410 genomes (accessed on 13 Jan 2022) and only genomes with relevant metadata (year of collection, country and source type) and availability of raw reads were included ( $n = 790$ ). *E. coli* ST410 genomes used by a previous study<sup>15</sup> that were not present in the Enterobase data set ( $n = 84$ ) were also included. SRA and ENA accession numbers were extracted and raw reads for the genomes were downloaded from the European Nucleotide Archive. Metadata for the genomes are presented in Supplementary Data 2.

### Genome assembly, annotation, and characterisation

Illumina sequence reads were trimmed and assembled with Shovill v1.1.0 with default settings (<https://github.com/tseemann/shovill>) and assemblies were assessed for contamination and completeness using QUAST v5.0.2, CheckM v1.1.3 and Centrifuge v1.0.4<sup>42–44</sup>. Assemblies with a genome size greater than six million base pairs, an N50 smaller than 15,000 or a genome contamination greater than 2% were excluded from further analysis. A de novo hybrid of assembly of both Illumina reads and Nanopore reads was carried out using Unicycler v0.4.8 with default settings<sup>45</sup>. Prokka v1.14.0 was used to annotate the genome sequences<sup>46</sup>. Acquired ARGs and mutations that confer resistance were identified using abritAMR v1.0.13, an ISO-certified bioinformatics platform for genomics-based bacterial AMR detection<sup>47</sup>. Plasmid replicons, virulence factors and serotypes were identified with ABRicate v1.0.1 (<https://github.com/tseemann/abricate>), using the Resfinder<sup>48</sup>, Plasmidfinder<sup>49</sup>, Ecoli-vf and EcoH<sup>27</sup> databases (updated on 15 September 2022) with default parameters. Multilocus sequence types (MLSTs) were determined using MLST v2.19.0 (<https://github.com/tseemann/mlst>) with the “ecoli” scheme. *fimH* types were identified with FimTyper v1.0<sup>50</sup>.

Protein sequences of GyrA, GyrB, ParC, ParE, FtsI (PBP3), and FyuA were obtained from all genomes by performing a TBLASTN query of a representative protein sequence for each of the six proteins. Amino acid substitutions in the quinolone resistance-determining regions (QRDRs; in GyrA, GyrB, ParC and ParE) and insertion in FtsI were identified by comparing their protein sequences with the corresponding protein sequences from quinolone- and fluoroquinolone-susceptible *E. coli* K-12 MG1655 GenBank: [NC\\_000913.3](https://www.ncbi.nlm.nih.gov/nuccore/NC_000913.3).

### Phylogenetic analysis of *E. coli* ST410

To construct a global phylogeny for ST410 *E. coli*, 956 genomes (109 from this study and 847 publicly available genomes that passed quality control) were used to generate a full-length whole-genome alignment using Snippy v4.6.0 (<https://github.com/tseemann/snippy>) with ST410 isolate YD786 (GenBank accession [CP013112.1](https://www.ncbi.nlm.nih.gov/nuccore/CP013112.1)) as the reference. The full-length whole-genome alignment was cleaned with the snippy-clean function and then used as an input to Gubbins v2.4.1<sup>51</sup> for identifying and filtering regions of homologous recombination. Variant sites in the alignment were extracted using SNP-sites v2.5.1<sup>52</sup> and a maximum-likelihood (ML) phylogenetic tree was reconstructed using IQ-tree<sup>53</sup> accounting for constant sites in the alignment and run for 1000 bootstraps using the extended model selection function. The resulting tree was annotated in iTOL<sup>54</sup>. The phylogeny of the ST410 CREC isolated from a Chinese children’s hospital with potential outbreaks was

also constructed with the complete genome of isolate 19-7 as the reference using the methods described above.

Analysis of population structure within the ML phylogeny was conducted using Fastbaps v1.0 to identify major clusters under default parameters<sup>32</sup>. Pairwise SNP distances between isolate genomes were calculated using snp-dists v0.8.2 (<https://github.com/tseemann/snp-dists>) using the recombination removed alignment from Gubbins as input.

### Comparative genomic analysis of B4/H24RxC and B5/H24RxC clones

Prokka-annotated sequences of both B4/H24RxC ( $n = 214$ ) and B5/H24RxC ( $n = 174$ ) were analysed with Panaroo v1.2.7<sup>55</sup> under default settings to infer core and pangenomes. Clone association analysis of the genes in the pangenome matrix was performed with Scoary v1.6.16<sup>56</sup> with a maximum Benjamini–Hochberg adjusted  $P$  value of 1E-30.

Using the method described above, a smaller scale core-genome SNP phylogeny was reconstructed only for the genomes in these two clones with the complete genome of a previously reported isolate 020026 (B4/H24RxC; Genbank: [CP034954](https://www.ncbi.nlm.nih.gov/nuccore/CP034954) to [CP034958](https://www.ncbi.nlm.nih.gov/nuccore/CP034958)) used as the reference<sup>16</sup> and isolate A8 (B3/H24Rx) from this study as outgroup. SNPs specific to the B5/H24RxC MDR clone were identified by feeding the entire SNP matrix of all genomes of both clones into Scoary v1.6.16<sup>56</sup> using the same settings as described above. Gubbins v2.4.1<sup>51</sup> was used to determine whether the clone-specific genes and SNPs were due to recombination. The exact sequence locations for phage regions in the complete genomes of reference strains 19-7 and 020026 were identified using the PHASTER server<sup>57</sup>.

Plasmid sequences were examined and manually annotated in Gene Construction Kit v4.5.1 (Textco Biosoftware, Raleigh, USA). Plasmid replicons were initially identified with the PlasmidFinder database as described above, and PubMLST was used to sub-type F-type replicons (<https://pubmlst.org/organisms/plasmid-mlst>). Insertion sequences were identified using the ISFinder database (<https://isfinder.biotoul.fr/>).

The sequence of the O-antigen Onovel1 identified in the majority of B5/H24RxC isolates was used to perform a web nucleotide BLAST. The top match was a previously reported O-antigen Ogn5<sup>27</sup>. Pairwise alignment and visualisation of the two sequences were carried out using clinker v0.0.28<sup>58</sup>.

### Coalescent analysis of *E. coli* ST410

Treemmer v0.3<sup>59</sup> was used to reduce the ST410 global phylogeny to 500 genomes, maintaining ~95% of the original genetic diversity. The subsampled genomes had also maintained the diversity in year of isolation and country of origin (Fig. S4). A time-calibrated phylogeny for ST410 was reconstructed using the SNP alignment data generated from the raw sequence data of the selected 500 genomes as described above. Each sequence in the alignment was annotated with the year of isolation. The presence of a temporal signal in the data was investigated by inferring linear relationship between root-to-tip distances of the phylogenetic branches and the year of sample isolation using TempEst v1.5.3<sup>60</sup>, which revealed a correlation coefficient of  $R^2 = 0.32$ . Coalescence-based analysis was performed with BEAST v2.6.6<sup>61</sup>. In order to identify the most suitable model, analyses were performed using different substitution models (GTR and HKY), strict and relaxed molecular clock, and different demographic models including Bayesian Skyline, constant population and exponential population. Model selection was performed using Nested Sampling<sup>62</sup> v1.1.0 within the BEAST2 package with a particle count of 1, sub chain length of 5000, and Epsilon of  $1.0 \times 10^{-12}$ . The best fit model was estimated to be a GTR model and a relaxed molecular clock with a Bayesian Skyline population model (Table S4). Three replicates for this model were run for 700 million Markov chain Monte Carlo (MCMC) iterations, sampling every

50,000 states. Log files were combined with a 10% burn-in using LogCombiner v2.6.6 and assessed for convergence by checking the effective sample size (ESS, >200 for each parameter) using Tracer v1.7.2. A maximum clade credibility tree summarising the posterior sample of trees in the combined MCMC runs was generated with TreeAnnotator v2.6.6. The resulting tree was annotated and visualised with iTOL<sup>54</sup> and ggtree v3.2.1<sup>63</sup>.

### Growth curves and qPCR-based competition assays

An optical growth analyser (BioTek Epoch2, USA) was used to monitor the growth rate of ST410 strains of B4/H24RxC and B5/H24RxC clones. Based on their availability and their genetic characteristics, B4/H24RxC isolates (005828 and 045869) and B5/H24RxC isolates (18-13, 19-3, 19-17, 20-11, 20-25, and 20-30) were selected for the assay. Briefly, overnight cultures in LB were adjusted to a turbidity equivalent to that of a 0.5 McFarland standard and then inoculated 1:1000 in fresh LB, half strength LB or 1/10 strength LB. For each strain, 300  $\mu$ l of inoculated medium was added into wells of the microplate in triplicate. Fresh medium was also added to three wells acting as blank controls. Cultures were incubated at 37 °C with continuous shaking for 24 h and OD<sub>600</sub> was measured every 30 min. Growth curves for each strain during the exponential phase were analysed with the GraphPad Prism 9 software. The doubling time was calculated from growth curve using fitted curve of sigmoid function with a Python script available at [https://github.com/huoww07/calulate\\_bacteria\\_doubling\\_time](https://github.com/huoww07/calulate_bacteria_doubling_time). Strain ATCC 25922 was included as a growth control in the assay. Two-tailed unpaired Student's *t* test was employed to measure the difference between the doubling time of the two clones.

The growth dynamics of both clones were further investigated using a qPCR-based head to head competition assay in LB. In each competition group, a *yodB* positive isolate of the B4/H24RxC clone was in competition with a *fyuA* positive isolate of the B5/H24RxC clone. For instance, overnight cultures of isolate 005828 (B4/H24RxC) and 18-13 (B5/H24RxC) in LB were adjusted to a turbidity equivalent to that of a 0.5 McFarland standard and equal amount of the adjusted cultures were inoculated 1:200 into the same tube of fresh LB. Culture was incubated at 37 °C with continuous shaking for 24 h and samples were taken at hour 0, 4, 8, 12, and 24 for DNA extraction using the Bacterial DNA kit D3350 (Omega BioTek, USA). qPCR was performed using the HiScript III RT SuperMix for qPCR kit (Vazyme, China) on a Roche LightCycler PCR instrument (Roche, Sweden) using *yodB* and *fyuA* primers. Relative amount of the genes was calculated using 16S rRNA as the reference. The 2 <sup>$\Delta\Delta C_t$</sup>  method was used to determine the ratio of *fyuA* to *yodB*. The primers used were 16srRNA-F: GGAAGAAGCTTGCT TCTTTGCTGAC and 16srRNA-R: AGCCCGGGGATTTACATCTGACT TA; *yodB*-F: GGTGGACAATAGAAGGAT and *yodB*-R: TTATCTGCT-GATGCGAGAA; *fyuA*-F: CAGTAGGCACGATGTTGTA and *fyuA*-R: GCTATCCGCAGGCTATATG.

### Virulence assay

The wax moth (*Galleria mellonella*) larvae assay was based on a previously described method<sup>16,64</sup>. The larvae (~35 days after hatching and of 300  $\pm$  50 mg) were purchased from Huiyude Biotechnology (Tianjin, China). B4/H24RxC isolates (005828 and 045869) and B5/H24RxC isolates (18-13, 19-3, 19-17, 20-11, 20-25 and 20-30) were assessed for their virulence. Hypervirulent *Klebsiella pneumoniae* strain K1088<sup>65</sup> and hypervirulent *Acinetobacter baumannii* strain AB5075<sup>66</sup> were used as positive controls while *Acinetobacter baumannii* ATCC 19606 and PBS were used as negative controls. Single bacterial colonies were used to inoculate 5 ml of LB and incubated at 37 °C with 200 rpm shaking for about 6 to 8 h to reach the exponential phase. Cultures then were adjusted to a turbidity equivalent to 10<sup>8</sup> CFU/ml in PBS as pre-determined by plate count and further diluted 1:10 to obtain cultures of 10<sup>7</sup> CFU/ml. For each strain, three groups of larvae (*n* = 10 in each

group) were injected with 20  $\mu$ l of aliquots of diluted culture via the last left proleg using a Disposable Sterile Insulin Syringe U-100 (B.Braun, Germany). The infected larvae were kept at 37 °C during the course of the assay and survival of the larvae was recorded at hour 6, 12, 18, 24, 36, 48 and 72. Survival curves were generated and analysed using GraphPad Prism 9 software.

### Iron source growth assay

The growth of B5/H24RxC isolates (18-13, 19-3, 19-17, 20-11, 20-25, and 20-30) were evaluated under different iron sources on LB agar plates using B4/H24RxC isolates (005828 and 045869) as controls as described previously with modifications<sup>16,67</sup>. Briefly, the MIC of 2'-dipyridyl (DIP; Macklin, China) was determined using a previously described method<sup>16</sup> and all the isolates had an MIC of 300  $\mu$ M. For each isolate, a single colony from an overnight LB agar plate was cultured in LB broth containing 200  $\mu$ M DIP for 6 h to limit the growth of the isolate. Bacterial cells were then collected by centrifugation at 4000  $\times$  *g*, washed and resuspended in PBS to a turbidity equivalent to that of a 0.5 McFarland standard. About 10<sup>5</sup> CFU of each isolate were then spread onto LB agar plates containing DIP at the MIC (300  $\mu$ M for all isolates). Iron sources (10  $\mu$ l) of different concentrations including bovine serum albumin (10 mg/ml, 25 mg/ml, 50 mg/ml), FeCl<sub>2</sub> (1 mM, 5 mM, 10 mM), hemin (0.01 mM, 0.1 mM, 1 mM), haemoglobin (1 mg/ml, 10 mg/ml, 50 mg/ml), holo-transferrin (10 mg/ml, 25 mg/ml, 50 mg/ml), and lactoferrin (10 mg/ml, 25 mg/ml, 50 mg/ml) were spotted directly onto the plates and were incubated 48 to 72 h at 37 °C. Plates without iron sources added were used as negative control. The growth of bacteria was detected by visual inspection. All iron sources were purchased from Macklin, China.

### Biofilm formation assays

Biofilm formation assays were performed on the strains above following a previously described method<sup>31</sup>. For each strain, bacterial cells harvested from overnight culture were resuspended in PBS to a turbidity equivalent to that of a 0.5 McFarland standard and further inoculated 1:100 into fresh LB medium. Aliquots (200  $\mu$ l, *n* = 12) of the inoculated LB were added into a 96-well polystyrene microplate and were incubated at 37 °C for 24 h. The microplate was then washed three times with PBS and air-dried for 30 mins. Biofilms in the wells were stained with 250  $\mu$ l 1% (w/v) crystal violet for 20 mins at room temperature. The stained wells were again washed with distilled water to remove unbound stain and allowed to dry for 30 mins. The stained biofilms were treated with 200  $\mu$ l 33% acetic acid and OD<sub>570</sub> was recorded using a fluorometer plate reader (BioTek Epoch2, USA). *Acinetobacter baumannii* ATCC 27853 and fresh LB were used as positive and negative controls. Optical density cutoff (OD<sub>c</sub>) was calculated as the average OD of the negative control plus three times the standard deviation of the negative control as proposed previously<sup>30,31</sup>. An average OD value  $\leq$  OD<sub>c</sub>, >OD<sub>c</sub> and  $\leq 2 \times$  OD<sub>c</sub>, >2  $\times$  OD<sub>c</sub> and  $\leq 4 \times$  OD<sub>c</sub>, and >4  $\times$  OD<sub>c</sub> indicates non-, weak, moderate and strong biofilm production, respectively.

### Statistics and reproducibility

Statistical analyses were performed using GraphPad Prism (version 9.3.1). Level of significance between two groups was assessed with two-tailed unpaired Student's *t* test. Survival analysis for wax moth larvae infected with different bacterial strains was performed using Log-rank (Mantel–Cox) test with GraphPad Prism (version 9.3.1). All information on sample sizes and statistics can be found in the figure legends and the reporting summary. No statistical method was used to predetermine sample size. No data were excluded from the analyses. Wax moth larvae were randomly allocated into different groups. Other experiments were not randomised. The investigators were not blinded to allocation during experiments and outcome assessment.

### Data visualisation

Bar charts, violin plots and line charts were generated using GraphPad Prism (version 9.3.1). Maps were generated using ggplot2<sup>68</sup> and sf<sup>69</sup> packages. Phylogenetic trees were annotated and visualised with iTOL<sup>54</sup>, ggtree v3.2.1<sup>63</sup> or Phandango<sup>70</sup>. Pairwise alignment and visualisation of the genomic sequences were carried out using clinker v0.0.28<sup>58</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Sequence data and genome assemblies generated in this study have been submitted to GenBank under the BioProject [PRJNA931432](#) and [PRJNA951454](#). The individual Illumina sequence read accession numbers of ST410 isolates are listed in Supplementary Data 2. The accession numbers for the nanopore sequenced isolates in this study can be found in Table S1. Source data are provided with this paper.

### References

- Murray C. J. L. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
- Hickman R. A. et al. Exploring the antibiotic resistance burden in livestock, livestock handlers and their non-livestock handling contacts: a one health perspective. *Front. Microbiol.* **12**, 651461 (2021).
- Magiorakos, A. P. et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* **18**, 268–281 (2012).
- Papp-Wallace, K. M., Endimiani, A., Taracila, M. A. & Bonomo, R. A. Carbapenems: past, present, and future. *Antimicrob. Agents Chemother.* **55**, 4943–4960 (2011).
- Meletis, G. Carbapenem resistance: overview of the problem and future perspectives. *Ther. Adv. Infect. Dis.* **3**, 15–21 (2016).
- Cui, X., Zhang, H. & Du, H. Carbapenemases in Enterobacteriaceae: detection and antimicrobial therapy. *Front. Microbiol.* **10**, 1823 (2019).
- Nordmann, P., Naas, T. & Poirel, L. Global spread of carbapenemase-producing Enterobacteriaceae. *Emerg. Infect. Dis.* **17**, 1791–1798 (2011).
- Canton, R. et al. Rapid evolution and spread of carbapenemases among Enterobacteriaceae in Europe. *Clin. Microbiol. Infect.* **18**, 413–431 (2012).
- Li, F. et al. Genetic characterization of carbapenem-resistant *Escherichia coli* from China, 2015–2017. *BMC Microbiol.* **21**, 248 (2021).
- Grundmann, H. et al. Occurrence of carbapenemase-producing *Klebsiella pneumoniae* and *Escherichia coli* in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. *Lancet Infect. Dis.* **17**, 153–163 (2017).
- Wu, W. et al. NDM metallo- $\beta$ -lactamases and their bacterial producers in health care settings. *Clin. Microbiol. Rev.* **32**, e00115–e00118 (2019).
- Martirosov, D. M. & Lodise, T. P. Emerging trends in epidemiology and management of infections caused by carbapenem-resistant Enterobacteriaceae. *Diagn. Microbiol. Infect. Dis.* **85**, 266–275 (2016).
- Peirano G. et al. Genomic epidemiology of global carbapenemase-producing *Escherichia coli*, 2015–2017. *Emerg. Infect. Dis.* **28**, 924–931 (2022).
- Zhang, R. et al. Nationwide surveillance of clinical carbapenem-resistant Enterobacteriaceae (CRE) strains in China. *EBioMedicine* **19**, 98–106 (2017).
- Roer L. et al. *Escherichia coli* sequence type 410 is causing new international high-risk clones. *MSphere*; **3**, e00337–18 (2018).
- Feng, Y. et al. Key evolutionary events in the emergence of a globally disseminated, carbapenem resistant clone in the *Escherichia coli* ST410 lineage. *Commun. Biol.* **2**, 322 (2019).
- Chen L., Peirano G., Kreiswirth B. N., Devlin R., Pitout J. D. D. Acquisition of genomic elements were pivotal for the success of *Escherichia coli* ST410. *J. Antimicrob. Chemother.* **77**, 3399–3407 (2022).
- Alm, R. A., Johnstone, M. R. & Lahiri, S. D. Characterization of *Escherichia coli* NDM isolates with decreased susceptibility to aztreonam/avibactam: role of a novel insertion in PBP3. *J. Antimicrob. Chemother.* **70**, 1420–1428 (2015).
- Zhang Y., Kashikar A., Brown C. A., Denys G., Bush K. Unusual *Escherichia coli* PBP 3 insertion sequence identified from a collection of carbapenem-resistant Enterobacteriaceae tested in vitro with a combination of ceftazidime-, ceftaroline-, or aztreonam-avibactam. *Antimicrob. Agents Chemother.* **61**, e00389–17 (2017).
- Sun, P. et al. Characterization Of *bla*<sub>NDM-5</sub>-positive *Escherichia coli* prevalent in a University Hospital In Eastern China. *Infect. Drug Resist.* **12**, 3029–3038 (2019).
- Al-Farsi, H. M. et al. Clinical and molecular characteristics of carbapenem non-susceptible *Escherichia coli*: a nationwide survey from Oman. *PLoS One* **15**, e0239924 (2020).
- Mahazu, S. et al. Possible dissemination of *Escherichia coli* sequence type 410 closely related to B4/H24RxC in Ghana. *Front. Microbiol.* **12**, 770130 (2021).
- Nadimpalli, M. L. et al. *Escherichia coli* ST410 among humans and the environment in Southeast Asia. *Int. J. Antimicrob. Agents.* **54**, 228–232 (2019).
- Gorrie, C. L. et al. Key parameters for genomics-based real-time detection and tracking of multidrug-resistant bacteria: a systematic analysis. *Lancet Microbe* **2**, e575–e583 (2021).
- Schubert, S., Rakin, A. & Heesemann, J. The *Yersinia* high-pathogenicity island (HPI): evolutionary and functional aspects. *Int. J. Med. Microbiol.* **294**, 83–94 (2004).
- Carniel, E. The *Yersinia* high-pathogenicity island: an iron-uptake island. *Microbes Infect.* **3**, 561–569 (2001).
- Ingle D. J. et al. In silico serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Microb. Genom.* **2**, e000064 (2016).
- Iguchi A., von Mentzer A., Kikuchi T., Thomson N. R. An untypeable enterotoxigenic *Escherichia coli* represents one of the dominant types causing human disease. *Microb. Genom.* **3**, e000121 (2017).
- Schubert, S., Picard, B., Gouriou, S., Heesemann, J. & Denamur, E. *Yersinia* high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. *Infect. Immun.* **70**, 5335–5337 (2002).
- Stepanovic, S. et al. Quantification of biofilm in microtiter plates: overview of testing conditions and practical recommendations for assessment of biofilm production by staphylococci. *Apmis* **115**, 891–899 (2007).
- Hassan, A. et al. Evaluation of different detection methods of biofilm formation in the clinical isolates. *Braz. J. Infect. Dis.* **15**, 305–311 (2011).
- Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* **47**, 5539–5549 (2019).
- Hu, J., Kan, B. A., Liu, Z. H. & Yu, S. Y. Enterotoxigenic *Escherichia coli* isolated from Chinese diarrhea patients with high-pathogenicity island of *Yersinia* is involved in synthesis of siderophore yersiniabactin. *World J. Gastroenterol.* **11**, 5816–5820 (2005).

34. Karch, H. et al. A genomic island, termed high-pathogenicity island, is present in certain non-O157 shiga toxin-producing *Escherichia coli* clonal lineages. *Infect. Immun.* **67**, 5994–6001 (1999).
35. Marschall, J. et al. Both host and pathogen factors predispose to *Escherichia coli* urinary-source bacteremia in hospitalized patients. *Clin. Infect. Dis.* **54**, 1692–1698 (2012).
36. Koh, E. I., Hung, C. S. & Henderson, J. P. The yersiniabactin-associated ATP binding cassette proteins YbtP and YbtQ enhance *Escherichia coli* fitness during high-titer cystitis. *Infect. Immun.* **84**, 1312–1319 (2016).
37. Robinson, A. E., Heffernan, J. R. & Henderson, J. P. The iron hand of uropathogenic *Escherichia coli*: the role of transition metal control in virulence. *Future Microbiol.* **13**, 813–829 (2018).
38. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13**, 269–284 (2015).
39. Stenutz, R., Weintraub, A. & Widmalm, G. The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol. Rev.* **30**, 382–403 (2006).
40. Zong, Z. Y., Fenn, S., Connor, C., Feng, Y. & McNally, A. Complete genomic characterization of two *Escherichia coli* lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital. *J. Antimicrob. Chemother.* **73**, 2340–2346 (2018).
41. Martinez-Gil M. et al. YeeJ is an inverse autotransporter from *Escherichia coli* that binds to peptidoglycan and promotes biofilm formation. *Sci. Rep.* **7**, 11326 (2017).
42. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* **26**, 1721–1729 (2016).
43. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
44. Mikheenko, A., Prijibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018).
45. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, e1005595 (2017).
46. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
47. Sherry, N. L. et al. An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. *Nat. Commun.* **14**, 60 (2023).
48. Zankari, E. et al. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).
49. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
50. Roer, L. et al. Development of a web tool for *Escherichia coli* subtyping based on *fimH* alleles. *J. Clin. Microbiol.* **55**, 2538–2543 (2017).
51. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
52. Page, A. J. et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* **2**, e000056 (2016).
53. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
54. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
55. Tonkin-Hill, G. et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
56. Brynildsrud, O., Bohlin, J., Scheffer, L. & Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* **17**, 238 (2016).
57. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
58. Gilchrist, C. L. M. & Chooi, Y. H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
59. Menardo, F. et al. Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics* **19**, 164 (2018).
60. Rambaut, A., Lam T. T., Carvalho L. M., Pybus O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**: vew007 (2016).
61. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
62. Russel, P. M., Brewer, B. J., Klaere, S. & Bouckaert, R. R. Model selection and parameter inference in phylogenetics using nested sampling. *Syst. Biol.* **68**, 219–233 (2019).
63. Yu, G. C., Smith, D. K., Zhu, H. C., Guan, Y. & Lam, T. T. Y. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
64. Jonsson, R., Struve, C., Jenssen, H. & Kroghfelt, K. A. The wax moth *Galleria mellonella* as a novel model system to study Enteroga-gregative *Escherichia coli* pathogenesis. *Virulence* **8**, 1894–1899 (2017).
65. Gu, D. X. et al. A fatal outbreak of ST11 carbapenem-resistant hypervirulent *Klebsiella pneumoniae* in a Chinese hospital: a molecular epidemiological study. *Lancet Infect. Dis.* **18**, 37–46 (2018).
66. Jacobs, A. C. et al. AB5075, a highly virulent isolate of *Acinetobacter baumannii*, as a model strain for the evaluation of pathogenesis and antimicrobial treatments. *Mbio* **5**, e01076–14 (2014).
67. Cernat, R. C. & Scott, K. P. Evaluation of novel assays to assess the influence of different iron sources on the growth of *Clostridium difficile*. *Anaerobe* **18**, 298–304 (2012).
68. Wickham H. ggplot2: elegant graphics for data analysis. *Use R*: 1–212 (2009).
69. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R. J.* **10**, 439–446 (2018).
70. Hadfield, J. et al. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* **34**, 292–293 (2018).

## Acknowledgements

This work was undertaken as part of the DETECTIVE research project funded by the National Natural Science Foundation of China (8181101332) and the Medical Research Council (MR/S013660/1). W.v.S. was also supported by a Wolfson Research Merit Award (WM160092). We thank Dr. Andries J. van Tonder for guidance on the bioinformatics analysis.

## Author contributions

C.Z., M.A.H., A.M., W.v.S., and X.B. initiated and designed the study. X.B., Y.G., R.A.M., and E.L.D. performed the experiments and analyses. B.L. and L.Y. performed experiments. Y.G., B.L., J.L., N.H., S.S., and Y.L. collected and provided the bacterial isolates. X.B. wrote the manuscript with input from C.Z., M.A.H., A.M., W.v.S., Y.G., and R.A.M. All the authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-43854-3>.

**Correspondence** and requests for materials should be addressed to Mark A. Holmes or Chao Zhuo.

**Peer review information** *Nature Communications* thanks the anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024