Chromosomal integration of *bla*$_{CTX-M}$ genes in diverse *Escherichia coli* isolates recovered from river water in Japan

Ryota Gomi[a*], Masaki Yamamoto[b], Michio Tanaka[b], Yasufumi Matsumura[b]

[a]Department of Environmental Engineering, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, 615-8540, Kyoto, Japan

[b]Department of Clinical Laboratory Medicine, Graduate School of Medicine, Kyoto University, 54 Shogoin-kawahara-cho, Sakyo-ku, 606-8507, Kyoto, Japan

[*]Corresponding author

E-mail address: gomi.ryota.34v@kyoto-u.jp; Tel: +81-75-383-3354; Fax: +81-75-383-3358

Abstract

Occurrence of extended-spectrum β-lactamase (ESBL)-producing *Escherichia coli* (ESBLEC) in environmental waters is of great concern. However, unlike clinical ESBLEC, their genetic characteristics, in particular the genetic contexts of ESBL genes, are not well understood. In this study, we sequenced and analyzed the genomes of CTX-M-producing *E. coli* isolates recovered from river water to fully characterize the genetic contexts of $bla_{\text{CTX-M}}$ genes. Among the 14 isolates with completed genomes, $bla_{\text{CTX-M}}$ genes were detected on the chromosome in nine isolates. All but one chromosomal $bla_{\text{CTX-M}}$ genes were associated with IS*Ecp1* and were carried on different transposition units ranging in size from 2,855 bp to 11,093 bp; the exception, $bla_{\text{CTX-M-2}}$, was associated with IS*CR1*. The remaining five isolates carried $bla_{\text{CTX-M}}$ genes on epidemic IncI1 plasmids of different sequence types (STs) (ST3, ST16, ST113, and ST167) (n = 4) or on an IncB/O/K/Z plasmid (n = 1). This study revealed that environmental *E. coli* carry $bla_{\text{CTX-M}}$ genes in diverse genetic contexts. Apparent high prevalence of chromosomal $bla_{\text{CTX-M}}$ potentially indicates that some *E. coli* can stably maintain $bla_{\text{CTX-M}}$ genes in environmental waters, though further studies are needed to confirm this.

Keywords

Escherichia coli, $bla_{\text{CTX-M}}$, genome sequencing, antibiotic resistance, river water, one health

33 **1. Introduction**

34 Occurrence of antibiotic-resistant bacteria (ARB) in surface waters is a public health concern.
35 Among ARB, extended-spectrum-β-lactamase (ESBL)-producing *Escherichia coli* (ESBLEC)
36 represent a major threat to public health, because ESBLEC can be important causes of
37 community-onset infections for which treatment options are limited (Pitout, 2010; Pitout and
38 Laupland, 2008). ESBL enzymes are divided into several families (e.g., CTX-M and ESBL
39 variants of TEM and SHV), among which CTX-M enzymes have been recognized as the most
40 common ESBL type (Castanheira et al., 2021). Previous studies have reported the
41 predominance of CTX-M ESBLs among environmental *E. coli*. For example, CTX-M-14(-
42 like) was prevalent among ESBLEC in surface waters in China, Japan, and Korea, and CTX-
43 M-15 was prevalent among ESBLEC in surface waters in France, Germany, and Switzerland
44 (Falgenhauer et al., 2021; Girlich et al., 2020; Jang et al., 2013; Liu et al., 2018; Miyagi and
45 Hirai, 2019; Zurfluh et al., 2013).

46 ESBL genes are often found on plasmids, which facilitates their horizontal spread (Castanheira
47 et al., 2021). However, some previous studies also detected ESBL genes on chromosomes and
48 suggested that integration of ESBL genes into chromosomes could contribute to the
49 stabilization and maintenance of ESBL genes in the absence of antibiotic selection pressure
50 (Hirai et al., 2013; Nair et al., 2021; Rodriguez et al., 2014). Thus, information on the location
51 (chromosome or plasmid) of ESBL genes is important to predict the mobility and stability of

52 these genes. Although, previous studies have reported the occurrence of ESBLEC in
53 environmental waters, limited information exists regarding the locations and genetic contexts
54 of ESBL genes in environmental ESBLEC (Falgenhauer et al., 2021; Girlich et al., 2020; Liu
55 et al., 2018). Here, we completely sequenced the genomes of CTX-M-producing *E. coli* isolates
56 recovered from river water in Japan, aiming to elucidate the locations and genetic contexts of
57 the $bla_{\text{CTX-M}}$ genes in environmental ESBLEC. The genetic contexts of chromosomal $bla_{\text{CTX-M}}$
58 genes in the genomes of environmental *E. coli* retrieved from the NCBI Reference Sequence
59 Database (RefSeq) were also characterized.

60

61 **2. Materials and methods**

62 **2.1. Bacterial isolates.**

63 We previously collected 531 *E. coli* isolates from the Yamato River in Japan between 2011
64 and 2013 and detected 18 isolates with $bla_{\text{CTX-M}}$ genes (Gomi et al., 2014; Gomi et al., 2017).
65 Briefly, we collected a total of 27 water samples from 10 sites in the Yamato River in August
66 and November of 2011, February of 2012, and October of 2013 (note that water samples were
67 not collected from all sampling sites on each sampling occasion). The samples were processed
68 using the membrane filter method with XM-G agar (Nissui, Tokyo, Japan) to obtain *E. coli*
69 isolates. In total, 531 *E. coli* isolates were obtained, and a subset of isolates (n = 155) was
70 selected and subjected to Illumina short-read sequencing. Among the sequenced isolates, we

71 detected 18 isolates that carried bla_{CTX-M} genes. These 18 isolates were analyzed as described

72 below.

73 **2.2. Antibiotic susceptibility testing.**

74 Antibiotic susceptibility was assessed by microdilution using the dry plate Eiken assay (Eiken,

75 Tokyo, Japan) as described previously (Gomi et al., 2017) (see **Table S1** for the antimicrobial

76 agents used). The results were interpreted according to CLSI criteria (CLSI, 2020) and

77 EUCAST epidemiological cutoff (ECOFF) values (http://www.eucast.org). ESBL production

78 was confirmed following the CLSI guidelines (CLSI, 2020). A detailed description of antibiotic

79 susceptibility testing is provided in the Supplementary Materials and methods.

80 **2.3. Genome sequencing and assembly.**

81 DNA was extracted from each isolate using a DNeasy PowerSoil Pro Kit (Qiagen, Hilden,

82 Germany). The short-read sequencing library was prepared using the Illumina DNA Prep

83 (Illumina, San Diego, CA), and the library was sequenced on an Illumina NovaSeq 6000 system

84 with 250-bp paired-end sequencing (although we obtained Illumina reads for the isolates in our

85 previous study (Gomi et al., 2017), the Illumina reads had low depth/uneven coverage and

86 hence we sequenced the isolates again with Illumina technology in the present study). The

87 long-read sequencing library was prepared using the SQK-LSK109 kit (Oxford Nanopore

88 Technologies, Oxford, UK), and the prepared library was sequenced on the MinION with a

89 FLO-MIN106 flow cell.

90 Illumina short reads were subsampled using seqtk (v1.3, https://github.com/lh3/seqtk), and the
91 subsampled reads were trimmed with fastp (v0.20.0) (Chen et al., 2018). Oxford Nanopore
92 Technologies (ONT) long reads were filtered with Filtlong (v0.2.0,
93 https://github.com/rrwick/Filtlong) to remove reads with low quality or short length. Hybrid
94 assembly of Illumina short reads and ONT long reads was performed by employing a short-
95 read-first approach, i.e., Unicycler (v0.4.8) with the --no_correct option (Wick et al., 2017).
96 Hybrid assembly was also performed using a long-read-first approach, i.e., long-read assembly
97 by Flye (v2.9) (Kolmogorov et al., 2019) followed by long-read polishing using Medaka
98 (v1.4.4, https://github.com/nanoporetech/medaka) and short-read polishing using Polypolish
99 (v0.4.3) (Wick and Holt, 2022). Assembly graphs of both hybrid assemblies were visualized
100 using Bandage (v0.8.1) for each isolate (Wick et al., 2015), and a better assembly was chosen
101 as the final assembly.

102 **2.4. Genomic analysis.**

103 Genomes were annotated using the RAST server (Aziz et al., 2008), ISfinder (Siguier et al.,
104 2006), ResFinder 4.1 (Bortolaia et al., 2020), and the BLAST algorithm
105 (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Plasmid replicons were detected using
106 PlasmidFinder 2.1 (Carattoli et al., 2014). Relaxase types, mate-pair formation types, and
107 transferability of plasmids were determined using MOB-typer (Robertson and Nash, 2018).

108 IncI1 plasmids were annotated using plasmids R64, ColIb-P9, and R621a as references

109 (Carattoli et al., 2021).

110 **2.5. Characterization of environmental *E. coli* genomes with *bla*$_{CTX-M}$ genes in RefSeq.**

111 RefSeq *E. coli* genomes with an assembly level of "complete genome" (n = 1903) were

112 downloaded in January 2022. Metadata information, such as collection date and isolation

113 source, was extracted from the GenBank files using an in-house Python script, and genomes of

114 *E. coli* from environmental waters were retrieved (n = 53). Antibiotic resistance genes were

115 detected using ABRicate (https://github.com/tseemann/abricate), and genomes carrying *bla*$_{CTX-}$

116 $_{M}$ on the chromosome were analyzed as described above.

117 **2.6. Accession numbers.**

118 The complete genomes and sequence read files were deposited in GenBank and the NCBI SRA

119 under BioProject PRJNA800231 (also see **Table S2** for the accession number of each replicon

120 for the completed genomes and **Table S1** for SRA accession numbers of the remaining

121 genomes).

122

123 **3. Results and discussion**

124 **3.1. Basic characteristics of *E. coli* isolates sequenced in this study.**

125 All 18 *bla*$_{CTX-M}$-carrying isolates characterized in the present study were non-susceptible/non-

126 wild type to ampicillin, aztreonam, cefazolin, cefepime, cefotaxime, and cefpodoxime, while

127 all were susceptible/wild type to amoxicillin-clavulanic acid, piperacillin-tazobactam,
128 imipenem, meropenem, amikacin, and colistin. Non-susceptibility/non-wild type rates for the
129 remaining antibiotics ranged from 6% (fosfomycin) to 89% (ampicillin-sulbactam) (**Table S1**).
130 ESBL production was confirmed in all the isolates. The 18 isolates were highly diverse,
131 comprising 16 different sequence types (STs), namely ST38 (n = 2), ST69 (n = 2), and a single
132 isolate from each of ST10, ST46, ST117, ST216, ST354, ST457, ST744, ST1193, ST2003,
133 ST3107, ST6214, ST6215, ST6216, and ST6220 (**Table S1**). The $bla_{\text{CTX-M}}$ genes detected were
134 also diverse, namely $bla_{\text{CTX-M-14}}$ (n = 8), $bla_{\text{CTX-M-2}}$ (n = 2), $bla_{\text{CTX-M-15}}$ (n = 2), $bla_{\text{CTX-M-24}}$ (n = 2),
135 $bla_{\text{CTX-M-55}}$ (n = 2), $bla_{\text{CTX-M-1}}$ (n = 1), $bla_{\text{CTX-M-3}}$ (n = 1), and $bla_{\text{CTX-M-8}}$ (n = 1). One isolate
136 (KTa008) carried two copies of the $bla_{\text{CTX-M-14}}$ gene. In Japan, ESBLEC are frequently detected
137 among *E. coli* infections. For example, 22% of extraintestinal pathogenic *E. coli* isolates were
138 positive for one or more ESBL genes in a study conducted at 10 acute-care hospitals
139 (Matsumura et al., 2017a). Like in other countries, these ESBLEC infections in Japan have
140 been shown to be mainly due to CTX-M-producing ST131 (Matsumura et al., 2017b). However,
141 the ESBLEC isolates analyzed in the present study belonged to a wide variety of STs and none
142 belonged to ST131, which indicates that the clonal composition of ESBLEC between
143 environmental settings and clinical settings may be different in Japan.
144 Of the 18 isolates, the genomes of 14 could be completed. These genomes carried 0 to 8
145 plasmids (see **Table S2** for details). The genomes of the remaining four isolates could not be

146 completed, probably due to the presence of long repeats in the genomes. The location and
147 genetic contexts of the *bla*$_{CTX-M}$ genes were determined for the 14 isolates with completed
148 genomes. Of these 14, nine carried *bla*$_{CTX-M}$ genes on chromosomes and five carried *bla*$_{CTX-M}$
149 genes on plasmids.

150 **3.2. Genetic contexts of chromosomal *bla*$_{CTX-M}$ genes.**

151 In eight of the nine isolates which carried *bla*$_{CTX-M}$ genes on chromosomes, the *bla*$_{CTX-M}$ genes
152 were associated with IS*Ecp1* and carried on transposition units (TPU) of different sizes (2,855
153 bp to 11,093 bp) (**Figure 1**). These TPU were flanked by AT-rich 5-bp target site duplications
154 (TSD). None of the TPU were identical, and the insertion sites of the TPU were different in all
155 cases, indicating that different transposition events contributed to the chromosomal integration
156 of each TPU. Notably, the TPU in KFu023 contained multiple resistance genes, including
157 *bla*$_{CTX-M-15}$, *aac(6')-Ib-cr*, *bla*$_{OXA-1}$, and *ΔcatB3*. This seems to highlight the ability of IS*Ecp1* to
158 capture sequences of different lengths in different transposition events and thus simultaneously
159 mobilize antibiotic resistance genes of different origins (Partridge et al., 2018). In most cases,
160 these TPU were directly inserted into chromosomes. However, in KKa019, the TPU was part
161 of a 35,625 bp multiresistance region bracketed by two copies of IS*26* (annotated as IS*15DI* in
162 the ISFinder database), and the whole multiresistance region was flanked by 8-bp TSD
163 (CGTTGCCG) in the chromosome (**Figure 2a**). KTa008 carried two copies of *bla*$_{CTX-M-14}$,
164 which were separated by ~1,169 kbp in the chromosome. Analysis of the genetic contexts of

165 these *bla*$_{CTX-M-14}$ genes indicated transposition of an already chromosomally-located TPU and

166 an adjacent region into another chromosomal location (**Figure S1**). This type of 'recurrent'

167 transposition was previously reported in clinical ESBLEC, though the TPU and insertion sites

168 were different (Hamamoto and Hirai, 2019).

169 The above results indicate that analysis of sequences immediately upstream and downstream

170 of *bla*$_{CTX-M}$ (e.g., identification of IS*Ecp1* upstream and *orf477* downstream of *bla*$_{CTX-M}$) is not

171 sufficient to track the movement and evolution of *bla*$_{CTX-M}$, since TPU can be very long

172 (sometimes > 10,000 bp) or can be inserted into a large multiresistance region. This can be

173 overcome, for example, by determining the complete genomes as shown in the present study.

174 We searched for *E. coli* genomes in the public database which carry the same TPU in the same

175 chromosomal location as those identified in the present study. BLASTN searches were

176 performed against the nucleotide collection (nr/nt) database in January 2022 using the TPU and

177 surrounding sequences identified in the present study as query sequences. Three of the nine

178 TPU/insertion site combinations were found in the database, while the remaining six were not

179 found, potentially corresponding to novel integration events (**Table S3**).

180 In KOr019, *bla*$_{CTX-M-2}$ was associated with IS*CR1* and situated within a complex class 1 integron,

181 which is congruent with previous studies (Canton et al., 2012; Partridge et al., 2018). This

182 complex class 1 integron was closely related to a recently reported transposon Tn*7337* (**Figure**

183 **2b**) (Coppola et al., 2022). KOr019 carried the IRt (inverted repeat at *tni* end)-IS*6100*-IRt

184 structure, but this structure was truncated by IS*26* in Tn*7337*. The 17,747 bp structure from IRi

185 (inverted repeat at *intI*1 end) to IRt was inserted into the chromosome of KOr019, creating 5-

186 bp TSD of AAATG. This insertion seems to have been mediated by Tni proteins provided *in*

187 *trans*, since the *tni* genes in the complex class 1 integron were absent (Partridge et al., 2018).

188 **3.3. Genetic contexts of bla_{CTX-M} genes on plasmids.**

189 The five completed plasmids with bla_{CTX-M} genes belonged to IncI1 (n = 4) and IncB/O/K/Z (n

190 = 1) (**Figure 3a**). IncI1 plasmids were typed as four different STs, namely ST3, ST16, ST113,

191 and ST167, by IncI1 pMLST (Carattoli et al., 2014). TPU were inserted into plasmid backbones

192 in the IncI1/ST3, IncI1/ST16, and IncI1/ST167 plasmids. Interestingly, the IncI1/ST16 and

193 IncI1/ST167 plasmids carried the same TPU (except for insertion of IS*1* in the IncI1/ST16

194 plasmid) inserted into the same location. IncI1/ST167 is a single-locus variant of IncI1/ST16,

195 which indicates the TPU was inserted into the plasmid backbone before the ST diversification.

196 In the IncI1/ST113 plasmid, an IS26-ΔIS*10R*-$bla_{CTX-M-8}$-IS*26* structure was inserted into the

197 plasmid. BLASTN analysis identified IncI1 plasmids belonging to the same IncI1-STs and

198 carrying the same TPU/IS*26* structures inserted into the same locations from different countries

199 (e.g., MH847453 from France for $bla_{CTX-M-1}$-carrrying IncI1/ST3, CP056650 from the UK for

200 $bla_{CTX-M-55}$-carrrying IncI1/ST16, KY964068 from Portugal for $bla_{CTX-M-8}$-carrrying

201 IncI1/ST113, CP021208 from China for $bla_{CTX-M-55}$-carrrying IncI1/ST167), which indicates

202 circulation of epidemic CTX-M IncI1 plasmids in geographically distant countries.

203 Comparison of IncI1 plasmids revealed the replication region, the leading region, and the
204 transfer regions in the IncI1 reference plasmid R64 were well conserved in all IncI1 plasmids
205 detected in this study, while many genes in the stability region were lacking (**Figure 3b**). In
206 the IncB/O/K/Z plasmid, $bla_{\text{CTX-M-14}}$ was linked to ΔTn*1721* and a class 1 integron. Similar
207 genetic contexts were previously reported (e.g., JF701188).

208 **3.4. Chromosomal location of $bla_{\text{CTX-M}}$ in the RefSeq assemblies.**

209 Among the 53 RefSeq environmental *E. coli* genomes, 20 carried $bla_{\text{CTX-M}}$. Sixteen carried
210 $bla_{\text{CTX-M}}$ on plasmids and four carried $bla_{\text{CTX-M}}$ on chromosomes (**Table S4**). None of the 20
211 RefSeq genomes with $bla_{\text{CTX-M}}$ belonged to ST131, again indicating a potential difference in
212 clonal composition of ESBLEC in environmental settings and clinical settings. The four
213 genomes with chromosomal $bla_{\text{CTX-M}}$ belonged to ST38 (n = 2), ST130 (n = 1) and ST648 (n =
214 1) and originated from various countries (Japan, New Zealand, Norway, and Switzerland),
215 which indicates that diverse *E. coli* isolates carrying chromosomal $bla_{\text{CTX-M}}$ are also present in
216 environmental waters in other countries. In all four genomes, $bla_{\text{CTX-M}}$ was associated with TPU
217 ranging from 2,971 bp to 13,552 bp. A TPU containing multiple resistance genes besides
218 $bla_{\text{CTX-M}}$ and a TPU inserted into a multiresistance region were detected, again highlighting the
219 importance of these mechanisms in the emergence of multidrug resistant *E. coli* in the
220 environment (see **Figure S2** for the analysis of the genetic contexts of $bla_{\text{CTX-M}}$ in these
221 genomes).

222 ST38, which was detected in two RefSeq genomes containing chromosomal bla_{CTX-M}, was also

223 detected in two isolates with chromosomal bla_{CTX-M} sequenced in this study. Interestingly, ST38

224 was also found to be prevalent in some previous studies reporting chromosomal bla_{CTX-M}

225 (Guenther et al., 2017; Guiral et al., 2011; Rodriguez et al., 2014). The reason for the

226 prevalence of ST38 among *E. coli* with chromosomal bla_{CTX-M} is unknown. However, multiple

227 integration events seem to have been involved in the acquisition of chromosomal bla_{CTX-M} in

228 ST38 (see **Figure 1**, **Figure 2a**, and **Figure S2**), which potentially indicates ST38 has genetic

229 mechanisms that can promote chromosomal integration of bla_{CTX-M}.

230 **3.5. Study limitations.**

231 This study has some limitations. The *E. coli* isolates analyzed in the present study were

232 collected between 2011 and 2013 (i.e., ~10 years ago). Additionally, a limited number of

233 complete genomes were determined (n = 14), and the isolates were collected regionally, not

234 nationwide. However, we also included RefSeq environmental *E. coli* genomes from various

235 countries with various collection dates in our analysis, which might mitigate these limitations.

236 **4. Conclusions.**

237 Here, we demonstrated that the genetic contexts of bla_{CTX-M} genes in environmental *E. coli* are

238 highly diverse in terms of the associated mobile elements (various TPU, IS*CR1*, and IS*26*) and

239 integration sites. This study also revealed that the chromosomal locations of bla_{CTX-M} genes

240 were unexpectedly frequent in environmental ESBLEC in Japan. Chromosomal integration of

241 *bla*$_{CTX-M}$ might allow *E. coli* to stably maintain *bla*$_{CTX-M}$ in environmental waters, though further

242 studies are needed to confirm this.

243

244 **Acknowledgements**

245 We acknowledge the NGS core facility of the Genome Information Research Center at the

246 Research Institute for Microbial Diseases of Osaka University for the support in DNA

247 sequencing. Computations were partially performed on the NIG supercomputer at ROIS

248 National Institute of Genetics. This work was supported by the Environment Research and

249 Technology Development Fund (JPMEERF20205006) of the Ministry of the Environment,

250 Japan. This study was also supported in-part by the Program for the Development of Next-

251 generation Leading Scientists with Global Insight (L-INSIGHT), sponsored by the Ministry of

252 Education, Culture, Sports, Science and Technology (MEXT), Japan.

253

254

255 **References**

256 Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid

257 annotations using subsystems technology. BMC Genomics 2008; 9: 75.

258 Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for
259 predictions of phenotypes from genotypes. J Antimicrob Chemother 2020; 75: 3491-
260 3500.
261 Canton R, Gonzalez-Alba JM, Galan JC. CTX-M Enzymes: Origin and Diffusion. Front
262 Microbiol 2012; 3: 110.
263 Carattoli A, Villa L, Fortini D, Garcia-Fernandez A. Contemporary IncI1 plasmids involved in
264 the transmission and spread of antimicrobial resistance in Enterobacteriaceae. Plasmid
265 2021; 118: 102392.
266 Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, et al. In silico
267 detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence
268 typing. Antimicrob Agents Chemother 2014; 58: 3895-903.
269 Castanheira M, Simner PJ, Bradford PA. Extended-spectrum beta-lactamases: an update on
270 their characteristics, epidemiology and detection. JAC Antimicrob Resist 2021; 3:
271 dlab092.
272 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
273 Bioinformatics 2018; 34: i884-i890.
274 CLSI. *Performance Standards for Antimicrobial Susceptibility Testing—Thirtieth Edition:*
275 *M100*. 2020.

276 Coppola N, Cordeiro NF, Trenchi G, Esposito F, Fuga B, Fuentes-Castillo D, et al. Imported
277 One-Day-Old Chicks as Trojan Horses for Multidrug-Resistant Priority Pathogens
278 Harboring *mcr-9*, *rmtG*, and Extended-Spectrum beta-Lactamase Genes. Appl Environ
279 Microbiol 2022; 88: e0167521.
280 Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain,
281 loss and rearrangement. PLoS One 2010; 5: e11147.
282 Falgenhauer L, Zur Nieden A, Harpel S, Falgenhauer J, Domann E. Clonal CTX-M-15-
283 Producing *Escherichia coli* ST-949 Are Present in German Surface Water. Front
284 Microbiol 2021; 12: 617349.
285 Girlich D, Bonnin RA, Naas T. Occurrence and Diversity of CTX-M-Producing *Escherichia*
286 *coli* From the Seine River. Front Microbiol 2020; 11: 603578.
287 Gomi R, Matsuda T, Matsui Y, Yoneda M. Fecal source tracking in water by next-generation
288 sequencing technologies using host-specific *Escherichia coli* genetic markers. Environ
289 Sci Technol 2014; 48: 9616-23.
290 Gomi R, Matsuda T, Matsumura Y, Yamamoto M, Tanaka M, Ichiyama S, et al. Whole-
291 Genome Analysis of Antimicrobial-Resistant and Extraintestinal Pathogenic
292 *Escherichia coli* in River Water. Appl Environ Microbiol 2017; 83: e02703-16.

293 Guenther S, Semmler T, Stubbe A, Stubbe M, Wieler LH, Schaufler K. Chromosomally
294 encoded ESBL genes in *Escherichia coli* of ST38 from Mongolian wild birds. J
295 Antimicrob Chemother 2017; 72: 1310-1313.
296 Guiral E, Mendez-Arancibia E, Soto SM, Salvador P, Fabrega A, Gascon J, et al. CTX-M-15-
297 producing enteroaggregative *Escherichia coli* as cause of travelers' diarrhea. Emerg
298 Infect Dis 2011; 17: 1950-3.
299 Hamamoto K, Hirai I. Characterisation of chromosomally-located $bla_{\text{CTX-M}}$ and its surrounding
300 sequence in CTX-M-type extended-spectrum beta-lactamase-producing *Escherichia*
301 *coli* isolates. J Glob Antimicrob Resist 2019; 17: 53-57.
302 Hirai I, Fukui N, Taguchi M, Yamauchi K, Nakamura T, Okano S, et al. Detection of
303 chromosomal $bla_{\text{CTX-M-15}}$ in *Escherichia coli* O25b-B2-ST131 isolates from the Kinki
304 region of Japan. Int J Antimicrob Agents 2013; 42: 500-6.
305 Jang J, Suh YS, Di DY, Unno T, Sadowsky MJ, Hur HG. Pathogenic *Escherichia coli* strains
306 producing extended-spectrum beta-lactamases in the Yeongsan River basin of South
307 Korea. Environ Sci Technol 2013; 47: 1128-36.
308 Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat
309 graphs. Nat Biotechnol 2019; 37: 540-546.

310 Liu H, Zhou H, Li Q, Peng Q, Zhao Q, Wang J, et al. Molecular characteristics of extended-
311 spectrum beta-lactamase-producing *Escherichia coli* isolated from the rivers and lakes
312 in Northwest China. BMC Microbiol 2018; 18: 125.
313 Matsumura Y, Noguchi T, Tanaka M, Kanahashi T, Yamamoto M, Nagao M, et al. Population
314 structure of Japanese extraintestinal pathogenic *Escherichia coli* and its relationship
315 with antimicrobial resistance. J Antimicrob Chemother 2017a; 72: 1040-1049.
316 Matsumura Y, Pitout JDD, Peirano G, DeVinney R, Noguchi T, Yamamoto M, et al. Rapid
317 Identification of Different *Escherichia coli* Sequence Type 131 Clades. Antimicrob
318 Agents Chemother 2017b; 61.
319 Miyagi K, Hirai I. A survey of extended-spectrum beta-lactamase-producing
320 *Enterobacteriaceae* in environmental water in Okinawa Prefecture of Japan and
321 relationship with indicator organisms. Environ Sci Pollut Res Int 2019; 26: 7697-7710.
322 Nair S, Chattaway M, Langridge GC, Gentle A, Day M, Ainsworth EV, et al. ESBL-producing
323 strains isolated from imported cases of enteric fever in England and Wales reveal
324 multiple chromosomal integrations of $bla_{\text{CTX-M-15}}$ in XDR *Salmonella* Typhi. J
325 Antimicrob Chemother 2021; 76: 1459-1466.
326 Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with
327 Antimicrobial Resistance. Clin Microbiol Rev 2018; 31.

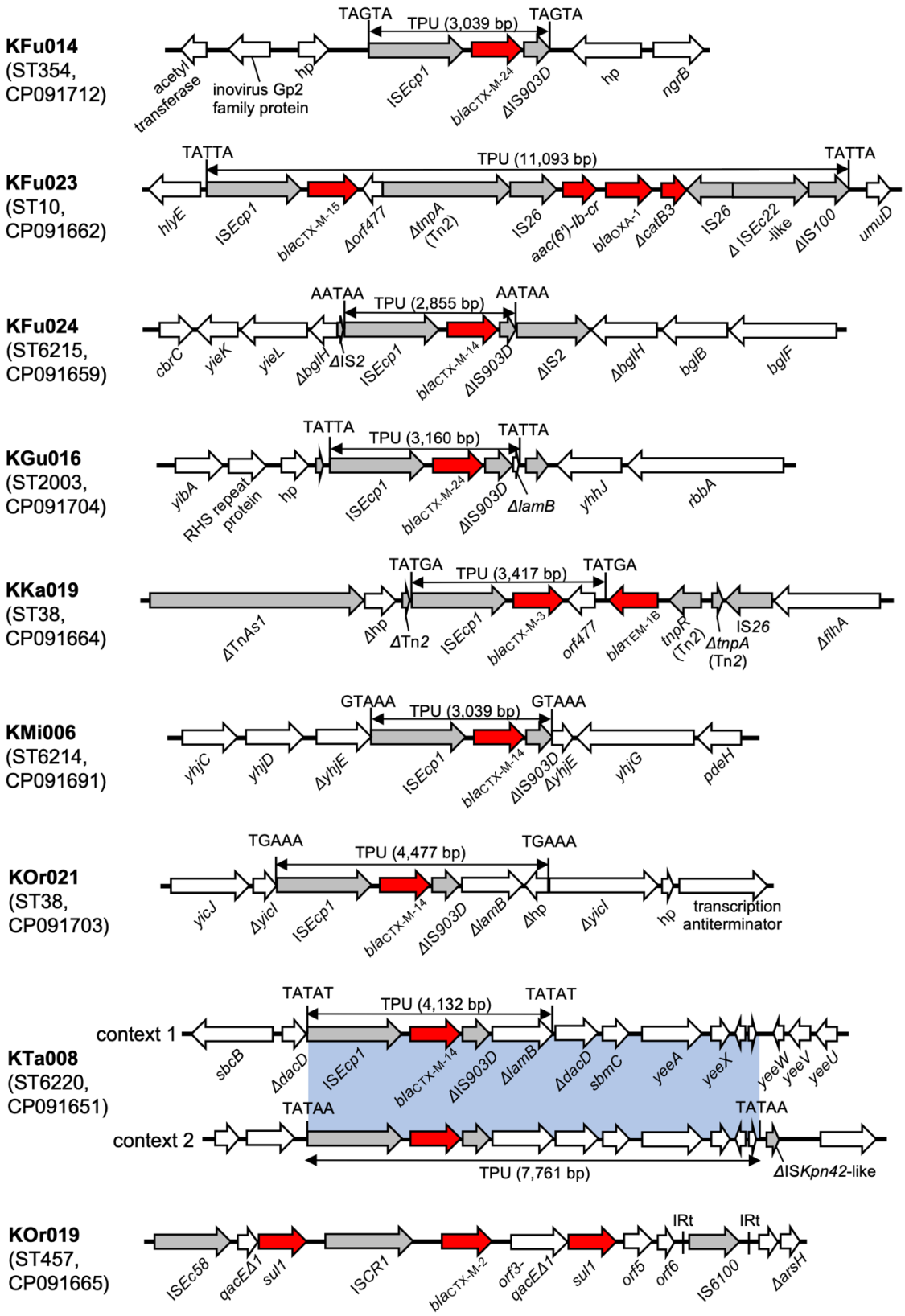
328 Pitout JD. Infections with extended-spectrum beta-lactamase-producing enterobacteriaceae:
329 changing epidemiology and drug treatment choices. Drugs 2010; 70: 313-33.
330 Pitout JD, Laupland KB. Extended-spectrum beta-lactamase-producing Enterobacteriaceae: an
331 emerging public-health concern. Lancet Infect Dis 2008; 8: 159-66.
332 Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of
333 plasmids from draft assemblies. Microb Genom 2018; 4: e000206.
334 Rodriguez I, Thomas K, Van Essen A, Schink AK, Day M, Chattaway M, et al. Chromosomal
335 location of $bla_{\text{CTX-M}}$ genes in clinical isolates of *Escherichia coli* from Germany, The
336 Netherlands and the UK. Int J Antimicrob Agents 2014; 43: 553-7.
337 Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for
338 bacterial insertion sequences. Nucleic Acids Res 2006; 34: D32-6.
339 Wick RR, Holt KE. Polypolish: Short-read polishing of long-read bacterial genome assemblies.
340 PLoS Comput Biol 2022; 18: e1009802.
341 Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies
342 from short and long sequencing reads. PLoS Comput Biol 2017; 13: e1005595.
343 Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome
344 assemblies. Bioinformatics 2015; 31: 3350-2.

345 Zurfluh K, Hachler H, Nuesch-Inderbinen M, Stephan R. Characteristics of extended-spectrum
346 beta-lactamase- and carbapenemase-producing Enterobacteriaceae Isolates from rivers
347 and lakes in Switzerland. Appl Environ Microbiol 2013; 79: 3021-6.

348

349

350 **FIGURES**



351

352 **Figure 1**. Genetic contexts of chromosomal *bla*$_{CTX-M}$ genes. Red arrows indicate antibiotic

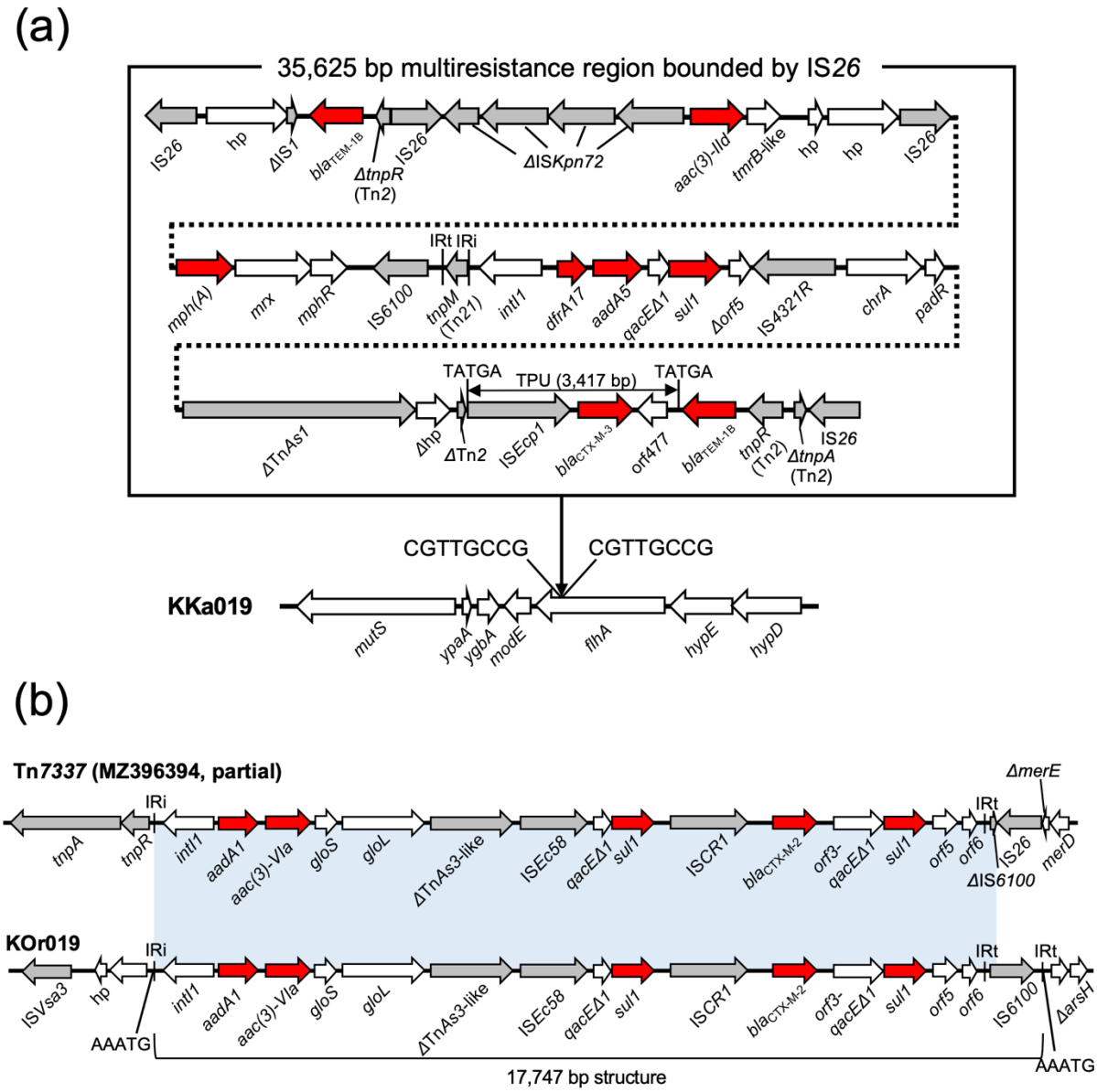
353 resistance genes, gray arrows indicate mobile elements, and white arrows indicate other

354 genes. The blue shaded box indicates the regions with 100% nucleotide identity. The entire

355 resistance regions of KKa019 and KOr019 are shown in **Figure 2**. 5-bp TSD are shown next

356 to each TPU. hp: hypothetical protein.

357



358

359 **Figure 2.** (a) A multiresistance region on the chromosome of KKa019. Insertion of the

360 resistance region into *flhA* generated 8-bp TSD of CGTTGCCG. A BLASTN search against

361 the nucleotide collection (nr/nt) database using the 35,625 bp multiresistance region as a

362 query sequence identified some related structures, but none showed >90% coverage. (b) A

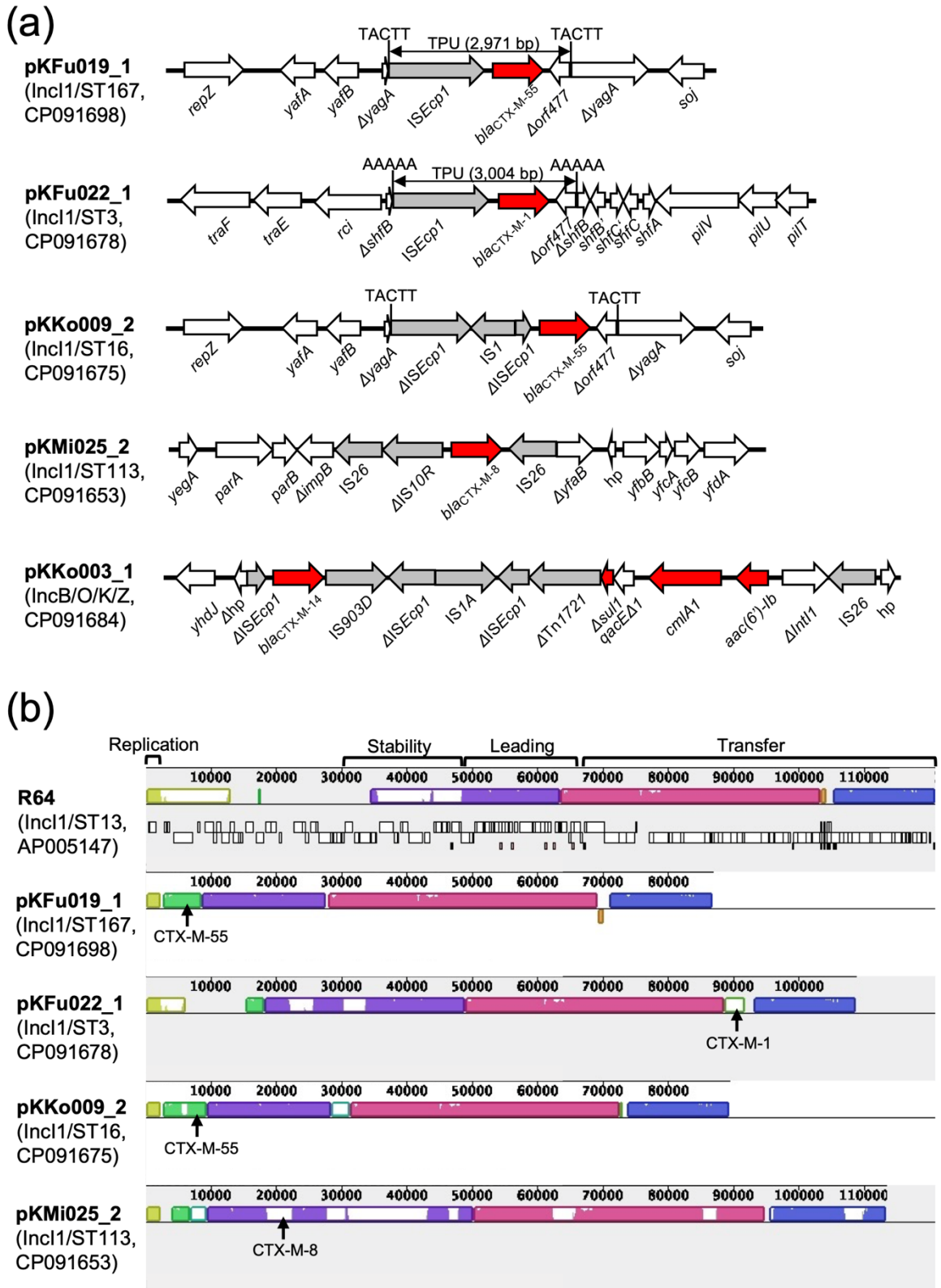
363 multiresistance region on the chromosome of KOr019. Tn*7337* is shown for the purpose of

364 comparison. The 17,747 bp structure from IRi to IRt was inserted into an intergenic region in

365 the KOr019 chromosome, generating 5-bp TSD of AAATG. The light blue shaded box

366 indicates regions with >99% nucleotide identity.

367



368

369 **Figure 3**. (a) Genetic contexts of *bla*$_{\text{CTX-M}}$ genes on plasmids. (b) Mauve comparison of IncI1

370 plasmids (Darling et al., 2010). IncI1 reference plasmid R64 is shown at the top. Four major

371 regions in R64 (replication, stability, leading, and conjugative transfer) are indicated.

372 Homologous segments are shown as colored blocks. Locations of $bla_{\text{CTX-M}}$ genes are indicated

373 with arrows.