# Specificities and commonalities of carbapenemase producing *Escherichia coli* isolated in France from 2012 to 2015.

[1,2]Rafael Patiño-Navarrete[#], [1,2]Isabelle Rosinski-Chupin[#], [1,2]Nicolas Cabanel, [1,2,3]Pengdbamba Dieudonné Zongo, [1,4]Mélanie Héry, [1,4]Saoussen Oueslati, [1,4]Delphine Girlich, [1,4,5,6]Laurent Dortet, [1,4,6]Rémy A Bonnin, [1,4,5,6]Thierry Naas[$] and [1,2,*]Philippe Glaser[$].

[$ and #] Contributed equally

[1]Unité EERA, Institut Pasteur, APHP, Université Paris Saclay, [2]UMR3525, CNRS, 28 rue du Dr Roux, 75015, Paris, France; [3]Sorbonne Université, Paris, France, [4] Team "Resist" INSERM U1184 "Immunology of Viral, Auto-Immune, Hematological and Bacterial diseases (IMVA-HB),"; [5]Department of Bacteriology-Hygiene, Bicêtre Hospital, APHP, [6]Associated French National Reference Center for Antibiotic Resistance, Le Kremlin-Bicêtre, France.

*Corresponding author:

Philippe GLASER,

Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15,

Tel + 33 1 45 68 89 96 - E-mail: pglaser@pasteur.fr

Running title: Carbapenemase producing *E. coli* from the French NRC

**ABSTRACT**

Carbapenemase-producing *Escherichia coli* (CP-*Ec*) represent a major public health threat with a risk of dissemination in the community as it has occurred for lineages producing extended spectrum ß-lactamases. To characterize the extend of CP-*Ec* spread in France, isolates from screening and infection samples received at the French National Reference Centre laboratory (F-NRC) for carbapenemase-producing Enterobacterales were investigated. Six hundred and ninety one CP-*Ec* isolates collected between 2012 and 2015 and 22 before were fully sequenced. Analysis of their genome sequences revealed some disseminating multidrug resistant (MDR) lineages frequently acquiring diverse carbapenemase genes mainly belonging to clonal complex (CC) 23 (ST 410) and CC10 (ST10, ST167) and sporadic isolates including rare ST131 isolates (n=17). However, the most represented ST was ST38 (n=92) with four disseminated lineages carrying $bla_{\text{OXA-48-like}}$ genes inserted in the chromosome. Globally, the most frequent carbapenemase gene (n=457) was $bla_{\text{OXA-48}}$. It was also less frequently associated with MDR isolates being the only resistance gene in 119 isolates. Thus, outside the ST38 clades, its acquisition was frequently sporadic with no sign of dissemination, reflecting the circulation of the IncL plasmid pOXA-48 in France and its high frequency of conjugation. In contrast $bla_{\text{OXA-181}}$ or $bla_{\text{NDM}}$ genes were often associated with the evolution of MDR *E. coli* lineages characterized by mutations in *ftsI* and *ompC.*

**IMPORTANCE**

Carbapenemase-producing *Escherichia coli* (CP-*Ec*) might be difficult to detect, as minimal inhibitory concentrations can be very low. However, their absolute number and their proportion among carbapenem-resistant *Enterobacterales* have been increasing, as reported by WHO and national surveillance programs. This suggests a still largely uncharacterized community spread of these isolates. Here we have characterized the diversity and evolution of CP-*Ec* isolated in France before 2016. We show that carbapenemase genes are associated with a wide variety of *E. coli* genomic backgrounds and a small number of dominant phylogenetic lineages. In a significant proportion of CP-*Ec*, the most frequent carbapenemase gene $bla_{\text{OXA-48}}$, was detected in isolates lacking any other resistance gene, reflecting the dissemination of pOXA-48 plasmids, likely in the absence of any antibiotic pressure. In contrast carbapenemase gene transfer may also occur in multi-drug resistant *E. coli*, ultimately giving rise to at-risk lineages encoding carbapenemases with a high potential of dissemination.

54 **INTRODUCTION**

55 *Escherichia coli* is one of the first causes of diverse bacterial infections in the community and in the
56 hospital. In particular, it is the most frequent cause of urinary tract infection (UTI) and 50-60% of
57 women will suffer at least once a UTI during her life (1). Therefore, multidrug resistance (MDR) in *E. coli*
58 is a major public health issue making *E. coli* infections more difficult to treat. In addition, as
59 carbapenemase-producing *Enterobacterales* (CPE) are increasingly detected (2, 3) and *E. coli* is an
60 ubiquitous member of the normal gut microbiome in humans, carbapenemase-producing *E. coli* (CP-
61 *Ec*) are also becoming a major actor for the global dissemination of carbapenemase genes (4).

62 The emergence and spread of carbapenem-resistant Gram-negative bacteria is mainly linked to the
63 widespread dissemination through horizontal gene transfer (HGT) of mobile genetic elements (MGEs)
64 encoding carbapenemases. These carbapenemases belong to Ambler class A (mainly KPC-type), class
65 B (metallo-β-lactamases IMP-, VIM- and NDM-types) or class D (OXA-48-like enzymes) of beta-
66 lactamases (5). The global epidemiology of extended spectrum ß-lactamases (ESBL) producing *E. coli*
67 has been characterized through multiple studies, revealing in Western countries the major
68 contribution of the sequence type (ST)131 lineage in the high prevalence of $bla_{\text{CTX-M}}$ family ESBL genes
69 (6). Much less is known with respect to CP-*Ec*.

70 Since in 2012, the French National Reference Centre Laboratory for carbapenemase-producing
71 *Enterobacterales* (F-NRC) has experienced a steady increase in the number of CP-*Ec* isolates received
72 each year (2). A multi locus sequence typing (MLST) analysis of isolates received in 2012 and 2013
73 revealed a broad diversity of STs, as the 140 analyzed isolates belonged to 50 different STs. However,
74 a few STs were over-represented (7), such as the ST38 (24 isolates) and the ST410 (10 isolates). In that
75 study, only one isolate belonged to ST131, contrasting with the situation in the UK where ST131
76 isolates represented 10% of the CP-*Ec* isolates received between 2014 and 2016 by Public Health
77 England (8). On the other hand, a genome based survey of CPE in the Netherlands between 2014 and
78 2018, revealed that the 264 received *E. coli* isolates belonged to 87 different STs, with three dominant
79 lineages, ST38 (n=46), ST167 (n=22) and ST405 (n=16) representing 32 % of the isolates (9). F-NRC
80 isolates also showed a predominance of isolates producing OXA-48-like carbapenemases followed by
81 NDM-producing isolates and suggested clonally related isolates among ST38 OXA-48-producing
82 isolates and ST410 OXA-181-producing isolates, respectively (7).

83 As whole-genome sequencing (WGS) significantly increases our ability to infer phylogenetic
84 relationships between isolates, we recently sequenced 50 ST410 CP-*Ec* isolates received by the F-NRC
85 between 2011 and 2015 (10) and found that 72% of them belonged to a newly described ST410 lineage
86 producing OXA-181 (11). We showed that this clade is characterized by mutations in the two porin
87 genes *ompC* and *ompF* leading to a decreased outer membrane permeability to certain ß-lactams and
88 by a four-codon duplication (YRIN) in the *ftsI* gene encoding the penicillin binding protein 3 (PBP3)

89 leading to a decreased susceptibility to ß-lactams targeting this PBP (10). After a thorough analysis of
90 CP-*Ec* genome sequences from public databases for mutations in these three genes, we proposed that
91 CP-*Ec* followed three different evolutionary trajectories. In some lineages which are enriched in CP-*Ec*
92 isolates and have disseminated globally, acquisition of carbapenem resistance genes might have been
93 facilitated by the mutations in porin genes and in *ftsI*. In ST38, the genetic background and in particular
94 a specific *ompC* allele with reduced permeability to some ß-lactams may have similarly facilitated the
95 acquisition of carbapenemase genes. In contrast, other CP-*Ec* isolates including from ST131 might have
96 occurred sporadically following the acquisition of plasmids encoding carbapenemase and with no clue
97 of dissemination (10).
98 Here we thoroughly characterized the diversity of CP-*Ec* circulating in France by sequencing the
99 genome of almost all isolates received by the F-NRC from its creation until 2015 (Table S1). By
100 combining whole genome phylogeny with the addition of *E. coli* genome sequences publicly available
101 (Table S2), and tracking mutations in candidate genes, we show that three different situations are
102 encountered. The high transmission potency of the IncL pOXA-48 plasmids has led to a high frequency
103 of OXA-48-producing isolates, often characterized by susceptibility to most non-ß-lactam antibiotics.
104 On the other hand, an increasing number of CP-*Ec* lineages, mainly producing OXA-181 and NDM
105 carbapenemases, are observed in France as in other Eurpean countries. These lineages are multi-drug
106 (MDR) or extensively drug resistant (XRD) lineages. They are strongly mutated in quinolone resistance
107 determinants and are often mutated in *ftsI* and in porin genes. Finally, the rapid dissemination of four
108 OXA-48/OXA-244 ST38 lineages might have been favored by the chromosomal integration of the
109 carbapenemase gene.
110
111 **RESULTS**
112
113 **CP-*Ec* isolates collected until 2015 by the French National Reference Center**.
114 Isolates analyzed in this work were sent to the F-NRC laboratory on a voluntary basis from private and
115 public clinical laboratories from different parts of France mainly between the years 2012 to 2015.
116 During this period, we encountered a regular increase in the number of isolates received each year
117 (Fig 1A) corresponding to an increasing number of isolated strains and an increasing frequency of
118 isolates sent to the F-NRC. 713 CP-*Ec* isolates, including 22 collected from 2001 to 2011 were submitted
119 to WGS. The 691 sequenced isolates of the 2012-2015 period represented 87.5 % of the 790 CP-*Ec*
120 isolates received by the F-NRC during this period.
121 The majority of the sequenced isolates (66.5%, 474/713) were from rectal swab screening of patients
122 suspected of carrying CPE (patients repatriated from an hospital abroad, patients that have visited a
123 foreign country within the last six months, contact patients of a former carrier, or a previously known

124 carrier), 24% (172/713) were considered to be responsible for infection (isolated from clinical
125 samples), and the source was unknown for 67 isolates (Fig. 1A, Table S1). During the four-year period,
126 The number of infection-related isolates was found to increase more slowly than the number of
127 screening or of unknown source isolates. A nearly two-fold decrease in the proportion of clinical versus
128 total isolates was observed between 2013 and 2014 (Fig. 1B). Among the 172 isolates associated with
129 disease, 122 were isolated from urine (71%), 15 from blood samples, eight from deep site samples,
130 eight from wound samples, nine from the respiratory tract, three from vaginal samples, three from
131 pus, three from bile, and one from the skin of a newborn. All these isolates were previously identified
132 as carbapenemase producers by PCR.
133
134 **Diversity of CP-*Ec* isolates as assessed by whole genome sequencing.**
135 The genome sequences of the 713 CP-*Ec* isolates were first analyzed following *de novo* assembly. For
136 each isolate, we determined its MLST type (Achtman scheme), its phylogroup (ClermonTyping) and its
137 antibiotic resistance genes (ARG) content. We also searched for mutations in the quinolone resistance
138 determining regions (QRDRs) of *gyrA*, *parC* and *parE,* and for mutations in *ftsI*, *ompC* and *ompF* we
139 previously identified as associated with CP-*Ec* disseminated lineages (10) (table S1). F-NRC CP-*Ec* were
140 assigned to 168 different STs including six new allelic combinations. ST38 was the most prevalent ST
141 (n=92, 12.9 %), followed by ST10 (n=67, 9.4 %), ST410 (n=64, 9 %), and ST167 (n=34, 4.8 %). Ten
142 additional STs were represented by at least 10 isolates (Fig. S1A, table S1) and 154 STs with less than
143 ten isolates including 102 STs with a single isolate.
144 We next performed a core genome phylogeny following read mapping using strain MG1655
145 (NC_000913.3) as reference genome sequence. The phylogenetic tree, estimated from 372,238 core
146 SNPs was consistent with the results of phylogroup determination by using *in silico* ClermonTyping (12)
147 except for a few isolates (Fig. 2). In agreement with the MLST-based analysis, this tree showed a broad
148 diversity of CP-*Ec* isolates belonging to the eight phylogroups and three dominant clades
149 corresponding to CC10 (phylogroup A; including ST10, ST167 and ST617), CC23 (phylogroup C;
150 including ST410, ST88 and ST90) and ST38 (phylogroup D) with 161, 97, and 98 isolates respectively,
151 accounting for 49.9 % of the F-NRC CP-*Ec* isolates analyzed (Fig. 2). Phylogroup B2 isolates represented
152 11 % of the total isolates (n=80) and only 17 CP-ST131 isoaltes were identified. Fluctuations in the
153 proportions of the main STs could be observed during the four years of the analysis but no clear
154 tendency could be identified (Fig. S1A).
155 Infection-related and screening isolates were intermixed throughout the phylogeny (Fig. 2). However,
156 an enrichment in infection-related isolates was observed in phylogroup C (Pearson's Chi-squared test,
157 $p<0.02$, ddl1) and phylogroup B2 (Pearson's Chi-squared test $p<0.0005$, ddl1) (Fig. S1B). In phylogroup

158 B2, 5 out of 9 ST127 isolates, 5 out of 17 ST131 isolates and 7 out of 8 ST636 were responsible for UTIs
159 (Table S1).
160
161 **Diversity of antibiotic resistance genes carried by CP-*Ec* isolates**
162 The number of acquired ARGs was found to vary between 1 and 26 among the 713 CP-*Ec* isolates. The
163 median was higher in phylogroup C isolates (m=16) and lower in isolates from phylogroup B2 (m=3)
164 compared to other phylogroups (Medians: A: 9; B1: 9; D: 11) (Fig. 3A). An ESBL of the CTX-M family
165 was present in 40.7% (N=290) of the isolates, with a predominance of $bla_{\text{CTX-M-15}}$ gene (N=205) (Table
166 S1). Mutations in g*yrA*, *parC* and/or *parE* potentially leading to fluoroquinolone (FQ) resistance
167 occurred in 425 CP-*Ec* isolates (59.6%), with mutations in *gyrA*, *parC* and *parE* identified in 412 isolates,
168 309 and 261 isolates respectively (Table S1). Up to five mutations in QRDRs were identified in 19
169 isolates and 250 isolates had four mutations in QRDR, suggesting they have been submitted to a long-
170 term evolution under antibiotic pressure including FQ (Table S1). Globally a higher number of
171 mutations in QRDR was associated with a higher number of resistance genes (Fig. 3B), furthering the
172 link between the number of QRDR mutations and a likely evolution under antibiotic selective pressure
173 for CP-*Ec* isolates.
174 Four isolates from 2015, two ST648 collected in a same hospital at one-week interval, one ST216 and
175 one ST744, encoded a *mcr-9* gene conferring resistance to colistin (Table S1). Although not associated
176 with infection, the four isolates were MDR, carrying 13 to 20 acquired ARGs in addition to *mcr-9* and
177 carbapenemase genes ($bla_{\text{VIM-1}}$ for three of them, $bla_{\text{NDM-1}}$ and $bla_{\text{OXA-48}}$ for one). Determination of the
178 MIC for colistin for these four isolates were at the resistance breakpoint (MIC=2 mg/l).
179
180 **Diversity of carbapenemase genes**
181 Carbapenemase genes identified by WGS (Table S1) were in agreement with molecular data collected
182 by the F-NRC laboratory. The $bla_{\text{OXA-48}}$ and $bla_{\text{OXA-181}}$ genes were the most frequent carbapenemase
183 genes detected in 464 (65%) and 101 (14.1%) *E. coli* isolates respectively. Twenty-nine other isolates
184 carried minor $bla_{\text{OXA-48-like}}$ genes (Table S1). $bla_{\text{NDM}}$ family genes were detected in 14.9% (106/713) of
185 the CP-*Ec* isolates ($bla_{\text{NDM-5}}$ in 49 isolates and $bla_{\text{NDM-1}}$ in 41). Three additional $bla_{\text{NDM}}$ alleles were
186 identified: $bla_{\text{NDM-7}}$ (n=10) including a variant coding for a NDM7-like carbapenemase with a S24G
187 mutation, $bla_{\text{NDM-4}}$ (n=5), and $bla_{\text{NDM-6}}$ (n=1). Four different $bla_{\text{VIM}}$ alleles were detected in 18 isolates:
188 $bla_{\text{VIM-1}}$ (n=8), $bla_{\text{VIM-4}}$ (n=8), $bla_{\text{VIM-2}}$ (n=1), and $bla_{\text{VIM-29}}$ (n=1). Only five *E. coli* isolates expressed $bla_{\text{KPC-}}$
189 $_{2}$ (n=2) or $bla_{\text{KPC-3}}$ (n=3) alleles. In twelve isolates, two different carbapenemase genes were found;
190 $bla_{\text{OXA-48}}$ / $bla_{\text{NDM-1}}$ was the most frequent co-occurrence (n=5) (Table S1).
191 The analysis of the number of additional ARGs and of mutations in QRDR regions showed that,
192 compared with other carbapenemase genes the presence of $bla_{\text{OXA-48}}$ was frequently associated with

193 less resistant isolates (Fig. 4A, 4B). In particular it was the only ARG in 117 isolates among the 457
194 (24.5%) encoding this carbapenemase, including 23 infection-related isolates. Among those isolates,
195 only five showed mutations in QRDR. Conversely, only two $bla_{NDM-1}$ gene carrying isolates among 42
196 and one $bla_{NDM-5}$ gene carrying isolates among 49 carried no other ARG.
197 A temporal analysis of the proportion of isolates displaying $bla_{OXA-48}$ did not reveal significant variations
198 (Pearson's Chi-squared test, p=0.05, ddl= 3) during the 2012-2015 period (Fig. 4C). In contrast, the
199 frequency of $bla_{OXA-181}$ relatively to other alleles was found to significantly increase from 0 % in 2012
200 to 18.9% in 2015 (Pearson's Chi-squared test, $p < 0.0002$ ddl=3) (Fig. 4C).
201 We observed some association between the carbapenemase gene and the ST. Among the 22 STs with
202 at least seven F-NRC CP-*Ec* isolates, fourteen predominantly displayed $bla_{OXA-48}$ (Fig. 5A); $bla_{OXA-48}$ was
203 also predominant in STs represented by six or less isolates. In three lineages, ST410, ST940, and
204 ST1284, $bla_{OXA-181}$ was the predominant allele. These three STs grouped 74% of the $bla_{OXA-181}$ gene
205 carrying *E. coli* isolates from the F-NRC, and ST410 accounted for 50.5% of them. Finally, $bla_{OXA-204}$ was
206 the most prevalent allele in ST90 and $bla_{NDM-5}$ the predominant allele in ST636.
207
208 **Characteristics of the STs preferentially associated with F-NRC CP-*Ec* isolates.**
209 Eight among the 14 STs with more than 10 CP-*Ec* isolates (ST410, ST167, ST940, ST405, ST617, ST90,
210 ST1284, ST101) were characterized by a larger number of ARGs (median ≥ 10) and a larger number of
211 mutations in QRDR as compared to ST38, ST10, ST131, ST69, ST88 and ST represented by less than
212 seven isolates (Fig. 5 B, C). ST648 isolates have an intermediate position, being highly mutated in QRDR
213 but less rich in ARGs. Analysis of CP-*Ec* isolates from the F-NRC for polymorphisms in *ftsI* revealed that
214 131 (18.4%) had a four-AA-insertion between positions 333 and 334 of PBP3, the insertion being
215 particularly prevalent in ST101, ST167, ST410 and ST1284 (Sup. Table 1 and Fig. 5D). Similarly, *ompC*
216 alleles we previously characterized as modifying susceptibility to antibiotics were more prevalent in
217 some STs: an ST38-like *ompC* allele resulting from a recombination event or a G138D mutation were
218 present in 33 and 26 isolates belonging to six different ST, whereas the R195L mutation was observed
219 in 50 isolates all from ST410 (Table S1 and Fig. 5E). In addition, 144 isolates belonging to phylogroups
220 D and F possessed the ST38-like allele inherited vertically. On the other hand the *ompF* gene was
221 pseudogeneized in 47 isolates (Table S1).
222 To analyze the phylogenetic relationships between F-NRC Cp-*Ec* and other Cp-*Ec* isolates collected
223 worldwide, we built ST-based maximum-likelihood trees for the 14 STs with more than 10 isolates and
224 included to this analysis genome sequences publicly available (Table S2). Together with the analysis of
225 QRDR mutations and mutations in *ftsI* and *ompC*, this showed that 51 (80%) of the ST410 (Fig.S2A), 17
226 (49%) of the ST167 (Sup Fig.S2B), 3 (33%) of the ST405 (Sup Fig.S2C) and 8 (62%) of the ST101 (Fig.
227 S2D) F-NRC CP-*Ec* belonged to internationally disseminating MDR subclades enriched in CP we

228 previously identified (10). Subclades characterized by mutations in *ftsI* and *ompC,* mainly expressed
229 $bla_{OXA-181}$ (ST410) or $bla_{NDM}$ (other lineages). Other isolates belonging to these ST were dispersed on
230 their respective phylogenies and with no sign of clonal dissemination for most of them.
231 ST38 corresponded to the most represented ST (n=92) among the studied CP-*Ec*. The phylogenetic
232 reconstruction together with 314 additional non redundant CP-*Ec* genome sequences retrieved from
233 Enterobase (http://enterobase.warwick.ac.uk/) and 150 non-CP isolates from NCBI provided
234 evidence that 80.4% (n=67) of the French isolates clustered into four clades (Fig. 6). Three of them, G1
235 (n=35), G2 (n=25) and G4 (n=4) only contained isolates encoding $bla_{OXA-48}$, while G3 (n=7), included
236 isolates expressing $bla_{OXA-48}$ or its single nucleotide derivative $bla_{OXA-244.}$ All but one isolates in G1
237 expressed $bla_{CTX-M-24}$, while isolates of G3 expressed $bla_{CTX-M-14b}$. The phylogenetic analysis provided
238 evidence of worldwide dissemination of G1, G3 and G4 clades and multiple introduction in France.
239 Strickingly, among the 28 isolates of the G2 clade 24 were collected in France and four in the
240 Netherlands, suggesting at this time a more regional dissemination. In none of the isolates of the four
241 clades, an IncL plasmid, that generally encode $bla_{OXA-48}$ (13), could be identified by PlasmidFinder (14).
242 It suggests a chromosomal integration as previously shown among ST38 isolates collected in the UK
243 (15). Analysis of two isolates of the G1 and G3 cluster (G1: GCA_900607445.1; G3: GCA_004759025.1),
244 whose genome sequences were completely assembled, confirmed that $bla_{OXA-48}$ and $bla_{OXA-244}$, in these
245 two isolates respectively, were chromosomally integrated. To assess the genetic support of $bla_{OXA-48}$ in
246 the two other clades, we have determined the complete genome sequence of two isolates belonging
247 to G2 (CNR65D6) and G4 (CNR85I8) by combining long-read Pacbio and Illumina sequencing and found
248 that both possess as chromosomally inserterted $bla_{OXA-48}$ gene. A fifth clade, composed of seven closely
249 related French isolates corresponded to a possible outbreak in the East of Paris between December
250 2014 and December 2015. These isolates are predicted to be highly resistant as they are carrying in
251 addition to $bla_{VIM-4,}$ $bla_{CTX-M-15}$, as well as 15 to 18 additional ARGs and five mutations in QRDRs.
252 ST10 *E. coli* isolates are commensals of a variety of mammals and bird species (16) and studies in
253 different contexts have shown that this ST is also associated with ARGs carriage (17, 18). The 67 F-NRC
254 ST10 CP-*Ec* isolates were distributed throughout the ST10 phylogeny. Nonetheless 31 isolates
255 belonged to a clade enriched with carbapenemase producers (Fig. S3A, in green) that mainly expressed
256 $bla_{OXA-48}$, with the exception of three isolates expressing $bla_{OXA-181}$, three expressing $bla_{NDM-5}$ and one
257 and $bla_{VIM-4}$. Two of the $bla_{OXA-181}$ and the three $bla_{NDM-5}$ expressing isolates were closely related, shared
258 a CTX-M-27 ESBL, four mutations in QRDR and a *ftsI* allele with a four AA YRIN insertion (Fig. S3A, in
259 red).
260 ST131 isolates are major contributors for extra-intestinal infections and the main responsible for the
261 dissemination of CTX-M-15 ESBL (19, 20). ST131 evolution has been thoroughly analyzed and four main

262 lineages A, B, C1 and C2 have been described (21). We identified only 17 ST131 CP-*Ec* isolates received
263 by the F-NRC (Fig. 2), ten belonged to lineage A, while the others were scattered in the three other
264 lineages (Fig. S3B). The main carbapenemase associated with these isolates was OXA-48 (N=10). We
265 also did not observe any association with the carriage of $bla_{\text{CTX-M}}$ genes, as only four ST131 CP-*Ec*
266 isolates carryied a $bla_{\text{CTX-M-15}}$ gene. These results are in accordance with the UK survey (8) and our
267 previous observations from sequences retrieved from public databases (10).
268 ST940 (phylogroup B1) and ST1284 (phylogroup A) represented 2.4% (n=17) and 1.9% (n=14) of the
269 sequenced isolates while they were poorly represented in sequence databases. For instance, in May,
270 2021, only 83 and 103 sequences genome sequences could be retrieved from Enterobase for these ST
271 respectively, but with 43% (n=36) and 22% (n=22) carrying a carbapenemase gene respectively. This
272 suggested that ST940 and ST1284 were associated with the dissemination of carbapenemase genes.
273 Phylogenetic trees were drawn by adding the non-redundant sequences from enterobase to those of
274 the F-NRC and NCBI. For ST940 (Fig. S4A) this revealed a well-differentiated sub-clade of 18 isolates
275 from Asia, Australia and Europe carrrying $bla_{\text{NDM-5}}$ and sharing mutations in *ftsI* and *ompC*. Isolates
276 from the F-NRC were distributed on the tree, showing that the over-representation of ST940 in France
277 did not result from local outbreaks. Three F-NRC isolates belonged to a second smaller sub-clade of
278 $bla_{\text{NDM}}$ expressing isolates, also mutated in *ftsI* and o*mpC*. In contrast, in ST1284, thirteen encoding
279 $bla_{\text{OXA-181}}$, were found to be closely related (Fig. S4B). Metadata analysis showed that 12 out of them
280 were recovered from the same health facility in the Paris suburb during a one-month period in 2015,
281 demonstrating a local outbreak origin. A fourteenth unrelated isolate expressed $bla_{\text{NDM-5}}$. The F-NRC
282 isolates belonged to a clade characterized by four mutations in QRDR determinants, a YRIN duplication
283 in PBP3, a G137D mutation in OmpC and containing additional OXA-181- or NDM-producing isolates.
284 Among the fifteen ST90 CP-*Ec* isolates, nine encoding the $bla_{\text{OXA-204}}$ allele were isolated in hospitals East
285 of Paris between August 2012 to April 2013 and suspected to be associated with the use of a
286 contaminated endoscope (22). Isolates closely related to this outbreak reappeared in 2014 (three
287 times), and in 2015 (twice), also in East of Paris located hospitals. They belonged to a MDR clade
288 characterized by four mutations in QRDR (Fig. S5A). Finally, in the four last STs with more than 10 CP-
289 *Ec* isolates (ST69, ST88, ST617, and ST648), French CP-*Ec* isolates were distributed throughout their
290 respective phylogenetic trees (FigS5B and C,S6A and B) with no sign of clonal dissemination, except for
291 a small cluster of five ST69 from two geographical regions and expressing *bla*OXA-48.
292 In addition to potential outbreaks detected in ST38 ($bla_{\text{VIM-4}}$), in ST90 ($bla_{\text{OXA-204}}$) and ST1284 ($bla_{\text{OXA-}}$
293 $_{\text{181}}$), we also found evidence for another potential outbreak among isolates belonging to the ST359.
294 The five isolates encoded $bla_{\text{OXA-48}}$ and $bla_{\text{CTX-M-32}}$ and had three mutations in QRDR. They were isolated
295 during January 2014, four of them in the south-eastern part of France and one in the Parisian suburbs.
296

297 **DISCUSSION**

298 The prevalence of CP-*Ec* is increasing worldwide. However whether this reveals conjugation event in
299 *E. coli* isolates by circulating CP-encoding plasmids or the emergence and dissemination of at-risk
300 clones is still largely unknown. Here, we have analyzed, by using WGS, 691 (87.8%) of the CP-*Ec* isolates
301 received by the F-NRC during the period 2012-2015 and 22 isolates from the Bicetre hospital strain
302 collection (2001-2011) to characterize both the diversity of carbapenemase genes and of CP-*Ec* isolates
303 in France. Altogether 713 CP-*Ec* were sequenced, representing to our knowledge the most extended
304 collection of CP-*Ec* sequenced, published to date.

305 The number of CP-*Ec* isolates sent to the F-NRC was found to strongly increase during the four-year
306 period of analysis, which might be a consequence of an increased circulation of CP-*Ec* in France but
307 also of an increased screening of potential CPE carriers at their admission at hospital. Indeed, while
308 the number of infection-related CP-*Ec* isolates was regularly increasing, their proportion compared to
309 the total number of received and sequenced isolates decreased by two-fold, with a clear change
310 observed between 2013 and 2014 (Fig. 1). This is likely a consequence of the implementation of the
311 recommendations on MDR screening of the French Public Health Advisory Board 2013 (23). However,
312 as CP-*Ec* isolates were sent on a voluntary basis by clinical laboratories, we cannot exclude some
313 sampling bias.

314 The most predominant carbapenemase allele was $bla_{\text{OXA-48}}$ (65%) followed by $bla_{\text{OXA-181}}$ (14.1%) which
315 detection constantly increased from no cases in 2012 to 73 in 2015. Next was $bla_{\text{NDM-5}}$ that was found
316 to progressively substitute to the $bla_{\text{NDM-1}}$ gene in frequency of detection (Fig. 4). The predominance
317 of $bla_{\text{OXA-48-like}}$ genes and particularly of $bla_{\text{OXA-48}}$, among CP-*Ec* was also noted in other studies (9, 24,25)
318 although in lower proportions. The small number of $bla_{\text{KPC}}$ genes detected in these studies is in contrast
319 with the situation reported in *K. pneumoniae* for many European countries (26). However, $bla_{\text{OXA-48}}$
320 was found the most prevalent allele in *K. pneumoniae* in France (2) and in the Netherlands (9). This
321 difference might result from a lower capacity for $bla_{\text{KPC}}$ encoding plasmids to conjugate to *E. coli* as
322 compared to pOXA-48 or to a higher fitness cost. It might also be linked to sampling differences as
323 different thresholds of carbapenem susceptibility might have been used to collect CRE isolates among
324 studies. Indeed OXA-48-like carbapenemases are generally associated with lower levels of carbapenem
325 resistance than KPC.

326 The number of ARGs identified in the CP-*Ec* isolates from the F-NRC varied from one (the
327 carbapenemase gene only) to 26, showing that carbapenemase genes were acquired not only in MDR
328 isolates but also in an *E. coli* population that can be considered as "naïve" relative to an evolution
329 under antibiotic selective pressure and resident of the intestinal microbiota. This was particularly true
330 for isolates producing $bla_{\text{OXA-48}}$ that were generally associated with a lower number of resistance genes
331 than isolates producing other carbapenemases, irrespectively of their clinical status (Fig. 4A). In

332 particular 118 out of the 121 (97.5%) isolates with no other ARG than the carbapenemase gene carried
333 *bla*$_{OXA-48}$. Therefore, the predominance of *bla*$_{OXA-48}$ among the carbapenemase genes identified in the
334 F-NRC collection might mainly rely on the higher conjugative transfer rate of pOXA-48 related plasmids
335 compared to those encoding other carbapenemases (27). pOXA48 has indeed been shown to rapidly
336 conjugate among *Enterobacterales* in hospitalized patients (28). This could also explain the high
337 frequency of *bla*$_{OXA-48}$ carrying isolates belonging to ST10, ST69, ST88, ST127, ST12 and ST58 (Fig. 5),
338 that carried a small number of resistance genes and QRDR mutations and among which four were
339 previously identified among the most frequent *E. coli* ST characterized from fecal samples (29).
340 On the other hand, MDR CP-*Ec* isolates may result from the transfer of carbapenem resistance genes
341 into isolates already selected through multiple exposures to antibiotic treatments that have already
342 acquired multi-resistance plasmids or chromosomal mutations increasing their intrinsic drug
343 resistance, such as mutations in QRDR, in *ftsI* and/or in porin genes. Alternatively, the circulation of
344 carbapenem-resistant lineages, such as the OXA-181-producing ST410 *E. coli* lineage, showing a stable
345 association between the lineage and the resistance gene, may also occur and significantly contribute
346 to the number of CP-*Ec* isolates collected. Discriminating between both alternatives would require a
347 more extended comparison of plasmids carried in these lineages after long-read sequencing similarly
348 to what been done in *K. pneumoniae* (30). Strikingly, the analysis of *ftsI* alleles coding for a PBP3 with
349 the four amino-acid duplication revealed a contrasted situation among CP-*Ec* isolates collected in
350 France, with the duplication in *ftsI* found in 72.3%, 77.8% and 67% of the isolates carrying *bla*$_{NDM-5}$
351 (34/47), *bla*$_{NDM-7}$ (7/9) and *bla*$_{OXA-181}$ (68/101) genes but in only 0.7% (4/458) of the isolates carrying
352 *bla*$_{OXA-48}$. We previously proposed that the fixation of mutations reducing the intrinsic susceptibiliy to
353 carbapenems might have favored the efficient conjugative transfer of plasmids carrying the
354 carbapenemase genes from other CPE species in the gut (10). The transconjugant would be selected
355 in a context of low biliary excretion of carbapenems or other ß-lactams after parenteral administration.
356 Increasing the proportion of the donor and recipient bacteria would therefore be less necessary for
357 plasmids with high conjugative rates, such as pOXA-48 (27). Alternatively, these mutations might also
358 have favored the plasmid maintenance by increasing the resistance level and selection of the isolates
359 during ß-lactam or carbapenem treament. Of particular concern in France are the OXA-181 producing
360 ST410 and the NDM (NDM-1, -5 and -7) ST167 lineages that significantly contribute to the circulation
361 of carbapenem resistance. However our study also reveals smaller lineages, that although less
362 frequently encountered in international studies, are nevertheless circulating. In contrast we did not
363 obtain evidence for clonal dissemination of CP-*Ec* ST131 lineages derived from the B2 clade responsible
364 for *bla*$_{CTX-M-15}$ gene dissemination (19, 20).
365 Despite their high frequency among the F-NRC collection, the ST38 isolates do not enter into one of
366 the two previous categories. Indeed, while they mainly express *bla*$_{OXA-48}$ or its single-point-mutant

367 derivative *bla*$_{OXA-244}$ (31), our phylogenetic analysis revealed that a majority of them belong to four
368 different lineages, including one mostly associated with a rapid dissemination in France. In contrast
369 with OXA-181 and NDM-producing lineages previously described, these lineages are associated with a
370 moderate number of ARGs and QRDR mutations. None of the ST38 isolates is mutated in *ftsI*, however
371 we previously demonstrated that all are characterized by an *ompC* allele that encodes a porin with a
372 reduced permeability to certain ß-lactams and has disseminated into unrelated lineages by
373 homologous recombination (10). A common feature of the four lineages is the integration of part of
374 the pOXA-48 plasmid carrying *bla*$_{OXA-48}$. This could have reduced a fitness cost of this plasmid and
375 facilitated the clonal expansion of these lineages. However, ST38 isolates were identified as
376 unfrequent colonizer of the GI tract (29, 32). Therefore additional features of this CP-*Ec* ST38 lineages
377 might have contributed to its dissemination.
378 In conclusion, by analysing genome sequences of CP-*Ec* isolates collected by the F-NRC we showed that
379 MDR lineages, enriched in carbapenemase-producing isolates are circulating in France, some of them,
380 such as the ST1284 isolates being associated with outbreaks. It also suggests that the evolutionary
381 trajectory may depend on the carbapenemase gene, *bla*$_{OXA-181}$ or *bla*$_{NDM}$ genes being more frequently
382 associated with the evolution of MDR *E. coli* lineages characterized by mutations in *ftsI* and *ompC*.
383 Surveillance of these mutations may be an important parameter in controling the circulation of MDR
384 lineages. On the other hand, carbapenemase genes are also frequently acquired through plasmid
385 dissemination from other bacterial species. Depending on the resistance background of the receiver
386 *E. coli*, this may lead to XDR or to isolates sensitive to a broad range of antibiotics. Finally, we also
387 observed a strong and rapid dissemination of ST38 isolates that might have been favored by a reduced
388 susceptibility to carbapenems linked to the ST38 *ompC* allele and by the chromosomal integration of
389 the carbapenemase gene. These results strengthen our model of different evolutionary trajectories
390 associated with the gain of carbapenemase genes (10). They also show that systematic genome
391 sequencing of CP-*Ec* and at a larger of CPE isolates, irrespective of their clinical or resistance status is
392 able to provide useful information not only on the circulation of MDR lineages, but also on the
393 propagation of resistance genes through horizontal gene transfer.
394
395 **MATERIAL AND METHODS**
396 **Isolate collection and sequencing.**
397 CP-*Ec* isolates analyzed in this study were collected by the F-NRC, mainly between the years 2012 to
398 2015. Twenty-two additional isolates we have received before the creation of the F-NRC in 2012, were
399 included in the analysis. Information on these isolates as the year of isolation, the region and
400 department in France and summary of their genomic features are reported in Table S1.
401

402 **Whole genome sequencing and analyses.**

403 DNA were extracted by using the Qiagen Blood and Tissue DNA easy kit. Sequencing libraries were
404 constructed by using the Nextera XT kit (Illumina) following the manufacturer instruction, and
405 sequenced with the Illumina HiSeq2500 or NextSeq500. FastQ files were trimmed for adaptors and
406 low-quality bases (setting the minimum base quality threshold to 25) with the Cutadapt fork Atropos
407 (33). *De novo* assemblies were generated from the trimmed reads with SPAdes v3.12.0 (34), using k-
408 mer sizes of 51, 71, 81, and 91, the coverage cut-off option was set to "auto" and the "careful" option
409 was activated. QUAST v2.2 (35) was used to assess the assembly quality and contigs shorter than 500
410 bp were filtered out for phylogenetic analyses. The ST38 isolates CNR65D6 and CNR85I8 were
411 sequenced to completion by using the long-read Pacbio technology. PacBio reads were assembled with
412 the RS_HGAP_Assembly.3 protocol from the SMRT analysis toolkit v2.3 (36) and with Canu (37),
413 polished with Quiver (36) and manually corrected by mapping Illumina reads using Breseq (38).
414 Assembled genomes were annotated with Prokka v1.9 (39). To analyze the F-NRC CP*Ec* isolates
415 belonging to the main ST recovered during the analysis in a more global context, their genome
416 sequences were combined to the assembled genome sequences from the same ST retrieved from the
417 NCBI database (July, 2019) and in some STs from Enterobase (http://enterobase.warwick.ac.uk/, April
418 2021). Retrieved genomes were annotated in the same way as the genome sequences of the F-NRC
419 isolates (Table S2).

420 To generate a core-genome phylogeny tree of all isolates, the best reference genome was first selected
421 among a set of 18 genomes analyzed by Touchon et al. (40) using the software refRank (41) and three
422 subsets of 100 randomly selected sequences from our study as an input. This led to the selection of
423 the genome sequence of strain MG1655-K12 (NC_000913.3) as the reference for read mapping and
424 SNP identification. Sequence reads for the 17 reference genomes (40) as well as for the genome
425 sequence of *Escherichia fergunsonii* strain ATCC 35469 (NC_011740.1), used as an outgroup were
426 simulated with ART (42). Trimmed sequencing reads were mapped against the MG1655-K12 genome
427 with BWA-MEM algorithm of the BWA v0.7.4 package (43). For SNP calling the Genome Analysis Toolkit
428 (GATK) v3.6.0 (44) was used with the following criteria, a minimum depth coverage (DP) of 10, a quality
429 by depth (QD) bigger than 2, a fisher strain bias (FS) below 60, a root mean square of the mapping
430 quality (MQ) above 40, and the mapping quality rank sum test (MQRankSum) and the read position
431 rank sum test (ReadPosRankSum) greater than -12.5 and -8 respectively. A Maximum-Likelihood tree
432 was estimated with RAxML v8.28 (45) using core-genome SNPs after removing positions in the
433 accessory genome identified with the filter_BSR_variome.py script from the LS-BSR pipeline (46).

434 For ST with more than 10 CP-*Ec* isolates from the F-NRC, a core genome alignment was generated with
435 Parsnp (47), by using a finished genome sequence as reference. A closely related isolate outside the
436 ST lineage was selected from the global phylogenetic tree including all F-NRC isolates and used as

437 outgroup to root the ST phylogenetic trees. Maximum-Likelihood (ML) trees were generated with
438 RAxML v8.28 (45) after removing regions of recombination with Gubbins (48). All graphic
439 representations were performed by using ITOL (49).
440
441 **MLST type, phylogroups, resistance gene and plasmid replicon identification.**
442 Sequence type was assigned to each assembly through a python script that relies on BLAST (10).
443 Antibiotic resistance genes (ARGs) and mutations in *gyrA, parC* and *parE* quinolone resistance
444 determining regions (QRDR) were identified with Resfinder 4.0 and PointFinder (50) run in local
445 respectively. The scripts and database (Retrieved: January 4, 2021) were downloaded from the
446 repositories of the Centre for Genomic Epidemiology (https://bitbucket.org/genomicepidemiology/).
447 The identified ARGs were manually reviewed to eliminate potential redundant ARG predicted at the
448 same genomic position. *mdf*(A) that is present in *E. coli* core genome was not taken into account.
449 Phylogroups were assigned by using EzClermont (51) run in local. Plasmid replicons were identified
450 with plasmidfinder run on each assembly (14). *ftsI*, *ompC and ompF* CDS sequences and *ompF*
451 promoter sequences were identified by using BlastN. Translated or nucleotide sequences were
452 clustered by cd-hit (cd-hit-v4.8.1) with an amino acid (FtsI, OmpC, OmpF) or nucleotide sequence
453 (*ompF* promoter) identity threshold of 1. Sequences of each cluster were aligned to detect mutations
454 in regions of interest: four amino-acid insertions between P333 and Y334 and E349K and I532L
455 mutations (FtsI), mutation modifying charge in L3 constriction loop, R195L mutation, nonsense or
456 frameshift mutations or OmpC sequence clustering with ST-38 OmpC sequences (OmpC), nonsense
457 mutations (OmpF) or mutations affecting one of the OmpR-boxes: mutation -46T/C in OmpR-F3 box
458 and mutation -75ΔT in OmpR-F2 box (*ompF* promoter).
459
460 **Statistical analysis.**
461 Pearson's chi-squared tests were performed by using standard libraries contained within the R
462 statistics package (http://www.R-project.org)
463
464 **Availability of data.**
465 All sequence data have been deposited at DDBJ/EMBL/GenBank (BioProject PRJEB46636) and
466 bioSample identifiers for the Illumina sequence data are listed in Table S1. Complete genome
467 sequences of CNR65D6 and CNR85I8 and the long-read sequencing data have been deposited at
468 DDBJ/EMBL/GenBank with the accession number ERZ3517884 (PacBio reads, ERR6414227) and
469 ERZ3518335 (ERS6682837 PacBio reads) respectively.
470
471

**ACKNOWLEDGMENTS**

This work was supported by grants from the French National Research Agency (ANR-10-LABX-62-IBEID, ANR-10-LABX-33, INCEPTION project (PIA/ANR-16-CONV-0005)) and from the European Union's Horizon 2020 research and Innovation Program under grant agrement No 773830 (Project ARDIG, One Health EJP). Pengdbamba Dieudonné Zongo is a scholar of Ed525 CDV, Sorbonne Université, Paris.

**REFERENCES**

1. Medina M, Castillo-Pino E. 2019. An introduction to the epidemiology and burden of urinary tract infections. Ther Adv Urol 11:1756287219832172.
2. Dortet L, Cuzon G, Ponties V, Nordmann P. 2017. Trends in carbapenemase-producing Enterobacteriaceae, France, 2012 to 2014. Euro Surveill 22.
3. French National Reference Centre for antibiotic resistance, activity report 2019 - 2020. https://wwwcnr-resistance-antibiotiquesfr/bilans-dactivitehtml.
4. Saleem AF, Allana A, Hale L, Diaz A, Salinas R, Salinas C, Qureshi SM, Hotwani A, Rahman N, Khan A, Zaidi AK, Seed PC, Arshad M. 2020. The Gut of Healthy Infants in the Community as a Reservoir of ESBL and Carbapenemase-Producing Bacteria. Antibiotics (Basel) 9:286.
5. Bush K, Jacoby GA. 2010. Updated functional classification of beta-lactamases. Antimicrob Agents Chemother 54:969-76.
6. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW, Kasarskis A, Batty EM, Kos V, Wilson DJ, Phetsouvanh R, Wyllie D, Sokurenko E, Manges AR, Johnson TJ, Price LB, Peto TE, Johnson JR, Didelot X, Walker AS, Crook DW. 2016. Evolutionary History of the Global Emergence of the *Escherichia coli* Epidemic Clone ST131. MBio 7:e02162.
7. Gauthier L, Dortet L, Cotellon G, Creton E, Cuzon G, Ponties V, Bonnin RA, Naas T. 2018. Diversity of carbapenemase-producing *Escherichia coli* isolates, France 2012-2013. Antimicrob Agents Chemother doi:10.1128/aac.00266-18.
8. Ellaby N, Doumith M, Hopkins KL, Woodford N, Ellington MJ. 2019. Emergence of diversity in carbapenemase-producing *Escherichia coli* ST131, England, January 2014 to June 2016. Euro Surveill 24:1800627
9. van der Zwaluw K, Witteveen S, Wielders L, van Santen M, Landman F, de Haan A, Schouls LM, Bosch T. 2020. Molecular characteristics of carbapenemase-producing *Enterobacterales* in the Netherlands; results of the 2014-2018 national laboratory surveillance. Clin Microbiol Infect 26:1412.e7-1412.e12.
10. Patino-Navarrete R, Rosinski-Chupin I, Cabanel N, Gauthier L, Takissian J, Madec JY, Hamze M, Bonnin RA, Naas T, Glaser P. 2020. Stepwise evolution and convergent recombination underlie the global dissemination of carbapenemase-producing *Escherichia coli*. Genome Med 12:10.
11. Roer L, Overballe-Petersen S, Hansen F, Schønning K, Wang M, Røder BL, Hansen DS, Justesen US, Andersen LP, Fulgsang-Damgaard D, Hopkins KL, Woodford N, Falgenhauer L, Chakraborty T, Samuelsen Ø, Sjöström K, Johannesen TB, Ng K, Nielsen J, Ethelberg S, Stegger M, Hammerum AM, Hasman H. 2018. *Escherichia coli* Sequence Type 410 Is Causing New International High-Risk Clones. mSphere 3:e00337-18
12. Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. 2018. ClermonTyping: an easy-to-use and accurate *in silico* method for *Escherichia* genus strain phylotyping. Microb Genom 4:0.000192
13. Hendrickx APA, Landman F, de Haan A, Witteveen S, van Santen-Verheuvel MG, Schouls LM, The Dutch Cpe Surveillance Study G. 2021. bla (OXA-48)-like genome architecture among carbapenemase-producing *Escherichia coli* and *Klebsiella pneumoniae* in the Netherlands. Microb Genom 7:000512

519 14. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup
520 F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid
521 multilocus sequence typing. Antimicrob Agents Chemother 58:3895-903.
522 15. Turton JF, Doumith M, Hopkins KL, Perry C, Meunier D, Woodford N. 2016. Clonal expansion
523 of *Escherichia coli* ST38 carrying a chromosomally integrated OXA-48 carbapenemase gene. J
524 Med Microbiol 65:538-46.
525 16. Leverstein-van Hall MA, Dierikx CM, Cohen Stuart J, Voets GM, van den Munckhof MP, van
526 Essen-Zandbergen A, Platteel T, Fluit AC, van de Sande-Bruinsma N, Scharinga J, Bonten MJ,
527 Mevius DJ. 2011. Dutch patients, retail chicken meat and poultry share the same ESBL genes,
528 plasmids and strains. Clin Microbiol Infect 17:873-80.
529 17. Oteo J, Diestra K, Juan C, Bautista V, Novais A, Pérez-Vázquez M, Moyá B, Miró E, Coque TM,
530 Oliver A, Cantón R, Navarro F, Campos J. 2009. Extended-spectrum beta-lactamase-producing
531 *Escherichia coli* in Spain belong to a large variety of multilocus sequence typing types, including
532 ST10 complex/A, ST23 complex/A and ST131/B2. Int J Antimicrob Agents 34:173-6.
533 18. Manges AR, Mende K, Murray CK, Johnston BD, Sokurenko EV, Tchesnokova V, Johnson JR.
534 2017. Clonal distribution and associated characteristics of *Escherichia coli* clinical and
535 surveillance isolates from a military medical center. Diagn Microbiol Infect Dis 87:382-385.
536 19. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M,
537 Chattopadhyay S, Stegger M, Andersen PS, Pearson T, Riddell K, Rogers P, Scholes D, Kahl B,
538 Keim P, Sokurenko EV. 2013. The Epidemic of Extended-Spectrum-beta-Lactamase-Producing
539 *Escherichia coli* ST131 Is Driven by a Single Highly Pathogenic Subclone, H30-Rx. MBio
540 4:e00377-13
541 20. Nicolas-Chanoine MH, Bertrand X, Madec JY. 2014. *Escherichia coli* ST131, an intriguing clonal
542 group. Clin Microbiol Rev 27:543-74.
543 21. Decano AG, Downing T. 2019. An *Escherichia coli* ST131 pangenome atlas reveals population
544 structure and evolution across 4,071 isolates. Sci Rep 9:17394.
545 22. Potron A, Bernabeu S, Cuzon G, Pontiès V, Blanchard H, Seringe E, Naas T, Nordmann P, Dortet
546 L. 2017. Analysis of OXA-204 carbapenemase-producing Enterobacteriaceae reveals possible
547 endoscopy-associated transmission, France, 2012 to 2014. Euro Surveill 22:17-00048
548 23. Lepelletier D, Berthelot P, Lucet JC, Fournier S, Jarlier V, Grandbastien B. 2015. French
549 recommendations for the prevention of 'emerging extensively drug-resistant bacteria' (eXDR)
550 cross-transmission. J Hosp Infect 90:186-95.
551 24. Ortega A, Sáez D, Bautista V, Fernández-Romero S, Lara N, Aracil B, Pérez-Vázquez M, Campos
552 J, Oteo J. 2016. Carbapenemase-producing *Escherichia coli* is becoming more prevalent in
553 Spain mainly because of the polyclonal dissemination of OXA-48. J Antimicrob Chemother
554 71:2131-8.
555 25. Grundmann H, Glasner C, Albiger B, Aanensen DM, Tomlinson CT, Andrasević AT, Cantón R,
556 Carmeli Y, Friedrich AW, Giske CG, Glupczynski Y, Gniadkowski M, Livermore DM, Nordmann
557 P, Poirel L, Rossolini GM, Seifert H, Vatopoulos A, Walsh T, Woodford N, Monnet DL. 2017.
558 Occurrence of carbapenemase-producing *Klebsiella pneumoniae* and *Escherichia coli* in the
559 European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective,
560 multinational study. Lancet Infect Dis 17:153-163.
561 26. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, Abudahab K, Goater R, Giani T,
562 Errico G, Aspbury M, Sjunnebo S, Feil EJ, Rossolini GM, Aanensen DM, Grundmann H. 2019.
563 Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial
564 spread. Nat Microbiol 4:1919-1929.
565 27. Potron A, Poirel L, Nordmann P. 2014. Derepressed transfer properties leading to the efficient
566 spread of the plasmid encoding carbapenemase OXA-48. Antimicrob Agents Chemother
567 58:467-71.
568 28. León-Sampedro R, DelaFuente J, Díaz-Agero C, Crellen T, Musicha P, Rodríguez-Beltrán J, de la
569 Vega C, Hernández-García M, López-Fresneña N, Ruiz-Garbajosa P, Cantón R, Cooper BS, San

570 Millán Á. 2021. Pervasive transmission of a carbapenem resistance plasmid in the gut
571 microbiota of hospitalized patients. Nat Microbiol 6:606-616.
572 29. Matsui Y, Hu Y, Rubin J, de Assis RS, Suh J, Riley LW. 2020. Multilocus sequence typing of
573 *Escherichia coli* isolates from urinary tract infection patients and from fecal samples of healthy
574 subjects in a college community. Microbiologyopen 9:1225-1233.
575 30. David S, Cohen V, Reuter S, Sheppard AE, Giani T, Parkhill J, Rossolini GM, Feil EJ, Grundmann
576 H, Aanensen DM. 2020. Integrated chromosomal and plasmid sequence analyses reveal
577 diverse modes of carbapenemase gene spread among *Klebsiella pneumoniae*. Proc Natl Acad
578 Sci U S A 117:25043-25054.
579 31. Emeraud C, Girlich D, Bonnin RA, Jousset AB, Naas T, Dortet L. 2021. Emergence and Polyclonal
580 Dissemination of OXA-244-Producing *Escherichia coli*, France. Emerg Infect Dis 27:1206-1210.
581 32. Massot M, Daubié AS, Clermont O, Jauréguy F, Couffignal C, Dahbi G, Mora A, Blanco J, Branger
582 C, Mentré F, Eddi A, Picard B, Denamur E, The Coliville G. 2016. Phylogenetic, virulence and
583 antibiotic resistance characteristics of commensal strain populations of *Escherichia coli* from
584 community subjects in the Paris area in 2010 and evolution over 30 years. Microbiology
585 (Reading) 162:642-650.
586 33. Didion JP, Martin M, Collins FS. 2017. Atropos: specific, sensitive, and speedy trimming of
587 sequencing reads. PeerJ 5:e3720.
588 34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,
589 Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.
590 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell
591 sequencing. J Comput Biol 19:455-77.
592 35. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome
593 assemblies. Bioinformatics 29:1072-5.
594 36. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A,
595 Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome
596 assemblies from long-read SMRT sequencing data. Nat Methods 10:563-9.
597 37. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and
598 accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res
599 27:722-736.
600 38. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes
601 from next-generation sequencing data using breseq. Methods Mol Biol 1151:165-88.
602 39. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.
603 40. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S,
604 Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M,
605 Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguenec C, Lescat
606 M, Mangenot S, Martinez-Jehanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf
607 CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Medigue C, Rocha EP, Denamur E. 2009.
608 Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive
609 paths. PLoS Genet 5:e1000344.
610 41. Becker L, Fuchs S, Pfeifer Y, Semmler T, Eckmanns T, Korr G, Sissolak D, Friedrichs M, Zill E,
611 Tung ML, Dohle C, Kaase M, Gatermann S, Rüssmann H, Steglich M, Haller S, Werner G. 2018.
612 Whole Genome Sequence Analysis of CTX-M-15 Producing *Klebsiella Isolates* Allowed
613 Dissecting a Polyclonal Outbreak Scenario. Front Microbiol 9:322.
614 42. Huang W, Li L, Myers JR, Marth GT. 2012. ART: a next-generation sequencing read simulator.
615 Bioinformatics 28:593-4.
616 43. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
617 Bioinformatics 25:1754-60.
618 44. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler
619 D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce
620 framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-303.

621 45. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
622 phylogenies. Bioinformatics 30:1312-3.
623 46. Sahl JW, Caporaso JG, Rasko DA, Keim P. 2014. The large-scale blast score ratio (LS-BSR)
624 pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ
625 2:e332.
626 47. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome
627 alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol
628 15:524.
629 48. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015.
630 Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome
631 sequences using Gubbins. Nucleic Acids Res 43:e15.
632 49. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
633 developments. Nucleic Acids Res 47:W256-w259.
634 50. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo
635 AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K,
636 Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett
637 NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S,
638 Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from
639 genotypes. J Antimicrob Chemother doi:10.1093/jac/dkaa345.
640 51. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. 2020. Easy phylotyping of Escherichia
641 coli via the EzClermont web app and command-line tool. Access Microbiol 2:acmi000143.
642

643 **Figure legends**

644 **Figure 1. Origin of the isolates received by the F-NRC during the 2012-2015 period**. **A**. Absolute
645 number of isolates per year from infection, screening or unknown origins, differentiating between
646 isolates that have or not been sequenced. **B**. Proportions of isolates from infection or from screening
647 or unknown origins showing that isolates from infection origin represent the same proportions among
648 total isolates that have been received by the F-NRC and isolates that have sequenced.

649

650 **Figure 2: Core genome phylogeny of CP-*Ec* isolates received by the F-NRC**. ML phylogeny of the 713
651 CP-*Ec* isolates was built with RAxML (45) from 372,238 core SNPs after sequence alignment on
652 MG1655-K12 (NC_000913.3) selected as the best reference. The genome sequence of *Escherichia*
653 *fergunsonii* strain ATCC 35469 (NC_011740.1) was used as an outgroup for the phylogenetic analysis.
654 Genomes from Touchon et al. (40) were also incorporated into the analysis. Genomic features are
655 indicated as in the figure key (left) from the inside to the outside circles: carbapenemases of the OXA,
656 NDM and other types, CTX-M ESBL, mutations in *gyrA* and *parC* QRDR region (FQ resistance), origin,
657 main ST, phylogroups. Reference genome from Touchon *et al*. are indicated as Touchon.

658

659 **Figure 3. Resistance gene content of the F-NRC CP-*Ec* isolates**. **A**. Box plot representation of the ARGs
660 content according to the phylogroup. The number of isolates belonging to each phylogroup is indicated
661 between brackets. The limits of the box indicate the lower and upper quartile. Outliers are indicated
662 by points above the maximal value. **B**: Relationships between the ARGs number and the number of

663 mutations in QRDR. In the box plot representations, the median is indicated by an horizontal bar and
664 the mean by a cross.
665
666 **Figure 4. Carbapenemase gene content of the F-NRC CP-*Ec* isolates**. **A**. Relationship between the
667 carbapenemase allele and the ARGs content. In the box plot representation, the median is indicated
668 by an horizontal bar and the mean by a cross. The limits of the box indicate the lower and upper
669 quartile. Outliers are indicated by points above the maximal value. **B**. Number of isolates carrying a
670 specific carbapenemase allele according to the number of QRDR mutations in their genomes. **C**. Per-
671 year evolution of the percentage of isolates carrying the different carbapenemase alleles in the 2012-
672 2015 period. Year of isolation are indicated as in the figure key (right). Given their small number (n=22),
673 strains isolated before 2012 are not indicated**.**
674
675 **Figure 5**. Features of the main ST associated with CP-*Ec* isolates in France. **A**. Proportion of the isolates
676 according to the carbapenemase allele in each ST; **B**. Box plot representation of the ARGs content
677 according to the ST; **C.** Proportion of isolates according to the number of mutations in QRDR; **D** :
678 Proportion of isolates with a specific PBP3 allele. PBP3 characterized by a YRIP, YRIK or YRIN insertion
679 between positions 333 and 334, the YRIN insertion is generally associated with a A413V mutation
680 (YRIN-L) and less frequently with a E349K mutation (YRIN-K-L) (10); **E.** Proportion of isolates carrying
681 specific mutations in *ompC:* R195L, G138D or acquisition of a ST38-like *ompC* allele by recombination
682 (10). For each ST, the phylogroup, the total number of isolates and the number of isolates associated
683 with urinary tract infections are indicated in the upper panel. Intermediate (Inter) and minor ST
684 indicate ST represented by 4-6 and 1-3 isolates, respectively.
685
686 **Figure 6. Core genome phylogeny of ST38 isolates.** ML Phylogeny was based on genome sequences
687 of 92 CP-*Ec* from the F-NRC and 464 genome sequences retrieved from Enterobase and from the NCBI
688 database including 331 carrying a carbapenemase gene. A core genome (2,900,000 nt) alignment of
689 the *de novo* assemblies on the sequence of GCA_005886035.1 used as a reference was performed by
690 using Parsnp (47); ML phylogeny was built with RAxML (45) from 6,170 core SNPs after removing
691 recombined regions with Gubbins (48). The genome sequence of CNRC6O47 (ST963) was used as an
692 outgroup for the phylogenetic analysis. F-NRC isolates are indicated by red triangles (inner circle).
693 Other genomic features are indicated as indicated in the figure key (left) from the inside to the outside
694 circles: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA* and *parC*
695 QRDR region (FQ resistance), number of ARGs, mutations in *ftsI*, *ompC*, *ompF*, geographical origin. The
696 four OXA-48-like clades (G1, G2, G3, G4) clustering most French isolates are colored in blue, red, violet

697 and green. A fifth clade (G5) corresponding to a possible outbreak of VIM-4 isolates in the East of Paris
698 is colored in brown
699
700 **Supplementary figures**
701 **Figure S1**. Per-year analysis of the origin of the isolates received by the F-NRC. **A**. Proportions of the
702 isolates belonging to the different ST. The total number of isolates is indicated between brackets. The
703 large proportion of ST90 isolates collected in 2012 is linked to a local outbreak. **B**. Proportions of the
704 isolates collected in screening and infection situations as a function of the year and the phylogroup.
705 The absolute number of isolates for each class is indicated inside the bars.
706
707 **Figure S2**. Core genome phylogenies of the main ST characterized by clades disseminating
708 internationally. A. ST410; B. ST167; C. ST405; D. ST101. Phylogenies were based on genome sequences
709 of A: 64 CP-*Ec* from the F-NRC and 146 genome sequences from the NCBI database;B: 35 CP-*Ec* from
710 the F-NRC and 134 genome sequences from the NCBI database; C. 15 CP-*Ec* from the F-NRC and 145
711 genome sequences from the NCBI database; D. 13 CP-*Ec* from the F-NRC and 194 genome sequences
712 from the NCBI database. Core genome (ST410: 3,522,000 nt; ST167: 3,276,000 nt; ST405: 3,685,000
713 nt; ST101: 3,774,000 nt) alignments of the *de novo* assemblies on the sequences of GCA_001442495.1
714 (ST410), GCA_003028815.1 (ST167), GCA_002142675.1 (ST405), GCA_002163655.1 (ST101) used as
715 reference were performed by using Parsnp (47); ML phylogeny was built with RAxML (45) from 6,176
716 (ST410), 10,393 (ST167), 37,482 (ST405), 19,967(ST101) core SNPs after removing recombined regions
717 with gubbins (48). The genome sequences of CNR93E7 (ST88), CNR93D10 (ST746), CNR73I9 (ST115),
718 CNR93I2 (ST906) were used as outgroups for the phylogenetic analyses of ST410, ST167, ST405 and
719 ST101 respectively. The origin from the F-NRC is indicated by red triangles close to the isolate name.
720 Other genomic features are indicated as indicated in the figure key (left) from the inside to the outside
721 lines: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA* and *parC* QRDR
722 region (FQ resistance), number of ARGs, mutations in *ftsI*, *ompC*, *ompF*, geographical origin.
723
724 **Figure S3**. Core genome phylogenies of ST10 and ST131 isolates. **A.** ST10 phylogeny based on the
725 genome sequences of 67 CP-*Ec* from the F-NRC and 153 genome sequences from the NCBI. **B.** ST131
726 phylogeny based on the genome sequences of 17 CP-*Ec* from the F-NRC and 462 genome sequences
727 from the NCBI. Core genome (ST10: 813,000 nt; ST131: 1,641,000 nt) alignments of the *de novo*
728 assemblies on the sequence of MG1655 (ST10) or GCA_000285655.3 (ST131) used as references were
729 performed by using Parsnp (47); ML phylogeny was built with RAxML (45) from 20,245 (ST10) and
730 14,366 (ST131) core SNPs after removing recombined regions with Gubbins (48). The genome
731 sequences of CNR93D10 (ST746) and CNRAL47G10 (ST640) were used as outgroups for the

732 phylogenetic analyses of ST10 and ST131, respectively. F-NRC isolates are indicated by red triangles
733 (inner circle). Other genomic features are indicated as indicated in the figure key (left) from the inside
734 to the outside circles: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA*
735 and *parC* QRDR region (FQ resistance), number of ARGs, mutations in *ftsI*, *ompC*, *ompF*, geographical
736 origin. The ST10 clade enriched in CP-*Ec* is indicated in green and the sub-clade with the YRIN insertion
737 in PBP3 in red. Note that the circle for *ftsI* mutations is absent in ST131, as no YRIN-like duplication in
738 PBP3 was identified among the analyzed isolates.
739
740 **Figure S4.** Core genome phylogeny of ST940 and ST1284 isolates. **A:** ST940 phylogeny based on
741 genome sequences of 17 CP-*Ec* from the F-NRC and 88 genome sequences from Enterobase and the
742 NCBIdatabase **B:** ST1284 phylogeny based on genome sequences of 14 CP-*Ec* from the F-NRC and 60
743 genome sequences from Enterobase and the NCBI database. A core genome (ST940: 3,639,000 nt;
744 ST1284: 3,779,000 nt) alignment of the *de novo* assemblies on the sequence of ST940: ESC_LB2149
745 and ST1284: ESC_LB2152AA (from Enterobase) used as references was performed by using Parsnp
746 (47); ML phylogeny was built with RAxML (45) from 13,859 (ST940) and 2,009 (ST1284) core SNPs after
747 removing recombined regions with Gubbins (48). The genome sequences of CNR98G1 (ST3022) and
748 MG1655 (ST10) were used as outgroups for the phylogenetic analyses of ST940 and ST1284,
749 respectively. F-NRC isolates are indicated by red triangles first column on the left. Other genomic
750 features are as indicated in the figure key (left) from the left to the right columns: carbapenemases of
751 the OXA, NDM and other types, number of mutations in *gyrA* and *parC* QRDR region (FQ resistance),
752 number of ARGs, mutations in *ftsI*, *ompC*, *ompF*, geographical origin.
753
754 **Figure S5**. Core genome phylogenies of ST69, ST88 and ST90. isolates. **A:** ST90 phylogeny based on the
755 15 CP-*Ec* from the F-NRC and 38 genome sequences from the NCBI database**; B:** ST69 phylogeny based
756 on the 16 CP-*Ec* from the F-NRC and 244 genome sequences from the NCBI database; **C.** ST88
757 phylogeny based on the 15 CP-*Ec* from the F-NRC and 99 genome sequences from the NCBI database.
758 Core genome (ST69: 2,735,000 nt; ST88: 3,472,000 nt; ST90: 3,801,000nt) alignments of the *de novo*
759 assemblies on the sequence of GCA_002443135.1 (ST69), GCA_002812685.1 (ST88) or
760 GCA_001900635.1 (ST90) used as references were performed by using Parsnp (47); ML phylogeny was
761 built with RAxML (45) from 4,596 (ST69), 12,557 (ST88) or 2,878 (ST90) core SNPs after removing
762 recombined regions with Gubbins (48). The genome sequences of CNR33D9 (ST394), CNR83B9 (ST410)
763 and CNR88B9 (ST847) were used as outgroups for the phylogenetic analyses of ST69, ST88 and ST90,
764 respectively. F-NRC isolates are indicated by red triangles (first column on the left). Other genomic
765 features are indicated as indicated in the figure key (left) from the left to the right columns:

766 carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA* and *parC* QRDR
767 region (FQ resistance), number of ARGs, mutations in *ftsI*, *ompC*, *ompF*, geographical origin.
768
769 **Figure S6**. Core genome phylogeny of ST617 and ST648 isolates based on genome sequences of **A**:
770 ST617 genome sequences of 15 CP-*Ec* from the F-NRC and 65 genome sequences from the NCBI
771 database; **B**: ST648 genome sequences of 14 CP-*Ec* from the F-NRC and 126 genome sequences from
772 the NCBI database. A core genome (ST617: 3,407,000; ST648: 3,501,000 nt) alignment of the *de novo*
773 assemblies on the sequences of GCA_002142695.1 (ST617) or GCA_004138645.1 (ST648) used as
774 references was performed by using Parsnp (47); ML phylogeny was built with RAxML (45) from 15,030
775 (ST617) or 3,782 (ST648) core SNPs after removing recombined regions with Gubbins (48). The genome
776 sequences of CNR93D10 (ST746) and CNR71A8 (ST1485) were used as outgroups for the phylogenetic
777 analyses of ST617 and ST648, respectively. F-NRC isolates are indicated by red triangles close to the
778 isolate name. Other genomic features are indicated as indicated in the figure key (left) from the left to
779 the right columns: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA*
780 and *parC* QRDR region (FQ resistance), number of ARGs, mutations in *ftsI*, *ompC*, *ompF*, geographical
781 origin.
782
783 Supplementary Tables
784 **Table S1**: Main characteristics of the F-NRC CP-Ec
785 **Table S2**: Main characteristics of genome sequences used for phylogenetic reconstructions

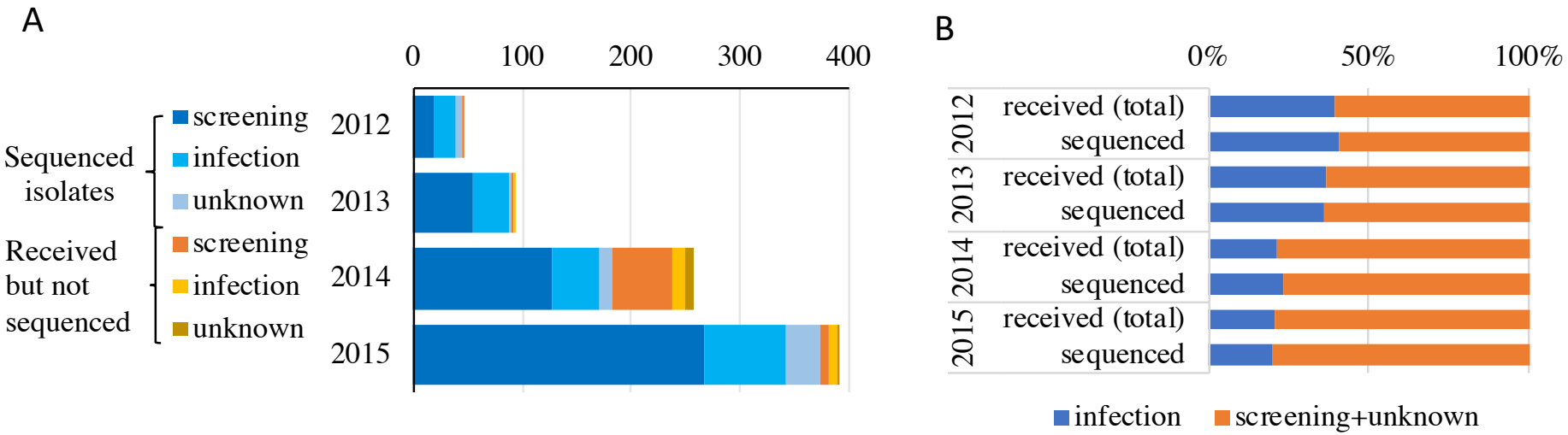

A
0
100
200
300
400
Sequenced isolates
screening
infection
unknown
Received but not sequenced
screening
infection
unknown
2012
2013
2014
2015
B
0%
50%
100%
2012
received (total)
sequenced
2013
received (total)
sequenced
2014
received (total)
sequenced
2015
received (total)
sequenced
infection
screening+unknown

Fig. 1

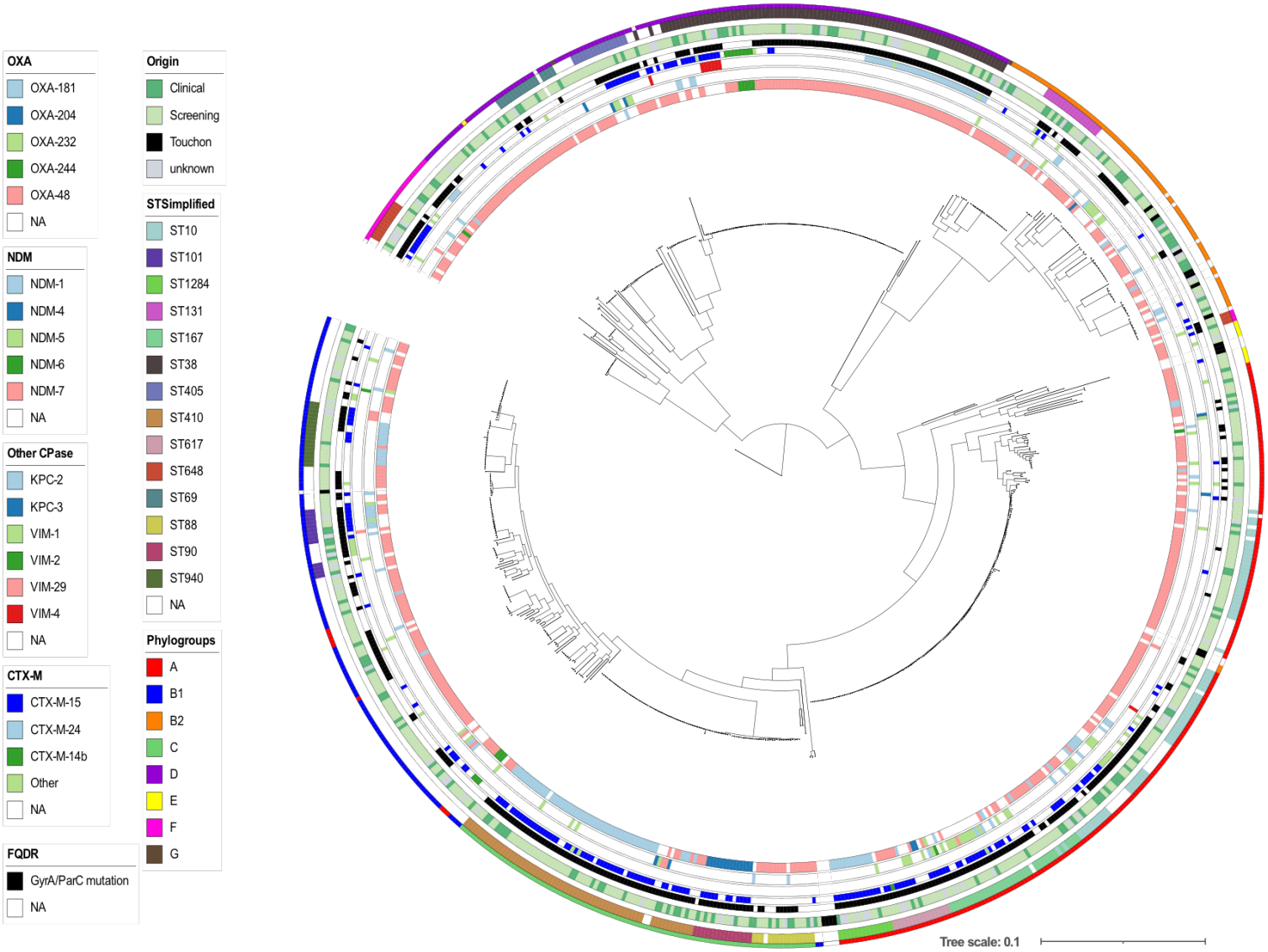

OXA
OXA-181
OXA-204
OXA-232
OXA-244
OXA-48
NA
NDM
NDM-1
NDM-4
NDM-5
NDM-6
NDM-7
NA
Other CPase
KPC-2
KPC-3
VIM-1
VIM-2
VIM-29
VIM-4
NA
CTX-M
CTX-M-15
CTX-M-24
CTX-M-14b
Other
NA
FQDR
GyrA/ParC mutation
NA
Origin
Clinical
Screening
Touchon
unknown
STSimplified
ST10
ST101
ST1284
ST131
ST167
ST38
ST405
ST410
ST617
ST648
ST69
ST88
ST90
ST940
NA
Phylogroups
A
B1
B2
C
D
E
F
G
Tree scale: 0.1

Fig. 2

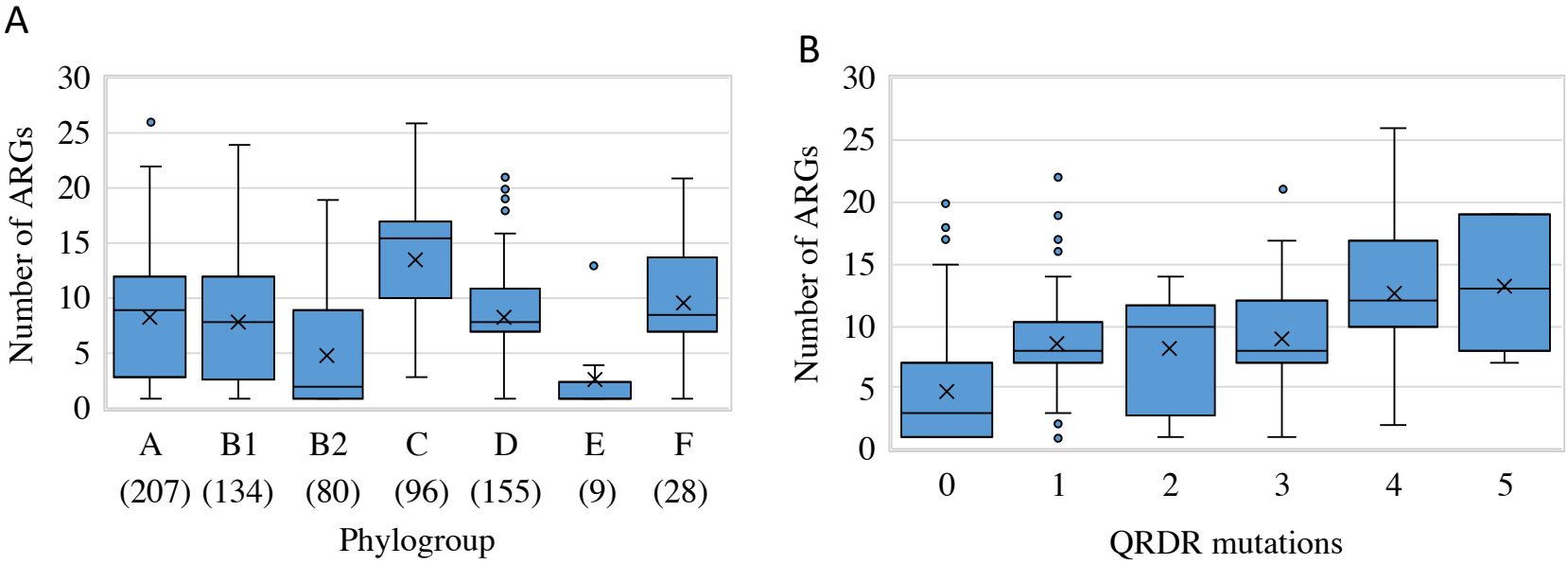

A
Number of ARGs
30
25
20
15
10
5
0
A
(207)
B1
(134)
B2
(80)
C
(96)
D
(155)
E
(9)
F
(28)
Phylogroup
B
Number of ARGs
30
25
20
15
10
5
0
0
1
2
3
4
5
QRDR mutations

Fig. 3

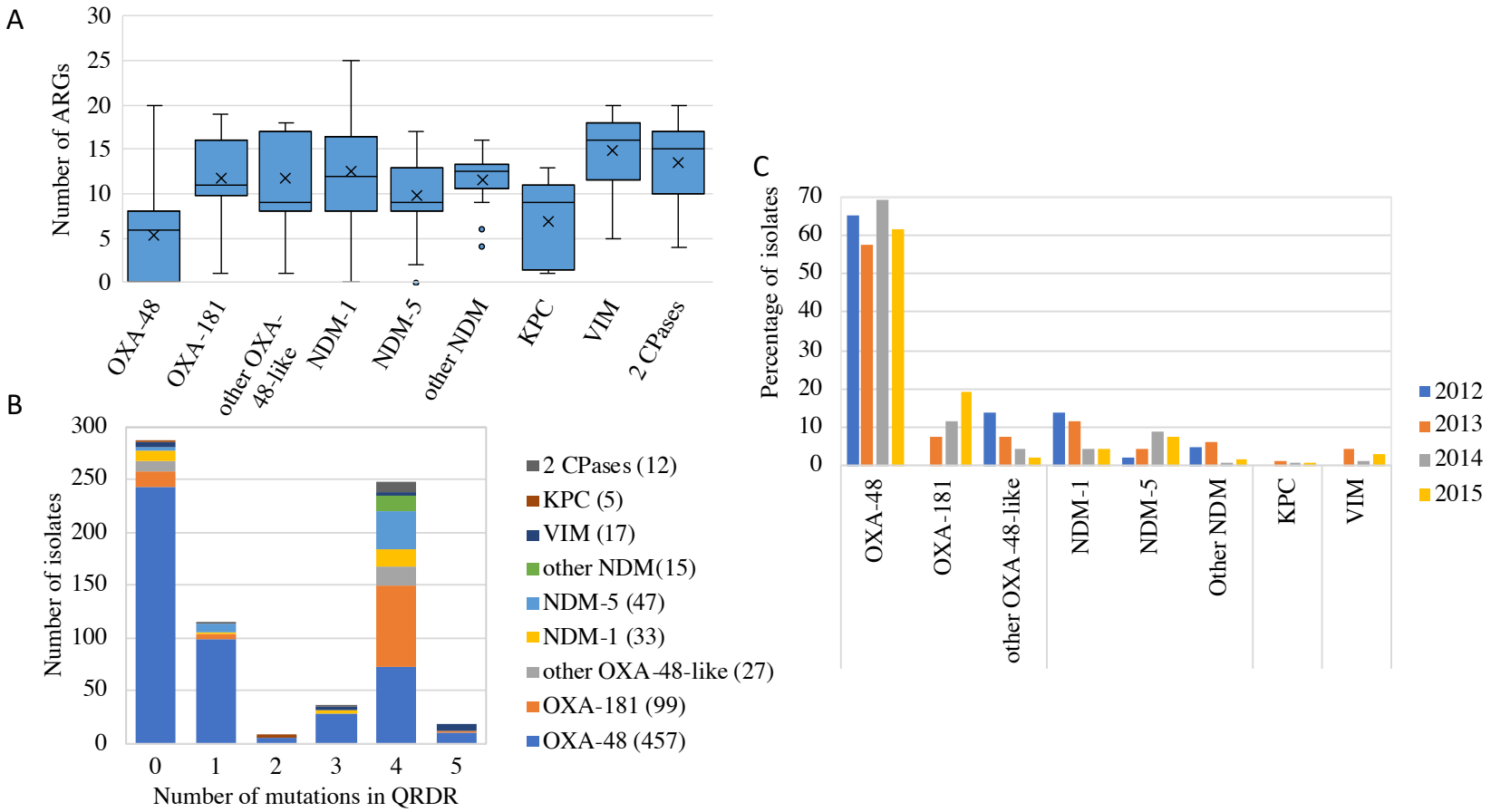

A
Number of ARGs
30
25
20
15
10
5
0
OXA-48
OXA-181
other OXA-48-like
NDM-1
NDM-5
other NDM
KPC
VIM
2 CPases
B
Number of isolates
300
250
200
150
100
50
0
0
1
2
3
4
5
Number of mutations in QRDR
2 CPases (12)
KPC (5)
VIM (17)
other NDM(15)
NDM-5 (47)
NDM-1 (33)
other OXA-48-like (27)
OXA-181 (99)
OXA-48 (457)
C
Percentage of isolates
70
60
50
40
30
20
10
0
OXA-48
OXA-181
other OXA-48-like
NDM-1
NDM-5
Other NDM
KPC
VIM
2012
2013
2014
2015

Fig. 4

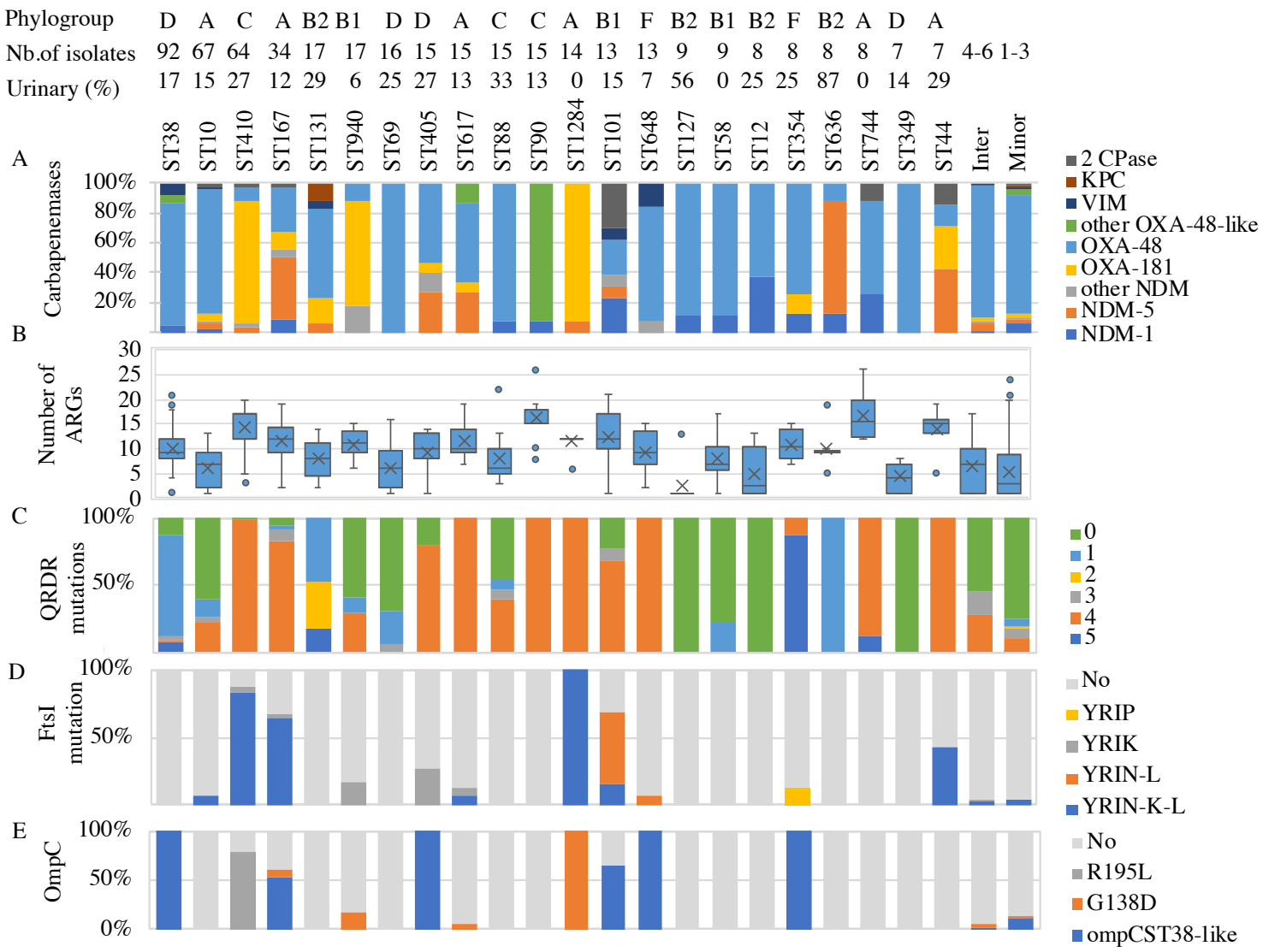

Phylogroup
Nb.of isolates
Urinary (%)
ST38 ST10 ST410 ST167 ST131 ST940 ST69 ST405 ST617 ST88 ST90 ST1284 ST101 ST648 ST127 ST58 ST12 ST354 ST636 ST744 ST349 ST44 Inter Minor
A
Carbapenemases
100%
80%
60%
40%
20%
2 CPase
KPC
VIM
other OXA-48-like
OXA-48
OXA-181
other NDM
NDM-5
NDM-1
B
Number of ARGs
30
25
20
15
10
5
0
C
QRDR mutations
100%
50%
0
1
2
3
4
5
D
FtsI mutation
100%
50%
No
YRIP
YRIK
YRIN-L
YRIN-K-L
E
OmpC
100%
50%
0%
No
R195L
G138D
ompCST38-like

Fig. 5

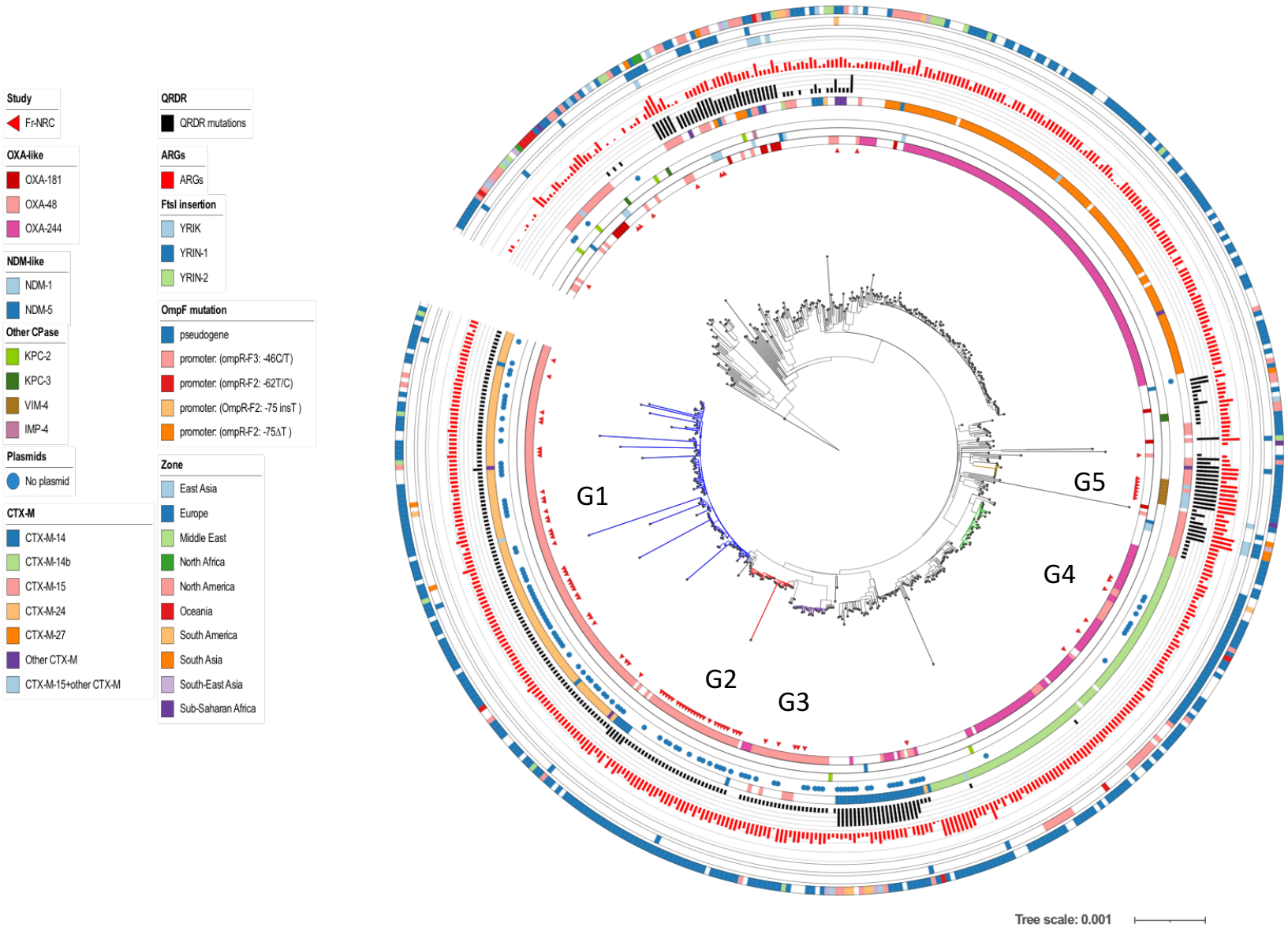
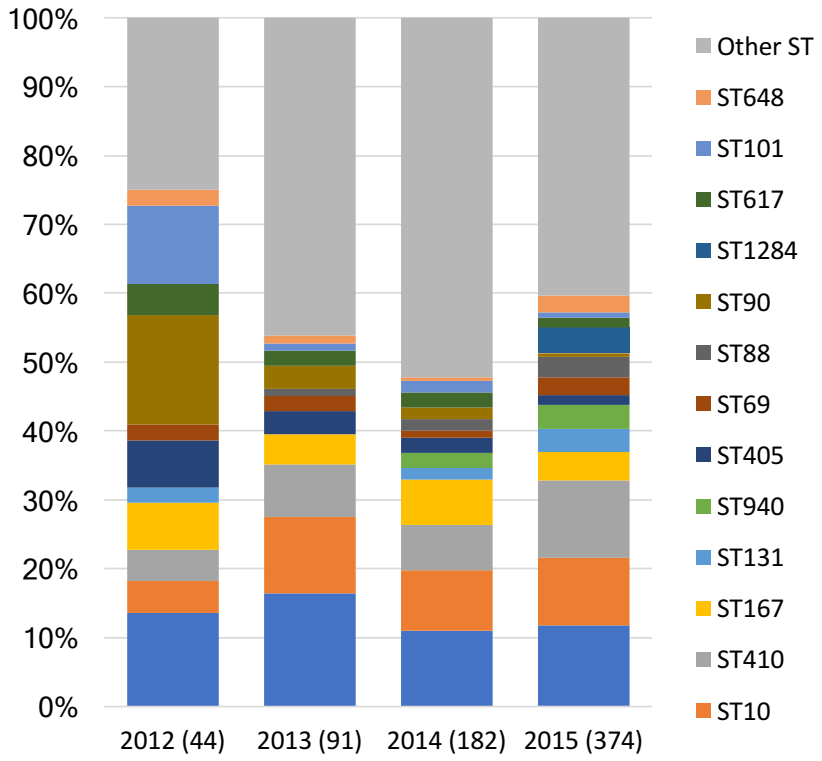

Study
Fr-NRC
OXA-like
OXA-181
OXA-48
OXA-244
NDM-like
NDM-1
NDM-5
Other CPase
KPC-2
KPC-3
VIM-4
IMP-4
Plasmids
No plasmid
CTX-M
CTX-M-14
CTX-M-14b
CTX-M-15
CTX-M-24
CTX-M-27
Other CTX-M
CTX-M-15+other CTX-M
QRDR
QRDR mutations
ARGs
ARGs
FtsI insertion
YRIK
YRIN-1
YRIN-2
OmpF mutation
pseudogene
promoter: (ompR-F3: -46C/T)
promoter: (ompR-F2: -62T/C)
promoter: (OmpR-F2: -75 insT )
promoter: (ompR-F2: -75ΔT )
Zone
East Asia
Europe
Middle East
North Africa
North America
Oceania
South America
South Asia
South-East Asia
Sub-Saharan Africa
G1
G2
G3
G4
G5
Tree scale: 0.001

Fig. 6

A


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
2012 (44)
2013 (91)
2014 (182)
2015 (374)
Other ST
ST648
ST101
ST617
ST1284
ST90
ST88
ST69
ST405
ST940
ST131
ST167
ST410
ST10

B

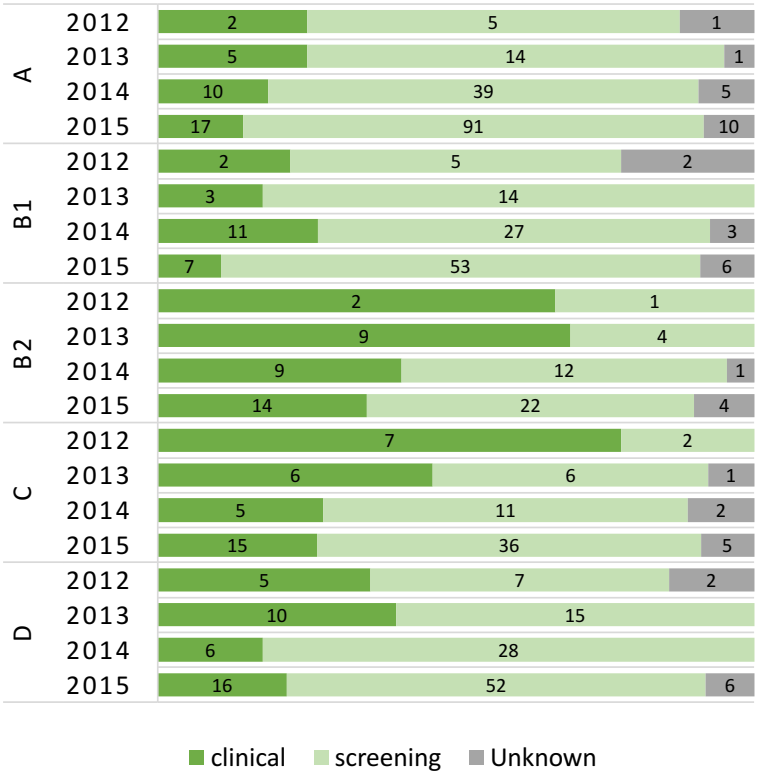

A
2012 2 5 1
2013 5 14 1
2014 10 39 5
2015 17 91 10
B1
2012 2 5 2
2013 3 14
2014 11 27 3
2015 7 53 6
B2
2012 2 1
2013 9 4
2014 9 12 1
2015 14 22 4
C
2012 7 2
2013 6 6 1
2014 5 11 2
2015 15 36 5
D
2012 5 7 2
2013 10 15
2014 6 28
2015 16 52 6
clinical
screening
Unknown

Fig. S1

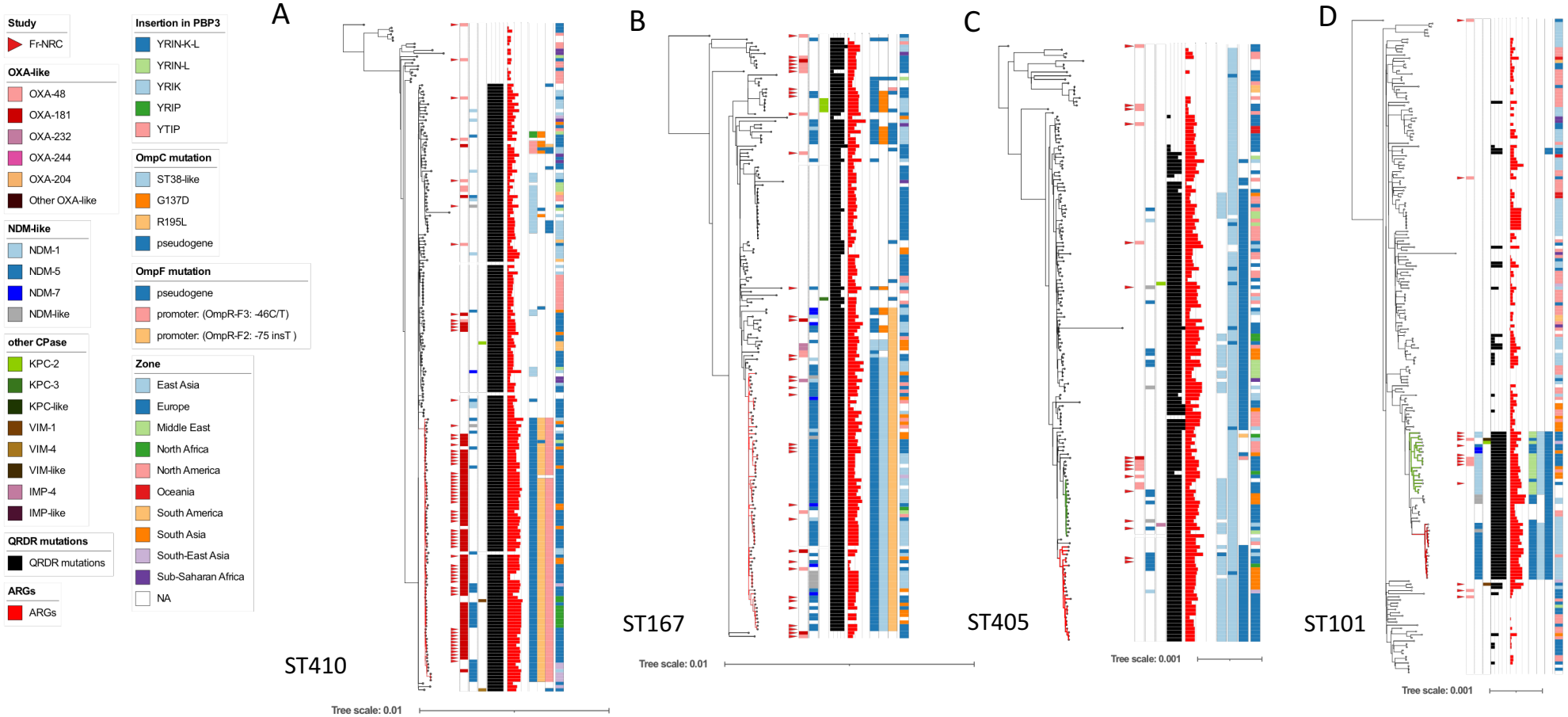

Study
Fr-NRC
OXA-like
OXA-48
OXA-181
OXA-232
OXA-244
OXA-204
Other OXA-like
NDM-like
NDM-1
NDM-5
NDM-7
NDM-like
other CPase
KPC-2
KPC-3
KPC-like
VIM-1
VIM-4
VIM-like
IMP-4
IMP-like
QRDR mutations
QRDR mutations
ARGs
ARGs
Insertion in PBP3
YRIN-K-L
YRIN-L
YRIK
YRIP
YTIP
OmpC mutation
ST38-like
G137D
R195L
pseudogene
OmpF mutation
pseudogene
promoter: (OmpR-F3: -46C/T)
promoter: (OmpR-F2: -75 insT )
Zone
East Asia
Europe
Middle East
North Africa
North America
Oceania
South America
South Asia
South-East Asia
Sub-Saharan Africa
NA
A
ST410
Tree scale: 0.01
B
ST167
Tree scale: 0.01
C
ST405
Tree scale: 0.001
D
ST101
Tree scale: 0.001

Fig. S2

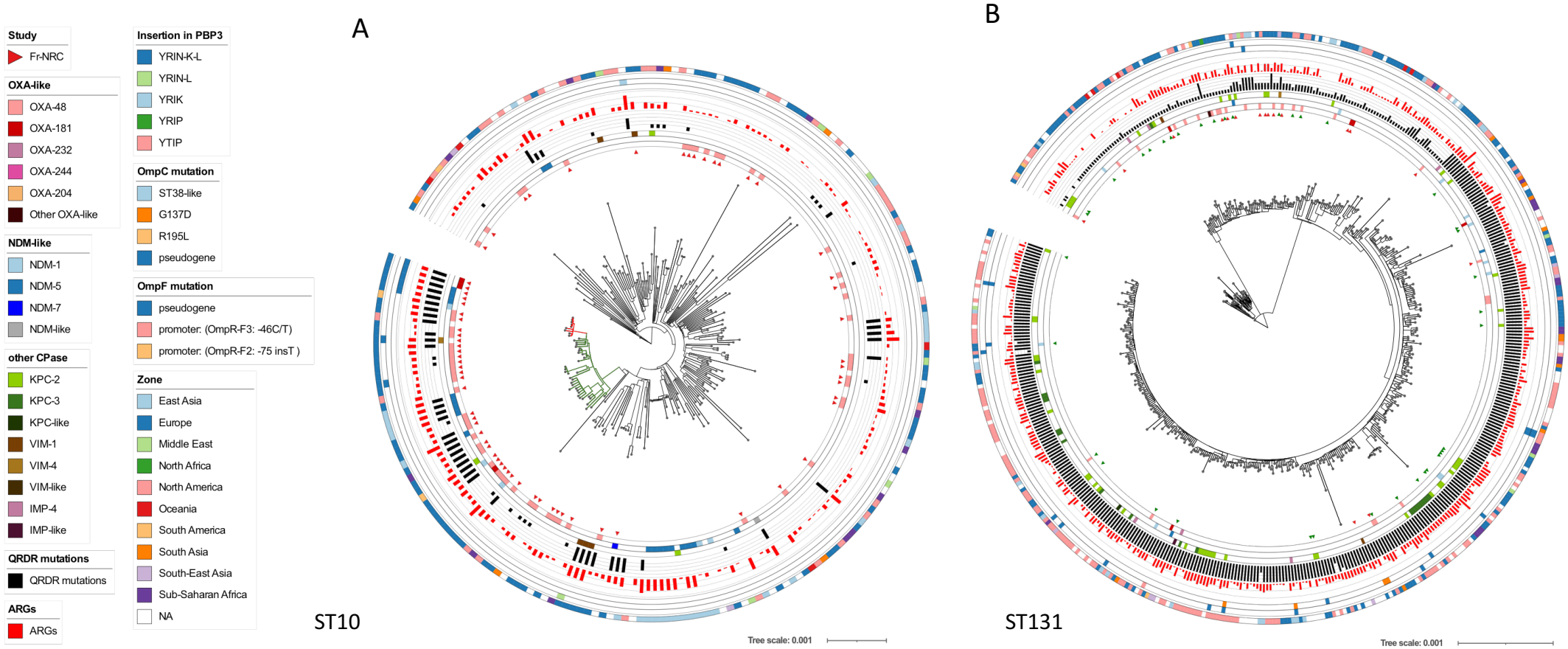

Study
Fr-NRC
OXA-like
OXA-48
OXA-181
OXA-232
OXA-244
OXA-204
Other OXA-like
NDM-like
NDM-1
NDM-5
NDM-7
NDM-like
other CPase
KPC-2
KPC-3
KPC-like
VIM-1
VIM-4
VIM-like
IMP-4
IMP-like
QRDR mutations
QRDR mutations
ARGs
ARGs
Insertion in PBP3
YRIN-K-L
YRIN-L
YRIK
YRIP
YTIP
OmpC mutation
ST38-like
G137D
R195L
pseudogene
OmpF mutation
pseudogene
promoter: (OmpR-F3: -46C/T)
promoter: (OmpR-F2: -75 insT )
Zone
East Asia
Europe
Middle East
North Africa
North America
Oceania
South America
South Asia
South-East Asia
Sub-Saharan Africa
NA
A
ST10
Tree scale: 0.001
B
ST131
Tree scale: 0.001

Fig. S3

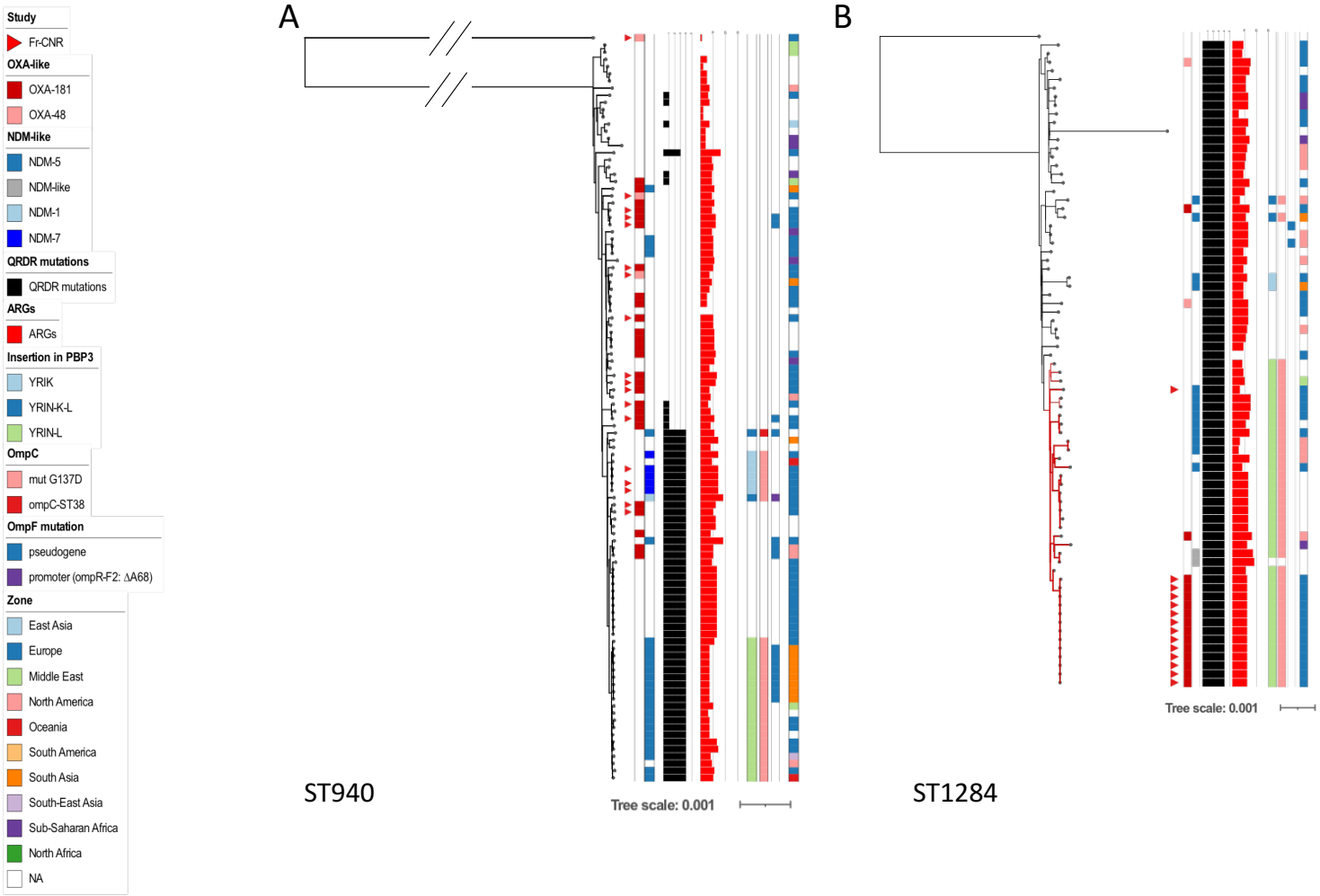

Study
Fr-CNR
OXA-like
OXA-181
OXA-48
NDM-like
NDM-5
NDM-like
NDM-1
NDM-7
QRDR mutations
QRDR mutations
ARGs
ARGs
Insertion in PBP3
YRIK
YRIN-K-L
YRIN-L
OmpC
mut G137D
ompC-ST38
OmpF mutation
pseudogene
promoter (ompR-F2: ΔA68)
Zone
East Asia
Europe
Middle East
North America
Oceania
South America
South Asia
South-East Asia
Sub-Saharan Africa
North Africa
NA
A
ST940
Tree scale: 0.001
B
ST1284
Tree scale: 0.001

Fig. S4

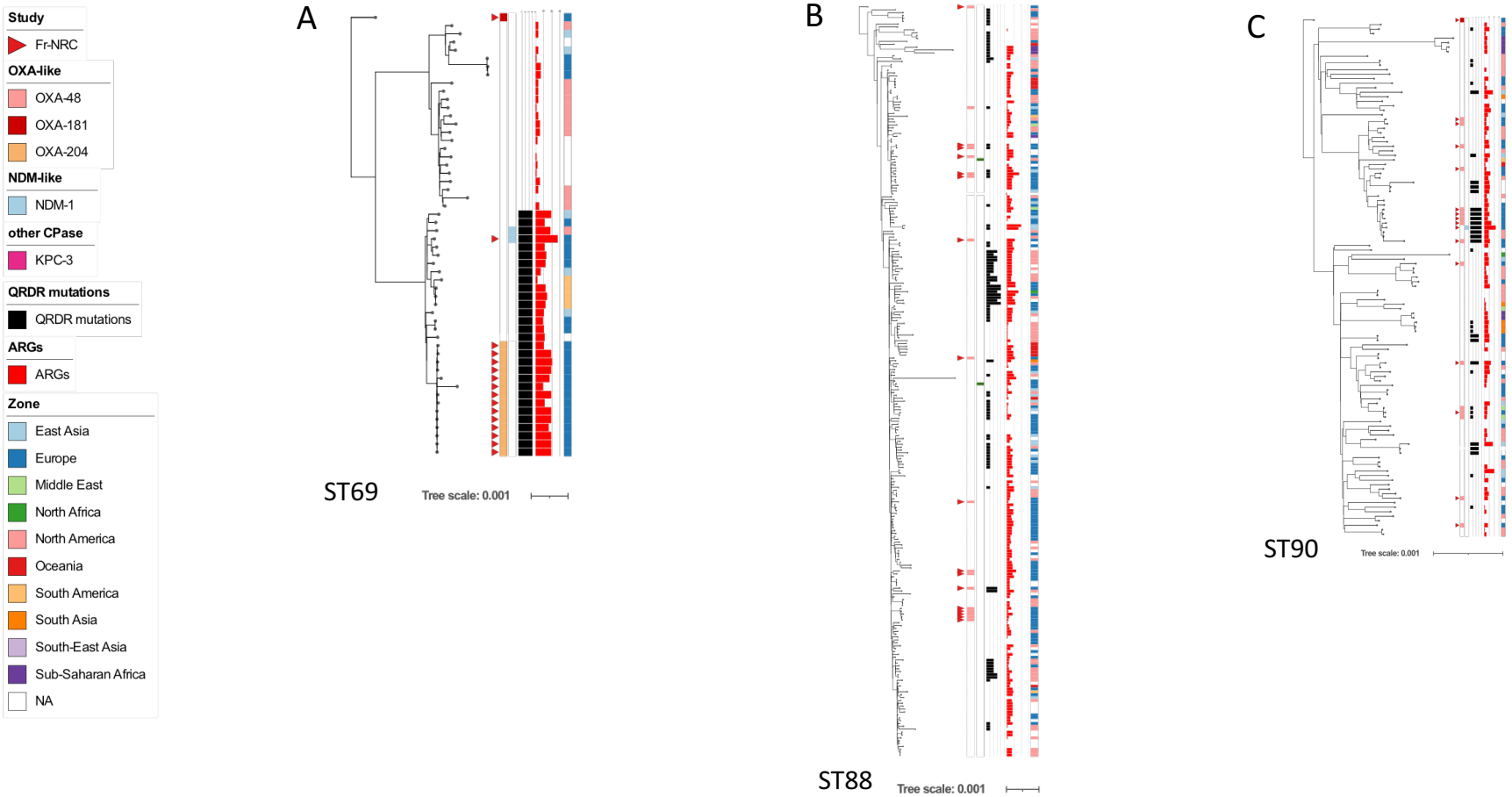

Study
Fr-NRC
OXA-like
OXA-48
OXA-181
OXA-204
NDM-like
NDM-1
other CPase
KPC-3
QRDR mutations
QRDR mutations
ARGs
ARGs
Zone
East Asia
Europe
Middle East
North Africa
North America
Oceania
South America
South Asia
South-East Asia
Sub-Saharan Africa
NA
A
ST69
Tree scale: 0.001
B
ST88
Tree scale: 0.001
C
ST90
Tree scale: 0.001

Fig. S5

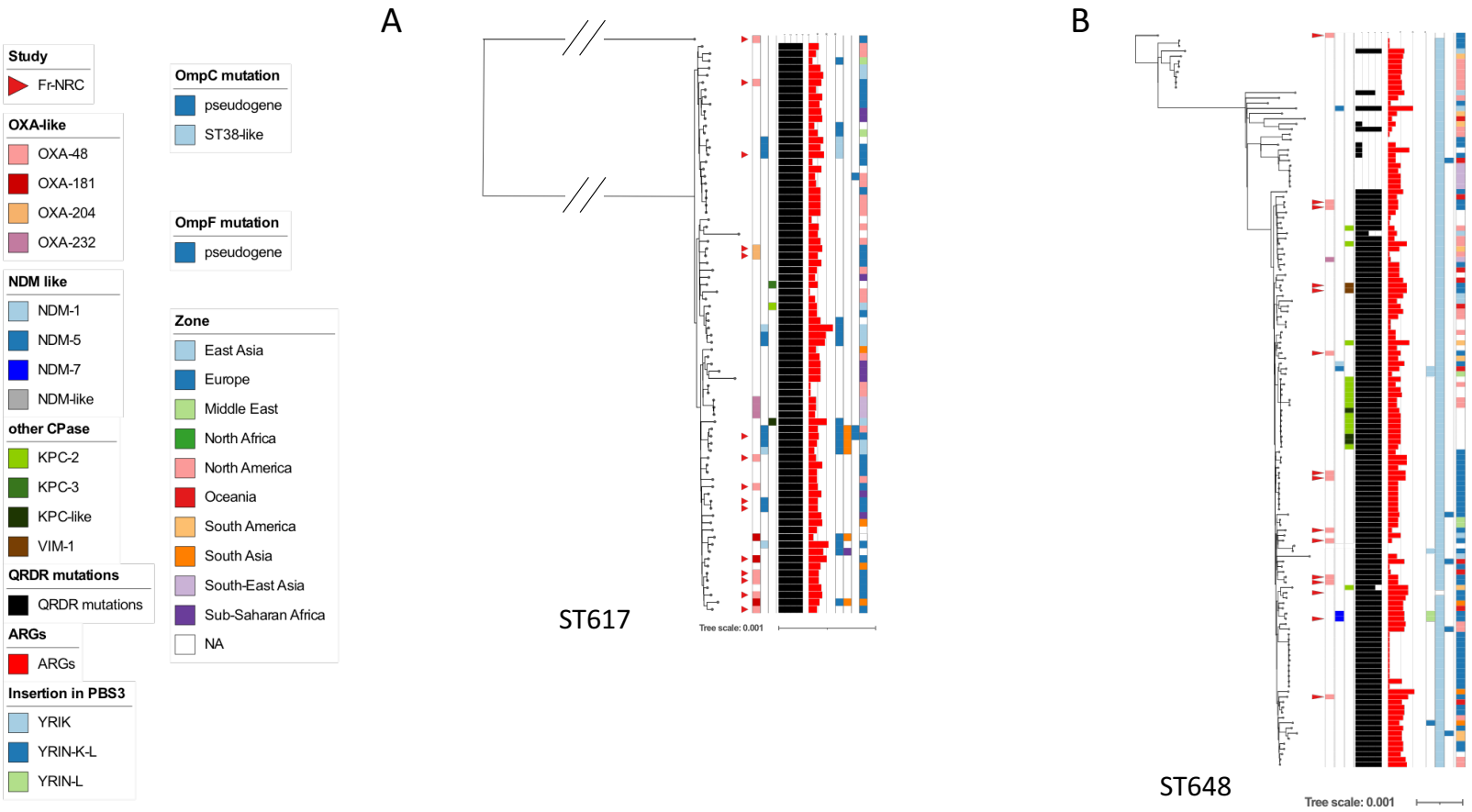

A
B
Study
Fr-NRC
OXA-like
OXA-48
OXA-181
OXA-204
OXA-232
NDM like
NDM-1
NDM-5
NDM-7
NDM-like
other CPase
KPC-2
KPC-3
KPC-like
VIM-1
QRDR mutations
QRDR mutations
ARGs
ARGs
Insertion in PBS3
YRIK
YRIN-K-L
YRIN-L
OmpC mutation
pseudogene
ST38-like
OmpF mutation
pseudogene
Zone
East Asia
Europe
Middle East
North Africa
North America
Oceania
South America
South Asia
South-East Asia
Sub-Saharan Africa
NA
ST617
Tree scale: 0.001
ST648
Tree scale: 0.001

Fig. S6