

Specificities and Commonalities of Carbapenemase-Producing Escherichia coli Isolated in France from 2012 to 2015

Rafael Patiño-Navarrete, Isabelle Rosinski-Chupin, Nicolas Cabanel, Pengdbamba Dieudonné Zongo, Mélanie Héry, Saoussen Oueslati, Delphine Girlich, Laurent Dortet, Rémy A Bonnin, Thierry Naas, et al.

▶ To cite this version:

Rafael Patiño-Navarrete, Isabelle Rosinski-Chupin, Nicolas Cabanel, Pengdbamba Dieudonné Zongo, Mélanie Héry, et al.. Specificities and Commonalities of Carbapenemase-Producing Escherichia coli Isolated in France from 2012 to 2015. 2022. pasteur-03519575v1

HAL Id: pasteur-03519575 https://pasteur.hal.science/pasteur-03519575v1

Preprint submitted on 10 Jan 2022 (v1), last revised 10 Jun 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 Specificities and commonalities of carbapenemase producing *Escherichia coli* isolated in France from 2 2012 to 2015. 3 ^{1,2}Rafael Patiño-Navarrete[#], ^{1,2}Isabelle Rosinski-Chupin[#], ^{1,2}Nicolas Cabanel, ^{1,2,3}Pengdbamba 4 Dieudonné Zongo, ^{1,4}Mélanie Héry, ^{1,4}Saoussen Oueslati, ^{1,4}Delphine Girlich, ^{1,4,5,6}Laurent 5 Dortet, ^{1,4,6}Rémy A Bonnin, ^{1,4,5,6}Thierry Naas^{\$} and ^{1,2,*}Philippe Glaser^{\$}. 6 7 8 ^{\$ and #} Contributed equally 9 ¹Unité EERA, Institut Pasteur, APHP, Université Paris Saclay, ²UMR3525, CNRS, Université de Paris 28 10 rue du Dr Roux, 75015, Paris, France; ³Sorbonne Université, Paris, France, ⁴ Team "Resist" INSERM 11 U1184 "Immunology of Viral, Auto-Immune, Hematological and Bacterial diseases (IMVA-12 HB),"; ⁵Department of Bacteriology-Hygiene, Bicêtre Hospital, APHP, ⁶Associated French National 13 14 Reference Center for Antibiotic Resistance, Le Kremlin-Bicêtre, France. 15 16 17 *Corresponding author: 18 Philippe GLASER, 19 Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15, 20 Tel + 33 1 45 68 89 96 - E-mail: pglaser@pasteur.fr 21 22 Running title: Carbapenemase producing E. coli from the French NRC 23

24 ABSTRACT

25 Carbapenemase-producing Escherichia coli (CP-Ec) represent a major public health threat with a risk 26 of dissemination in the community as it has occurred for lineages producing extended spectrum ß-27 lactamases. To characterize the extend of CP-Ec spread in France, isolates from screening and infection 28 samples received at the French National Reference Centre laboratory (F-NRC) for carbapenemase-29 producing Enterobacterales were investigated. Six hundred and ninety one CP-Ec isolates collected 30 between 2012 and 2015 and 22 before were fully sequenced. Analysis of their genome sequences 31 revealed some disseminating multidrug resistant (MDR) lineages frequently acquiring diverse 32 carbapenemase genes mainly belonging to clonal complex (CC) 23 (ST 410) and CC10 (ST10, ST167) 33 and sporadic isolates including rare ST131 isolates (n=17). However, the most represented ST was ST38 34 (n=92) with four disseminated lineages carrying $bla_{OXA-48-like}$ genes inserted in the chromosome. 35 Globally, the most frequent carbapenemase gene (n=457) was bla_{OXA-48}. It was also less frequently 36 associated with MDR isolates being the only resistance gene in 119 isolates. Thus, outside the ST38 37 clades, its acquisition was frequently sporadic with no sign of dissemination, reflecting the circulation 38 of the IncL plasmid pOXA-48 in France and its high frequency of conjugation. In contrast bla_{OXA-181} or 39 bla_{NDM} genes were often associated with the evolution of MDR E. coli lineages characterized by 40 mutations in *ftsI* and *ompC*.

41

42 IMPORTANCE

43 Carbapenemase-producing Escherichia coli (CP-Ec) might be difficult to detect, as minimal inhibitory 44 concentrations can be very low. However, their absolute number and their proportion among 45 carbapenem-resistant Enterobacterales have been increasing, as reported by WHO and national 46 surveillance programs. This suggests a still largely uncharacterized community spread of these isolates. 47 Here we have characterized the diversity and evolution of CP-Ec isolated in France before 2016. We 48 show that carbapenemase genes are associated with a wide variety of *E. coli* genomic backgrounds 49 and a small number of dominant phylogenetic lineages. In a significant proportion of CP-Ec, the most 50 frequent carbapenemase gene bla_{OXA-48}, was detected in isolates lacking any other resistance gene, 51 reflecting the dissemination of pOXA-48 plasmids, likely in the absence of any antibiotic pressure. In 52 contrast carbapenemase gene transfer may also occur in multi-drug resistant *E. coli*, ultimately giving 53 rise to at-risk lineages encoding carbapenemases with a high potential of dissemination.

54 INTRODUCTION

Escherichia coli is one of the first causes of diverse bacterial infections in the community and in the hospital. In particular, it is the most frequent cause of urinary tract infection (UTI) and 50-60% of women will suffer at least once a UTI during her life (1). Therefore, multidrug resistance (MDR) in *E. coli* is a major public health issue making *E. coli* infections more difficult to treat. In addition, as carbapenemase-producing *Enterobacterales* (CPE) are increasingly detected (2, 3) and *E. coli* is a ubiquitous member of the human gut microbiome, carbapenemase-producing *E. coli* (CP-*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*_{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 analyzed 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).

Whole-genome sequencing (WGS) significantly increases our ability to infer phylogenetic relationships between isolates. By sequencing 50 ST410 CP-*Ec* isolates received by the F-NRC between 2011 and 2015 (10) we showed that 36 (72%) belonged to a newly described ST410 lineage producing OXA-181 (11). We showed that this clade is characterized by mutations in the two porin genes *ompC* and *ompF* leading to a decreased outer membrane permeability to certain ß-lactams and by a four-codon duplication (YRIN) in the *ftsI* gene encoding the penicillin binding protein 3 (PBP3) leading to a

89 decreased susceptibility to ß-lactams targeting this PBP (10). After a thorough analysis of CP-Ec 90 genome sequences from public databases for mutations in these three genes, we proposed that CP-Ec 91 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 *ftsl*. 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 in 2012 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 European countries. These lineages are multi-drug 106 (MDR) or extensively drug resistant (XRD) lineages, they show several mutations in quinolone 107 resistance determining regions (QRDR) and are often mutated in *ftsl* and in porin genes. Finally, the 108 rapid dissemination of four OXA-48/OXA-244 ST38 lineages might have been favored by the 109 chromosomal integration of the 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 QRDRs of gyrA, parC 138 and parE, and for mutations in ftsI, ompC and ompF we previously identified as associated with CP-Ec 139 disseminated lineages (10) (table S1). F-NRC CP-Ec were assigned to 168 different STs including six new 140 allelic combinations. ST38 was the most prevalent ST (n=92, 12.9 %), followed by ST10 (n=67, 9.4 %), 141 ST410 (n=64, 9%), and ST167 (n=34, 4.8%). Ten additional STs were represented by at least 10 isolates 142 (Fig. S1A, table S1) and 154 STs with less than ten isolates including 102 STs with a single isolate.

143 We next performed a core genome phylogeny following read mapping using strain MG1655 144 (NC_000913.3) as reference genome sequence. The phylogenetic tree, estimated from 372,238 core 145 SNPs was consistent with the results of phylogroup determination by using *in silico* ClermonTyping (12) 146 except for a few isolates (Fig. 2). In agreement with the MLST-based analysis, this tree showed a broad 147 diversity of CP-Ec isolates belonging to the eight phylogroups and three dominant clades 148 corresponding to CC10 (phylogroup A; including ST10, ST167 and ST617), CC23 (phylogroup C; 149 including ST410, ST88 and ST90) and ST38 (phylogroup D) with 161, 97, and 98 isolates respectively, 150 accounting for 49.9 % of the F-NRC CP-Ec isolates analyzed (Fig. 2). Phylogroup B2 isolates represented 151 11 % of the total isolates (n=80) and only 17 CP-ST131 isoaltes were identified. Fluctuations in the 152 proportions of the main STs could be observed during the four years of the analysis but no clear 153 tendency could be identified (Fig. S1A).

Infection-related and screening isolates were intermixed throughout the phylogeny (Fig. 2). However,
 an enrichment in infection-related isolates was observed in phylogroup C (Pearson's Chi-squared test,
 p<0.02, ddl1) and phylogroup B2 (Pearson's Chi-squared test p<0.0005, ddl1) (Fig. S1B). In phylogroup

B2, 5 out of 9 ST127 isolates, 5 out of 17 ST131 isolates and 7 out of 8 ST636 were responsible for UTIs(Table S1).

159

160 Diversity of antibiotic resistance genes carried by CP-Ec isolates

161 The number of acquired ARGs was found to vary between 1 and 26 among the 713 CP-Ec isolates. The 162 median was higher in phylogroup C isolates (m=16) and lower in isolates from phylogroup B2 (m=3) 163 compared to other phylogroups (Medians: A: 9; B1: 9; D: 11) (Fig. 3A). An ESBL of the CTX-M family 164 was present in 40.7% (N=290) of the isolates, with a predominance of *bla*_{CTX-M-15} gene (N=205) (Table 165 S1). Mutations in gyrA, parC and/or parE potentially leading to fluoroquinolone (FQ) resistance 166 occurred in 425 CP-Ec isolates (59.6%), with mutations in gyrA, parC and parE identified in 412 isolates, 167 309 and 261 isolates respectively (Table S1). Up to five mutations in QRDRs were identified in 19 168 isolates and 250 isolates had four mutations in QRDR, suggesting they have been submitted to a long-169 term evolution under antibiotic pressure including FQ (Table S1). Globally a higher number of 170 mutations in QRDR was associated with a higher number of resistance genes (Fig. 3B), furthering the 171 link between the number of QRDR mutations and a likely evolution under antibiotic selective pressure 172 for CP-Ec isolates.

Four isolates from 2015, two ST648 collected in a same hospital at one-week interval, one ST216 and one ST744, encoded a *mcr-9* gene conferring resistance to colistin (Table S1). Although not associated with infection, the four isolates were MDR, carrying 13 to 20 acquired ARGs in addition to *mcr-9* and carbapenemase genes (*bla*_{VIM-1} for three of them, *bla*_{NDM-1} and *bla*_{OXA-48} for one). Determination of the MIC for colistin for these four isolates were at the resistance breakpoint (MIC=2 mg/l).

178

179 Diversity of carbapenemase genes

180 Carbapenemase genes identified by WGS (Table S1) were in agreement with molecular data collected

181 by the F-NRC laboratory. The bla_{OXA-48} and $bla_{OXA-181}$ genes were the most frequent carbapenemase

182 genes detected in 464 (65%) and 101 (14.1%) E. coli isolates respectively. Twenty-nine other isolates

- 183 carried minor *bla*_{OXA-48-like} genes (Table S1). *bla*_{NDM} family genes were detected in 14.9% (106/713) of
- 184 the CP-Ec isolates (bla_{NDM-5} in 49 isolates and bla_{NDM-1} in 41). Three additional bla_{NDM} alleles were
- 185 identified: bla_{NDM-7} (n=10) including a variant coding for a NDM7-like carbapenemase with a S24G
- 186 mutation, bla_{NDM-4} (n=5), and bla_{NDM-6} (n=1). Four different bla_{VIM} alleles were detected in 18 isolates:
- 187 *bla*_{VIM-1} (n=8), *bla*_{VIM-4} (n=8), *bla*_{VIM-2} (n=1), and *bla*_{VIM-29} (n=1). Only five *E. coli* isolates expressed *bla*_{KPC-}
- 188 $_{2}$ (n=2) or bla_{KPC-3} (n=3) alleles. In twelve isolates, two different carbapenemase genes were found;
- 189 $bla_{OXA-48} / bla_{NDM-1}$ was the most frequent co-occurrence (n=5) (Table S1).
- 190 The analysis of the number of additional ARGs and of mutations in QRDR regions showed that, 191 compared with other carbapenemase genes the presence of *bla*_{OXA-48} was frequently associated with

less resistant isolates (Fig. 4A, 4B). In particular it was the only ARG in 117 isolates among the 457
(24.5%) encoding this carbapenemase, including 23 infection-related isolates. Among those isolates,
only five showed mutations in QRDR. Conversely, only two *bla*_{NDM-1} gene carrying isolates among 42
and one *bla*_{NDM-5} gene carrying isolates among 49 carried no other ARG.

A temporal analysis of the proportion of isolates displaying bla_{OXA-48} did not reveal significant variations (Pearson's Chi-squared test, p=0.05, ddl= 3) during the 2012-2015 period (Fig. 4C). In contrast, the frequency of $bla_{OXA-181}$ relatively to other alleles was found to significantly increase from 0 % in 2012 to 18.9% in 2015 (Pearson's Chi-squared test, p < 0.0002 ddl=3) (Fig. 4C).

- We observed some association between the carbapenemase gene and the ST. Among the 22 STs with at least seven F-NRC CP-*Ec* isolates, fourteen predominantly displayed bla_{OXA-48} (Fig. 5A); bla_{OXA-48} was also predominant in STs represented by six or less isolates. In three lineages, ST410, ST940, and ST1284, $bla_{OXA-181}$ was the predominant allele. These three STs grouped 74% of the $bla_{OXA-181}$ gene carrying *E. coli* isolates from the F-NRC, and ST410 accounted for 50.5% of them. Finally, $bla_{OXA-204}$ was the most prevalent allele in ST90 and bla_{NDM-5} the predominant allele in ST636.
- 206

207 Characteristics of the STs preferentially associated with F-NRC CP-Ec isolates.

208 Eight among the 14 STs with more than 10 CP-Ec isolates (ST410, ST167, ST940, ST405, ST617, ST90, 209 ST1284, ST101) were characterized by a larger number of ARGs (median \geq 10) and a larger number of 210 mutations in QRDR as compared to ST38, ST10, ST131, ST69, ST88 and ST represented by less than 211 seven isolates (Fig. 5 B, C). ST648 isolates have an intermediate position, being highly mutated in QRDR 212 but less rich in ARGs. Analysis of CP-Ec isolates from the F-NRC for polymorphisms in ftsl revealed that 213 131 (18.4%) had a four-AA-insertion between positions 333 and 334 of PBP3, the insertion being 214 particularly prevalent in ST101, ST167, ST410 and ST1284 (Sup. Table 1 and Fig. 5D). Similarly, ompC 215 alleles we previously characterized as modifying susceptibility to antibiotics were more prevalent in 216 some STs: an ST38-like *ompC* allele resulting from a recombination event or a G138D mutation were 217 present in 33 and 26 isolates belonging to six different ST, whereas the R195L mutation was observed 218 in 50 isolates all from ST410 (Table S1 and Fig. 5E). In addition, 144 isolates belonging to phylogroups 219 D and F possessed the ST38-like allele inherited vertically. On the other hand the ompF gene was 220 pseudogeneized in 47 isolates (Table S1).

To analyze the phylogenetic relationships between F-NRC Cp-*Ec* and other Cp-*Ec* isolates collected worldwide, we built ST-based maximum-likelihood trees for the 14 STs with more than 10 isolates and included to this analysis genome sequences publicly available (Table S2). Together with the analysis of QRDR mutations and mutations in *ftsl* and *ompC*, this showed that 51 (80%) of the ST410 (Fig.S2A), 17 (49%) of the ST167 (Sup Fig.S2B), 3 (33%) of the ST405 (Sup Fig.S2C) and 8 (62%) of the ST101 (Fig. S2D) F-NRC CP-*Ec* belonged to internationally disseminating MDR subclades enriched in CP we

previously identified (10). Subclades characterized by mutations in *ftsl* and *ompC*, mainly expressed
 *bla*_{OXA-181} (ST410) or *bla*_{NDM} (other lineages). Other isolates belonging to these ST were dispersed on
 their respective phylogenies and with no sign of clonal dissemination for most of them.

230 ST38 corresponded to the most represented ST (n=92) among the studied CP-Ec. The phylogenetic 231 reconstruction together with 314 additional non redundant CP-Ec genome sequences retrieved from 232 Enterobase (http://enterobase.warwick.ac.uk/) and 150 non-CP isolates from NCBI provided 233 evidence that 80.4% (n=67) of the French isolates clustered into four clades (Fig. 6). Three of them, G1 234 (n=35), G2 (n=25) and G4 (n=4) only contained isolates encoding *bla*_{OXA-48}, while G3 (n=7), included 235 isolates expressing bla_{OXA-48} or its single nucleotide derivative bla_{OXA-244}. All but one isolates in G1 236 expressed bla_{CTX-M-24}, while isolates of G3 expressed bla_{CTX-M-14b}. The phylogenetic analysis provided 237 evidence of worldwide dissemination of G1, G3 and G4 clades and multiple introduction in France. 238 Strickingly, among the 28 isolates of the G2 clade, 24 were collected in France and four in the 239 Netherlands, suggesting at this time a more regional dissemination. In none of the isolates of the four 240 clades, an IncL plasmid, that generally encode bla_{OXA-48} (13), could be identified by PlasmidFinder (14). 241 It suggests a chromosomal integration as previously shown among ST38 isolates collected in the UK 242 (15). Analysis of representative isolates of the G1 and G3 cluster (G1: GCA 005886035.1; G3: 243 GCA 004759025.1), whose genome sequences were completely assembled, confirmed that bla_{OXA-48} 244 and *bla*_{OXA-244}, in these two isolates respectively, were chromosomally integrated. To assess the genetic 245 support of *bla*_{OXA-48} in the two other clades, we have determined the complete genome sequence of 246 two isolates belonging to G2 (CNR65D6) and G4 (G4:CNR85I8) by combining long-read Pacbio and 247 Illumina sequencing and found that both possess a chromosomally inserted bla_{DXA-48} gene. pOXA-48 248 plasmid sequences of various lengths were cointegrated with bla_{OXA-48}, the whole sequence being 249 surrounded by IS1 insertion sequences with different insertion sites for the four clades (Fig. 7). BLASTN 250 analyses on the assembled Illumina sequences of other isolates of each clade showed that isolates of 251 the same clade share the same insertion sites (see material and methods). A fifth clade, composed of 252 seven closely related French isolates corresponded to a possible outbreak in the East of Paris between 253 December 2014 and December 2015. These isolates are predicted to be highly resistant as they are 254 carrying in addition to bla_{VIM-4}, bla_{CTX-M-15}, as well as 15 to 18 additional ARGs and five mutations in 255 QRDRs.

ST10 *E. coli* isolates are commensals of a variety of mammals and bird species (16) and studies in different contexts have shown that this ST is also associated with ARGs carriage (17, 18). The 67 F-NRC ST10 CP-*Ec* isolates were distributed throughout the ST10 phylogeny. Nonetheless 31 isolates belonged to a clade enriched with carbapenemase producers (Fig. S3A, in green) that mainly expressed *bla*_{OXA-48}, with the exception of three isolates expressing *bla*_{OXA-181}, three expressing *bla*_{NDM-5} and one and *bla*_{VIM-4}. Two of the *bla*_{OXA-181} and the three *bla*_{NDM-5} expressing isolates were closely related, shared

a CTX-M-27 ESBL, four mutations in QRDR and a *ftsl* allele with a four AA YRIN insertion (Fig. S3A, in
 red).

264 ST131 isolates are major contributors for extra-intestinal infections and the main responsible for the 265 dissemination of CTX-M-15 ESBL (19, 20). ST131 evolution has been thoroughly analyzed and four main 266 lineages A, B, C1 and C2 have been described (21). We identified only 17 ST131 CP-Ec isolates received 267 by the F-NRC (Fig. 2), ten belonged to lineage A, while the others were scattered in the three other 268 lineages (Fig. S3B). The main carbapenemase associated with these isolates was OXA-48 (N=10). We 269 also did not observe any association with the carriage of *bla*_{CTX-M} genes, as only four ST131 CP-*Ec* 270 isolates carryied a $bla_{CTX-M-15}$ gene. These results are in accordance with the UK survey (8) and our 271 previous observations from sequences retrieved from public databases (10).

272 ST940 (phylogroup B1) and ST1284 (phylogroup A) represented 2.4% (n=17) and 1.9% (n=14) of the 273 sequenced isolates while they were poorly represented in sequence databases. For instance, in May, 274 2021, only 83 and 103 sequences genome sequences could be retrieved from Enterobase for these ST 275 respectively, but with 43% (n=36) and 22% (n=22) carrying a carbapenemase gene respectively. This 276 suggested that ST940 and ST1284 were associated with the dissemination of carbapenemase genes. 277 Phylogenetic trees were drawn by adding the non-redundant sequences from enterobase to those of 278 the F-NRC and NCBI. For ST940 (Fig. S4A) this revealed a well-differentiated sub-clade of 18 isolates 279 from Asia, Australia and Europe carrrying bla_{NDM-5} and sharing mutations in *ftsl* and *ompC*. Isolates 280 from the F-NRC were distributed on the tree, showing that the over-representation of ST940 in France 281 did not result from local outbreaks. Three F-NRC isolates belonged to a second smaller sub-clade of 282 bla_{NDM} expressing isolates, also mutated in *fts1* and ompC. In contrast, in ST1284, thirteen encoding 283 $bla_{OXA-181}$, were found to be closely related (Fig. S4B). Metadata analysis showed that 12 out of them 284 were recovered from the same health facility in the Paris suburb during a one-month period in 2015, 285 demonstrating a local outbreak origin. A fourteenth unrelated isolate expressed bla_{NDM-5}. The F-NRC 286 isolates belonged to a clade characterized by four mutations in QRDR determinants, a YRIN duplication 287 in PBP3, a G137D mutation in OmpC and containing additional OXA-181- or NDM-producing isolates. 288 Among the fifteen ST90 CP-Ec isolates, nine encoding the bla_{OXA-204} allele were isolated in hospitals East 289 of Paris between August 2012 to April 2013 and suspected to be associated with the use of a 290 contaminated endoscope (22). Isolates closely related to this outbreak reappeared in 2014 (three

characterized by four mutations in QRDR (Fig. S5A). Finally, in the four last STs with more than 10 CP-

times), and in 2015 (twice), also in East of Paris located hospitals. They belonged to a MDR clade

293 Ec isolates (ST69, ST88, ST617, and ST648), French CP-Ec isolates were distributed throughout their

respective phylogenetic trees (FigS5B and C,S6A and B) with no sign of clonal dissemination, except for

a small cluster of five ST69 from two geographical regions and expressing *bla*OXA-48.

291

296 In addition to potential outbreaks detected in ST38 (blavIII-4), in ST90 (blavA-204) and ST1284 (blavA-297 181), we also found evidence for another potential outbreak among isolates belonging to the ST359. 298 The five isolates encoded *bla*_{OXA-48} and *bla*_{CTX-M-32} and had three mutations in QRDR. They were isolated 299 during January 2014, four of them in the south-eastern part of France and one in the Parisian suburbs. 300

301 DISCUSSION

302 The prevalence of CP-Ec is increasing worldwide. However whether this reveals conjugation event in 303 E. coli isolates by circulating CP-encoding plasmids or the emergence and dissemination of at-risk 304 clones is still largely unknown. Here, we have analyzed, by using WGS, 691 (87.8%) of the CP-Ec isolates 305 received by the F-NRC during the period 2012-2015 and 22 isolates from the Bicêtre hospital strain 306 collection (2001-2011) to characterize both the diversity of carbapenemase genes and of CP-Ec isolates 307 in France. Altogether 713 CP-Ec were sequenced, representing to our knowledge the most extended 308 collection of CP-*Ec* sequenced, published to date.

309 The number of CP-Ec isolates sent to the F-NRC was found to strongly increase during the four-year 310 period of analysis, which might be a consequence of an increased circulation of CP-Ec in France but 311 also of an increased screening of potential CPE carriers at their admission at hospital. Indeed, while 312 the number of infection-related CP-Ec isolates was regularly increasing, their proportion compared to 313 the total number of received and sequenced isolates decreased by two-fold, with a clear change 314 observed between 2013 and 2014 (Fig. 1). This is likely a consequence of the implementation of the 315 recommendations on MDR screening of the French Public Health Advisory Board 2013 (23). However, 316 as CP-Ec isolates were sent on a voluntary basis by clinical laboratories, we cannot exclude some 317 sampling bias.

318 The most predominant carbapenemase allele was bla_{OXA-48} (65%) followed by bla_{OXA-181} (14.1%) which 319 detection constantly increased from no cases in 2012 to 73 in 2015. Next was bla_{NDM-5} that was found 320 to progressively substitute to the *bla*_{NDM-1} gene in frequency of detection (Fig. 4). The predominance 321 of $bla_{OXA-48-like}$ genes and particularly of bla_{OXA-48} , among CP-*Ec* was also noted in other studies (9, 24,25) 322 although in lower proportions. The small number of bla_{KPC} genes detected in these studies is in contrast 323 with the situation reported in K. pneumoniae for many European countries (26). However, bla_{0XA-48} 324 was found the most prevalent allele in K. pneumoniae in France (2) and in the Netherlands (9). This 325 difference might result from a lower capacity for blakpc encoding plasmids to conjugate to E. coli as 326 compared to pOXA-48 or to a higher fitness cost. It might also be linked to sampling differences as 327 different thresholds of carbapenem susceptibility might have been used to collect CRE isolates among 328 studies. Indeed OXA-48-like carbapenemases are generally associated with lower levels of carbapenem 329 resistance than KPC. This analysis was performed on isolates collected between 2012 and 2015 and 330 revealed early trends in the evolution of CP-Ec isolated in France. It will be essential in the

characterization of the genomic evolution of CP-*Ec* in more recent years, which remains to beperformed.

333 The number of ARGs identified in the CP-Ec isolates from the F-NRC varied from one (the 334 carbapenemase gene only) to 26, showing that carbapenemase genes were acquired not only in MDR 335 isolates but also in an E. coli population that can be considered as "naïve" relative to an evolution 336 under antibiotic selective pressure and resident of the intestinal microbiota. This was particularly true 337 for isolates producing bla_{OXA-48} that were generally associated with a lower number of resistance genes 338 than isolates producing other carbapenemases, irrespectively of their clinical status (Fig. 4A). In 339 particular 118 out of the 121 (97.5%) isolates with no other ARG than the carbapenemase gene carried 340 bla_{OXA-48} . Therefore, the predominance of bla_{OXA-48} among the carbapenemase genes identified in the 341 F-NRC collection might mainly rely on the higher conjugative transfer rate of pOXA-48 related plasmids 342 compared to those encoding other carbapenemases (27). pOXA-48 has indeed been shown to rapidly 343 conjugate among *Enterobacterales* in hospitalized patients (28). This could also explain the high 344 frequency of bla_{OXA-48} carrying isolates belonging to ST10, ST69, ST88, ST127, ST12 and ST58 (Fig. 5), 345 that carried a small number of resistance genes and QRDR mutations and among which four were 346 previously identified among the most frequent *E. coli* ST characterized from fecal samples (29).

347 On the other hand, MDR CP-Ec isolates may result from the transfer of carbapenem resistance genes 348 into isolates already selected through multiple exposures to antibiotic treatments that have already 349 acquired multi-resistance plasmids or chromosomal mutations increasing their intrinsic drug 350 resistance, such as mutations in QRDR, in *fts1* and/or in porin genes. Alternatively, the circulation of 351 carbapenem-resistant lineages, such as the OXA-181-producing ST410 E. coli lineage, showing a stable 352 association between the lineage and the resistance gene, may also occur and significantly contribute 353 to the number of CP-Ec isolates collected. Discriminating between both alternatives would require a 354 more extended comparison of plasmids carried in these lineages after long-read sequencing similarly 355 to what been done in K. pneumoniae (30). Strikingly, the analysis of ftsl alleles coding for a PBP3 with 356 the four amino-acid duplication revealed a contrasted situation among CP-Ec isolates collected in 357 France, with the duplication in *ftsI* found in 72.3%, 77.8% and 67% of the isolates carrying *bla*_{NDM-5} 358 (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 359 bla_{0XA-48} . We previously proposed that the fixation of mutations reducing the intrinsic susceptibility to 360 carbapenems might have favored the efficient conjugative transfer of plasmids carrying the 361 carbapenemase genes from other CPE species in the gut (10). Transconjugant would be selected in a 362 context of low biliary excretion of carbapenems or other ß-lactams after parenteral administration. 363 Increasing the proportion of the donor and recipient bacteria would therefore be less necessary for 364 plasmids with high conjugative rates, such as pOXA-48 (27). Alternatively, these mutations might also 365 have favored the plasmid maintenance by increasing the resistance level and selection of the isolates

during β-lactam or carbapenem treament. Of particular concern in France are the OXA-181 producing
ST410 and the NDM (NDM-1, -5 and -7) ST167 lineages that significantly contribute to the circulation
of carbapenem resistance. However our study also reveals smaller lineages, that although less
frequently encountered in international studies, are nevertheless circulating. In contrast we did not
obtain evidence for clonal dissemination of CP-*Ec* ST131 lineages derived from the B2 clade responsible
for *bla*_{CTX-M-15} gene dissemination (19, 20).

372 Despite their high frequency among the F-NRC collection, the ST38 isolates do not enter into one of 373 the two previous categories. Indeed, while they mainly express bla_{OXA-48} or its single-point-mutant 374 derivative bla_{OXA-244} (31), our phylogenetic analysis revealed that a majority of them belong to four 375 different lineages, including one mostly associated with a rapid dissemination in France. In contrast 376 with OXA-181 and NDM-producing lineages previously described, these lineages are associated with a 377 moderate number of ARGs and QRDR mutations. None of the ST38 isolates is mutated in *ftsI*, however 378 we previously demonstrated that all are characterized by an *ompC* allele that encodes a porin with a 379 reduced permeability to certain ß-lactams and has disseminated into unrelated lineages by 380 homologous recombination (10). A common feature of the four lineages is the integration of part of 381 the pOXA-48 plasmid carrying bla_{OXA-48}. This could have reduced a fitness cost of this plasmid and 382 facilitated the clonal expansion of these lineages. However, ST38 isolates were identified as 383 unfrequent colonizer of the GI tract (29, 32). Therefore additional features of these CP-Ec ST38 lineages 384 might have contributed to their dissemination.

385 In conclusion, by analysing genome sequences of CP-Ec isolates collected by the F-NRC we showed that 386 MDR lineages, enriched in carbapenemase-producing isolates are circulating in France, some of them, 387 such as ST1284 isolates being associated with outbreaks. It also suggests that the evolutionary 388 trajectory may depend on the carbapenemase gene, $bla_{OXA-181}$ or bla_{NDM} genes being more frequently 389 associated with the evolution of MDR E. coli lineages characterized by mutations in ftsl and ompC. 390 Surveillance of these mutations may be an important parameter in controling the circulation of MDR 391 lineages. On the other hand, carbapenemase genes are also frequently acquired through plasmid 392 dissemination from other bacterial species. Depending on the resistance background of the receiver 393 E. coli, this may lead to XDR or to isolates sensitive to a broad range of antibiotics. Finally, we also 394 observed a strong and rapid dissemination of ST38 isolates that might have been favored by a reduced 395 susceptibility to carbapenems linked to the ST38 ompC allele and by the chromosomal integration of 396 the carbapenemase gene. These results strengthen our model of different evolutionary trajectories 397 associated with the gain of carbapenemase genes (10). They also show that systematic genome 398 sequencing of CP-Ec and at a larger of CPE isolates, irrespective of their clinical or resistance status is 399 able to provide useful information not only on the circulation of MDR lineages, but also on the 400 propagation of resistance genes through horizontal gene transfer.

401

402 MATERIAL AND METHODS

403 Isolate collection and sequencing.

404 CP-*Ec* isolates analyzed in this study were collected by the F-NRC, mainly between the years 2012 to 405 2015. Twenty-two additional isolates we have received before the creation of the F-NRC in 2012, were 406 included in the analysis. Information on these isolates as the year of isolation, the region and 407 department in France and summary of their genomic features are reported in Table S1.

408

409 Whole genome sequencing and analyses.

410 DNA were extracted by using the Qiagen Blood and Tissue DNA easy kit. Sequencing libraries were 411 constructed by using the Nextera XT kit (Illumina) following the manufacturer instruction, and 412 sequenced with the Illumina HiSeq2500 or NextSeq500. FastQ files were trimmed for adaptors and 413 low-quality bases (setting the minimum base quality threshold to 25) with the Cutadapt fork Atropos 414 (33). De novo assemblies were generated from the trimmed reads with SPAdes v3.12.0 (34), using k-415 mer sizes of 51, 71, 81, and 91, the coverage cut-off option was set to "auto" and the "careful" option 416 was activated. QUAST v2.2 (35) was used to assess the assembly quality and contigs shorter than 500 417 bp were filtered out for phylogenetic analyses. The ST38 isolates CNR65D6 and CNR85I8 were 418 sequenced to completion by using the long-read Pacbio technology. PacBio reads were assembled with 419 the RS_HGAP_Assembly.3 protocol from the SMRT analysis toolkit v2.3 (36) and with Canu (37), 420 polished with Quiver (36) and manually corrected by mapping Illumina reads using Breseq (38). 421 Assembled genomes were annotated with Prokka v1.9 (39). To analyze the F-NRC CPEc isolates 422 belonging to the main ST recovered during the analysis in a more global context, their genome 423 sequences were combined to the assembled genome sequences from the same ST retrieved from the 424 NCBI database (July, 2019) and in some STs from Enterobase (http://enterobase.warwick.ac.uk/, April 425 2021). Retrieved genomes were annotated in the same way as the genome sequences of the F-NRC 426 isolates (Table S2).

427 To generate a core-genome phylogeny tree of all isolates, the best reference genome was first selected 428 among a set of 18 genomes analyzed by Touchon et al. (40) using the software refRank (41) and three 429 subsets of 100 randomly selected sequences from our study as an input. This led to the selection of 430 the genome sequence of strain MG1655-K12 (NC 000913.3) as the reference for read mapping and 431 SNP identification. Sequence reads for the 17 reference genomes (40) as well as for the genome 432 sequence of Escherichia fergunsonii strain ATCC 35469 (NC_011740.1), used as an outgroup were 433 simulated with ART (42). Trimmed sequencing reads were mapped against the MG1655-K12 genome 434 with BWA-MEM algorithm of the BWA v0.7.4 package (43). For SNP calling the Genome Analysis Toolkit 435 (GATK) v3.6.0 (44) was used with the following criteria, a minimum depth coverage (DP) of 10, a quality

by depth (QD) bigger than 2, a fisher strain bias (FS) below 60, a root mean square of the mapping quality (MQ) above 40, and the mapping quality rank sum test (MQRankSum) and the read position rank sum test (ReadPosRankSum) greater than -12.5 and -8 respectively. A Maximum-Likelihood tree was estimated with RAxML v8.28 (45) using core-genome SNPs after removing positions in the accessory genome identified with the filter_BSR_variome.py script from the LS-BSR pipeline (46).

For ST with more than 10 CP-*Ec* isolates from the F-NRC, a core genome alignment was generated with Parsnp (47), by using a finished genome sequence as reference. A closely related isolate outside the ST lineage was selected from the global phylogenetic tree including all F-NRC isolates and used as outgroup to root the ST phylogenetic trees. Maximum-Likelihood (ML) trees were generated with RAXML v8.28 (45) after removing regions of recombination with Gubbins (48). All graphic representations were performed by using ITOL (49).

447

448 MLST type, phylogroups, resistance gene and plasmid replicon identification.

449 Sequence type was assigned to each assembly through a python script that relies on BLAST (10). 450 Antibiotic resistance genes (ARGs) and mutations in qyrA, parC and parE quinolone resistance 451 determining regions (QRDR) were identified with Resfinder 4.0 and PointFinder (50) run in local 452 respectively. The scripts and database (Retrieved: January 4, 2021) were downloaded from the 453 repositories of the Centre for Genomic Epidemiology (https://bitbucket.org/genomicepidemiology/). 454 The identified ARGs were manually reviewed to eliminate potential redundant ARG predicted at the 455 same genomic position. *mdf*(A) that is present in *E. coli* core genome was not taken into account. 456 Phylogroups were assigned by using EzClermont (51) run in local. Plasmid replicons were identified 457 with plasmidfinder run on each assembly (14). ftsl, ompC and ompF CDS sequences and ompF 458 promoter sequences were identified by using BlastN. Translated or nucleotide sequences were 459 clustered by cd-hit (cd-hit-v4.8.1) with an amino acid (FtsI, OmpC, OmpF) or nucleotide sequence (*ompF* promoter) identity threshold of 1. Sequences of each cluster were aligned to detect mutations 460 461 in regions of interest: four amino-acid insertions between P333 and Y334 and E349K and I532L 462 mutations (FtsI), mutation modifying charge in L3 constriction loop, R195L mutation, nonsense or 463 frameshift mutations or OmpC sequence clustering with ST-38 OmpC sequences (OmpC), nonsense 464 mutations (OmpF) or mutations affecting one of the OmpR-boxes: mutation -46T/C in OmpR-F3 box 465 and mutation $-75\Delta T$ in OmpR-F2 box (*ompF* promoter).

466

467 Determination of *bla*_{OXA-48} insertion site in ST38 clades 1, 2, 3 and 4

The genome sequences of GCA_005886035.1 (CP040390.1), CNR65D6, GCA_004759025.1
(CP038505.1), and CNR85I8 were chosen as representative genomes of clusters 1, 3, 2 and 4,
respectively and aligned, by using Blastn, against the sequence of GCA_001900295.1 (CP010116.1), a

471 ST38 isolate that does not express a carbapenemase and whose genome sequence was complete. The 472 regions surrounding the bla_{OXA-48} insertion sites in each clade were more precisely aligned against the 473 corresponding sequences of other representative genomes by using Easyfig (Easyfig mac 2.1) (52) 474 (Fig. S3). To perform a blast analysis of insertion sites by using contigs generated from Illumina 475 sequences, smaller sequences were chosen as queries as following: nucleotides (nt) 1240133..1240934, nt 964173..965072, nt 419599..421606, nt 4206783..4207690 from 476 477 GCA_001900295.1 to screen for native sequences corresponding to the four *bla*_{OXA-48} insertion site in 478 G1, G2, G3 and G4 clades and sequences nt 4460332..4460759 from CP040390.1, nt 4763987..4764506 479 from CNR65D6, nt 100784..102087 from CP038505.1 and nt 1206426..1206985 from CNR85I8 to 480 screen for the integration site plus IS1 insertion as occurring in G1, G2, G3 and G4 clades respectively. 481

482 Statistical analysis.

483 Pearson's chi-squared tests were performed by using standard libraries contained within the R
484 statistics package (http://www.R-project.org)

485

486 Availability of data.

487 All sequence data have been deposited at DDBJ/EMBL/GenBank (BioProject PRJEB46636) and 488 bioSample identifiers for the Illumina sequence data are listed in Table S1. Complete genome 489 sequences of CNR65D6 and CNR85I8 and the long-read sequencing data have been deposited at 490 DDBJ/EMBL/GenBank with the accession number ERZ3517884 (PacBio reads, ERR6414227) and 491 ERZ3518335 (ERS6682837 PacBio reads) respectively.

492 493

494 ACKNOWLEDGMENTS

495 This work was supported by grants from the French National Research Agency (ANR-10-LABX-62-IBEID,

496 ANR-10-LABX-33, INCEPTION project (PIA/ANR-16-CONV-0005)) and from the European Union's

497 Horizon 2020 research and Innovation Program under grant agrement No 773830 (Project ARDIG, One

- 498 Health EJP). Pengdbamba Dieudonné Zongo is a scholar of Ed525 CDV, Sorbonne Université, Paris.
- 499

500 **REFERENCES**

- Medina M, Castillo-Pino E. 2019. An introduction to the epidemiology and burden of urinary tract infections. Ther Adv Urol 11:1756287219832172.
- 5032.Dortet L, Cuzon G, Ponties V, Nordmann P. 2017. Trends in carbapenemase-producing504Enterobacteriaceae, France, 2012 to 2014. Euro Surveill 22.
- 5053.French National Reference Centre for antibiotic resistance, activity report 2019 2020.506https://wwwcnr-resistance-antibiotiquesfr/bilans-dactivitehtml.

- 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.
- 5105.Bush K, Jacoby GA. 2010. Updated functional classification of beta-lactamases. Antimicrob511Agents Chemother 54:969-76.
- 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.
- 516 7. Gauthier L, Dortet L, Cotellon G, Creton E, Cuzon G, Ponties V, Bonnin RA, Naas T. 2018.
 517 Diversity of carbapenemase-producing *Escherichia coli* isolates, France 2012-2013. Antimicrob
 518 Agents Chemother doi:10.1128/aac.00266-18.
- 5198.Ellaby N, Doumith M, Hopkins KL, Woodford N, Ellington MJ. 2019. Emergence of diversity in520carbapenemase-producing *Escherichia coli* ST131, England, January 2014 to June 2016. Euro521Surveill 24:1800627
- 5229.van der Zwaluw K, Witteveen S, Wielders L, van Santen M, Landman F, de Haan A, Schouls LM,523Bosch T. 2020. Molecular characteristics of carbapenemase-producing *Enterobacterales* in the524Netherlands; results of the 2014-2018 national laboratory surveillance. Clin Microbiol Infect52526:1412.e7-1412.e12.
- 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.
- 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
- 53412.Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. 2018. ClermonTyping: an535easy-to-use and accurate *in silico* method for *Escherichia* genus strain phylotyping. Microb536Genom 4:0.000192
- 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
- 541 14. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup
 542 F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid
 543 multilocus sequence typing. Antimicrob Agents Chemother 58:3895-903.
- Turton JF, Doumith M, Hopkins KL, Perry C, Meunier D, Woodford N. 2016. Clonal expansion
 of *Escherichia coli* ST38 carrying a chromosomally integrated OXA-48 carbapenemase gene. J
 Med Microbiol 65:538-46.
- Leverstein-van Hall MA, Dierikx CM, Cohen Stuart J, Voets GM, van den Munckhof MP, van
 Essen-Zandbergen A, Platteel T, Fluit AC, van de Sande-Bruinsma N, Scharinga J, Bonten MJ,
 Mevius DJ. 2011. Dutch patients, retail chicken meat and poultry share the same ESBL genes,
 plasmids and strains. Clin Microbiol Infect 17:873-80.
- Oteo J, Diestra K, Juan C, Bautista V, Novais A, Pérez-Vázquez M, Moyá B, Miró E, Coque TM,
 Oliver A, Cantón R, Navarro F, Campos J. 2009. Extended-spectrum beta-lactamase-producing
 Escherichia coli in Spain belong to a large variety of multilocus sequence typing types, including
 ST10 complex/A, ST23 complex/A and ST131/B2. Int J Antimicrob Agents 34:173-6.
- 55518.Manges AR, Mende K, Murray CK, Johnston BD, Sokurenko EV, Tchesnokova V, Johnson JR.5562017. Clonal distribution and associated characteristics of *Escherichia coli* clinical and557surveillance isolates from a military medical center. Diagn Microbiol Infect Dis 87:382-385.

- Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M,
 Chattopadhyay S, Stegger M, Andersen PS, Pearson T, Riddell K, Rogers P, Scholes D, Kahl B,
 Keim P, Sokurenko EV. 2013. The Epidemic of Extended-Spectrum-beta-Lactamase-Producing
 Escherichia coli ST131 Is Driven by a Single Highly Pathogenic Subclone, H30-Rx. MBio
 4:e00377-13
- 56320.Nicolas-Chanoine MH, Bertrand X, Madec JY. 2014. Escherichia coli ST131, an intriguing clonal564group. Clin Microbiol Rev 27:543-74.
- 56521.Decano AG, Downing T. 2019. An *Escherichia coli* ST131 pangenome atlas reveals population566structure and evolution across 4,071 isolates. Sci Rep 9:17394.
- Potron A, Bernabeu S, Cuzon G, Pontiès V, Blanchard H, Seringe E, Naas T, Nordmann P, Dortet
 L. 2017. Analysis of OXA-204 carbapenemase-producing Enterobacteriaceae reveals possible
 endoscopy-associated transmission, France, 2012 to 2014. Euro Surveill 22:17-00048
- 570 23. Lepelletier D, Berthelot P, Lucet JC, Fournier S, Jarlier V, Grandbastien B. 2015. French
 571 recommendations for the prevention of 'emerging extensively drug-resistant bacteria' (eXDR)
 572 cross-transmission. J Hosp Infect 90:186-95.
- 573 24. Ortega A, Sáez D, Bautista V, Fernández-Romero S, Lara N, Aracil B, Pérez-Vázquez M, Campos
 574 J, Oteo J. 2016. Carbapenemase-producing *Escherichia coli* is becoming more prevalent in
 575 Spain mainly because of the polyclonal dissemination of OXA-48. J Antimicrob Chemother
 576 71:2131-8.
- 577 25. Grundmann H, Glasner C, Albiger B, Aanensen DM, Tomlinson CT, Andrasević AT, Cantón R,
 578 Carmeli Y, Friedrich AW, Giske CG, Glupczynski Y, Gniadkowski M, Livermore DM, Nordmann
 579 P, Poirel L, Rossolini GM, Seifert H, Vatopoulos A, Walsh T, Woodford N, Monnet DL. 2017.
 580 Occurrence of carbapenemase-producing *Klebsiella pneumoniae* and *Escherichia coli* in the
 581 European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective,
 582 multinational study. Lancet Infect Dis 17:153-163.
- 583 26. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, Abudahab K, Goater R, Giani T,
 584 Errico G, Aspbury M, Sjunnebo S, Feil EJ, Rossolini GM, Aanensen DM, Grundmann H. 2019.
 585 Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial
 586 spread. Nat Microbiol 4:1919-1929.
- Potron A, Poirel L, Nordmann P. 2014. Derepressed transfer properties leading to the efficient
 spread of the plasmid encoding carbapenemase OXA-48. Antimicrob Agents Chemother
 58:467-71.
- León-Sampedro R, DelaFuente J, Díaz-Agero C, Crellen T, Musicha P, Rodríguez-Beltrán J, de la
 Vega C, Hernández-García M, López-Fresneña N, Ruiz-Garbajosa P, Cantón R, Cooper BS, San
 Millán Á. 2021. Pervasive transmission of a carbapenem resistance plasmid in the gut
 microbiota of hospitalized patients. Nat Microbiol 6:606-616.
- 59429.Matsui Y, Hu Y, Rubin J, de Assis RS, Suh J, Riley LW. 2020. Multilocus sequence typing of595*Escherichia coli* isolates from urinary tract infection patients and from fecal samples of healthy596subjects in a college community. Microbiologyopen 9:1225-1233.
- 597 30. David S, Cohen V, Reuter S, Sheppard AE, Giani T, Parkhill J, Rossolini GM, Feil EJ, Grundmann
 598 H, Aanensen DM. 2020. Integrated chromosomal and plasmid sequence analyses reveal
 599 diverse modes of carbapenemase gene spread among *Klebsiella pneumoniae*. Proc Natl Acad
 600 Sci U S A 117:25043-25054.
- 60131.Emeraud C, Girlich D, Bonnin RA, Jousset AB, Naas T, Dortet L. 2021. Emergence and Polyclonal602Dissemination of OXA-244-Producing *Escherichia coli*, France. Emerg Infect Dis 27:1206-1210.
- Massot M, Daubié AS, Clermont O, Jauréguy F, Couffignal C, Dahbi G, Mora A, Blanco J, Branger
 C, Mentré F, Eddi A, Picard B, Denamur E, The Coliville G. 2016. Phylogenetic, virulence and
 antibiotic resistance characteristics of commensal strain populations of *Escherichia coli* from
 community subjects in the Paris area in 2010 and evolution over 30 years. Microbiology
 (Reading) 162:642-650.
- 60833.Didion JP, Martin M, Collins FS. 2017. Atropos: specific, sensitive, and speedy trimming of609sequencing reads. PeerJ 5:e3720.

- 610 34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,
 611 Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.
 612 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell
 613 sequencing. J Comput Biol 19:455-77.
- 61435.Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome615assemblies. Bioinformatics 29:1072-5.
- 616 36. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A,
 617 Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome
 618 assemblies from long-read SMRT sequencing data. Nat Methods 10:563-9.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722-736.
- 622 38. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes
 623 from next-generation sequencing data using breseq. Methods Mol Biol 1151:165-88.
- 624 39. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.
- 40. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S,
 Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M,
 Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguenec C, Lescat
 M, Mangenot S, Martinez-Jehanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf
 CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Medigue C, Rocha EP, Denamur E. 2009.
 Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive
 paths. PLoS Genet 5:e1000344.
- Becker L, Fuchs S, Pfeifer Y, Semmler T, Eckmanns T, Korr G, Sissolak D, Friedrichs M, Zill E,
 Tung ML, Dohle C, Kaase M, Gatermann S, Rüssmann H, Steglich M, Haller S, Werner G. 2018.
 Whole Genome Sequence Analysis of CTX-M-15 Producing *Klebsiella Isolates* Allowed
 Dissecting a Polyclonal Outbreak Scenario. Front Microbiol 9:322.
- 42. Huang W, Li L, Myers JR, Marth GT. 2012. ART: a next-generation sequencing read simulator.
 Bioinformatics 28:593-4.
- 43. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
 Bioinformatics 25:1754-60.
- 640 44. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler
 641 D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce
 642 framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-303.
- 643 45. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
 644 phylogenies. Bioinformatics 30:1312-3.
- 645 46. Sahl JW, Caporaso JG, Rasko DA, Keim P. 2014. The large-scale blast score ratio (LS-BSR)
 646 pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ
 647 2:e332.
- 648 47. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome
 649 alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol
 650 15:524.
- 48. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015.
 652 Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res 43:e15.
- 65449.Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new655developments. Nucleic Acids Res 47:W256-w259.
- Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo
 AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K,
 Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett
 NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S,
 Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from
 genotypes. J Antimicrob Chemother doi:10.1093/jac/dkaa345.

51. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. 2020. Easy phylotyping of Escherichia
coli via the EzClermont web app and command-line tool. Access Microbiol 2:acmi000143.
52. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer.
Bioinformatics 27:1009-10.

- 666
- 667

668 Figure legends

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

674

675 Figure 2: Core genome phylogeny of CP-Ec isolates received by the F-NRC. ML phylogeny of the 713 676 CP-Ec isolates was built with RAxML (45) from 372,238 core SNPs after sequence alignment on 677 MG1655-K12 (NC_000913.3) selected as the best reference. The genome sequence of Escherichia 678 fergunsonii strain ATCC 35469 (NC 011740.1) was used as an outgroup for the phylogenetic analysis. 679 Genomes from Touchon et al. (40) were also incorporated into the analysis. Genomic features are 680 indicated as in the figure key (left) from the inside to the outside circles: carbapenemases of the OXA, 681 NDM and other types, CTX-M ESBL, mutations in gyrA and parC QRDR region (FQ resistance), origin, 682 main ST, phylogroups. Reference genome from Touchon et al. are indicated as Touchon.

683

Figure 3. Resistance gene content of the F-NRC CP-Ec isolates. A. Box plot representation of the ARGs
content according to the phylogroup. The number of isolates belonging to each phylogroup is indicated
between brackets. The limits of the box indicate the lower and upper quartile. Outliers are indicated
by points above the maximal value. B: Relationships between the ARGs number and the number of
mutations in QRDR. In the box plot representations, the median is indicated by an horizontal bar and
the mean by a cross.

690

Figure 4. Carbapenemase gene content of the F-NRC CP-*Ec* isolates. A. Relationship between the carbapenemase allele and the ARGs content. In the box plot representation, the median is indicated by an horizontal bar and the mean by a cross. The limits of the box indicate the lower and upper quartile. Outliers are indicated by points above the maximal value. **B**. Number of isolates carrying a specific carbapenemase allele according to the number of QRDR mutations in their genomes. **C**. Peryear evolution of the percentage of isolates carrying the different carbapenemase alleles in the 2012-

697 2015 period. Year of isolation are indicated as in the figure key (right). Given their small number (n=22),
698 strains isolated before 2012 are not indicated.

699

700 Figure 5. Features of the main ST associated with CP-Ec isolates in France. A. Proportion of the isolates 701 according to the carbapenemase allele in each ST; B. Box plot representation of the ARGs content 702 according to the ST; C. Proportion of isolates according to the number of mutations in QRDR; D : 703 Proportion of isolates with a specific PBP3 allele. PBP3 characterized by a YRIP, YRIK or YRIN insertion 704 between positions 333 and 334, the YRIN insertion is generally associated with a A413V mutation 705 (YRIN-L) and less frequently with a E349K mutation (YRIN-K-L) (10); E. Proportion of isolates carrying 706 specific mutations in *ompC*: R195L, G138D or acquisition of a ST38-like *ompC* allele by recombination 707 (10). For each ST, the phylogroup, the total number of isolates and the number of isolates associated 708 with urinary tract infections are indicated in the upper panel. Intermediate (Inter) and minor ST 709 indicate ST represented by 4-6 and 1-3 isolates, respectively.

710

711 Figure 6. Core genome phylogeny of ST38 isolates. ML Phylogeny was based on genome sequences 712 of 92 CP-Ec from the F-NRC and 464 genome sequences retrieved from Enterobase and from the NCBI 713 database including 331 carrying a carbapenemase gene. A core genome (2,900,000 nt) alignment of 714 the *de novo* assemblies on the sequence of GCA 005886035.1 used as a reference was performed by 715 using Parsnp (47); ML phylogeny was built with RAxML (45) from 6,170 core SNPs after removing 716 recombined regions with Gubbins (48). The genome sequence of CNRC6O47 (ST963) was used as an 717 outgroup for the phylogenetic analysis. F-NRC isolates are indicated by red triangles (inner circle). 718 Other genomic features are indicated as indicated in the figure key (left) from the inside to the outside 719 circles: carbapenemases of the OXA, NDM and other types, absence of any plasmid of any Inc-type as 720 identified by using PlasmidFinder (14), ESBL of the CTX-M types, number of mutations in gyrA and parC QRDRs (FQ resistance), number of ARGs, mutations in *fts1*, *ompC*, *ompF*, geographical origin. The four 721 722 OXA-48-like clades (G1, G2, G3, G4) clustering most French isolates are colored in blue, red, violet and 723 green. A fifth clade (G5) corresponding to a possible outbreak of VIM-4 isolates in the East of Paris is 724 colored in brown

725

Figure 7. *bla*_{OXA-48} integration sites of ST38 CP-*Ec* clades 1, 2, 3 and 4. The genome sequences of GCA_005886035.1 (CP040390.1), CNR65D6, GCA_004759025.1 (CP038505.1), and CNR85I8 were chosen as representative genomes of clusters 1, 2, 3 and 4, respectively and aligned, by using Blastn, against the sequence of GCA_001900295.1 (CP010116.1), a ST38 isolate that does not carry a carbapenemase gene and whose genome sequence was complete. Gene annotations were derived from MG1655 (NC_000913.3). The regions surrounding the *bla*_{OXA-48} insertion sites in each clade were

more precisely aligned against the corresponding sequences of CP010116.1 and of representative genomes of the other clades by using Easyfig (52). *bla*_{OXA-48} is represented by a brown arrow, cointegrated genes of pOXA-48 origin with light green arrows and the IS1 transposase with a green arrow. Sequences used to perform the BLAST analysis on Illumina contigs are underlined in red. It is to note that, in the G3 clade, *bla*_{OXA-48} integration occurs into a genomic island inserted at tRNASer. Another genomic island is present in G1-type strains but does not carry *bla*_{OXA-48}.

738

739





















С

Fig.4



Phylogroup D A C A B2 B1 D D A C C A B1 F B2 B1 B2 F B2 A D A

Fig. 5





Tree scale: 0.001



Fig. 7

А



Figure S1. **Per-year analysis of the origin of the isolates received by the F-NRC. A**. Proportions of the isolates belonging to the different ST. The total number of isolates is indicated between brackets. The large proportion of ST90 isolates collected in 2012 is linked to a local outbreak. **B**. Proportions of the isolates collected in screening and infection situations as a function of the year and the phylogroup. The absolute number of isolates for each class is indicated inside the bars.

В



Figure S2. Core genome phylogenies of the main ST characterized by clades disseminating internationally. A. ST410; B. ST167; C. ST405; D. ST101. Phylogenies were based on genome sequences of A: 64 CP-*Ec* from the F-NRC and 146 genome sequences from the NCBI database; B: 35 CP-*Ec* from the F-NRC and 134 genome sequences from the NCBI database; C. 15 CP-*Ec* from the F-NRC and 145 genome sequences from the NCBI database; D. 13 CP-*Ec* from the F-NRC and 194 genome sequences from the NCBI database. Core genome (ST410: 3,522,000 nt; ST167: 3,276,000 nt; ST405: 3,685,000 nt; ST101: 3,774,000 nt) alignments of the *de novo* assemblies on the sequences of GCA_001442495.1 (ST410), GCA_003028815.1 (ST167), GCA_002142675.1 (ST405), GCA_002163655.1 (ST101) used as reference were performed by using Parsnp; ML phylogeny was built with RAxML from 6,176 (ST410), 10,393 (ST167), 37,482 (ST405), 19,967(ST101) core SNPs after removing recombined regions with Gubbins. The genome sequences of CNR93E7 (ST88), CNR93D10 (ST746), CNR73I9 (ST115), CNR93I2 (ST906) were used as outgroups for the phylogenetic analyses of ST410, ST167, ST405 and ST101 respectively. The origin from the F-NRC is indicated by red triangles close to the isolate name. Other genomic features are indicated as indicated in the figure key (left) from the inside to the outside lines: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA* and *parC* QRDR region (FQ resistance), number of ARGs, mutations in *ftsl, ompC, ompF*, geographical origin.



Figure S3. Core genome phylogenies of ST10 and ST131 isolates. A. ST10 phylogeny based on the genome sequences of 67 CP-*Ec* from the F-NRC and 153 genome sequences from the NCBI. **B.** ST131 phylogeny based on the genome sequences of 17 CP-*Ec* from the F-NRC and 462 genome sequences from the NCBI. Core genome (ST10: 813,000 nt; ST131: 1,641,000 nt) alignments of the *de novo* assemblies on the sequence of MG1655 (ST10) or GCA_000285655.3 (ST131) used as references were performed by using Parsnp; ML phylogeny was built with RAxML from 20,245 (ST10) and 14,366 (ST131) core SNPs after removing recombined regions with Gubbins. The genome sequences of CNR93D10 (ST746) and CNRAL47G10 (ST640) were used as outgroups for the phylogenetic analyses of ST10 and ST131, respectively. F-NRC isolates are indicated by red triangles (inner circle). Other genomic features are indicated as indicated in the figure key (left) from the inside to the outside circles: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA* and *parC* QRDR region (FQ resistance), number of ARGs, mutations in *ftsl, ompC, ompF,* geographical origin. The ST10 clade enriched in CP-*Ec* is indicated in green and the sub-clade with the YRIN insertion in PBP3 in red. Note that the circle for *ftsl* mutations is absent in ST131, as no YRIN-like duplication in PBP3 was identified among the analyzed isolates.



Figure S4. Core genome phylogeny of ST940 and ST1284 isolates. A: ST940 phylogeny based on genome sequences of 17 CP-*Ec* from the F-NRC and 88 genome sequences from Enterobase and the NCBIdatabase **B:** ST1284 phylogeny based on genome sequences of 14 CP-*Ec* from the F-NRC and 60 genome sequences from Enterobase and the NCBI database. A core genome (ST940: 3,639,000 nt; ST1284: 3,779,000 nt) alignment of the *de novo* assemblies on the sequence of ST940: ESC_LB2149 and ST1284: ESC_LB2152AA (from Enterobase) used as references was performed by using Parsnp; ML phylogeny was built with RAxML from 13,859 (ST940) and 2,009 (ST1284) core SNPs after removing recombined regions with Gubbins. The genome sequences of CNR98G1 (ST3022) and MG1655 (ST10) were used as outgroups for the phylogenetic analyses of ST940 and ST1284, respectively. F-NRC isolates are indicated by red triangles first column on the left. Other genomic features are as indicated in the figure key (left) from the left to the right columns: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA* and *parC* QRDR (FQ resistance), number of ARGs, mutations in *ftsl, ompC, ompF*, geographical origin.



Figure S5. **Core genome phylogenies of ST69, ST88 and ST90. isolates. A:** ST90 phylogeny based on the 15 CP-*Ec* from the F-NRC and 38 genome sequences from the NCBI database; **B:** ST69 phylogeny based on the 16 CP-*Ec* from the F-NRC and 244 genome sequences from the NCBI database; **C.** ST88 phylogeny based on the 15 CP-*Ec* from the F-NRC and 99 genome sequences from the NCBI database. Core genome (ST69: 2,735,000 nt; ST88: 3,472,000 nt; ST90: 3,801,000nt) alignments of the *de novo* assemblies on the sequence of GCA_002443135.1 (ST69), GCA_002812685.1 (ST88) or GCA_001900635.1 (ST90) used as references were performed by using Parsnp; ML phylogeny was built with RAxML from 4,596 (ST69), 12,557 (ST88) or 2,878 (ST90) core SNPs after removing recombined regions with Gubbins. The genome sequences of CNR33D9 (ST394), CNR83B9 (ST410) and CNR88B9 (ST847) were used as outgroups for the phylogenetic analyses of ST69, ST88 and ST90, respectively. F-NRC isolates are indicated by red triangles (first column on the left). Other genomic features are indicated as indicated in the figure key (left) from the left to the right columns: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA* and *parC* QRDR (FQ resistance), number of ARGs, mutations in *ftsI, ompC, ompF*, geographical origin.



Figure S6. **Core genome phylogeny of ST617 and ST648 isolates A**: ST617 genome sequences of 15 CP-*Ec* from the F-NRC and 65 genome sequences from the NCBI database; **B**: ST648 genome sequences of 14 CP-*Ec* from the F-NRC and 126 genome sequences from the NCBI database. A core genome (ST617: 3,407,000; ST648: 3,501,000 nt) alignment of the *de novo* assemblies on the sequences of GCA_002142695.1 (ST617) or GCA_004138645.1 (ST648) used as references was performed by using Parsnp; ML phylogeny was built with RAXML from 15,030 (ST617) or 3,782 (ST648) core SNPs after removing recombined regions with Gubbins. The genome sequences of CNR93D10 (ST746) and CNR71A8 (ST1485) were used as outgroups for the phylogenetic analyses of ST617 and ST648, respectively. F-NRC isolates are indicated by red triangles close to the isolate name. Other genomic features are indicated as indicated in the figure key (left) from the left to the right columns: carbapenemases of the OXA, NDM and other types, number of mutations in *gyrA* and *parC* QRDR (FQ resistance), number of ARGs, mutations in *ftsl, ompC, ompF*, geographical origin.