Success of *Escherichia coli* O25b:H4 ST131 clade C associated with a decrease in virulence

Marion DUPRILOT[a,b]#, Alexandra BARON[a], François BLANQUART[a,c], Sara DION[a], Philippe LETTÉRON[d], Saskia-Camille Flament-Simon[e], Olivier CLERMONT[a], Erick DENAMUR[a,f], Marie-Hélène NICOLAS-CHANOINE[a]#

[a]: Université de Paris, INSERM, IAME, 75018 Paris

[b]: APHP, Laboratoire de Microbiologie, Hôpital Beaujon, 92110 Clichy

[c]: Centre for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France

[d]: Université Paris Diderot UMR 1149 Inserm - ERL CNRS 8252 75018 Paris

[e]: Laboratorio de Referencia de *Escherichia coli* (LREC), Departamento de Microbioloxía e Parasitoloxía, Facultade de Veterinaria, Universidade de Santiago de Compostela (USC), Lugo, Spain.

[f]: APHP, Laboratoire de Génétique Moléculaire, Hôpital Bichât, 75018 Paris

#Address correspondence to:

Marion DUPRILOT, e-mail: marion.duprilot@gmail.com

Marie-Hélène NICOLAS-CHANOINE, e-mail: marie-helene.nicolas-chanoine@inserm.fr

Running Head: Loss of virulence in the global emergent ST131 clade C

Abstract word count: 248

Text word count: 5256

Abstract

Escherichia coli of sequence type (ST) 131 resistant to fluoroquinolones and producer of CTX-M-15 is globally one of the major extraintestinal pathogenic *E. coli* (ExPEC). ST131 phylogenesis showed that multidrug-resistant ST131 strains belong to a clade called C, descending from an ancestral clade called B, comprising mostly antibiotic-susceptible strains. Antibiotic resistance could appear as one of the keys of the clade C global success. We hypothesized that other features of ST131 clade C could contribute to this success since other major global ExPEC clones (ST73, ST95) are mostly antibiotic-susceptible. To test this hypothesis, we measured the growth abilities, early biofilm formation and virulence-factor content of a collection of clade B and clade C strains. Moreover, using competition assays, we measured the capacity of selected representative strains of clades B and C to colonize the mouse intestine and urinary tract, and to kill mice in a septicemia model. Clade B and C strains had similar growth ability. However, clade B strains were more frequently early biofilm producers, expressed mostly faster their type 1 fimbriae and displayed more virulence factor-encoding genes than clade C strains. Clade B outcompeted clade C in the gut and/or urinary tract colonization models and in the septicemia model. These results strongly suggest that clade C strain evolution includes a loss of virulence, *i.e.* a process that could enhance micro-organism persistence in a given host and thus optimize transmission. This process, associated with acquired antibiotic-resistance, could ensure clade C strain survival in environments under antibiotic pressure.

Importance

Extraintestinal pathogen *Escherichia coli* (ExPEC) are virulent but mostly antibiotic-susceptible. One worrying exception is ST131, a major multidrug resistant ExPEC clone that has spread worldwide since the 2000s. To contain the emergence of this threatening clone, we need to understand what factors favored its emergence and dissemination. Here, we investigated whether multidrug-resistant ST131 had advantageous phenotypic properties beyond multidrug resistance. To this end, we competed the emergent multidrug-resistant ST131 with its antibiotic-susceptible ancestor in different conditions: biofilm production, *in vivo* colonization and virulence experiments. In all *in vivo* competitions, we found that multidrug-resistant ST131 was losing to its ancestor, suggesting a lesser virulence of multidrug-resistant ST131. It was previously described that losing virulence can increase micro-organism persistence in some populations and subsequently its level of transmissibility. Thus, a decreased level of virulence, associated with multidrug resistance, could explain the global success of ST131.

Introduction

In the last two decades, *Escherichia coli* O25b:H4 of sequence type (ST) 131 has emerged worldwide among human extraintestinal pathogenic *E. coli* (ExPEC) (1, 2). This multidrug resistant clone has become a major public health issue. The evolutionary history of this successful ST has been reconstructed in details. Within O25b:H4 ST131, two clades called B and C were distinguished on the basis of their *fimH* gene allele, *H*22 and *H*30, respectively (3, 4). Through a time-calibrated phylogeny, Ben Zakour *et al.* showed that clade B is the ancestor of clade C and that sequential mutational events have shaped the evolution of ST131 from the 1950s (5). This evolutionary scenario shows the diversification of clade B into different subclades (from B1 to B5, then the intermediate B0) and notably clade C. Clade C also diversified into different subclades characterized by the acquisition of the *fimH*30 allele around 1980 (subclade C0), followed by the acquisition of the *gyrA*-1AB and *parC*-1aAB alleles encoding fluoroquinolone resistance that occurred around 1987 (subclades C1 and C2). All subclade C2 strains and some of subclade C1 forming C1-M27 cluster (6) are additionally resistant to extended spectrum cephalosporins (ESC) due to the production of extended spectrum β-lactamases (ESBL) of CTX-M type, CTX-M-15 and CTX-M-27, respectively. Epidemiologic studies mostly determined ST131 prevalence among *E. coli* isolates resistant to fluoroquinolones and/or producers of ESBL (7). Studies designed to assess the relative frequencies of clade B and clade C ST131 strains are rare (8–10) and those to assess the relative frequencies of B and C subclades strains do not exist. Recently, Kallonen *et al.* depicted 14.4% of ST131 clades among bacteremia *E. coli* isolates systematically collected between 2001 and 2012 in England (9). Using the English ST131 isolates' whole genome sequences, we were able to assess the relative frequencies of subclades B and C in the present study.

The evolutionary success of ST131 is still largely unexplained. Resistance to antibiotics, notably to fluoroquinolones and ESC may explain the success of clade C strains, as these resistances were shown not to impact the clade C growth fitness (11, 12). However, clones susceptible to antibiotics, such as ST73 and ST95, are as successful as ST131 clade C among bacteremia, suggesting that factors other than antibiotic resistance can participate in the success of a clone (9),

The aim of our study was thus to investigate the potential role of factors other than antibiotic resistance, notably growth rate, biofilm formation, colonization ability and virulence, in the success of ST131 clade C strains. To this purpose, we first established a collection of 39 *E. coli* O25b:H4 ST131 comprising representatives of the ST131 clade and subclade genetic diversity and tested them in a variety of *in vitro* phenotypic tests. We then selected three strains for further *in vitro* and *in vivo* competition assays, including various mouse models.

Materials and methods

Technical details for each section are available in the supplemental material Text S1.

Bacterial strains

The 39 studied O25b:H4 ST131 *E. coli* strains comprised 18 *fimH*22 and 21 *fimH*30 strains obtained between 1993 and 2012 from different geographic origins and sources (Table S1). Nalidixic acid and ciprofloxacin resistance had been determined by the agar diffusion method and interpreted following the 2015 EUCAST recommendations (www.eucast.org) and ESBL production by the double disk synergy test (13). CFT073 and *E. coli* K-12 MG1655 strains were used as positive and negative controls, respectively, in the septicemia mouse model.

Genome sequencing and analysis

Whole genome sequencing (WGS) of our 39 strains was performed (Tables S2). All genomes were analyzed for plasmid content and typing [Plasmid Finder, with identity >95%, and

pMLST (14)]. Then, Abricate (15) was used to detect genes encoding antibiotic resistance with Resfinder (16) and virulence factors with a custom virulence database composed of VirulenceFinder (17), the virulence factor database (VFDB) (18), and classical genes characterizing ExPEC. Virotypes were determined as previously described (19). All contigs were submitted to the MicroScope Platform (20) for further gene investigation such as *gyrA/parC* alleles and genes involved in the biofilm formation (21). When necessary, presence or absence of some genes was controlled by PCR.

We investigated how the different subclades B and C of ST131 were represented in our collection and how they evolved in frequency over time in a larger collection from Kallonen *et al.* (9). To that end, we complemented our 39 genomes with 218+21 genomes from two published studies [218 from Kallonen *et al.* (9) ,21 from Ben Zakour *et al.* (5)]. The phylogenetic tree including all the strains was constructed from non-recombinant single nucleotide polymorphisms (SNPs) of core genomes using maximum likelihood. To infer the linear trend of Kallonen *et al.*'s isolates over 2001-2012, we fitted a logistic model to the frequency of each subclade as a function of time (in years).

Gene deletion and gene complementation

Replacement by a kanamycin resistance cassette was used to inactivate the chromosomal *fimB* gene of S250 and CES131C strains and the plasmidic *aadA2* gene of CES131C strain following a strategy adapted from Datsenko and Wanner (22). Primers and plasmids used are listed in Tables S3 and S4. Complementation of Δ*fimB* mutants were performed by cloning the promoter and encoding regions of the parental *fimB* genes into pSC-A-amp/kan (Table S4) by using the StrataClone PCR Cloning Kit (Agilent Technologies, Massy, France). The recombinant plasmid was then electroporated into competent strains S250Δ*fimB*:: or CES131CΔ*fimB*:: and transformants were selected on lysogeny broth (LB) (Invitrogen, Carlsbad, California, USA) agar plates containing 100 mg/L of kanamycin. The empty

plasmid pSC-A-amp/kan-Control_Insert was used as negative control. All mutants were confirmed by PCR and sequencing.

Determination of early biofilm formation

The primary step (5 h incubation) of biofilm formation was measured using BioFilm Ring Test® (BioFilm Control, Saint-Beauzire, France) according to the manufacturer's recommendations and as previously described (23). The biofilm formation index (BFI), with values ranging from 20 (absence of biofilm formation) to 0 (high biofilm formation), is inversely proportional to the biofilm formation ability. A BFI value of 10 was chosen as the biofilm production cut-off (biofilm formation: BFI≤10, no biofilm formation: BFI>10). The experiments were repeated at least three times for each strain. BHI broth without any strains was used as negative control, and strains S250 and 39, previously described with this method as strong and never biofilm producers, respectively, as control strains (24).

Yeast cell agglutination assay

Expression of type 1 fimbriae was assessed by using the yeast cell (*Saccharomyces cerevisiae*) agglutination assay as previously described (25), and after adaptations to highlight the early expression of type 1 fimbriae, by using 10 µL of a pellet obtained after centrifugation (3000 g, 10 min) of 3 mL of LB broth culture after 2 and 5 h-incubations for the agglutination assay. Based on the biofilm formation and the sequence of the *fim* operon, strains S250 and H1447 were used as positive and negative controls, respectively.

Congo red assays

Curli and cellulose production was assessed by using stationary-phase cultures spotted onto LB low salt (Invitrogen) plates supplemented with 40 µg/mL Congo red (Sigma Aldrich, St-Quentin Fallavier, France) at 30°C during 48 h, as previously described (26). The production of curli fimbriae and cellulose resulted in the red, dry and rough (rdar) colony morphotype.

Colicin and/or phage production

Colicin and/or phage production was detected by plaque lysis assays, with *E. coli* K-12 used as the sensitive strain as previously described (27). The assay was considered as positive when tested strains were surrounded by a halo, corresponding to the growth inhibition of *E. coli* K-12 strain.

Fitness assays

Individual fitness assays

Fitness assay was performed as previously described in LB broth (28) by using an automatic spectrophotometer (Tecan Infinite F200 Pro) that measures the OD_{600} in each well every 5 min over a period of 24 h. The experiment was repeated three times. Growth curves were then analyzed and maximum growth rates was calculated and expressed in h^{-1}.

Competition experiments

Selected strains of our collection and constructed mutants were submitted in pairs to competition in different conditions.

In vitro planktonic competitions

To determine the relative fitness of strains in planktonic conditions, they were grown in couple in LB broth during 100 generations, *i.e.* during ten days, as previously described (29). Competitive indexes (CI) were obtained using the following formula: $\log CFU[(isolate1/isolate2)T_x/(isolate1/isolate2)T_0]$, isolate1 being the first strain cited and isolate2 the second one in figure legends.

In vitro competitions in biofilm conditions

To determine the relative fitness of strains in biofilm conditions, they were grown at 1:1 ratio in mixed colonies during seven days in aerobic and anaerobic conditions, with adaptations to what's previously described (30). For each competition, the experience was repeated three times. CIs were obtained as described above.

Intestinal colonization mouse model

Six-week-old female mice CD-1 from Charles River® (L'Arbresle, France) pre-treated with streptomycin before inoculation of challenging strains were used to assess strains' relative ability to colonize the mouse intestine, as previously described (28). At least five mice were used for each competition. At days 1, 4 and 7, the intestinal population of *E. coli* was estimated by plating dilutions of weighed fresh feces on LB agar and LB agar with appropriate antibiotics (kanamycin 50 mg/L or ciprofloxacin 1 mg/L). CIs were obtained as described above.

Septicemia mouse model

Female mice OF1 of 14-16 g (4-week-old) from Charles River® were used to assess the strains' relative ability to cause sepsis, as previously described (31). After subcutaneous inoculation of individual strain or mixed strains at a 1:1 ratio, time to death was monitored during seven days. Five to ten mice were used for each assay. In all experiments, CFT073 strain and *E. coli* K-12 MG1655 strain were used as positive and negative control, respectively (32). Kaplan-Meier estimates of mouse survival were performed for individual strain inoculations. In competitions assays, the spleen of all spontaneously dead mice was collected, weighed and pounded in physiological water and infecting population of *E. coli* was estimated by plating dilutions of spleen suspensions on LB agar and LB agar with appropriate antibiotics (kanamycin 50 mg/L, ciprofloxacin 1 mg/L or ampicillin 100 mg/L). CIs were obtained as described above.

Urinary tract infection mouse model

CBA female mice of 8-22g (8-week-old) from Janvier® (Le Genest-Saint-Isle, France) were used to assess strains' relative ability to cause an ascending unobstructed urinary tract infection, as previously developed in our lab (33). Ten mice were used for each competition assay, and the experiments were repeated twice. CIs were obtained as described above.

Statistical analysis

Wilcoxon rank sum test was used to compare the average number of virulence factor (VF) per clade and Fisher's exact test was used to test the distribution of each VF-encoding gene between clades. Individual fitness measurement was estimated by a mixed model with random effect on the strain to take into account the triplicate determination. For biofilm data, two-way repeated measures ANOVA was performed, followed by Tukey's range test when necessary (if there were three groups). The association between groups (virotypes, biofilm phenotypes, yeast agglutination test and Congo red assay) was assessed by Fisher's exact test. For the *in vitro* and *in vivo* competitions, a non-parametric Wilcoxon test on paired data was conducted on CI values, and P values were corrected for multiple comparisons by the Benjamini-Hochberg procedure (34) when necessary. Mouse survival differences were determined by Log-Rank test. A significance level of 0.05 was used for all tests. All statistical analyses were carried out with R software (35).

To visualize all data of our strain collection in one graph, a multiple correspondence analysis (MCA) was realized with R software (35). The variables included in the analysis were strain source, *fimH* allele, fluoroquinolone susceptibility, ESBL production, and biofilm formation phenotype. The distance between the variables and the origin of the graph measures the quality of their representation and is estimated by the square cosine (cos2): variables far from the origin of the graph are discriminant and got a high cos2 and variables close to the origin are not perfectly depicted and got a low cos2 (Figure S1A).

Ethics statement

Colonization, septicemia and pyelonephritis murine protocols (n°APAFIS#4948-2016021216251548 v4, n°APAFIS#4951-2016020515004032 v2 and n°APAFIS#4950-2016021211417682 v4, respectively) were approved by the French Ministry of Research and by the ethical committee for animal experiments.

Data availability

Raw sequences of the 39 isolates were deposited in GenBank under BioProject accession number PRJNA320043 and PRJNA566165.

Results

Evolutionary success of the C1 and C2 subclade strains in extraintestinal infections

Using both the genome sequences of the ST131 strains published by Kallonen *et al.* (9) and the genome sequences of strains belonging to the different B and C subclades (5), we characterized the B and C subclades of Kallonen *et al.*'s O25b:H4 ST131 strains (Figure S2). By investigating trends in O25b:H4 ST131 subclade frequency over the period 2001-2012 (Figures 1A, 1B, 1C) in extraintestinal infections, we observed that (i) B4 and B5 subclade strains increased more than the other B subclade strains, (ii) clade C strains increased more than clade B strains and (iii) C2 subclade strains displayed the most important increase. Thus, although both clades B and C were maintained over this time period, clade C appears more evolutionarily successful than clade B.

Establishment of a collection of representative strains of the ST131 O25b clades and subclades

Clades and subclades

To determine which previously established ST131 clades and subclades our 39 strains belonged to, a phylogenetic tree was inferred using core genomes exempt of recombinant regions (Figure S2). Our collection comprised strains belonging to each subclade B (except for B2 and B0) and each subclade C. Among the 19 clade B strains, 17 displayed the expected *fimH*22, one a fim*H22*-like variant, and one, CES131C strain belonging to subclade B4, the

*fimH*30 variant that had been described as specific of clade C strains. For these reasons, CES131C strain was called "Hybrid" in the rest of the study.

Plasmids

All strains but one (strain 208) harbored an IncF plasmid with a FII, FIA and FIB replicon allele composition corresponding to the one previously described for clade B and C1 and C2 subclades strains (12), except that one C1 subclade strain (CES9C) lacked the FIB replicon as described in C2 subclade strains (Table S5).

Genes encoding antibiotic resistance

As shown in Table S5, our collection comprised strains that displayed a diverse antibiotic resistance-encoding gene content in both clades B and C, with a gradual accumulation of these genes from clade B to clade C.

VF-encoding genes

Among the genes classically sought for in ExPEC (19), those found in our strains are presented according to subclades in Table S6. The average number of VF-encoding genes was higher in clade B strains (17±2.1) than in clade C strains (14.3±1.3) ($P<0.0001$). The *hlyF*, *cdtB*, *iroN*, *kpsMII*, *cvaC*, *iss* and *ibeA* genes were significantly more frequent among clade B strains than among clade C strains whereas the opposite was found for the *sat* gene. Virotypes (19) were significantly associated with subclades ($P<0.00001$) and virotypes not yet described were displayed by Hybrid, one C1 subclade and three C2 subclade strains.

Individual strain phenotypic assays *in vitro*

Maximum growth rate (MGR)

We assessed the fitness of our 39 strains by measuring MGR in planktonic conditions (Figure 3). MGR did not differ significantly between clade B and clade C strains ($P = 0.08$), and between clinical and feces isolates ($P=0.07$) (Figures 3A and 3B). Within clade B strains, those resistant to nalidixic acid seemed to have a lower MGR than the susceptible ones

297 (Figure 3C), but this difference was not significant (P=0.2). Regarding clade C strains, no
298 significant MGR differences were observed between nalidixic acid/ciprofloxacin-resistant
299 strains and susceptible strains (P=0.9) (Figure 3D).
300 Kinetic of early biofilm formation and expression of type 1 and curli fimbriae
301 Clade B strains were more frequently biofilm producers than clade C strains at 2 and 5 h
302 (P<0.001). We found three significantly different phenotypes of early biofilm production
303 (P<0.0001) within our 39 strains (Figure 2, Table S5): early and persistent producers (BFI≤10
304 at 2, 3 and 5 h), delayed producers (BFI>10 at 2 and 3 h but ≤10 at 5 h) and never producers
305 (BFI>10 at 2, 3 and 5 h). As BFI values after 2 and 3 h of incubation classified a given strain
306 into the same biofilm production phenotype (Figure 2A), only BFI values obtained after 2 and
307 5 h of incubation were considered for further analyses. The 20 clade C strains were either
308 delayed producers (n=8) or never producers (n=12). There was no significant association
309 between subclades C and biofilm phenotype. Clade B strains produced biofilm more
310 frequently than clade C strains. Five (71%) of the B5 strains were early producers and two
311 (29%) were delayed producers, whereas all eight B4 strains were delayed or never producers.
312 This difference between B5 and B4 strains was statistically significant (P=0.02).
313 As the ability to form early biofilm is linked to type 1 fimbriae and curli fimbriae expression
314 (36), we studied these expressions in our strains by using the yeast agglutination test and
315 analyzing colony aspect of bacteria spotted on Congo red agar, respectively. After 2 and 5 h
316 of shaking growth, we observed that the early biofilm producers (n=8) expressed their type 1
317 fimbriae more quickly than the delayed and never producers (P≤0.0001) (Table 1). Applying
318 the standard yeast agglutination test as Totsika *et al.* (25) in which fimbriae expression is
319 assessed after ≥ 24 h, no significant differences were observed for type 1 fimbriae expression
320 according to biofilm phenotypes or clade membership, in both shaking and static conditions
321 (Table 1). Production of curli fimbriae and cellulose was explored (Table 2). There was a

significant association between Congo red morphotype and early biofilm formation capacity (P<0.0001). Significantly more clade B (47%) than clade C strains (15%) were positive for the Congo red assay (P<0.05).

Genes potentially linked to different phenotypes of early biofilm formation

To investigate the genetic factors potentially involved in differences in biofilm formation, we examined the diversity in the *fim* operon across all strains, knowing that Fim proteins participate in biofilm formation (21) and that an insertion sequence ISEc55 has been described in the clade C strain *fimB* gene (25) that encodes a co-factor of the type 1 fimbriae synthesis regulation. We found at least one non-synonymous SNP per gene in the *fim* operon between clade B and clade C strains (Figure S3). The entire *fim* operon was the same in all clade C strains, including ISEc55 within the *fimB* gene. Within clade B, two strains (H1447 and 001-001) found as delayed biofilm producers possessed a partially deleted *fim* operon. Hybrid, which was also a delayed biofilm producer, displayed a *fim* operon completely different from that of the other clade B strains, showing a perfect match with the *fim* operon of clade C strains, except for two SNPs and the absence of ISEc55 within its *fimB* gene (Figure S3). In order to assess the role of the *fimB* gene inactivation in clade C, we constructed Δ*fimB* variants from an early biofilm producer B1 strain (S250 strain), and a delayed biofilm producer (Hybrid). Both Δ*fimB* variants lost their ability to form biofilm during the 5h-time period (Figures 2B and 2C). Complementation with the wild *fimB* gene restored the biofilm production phenotype displayed by the parental strain of the two variants (Figures 2B and 2C). Δ*fimB* variants were also submitted to yeast agglutination test and Congo red assay. Both Δ*fimB* variants were negative for early yeast agglutination test, unlike parent strains, but were positive for standard test. No changes were observed on Congo red phenotype for both variants (data not shown).

Further genes and/or deduced proteins previously described as involved in *E. coli* biofilm formation (21) were compared between our strains by taking S250 strain as reference (data not shown). We observed identical protein sequence modifications in all B4 strains that were all delayed biofilm producers: substitution K142Q in FlgD that is required for flagellar hook assembly (37), the presence of the *nanC* gene (absent in the other clade B strains) and encoding N-acetylneuraminic acid porin of which regulators also control the *fimB* expression (38), and a 7-bp deletion at the end of the *ydaM* gene encoding a diguanylate cyclase playing a role in the curli biosynthesis by inducing the transcription of the regulator CsgD.

Strain selection for competition assays

In order to compare more accurately the fitness of clade B and C strains, we turned to competitions assays. To select candidates for competitions among our 39 strains, we used the phylogenetic history (Figure S2), the multiple correspondence analysis (Figure S1B), acquired antibiotic resistance (Table S5) and MGR values. Among clade B strains, we selected strain S250 *i.e.* one of the two most ancestral strains (subclade B1) that we called "Ancestor". Among clade C strains, we selected a C1 subclade strain, CES164C, harboring the clade C-characteristic fluoroquinolone resistance and a limited plasmid-mediated resistance (amoxicillin resistance related to TEM enzyme), to avoid the potential burden of additional resistances, such as that related to the production of CTX-M-15 in each C2 subclade strain. Accordingly, we called CES164C strain "Emergent". We also retained Hybrid (subclade B4) that harbors the first trait having characterized the emergence of clade C strains *i.e.* the *fimH*30 allele.

As shown in Table 3, Ancestor, Emergent and Hybrid displayed a similar average MGR in LB (Wilcoxon Rank Test, $P>0.1$), were non-colicin/phage producers, but displayed differences in early biofilm formation, type 1 fimbriae expression and Congo red phenotype, as well as in the number and content of genes encoding antibiotic resistance and VFs (Table 3 and Figure

4). However, some VF-encoding genes were present only in Ancestor: the invasin-encoding *ibeA* gene, the salmochelin-encoding *iroBCDEN* cluster, and the *etsC* gene encoding a putative secretion system I, whereas some others were only present in Hybrid: the *afaACDE* and *draP* encoding Afa/Dr adhesins, the *daaF* gene encoding the F1845 fimbrial adhesin, the *celB* gene encoding a major carbohydrate active-transport system and the *mcbA* gene encoding a bacteriocin.

To assess the potential role of the *fimB* gene in the fitness of our strains, we also retained Δ*fimB*::Kana variants of Ancestor and Hybrid, the absence of kanamycin cassette fitness cost having been checked *in vivo* by competing AncestorΔ*fimB*::Kana against AncestorΔ*fimB*:: and HybridΔ*fimB*::Kana against HybridΔ*fimB*:: (see Figures 6 and 8). We also checked the absence of fitness cost of the Δ*aada2*::Kana variant by competing HybridΔ*aada2*::Kana against Hybrid in planktonic conditions in LB, as we were obliged to use only streptomycin susceptible strains in the intestinal colonization model (data not shown).

Competition assays

In planktonic conditions

No significant difference was observed between the relative fitness of Ancestor, Emergent and Hybrid in aerobic batch LB growth as CIs did not exceed 1 or -1 log (Figure 5A). The absence of the *fimB* gene did not impact the relative fitness of Ancestor and Hybrid (Figure 5A).

In mature biofilm conditions

No significant difference was observed between the relative fitness of Ancestor, Emergent and Hybrid in aerobic and anaerobic biofilm growth conditions as CIs did not exceed 1 or -1 log (Figure 5B). However, a small competitive advantage was observed for both Ancestor and Hybrid against Emergent, after 7 days in anaerobic biofilm conditions. Lack of the *fimB* gene did not impact the relative fitness of Ancestor and Hybrid in these conditions (Figure 5B).

Intestinal colonization mouse model

Ancestor displayed a better intestinal colonization capacity than Emergent. Following the streptomycin treatment, none of the mice used in this experiment had Enterobacteriaceae in feces at the time of tested bacteria inoculation. Feces sampled at each studied time contained between 10^5 and 10^{10} *E. coli* per gram showing a satisfactory colonization rate of mice (data not shown). Ancestor outcompeted significantly Emergent after seven days post-inoculation by at least 1.5 log CFU (P = 0.02) (Figure 6), whereas there were no significant differences between Ancestor and Hybrid, and Emergent and Hybrid. Lack of the *fimB* gene did not seem to impact the colonization ability of Ancestor and Hybrid (Figure 6).

Septicemia mouse model

In terms of *in vivo* virulence expression, Ancestor, Emergent and Hybrid were "killers", as they killed between 10 and 80% of the mice 24 h after inoculation (39). Kaplan-Meier survival curves showed significantly different killing patterns (Figure 7): Ancestor killed 70% of the mice in 24 h and 100% in 26 h whereas Emergent and Hybrid, which killed 0 and 20% of the mice in 24 h, respectively, required four days to kill 100% of the mice (P<0.001). No significant difference was observed between the positive control CFT073 killing pattern and the Ancestor's one. The lack of the *fimB* gene in Ancestor and Hybrid did not impact their killing pattern (Figure 7). In order to observe subtler differences in the *in vivo* virulence, competitions in a 1:1 ratio were performed in this mouse model. Spleen bacterial analyses showed that Ancestor outcompeted by 1 and 2 log Hybrid and Emergent, respectively (P=0.002), and Hybrid outcompeted Emergent by 1 log (P=0.002) (Figure 8). Lack of the *fimB* gene in Ancestor and Hybrid did not impact their *in vivo* virulence in this mouse model (Figure 8).

Urinary tract infection mouse model

Ancestor displayed a better urinary tract colonization capacity than Emergent. In total, 30

mice were inoculated, with an overall infection rate of 87%, and no death was observed 48 h after the inoculation. Bladders and kidneys collected 48 h after inoculation contained an average of 9.10^7 CFU/g (range: 4.10^4-2.10^9 CFU/g) and of 5.10^5 CFU/g (range: 6.10^1-2.10^6 CFU/g), respectively, showing a satisfactory colonization rate of mice (data not shown). Ancestor outcompeted Emergent by about 2 log in the bladder and the kidneys (P=0.004) (Figure 9). Hybrid outcompeted Emergent by about 2.5 log in the bladder (P=0.008) but not in the kidneys (Figure 8). No significant difference was observed between Ancestor and Hybrid, even though Hybrid was predominant of about 2 log more in average in the kidneys (Figure 9).

Discussion

The purpose of this work was to examine the phenotypic differences between clade B and clade C strains regarding growth rate, early biofilm formation, colonization ability and virulence, in order to estimate their potential role in the success of ST131 clade C strains. The originality and strength of our study was to analyze these phenotypic differences in various i*n vitro* and *in vivo* models and under competition conditions. The limited number of strains submitted to *in vivo* competitions was determined by ethical concerns related to animal experimentations. Transposition of our *in vivo* results to human has to be made with caution because the animal models do not strictly reflect the infection physiological process observed in humans, particularly for the UTI model in which bacteria urethra ascent is absent.

Our ST131 collection comprised representative strains of subclade B and C, except for one, called Hybrid, which belonged to subclade B4 and harbored a *fimH*30 allele. This exceptional feature has already been reported in another subclade B4 strain (5), indicating that *fim* operon

445 recombination that shaped the clade C evolution had also occurred among clade B strains, but
446 apparently with a lower adaptive success in the genetic background of clade B than in that of
447 clade C.

448 No growth rate difference was detected between our clade B and clade C strains and between
449 fluoroquinolone susceptible and resistant strains. The latter finding is consistent with the fact
450 that, among the mutations leading to fluoroquinolone-resistance, those observed in ST131
451 displayed the lowest fitness cost (11). Moreover, no growth rate difference was observed
452 between clinical and commensal strains of our collection. This might suggest that ST131 is
453 adapted to both commensalism and extraintestinal virulence, supporting the idea that
454 virulence could be a derivative of commensalism (29).

455 The genetic and phenotypic variability of ST131 subclades were rarely explored in details in
456 previous literature. Our analyses of Kallonen *et al.*'s blood *E. coli* isolates showed that C1 and
457 C2 subclade strains increased in frequency over time and that clade B strains, notably those of
458 B4 and B5 subclades, persisted as well, representing 16.5% of the ST131 isolates.
459 One of the most interesting findings of our study was that clade C strains were less virulent
460 and had lower intestinal and urinary colonization capacities than clade B strains. Regarding
461 intestinal colonization, no comparison exists in the literature between these two clades, but
462 contradictory results were obtained considering intestinal cell adhesion capacity (40). In terms
463 of virulence, conflicting results were published, as, on the one hand, fluoroquinolone
464 susceptible ST131 strains were shown to kill 50% of the mice when resistant ST131 strains
465 killed 20% (41), and on the other hand, strains of virotype D, *i.e.* clade B-specific virotype,
466 were shown to be less virulent than strains of other virotypes associated with clade C (39).
467 Regarding urinary colonization, a clade C ST131 strain compared with strains belonging to
468 well-known UPEC STs in separate mouse urinary tract infection (UTI) models, was shown to
469 have a higher UTI ability (28). Here, we demonstrated that a representative subclade B1 strain

(Ancestor) outcompeted a representative clade C strain (Emergent) in a UTI model. Moreover, the "Hybrid" B4 subclade strain (harboring the *fimH*30 allele) outcompeted the clade C strain in the bladder but not in the kidneys. Interestingly, Hybrid strain also displayed intermediate phenotypes with regard to intestine colonization and sepsis in mice, which could illustrate the gradual loss of virulence of ST131 over its phylogenetic evolution.

As previously described, our clade B strains were more frequently early biofilm producers than clade C strains (23, 24, 42). Knocking out the *fimB* gene in clade B strains resulted in loss of both early biofilm production capacity and early type 1 fimbriae expression. It was shown that clade C ST131 mutants devoid of type 1 pili expression had a lower intestinal and urinary tract colonization abilities compared to parent strains (25, 40). We subjected Δ*fimB* clade B mutants to competition experiments with parent strains. No differences were detected neither during growth in mature biofilm conditions nor during mouse intestinal colonization and septicemia model, suggesting that early biofilm production kinetic doesn't seem to contribute to the better colonization abilities or to the higher virulence observed in clade B strains. In the same way, *E. coli* biofilm formation capacity on abiotic surfaces was not correlated to the ability to durably colonize mouse intestine (29). However, to confirm this hypothesis, competitions between Ancestor Δ*fimB* mutant and Emergent would be required.

It may appear surprising that the successful clade C had lower virulence and lower competitive ability in the intestine and in a UTI model. We propose two explanations. First, the resistance to fluoroquinolones displayed by clade C may come at a direct cost (43). This cost would not be revealed in maximum growth rate in planktonic conditions, but would manifest as a lower colonization ability. In spite of this cost, antibiotic resistance allowed clade C to spread. Second, the lower virulence may actually confer an additional evolutionary advantage to clade C. According to the trade-off theory (44), host exploitation by a pathogen evolves to an optimal level under a balance between the benefits in terms of transmission and

the costs in terms of host mortality (45). Thus, lower virulence could confer improved fitness overall by avoiding symptomatic infections (46, 47), even if it comes at the cost of lower colonisation ability.

To conclude, we revealed substantial phenotypic differences in different clades of ST131, with the successful and multi-resistant clade C being notably less able to form biofilm, and less able to colonize the intestine and the urinary tract than its ancestor clade B. Whether these phenotypic differences represent a cost in spite of which clade C is successful, or an additional adaptation, remains an open question that would be interesting to explore in further work.

Acknowledgments

We thank Christophe Beloin for his advices about biofilm exploration. We thank Meril Massot, Marie Vigan and Cedric Laouénan for their help with statistical analyses. We also thank Olivier Tenaillon and Antoine Bridier-Nahmias for genomic analyses assistance. Finally, we thank Johann Beghain, Mélanie Magnan and Françoise Chau for their technical assistance in this work.

This study was supported by a grant from the project JPI-EC-AMR 2016 with the French Agence Nationale de la Recherche (ANR) as sponsor (N°. ANR-16-JPEC-0002-04) to MHNC and by the “Fondation pour la Recherche Médicale” (Equipe FRM 2016, grant number DEQ20161136698) to ED. S.C.F.S. acknowledges the FPU programme for her grant (FPU15/02644) from the Secretaría General de Universidades, Spanish Ministerio de Educación, Cultura y Deporte.

References

519 1. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L,
520 Billig M, Chattopadhyay S, Stegger M, Andersen PS, Pearson T, Riddell K, Rogers P,
521 Scholes D, Kahl B, Keim P, Sokurenko EV. 2013. The epidemic of extended-spectrum-
522 β-bactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic
523 subclone, *H*30-Rx. mBio 4:e00377-13.

524 2. Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, Riddell K,
525 Rogers P, Qin X, Butler-Wu S, Price LB, Aziz M, Nicolas-Chanoine M-H, Debroy C,
526 Robicsek A, Hansen G, Urban C, Platell J, Trott DJ, Zhanel G, Weissman SJ, Cookson
527 BT, Fang FC, Limaye AP, Scholes D, Chattopadhyay S, Hooper DC, Sokurenko EV.
528 2013. Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia*
529 *coli*. J Infect Dis 207:919–928.

530 3. Olesen B, Frimodt-Møller J, Leihof RF, Struve C, Johnston B, Hansen DS, Scheutz F,
531 Krogfelt KA, Kuskowski MA, Clabots C, Johnson JR. 2014. Temporal trends in
532 antimicrobial resistance and virulence-associated traits within the *Escherichia coli*
533 sequence type 131 clonal group and its *H*30 and *H*30-Rx subclones, 1968 to 2012.
534 Antimicrob Agents Chemother 58:6886–6895.

535 4. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM,
536 Phan M-D, Gomes Moriel D, Peters KM, Davies M, Rogers BA, Dougan G, Rodriguez-
537 Baño J, Pascual A, Pitout JDD, Upton M, Paterson DL, Walsh TR, Schembri MA,
538 Beatson SA. 2014. Global dissemination of a multidrug resistant *Escherichia coli* clone.
539 Proc Natl Acad Sci U S A 111:5694–5699.

540 5. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW,
541 Stanton-Cook M, Schembri MA, Beatson SA. 2016. Sequential acquisition of virulence

542 and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131.
543 mBio 7:e00347-00316.

544 6. Matsumura Y, Pitout JDD, Gomi R, Matsuda T, Noguchi T, Yamamoto M, Peirano G,
545 DeVinney R, Bradford PA, Motyl MR, Tanaka M, Nagao M, Takakura S, Ichiyama S.
546 2016. Global *Escherichia coli* sequence type 131 clade with $bla_{\text{CTX-M-27}}$ gene. Emerg
547 Infect Dis 22:1900–1907.

548 7. Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. 2014. *Escherichia coli* ST131, an
549 intriguing clonal group. Clin Microbiol Rev 27:543–574.

550 8. Morales-Barroso I, López-Cerero L, Molina J, Bellido M, Navarro MD, Serrano L,
551 González-Galán V, Praena J, Pascual A, Rodríguez-Baño J. 2017. Bacteraemia due to
552 non-ESBL-producing *Escherichia coli* O25b:H4 sequence type 131: insights into risk
553 factors, clinical features and outcomes. Int J Antimicrob Agents 49:498–502.

554 9. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ,
555 Parkhill J. 2017. Systematic longitudinal survey of invasive *Escherichia coli* in England
556 demonstrates a stable population structure only transiently disturbed by the emergence of
557 ST131. Genome Res 27:1437–1449.

558 10. Blanco J, Mora A, Mamani R, López C, Blanco M, Dahbi G, Herrera A, Blanco JE,
559 Alonso MP, García-Garrote F, Chaves F, Orellana MÁ, Martínez-Martínez L, Calvo J,
560 Prats G, Larrosa MN, González-López JJ, López-Cerero L, Rodríguez-Baño J, Pascual
561 A. 2011. National survey of *Escherichia coli* causing extraintestinal infections reveals
562 the spread of drug-resistant clonal groups O25b:H4-B2-ST131, O15:H1-D-ST393 and
563 CGA-D-ST69 with high virulence gene content in Spain. J Antimicrob Chemother
564 66:2011–2021.

11. Huseby DL, Pietsch F, Brandis G, Garoff L, Tegehall A, Hughes D. 2017. Mutation supply and relative fitness shape the genotypes of ciprofloxacin-resistant *Escherichia coli*. Mol Biol Evol 34:1029–1039.

12. Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, Flores-Figueroa C, Aziz M, Stoesser N, Sokurenko E, Price LB, Johnson JR. 2016. Separate F-type plasmids have shaped the evolution of the *H*30 subclone of *Escherichia coli* sequence type 131. mSphere 1:e00121-16.

13. Jarlier V, Nicolas MH, Fournier G, Philippon A. 1988. Extended broad-spectrum beta-lactamases conferring transferable resistance to newer beta-lactam agents in Enterobacteriaceae: hospital prevalence and susceptibility patterns. Rev Infect Dis 10:867–878.

14. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895–3903.

15. Seemann T. 2018. :mag_right: :pill: Mass screening of contigs for antimicrobial and virulence genes: tseemann/abricate. Perl.

16. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644.

17. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol 52:1501–1510.

588 18. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined
589 dataset for big data analysis—10 years on. Nucleic Acids Res 44:D694–D697.

590 19. Dahbi G, Mora A, Mamani R, López C, Alonso MP, Marzoa J, Blanco M, Herrera A,
591 Viso S, García-Garrote F, Tchesnokova V, Billig M, de la Cruz F, de Toro M, González-
592 López JJ, Prats G, Chaves F, Martínez-Martínez L, López-Cerezo L, Denamur E, Blanco
593 J. 2014. Molecular epidemiology and virulence of *Escherichia coli* O16:H5-ST131:
594 comparison with *H*30 and *H*30-Rx subclones of O25b:H4-ST131. Int J Med Microbiol
595 304:1247–1257.

596 20. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, Rouy Z, Roche D,
597 Salvignol G, Scarpelli C, Médigue C. 2009. MicroScope: a platform for microbial
598 genome annotation and comparative genomics. Database J Biol Databases Curation
599 2009:bap021.

600 21. Niba ETE, Naka Y, Nagase M, Mori H, Kitakawa M. 2007. A genome-wide approach to
601 identify the genes involved in biofilm formation in *Escherichia coli*. DNA Res Int J
602 Rapid Publ Rep Genes Genomes 14:237–246.

603 22. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in
604 *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645.

605 23. Flament-Simon S-C, Duprilot M, Mayer N, García V, Alonso MP, Blanco J, Nicolas-
606 Chanoine M-H. 2019. Association between kinetics of early biofilm formation and
607 clonal lineage in *Escherichia coli*. Front Microbiol 10:1183.

608 24. Nicolas-Chanoine M-H, Petitjean M, Mora A, Mayer N, Lavigne J-P, Boulet O, Leflon-
609 Guibout V, Blanco J, Hocquet D. 2017. The ST131 *Escherichia coli H*22 subclone from

610 human intestinal microbiota: Comparison of genomic and phenotypic traits with those of
611 the globally successful *H*30 subclone. BMC Microbiol 17:71.

612 25. Totsika M, Beatson SA, Sarkar S, Phan M-D, Petty NK, Bachmann N, Szubert M,
613 Sidjabat HE, Paterson DL, Upton M, Schembri MA. 2011. Insights into a multidrug
614 resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome
615 analysis and virulence mechanisms. PLOS ONE 6:e26578.

616 26. Dudin O, Geiselmann J, Ogasawara H, Ishihama A, Lacour S. 2014. Repression of
617 flagellar genes in exponential phase by CsgD and CpxR, two crucial modulators of
618 *Escherichia coli* biofilm formation. J Bacteriol 196:707–715.

619 27. Smati M, Clermont O, Bleibtreu A, Fourreau F, David A, Daubié A-S, Hignard C,
620 Loison O, Picard B, Denamur E. 2015. Quantitative analysis of commensal *Escherichia*
621 *coli* populations reveals host-specific enterotypes at the intra-species level.
622 MicrobiologyOpen 4:604–615.

623 28. Vimont S, Boyd A, Bleibtreu A, Bens M, Goujon J-M, Garry L, Clermont O, Denamur
624 E, Arlet G, Vandewalle A. 2012. The CTX-M-15-producing *Escherichia coli* clone
625 O25b:H4-ST131 has high intestine colonization and urinary tract infection abilities. PloS
626 One 7:e46547.

627 29. Diard M, Garry L, Selva M, Mosser T, Denamur E, Matic I. 2010. Pathogenicity-
628 associated islands in extraintestinal pathogenic *Escherichia coli* are fitness elements
629 involved in intestinal colonization. J Bacteriol 192:4885–4893.

630 30. Saint-Ruf C, Garfa-Traoré M, Collin V, Cordier C, Franceschi C, Matic I. 2014. Massive
631 diversification in aging colonies of *Escherichia coli*. J Bacteriol 196:3059–3073.

632 31. Smati M, Magistro G, Adiba S, Wieser A, Picard B, Schubert S, Denamur E. 2017.
633 Strain-specific impact of the high-pathogenicity island on virulence in extra-intestinal
634 pathogenic *Escherichia coli*. Int J Med Microbiol IJMM 307:44–56.

635 32. Johnson JR, Clermont O, Menard M, Kuskowski MA, Picard B, Denamur E. 2006.
636 Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits,
637 phylogenetic group, and ecological source. J Infect Dis 194:1141–1150.

638 33. Labat F, Pradillon O, Garry L, Peuchmaur M, Fantin B, Denamur E. 2005. Mutator
639 phenotype confers advantage in *Escherichia coli* chronic urinary tract infection
640 pathogenesis. FEMS Immunol Med Microbiol 44:317–321.

641 34. Holm S. 1979. A simple sequentially rejective multiple test procedure. Scand J Stat
642 6:65–70.

643 35. R development core team. 2014. R: A language and environment for statistical
644 computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-
645 07-0, URL: http://www.R-project.org.

646 36. Fronzes R, Remaut H, Waksman G. 2008. Architectures and biogenesis of non-flagellar
647 protein appendages in Gram-negative bacteria. EMBO J 27:2271–2280.

648 37. Ohnishi K, Ohto Y, Aizawa S, Macnab RM, Iino T. 1994. FlgD is a scaffolding protein
649 needed for flagellar hook assembly in *Salmonella typhimurium*. J Bacteriol 176:2272–
650 2281.

651 38. Condemine G, Berrier C, Plumbridge J, Ghazi A. 2005. Function and expression of an
652 N-acetylneuraminic acid-inducible outer membrane channel in *Escherichia coli*. J
653 Bacteriol 187:1959–1965.

654 39. Mora A, Dahbi G, López C, Mamani R, Marzoa J, Dion S, Picard B, Blanco M, Alonso
655 MP, Denamur E, Blanco J. 2014. Virulence patterns in a murine sepsis model of ST131
656 *Escherichia coli* clinical isolates belonging to serotypes O25b:H4 and O16:H5 are
657 associated to specific virotypes. PLoS ONE 9:e87025.

658 40. Sarkar S, Hutton ML, Vagenas D, Ruter R, Schüller S, Lyras D, Schembri MA, Totsika
659 M. 2018. Intestinal colonization traits of pandemic multidrug-resistant *Escherichia coli*
660 ST131. J Infect Dis 218:979–990.

661 41. Johnson JR, Porter SB, Zhanel G, Kuskowski MA, Denamur E. 2012. Virulence of
662 *Escherichia coli* clinical isolates in a murine sepsis model in relation to sequence type
663 ST131 status, fluoroquinolone resistance, and virulence genotype. Infect Immun
664 80:1554–1562.

665 42. Pantel A, Dunyach-Remy C, Essebe CN, Mesureur J, Sotto A, Pagès J-M, Nicolas-
666 Chanoine M-H, Lavigne J-P. 2016. Modulation of membrane influx and efflux in
667 *Escherichia coli* sequence type 131 has an impact on bacterial motility, biofilm
668 formation, and virulence in a *Caenorhabditis elegans* model. Antimicrob Agents
669 Chemother 60:2901–2911.

670 43. Basra P, Alsaadi A, Bernal-Astrain G, O'Sullivan ML, Hazlett B, Clarke LM,
671 Schoenrock A, Pitre S, Wong A. 2018. Fitness tradeoffs of antibiotic resistance in
672 extraintestinal pathogenic *Escherichia coli*. Genome Biol Evol 10:667–679.

673 44. Anderson RM, May RM. 1982. Coevolution of hosts and parasites. Parasitology 85 (Pt
674 2):411–426.

675 45. Diard M, Hardt W-D. 2017. Evolution of bacterial virulence. FEMS Microbiol Rev
676 41:679–697.

677 46. Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K, Nordstrom L, Gauld L, Weaver B,
678 Rolland D, Statham S, Horwinski J, Sariya S, Davis GS, Sokurenko E, Keim P, Johnson
679 JR, Price LB. 2018. *Escherichia coli* ST131-*H*22 as a foodborne uropathogen. mBio
680 9:e00470–18, /mbio/9/4/mBio.00470–18.atom.

681 47. Salvador E, Wagenlehner F, Köhler C-D, Mellmann A, Hacker J, Svanborg C, Dobrindt
682 U. 2012. Comparison of asymptomatic bacteriuria *Escherichia coli* isolates from healthy
683 individuals *versus* those from hospital patients shows that long-term bladder
684 colonization selects for attenuated virulence phenotypes. Infect Immun 80:668–678.

685

686 ## Figure legend

687 **Figure 1. Dynamics of subclades of O25b:H4 ST131 *E. coli* isolated from bacteremia over**
688 **a 11-year sampling period [data from Kallonen *et al* (9)].** A. Cumulated frequencies of
689 ST131 subclades as a function of time, from the most ancestral subclade B1 at the bottom to
690 the derived subclade C2 on top. B. Frequency of ST131 subclades as a function of time.
691 Subclades are represented by colors. C. Inferred linear trends of ST131 subclade frequencies
692 (among all *E. coli* strains) as a function of time over 2001-2012.

693 **Figure 2. Early biofilm production.** Biofilm Ring Test® was used to measure biofilm
694 formation after 2, 3 and 5h of static incubation. BFI values are inversely proportional to the
695 biofilm formation capacity. A. Biofilm formation for the 39 *E. coli* ST131 strains. Clade B:
696 continuous line; clade C: dashed line. Biofilm groups: blue lines: never producers; green
697 lines: delayed producers; orange lines: early producers. B and C. Biofilm measurements for
698 the *fimB* gene knockout experiments on Ancestor and Hybrid strains, respectively. Δ*fimB*:
699 Δ*fimB*::; Δ*fimB*p: Δ*fimB*::pSC-A, strain complemented with an empty plasmid; Δ*fimB*c:
700 Δ*fimB*::pSC-A-*fimB*, strain complemented with its *fimB* gene.

Figure 3. Maximum growth rate of the 39 *E. coli* strains in LB. Maximum growth rate is calculated from three independent culture assays and expressed in h^{-1}. Boxplots indicate maximum growth rate distribution across strains, and horizontal black bars indicate median values. The upper and lower ends of the box correspond to the upper and lower quartiles, respectively. Error bars represent the standard error of the mean of three experiments. Outliers are represented by dots. Strains were grouped into two categories: A. Clade B and clade C strains; B. Clinical and commensal (feces) strains; C. Nalidixic acid-resistant (R) and susceptible (S) strains within clade B; D. Nalidixic acid-resistant (R) and susceptible (S) strains within clade C. No significant difference was observed in each comparison (Wilcoxon test)

Figure 4. Virulence factors (VF)-encoding genes in Ancestor, Emergent and Hybrid. Each strain is represented by a circle, VF encoding genes specific of a strain are indicated in the unshared part of the circle, while genes in common are indicated in the intersecting region between the two or three involved strains. Among the 67 genes found at least once in these three strains, the 42 common genes are as follows: *fdeC, fimACDEFGHI, yfcV, yagVWXYZ ,ykgK, chuA, fyuA, entABCEFS, fepABCDG, fes, irp2, kpsDEMMTII, K5, iss, malX, usp, ompT, ompA, aslA.*

Figure 5. *In vitro* competitions. Dots shapes indicate the couple of strains in competition in panels A and B. Competitive index (CI) is expressed in log, and CI above zero (blue line) means that isolate 1 outcompetes isolate 2, and conversely. An interval of ±1 log was considered as the limit of discrimination. Error bars represent the standard error of the mean of a least three experiments. A. Competitions at an initial ratio of 1:1 in planktonic conditions in LB medium during 100 generations. B. Competitions at an initial ratio of 1:1 in mature biofilm conditions (bacteria spotted on BHI agar plate) during 7 days, in aerobic (continuous line) and anaerobic atmosphere (dashed line).

Figure 6. Competition in an intestinal colonization mouse model. Competitive index (CI) is expressed in log, and CI above zero means that isolate 1 outcompetes isolate 2, and conversely. An interval of ±1 log was considered as the limit of discrimination. Each dot represents a mouse at given times: light blue: day 1, dark blue: day 4 and green: day 7 post-inoculation. Black diamonds depict mean values of CI. Kanamycin resistance cassette: control competition evaluating the cost of the Kanamycin resistance cassette, with AncestorΔ*fimB*:: or HybridΔ*fimB*:: as isolate 1, and AncestorΔ*fimB*::Kana or HybridΔ*fimB*::Kana as isolate 2, respectively. * P=0.02, generated by the Wilcoxon signed-rank test and corrected for multiple comparisons by the Benjamini-Hochberg procedure.

Figure 7. Kaplan-Meier survival curves of mice injected with *E. coli* ST131 strains. Strains are represented by line types.

Figure 8. Competition in septicemia mouse model determined by spleen bacterial load. Competitive index (CI) is expressed in log, and CI above zero means that isolate 1 outcompetes isolate 2, and conversely. An interval of ±1 log was considered as the limit of discrimination. Kanamycin resistance cassette: control competition evaluating the cost of the Kanamycin resistance cassette, with AncestorΔ*fimB*:: or HybridΔ*fimB*:: as isolate 1, and AncestorΔ*fimB*::Kana or HybridΔ*fimB*::Kana as isolate 2 respectively. ** P=0.002, generated by the Wilcoxon signed-rank test.

Figure 9. Competition in urinary tract infection mouse model determined by bladder and kidney bacterial load. Competitive index (CI) is expressed in log, and CI above zero means that isolate 1 outcompetes isolate 2, and conversely. An interval of ±1 log was considered as the limit of discrimination. ** P<0.01, generated by the Wilcoxon signed-rank test and corrected for multiple comparisons by the Benjamini-Hochberg procedure.

751 **Table 1. Yeast agglutination test applied to the 39 strains in different growth conditions**

752 **and at different time points according to biofilm production phenotype and clade type.**

Number (%) of strains positive for the yeast agglutination test						
Group	Early test		Standard[a] test			
	Shaking culture		Shaking culture	Static subculture		
	2h	5h	24h	24h	48h	72h
Biofilm ++ (n=8)	8 (100)	8 (100)	8 (100)	8 (100)	8 (100)	8 (100)
Biofilm -+ (n=16)	4 (25) *	6 (38) *	14 (89)	14 (89)	14 (89)	14 (89)
Biofilm -- (n=15)	1 (7)	1 (7)	13 (87)	14 (93)	15 (100)	15 (100)
Clade B (n=19)	13 (68) *	15 (79) *	17 (90)	17 (90)	17 (90)	17 (90)
Clade C (n=20)	0	0	18 (90)	19 (95)	20 (100)	20 (100)

[a]: Standard test was realized according to Totsika *et al.*[25]; ++: early biofilm producer; -+: delayed biofilm producer; --: never biofilm producer; * Fisher exact test P≤0.0001.

753

754

755

756

757

758

759

760

761

762

763

764

765

766 **Table 2. Colony color and morphotype on Congo red agar plates of the**

767 **39 strains according to biofilm production phenotype and clade type**

Number (%) of strains with a positive Congor red test		
Group	Congo red staining	Rdar morphotype
Biofilm ++ (n=8)	8 (100)	8 (100)
Biofilm -+ (n=16)	1 (6) *	1 (6) *
Biofilm -- (n=15)	4 (27)	3 (20)
Clade B (n=19)	9 (47) **	9 (47) **
Clade C (n=20)	3 (15)	2 (10)

++: early biofilm producer; -+: delayed biofilm producer; --: never biofilm producer; Rdar : red, dry and rough colonies; * Fisher exact test P<0.0001;

768 ** Fisher exact test P<0.001

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785 **Table 3. Comparative characteristics of the three strains selected for competition assays**

Selected strain	MGR (h^{-1})	Colicin/phage production	Biofilm formation phenotype	Type 1 fimbriae expression	Congo red assay phenotype	Antibiotic resistance encoding gene	Nb of VF encoding genes	Nb of ExPEC VF encoding genes
Ancestor (S250)	1.35±0,06	no	early	in 2h	rdar	/	53	14
Emergent (CES164C)	1.48±0,06	no	never	in 24h	red staining	*bla*$_{TEM}$, *gyrA1AB*, *parC1aAB*	52	14
Hybrid (CES131C)	1.47±0,09	no	delayed	in 2h	white staining	*bla*$_{TEM}$, *aac(3)-IId*, *aadA2*, *mphA*, *sul1*, *tetA*, *dfrA12*	63	16

786 rdar: red, dry and rough; VF: virulence factor

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

Figures

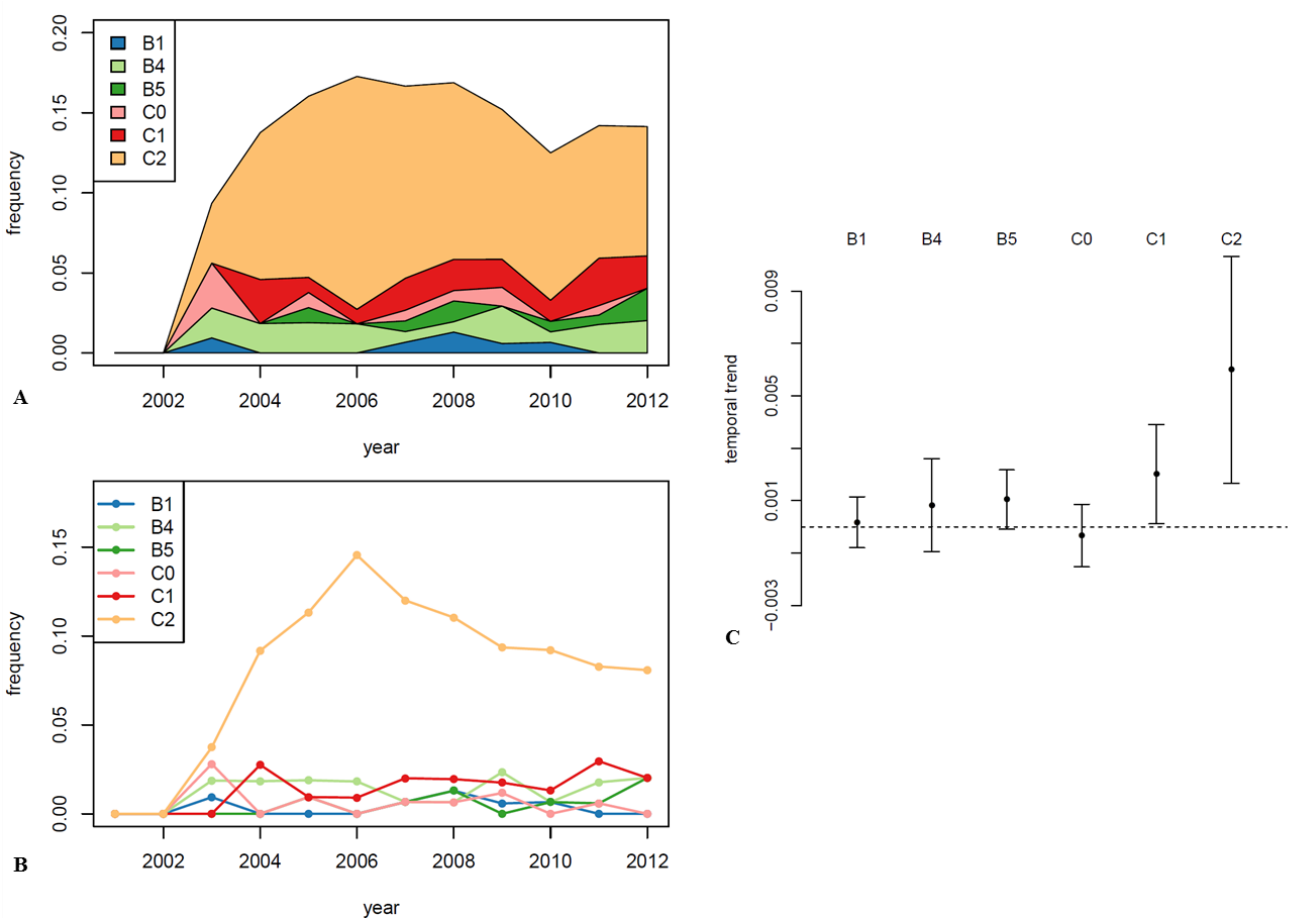


Figure 2. Dynamics of subclades of O25b:H4 ST131 *E. coli* isolated from bacteremia over a 11-year sampling period [data from Kallonen *et al* (9)]. A. Cumulated frequencies of ST131 subclades as a function of time, from the most ancestral subclade B1 at the bottom to the derived subclade C2 on top. B. Frequency of ST131 subclades as a function of time. Subclades are represented by colors. C. Inferred linear trends of ST131 subclade frequencies (among all *E. coli* strains) as a function of time over 2001-2012.

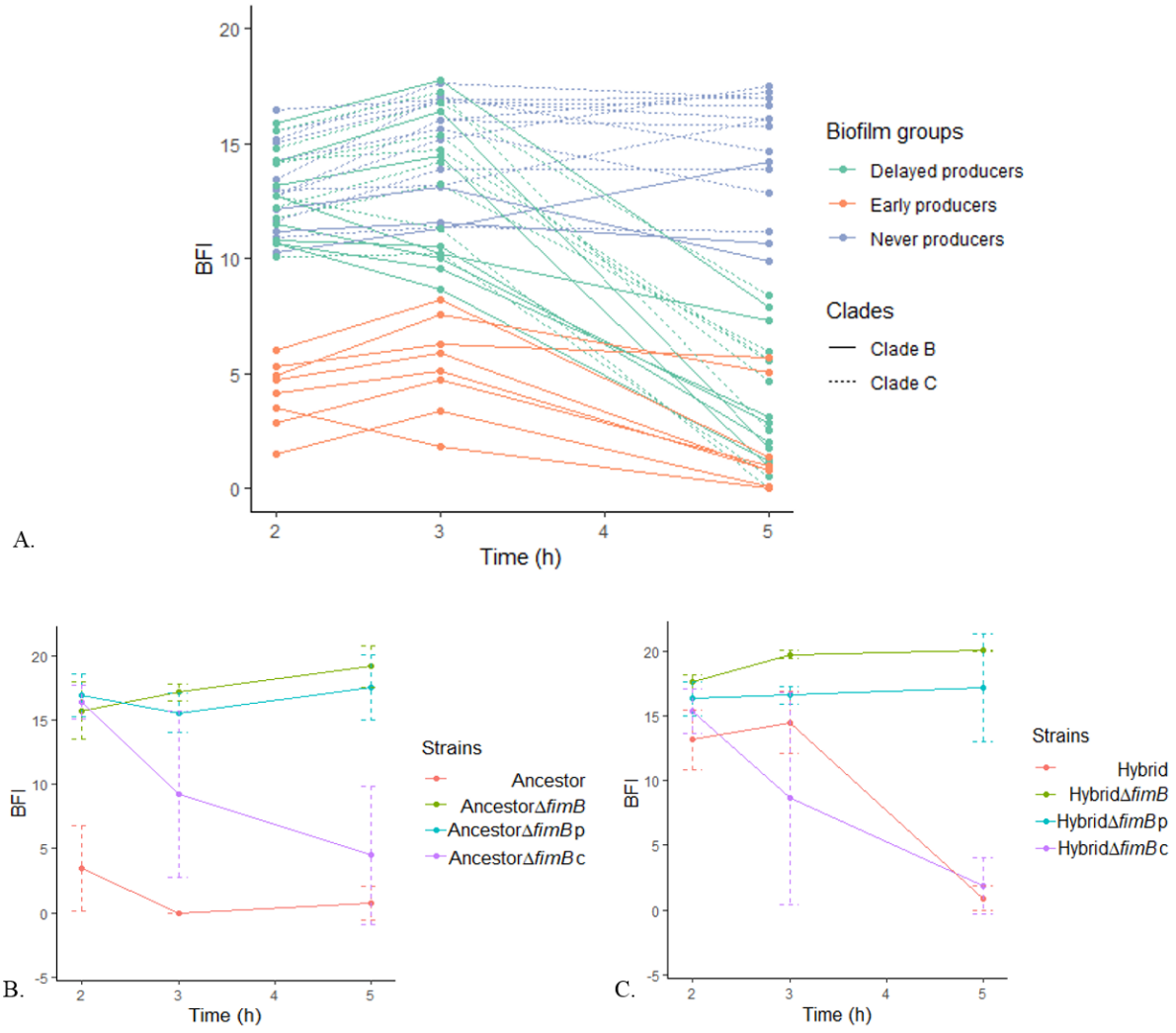


Figure 2. Early biofilm production. Biofilm Ring Test® was used to measure biofilm formation after 2, 3 and 5 h of static incubation. BFI values are inversely proportional to the biofilm formation capacity. A. Biofilm formation for the 39 *E. coli* ST131 strains. Clade B: continuous line; clade C: dashed line. Biofilm groups: blue lines: never producers; green lines: delayed producers; orange lines: early producers. B and C. Biofilm measurements for the *fimB* gene knockout experiments on Ancestor and Hybrid strains, respectively. Δ*fimB*: Δ*fimB*::; Δ*fimB*p: Δ*fimB*::pSC-A, strain complemented with an empty plasmid; Δ*fimB*c: Δ*fimB*::pSC-A-*fimB*, strain complemented with its wild type *fimB* gene.

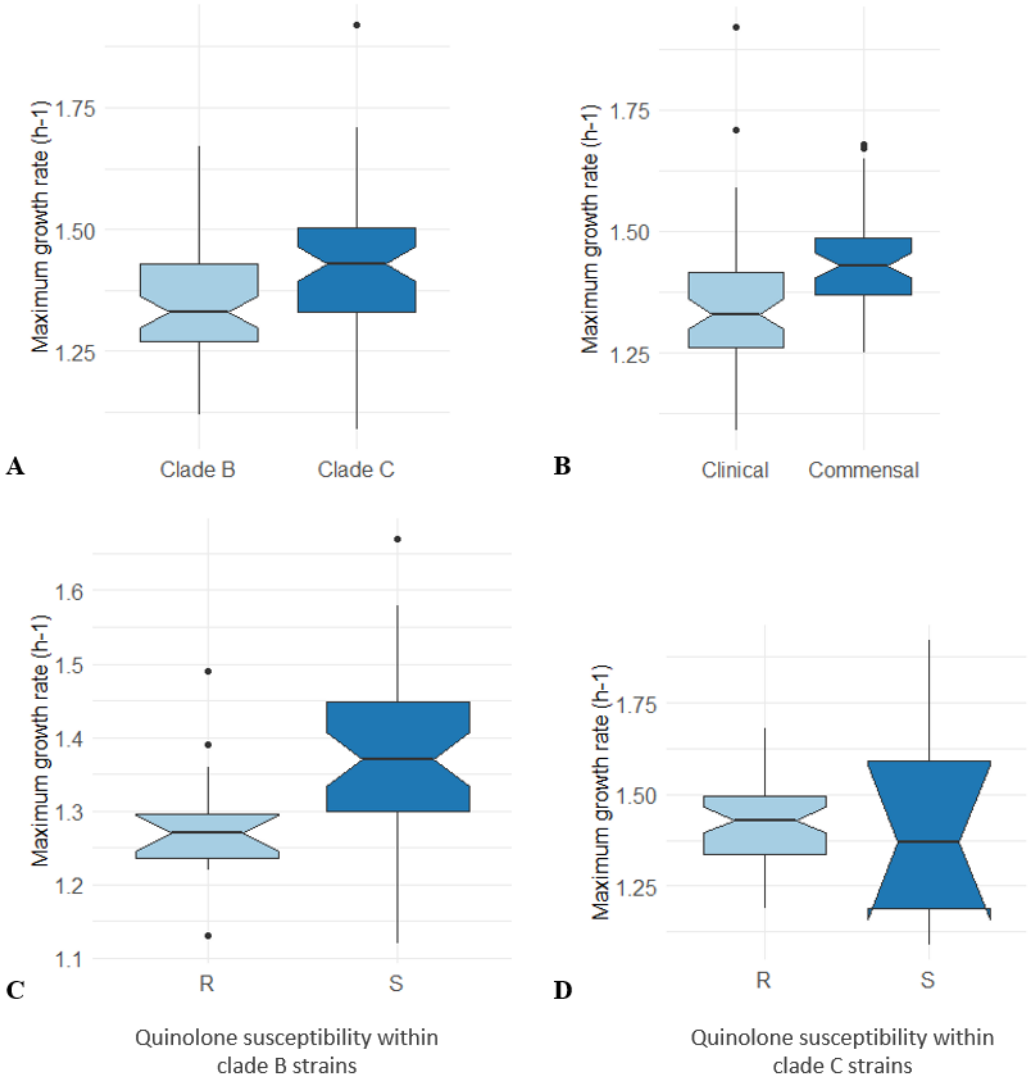


Figure 3. Maximum growth rate of the 39 *E. coli* strains in LB. Maximum growth rate is calculated from three independent culture assays and expressed in h^{-1}. Boxplots indicate maximum growth rate distribution across strains, and horizontal black bars indicate median values. The upper and lower ends of the box correspond to the upper and lower quartiles, respectively. Error bars represent the standard error of the mean of three experiments. Outliers are represented by dots. Strains were grouped into two categories: A. Clade B and clade C strains; B. Clinical and commensal (feces) strains; C. Nalidixic acid-resistant (R) and susceptible (S) strains within clade B; D. Nalidixic acid-resistant (R) and susceptible (S) strains within clade C. No significant difference was observed in each comparison (Wilcoxon test)

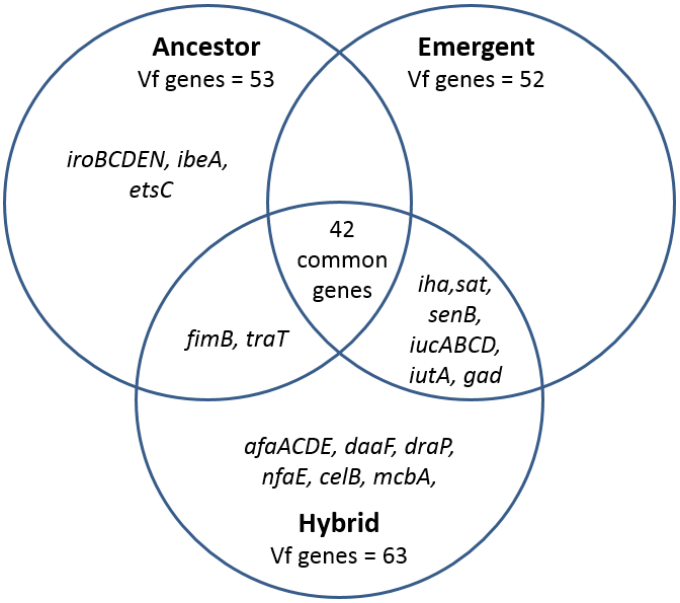
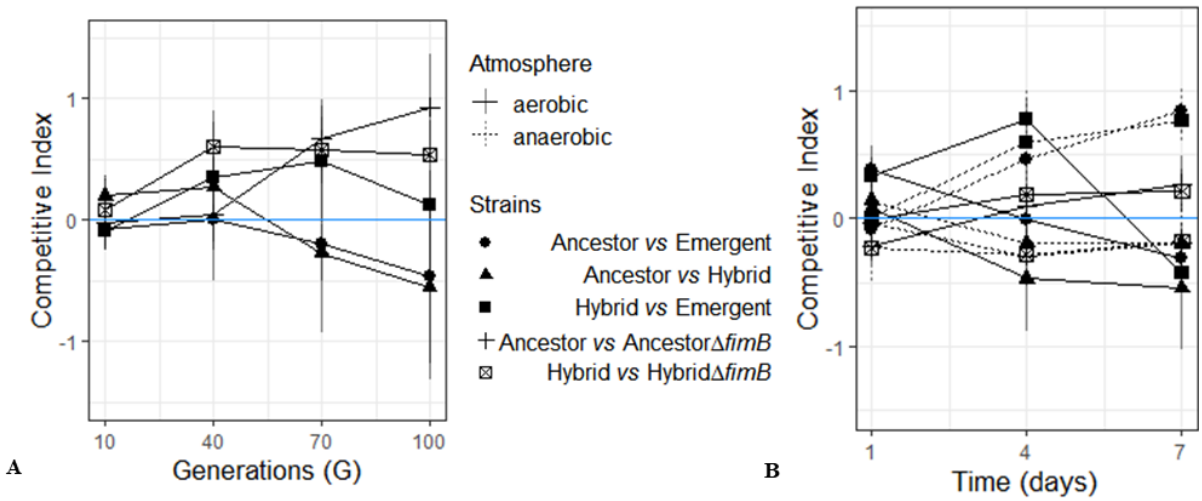


Figure 4. Virulence factors (VF)-encoding genes in Ancestor, Emergent and Hybrid. Each strain is represented by a circle, VF encoding genes specific of a strain are indicated in the unshared part of the circle, while genes in common are indicated in the intersecting region between the two or three involved strains. Among the 67 genes found at least once in these three strains, the 42 common genes are as follows: *fdeC*, *fimACDEFGHI*, *yfcV*, *yagVWXYZ* ,*ykgK*, *chuA*, *fyuA*, *entABCEFS*, *fepABCDG*, *fes*, *irp2*, *kpsDEMMTII*, *K5*, *iss*, *malX*, *usp*, *ompT*, *ompA*, *aslA*.



841

842 **Figure 5. *In vitro* competitions.** Dots shapes indicate the couple of strains in competition in
843 panels A and B. Competitive index (CI) is expressed in log, and CI above zero (blue line) means
844 that isolate 1 outcompetes isolate 2, and conversely. An interval of ±1 log was considered as
845 the limit of discrimination. Error bars represent the standard error of the mean of a least three
846 experiments. A. Competitions at an initial ratio of 1:1 in planktonic conditions in LB medium
847 during 100 generations. B. Competitions at an initial ratio of 1:1 in biofilm conditions (bacteria
848 spotted on BHI agar plate) during 7 days, in aerobic (continuous line) and anaerobic atmosphere
849 (dashed line).

850

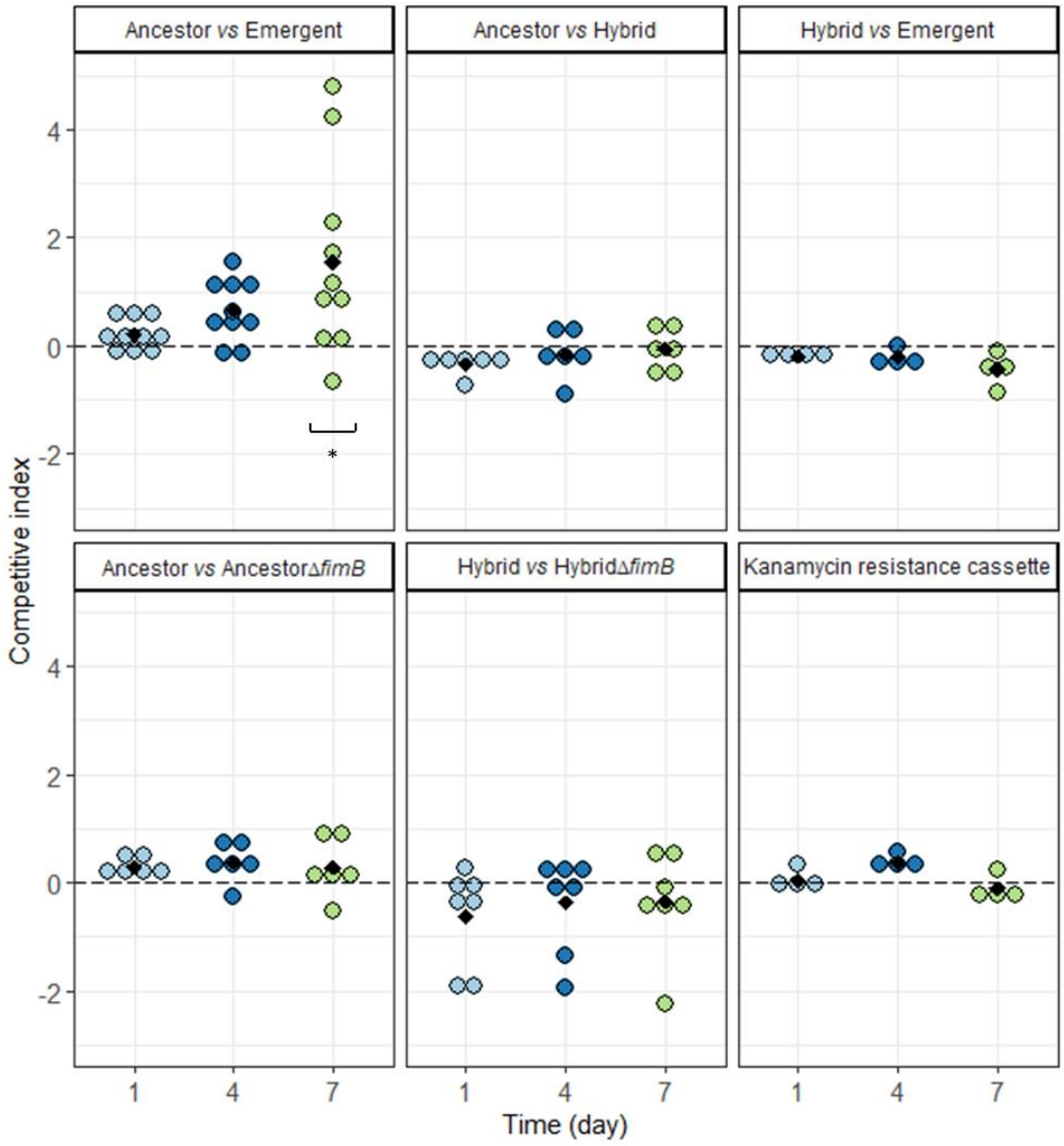
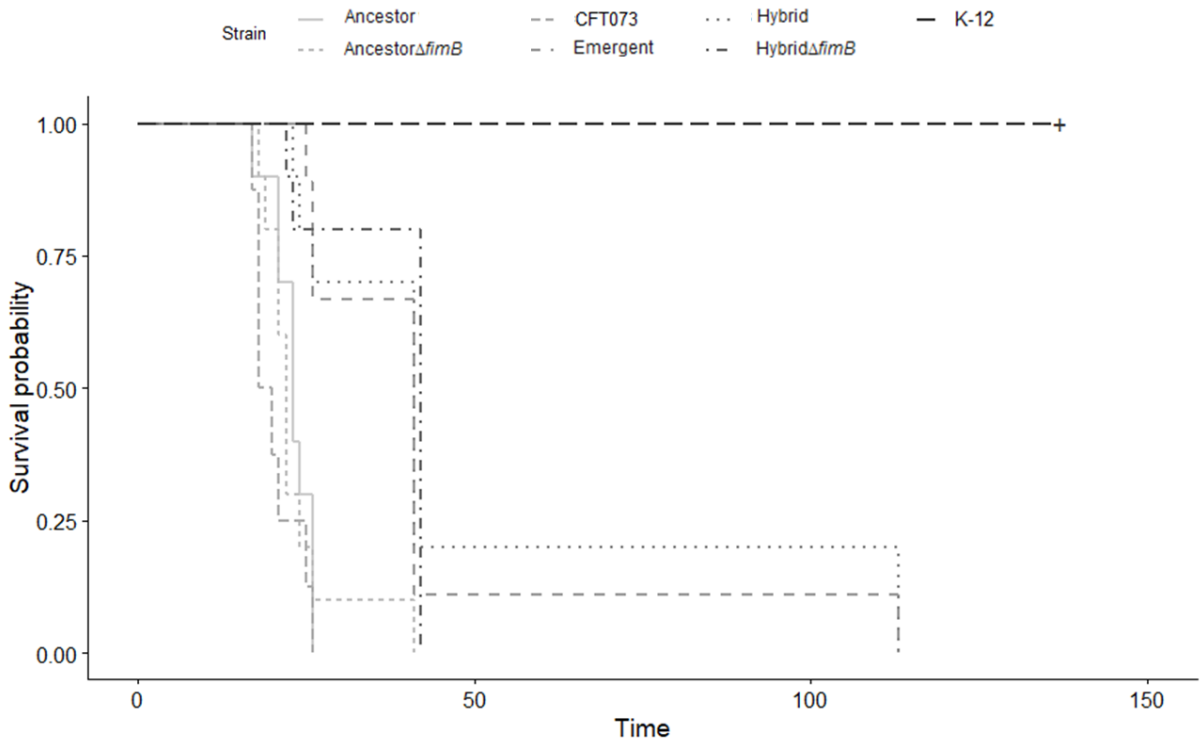


Figure 6. Competition in an intestinal colonization mouse model. Competitive index (CI) is expressed in log, and CI above zero means that isolate 1 outcompetes isolate 2, and conversely. An interval of ±1 log was considered as the limit of discrimination. Each dot represents a mouse at given times: light blue: day 1, dark blue: day 4 and green: day 7 post-inoculation. Black diamonds depict mean values of CI. Kanamycin resistance cassette: control competition evaluating the cost of the Kanamycin resistance cassette, with AncestorΔ*fimB*:: or HybridΔ*fimB*:: as isolate 1, and AncestorΔ*fimB*::Kana or HybridΔ*fimB*::Kana as isolate 2, respectively. * P=0.02, generated by the Wilcoxon signed-rank test and corrected for multiple comparisons by the Benjamini-Hochberg procedure.



862

863 **Figure 7. Kaplan-Meier survival curves of mice injected with *E. coli* ST131 strains.** Strains

864 are represented by line types.

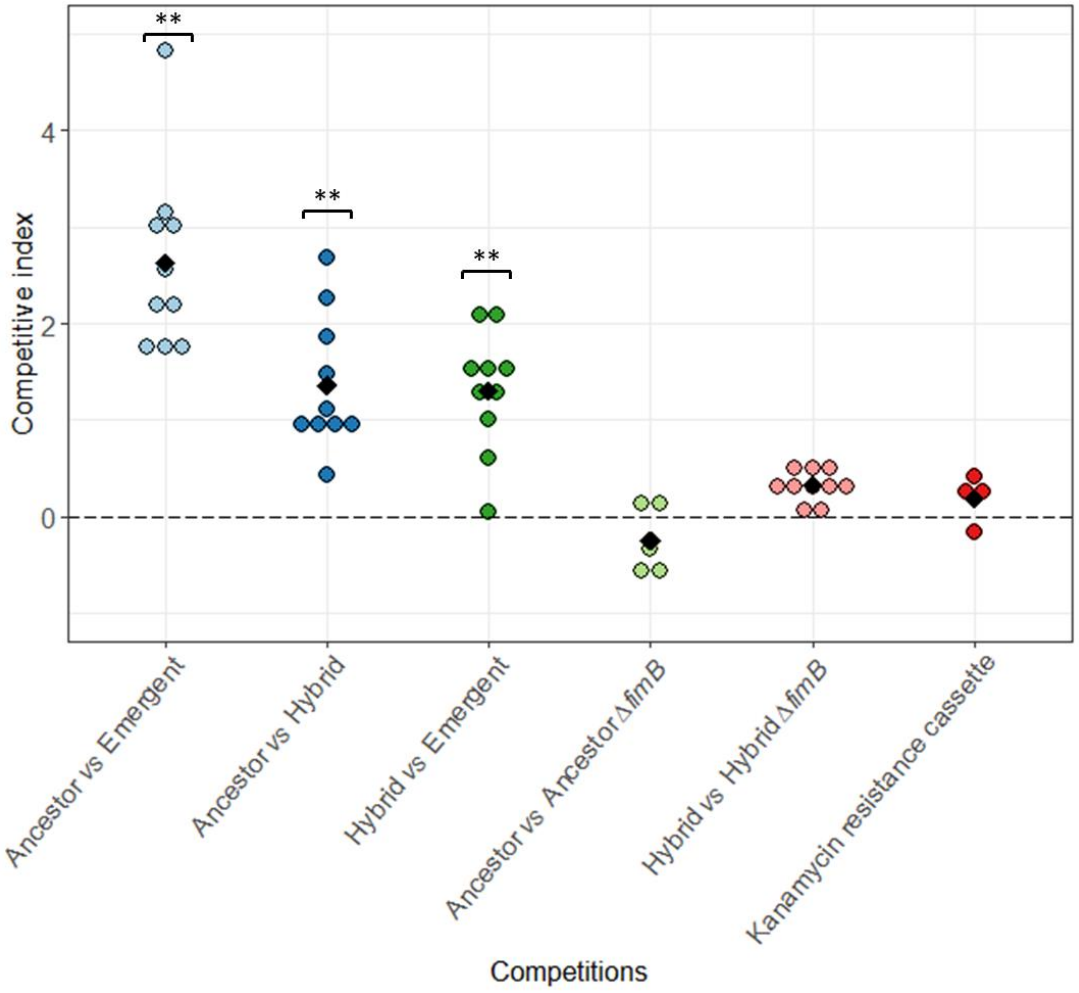
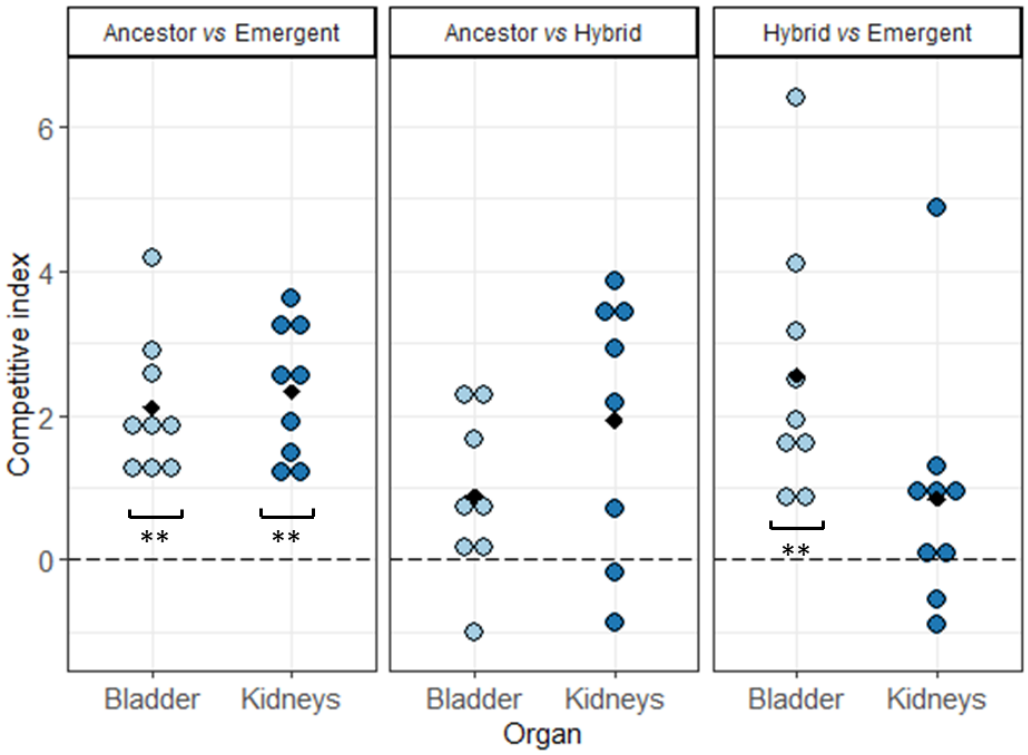


Figure 8. Competition in septicemia mouse model determined by spleen bacterial load. Competitive index (CI) is expressed in log, and CI above zero means that isolate 1 outcompetes isolate 2, and conversely. An interval of ±1 log was considered as the limit of discrimination. Kanamycin resistance cassette: control competition evaluating the cost of the Kanamycin resistance cassette, with Ancestor Δ*fimB*:: or Hybrid Δ*fimB*:: as isolate 1, and Ancestor Δ*fimB*::Kana or Hybrid Δ*fimB*::Kana as isolate 2 respectively. ** P=0.002, generated by the Wilcoxon signed-rank test.



873

874 **Figure 9. Competition in urinary tract infection mouse model determined by bladder and**
875 **kidneys bacterial load.** Competitive index (CI) is expressed in log, and CI above zero means
876 that isolate 1 outcompetes isolate 2, and conversely. An interval of ±1 log was considered as
877 the limit of discrimination. ** P<0.01, generated by the Wilcoxon signed-rank test and
878 corrected for multiple comparisons by the Benjamini-Hochberg procedure.

879

880

881

Supplemental material legend

Text S1. Technical details of the Materials and Methods section

Table S1. Strain collection: characteristics known at collection time

Table S2. Genome sequence quality

Table S3. Primers used in this study

Table S4. Plasmids used in this study

Table S5. Genotypic and phenotypic analyses of strain collection

Table S6. Main virulence factor (VF)-encoding genes

Figure S1. **Multiple correspondence analyses (MCA) of the 39 *E. coli* ST131 strains.** (A) The variables are as follows: the sources (fecal or clinical sample), the *fimH* allele (*H*22 or *H*30), fluoroquinolone (FQ) susceptibility (FQ_S: susceptible to FQ or FQ_R: resistant to FQ), the ESBL production (ESBL_P: ESBL producer or ESBL_N: non-ESBL producer), the biofilm production (early, delayed or never). cos2: square cosine (B) Plane projections on the Dim1-Dim2 of the variables and the strains. Strains or groups of strains are represented by colored circles: clade B in red and clade C in blue; circle size varies with the number of strain represented; variables are represented by black triangles. Locations of the three chosen strains, Ancestor, Hybrid and Emergent are pointed.

Figure S2. Maximum likelihood phylogenetic tree. This tree was built from non-recombinant SNPs using maximum likelihood, rooted on a clade A ST131 reference strain, SE15. Scale bar represents the average nucleotide substitutions per site. Branches are colored in accordance with subclades: B1 in yellow, B2 in pale orange, B3 in dark orange, B4 in red, B5 in pink, B0 in purple, C0 in pale blue, C1 in light green, C1-M27 in dark blue and C2 in dark green. Strains are named according to their origin : BZ for Ben Zakour *et al.*(5), K for Kallonen *et al.*(9) and D for this work.

907 **Figure S3. Schematic representation of the *fim* operon in a clade B strain, a clade C**
908 **strains and Hybrid strain of subclade B4.** Genes are depicted by arrows except for *fimS* that
909 is an invertible element, and the dotted lines represent ISEc55: the insertion sequence present
910 within the *fimB* gene of all clade C strains. A. Comparison between a representative clade B
911 strain and Hybrid . B. Comparison between Hybrid and a representative clade C strain. SNP:
912 single nucleotide polymorphisms and their effect: non-synonymous and synonymous. The
913 additional number indicates the presence of gap.