Two levels of specialization in bacteraemic *Escherichia coli* strains revealed by their comparison with commensal strains

O. CLERMONT[1,2], C. COUFFIGNAL[1,2,3], J. BLANCO[4], F. MENTRÉ[1,2,3], B. PICARD[1,5], E. DENAMUR[1,2,3]* and the COLIVILLE and COLIBAFI groups†

[1] *INSERM, IAME, UMR1137, Paris, France*
[2] *Université Paris Diderot, IAME, UMR1137, Sorbonne Paris Cité, Paris, France*
[3] *APHP, Hôpitaux Universitaires Paris Nord Val-de-Seine, Site Bichat Claude-Bernard, Paris, France*
[4] *Laboratorio de Referencia de E. coli (LREC), Departamento de Microbioloxía e Parasitoloxía, Facultade de Veterinaria, Universidade de Santiago de Compostela (USC), Lugo, Spain*
[5] *Université Paris Nord, IAME, UMR1137, Sorbonne Paris Cité, Bobigny, France*

SUMMARY

Bacteraemia caused by *Escherichia coli* are particularly frequent and severe, contrasting with the commensal character of the strains found in the digestive tract. A better understanding of the relationships between strains of both origins is needed to unravel the pathogenesis of this disease. Two hundred and forty-three commensal strains were compared to 243 bacteraemic strains isolated from adult hosts matched in terms of gender and age, and from similar location and epoch. Phylogenetic grouping, O-type determination, virulence factor content and antibiotic resistance were compared. Compared to commensal strains, the bacteraemic strains were characterized by a higher proportion of B2, C and D phylogroups, and a lower proportion of A, B1, E and F phylogroups. They also had a lower proportion of the B2 subgroup IV (STc141), a higher proportion of virulence factors, and a higher frequency of antibiotic resistance. These differences were more marked for the bacteraemic strains of urinary tract origin with the presence of specific clones, whereas the bacteraemic strains of digestive origin remained non-significantly different from the commensal strains, except for their antibiotic resistance. Thus, two levels of specialization from commensal strains were demonstrated in the bacteraemic strains: resistance to antibiotics in all cases, and virulence for those of urinary tract origin.

Key words: Antibiotic resistance, bacteraemia, commensal, *Escherichia coli*, phylogenetic group, portal of entry, virulence factor.

INTRODUCTION

Bacteraemia caused by *Escherichia coli* are particularly frequent and severe, with a case-fatality rate of 5–30% [1, 2]. Furthermore, the spread in recent years of isolates producing extended-spectrum *β*-lactamases (ESBLs) and carbapenemases that are often resistant to most of the available antibiotic classes have worsened the clinical impact of *E. coli* bacteraemia [3–5]. In contrast to this, *E. coli* remains the most prevalent aerobic bacteria of the human digestive tract [6], reaching a density of 10^8 colony-forming units per gram of faeces [7].

* Author for correspondence: Professor E. Denamur, IAME, UMR1137, INSERM, Université Paris Diderot, Site Xavier Bichat, 16, rue Henri Huchard, 75018 Paris, France.
(Email: erick.denamur@inserm.fr)
† Members of the COLIVILLE and COLIBAFI groups are given in the Appendix.

It is commonly accepted that the digestive tract is the reservoir of extra-intestinal pathogenic *E. coli* (ExPEC) strains [8–10], as well as of *E. coli* resistant strains [11]. In contrast to the uniqueness of this reservoir, *E. coli* bacteraemia is characterized by the variability of its pathophysiological processes, which determine the severity of disease [2]. Of these, two portals of entry are prevalent: the urinary tract during the course of pyelonephritis, prostatitis or renal abscess, and the digestive tract by translocation and various intra-abdominal infections [2].

Several epidemiological studies comparing pathogenic and commensal *E. coli* strains have documented that urinary tract infection (UTI) strains, especially those isolated from pyelonephritis, belong to particular clones characterized by specific phylogenetic backgrounds, a high content of virulence factors (VFs) and a few O-types, arguing for a specialization of these clones in the high diversity of commensal strains that exhibit a low level of VFs [12–20]. Less data are available for non-urinary portal-of-entry bacteraemic strains, especially from those with a digestive portal of entry [21–23]. Furthermore, these latter studies are based on a relatively small series of strains, and with one exception [23], the patients and control subjects used for the commensal strains were not carefully matched.

To gain insight into the pathogenesis of *E. coli* bacteraemia, we performed a large epidemiological study comparing 243 commensal strains to 243 bacteraemic strains isolated from cases with a variety of portal of entry, including both urinary tract and digestive tract, from adult hosts matched in terms of gender and age, and from similar location and epoch. We demonstrate that the strains isolated from bacteraemia are more resistant to antibiotics than the commensal intestinal strains. Furthermore, the bacteraemic strains of urinary tract origin are a subset of the diversity of the commensal intestinal strains that has been selected for, whereas the strains isolated from bacteraemia of digestive origin encompass the same diversity as the commensal strains.

MATERIALS AND METHODS

Strain collections

Two previously published collections of *E. coli sensu stricto* were used. One is the COLIVILLE collection of 243 *E. coli* strains isolated from the stools of 243 community adult subjects living in the Paris area in 2010 [24], which was thereafter named 'commensal collection' (approved by the ethics evaluation committee of Institut National de la Santé et de la Recherche Médicale; CCTIRS no. 09·243). These can be considered as 'true' commensal strains, as strict inclusion criteria were applied to selection of the subjects, excluding individuals who had undergone any antibiotic therapy in the preceding month and any hospitalization in the 3 months preceding inclusion. In order to avoid any epidemiological bias, the recruited subjects belonged to independent families. The other is the COLIBAFI collection of 373 *E. coli* clinical strains isolated from the blood of 373 patients during the course of *E. coli* bacteraemia in the Paris area in 2005 [2], which was thereafter named 'clinical collection' (approved by the ethics evaluation committee of the Hôpital Saint Louis, Paris, France; no. 2004–06). The patients included were from seven distinct hospitals and a large diversity of wards. The portal of entry of these bacteraemia was established according to the compatible clinical and/or radiographic features and the isolation of *E. coli* from the presumed source of infection. When *E. coli* could not be isolated from the presumed portal of entry, this latter was assigned on the basis of firm clinical suspicion, provided that all other possible sources of infection had been excluded. If the clinical data were ambiguous, the portal of entry was categorized as 'undetermined' [2].

As it is known that *E. coli* epidemiology is related to the age and the sex of the subjects [25], each strain of the commensal collection was paired with a clinical collection strain ($n = 243$). Matching with a 1:1 ratio has been done on gender and minimizing age, with a median age difference of 1 year and an interquartile range (IQR) of 0–9 years. Moreover, two subsets according to the urinary ($n = 138$) and digestive ($n = 60$) portal of entry for the *E. coli* strains of the clinical collection were assessed with their pairs of the commensal collection.

Strain characterization

A successive refinement strain-typing strategy was applied on the two collections from the seven main phylogenetic groupings (A, B1, B2, C, D, E, F) [26] to the characterization of particular clones of medical and epidemiological importance by a combination of PCR-based methods [27–29] and O-typing [30, 31]. Thus, the B1 clonal complex 87 (CC87) (Institut Pasteur MLST schema nomenclature corresponding to ST58 and ST155 in the Achtman schema [27]),

the 10 main B2 subgroups and the clonal group A (CGA) from the D phylogroup were determined in the strains. The exhaustive correspondence between this typing approach and STc membership according to the currently used Achtman MLST schema [32] is available in Clermont *et al.* [33].

The presence of 19 VFs representing the four main classes for *E. coli* extraintestinal virulence, including adhesins/invasin, iron capture systems, toxins and protectins was determined as in previous studies [31, 34]. The *ibeA* gene is implicated in invasion but also indirectly in adhesion [35].

The resistance to seven antibiotics was determined according to the recommendations of the Antimicrobial Committee of the French Society for Microbiology (http://www.sfm-microbiologie.org) [amoxicillin, amoxicillin-clavulanic acid, cefoxitin, cefotaxime, amikacin, ofloxacin, cotrimoxazole (sulfamethoxazole-trimethoprim)].

The bacterial characteristics listed above were already available for the strains of the commensal collection [24], except for CC87 membership. For the strains of the clinical collection, membership in the seven main *E. coli* phylogroups, membership in the B1, B2 and D main clonal complexes, presence of the *clbQ* VF-coding gene and O-antigen determination were performed in this study, with the remaining characters being already available [2].

Statistical analyses

The bacterial characteristics described above were compared between the commensal and clinical collections. The same characteristics were compared in the specific clonal complexes/subgroups of the B2 and D phylogroups. A comparison was also performed according to the portal of entry (urinary- and digestive-source bacteraemia) for the *E. coli* strains of the clinical collection with their paired *E. coli* strains of the commensal collection.

Comparisons between the two collections were performed using non-parametric tests (Wilcoxon rank-sum test or Fisher's exact test). As multiple tests were performed, the *P* values were adjusted using the Benjamini & Hochberg method [36] for each type of characteristic (phylogroup, VF, antibioresistance). For the phylogroup characteristics, if significant comparisons were found within all and each subgroup strains, the multiple test correction was performed at two levels: phylogenetic groups and subgroups. Analyses were performed with SAS

v. 9.4 (SAS Institute Inc., USA). Data are shown as medians with IQR for continuous variables except when otherwise stated, and number/percentage for categorical variables. All tests were two-sided with a type-I error fixed to 0·05.

RESULTS

Host characteristics

Two hundred and forty-three subjects and 243 patients were included in this study from the commensal and clinical cohorts, respectively. The sex ratio was 0·83 in each cohort, corresponding to 110 men and 133 women for both commensal and clinical cohorts. The mean ages were 56·4 years [±standard deviation (s.d.) 11·3, range 21–86] and 59·0 years (±s.d. 15·1, range 20–86) for the commensal and clinical cohorts, respectively. The portal of entry of the bacteraemia episodes in the clinical collection were urinary tract (138, 56·8%), digestive (60, 24·7%) and miscellaneous and undetermined (45, 18·5%).

Global comparison of the two collections of strains, i.e. bacteraemic *vs.* commensal strains

We first genotyped the strains of both collections at the phylogroup and clonal complex levels (Table 1). When the clinical collection strain phylogenetic distribution was compared to that of the commensal collection, clear differences were observed for all phylogroups. Phylogroups B2, C and D were over-represented in the clinical strains relative to the commensal strains, whereas the A, B1, E and F phylogroup strains were under-represented. Of note, within the B2 phylogroup strains, only subgroup IV (STc141) strains were under-represented in the clinical strains. Within the B2 phylogroup, we had more thoroughly characterized the strains at the clonal level by a combination of subgroup and O-type, as proposed previously [37, 38] (i.e. I-O25b) (Table 2, descriptive data). Some clones (I-O25b, VI-O4, VII-O23, IX-O1, IX-O45a) appeared to be over-represented in the clinical strains, whereas the IV-O2b clone was under-represented, and not a single VIII-O81 strain was isolated from the clinical samples. The proportion of D phylogroup strains that were CGA, an anti-microbial-resistant extra-intestinal pathogenic clonal complex, was not significantly different between the two collections, as well as the proportion of B1 phylogroup strains that were CC87, a newly reported

Table 1. *Phylogenic group and subgroup distribution in* E. coli *strains of the commensal and clinical collections*

Phylogenic groups/subgroups*	All strains				Matched commensal and urinary-source strains				Matched commensal and digestive-source strains†‡	
	n (%)		Adjusted P value for groups†	Adjusted P value for B2 subgroups‡	n (%)		Adjusted P value for groups†	Adjusted P value for B2 subgroups‡	n (%)	
	Commensal (n = 243)	Clinical (n = 243)			Commensal (n = 138)	Clinical (n = 138)			Commensal (n = 60)	Clinical (n = 60)
A	68 (28·0)	33 (13·6)	<0·001		41 (29·7)	7 (5·1)	<0·001		15 (25·0)	19 (31·7)
B1	31 (12·8)	18 (7·4)	0·07		16 (11·6)	7 (5·1)	0·24		9 (15·0)	7 (11·7)
CC87§	7 (19·4)	7 (36·9)			4 (20·0)	3 (33·3)			3 (33·3)	2 (20·0)
B2	84 (34·6)	129 (53·1)	<0·001		47 (34·1)	89 (64·5)	<0·001		21 (35·0)	22 (36·7)
I	12 (14·3)	21 (16·3)		0·85	7 (14·9)	8 (9·0)		0·45	3 (14·3)	6 (27·3)
II	18 (22·4)	32 (24·8)		0·85	9 (19·2)	23 (25·8)		0·45	5 (23·8)	4 (18·2)
IV	17 (20·2)	4 (3·1)		<0·001	10 (21·3)	1 (1·1)		<0·001	4 (19·1)	3 (13·6)
IX	15 (17·9)	34 (26·4)		0·73	11 (23·4)	29 (32·6)		0·45	2 (9·5)	3 (13·6)
Other subgroups	22 (26·2)	38 (29·5)		0·85	10 (21·3)	28 (31·5)		0·45	7 (33·3)	6 (27·3)
C	7 (2·9)	19 (7·8)	0·07		6 (4·4)	10 (7·3)	0·44		1 (1·7)	3 (5·0)
D	21 (8·6)	35 (14·4)	0·07		11 (8·0)	23 (16·7)	0·17		8 (13·3)	5 (8·3)
CGA‖	11 (52·4)	19 (54·3)			5 (45·5)	15 (65·2)			4 (50·0)	2 (40·0)
E	7 (2·9)	0 (0)	0·06		4 (2·9)	0 (0·0)	0·24		0 (0·0)	0 (0·0)
F	25 (10·3)	9 (3·7)	0·04		13 (9·4)	2 (1·5)	0·03		6 (10·0)	4 (6·7)

* The correspondence with the Achtman MLST schema [32] is as follow: CC87 = STc155; B2 subgroup I = STc131; B2 subgroup II = STc73; B2 subgroup IV = STc141; B2 subgroup IX = STc95; CGA = STc69 [33].
† Global *P* values for the comparison of all phylogroups between the two collections: all strains $P < 0.001$; matched urinary source $P < 0.001$; matched digestive source $P = 0.73$.
‡ Global *P* values for the comparison of B2 subgroups between the two collections: all strains $P = 0.002$; matched urinary source $P < 0.001$; matched digestive source $P = 0.83$.
§ Non-adjusted *P* value for the comparison of CC87 subgroup between the two collections: all strains $P = 0.32$; matched urinary source $P = 0.64$; not done for matched digestive source.
‖ Non-adjusted *P* value for the comparison of CGA subgroup between the two collections: all strains $P = 0.89$; matched urinary source $P = 0.27$; not done for matched digestive source.

Table 2. *Characterization of the main B2 clones using a combination of subgroup and O-type in* E. coli *strains of the commensal and clinical collections*

	All strains		Matched commensal and urinary source strains	
Clone ID*	Commensal (n = 84)	Clinical (n = 129)	Commensal (n = 47)	Clinical (n = 89)
I-O25b	3 (25)†	14 (62)	2	5
I-O6a	2 (17)	4 (19)	2	2
I-O9	1 (8)	0	0	0
II-O18	2 (11)	0	1	0
II-O22	1 (6)	3 (9)	0	1
II-O2a	0	3 (9)	0	3
II-O6a	13 (72)	24 (75)	6	19
III-O6a	2 (100)	5 (83)	1	4
IV-O2b	11 (65)	3 (75)	7	1
IV-O46	2 (12)	0	1	0
VI-O4	1 (100)	10 (91)	0	8
VII-O23	0	4 (40)	0	4
VII-O75	5 (83)	6 (60)	2	4
VIII-O81	5 (100)	0	3	0
IX-O1	6 (40)	19 (56)	3	16
IX-O18	4 (27)	3 (9)	4	1
IX-O2a	3 (20)	3 (9)	2	3
IX-O45a	0	7 (21)	0	7

* The roman number corresponds to the B2 subgroup. The correspondence with the Achtman MLST schema [32] is as follow: subgroup I = STc131; subgroup II = STc73; subgroup III = STc127; subgroup IV = STc141; subgroup VI = STc12; subgroup VII = STc14; subgroup VIII = STc452; subgroup IX = STc95 [33].

† The numbers in parentheses correspond to the percentages of each clone within the corresponding subgroup.

antimicrobial-resistant clonal complex of animal origin spreading in humans [27] (Table 1).

We next analysed the presence of 19 VFs in the two collections (Table 3). We observed a highly significant over-representation of the VFs involved in iron acquisition (all except *ireA*) in the clinical strains. This over-representation of VFs was less substantial for adhesins, protectins and toxins. Markedly, the *ibeA* invasin/adhesin coding gene was the only VF under-represented in the clinical strains (Table 3). Globally, the VF score was significantly higher in the clinical strains (8, IQR 5–11) than in the commensal strains (6, IQR 2–9). As it is well known that B2 strains are enriched in VFs [39], we then compared the VF proportions within the B2 strains (Table 3). It appeared that the over-representation of VFs in the clinical strains was less pronounced for the iron

acquisition VFs and adhesins, and disappeared for the protectins and toxins. The virulence score was similar [11 (IQR 8–13) and 11 (IQR 9–14)] for the commensal and clinical collections of the B2 subgroups, respectively (Table 3).

Finally, we compared the antibiotic resistance of the two collections (Fig. 1). For all the studied antibiotics, we observed a higher frequency of resistance in the clinical strains than in the commensal strains.

Comparisons of commensal *vs.* urinary portal of entry and *vs.* digestive portal of entry bacteraemic strains

The two main portals of entry of bacteraemia are the urinary tract and the digestive tract. We therefore undertook a comparison according to the portal of entry for the *E. coli* strains of the clinical collection (urinary tract = 138 strains, digestive tract = 60 strains) with their paired *E. coli* strains from the commensal collection.

When comparing the phylogenetic distribution, the differences observed between the strains of the two collections were even more marked for the urinary tract origin bacteraemia, with B2 phylogroup strains reaching 64·5% and A phylogroup strains decreasing to 5·1% (Table 1). Similarly, within the B2 phylogroup, clones I-O25b, II-O6a, VI-O4, IX-O1 and IX-O45a were over-represented in the bacteraemic strains of urinary origin (Table 2, descriptive data). In contrast, no significant difference was observed in the representation of the different phylogroups and subgroups between the commensal strains and bacteraemic strains of digestive origin (Table 1).

The comparison of the content of 19 VFs yielded similar trends to those observed for the phylogroup distribution. All five VFs involved in iron acquisition two adhesins, two protectins and two toxins were over-represented in urinary tract origin bacteraemia strains, whereas the *ibeA* gene was under-represented (Fig. 2a). Globally, the VF score of the urinary clinical strains was 11 (IQR 7–12), *vs.* 5 (IQR 2–9) (P = 0·006) for commensal strains. When comparing only B2 phylogenetic group strains, the over-representation of VFs was overall less pronounced, with a virulence score of 11 (IQR 8–12) and 11 (IQR 11–14) (P < 0·01) in the commensal and urinary clinical strains, respectively. The protectin, toxin and *iroN* gene content was not found to be differentially represented (Supplementary Table S1). Contrasting sharply with the strains of urinary tract origin, the clinical strains of digestive origin were not found to be different from the commensal

Table 3. *Virulence factors detected in* E. coli *strains of the commensal and clinical collections for all strains (n = 243 in each collection) and in the B2 phylogroup strains (n = 84 and n = 129 in each collection, respectively)*

Virulence traits	Function	All strains			B2 phylogroup strains		
		n (%)*			*n* (%)*		
		Commensal (*n* = 243)	Clinical (*n* = 243)	Adjusted *P* value†	Commensal (*n* = 84)	Clinical (*n* = 129)	Adjusted *P* value†
iha	Adhesin	83 (34·2)	88 (36·2)	0·75	30 (35·7)	60 (46·5)	0·77
papC	Adhesin	59 (24·3)	124 (51)	<0·01	38 (45·2)	91 (70·5)	<0·01
hra	Adhesin	53 (21·8)	75 (30·9)	0·19	32 (38·1)	44 (34·1)	0·77
sfa/foc	Adhesin	46 (18·9)	59 (24·3)	0·75	41 (48·8)	57 (44·2)	0·77
papGII‡	Adhesin	28 (11·5)	97 (39·9)	<0·01	19 (22·6)	73 (56·6)	<0·01
ibeA	Invasin/Adhesin	27 (11·1)	11 (4·5)	0·07	26 (31)	9 (7)	<0·01
papGIII‡	Adhesin	20 (8·2)	22 (9·1)	0·75	20 (23·8)	22 (17·1)	0·77
irp2	Iron acquisition	153 (63)	199 (81·9)	<0·01	78 (92·9)	129 (100)	0·03
fyuA	Iron acquisition	151 (62·1)	199 (81·9)	<0·01	78 (92·9)	129 (100)	0·03
iucC	Iron acquisition	112 (46·1)	170 (70)	<0·01	49 (58·3)	100 (77·5)	0·04
iroN	Iron acquisition	85 (35)	135 (55·6)	<0·01	62 (73·8)	99 (76·7)	0·77
ireA	Iron acquisition	49 (20·2)	72 (29·6)	0·14	27 (32·1)	58 (45)	0·68
ompT	Protectin	143 (58·8)	172 (70·8)	0·07	77 (91·7)	116 (89·9)	0·77
traT	Protectin	120 (49·4)	155 (64)	0·01	45 (53·6)	86 (67·2)	0·55
neuC	Protectin	58 (23·9)	48 (19·8)	0·75	36 (42·9)	41 (31·8)	0·77
usp	Toxin	94 (38·7)	125 (51·4)	0·06	79 (94)	120 (93)	0·77
sat	Toxin	64 (26·3)	79 (32·5)	0·75	22 (26·2)	55 (42·6)	0·20
clbQ	Toxin	49 (20·2)	56 (23)	0·75	49 (58·3)	56 (43·4)	0·43
hlyC	Toxin	35 (14·4)	65 (26·7)	0·01	34 (40·5)	60 (46·5)	0·77
cnf1	Toxin	29 (11·9)	41 (16·9)	0·75	29 (34·5)	41 (31·8)	0·77
Virulence score, median [IQR]		6 [2–9]	8 [5–11]	<0·0001§	11 [8–13]	11 [9–14]	0·10§

* Except where otherwise stated.
† Adjusted *P* value for the comparison of the 20 virulence factors within each subgroup of strains.
‡ Distinct alleles of the *papG* virulence gene.
§ Wilcoxon rank-sum test.

strains. Globally, the VF scores were 6 (IQR 2–10) and 5 (IQR 2–8) (*P* = 0·68) for the digestive clinical and commensal strains, respectively. This held true whether individual factors of all the strains were analysed (Fig. 2*b*), or when only B2 group strains were taken into consideration (Supplementary Table S1).

The differences observed for antibiotic resistance prevalence between the commensal and clinical collections were found to be present in the same proportions between the commensal and the bacteraemic strains of urinary and digestive origins (Table S2), contrasting with the analogous finding regarding virulence gene content.

DISCUSSION

In this work, we compared commensal and bacteraemic *E. coli* strains to gain insight into the pathogenesis of *E. coli* bacteraemia. The strengths of our epidemiological study were (i) the size of the sampling

(243 commensal *vs.* 243 bacteraemic strains), allowing statistical analysis of sub-populations, (ii) the unrelated character of the strains within each collection, (iii) the 'true' commensal character of the commensal collection [24], (iv) the matching of the two populations of hosts in terms of sex and age, and (v) the homogeneity of epoch and geographical location of the study.

The first characteristic that sharply differentiated the populations of commensal and clinical strains was their acquired resistance to all the antibiotics tested, whatever the plasmidic or chromosomal origin of this resistance (Fig. 1) and despite the fact that the acquired resistance of *E. coli* increases with time [24] and that the strains of the clinical collection were recruited 5 years before those of the commensal collection. Moreover, this difference was observed whatever the portal of entry of the bacteria (Supplementary Table S2). This difference can be due to different
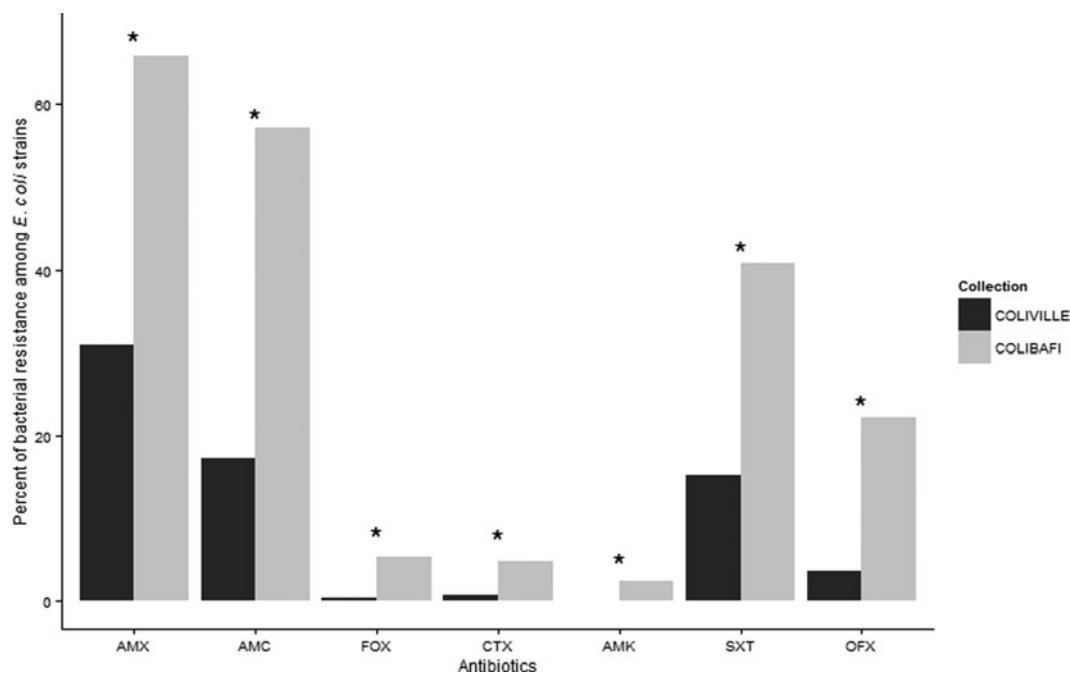
Fig. 1. Bacterial resistance to seven antibiotics in *E. coli* strains of the commensal (COLIVILLE) and clinical (COLIBAFI) collections for all strains (*n* = 243 in each collection). Histograms represent proportion of resistance to amoxicillin (AMX), amoxicillin-clavulanic acid (AMC), cefoxitin (FOX), cefoxatime (CTX), amikacin (AMK), cotrimoxazole (SXT) and ofloxacin (OFX). The asterisk (*) indicates a significant difference between the two collections.

ecological sources (commensal *vs.* clinical) or more probably to the antibiotic selective pressure applied on the microbiota of the patients or to the nosocomial contamination of the patients by resistant strains before the bacteraemic episode [40]. Thus, we evidenced 9·7% of patients receiving antibiotics within 2 weeks preceding the bacteraemia and 23·9% of nosocomial infections in the 243 COLIBAFI patients (Supplementary Table S3).

The second characteristic that distinguished the strains according to their commensal or clinical origin was their phylogenetic distribution and VF content, two traits previously demonstrated to be linked [39]. The newly described phylogroups E and F were under-represented in the clinical strains, whereas phylogroup C was over-represented (Table 1). This latter phylogroup encompasses strains virulent in a mouse model of sepsis, which was isolated during the course of a cluster of neonatal meningitis [41]. Moreover, this clonal complex is also prone to be antibiotic resistant [24]. An over-representation of strains belonging to the B2 and D phylogroups was noted in the clinical strains, as reported previously [23, 39].

Within phylogroup B2, the subgroup IV, corresponding to STc141 according to the Achtman schema [32], was under-represented in the clinical isolates, and no subgroup VIII (STc452) strain was identified in the clinical strains (Table 2). These particular subgroups could correspond to commensal clones with low level of human invasiveness. This hypothesis is corroborated by the fact that the strains of subgroup VIII are avirulent in the mouse model of sepsis [42]. The proportions of subgroups I (STc131), II (STc73) and IX (STc95) within the B2 phylogroup and of CGA (STc69) within the D phylogroup were similar in both commensal and clinical strains, despite that group B2 (and to a lesser extent group D) as a whole were more prevalent in the clinical isolates than the commensal isolates (Table 1). This is corroborated to the fact that these four STcs are also those most frequently isolated in extra-intestinal infections worldwide [43–45].

As expected, the proportion of VFs, especially those involved in iron acquisition, was higher in the clinical than commensal strains. However, this difference was less marked within the B2 phylogroup strains (Table 3). Of note, the strains of B2 subgroup IV often possessed the *ibeA* gene (data not shown), explaining its under-representation within the clinical strains.

The most marked feature of our study was that, except for antibiotic resistance, the observed differences between the characteristics of the commensal and bacteraemic strains depend on the portal of entry of
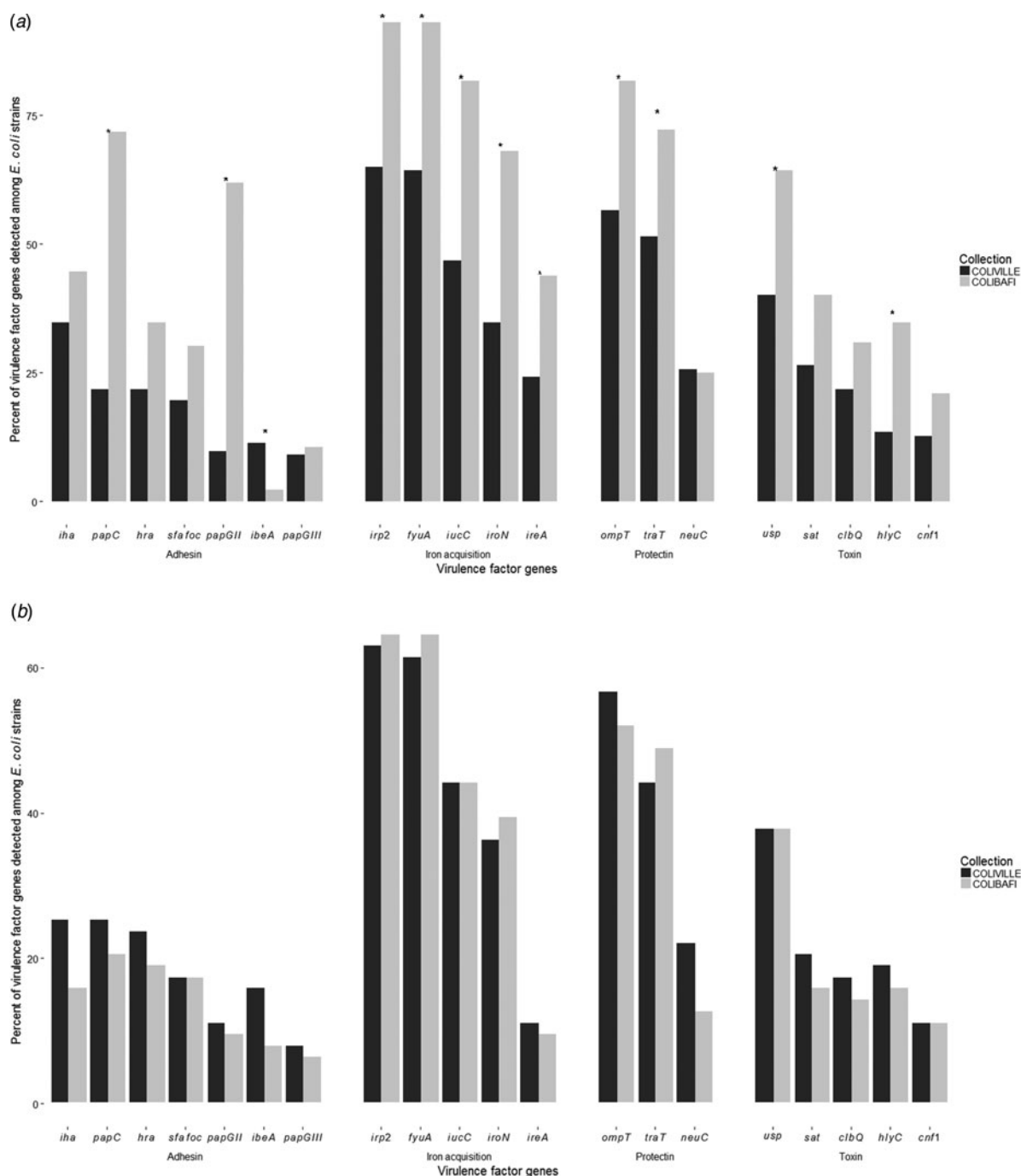
Fig. 2. Virulence factor genes detected in *E. coli* strains of the commensal (COLIVILLE) and clinical (COLIBAFI) collections according to the portal of entry of the bacteraemia. (*a*) Urinary portal of entry ($n = 138$ in each collection). (*b*) Digestive portal of entry ($n = 60$ in each collection). Histograms represent proportion of virulence factors grouped by function. The asterisk (*) indicates a significant difference between the two collections.

bacteraemia, i.e. urinary *vs.* digestive tract. Previous studies have identified differences in phylogenetic background and VF repertoire between urine-source *versus* other-source blood isolates [2, 21, 22]. A clear specialization of the strains isolated during a bacteraemia of urinary tract origin was revealed by a marked reduction

in prevalence of A phylogroup strains and an increase of specific phylogenetic group B2 clones (Table 2), exhibiting numerous VFs (Supplementary Table S1). These clones exhibiting O-specific antigens were previously found to be associated with UTI (O-UTI group) [46]. Conversely, no significant difference in terms of

phylogroup distribution or VF content was found between commensal and bacteraemic strains of digestive tract origin (Table 1 and Fig. 2b). Additionally, a relationship between the portal of entry and host characteristics was evidenced. The patients having a bacteraemia of digestive origin were more often males and with predisposing conditions such as solid cancer, tobacco addiction and immunosuppression than those having a bacteraemia of urinary origin (Supplementary Table S3).

Our data provide evidence for two distinct models [8] of pathophysiology in bacteraemia, according to the portal of entry and the host characteristics: (i) the special pathogenicity model in the case of urinary tract origin, implicating a subset of highly specialized strains selected from the intestinal tract which are able to cross several anatomical and physiological barriers and occurring mostly in females, and (ii) the prevalence model in the case of digestive origin, implying that more prevalent faecal *E. coli* clones are able to translocate into the bloodstream in hosts with predisposing conditions. This duality in pathogenesis is also linked to death at 28 days; with a urinary tract portal of entry and the presence of the *ireA* gene involved in iron capture being negatively correlated with death [2] (Supplementary Table S3).

This study had several limitations. First, despite the substantial size of the cohorts, their divisions into subgroups according to several variables resulted in the comparison of series of small size, which is not always statistically relevant. Second, we did not performed an exhaustive characterization of the clones, which could be refined by complete multilocus sequence typing and ultimately whole-genome sequencing. This fine-scale characterization of the strains will be the topic of a further study. Last, we did not focus our study on the relationship of host characteristics with bacterial characteristics and clinical outcomes.

CONCLUSIONS

By comparing bacteraemic to rigorously selected community commensal *E. coli* strains, we were able to demonstrate two levels of specialization in the bacteraemic strains: (i) an acquired resistance to antibiotics, which concerns the bacteraemic strains regardless of their portal of entry and (ii) an intrinsic virulence characterized by membership of specific clonal groups which are adapted to the urinary portal of entry. These epidemiological data provided additional evidence for niche specialization within the *E. coli* species.

APPENDIX

The COLIVILLE group is composed of the following medical practitioners who were involved in the recruitment of subjects for the study: Monique Allouche, Jean-Pierre Aubert, Isabelle Aubin, Ghislaine Audran, Dan Baruch, Philippe Birembaux, Max Budowski, Emilie Chemla, Alain Eddi, Marc Frarier, Eric Galam, Julien Gelly, Serge Joly, Jean-François Millet, Michel Nougairede, Nadja Pillon, Guy Septavaux, Catherine Szwebel, Philippe Vellard, Raymond Wakim, Xavier Watelet and Philippe Zerr.

The COLIBAFI group includes the following individuals. Clinical investigators are Michel Wolff, Loubna Alavoine, Xavier Duval, David Skurnik, Paul-Louis Woerther, Antoine Andremont (CHU Bichat-Claude-Bernard, Paris); Etienne Carbonnelle, Olivier Lortholary, Xavier Nassif (CHU Necker-Enfants Malades, Paris); Sophie Abgrall, Françoise Jaureguy, Bertrand Picard (CHU Avicenne, Bobigny); Véronique Houdouin, Yannick Aujard, Stéphane Bonacorsi, Edouard Bingen, Chloé Lemaitre, Romain Basmaci (CHU Robert-Debré, Paris); Agnès Meybeck, Guilène Barnaud, Catherine Branger (CHU Louis-Mourier, Colombes); Agnès Lefort, Bruno Fantin, Claire Bellier, Frédéric Bert, Marie-Hélène Nicolas-Chanoine (CHU Beaujon, Clichy); Bernard Page, Julie Cremniter, Jean-Louis Gaillard (CHU Ambroise-Paré, Boulogne-Billancourt); Bernard Garo, Séverine Ansart, Geneviève Herry-Arnaud, Didier Tandé (CHU Brest, Brest); Jean-Claude Renet, René Ze Bekolo, Renaud Verdon, Roland Leclercq (CHU Caen, Caen); Claire de Gialluly, Jean-Marc Besnier, Laurent Mereghetti, Roland Quentin (CHU Tours, Tours); Achille Kouatchet, Alain Mercat, Marie Laure Joly-Guillou (CHU Angers, Angers); Catherine Dalebroux, Pascal Chavanet, Catherine Neuwirth (CHU Dijon, Dijon); Camille Colliard, Martin Dary, Gilles Potel, Jocelyne Caillon (CHU Nantes, Nantes); Françoise Leturdu, Jean-Pierre Sollet, Gaëtan Plantefève (CH Argenteuil, Argenteuil); Agnès de Patureaux, Pierre Tattevin, Pierre-Yves Donnio (CHU Rennes, Rennes), all in France. Those responsible for bacterial genotyping are Erick Denamur, Olivier Clermont, Christine Amorin, Jeremy Glodt (INSERM, UMR722,

Université Paris-Diderot, Paris, France). Those responsible for methodology are Xavière Panhard, Ludovic Lassel, Quentin Dornic, France Mentré (AP-HP, Hôpital Bichat, UF de Biostatistiques, Paris, France), Estelle Marcault, Florence Tubach (CHU Bichat-Claude-Bernard, Paris, France).

SUPPLEMENTARY MATERIAL

ACKNOWLEDGMENTS

DECLARATION OF INTEREST

None.

REFERENCES

1. **Russo TA, Johnson JR.** Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes and Infection* 2003; **5**: 449–456.

2. **Lefort A, et al.** Host factors and portal of entry outweigh bacterial determinants to predict the severity of *Escherichia coli* bacteremia. *Journal of Clinical Microbiology* 2011; **49**: 777–783.

3. **Rodriguez-Bano J, et al.** Community-onset bacteremia due to extended-spectrum beta-lactamase-producing *Escherichia coli*: risk factors and prognosis. *Clinical Infectious Diseases* 2010; **50**: 40–48.

4. **Courpon-Claudinon A, et al.** Bacteraemia caused by third-generation cephalosporin-resistant *Escherichia coli* in France: prevalence, molecular epidemiology and clinical features. *Clinical Microbiology and Infection* 2011; **17**: 557–565.

5. **Doi Y, et al.** Community-associated extended-spectrum beta-lactamase-producing *Escherichia coli* infection in the United States. *Clinical Infectious Diseases* 2013; **56**: 641–648.

6. **Tenaillon O, et al.** The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology* 2010; **8**: 207–217.

7. **Smati M, et al.** Real-time PCR for quantitative analysis of human commensal *Escherichia coli* populations reveals a high frequency of subdominant phylogroups. *Applied and Environmental Microbiology* 2013; **79**: 5005–5012.

8. **Moreno E, et al.** Relationship between *Escherichia coli* strains causing acute cystitis in women and the fecal *E. coli* population of the host. *Journal of Clinical Microbiology* 2008; **46**: 2529–2534.

9. **Chen SL, et al.** Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. *Science Translational Medicine* 2013; **5**: 184ra60.

10. **Krawczyk B, et al.** Characterisation of *Escherichia coli* isolates from the blood of haematological adult patients with bacteraemia: translocation from gut to blood requires the cooperation of multiple virulence factors. *European Journal of Clinical Microbiology and Infectious Diseases* 2015; **34**: 1135–1143.

11. **Carlet J.** The gut is the epicentre of antibiotic resistance. *Antimicrobial Resistance and Infection Control* 2012; **1**: 39.

12. **Caugant DA, et al.** Genetic diversity and relationships among strains of *Escherichia coli* in the intestine and those causing urinary tract infections. *Progress in Allergy* 1983; **33**: 203–227.

13. **Johnson JR, et al.** Bacterial characteristics in relation to clinical source of *Escherichia coli* isolates from women with acute cystitis or pyelonephritis and uninfected women. *Journal of Clinical Microbiology* 2005; **43**: 6064–6072.

14. **Johnson JR, et al.** Host-pathogen relationships among *Escherichia coli* isolates recovered from men with febrile urinary tract infection. *Clinical Infectious Diseases* 2005; **40**: 813–822.

15. **Sabate M, et al.** Pathogenicity island markers in commensal and uropathogenic *Escherichia coli* isolates. *Clinical Microbiology and Infection* 2006; **12**: 880–886.

16. **Grude N, et al.** A comparison of phylogenetic group, virulence factors and antibiotic resistance in Russian and Norwegian isolates of *Escherichia coli* from urinary tract infection. *Clinical Microbiology and Infection* 2007; **13**: 208–211.

17. **Wang MC, et al.** Different roles of host and bacterial factors in *Escherichia coli* extra-intestinal infections. *Clinical Microbiology and Infection* 2009; **15**: 372–379.

18. **Jadhav S, et al.** Virulence characteristics and genetic affinities of multiple drug resistant uropathogenic *Escherichia coli* from a semi urban locality in India. *PLoS ONE* 2011; **6**: e18063.

19. **Kudinha T, et al.** Distribution of phylogenetic groups, sequence type ST131, and virulence-associated traits among *Escherichia coli* isolates from men with pyelonephritis or cystitis and healthy controls. *Clinical Microbiology and Infection* 2013; **19**: E173–180.

20. **Kudinha T, et al.** Genotypic and phenotypic characterization of *Escherichia coli* isolates from children with urinary tract infection and from healthy carriers. *Pediatric Infectious Disease Journal* 2013; **32**: 543–548.

21. **Johnson JR, et al.** Epidemiological correlates of virulence genotype and phylogenetic background among *Escherichia coli* blood isolates from adults with

diverse-source bacteremia. *Journal of Infectious Diseases* 2002; **185**: 1439–1447.

22. **Wang MC, *et al.*** The role of bacterial virulence and host factors in patients with *Escherichia coli* bacteremia who have acute cholangitis or upper urinary tract infection. *Clinical Infectious Diseases* 2002; **35**: 1161–1166.

23. **Sannes MR, *et al.*** Virulence factor profiles and phylogenetic background of *Escherichia coli* isolates from veterans with bacteremia and uninfected control subjects. *Journal of Infectious Diseases* 2004; **190**: 2121–2128.

24. **Massot M, *et al.*** 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* 2016; **162**: 642–650.

25. **Gordon DM, Stern SE, Collignon PJ.** Influence of the age and sex of human hosts on the distribution of *Escherichia coli* ECOR groups and virulence traits. *Microbiology* 2005; **151**: 15–23.

26. **Clermont O, *et al.*** The Clermont *Escherichia coli* phylotyping method revisited: improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports* 2013; **8**: 58–65.

27. **Skurnik D, *et al.*** Emergence of antimicrobial-resistant *Escherichia coli* of animal origin spreading in humans. *Molecular Biology and Evolution* 2016; **33**: 898–914.

28. **Clermont O, *et al.*** Development of an allele-specific PCR for *Escherichia coli* B2 sub-typing, a rapid and easy to perform substitute of multilocus sequence typing. *Journal of Microbiological Methods* 2014; **101**: 24–27.

29. **Johnson JR, *et al.*** Rapid and specific detection of *Escherichia coli* clonal group A by gene-specific PCR. *Journal of Clinical Microbiology* 2004; **42**: 2618–2622.

30. **Guinée PAM, Jansen WH.** *Escherichia coli* associated with neonatal diarrhoea in piglets and calves. In: Leeww PM, Guinée PAM, eds. *Laboratory Diagnosis in Neonatal Calf and Pig Diarrhoea, Current Topics in Veterinary and Animal Science.* The Hague: Martinus-Nijhoff, 1981, pp. 126–162.

31. **Clermont O, *et al.*** Animal and human pathogenic *Escherichia coli* strains share common genetic backgrounds. *Infection, Genetics and Evolution* 2011; **11**: 654–662.

32. **Wirth T, *et al.*** Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Molecular Microbiology* 2006; **60**: 1136–1151.

33. **Clermont O, Gordon D, Denamur E.** Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology* 2015; **161**: 980–988.

34. **Johnson JR, *et al.*** Molecular epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island. *Journal of Clinical Microbiology* 2008; **46**: 3906–3911.

35. **Cortes MA, *et al.*** Inactivation of *ibeA* and *ibeT* results in decreased expression of type 1 fimbriae in extraintestinal pathogenic *Escherichia coli* strain BEN2908. *Infection and Immunity* 2008; **76**: 4129–4136.

36. **Benjamini Y, Hochberg Y.** Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1995; **57**: 289–300.

37. **Bidet P, *et al.*** Combined multilocus sequence typing and O serogrouping distinguishes *Escherichia coli* subtypes associated with infant urosepsis and/or meningitis. *Journal of Infectious Diseases* 2007; **196**: 297–303.

38. **Burdet C, *et al.*** *Escherichia coli* bacteremia in children: age and portal of entry are the main predictors of severity. *Pediatric Infectious Disease Journal* 2014; **33**: 872–879.

39. **Picard B, *et al.*** The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infection and Immunity* 1999; **67**: 546–553.

40. **de Lastours V, *et al.*** Emergence of quinolone resistance in the microbiota of hospitalized patients treated or not with a fluoroquinolone. *Journal of Antimicrobial Chemotherapy* 2014; **69**: 3393–3400.

41. **Moissenet D, *et al.*** Meningitis caused by *Escherichia coli* producing TEM-52 extended-spectrum beta-lactamase within an extensive outbreak in a neonatal ward: epidemiological investigation and characterization of the strain. *Journal of Clinical Microbiology* 2010; **48**: 2459–2463.

42. **Clermont O, *et al.*** Evidence for a human-specific *Escherichia coli* clone. *Environmental Microbiology* 2008; **10**: 1000–1006.

43. **Bengtsson S, *et al.*** Sequence types and plasmid carriage of uropathogenic *Escherichia coli* devoid of phenotypically detectable resistance. *Journal of Antimicrobial Chemotherapy* 2012; **67**: 69–73.

44. **Gibreel TM, *et al.*** Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. *Journal of Antimicrobial Chemotherapy* 2012; **67**: 346–356.

45. **Alhashash F, *et al.*** Multidrug-resistant *Escherichia coli* bacteremia. *Emerging Infectious Diseases* 2013; **19**: 1699–1701.

46. **Johnson JR.** Virulence factors in *Escherichia coli* urinary tract infection. *Clinical Microbiology Review* 1991; **4**: 80–128.