

**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.**

Ménil Massot, Anne-Sophie Daubié, Olivier Clermont, Françoise Jauréguy, Camille Couffignal, Ghizlane Dahbi, Azucena Mora, Jorge Blanco, Catherine Branger, France Mentré, et al.

► **To cite this version:**

Ménil Massot, Anne-Sophie Daubié, Olivier Clermont, Françoise Jauréguy, Camille Couffignal, 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, Microbiology Society, 2016, 162 ((4)), pp.642-50. <10.1099/mic.0.000242.>. <inserm-01393146>

**HAL Id: inserm-01393146**

**<http://www.hal.inserm.fr/inserm-01393146>**

Submitted on 6 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**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 30years**

**Running title:** Commensal human *E. coli* in Paris

**Authors and affiliations:** MÉRIL Massot<sup>1,2,\*</sup>, Anne-Sophie Daubié<sup>1,3,4,\*</sup>, Olivier Clermont<sup>1,2,\*</sup>, Françoise Jauréguy<sup>1,3,4</sup>, Camille Couffignal<sup>1,2</sup>, Ghizlane Dahbi<sup>5</sup>, Azucena Mora<sup>5</sup>, Jorge Blanco<sup>5</sup>, Catherine Branger<sup>1,2</sup>, France Mentré<sup>1,2</sup>, Alain Eddi<sup>6</sup>, Bertrand Picard<sup>1,3,4</sup>, Erick Denamur<sup>1,2</sup>, and the COLIVILLE Group<sup>†</sup>

\* These authors have equally contributed to the work

<sup>1</sup> INSERM, IAME, UMR1137, Paris, France

<sup>2</sup> Univ Paris Diderot, IAME, UMR1137, Sorbonne Paris Cité, Paris, France

<sup>3</sup> APHP, Hôpitaux Universitaires Paris Seine Saint-Denis, Site Avicenne, Bobigny, France

<sup>4</sup> Univ Paris Nord, IAME, UMR1137, Sorbonne Paris Cité, Bobigny, France

<sup>5</sup> Laboratorio de Referencia de *E. coli* (LREC), Departamento de Microbiología e

Parasitología, Facultad de Veterinaria, Universidad de Santiago de Compostela (USC),

Lugo, Spain

<sup>6</sup> Département de Médecine Générale, Univ Paris Diderot, Sorbonne Paris Cité, Faculté de

Médecine, Paris, France

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

31

32 **Corresponding author:** Erick Denamur, IAME, UMR1137, INSERM, Université Paris  
33 Diderot, Site Xavier Bichat, 16, rue Henri Huchard, 75018 Paris, France. Tel: 33 1 57 27 77  
34 39; email: erick.denamur@inserm.fr

35

36 **Key words:***Escherichia coli*, phylogroup, microbiota, commensal

37

38 **Content category:** Environmental and Evolutionary Microbiology

39

40 **Word counts:**294words in the abstract,4019in the main text

41

42 **SUMMARY**

43

44 It is important to study commensal populations of *Escherichia coli* because they appear to be  
45 the reservoir of both extra-intestinal pathogenic *E. coli* and antibiotic resistant strains of *E.*  
46 *coli*. We studied 279 dominant faecal strains of *E. coli* from 243 adults living in the community in  
47 the Paris area in 2010. The phylogenetic group and sub-group [sequence type complex  
48 (STc)] of the isolates and the presence of 20 virulence genes were determined by PCR  
49 assays. The O-types and the resistance to 18 antibiotics were assessed phenotypically. The B2  
50 group was the most frequently recovered (34.0%), followed by the A group (28.7%), and other  
51 groups were rarer. The most prevalent B2 subgroups were II (STc73), IV (STc141), IX  
52 (STc95) and I (STc131) with 22.1%, 21.1%, 16.8% and 13.7%, respectively, of the B2 group  
53 strains. Virulence factors (VFs) were more common in B2 group than other strains. One or  
54 more resistance was found in 125 strains (44.8% of the collection) but only six (2.2% of the  
55 collection) were multiresistant; no extended-spectrum beta-lactamase-producing strain was  
56 isolated. The C phylogroup and clonal group A strains were the most resistant. No trade-off  
57 between virulence and resistance was evidenced. We compared these strains to collections of  
58 strains gathered in the same conditions 30 and 10 years ago. There has been a parallel and  
59 linked increase in the frequency of B2 group strains (from 9.4% in 1980 to 22.7% in 2000 and  
60 34.0% in 2010) and of VFs. Antibiotic resistance also increased, from 22.6% of strains  
61 resistant to at least one antibiotic in 1980 to 31.8% in 2000 and 44.8% in 2010; resistance to  
62 streptomycin, however, remained stable. Commensal human *E. coli* populations have clearly  
63 evolved substantially over time, presumably reflecting changes in human practices, and  
64 particularly increasing antibiotic use.

65

## 66 INTRODUCTION

67

68 *Escherichia coli* is a gut commensal of vertebrates, including humans(Berg, 1996). It can  
69 nevertheless cause a broad range of diseases from various diarrheal diseases to extra-intestinal  
70 diseases, and particularly urinary tract infections and bacteraemia(Russo & Johnson, 2003,  
71 Kaper *et al.*, 2004). The gut is considered to be the reservoir of strains causing these extra-  
72 intestinal pathologies, and both epidemiological and experimental studies have led to the  
73 notion that virulence could be a by-product of commensalism(Nowrouzian *et al.*, 2005, Le  
74 Gall *et al.*, 2007, Diard *et al.*, 2010).

75

76 Molecular typing, based on multilocus sequence typing (MLST), various PCR based  
77 approaches and, more recently, whole genome sequencing, has provided a good understanding  
78 of the phylogenetic framework of the *E. coli* species(Clermont *et al.*, 2015). The species *E.*  
79 *colisensu stricto* is largely clonal(Desjardins *et al.*, 1995), with seven main phylogenetic  
80 groups (A, B1, B2, C, D, E, and F) each composed of several clones/clonal complexes or sub-  
81 groups(Clermont *et al.*, 2013). *Escherichia* clades have been described: they are  
82 phenotypically indistinguishable from *E. coli* but the nucleotide sequences are highly  
83 divergent, with clade I strains being the most closely related to *E. coli sensu stricto*(Walk *et*  
84 *al.*, 2009). Membership of these phylogenetic entities is of particular interest, as there is a  
85 relationship between the genetic background of a strain and its virulence factors(VFs)  
86 (Escobar-Paramo *et al.*, 2004). For example, some sequence types (STs) within the B2  
87 phylogroup, as ST137, include strains with numerous extra-intestinal genes and high virulence  
88 potential(Messika *et al.*, 2012).

89

90 Recently, there have been some major shifts in the epidemiology of *E. coli*. Some clonal  
91 complexes, such as ST complex (STc)69, STc73, STc95 and STc131, have become more  
92 prevalent in extra-intestinal diseases(Bidet *et al.*, 2007, Bert *et al.*, 2010, Mahjoub-Messai *et*  
93 *al.*, 2011, Bengtsson *et al.*, 2012, Gibreel *et al.*, 2012, Alhashash *et al.*, 2013, Clermont *et al.*,  
94 2014). Some of these clonal complexes have members that spread resistance to various  
95 antibiotics such as the cotrimoxazole [clonal group A (CGA) or STc69](Manges *et al.*,  
96 2001) and third generation cephalosporins [ST131 producing extended-spectrum beta-  
97 lactamase (ESBL)](Nicolas-Chanoine *et al.*, 2014). The gut has been found to be at the hub of  
98 this resistance (Carlet, 2012) and a major driver of the spread of antibiotic resistance in the  
99 community (Woerther *et al.*, 2013).

100

101 In view of these major changes to *E. coli* epidemiology, it would be valuable to have recent  
102 and reliable data on human commensal strains. Unfortunately, such data are scarce and most  
103 relevant studies are based on specific populations, like hospitalised patients or travellers, and  
104 many are focused on resistant strains. We report a study of a collection of 279 *E. coli* strains  
105 gathered from adult subjects living in the community in the Paris area (hereafter referred to as  
106 the “COLIVILLE” collection). We compared these strains to collections of strains gathered in  
107 the same conditions 30 (Duriez *et al.*, 2001) and 10 (Escobar-Paramo *et al.*, 2004) years ago by  
108 our group.

109

## 110 **MATERIALS AND METHODS**

111

### 112 **Subjects of the COLIVILLE collection**

113 Subjects were recruited by general practitioners of the Department of General Practice of the  
114 University Paris Diderot from the region of Ile-de-France (Paris, France, and its suburban  
115 area) from May 2009 to December 2011. All participants lived in the community and  
116 volunteered to self-collect a faecal swab sample. The inclusion criteria were: age of 18 years  
117 or more, no history of gastrointestinal disease, no symptoms of immunosuppression, no  
118 antibiotic therapy in the previous month and no hospitalisation in the 3 months preceding  
119 inclusion. Written informed consent was obtained from each participant, and the study was  
120 approved by the ethics evaluation committee of Institut National de la Santé et de la  
121 Recherche Médicale (INSERM) (CCTIRS no. 09.243, CNIL no. 909277, and CQI no. 01-  
122 014).

123

### 124 **Isolation and characterisation of the COLIVILLE strains**

125 Faecal samples were self-collected by the subjects. Immediately after stool emission, a swab  
126 was dipped into the faeces, put in Amies transport medium (Medical Wire & Equipment  
127 Corsham, Wiltshire England) and sent by mail to the Avicenne hospital laboratory (Bobigny,  
128 France) (Smati *et al.*, 2013). The swabs were plated onto Drigalski agar plates upon arrival  
129 (Bio-Rad, Life Science, Marnes-la-Coquette, France). After 24 hours of incubation at 37°C,  
130 the plates were inspected for the phenotypic aspect of the lactose-positive colonies. In most  
131 cases, only one type of colony was recovered. However, in some cases, two co-dominant  
132 types of colony were retrieved. One colony, or if appropriate two phenotypically distinct  
133 colonies, were randomly picked, allowing the recovery of the dominant *E. coli* clones of

134 the commensal faecal microbiota from each subject (Smati *et al.*, 2013). These isolates were  
135 identified as *E. coli* using API 20E (bioMérieux, Marcy l’Etoile, France), and stored in  
136 glycerol stock solution at -80°C.

137

### 138 ***E. coli* genotyping**

139 DNA was extracted from colonies of each strain using the Wizard genomic DNA purification  
140 kit (Promega, France) following the manufacturer instructions. The *E. coli* phylogroup (A, B1,  
141 B2, C, D, E, F) and *Escherichia* clade of each isolate were determined by the PCR quadruplex  
142 (Clermont *et al.*, 2013) and *aes* and *chuA* allele-specific amplification (Clermont *et al.*, 2011)  
143 methods, respectively. PCR-based methods were used to identify strains of CGA clonal  
144 complex (STc69) within the D phylogroup (Johnson *et al.*, 2004) and to classify the B2  
145 phylogroup strains into the 10 main subgroups (I to X) (Clermont *et al.*, 2014, Clermont *et al.*,  
146 2015). Strains were screened for the presence of 20 VFs representative of the main classes of  
147 *E. coli* extraintestinal virulence determinants known, including adhesins (*iha*, *papC*,  
148 *hra*, *sfa/foc*, *papGII* and *papGIII* alleles of *papG*, and *ibeA*), iron capture systems (*irp2*, *fyuA*,  
149 *iucC*, *iroN* and *ireA*), protectins (*ompT*, *traT* and *neuC*, chromosomal) and toxins  
150 (*usp*, *sat*, *clbQ*, *hlyC*, and *cnfI*) (Johnson *et al.*, 2008, Clermont *et al.*, 2011). For each isolate, a  
151 virulence score, defined as the number of the 20 VFs tested that were present in that strain,  
152 was calculated [adapted from (Lefort *et al.*, 2011)].

153

### 154 **O-typing**

155 Classical determination of O antigens was carried out according to Guinée & Jansen (1981),  
156 with all O (O1 to O181) antisera available. Subtyping of O2, O6, O25 and O45 types was  
157 performed by PCR as previously described (Clermont *et al.*, 2011).

158

### 159 **Antibiotic resistance phenotypes**

160 The strains were tested for their antibiotic susceptibilities using a disk diffusion method  
161 according to the recommendations of the Antimicrobial Committee of the French Society for  
162 Microbiology (<http://www.sfm-microbiologie.org>). The following antimicrobial agents were  
163 tested: amoxicillin, amoxicillin-clavulanic acid, cefoxitin, cefotaxime, ertapenem, imipenem,  
164 streptomycin, gentamicin, tobramycin, netilmicin, amikacin, tetracycline, cotrimoxazole  
165 (sulfamethoxazole-trimethoprim), chloramphenicol, nalidixic acid, ofloxacin, fosfomicin, and  
166 nitrofurantoin. A strain was considered to be multiresistant if it was resistant to, at  
167 least, penicillins, cotrimoxazole and quinolones (Lefort *et al.*, 2011). For each isolate, a

168 resistance score was computed. The resistance score was defined as the number of antibiotic  
169 classes to which the strain was resistant, among the ten tested (penicillins, cephalosporins,  
170 carbapenems, aminoglycosides, tetracyclines, sulfonamides, amphenicols, quinolones,  
171 phosphonic acids, furans).

172

### 173 **Characteristics of other collections of commensal strains**

174 Two previously published collections of human commensal strains isolated from the Paris area  
175 were studied for comparison. The criteria of selection of the subjects and protocols of *E. coli*  
176 isolation were similar to those of the COLIVILLE collection, except for the number of  
177 isolates studied for each individual. The VDG collection was from subjects sampled in 1980  
178 and included 53 strains from 53 subjects (Duriez *et al.*, 2001) (one strain per individual); the  
179 AEM collection was obtained in 2000 and included 44 strains from 27 subjects (Escobar-  
180 Paramo *et al.*, 2004) (1 to 4 strains per individual). The phylogroup/subgroup membership, the  
181 presence of 10 VFs representative of the four main classes (*sfa/foc*, *papC*, *iroN*, *iucC*, *fyuA*,  
182 *neuC*, *traT*, *hlyC*, *cnfI* and *usp*), and the susceptibility to seven antibiotics (amoxicillin,  
183 amoxicillin-clavulanic acid, streptomycin, chloramphenicol, cotrimoxazole, tetracycline  
184 and nalidixic acid) of these strains were determined as above. For all the strains of the three  
185 collections, we calculated a 'reduced' virulence score, defined as the number of the 10 tested  
186 VFs present in the strain.

187

### 188 **Statistical analyses**

#### 189 *Study of the COLIVILLE strain determinants*

190 Several bacterial determinants were compared between phylogroups using Fisher exact tests  
191 and analyses of variance. Fisher exact tests were used to test associations between phylogroup  
192 and O-type and between phylogroup and VFs. Associations were tested for the most  
193 numerous O-types (10 or more strains), and all the studied VFs. Associations between the  
194 most frequent antibiotic resistances (at least five resistant strains) were similarly tested.

195 To evaluate the strength of association between each VF and B2 group strains, Fisher's exact  
196 test was used to compare the distribution of each VF between B2 group and all the other  
197 groups combined. As multiple tests were performed, the p values were adjusted using the  
198 Benjamini and Hochberg method (Benjamini & Hochberg, 1995). Virulence and resistance  
199 scores were compared between phylogroups by analyses of variance; if significant, Tuckey's  
200 test was used for pairwise comparisons between phylogroups. Virulence scores were  
201 compared between B2 subgroups in a similar way. The Pearson correlation test was used to



202 assess the relationship between virulence and resistance scores. The resistance scores of the  
203 clonal group A strains and the other phylogroup D strains were compared using a Wilcoxon-  
204 Mann-Whitney test.

#### 205 *Comparison with other commensal collections*

206 The proportion of phylogenetic groups and of resistant strains were compared between the three  
207 Parisian collections (VDG, AEM and COLIVILLE) by Fisher exact tests. The 'reduced'  
208 virulence scores of the three collections were compared using a linear regression model with  
209 the date of sampling and the phylogenetic group as predictors.

210 All statistical analyses were performed using R software (R version 3.0.2). Data are shown as  
211 means with standard deviations (SD) for continuous variables and number and percentage for  
212 categorical variables. All tests were two-sided with a 5% type I error.

213

## 214 **RESULTS**

215

### 216 **COLIVILLE subject population and strains**

217 Two hundred and forty four subjects fulfilling our inclusion criteria were enrolled; there were  
218 111 men and 133 women. The mean age was 57 years (11.4): 57 years for the men (12.5) and  
219 56 years for the women (10.3). The dominant *E. coli* clone was isolated from the stools of  
220 these participants by plating on selective Drigalski medium. *E. coli* colonies were  
221 recovered from all subjects. For 208 subjects, colonies were phenotypically homogeneous and  
222 one was randomly isolated for study. For 36 subjects, there were two phenotypically distinct  
223 types of colonies, and one representative of each type was isolated and studied. Thus, 280  
224 strains were isolated and phenotypically identified as *E. coli* using API 20E; 279 from 243  
225 subjects were identified as *E. coli sensu stricto* by the quadruplex PCR assay (Clermont *et al.*,  
226 2013) and one strain from a subject exhibiting phenotypically homogeneous colonies was  
227 identified as *Escherichia* clade IV (Clermont *et al.*, 2011) (Table S1). This last subject was not  
228 included in the subsequent analyses.

229

### 230 **Phylogenetic group/subgroup, O-type and VF distribution**

231 The PCR quadruplex method was used to classify all isolates among the seven main  
232 phylogroups. Group B2 was the most frequently recovered (34.0%), followed by group A  
233 (28.7%) and the group B1 (12.9%); members of groups F (9.7%), D (9.0%), C (3.2%) and E  
234 (2.5%) were less numerous (Table 1). Many of the group B2 strains were in four sub-groups:  
235 sub-group II (STc73) (22.1% of the B2 group), sub-group IV (STc141) (21.1%), sub-group IX

236 (STc95) (16.8%) and sub-group I (STc131) (13.7%) (Table 1). Four sub-group I strains  
237 belonged to ST131. Only 7.3% of the B2 strains were not assigned to a sub-group by our PCR-  
238 based assay (Clermont *et al.*, 2014). Forty-eight % of the D strains belonged to CGA (STc69).  
239

240 The standard O-typing assay failed to assign 22 strains to an O-type. Among the 257 typeable  
241 strains, 75 O-types were detected, 28 represented by single strains and 47 by at least two  
242 strains. There were at least 10 strains in each of five O-types: O6a, O1, O2b, O9 and  
243 O18. There was an association between O-type and phylogenetic group: most of the O6a, O1  
244 and O2b strains belonged to the B2, F and B2 groups, respectively (p value =  $2 \times 10^{-9}$ ), and  
245 most non-typable strains belonged to the A group. Three of the ST131 strains belonged to the  
246 ST131-O25b clone and one belonged to the ST131-O16 clone. Most strains of the B2 sub-  
247 groups II, IV, VI, VII and VIII were O6a-, O2b-, O4-, O75-, and O81-types, respectively,  
248 whereas the B2 subgroup I and IX strains were more heterogeneous (I: O25b, O9; IX: O1,  
249 O18, O2a). In the phylogroup D, CGA strains were O17 and O77 types, and these types were  
250 not detected among non-CGA strains (Table S1).

251  
252 The 279 strains were screened for 20 VF genes. VF genes involved in iron acquisition and  
253 protectins were more frequent than those for adhesins and toxins (Table 2). Three VF genes  
254 were found in more than half of the strains: *irp2* (60.9%) and *fuyA* (60.2%) both belonging to  
255 the high pathogenicity island (HPI) and involved in iron acquisition (Schubert *et al.*, 2009),  
256 and the protectin gene *ompT* (58.4%) (Table 2). The carriage of each of the 20 tested virulence  
257 factors differed significantly between phylogroups (global adjusted p values  $\leq 0.005$ , Table 2).  
258 Seventeen VF genes were significantly more prevalent among phylogroup B2 strains than  
259 among other strains (adjusted p values  $\leq 0.002$ ), whereas the three others (*iha*, *traT* and *sat*)  
260 were not (Table 2). Three VFs (*papGIII*, *clbQ* and *cnfI*) were found exclusively in the B2  
261 strains.

262 The virulence score for the strains was from 0 to 17, with a mean of 5.9 (4.5) (Table 1).  
263 Virulence scores differed significantly between the phylogroups (p value =  $10^{-16}$ ). The virulence  
264 score of the B2 strains was significantly higher than those of the A, B1, C, D, E and F strains  
265 (p value =  $10^{-13}$ ,  $10^{-13}$ ,  $10^{-6}$ ,  $10^{-13}$ ,  $10^{-8}$  and  $10^{-7}$ , respectively). D strains had a significantly  
266 higher virulence score than A and B1 strains (p value = 0.004 and 0.01, respectively) and F  
267 strains had a significantly higher virulence score than A strains (p value =  $10^{-10}$ ). The virulence  
268 score was significantly different between subgroups of the B2 group (p value =  $10^{-5}$ ): sub-  
269 groups II, IV and IX had significantly higher virulence scores, with means of 12.5 (2.3), 10.7

270 (3.0) and 11.2 (1.3), respectively, than the sub-group I with mean 7.2 (1.9) ( $p$  value=  $10^{-5}$ ,  
271 0.004 and 0.001, respectively).

272

### 273 **Antibiotic resistance phenotype**

274 Resistance to one or more of the 18 tested antibiotics was found in 125 (44.8%) of the 279  
275 strains (Table 1). Only 6 strains (2.2%) were multiresistant, *i.e.* resistant to, at least,  
276 amoxicillin, cotrimoxazole and nalidixic acid. Overall, 29.7% of the strains were resistant to  
277 amoxicillin and 16.1% to amoxicillin-clavulanic acid (Table 3). The only strain resistant  
278 to cefotaxime did not produce an ESBL but had a plasmid-encoded CMY-2 cephalosporinase  
279 (data not shown). No strain was resistant to carbapenems. Among the aminoglycosides,  
280 23.3% of the strains were resistant to streptomycin and only 0.7%, 0.7% and 0.4% of the  
281 strains were resistant to gentamicin, tobramycin and netilmicin, respectively (Table 3).  
282 Resistance to tetracycline was found in 21.9% of the strains, to cotrimoxazole in 15.8% and to  
283 chloramphenicol in 7.2%; 6.1% of strains were resistant to nalidixic acid and 3.2% to  
284 ofloxacin. Significant associations were found between penicillin resistance and resistance to  
285 the following antibiotics: chloramphenicol, cotrimoxazole, streptomycin, and tetracycline  
286 (adjusted  $p$  values  $< 0.05$ ). Associations between resistances to chloramphenicol,  
287 cotrimoxazole, streptomycin, and tetracycline were also significant (adjusted  $p$  values  $< 10^{-3}$ ).  
288 Resistance to nalidixic acid was significantly associated with resistance to amoxicillin-  
289 clavulanic acid, streptomycin, cotrimoxazole and chloramphenicol ( $p$  value= 0.049, 0.048,  
290 0.049 and 0.007, respectively).

291 Resistance scores were between 0 and 6, with a mean of 1.1 (1.5) (Table 1). There was a  
292 significant difference of the resistance score between phylogroups ( $p$  value= 0.001). The  
293 resistance score of the C strains was significantly higher than those of the A, B1, B2, E and F  
294 strains ( $p$  value= 0.002, 0.02, 0.002, 0.03 and 0.03, respectively). The mean resistance score  
295 of the group D strains was 1.7 (2.0); the mean resistance score of CGA strains was 2.7 (1.7) and  
296 was significantly higher than the mean score of other group D strains, which was 0.9 (1.8) ( $p$   
297 value= 0.009).

298

299 For the 279 strains as a whole, there was no significant correlation (Pearson coefficient= 0.06,  
300  $p$  value= 0.3; adjusted  $R^2$  of linear regression= 0.0006,  $p$  value= 0.3) between the virulence  
301 and the resistance scores.

302

### 303 **Comparison with other collections of commensal *E. coli***

304 We compared the COLIVILLE collection to collections obtained in 1980 (VDG) and 2000  
305 (AEM) in the same area. The most apparent observation was the increase in the proportion of  
306 group B2 strains from 9.4% in 1980 to 22.7% in 2000 and 34.0% in 2010, associated with a  
307 parallel decrease in the proportion of phylogroup A strains (52.8%, 27.3% and 28.7%,  
308 respectively,  $p$  value= 0.0002) (Fig. 1a). No CGA strain was evidenced in the VDG and AEM  
309 collections. The 'reduced' virulence score has increased significantly over time, from a mean of  
310 1.5 (2.0) in 1980, to 2.3 (2.2) in 2000 and 3.2 (2.6) in 2010 (Fig. 1b). Our model indicated that  
311 this increase was due to the change in the phylogenetic distribution between 1980 and 2010  
312 (effects of date of sampling were not significant), and in particular the increase in the  
313 proportion of the group B2 strains and the decrease in that of the A group strains. A more  
314 thorough analysis of the few B2 strains available from the VDG and AEM collections showed  
315 that classical extra-intestinal pathogenic *E. coli* (ExPEC) B2 subgroup strains (subgroup  
316 II/O6a and O22, subgroup III/O6a and subgroup IX/O1) with numerous VFs were already  
317 present (Table S2).

318 We also studied the evolution of antibiotic resistance in these populations (Fig. 1c). The  
319 proportion of strains resistant to at least one antibiotic increased from 22.6% in 1980 to 31.8%  
320 in 2000 and 44.8% in 2010 ( $p$  value=0.005). We observed the emergence of the resistance to  
321 amoxicillin and amoxicillin-clavulanic acid to attain 29.7% and 16.1%, respectively, by  
322 2010. A similar trend was observed for cotrimoxazole and tetracycline whereas the resistance  
323 to streptomycin remained generally stable.

324

## 325 **DISCUSSION**

326

327 A major issue when isolating commensal strains of an opportunistic pathogenic species, such  
328 as *E. coli*, is the quality of the epidemiological data available for the human host. We therefore  
329 applied very strict inclusion criteria for the subjects, excluding individuals with any antibiotic  
330 therapy in the previous month and any hospitalisation in the 3 months preceding inclusion; the  
331 subjects were all epidemiologically unrelated and were recruited in the same geographic area  
332 and over a short period of time. To our knowledge, this is the largest series of "true" human  
333 commensal strains to be thoroughly studied, with data including population genetics,  
334 virulence and antibiotic resistance data. This collection will be useful to the scientific  
335 community for subsequent comparison purposes (Table S1).

336

337 The high frequency of the phylogroup B2 strains in 2010 (Table 1) is inconsistent with recent  
338 descriptions of human commensal strains in industrialised countries (Bailey *et al.*, 2010,  
339 Bailey *et al.*, 2010, Tenaillon *et al.*, 2010). It contrasts with data obtained in developing  
340 countries where phylogroup A and B1 strains predominate (Unno *et al.*, 2009, Bailey *et al.*,  
341 2010, Li *et al.*, 2010, Tenaillon *et al.*, 2010, Lescat *et al.*, 2013, Tapader *et al.*, 2014). Group A  
342 strains are the best adapted to a wide range of human hosts, whatever their lifestyle, diet and  
343 hygiene status. Our study is one of the first to use the quadruplex PCR assay (Clermont *et al.*,  
344 2013) allowing the identification of the minor phylogenetic groups C, E and F. Few group C  
345 and E strains were found (3.2% and 2.5%, respectively) whereas almost 10% of  
346 strains were group F, such that they were more numerous than group D strains (9.0%). Half  
347 of the group D strains were CGA strains, a clonal group including ExPEC (Manges *et al.*,  
348 2001). Only one *Escherichia* clade was found, confirming the rarity of other *Escherichia*  
349 clades in human (Clermont *et al.*, 2011). We did not find any significant correlation between  
350 the phylogroup distribution and the age and sex of the subjects (data not shown), contrary to  
351 observations in Australia (Gordon *et al.*, 2005). We compared our 2010 collection to previous  
352 collections from the Paris area. The frequency of isolation of phylogroup B2 strains has  
353 increased with a parallel decrease of that of phylogroup A strains (Fig. 1a); this may be a  
354 consequence of modifications in the lifestyle of the Parisian population including changing  
355 food processing and hygiene procedures. Conversely, the prevalence of the minor groups (C,  
356 E and F) has remained stable over time and maybe less sensitive to these changes. The  
357 presence of VFs is linked to phylogenetic group membership, a well-known phenomenon  
358 (Escobar-Paramo *et al.*, 2004), and we observed an increase of the number of VFs with time,  
359 paralleling the changes in phylogroup distribution (Fig. 1b).

360  
361 Group B2 has various sub-groups or clonal complexes, three of them (sub-group I=STc131;  
362 sub-group II=STc73; sub-group IX=STc95) containing most ExPEC strains (Bengtsson *et al.*,  
363 2012, Gibreel *et al.*, 2012, Alhashash *et al.*, 2013, Clermont *et al.*, 2014). We found numerous  
364 strains of sub-groups II and IX among the commensal B2 strains, with the classical ExPEC O-  
365 types (*i.e.* sub-group II/O6a and sub-group IX/O1-O18-O2a) whereas few sub-group I strains,  
366 also frequently recovered among commensal strains, were the classical ST131/O25b-O16  
367 clones (Johnson *et al.*, 2014). We also found many B2 sub-group IV (STc141) strains  
368 exhibiting the O2b type. It is therefore possible that two types of B2 strains in the human  
369 microbiota can be distinguished: the classical ExPEC clones that have both the property of  
370 being highly virulent and the property of being good gut colonizers, the extra-intestinal

371 virulence being a by-product of commensalism (Nowrouzian *et al.*, 2005, Le Gall *et al.*, 2007,  
372 Diard *et al.*, 2010);and the commensal-specific strains, as has been reported for the O81 B2  
373 sub-group VIII (STc452) strains (Clermont *et al.*, 2008).

374

375 Overall, the strains we isolated exhibited a low level of antibiotic resistance (Table 1) relative  
376 to French clinical strains isolated over the same period (<http://www.onerba.org>). For example,  
377 resistances to amoxicillin, cefotaxime, gentamicin, cotrimoxazole and fluoroquinolones were  
378 51.8%, 7.6%, 5.6%, 23.9% and 14.4%, respectively, for clinical strains in 2010, but only  
379 29.7%, 0.4%, 0.7%, 15.8%, 6.1%, respectively, in the COLIVILLE collection. This low  
380 prevalence of resistance in our collection argues indirectly for the “true” commensal character  
381 of the strains. Beside the well-known CGA (Manges *et al.*, 2001), our work identifies, for the  
382 first time, phylogroup C as a potential reservoir of resistance among commensal strains (Table  
383 1). A strain of this phylogroup producing an ESBL was responsible for an extensive outbreak  
384 of digestive colonisation in a neonatal ward with one case of meningitis (Moissenet *et al.*,  
385 2010).

386

387 Although antibiotic resistance in our 2010 collection was less prevalent than among clinical  
388 strains (see above), it was higher than in the earlier similar collections (Fig. 1c). This is  
389 consistent with a recent study in a Parisian check-up centre where a 10-fold increase of ESBL-  
390 producing *E. coli* strains was observed between 2006 and 2011 (Nicolas-Chanoine *et al.*,  
391 2013). These effects are probably due to an increase in the exposure of the general population  
392 to antibiotic pressure (Woerther *et al.*, 2013). Contrasting with this trend of  
393 increasing resistance, the prevalence of resistance to streptomycin appears to be stable through  
394 time in the Parisian population. A similar level of resistance to streptomycin (23%) was  
395 reported for *E. coli* strains isolated in England in 1991 (Chiew *et al.*, 1998). Streptomycin  
396 therapy was frequent in the 1950s and 1960s but is not now widely used, and it largely  
397 restricted to tuberculosis. The non-disappearance of resistance to an (almost) unused antibiotic  
398 may be because the genes conferring this resistance are borne by genetic structures, including  
399 transposons and integrons, that possess other antibiotic resistance genes (Chiew *et al.*,  
400 1998). Indeed, we found an association in our collection between resistance to streptomycin and  
401 resistance to cotrimoxazole and chloramphenicol, all these resistances being mediated by co-  
402 localised integron-borne genes (Partridge *et al.*, 2009).

403

404 In conclusion, our data indicates that commensal human *E. coli* populations evolve  
405 substantially over time and that precise and repeated characterisations are required to  
406 document the emergence of virulent and/or resistant strains in the community.

407

## 408 **ACKNOWLEDGMENTS**

409

410 This work was partially supported by the grant CN2012/303 (Consellería de Cultura,  
411 Educación e Ordenación Universitaria, Xunta de Galicia and The European Regional  
412 Development Fund, ERDF).

413

## 414 **ABBREVIATIONS**

415

416 CGA: Clonal Group A, ExPEC: Extraintestinal Pathogenic *Escherichia coli*, STc: Sequence  
417 Type complex, ST: Sequence Type, VF: Virulence Factor, MLST: Multilocus Sequence  
418 Typing, ESBL: Extended Spectrum Beta-Lactamase, SD: Standard Deviation, HPI: High  
419 Pathogenicity Island.

420

## 421 **REFERENCES**

422

423 Alhashash F, Weston V, Diggle M & McNally A (2013) Multidrug-resistant *Escherichia coli*  
424 bacteremia. *Emerg Infect Dis***19**: 1699-1701.

425 Bailey JK, Pinyon JL, Anantham S & Hall RM (2010) Commensal *Escherichia coli* of  
426 healthy humans: a reservoir for antibiotic-resistance determinants. *J Med Microbiol***59**: 1331-  
427 1339.

428 Bailey JK, Pinyon JL, Anantham S & Hall RM (2010) Distribution of human commensal  
429 *Escherichia coli* phylogenetic groups. *J Clin Microbiol***48**: 3455-3456.

430 Bengtsson S, Naseer U, Sundsfjord A, Kahlmeter G & Sundqvist M (2012) Sequence types  
431 and plasmid carriage of uropathogenic *Escherichia coli* devoid of phenotypically detectable  
432 resistance. *J Antimicrob Chemother***67**: 69-73.

433 Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate: a practical and  
434 powerful approach to multiple testing. *J R Stat Soc Ser B Methodol***57**: 289-300.

435 Berg RD (1996) The indigenous gastrointestinal microflora. *Trends Microbiol***4**: 430-435.

436 Bert F, Johnson JR, Ouattara B, Leflon-Guibout V, Johnston B, Marcon E, Valla D, Moreau  
437 R & Nicolas-Chanoine MH (2010) Genetic diversity and virulence profiles of *Escherichia*

438 *coli* isolates causing spontaneous bacterial peritonitis and bacteremia in patients with  
439 cirrhosis. *J Clin Microbiol***48**: 2709-2714.

440 Bidet P, Mahjoub-Messai F, Blanco J, Dehem M, Aujard Y, Bingen E & Bonacorsi S (2007)  
441 Combined multilocus sequence typing and O serogrouping distinguishes *Escherichia coli*  
442 subtypes associated with infant urosepsis and/or meningitis. *J Infect Dis***196**: 297-303.

443 Carlet J (2012) The gut is the epicentre of antibiotic resistance. *Antimicrob Resist Infect*  
444 *Control***1**: 39.

445 Chiew YF, Yeo SF, Hall LM & Livermore DM (1998) Can susceptibility to an antimicrobial  
446 be restored by halting its use? The case of streptomycin versus *Enterobacteriaceae*. *J*  
447 *Antimicrob Chemother***41**: 247-251.

448 Clermont O, Gordon D & Denamur E (2015) Guide to the various phylogenetic classification  
449 schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology***161**: 980-  
450 988.

451 Clermont O, Christenson JK, Denamur E & Gordon DM (2013) The Clermont *Escherichia*  
452 *coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-  
453 groups. *Environ Microbiol Rep***8**: 58-65.

454 Clermont O, Gordon DM, Brisse S, Walk ST & Denamur E (2011) Characterization of the  
455 cryptic *Escherichia* lineages: rapid identification and prevalence. *Environ Microbiol***13**: 2468-  
456 2477.

457 Clermont O, Christenson JK, Daubie AS, Gordon DM & Denamur E (2014) Development of  
458 an allele-specific PCR for *Escherichia coli* B2 sub-typing, a rapid and easy to perform  
459 substitute of multilocus sequence typing. *J Microbiol Methods***101**: 24-27.

460 Clermont O, Lescat M, O'Brien CL, Gordon DM, Tenailon O & Denamur E (2008) Evidence  
461 for a human-specific *Escherichia coli* clone. *Environ Microbiol***10**: 1000-1006.

462 Clermont O, Olier M, Hoede C, Diancourt L, Brisse S, Keroudean M, Glodt J, Picard B,  
463 Oswald E & Denamur E (2011) Animal and human pathogenic *Escherichia coli* strains share  
464 common genetic backgrounds. *Infect Genet Evol***11**: 654-662.

465 Desjardins P, Picard B, Kaltenbock B, Elion J & Denamur E (1995) Sex in *Escherichia coli*  
466 does not disrupt the clonal structure of the population: evidence from random amplified  
467 polymorphic DNA and restriction-fragment-length polymorphism. *J Mol Evol***41**: 440-448.

468 Diard M, Garry L, Selva M, Mosser T, Denamur E & Matic I (2010) Pathogenicity-associated  
469 islands in extraintestinal pathogenic *Escherichia coli* are fitness elements involved in  
470 intestinal colonization. *J Bacteriol***192**: 4885-4893.



471 Duriez P, Clermont O, Bonacorsi S, Bingen E, Chaventre A, Elion J, Picard B & Denamur E  
472 (2001) Commensal *Escherichia coli* isolates are phylogenetically distributed among  
473 geographically distinct human populations. *Microbiology***147**: 1671-1676.

474 Escobar-Paramo P, Clermont O, Blanc-Potard AB, Bui H, Le Bouguenec C & Denamur E  
475 (2004) A specific genetic background is required for acquisition and expression of virulence  
476 factors in *Escherichia coli*. *Mol Biol Evol***21**: 1085-1094.

477 Escobar-Paramo P, Grenet K, Le Menac'h A, *et al.* (2004) Large-scale population structure of  
478 human commensal *Escherichia coli* isolates. *Appl Environ Microbiol***70**: 5698-5700.

479 Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ & Upton M (2012) Population  
480 structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli*  
481 from Northwest England. *J Antimicrob Chemother***67**: 346-356.

482 Gordon DM, Stern SE & Collignon PJ (2005) Influence of the age and sex of human hosts on  
483 the distribution of *Escherichia coli* ECOR groups and virulence traits. *Microbiology***151**: 15-  
484 23.

485 Guinée PAM & Jansen WH (1981) *Escherichia coli* associated with neonatal diarrhoea in  
486 piglets and calves. *Laboratory diagnosis in neonatal calf and pig diarrhoea, current topics in*  
487 *veterinary and animal science*, (Leeww PM & Guinée PAM, eds.), p. 126-162. Martinus-  
488 Nijhoff, The Hague.

489 Johnson JR, Owens K, Manges AR & Riley LW (2004) Rapid and specific detection of  
490 *Escherichia coli* clonal group A by gene-specific PCR. *J Clin Microbiol***42**: 2618-2622.

491 Johnson JR, Johnston B, Kuskowski MA, Nougayrede JP & Oswald E (2008) Molecular  
492 epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island. *J Clin*  
493 *Microbiol***46**: 3906-3911.

494 Johnson JR, Clermont O, Johnston B, Clabots C, Tchesnokova V, Sokurenko E, Junka AF,  
495 Maczynska B & Denamur E (2014) Rapid and specific detection, molecular epidemiology,  
496 and experimental virulence of the O16 subgroup within *Escherichia coli* sequence type 131. *J*  
497 *Clin Microbiol***52**: 1358-1365.

498 Kaper JB, Nataro JP & Mobley HL (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol***2**:  
499 123-140.

500 Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E & Tenailon O (2007)  
501 Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic  
502 group *Escherichia coli* strains. *Mol Biol Evol***24**: 2373-2384.

503 Lefort A, Panhard X, Clermont O, Woerther PL, Branger C, Mentre F, Fantin B, Wolff M &  
504 Denamur E (2011) Host factors and portal of entry outweigh bacterial determinants to predict  
505 the severity of *Escherichia coli* bacteremia. *J Clin Microbiol***49**: 777-783.

506 Lescat M, Clermont O, Woerther PL, *et al.* (2013) Commensal *Escherichia coli* strains in  
507 Guiana reveal a high genetic diversity with host-dependant population structure. *Environ*  
508 *Microbiol Rep***5**: 49-57.

509 Li B, Sun JY, Han LZ, Huang XH, Fu Q & Ni YX (2010) Phylogenetic groups and  
510 pathogenicity island markers in fecal *Escherichia coli* isolates from asymptomatic humans in  
511 China. *Appl Environ Microbiol***76**: 6698-6700.

512 Mahjoub-Messai F, Bidet P, Caro V, Diancourt L, Biran V, Aujard Y, Bingen E & Bonacorsi  
513 S (2011) *Escherichia coli* isolates causing bacteremia via gut translocation and urinary tract  
514 infection in young infants exhibit different virulence genotypes. *J Infect Dis***203**: 1844-1849.

515 Manges AR, Johnson JR, Foxman B, O'Bryan TT, Fullerton KE & Riley LW (2001)  
516 Widespread distribution of urinary tract infections caused by a multidrug-resistant  
517 *Escherichia coli* clonal group. *N Engl J Med***345**: 1007-1013.

518 Messika J, Magdoud F, Clermont O, Margetis D, Gaudry S, Roux D, Branger C, Dreyfuss D,  
519 Denamur E & Ricard JD (2012) Pathophysiology of *Escherichia coli* ventilator-associated  
520 pneumonia: implication of highly virulent extraintestinal pathogenic *Escherichia coli* strains.  
521 *Intensive Care Med***38**: 2007-2016.

522 Moissenet D, Salauze B, Clermont O, Bingen E, Arlet G, Denamur E, Merens A, Mitanchez  
523 D & Vu-Thien H (2010) Meningitis caused by *Escherichia coli* producing TEM-52 extended-  
524 spectrum beta-lactamase within an extensive outbreak in a neonatal ward: epidemiological  
525 investigation and characterization of the strain. *J Clin Microbiol***48**: 2459-2463.

526 Nicolas-Chanoine MH, Bertrand X & Madec JY (2014) *Escherichia coli* ST131, an intriguing  
527 clonal group. *Clin Microbiol Rev***27**: 543-574.

528 Nicolas-Chanoine MH, Gruson C, Bialek-Davenet S, *et al.* (2013) 10-Fold increase (2006-11)  
529 in the rate of healthy subjects with extended-spectrum beta-lactamase-producing *Escherichia*  
530 *coli* faecal carriage in a Parisian check-up centre. *J Antimicrob Chemother***68**: 562-568.

531 Nowrouzian FL, Wold AE & Adlerberth I (2005) *Escherichia coli* strains belonging to  
532 phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants.  
533 *J Infect Dis***191**: 1078-1083.

534 Partridge SR, Tsafnat G, Coiera E & Iredell JR (2009) Gene cassettes and cassette arrays in  
535 mobile resistance integrons. *FEMS Microbiol Rev***33**: 757-784.

536 Russo TA & Johnson JR (2003) Medical and economic impact of extraintestinal infections  
537 due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes*  
538 *Infect***5**: 449-456.

539 Schubert S, Darlu P, Clermont O, Wieser A, Magistro G, Hoffmann C, Weinert K, Tenaillon  
540 O, Matic I & Denamur E (2009) Role of intraspecies recombination in the spread of  
541 pathogenicity islands within the *Escherichia coli* species. *PLoS Pathog***5**: e1000257.

542 Smati M, Clermont O, Le Gal F, Schichmanoff O, Jaureguy F, Eddi A, Denamur E & Picard  
543 B (2013) Real-time PCR for quantitative analysis of human commensal *Escherichia coli*  
544 populations reveals a high frequency of subdominant phylogroups. *Appl Environ*  
545 *Microbiol***79**: 5005-5012.

546 Tapader R, Chatterjee S, Singh AK, Dayma P, Haldar S, Pal A & Basu S (2014) The high  
547 prevalence of serine protease autotransporters of *Enterobacteriaceae* (SPATEs) in  
548 *Escherichia coli* causing neonatal septicemia. *Eur J Clin Microbiol Infect Dis***33**: 2015-2024.

549 Tenaillon O, Skurnik D, Picard B & Denamur E (2010) The population genetics of  
550 commensal *Escherichia coli*. *Nat Rev Microbiol***8**: 207-217.

551 Unno T, Han D, Jang J, Lee SN, Ko G, Choi HY, Kim JH, Sadowsky MJ & Hur HG (2009)  
552 Absence of *Escherichia coli* phylogenetic group B2 strains in humans and domesticated  
553 animals from Jeonnam Province, Republic of Korea. *Appl Environ Microbiol***75**: 5659-5666.

554 Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM & Whittam TS (2009)  
555 Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol***75**: 6534-6544.

556 Woerther PL, Angebault C, Jacquier H, *et al.*(2013) Characterization of fecal extended-  
557 spectrum-beta-lactamase-producing *Escherichia coli* in a remote community during a long  
558 time period. *Antimicrob Agents Chemother***57**: 5060-5066.

559

560 **Table 1.** Phylogenic, virulence and resistance characteristics of the 279 *E. coli* strains of the COLIVILLE collection

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

Phylogenetic groups/subgroups		Strains N (%)	Mean of virulence score* ( $\pm$ SD <sup>†</sup> )	Resistant strains‡ N (%)	Mean resistance score§ ( $\pm$ SD)
A		80 (28.7)	2.5 ( $\pm$ 2.5)	28 (35.0)	0.9 ( $\pm$ 1.5)
B1		36 (12.9)	2.3 ( $\pm$ 2.1)	17 (47.2)	1.0 ( $\pm$ 1.4)
B2		95 (34.0)	10.4 ( $\pm$ 3.0)	43 (45.3)	0.9 ( $\pm$ 1.3)
I	ST131	4 (4.2#)	7.8 ( $\pm$ 1.3)	2 (50.0**)	1.3 ( $\pm$ 1.5)
	Non ST131	9 (9.5)	7.2 ( $\pm$ 1.9)	8 (88.9)	1.6 ( $\pm$ 1.7)
II		21 (22.1)	12.5 ( $\pm$ 2.3)	8 (38.1)	0.6 ( $\pm$ 1.0)
III		2 (2.1)	11.5 ( $\pm$ 0.7)	1 (50.0)	0.5 ( $\pm$ 0.7)
IV		20 (21.1)	10.7 ( $\pm$ 3.0)	7 (35.0)	0.6 ( $\pm$ 1.1)
V		1 (1.1)	13.0	1 (100.0)	3.0
VI		3 (3.2)	13.7 ( $\pm$ 0.6)	2 (66.7)	1.7 ( $\pm$ 2.1)
VII		6 (6.3)	10.0 ( $\pm$ 1.7)	3 (50.0)	0.5 ( $\pm$ 0.6)
VIII		5 (5.2)	5.8 ( $\pm$ 1.3)	4 (80.0)	1.0 ( $\pm$ 0.7)
IX		16 (16.8)	11.2 ( $\pm$ 1.3)	4 (25.0)	0.4 ( $\pm$ 0.9)
X		1 (1.1)	4.0	1 (100.0)	2.0
Unassigned		7 (7.3)	10.0 ( $\pm$ 3.7)	4 (57.1)	1.7 ( $\pm$ 2.0)
C		9 (3.2)	5.2 ( $\pm$ 2.0)	8 (88.9)	2.9 ( $\pm$ 1.8)
D		25 (9.0)	4.9 ( $\pm$ 3.2)	14 (56.0)	1.7 ( $\pm$ 2.0)
CGA <sup>††</sup>		12 (48.0)	5.7 ( $\pm$ 3.2)	10 (83.3)	2.7 ( $\pm$ 1.7)
Non CGA		13 (52.0)	4.2 ( $\pm$ 3.1)	4 (30.8)	0.9 ( $\pm$ 1.8)
E		7 (2.5)	3.6 ( $\pm$ 1.9)	2 (28.6)	0.6 ( $\pm$ 1.1)
F		27 (9.7)	6.9 ( $\pm$ 3.8)	13 (48.1)	1.2 ( $\pm$ 1.6)
Total		279 (100.0)	5.9 ( $\pm$ 4.5)	125 (44.8)	1.1 ( $\pm$ 1.5)

584 \* Virulence score, defined as the number of virulence genes present in the strain, among the 20 tested.

585 <sup>†</sup> SD, standard-deviation.

586 <sup>‡</sup> Resistant strain, defined as a strain resistant to at least one antimicrobial agent among the 18 tested.

587 <sup>§</sup> Resistance score, defined as the number of antibiotic classes (penicillins, cephalosporins, carbapenems, aminoglycosides, amphenicols, sulfonamides, tetracyclines, quinolones, furans, phosphonic acids) to which the strain is resistant among the 10 tested.

589 || The subgroups I, II, III, IV, V, VI, VII, VIII, IX and X correspond to the STc131, 73, 127, 141, 144, 12, 14, 452, 95 and 372, respectively (Clermont *et al.*, 2015). ST131, Sequence Type 131.

591 # Proportions of subgroups are reported as fractions of the respective phylogroups.

592 \*\* Proportions of resistant strains within the subgroup.  
593 †† CGA, clonal group A (STc69).  
594

595 **Table 2.** Virulence traits of the 279 *E. coli* strains of the COLIVILLE collection, with respect to the phylogenetic group of the strains

Virulence traits	Function	Strains N (%)	Phylogenetic groups						Global adjusted p values	Adjusted p values B2 versus other groups	
			A (n= 80)	B1 (n= 36)	B2 (n= 95)	C (n= 9)	D (n= 25)	E (n= 7)			F (n= 27)
<i>iha</i>	Adhesin	94 (33.7)	18	7	35	1	12	5	16	0.0004	0.4
<i>papC</i>	Adhesin	65 (23.3)	4	1	42	1	4	0	13	9x10 <sup>-12</sup>	2x10 <sup>-8</sup>
<i>hra</i>	Adhesin	61 (21.9)	11	5	36	2	6	1	0	7x10 <sup>-5</sup>	9x10 <sup>-6</sup>
<i>sfa/foc</i>	Adhesin	51 (18.3)	1	1	46	0	0	3	0	6x10 <sup>-16</sup>	5x10 <sup>-16</sup>
<i>papGII*</i>	Adhesin	32 (11.5)	0	0	22	0	4	0	6	6x10 <sup>-7</sup>	5x10 <sup>-5</sup>
<i>ibeA</i>	Adhesin	30 (10.8)	0	0	29	0	1	0	0	2x10 <sup>-10</sup>	5x10 <sup>-14</sup>
<i>papGIII*</i>	Adhesin	21 (7.5)	0	0	21	0	0	0	0	4x10 <sup>-7</sup>	5x10 <sup>-11</sup>
<i>irp2</i>	Iron acquisition	170 (60.9)	28	10	88	8	15	0	21	6x10 <sup>-16</sup>	5x10 <sup>-16</sup>
<i>fyuA</i>	Iron acquisition	168 (60.2)	27	9	88	8	15	0	21	6x10 <sup>-16</sup>	5x10 <sup>-16</sup>
<i>iucC</i>	Iron acquisition	128 (45.9)	26	7	58	4	14	5	14	6x10 <sup>-5</sup>	5x10 <sup>-4</sup>
<i>iroN</i>	Iron acquisition	95 (34.1)	5	7	71	4	3	3	2	6x10 <sup>-16</sup>	5x10 <sup>-16</sup>
<i>ireA</i>	Iron acquisition	57 (20.4)	12	3	30	4	2	2	4	0.005	0.002
<i>ompT</i>	Protectin	163 (58.4)	12	14	88	8	15	5	21	6x10 <sup>-16</sup>	5x10 <sup>-16</sup>
<i>traT</i>	Protectin	139 (49.8)	27	19	51	6	16	1	19	0.002	0.4
<i>neuC</i>	Protectin	66 (23.7)	9	0	41	0	1	0	15	2x10 <sup>-12</sup>	10 <sup>-7</sup>
<i>usp</i>	Toxin	106 (38.0)	1	0	90	0	0	0	15	6x10 <sup>-16</sup>	5x10 <sup>-16</sup>
<i>sat</i>	Toxin	72 (25.8)	15	1	25	1	12	0	18	5x10 <sup>-8</sup>	0.9
<i>clbQ</i>	Toxin	55 (19.7)	0	0	55	0	0	0	0	6x10 <sup>-16</sup>	5x10 <sup>-16</sup>
<i>hlyC</i>	Toxin	41 (14.7)	0	0	39	0	2	0	0	2x10 <sup>-15</sup>	5x10 <sup>-16</sup>
<i>cnfI</i>	Toxin	33 (11.8)	0	0	33	0	0	0	0	5x10 <sup>-13</sup>	5x10 <sup>-16</sup>

596 \* Distinct alleles of the *papG* virulence gene.

597

598 **Table 3.** Distribution of resistance to 18 antibiotics in the COLIVILLE collection (n= 279)

Antibiotic classes	Antibiotics	Strains N (%)
Penicillins	Amoxicillin	83 (29.7)
	Amoxicillin – clavulanic acid	45 (16.1)
Cephalosporins	Cefoxitin	3 (1.1)
	Cefotaxime	1 (0.4)
Carbapenems	Ertapenem	0 (0)
	Imipenem	0 (0)
Aminoglycosides	Streptomycin	65 (23.3)
	Gentamicin	2 (0.7)
	Tobramycin	2 (0.7)
	Netilmicin	1 (0.4)
	Amikacin	0 (0)
Tetracyclines	Tetracycline	61 (21.9)
Sulfonamides	Cotrimoxazole	44 (15.8)
Amphenicols	Chloramphenicol	20 (7.2)
Quinolones	Nalidixic acid	17 (6.1)
	Ofloxacin	9 (3.2)
Phosphonic acids	Fosfomycin	1 (0.4)
Furans	Nitrofurantoin	1 (0.4)

633 **FIGURE LEGENDS**

634

635 Figure 1: Phenotypic and genotypic characteristics of *E. coli* strains in each of the three  
636 collections. The COLIVILLE collection is the population isolated in (and labelled) “2010”.  
637 The VDG (“1980”) and AEM (“2000”) collections are from Duriez *et al.* (2001) and Escobar-  
638 Paramo *et al.* (2004), respectively. (a) Results are presented as a proportional stacked bar  
639 graph representing the phylogenetic distribution of *E. coli*. (b) Distribution of the  
640 ‘reduced’ virulence score. The black bars within each box plot show median values. The box  
641 covers the 25<sup>th</sup> percentile to the 75<sup>th</sup> percentile of the data. Bars above and below the box  
642 show 1.5 times the inter-quartile range. Dots located at some distance outside the bars  
643 correspond to outliers. (c) Histograms representing prevalence of resistance to amoxicillin  
644 (AMX), amoxicillin-clavulanic acid (AMC), streptomycin (S), chloramphenicol (C),  
645 cotrimoxazole (SXT), tetracycline (TE) and nalidixic acid (NAL), and antibiotic resistant  
646 strains.

647

648

649

650

651

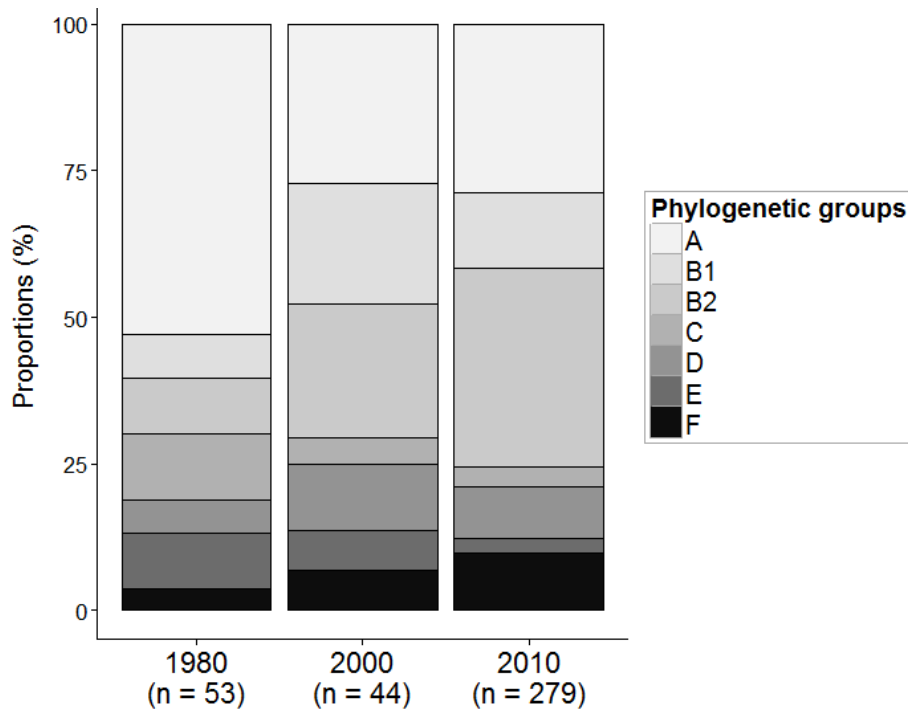
652

653

654

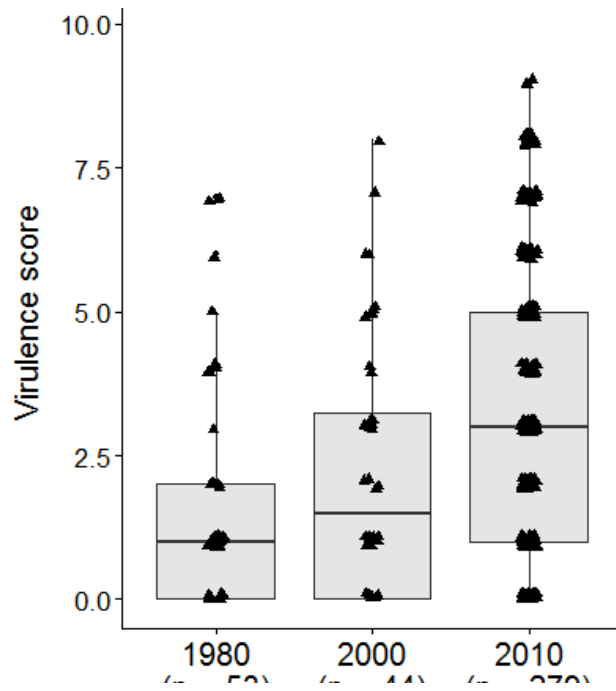


655 **Fig. 1a**



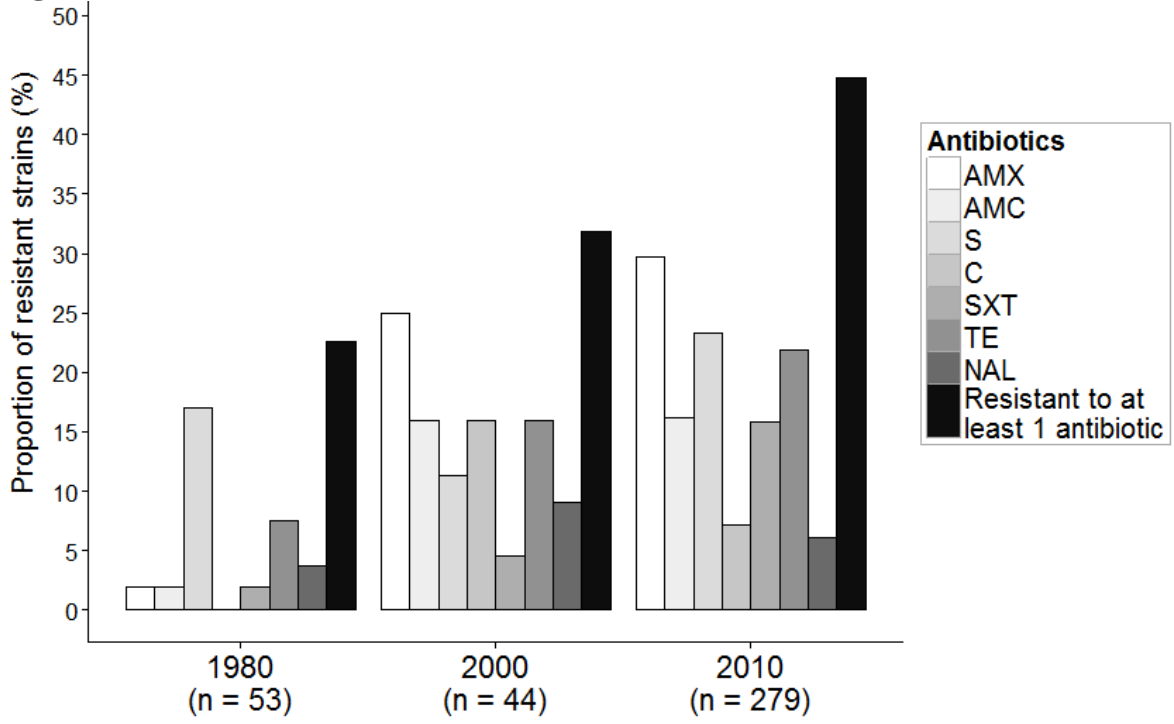
656  
657  
658

**Fig. 1b**



659  
660  
661  
662  
663  
664  
665  
666  
667

668 **Fig. 1c**



669  
670