# Commensal *Escherichia coli* strains in Guiana reveal a high genetic diversity with host-dependant population structure

Mathilde Lescat,[1,2,3] Olivier Clermont,[1,4]
Paul Louis Woerther,[5,6] Jérémy Glodt,[1,4] Sara Dion,[1,4]
David Skurnik,[5,6,7] Felix Djossou,[8] Claire Dupont,[5,6]
Gilles Perroz,[9] Bertrand Picard,[1,2,3]
François Catzeflis,[10] Antoine Andremont[5,6] and
Erick Denamur[1,4]*

[1]*UMR 722 INSERM, Paris, France.*

[2]*Université Paris Nord, PRES Sorbonne Paris Cité, Paris, France.*

[3]*Hôpital Avicenne, Assistance Publique-Hôpitaux de Paris, Bobigny, France.*

[4]*Université Paris Diderot, PRES Sorbonne Paris Cité, Paris, France.*

[5]*EA 3964 Université Paris Diderot, PRES Sorbonne Paris Cité, Paris, France.*

[6]*Centre National de Référence, Résistance bactérienne dans les flores commensales, Hôpital Bichat-Claude Bernard, Assistance Publique-Hôpitaux de Paris, Paris, France.*

[7]*Harvard Medical School, Channing Lab, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.*

[8]*Centre Hospitalier de Cayenne Andrée Rosemon, Cayenne, Guyane Française.*

[9]*UPS 2561, CNRS Guyane, Cayenne, Guyane Française.*

[10]*Laboratoire MIVEGEC, UMR IRD 224-CNRS 5290-Université Montpellier 1, Montpellier, F-34394, France.*

## Summary

**We undertook a large-scale epidemiological survey of commensal *Escherichia coli* in Trois-Sauts, an isolated village located in the south of French Guiana where human population exchanges are restricted and source of antibiotics controlled. Stools from 162 Wayampi Amerindians and rectal swabs from 33 human associated and 198 wild animals were collected in the close proximity of the village. The prevalence of *E. coli* was decreasing from humans (100%) to human associated (64%) and wild (45%) animals. A clear genetic structure between these three *E. coli* populations was observed with human strains belonging very rarely to B2 phylogroup (3.7%), exhibiting few virulence genes and bacteriocins but being antibiotic resistant whereas wild animal strains were characterized by 46.1% of B2 phylogroup belonging, with very unique and infrequent sequence types, numerous extraintestinal genes and bacteriocins but no antibiotic resistance; the human-associated animal strains being intermediate. Furthermore, an unexpected genetic diversity was observed among the strains, as the housekeeping gene nucleotide diversity per site of the Trois-Sauts's strains was higher than the one of reference strains representative of the known species diversity. The existence of such *E. coli* structured phylogenetic diversity within various hosts of a single localization has never been reported.**

## Introduction

*Escherichia coli* is a versatile bacterial species, alternatively found in its primary habitat where it is the main aerobic commensal of the gut of vertebrates (Tenaillon *et al.*, 2010), and in its secondary habitat in water and sediments (Savageau, 1983). It may also cause intestinal and extraintestinal diseases in many animal species and human (Donnenberg, 2002). Understanding what makes *E. coli* an occasionally devastating pathogen requires a better knowledge of its ecology as a commensal. *Escherichia coli* exhibits substantial clonal structure (Desjardins *et al.*, 1995), with at least seven main phylogenetic groups (A, B1, B2, C, D, E and F) (Jaureguy *et al.*, 2008; Touchon *et al.*, 2009; Moissenet *et al.*, 2010; Tenaillon *et al.*, 2010). More recently, strains indistinguishable from *E. coli* phenotypically, but highly divergent at the genetic level, have been identified and classified in five *Escherichia* clades (clades I to V) (Walk *et al.*, 2009). Numerous studies on commensal strains sampled among various human and animal populations demonstrated weak associations between host species and the prevalence of strains of the main phylogenetic groups (Gordon and Cowling, 2003; Escobar-Paramo *et al.*, 2006). Although the frequency of the genetic groups varies to

**Table 1.** Prevalence of *E. coli* for selected humans, human-associated animals and wild animals (*n* > 6).

| Population | Class | Order | Family | Species | Diet[a] | Body mass (kg)[b] | Sample size (*n*) | Prevalence (%) |
|---|---|---|---|---|---|---|---|---|
| Humans | *Mammalia* | *Hominidae* | *Primates* | *Homo sapiens sapiens* | Omnivorous | 60 | 162 | 100 |
| Human-associated animals | *Mammalia* | *Carnivora* | *Canidae* | *Canis lupus familiaris* | Omnivorous | 5 | 10 | 60 |
| Human-associated animals | *Aves* | *Galliformes* | *Phasianidae* | *Gallus gallus domesticus* | Omnivorous | 1.5 | 17 | 82 |
| Wild animals | *Mammalia* | *Rodentia* | *Echimiyidae* | *Proechimys cuvieri* | Herbivorous | 0.33 | 13 | 62 |
| Wild animals | *Mammalia* | *Rodentia* | *Muridae* | *Nectomys melanius* | Herbivorous | 0.18 | 32 | 75 |
| Wild animals | *Mammalia* | *Rodentia* | *Muridae* | *Oryzomys megacephalus* | Herbivorous | 0.03 | 14 | 43 |
| Wild animals | *Mammalia* | *Rodentia* | *Muridae* | *Oecomys bicolor* | Herbivorous | 0.028 | 17 | 41 |
| Wild animals | *Mammalia* | *Chiroptera* | *Phyllostomidae* | *Rhinophylla pumilio* | Omnivorous | 0.09 | 11 | 36 |
| Wild animals | *Mammalia* | *Chiroptera* | *Phyllostomidae* | *Carollia perspicillata* | Omnivorous | 0.017 | 9 | 11 |
| Wild animals | *Mammalia* | *Chiroptera* | *Mollossidae* | *Molossus molossus* | Carnivorous | 0.01 | 18 | 11 |

**a.** Diets were classified in three main categories: omnivorous, herbivorous and carnivorous.
**b.** Body masses were obtained by the average body mass of adult specimens per species.

some degree with the diet or body mass of the host from which they are isolated, other factors such as climate, year of sampling or the domestication status of the animals sampled (wild versus domesticated) also shape the genetic structure of *E. coli* (Gordon and Cowling, 2003; Escobar-Paramo *et al*., 2006). Likewise, *Escherichia* clades have been retrieved as commensal in digestive tracts of numerous species, with high and low prevalence in bird species and humans respectively (Clermont *et al*., 2011a).

However, these comparative epidemiological studies have only been performed among geographically distant populations sampled at various times, providing heterogeneous data with numerous biases. Thus, there is a clear need for studies considering *E. coli* in its diversity among its various hosts at a precise time within a single localization, avoiding confounding effects of geography, climate and date of sampling. Consequently, we undertook a large-scale epidemiological survey in an isolated Amerindian village, Trois-Sauts, where human population exchanges are restricted and source of antibiotics completely controlled. Humans, human-associated animals and wild animals were sampled in the close proximity of the Amerindian village. The aim of this work was to determine the phylogenetic distribution, the virulence potential and the antibiotic resistance of commensal *E. coli* strains from human and non-human vertebrate origins and to study the population genetic structure of *E. coli* in a spatially restricted area as compared with the whole species.

## Results and discussion

### *The prevalence of commensal* E. coli *in humans, human-associated animals and wild animals*

We performed two sampling campaigns in October 2006 and June 2008 in the village of Trois-Sauts, composed of four hamlets, and located south of French Guiana nearby

the Oyapock river source (Fig. S1). During these periods, we sampled 393 individuals comprising 162 adult Wayampi Amerindians, 33 human-associated animals [mainly chickens (17) and dogs (10)] living in the village and 198 wild animals [2 reptiles, 1 amphibian, 16 birds and 179 mammals mainly composed of rodents (99), chiropters (60) and marsupials (16)] living in or around the Trois-Sauts hamlets in the Amazonian forest (Fig. S1 and Table S1). Globally, the prevalence, defined as the fraction of hosts in which *E. coli* is present in the stools (Gordon and Cowling, 2003), is decreasing from humans (100%) to human associated (64%) and wild (45%) animals (Table 1). We found an effect of the body mass when comparing herbivorous rodents ($P = 0.0162$, $\chi^2$) (Table 1), the prevalence of *E. coli* increasing with the body mass.

To our knowledge, this work is, with the one of Gordon and Cowling (Gordon and Cowling, 2003), the only one to have sampled a large number of wild animals in their native environment. It largely confirms what has been observed in Australian vertebrates (Gordon and Cowling, 2003). It also demonstrates that the *E. coli* species is the most adapted to human host.

### *Phylogenetic diversity, virulence, bacteriocin and antibiotic resistance determinants of* E. coli *strains from human, human-associated animal and wild animal hosts*

Using a combination of allele-specific PCRs and multi-locus sequence typing (MLST) (see *Supporting information*), we were able to discriminate the 272 *E. coli* strains isolated from 162 humans, 21 human-associated animals and 89 wild animals (one strain per individual sampled) in seven main *E. coli* phylogenetic groups (A, B1, B2, C, D, E and F) whereas no *Escherichia* clade strain was detected (Table 2 and Fig. 1).

The phylogenetic distribution was clearly different among human strains (HS), human-associated animal strains (HAAS) and wild animal strains (WAS)

($P < 0.00001$, global $\chi^2$) (Table 2). The phylogenetic distribution within the HS was characterized by a majority of A phylogroup strains (54.3%), followed by B1 (23.5%) and D (13.6%) phylogroup strains. B2 phylogroup strains were particularly rare (3.7%), confirming the results of a precedent sampling campaign performed in 2001 in the same village (Grenet *et al.*, 2004), and contrasting with the high level of B2 group strains observed in Western populations (Tenaillon *et al.*, 2010). HAAS belong mainly to B1 phylogroup (52.4%) whereas wild animals exhibit a high proportion of B2 phylogroup strains (46.1%), followed by B1 (24.7%) and D phylogroup strains (15.7%). An influence of diet on the distribution of B2 phylogenetic group strains among animals (including human associated and wild animals) was evidenced with a high (49%) and low (20%) prevalence of B2 strains among herbivorous and omnivorous animals respectively ($P < 0.005$, $\chi^2$). Similar distribution has been observed in Australian animals with low body mass (Gordon, 2004). Indeed, among the 110 animals sampled in our work and positive for *E. coli*, only nine had body mass index over 10 kg (Table S1). The identification of the exact diet components involved in the various prevalence of the phylogroup strains is of high interest as phylogroup B2 strains have been suspected to be involved in colon cancer (Hayashi, 2006; Nougayrede *et al.*, 2006) and will benefit from experimental approaches as recently reported in rats (O'Brien and Gordon, 2011).

Extraintestinal virulence genes (*hly*, *cnf*, *aer*, *papC*, *iroN*, *traT*, *fyuA*, *sfa* and *kpsE*) were detected by PCR and an extraintestinal virulence score was determined for each strain (see *Supporting information*). This extraintestinal virulence score was higher for WAS compared with strains issued from the other populations ($P < 0.001$ in both cases, Student's test) (Table 2). More precisely, when compared with HS, WAS were characterized by higher frequency of factor coding genes *fyuA*, *papC*, *iroN*, *sfa* and *traT* ($P < 0.025$, $\chi^2$). The only exception was the *iucC* gene, which was more frequent in HS ($P = 0.0125$) (Table S2). An analysis of variance (ANOVA) of the distribution of the extraintestinal virulence scores showed a major B2 phylogroup belonging effect ($P = 2 \times 10^{-16}$), but no population effect. Thus, the higher extraintestinal virulence score of WAS is due to the higher proportion of B2 in WAS; B2 strains having more extraintestinal genes than non-B2 strains, as previously reported (Picard *et al.*, 1999).

The intrinsic extraintestinal virulence of B2 phylogroup WAS has then been evaluated in a septicaemia mouse model (Johnson *et al.*, 2006) using eight representative B2 group strains isolated from human-associated animals and wild animals (D32-HAA, S1.13-WA, S1.53, S1.56-WA, S1.74-WA, S2.2-WA, S2.19-WA, S2.21-WA) (Fig. 1 and Table S1). Four and three strains were classified as

**Table 2.** Characteristics of *E. coli* strains sampled from the humans, the human-associated animals and the wild animals.

| Population (n) | Phylogenetic group[a] | | | | | | | | Extraintestinal virulence score[b] | Intraintestinal virulence score[c] | Bacteriocin score[d] | Resistance to antibiotic score[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A n (%) | B1 n (%) | C n (%) | E n (%) | D n (%) | F n (%) | B2 n (%) | | | | | |
| Humans (162) | 88 (54.3) | 38 (23.5) | 4 (2.5) | 3 (1.9) | 22 (13.6) | 1 (0.6) | 6 (3.7) | 7.79 | 0.62 | 1.73 | 5.44 |
| Human associated animals (21) | 3 (14.3) | 11 (52.4) | 2 (9.5) | 1 (4.8) | 2 (9.5) | 0 (0) | 2 (9.5) | 4.76 | 0.95 | 0.95 | 4.61 |
| Wild animals (89) | 8 (9) | 22 (24.7) | 0 (0) | 2 (2.2) | 14 (15.7) | 2 (2.2) | 41 (46.1) | 17.98 | 1.12 | 6.07 | 0.35 |

a. Phylogenetic groups were firstly assigned to one of the seven groups A0, A1, B1, B2₂, B2₃, D₁ and D₂ using the triplex method (Clermont et al., 2000). Then, multilocus sequence typing from the Pasteur Institute scheme (Jaureguy et al., 2008) was performed on B2₂, B2₃, D₁ and D₂ strains; C group strains were identified among A₁ strains using a specific C group PCR. This procedure allows the assignation of strains among the seven phylogroups A, B1, C, E, D, F and B2.
b. Average extraintestinal virulence score was determined by calculating the average of all extraintestinal virulence scores obtained as the relative proportion of the number of detected extraintestinal virulence factors on the total number of tested extraintestinal virulence factors * 100.
c. Average intraintestinal virulence score was determined by calculating the average of all intraintestinal virulence scores obtained as the relative proportion of the number of detected intraintestinal virulence factors on the total number of tested intraintestinal virulence factors * 100.
d. Average bacteriocin score was determined by calculating the average of all bacteriocin scores obtained as the relative proportion of the number of detected bacteriocins on the total number of tested bacteriocins * 100.
e. Average resistance to antibiotic score was determined by calculating the average of all antibiotic resistance scores obtained as the relative proportion of the number of detected antibiotic resistance on the total number of tested antibiotic resistances * 100.
f. $P < 0.05$.
g. $P < 0.0001$.

**Fig. 1.** Phylogenetic trees of the 153 *E. coli* strains reconstructed from the partial sequences of eight housekeeping genes (http://www.pasteur.fr/mlst) by PHYML (Guindon *et al.*, 2005) and rooted on *E. fergusonii*. Boostrap values are indicated at the corresponding nodes only when they exceed 70%. White (H), grey (D) and black (S1 and S2) circles are indicated when the strains have human, human-associated animal and wild animal origins respectively. The ECOR strains are indicated as EC followed by the number of the ECOR strain. The reference strains ED1a, E2348/69, 536, TN03, 042 and EDL933 are included. Others representative strains are from Clermont et al. (2011b). The seven phylogenetic groups A, B1, B2, C, D, E and F are indicated on the right part of the figure. The STcs 131, 73, 92, 141, 144, 12, 14, 452 and 95, defined by the Achtman scheme (http://mlst.ucc.i.e./mlst/dbs/Ecoli) (Wirth *et al.*, 2006) and belonging to the phylogenetic group B2, correspond to the B2 subgroups I, II, III, IV, V, VI, VII, VIII and IX respectively (Le Gall *et al.*, 2007; Clermont *et al.*, 2011b). The E2348/69 strain is a representative of the EPEC-1 group. (A) Focuses on the phylogenetic relations between the strains of the phylogenetic group B2; (B) focuses on the relations between the strains of the other groups.
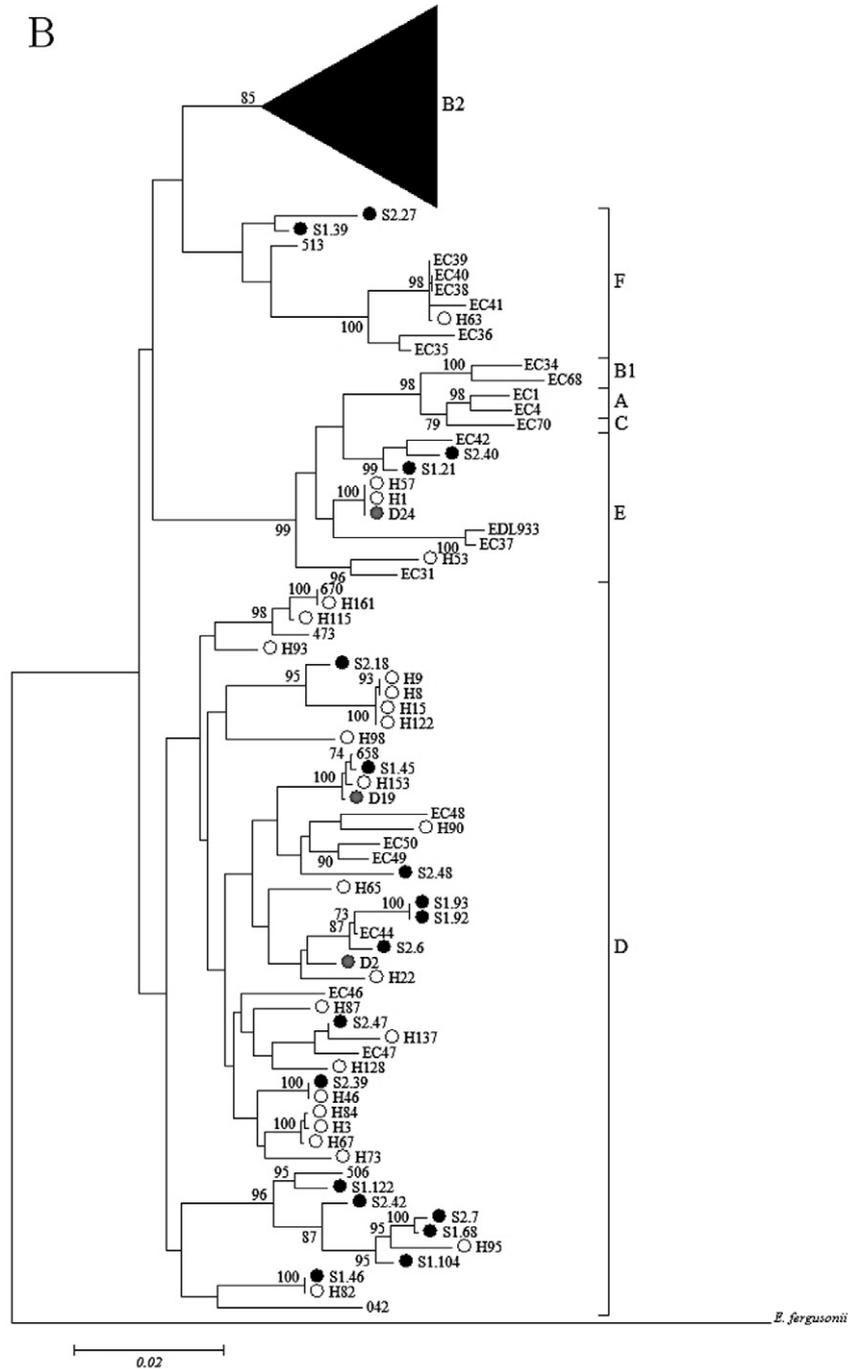
**Fig. 1.** *cont.*

killer and intermediate killer respectively (S1.13-WA, S1.53, S1.56-WA, S2.19-WA in one side and S1.74-WA, S2.2-WA and D32-HAA in the other side) and only one strain was non-killer (S2.21-WA). We did not evidenced in this set of B2 strains a link between the number of virulence genes and the number of killed mice. Such a link has been reported at the species level (Johnson *et al.*, 2000). This absence of correlation could be due (i) to the few number of studied strains and virulence determinants and/or (ii) to the fact that B2 group strains have an intrinsic property of virulence. In addition to the fact that virulence factors of *E. coli* strains may have been selected to face commensal lifestyle (the virulence by-product of commensalism hypothesis) (Le Gall *et al.*, 2007; Diard *et al.*, 2010), it could be proposed that the B2 phylogenetic background by itself has also been selected to allow an

efficient gut colonization, in accordance with the epidemiological data reported in Swedish infants (Nowrouzian *et al.*, 2005).

Bacteriocins, including colicins and microcins, were detected among the strains by a combined phenotypic and PCR-based genotyping assay and a bacteriocin score was determined (see *Supporting information*). The bacteriocin scores were higher for WAS than for HS and HAAS ($P = 0.0001$ and $P = 0.046$, respectively, Student's test) (Table 1). Specifically, colicin was significantly increased among WAS compared with HAAS and HS ($P < 0.01$, $\chi^2$ test) (Table S2), confirming a previous work comparing human and mammal animal commensal strains where colicin production was also increased in animal strains (Riley and Gordon, 1996). Among colicin coding genes, colicin B gene appeared also more prevalent in WAS than in HS ($P < 0.0001$, $\chi^2$ test) (Table S2). No significant differences were observed between microcin coding gene proportions. An ANOVA analysis of the distribution of the bacteriocin scores showed a trend for both B2 phylogroup belonging and population effects ($P = 0.05$).

Intraintestinal virulence genes representing the different intestinal *E. coli* pathotypes were detected by PCR and an intraintestinal virulence score was determined (see *Supporting information*). No significant difference in the score was noted among the three studied populations (Table 2). More precisely, *eae* was present in nine strains issued from the three populations. Among these *eae*-positive strains, four were positive for the *bfpA* gene, representing typical enteropathogenic *E. coli* (EPEC). These strains were exclusively issued from wild animal faeces (S1.6-WA, S1.11-WA, S1.76-WA and S2.1-WA). One HS (H27-H), one HAAS (D4-HAA) and one WAS (S1.88-WA) were positive for the *eltB* gene coding for the toxin LT, then classified as enterotoxinogenic *E. coli* (ETEC). Four HS were positive for *aatA* (coding for EAST1) and one HS for *afaD* genes, specific of the enteroaggregative *E. coli* (EAEC) and diffusely adherent *E. coli* (DAEC) respectively. Any strain was positive for genes specific of enterohaemorrhagic (EHEC) and enteroinvasive *E. coli* (EIEC) (Table S2). To our knowledge, this work is the first to report ETEC in wild animals, extending the animal reservoir of intestinal pathogenic *E. coli* strains.

The antimicrobial susceptibility of the strains to 32 antibiotics was determined using the disk-diffusion method and an antibiotic resistance score was determined (see *Supporting information*) (Table S3). WAS presented the lowest score compared with the two others ($P < 0.001$ in both cases, Student's test) (Table 2). Specifically, WAS were less resistant to β-lactamins (amoxicillin, amoxicillin-clavulanate, ticarcillin, piperacillin, first- and second-generation cephalosporin) and to sulfamethoxazole than HS and HAAS (all $P < 0.05$, $\chi^2$ test). An ANOVA analysis of

the distribution of the resistance scores showed a population effect ($P = 0.001$), but no B2 phylogroup belonging effect. We thus confirmed a strong anthropogenic influence on the level of antibiotic resistance, as previously proposed from collections sampled all over the world (Skurnik *et al.*, 2006). When compared with other human populations, the level of resistance in Wayampis, despite their traditional way of life associated with a lower level of hygiene compared with Western countries, is similar to those obtained in strains isolated from Western populations such as metropolitan French (Aubry-Damon *et al.*, 2004) which are submitted to comparable antibiotic pressures (1.08 treatment per subject per year) (Woerther *et al.*, 2010). Besides, resistance to antibiotics of Wayampi strains is lower than those observed in emerging countries in the dominant microbiota of individuals such as in India (Mathai *et al.*, 2008), Peru and Bolivia (Bartoloni *et al.*, 2008) where the utilization of antibiotics is uncontrolled (Morgan *et al.*, 2011).

*Multidimensional analysis of the data*

We first performed character mapping on the phylogenetic trees of the Fig. 1A and B to look at the patterns of horizontal gene transfer of virulence and resistance determinants. The virulence genes grouped in genomic islands (Bingen-Bidois *et al.*, 2002) and the antibiotic resistances encoded by mobile elements were compared according to the different clonal backgrounds (Table S4). A high level of mobility of the genetic elements was observed in the strain population as different genomic island and/or resistance gene contents were observed within a single clone, an observation made from strains originating worldwide (Escobar-Paramo *et al.*, 2004).

We then assessed the global relationships between the phylogenetic groups, the host origin (human, human-associated animal and wild animal), the presence of extraintestinal and intraintestinal virulence determinants and of bacteriocins, and the presence of resistance to antibiotics, by carrying out a factorial analysis of correspondence (FAC). The projections of the variables on the plane F1/F2, which accounted for 36.25% of the total variance, distinguished the phylogroup B2, wild animal origin, the presence of bacteriocins, the presence of extraintestinal and intraintestinal virulence determinants and the phylogroup F, which were projected on the positive values of the factor F1 from the phylogroup A, human origin, the presence of resistance to antibiotics, the absence of extraintestinal virulence determinant and bacteriocins and the phylogroup C, which were projected on the negative values of F1. Moreover, human-associated animal origin, phylogroups B1 and E were distinguished by the positive values of F2 (Fig. 2). The observed pattern is in part due to the non-random distribution of the phylo-
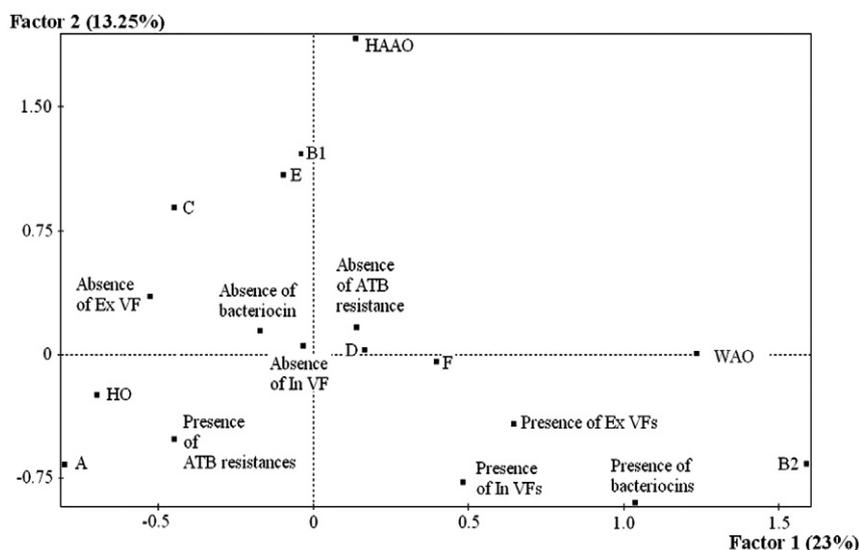
**Fig. 2.** Projections of the origin of the strains, the phylogenetic groups, the extra- and intraintestinal VFs, the presence of bacteriocins and the resistance to antibiotics detected in the 272 studied *E. coli* strains on the plane F1/F2 computed from the factorial analysis of correspondence. The abbreviations are as follows: A, B1, B2, D, C and E correspond to the phylogenetic groups; HO, HAAO and WAO correspond to human, human-associated animal and wild animal origins respectively; Ex: extraintestinal; In: intraintestinal; ATB: antibiotic; VF: virulence factor.

groups between human, human-associated and wild animals, with other characters as extraintestinal VFs driven by the phylogroup belonging.

*Phylogenetic diversity of sampled* E. coli *strains as compared with the species one*

A global visual inspection of the MLST trees (Fig. 1) revealed a huge diversity among the 96 B2, D, E and F strains studied. These strains belong to 76 distinct sequence types (STs) distributed among all the strains with a continuum of diversity. When considering the B2 strains, among the 6 human strains, one and two belonged to the frequently isolated subgroups II [ST complex (STc) 73 according to the Achtman scheme (http://mlst.ucc.ie)] and IX (STc95), respectively, and two to subgroup V (STc144) (Fig. 1) (Le Gall *et al.*, 2007). At the opposite, among the 41 B2 strains isolated from wild animal faeces, only 9 belonged to the previously described subgroups I (STc131), II, IV (STc141), V and VII

(STc14) (Le Gall *et al.*, 2007) (Fig. 1). The remaining 32 strains do not belong to any known subgroups (Le Gall *et al.*, 2007) (Fig. 1). This indicates that wild animal B2 strains correspond to particular and infrequently reported strains.

To better understand the origin of the observed diversity, we compared the ST diversity and the nucleotide diversity per site (Pi) of the 96 Trois Sauts region's strains to various strain collections (see *Supporting information*) (Table 3). We first compare our collection to the ECOR collection, which is representative of the species diversity as a whole (Ochman and Selander, 1984). Although the ST/strain ratio is higher in the ECOR collection, the Pi of the Guiana collection is significantly higher (Table 3). We then compare our collection to the Broad collection, which has been made to originate from numerous species and to a collection gathered in metropolitan France from human, domestic and wild animals (Table 3). In both cases, the Pi of the Guiana collection remains significantly higher than the other collection ones (Table 3).

**Table 3.** Diversity of the *E. coli* strains of phylogroups B2, D, E and F from various collections including the strains sampled in this work.

| Collection | Sample origin | Strains (*n*) | Host species (*n*) | ST (*n*) | Ratio host species (*n*)/strains (*n*) | Ratio STs (*n*)/strains (*n*) | Diversity per site (Pi) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| French Guiana (this work) | C[a] | 96 | 33 | 76 | 0.34 | 0.79 | 0.02655 | **b | **b | **b |
| ECOR | C/ExPEC | 31 | 9 | 28 | 0.29 | 0.90 | 0.02603 | | | |
| Broad[c] | C/ExPEC | 40 | 23 | 29 | 0.58 | 0.73 | 0.02564 | | | |
| ROAR[d] (Metropolitan France) | C | 137 | 25 | 85 | 0.18 | 0.62 | 0.02507 | | | |

**a.** C: commensal; ExPEC: extraintestinal pathogenic *E. coli.*.
**b.** $P < 0.0001$.
**c.** The Broad collection is available at http://www.broadinstitute.org/annotation/genome/escherichia_antibiotic_resistance/MultiHome.html.
**d.** This collection is composed of 38, 36 and 63 strains from human, domestic and wild animals respectively (D. Skurnik, O. Clermont, S. Brisse and E. Denamur, personal collection).

Collectively, these data show unreported strain diversity in a unique remote location in French Guiana with a host related genetic structure. More thorough analysis of the totality of the strains, as well as additional sampling campaigns in the same location, both underway, will be needed to decipher the forces shaping this structure.

## Acknowledgements

## References

Aubry-Damon, H., Grenet, K., Sall-Ndiaye, P., Che, D., Cordeiro, E., Bougnoux, M.E., *et al.* (2004) Antimicrobial resistance in commensal flora of pig farmers. *Emerg Infect Dis* **10:** 873–879.

Bartoloni, A., Pallecchi, L., Fiorelli, C., Di Maggio, T., Fernandez, C., Villagran, A.L., *et al.* (2008) Increasing resistance in commensal *Escherichia coli*, Bolivia and Peru. *Emerg Infect Dis* **14:** 338–340.

Bingen-Bidois, M., Clermont, O., Bonacorsi, S., Terki, M., Brahimi, N., Loukil, C., *et al.* (2002) Phylogenetic analysis and prevalence of urosepsis strains of *Escherichia coli* bearing pathogenicity island-like domains. *Infect Immun* **70:** 3216–3226.

Clermont, O., Bonacorsi, S., and Bingen, E. (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66:** 4555–4558.

Clermont, O., Gordon, D.M., Brisse, S., Walk, S.T., and Denamur, E. (2011a) Characterization of the cryptic *Escherichia* lineages: rapid identification and prevalence. *Environ Microbiol* **13:** 2468–2477.

Clermont, O., Olier, M., Hoede, C., Diancourt, L., Brisse, S., Keroudean, M., *et al.* (2011b) Animal and human pathogenic *Escherichia coli* strains share common genetic backgrounds. *Infect Genet Evol* **11:** 654–662.

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

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

Donnenberg, M. (2002) Escherichia coli *Virulence Mechanisms of Versatile Pathogen*. San Diego, CA, USA: Academic Press.

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

Escobar-Paramo, P., Le Menac'h, A., Le Gall, T., Amorin, C., Gouriou, S., Picard, B., *et al.* (2006) Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. *Environ Microbiol* **8:** 1975–1984.

Gordon, D.M. (2004) The influence of ecological factors on the distribution and the genetic structure of *Escherichia coli*. In *EcoSal* – Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*. Böck, A., Curtiss, R., III, Kaper, J.B., Karp, P.D., Neidhardt, F.C., Nyström, T., *et al.* (eds). Washington, DC, USA: ASM Press.

Gordon, D.M., and Cowling, A. (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* **149:** 3575–3586.

Grenet, K., Guillemot, D., Jarlier, V., Moreau, B., Dubourdieu, S., Ruimy, R., *et al.* (2004) Antibacterial resistance, Wayampis Amerindians, French Guyana. *Emerg Infect Dis* **10:** 1150–1153.

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005) PHYML Online – a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **33:** W557–W559.

Hayashi, T. (2006) Microbiology. Breaking the barrier between commensalism and pathogenicity. *Science* **313:** 772–773.

Jauréguy, F., Landreau, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G., *et al.* (2008) Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* **9:** 560.

Johnson, J.R., Kuskowski, M., Denamur, E., Elion, J., and Picard, B. (2000) Clonal origin, virulence factors, and virulence. *Infect Immun* **68:** 424–425.

Johnson, J.R., Clermont, O., Menard, M., Kuskowski, M.A., Picard, B., and Denamur, E. (2006) Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits, phylogenetic group, and ecological source. *J Infect Dis* **194:** 1141–1150.

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

Mathai, E., Chandy, S., Thomas, K., Antoniswamy, B., Joseph, I., Mathai, M., *et al.* (2008) Antimicrobial resistance surveillance among commensal *Escherichia coli* in rural and urban areas in Southern India. *Trop Med Int Health* **13:** 41–45.

Moissenet, D., Salauze, B., Clermont, O., Bingen, E., Arlet, G., Denamur, E., *et al.* (2010) 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. *J Clin Microbiol* **48:** 2459–2463.

Morgan, D.J., Okeke, I.N., Laxminarayan, R., Perencevich, E.N., and Weisenberg, S. (2011) Non-prescription antimicrobial use worldwide: a systematic review. *Lancet Infect Dis* **11:** 692–701.

Nougayrede, J.P., Homburg, S., Taieb, F., Boury, M., Brzusz-kiewicz, E., Gottschalk, G., *et al.* (2006) *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313:** 848–851.

Nowrouzian, F.L., Wold, A.E., and Adlerberth, I. (2005) *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J Infect Dis* **191:** 1078–1083.

O'Brien, C.L., and Gordon, D.M. (2011) Effect of diet and gut dynamics on the establishment and persistence of *Escherichia coli*. *Microbiology* **157:** 1375–1384.

Ochman, H., and Selander, R.K. (1984) Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* **157:** 690–693.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., *et al.* (1999) The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* **67:** 546–553.

Riley, M.A., and Gordon, D.M. (1996) The ecology and evolution of bacteriocins. *J Ind Microbiol* **17:** 151–158.

Savageau, M.A. (1983) *Escherichia coli* habitats, cell types, and moleculat mechanisms of gene control. *Am Nat* **122:** 732–744.

Skurnik, D., Ruimy, R., Andremont, A., Amorin, C., Rouquet, P., Picard, B., and Denamur, E. (2006) Effect of human vicinity on antimicrobial resistance and integrons in animal faecal *Escherichia coli*. *J Antimicrob Chemother* **57:** 1215–1219.

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

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., *et al.* (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* **5:** e1000344.

Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M., and Whittam, T.S. (2009) Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol* **75:** 6534–6544.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., *et al.* (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* **60:** 1136–1151.

Woerther, P.L., Angebault, C., Lescat, M., Ruppe, E., Skurnik, D., Mniai, A.E., *et al.* (2010) Emergence and dissemination of extended-spectrum beta-lactamase-producing *Escherichia coli* in the community: lessons from the study of a remote and controlled population. *J Infect Dis* **202:** 515–523.

## Supporting information

Additional Supporting Information may be found in the online version of this article:

**Experimental procedures.**

**Fig. S1.** Location of study site and sample collection points in Trois-Sauts, French Guiana. The study was located in a protected area where human access is restricted. Locations of human, human-associated animal and wild animal samples are indicated in white, grey and black circles, according to the Global Positioning System (GPS) collected for each sample. Circles are proportional to the numbers of sampled individuals.

**Table S1.** Main characteristics of individuals sampled during the first collect in October 2006 [humans (H), human-associated animals (HAA) and wild animals (S1-WA)] and the second collect in June 2008 [wild animals (S2-WA)] as well as the *E. coli* presence in the faeces.

**Table S2.** Presence of extraintestinal virulence factors, intraintestinal virulence factors, bacteriocins and corresponding scores of strains of *E. coli* issued from humans, human-associated animals and wild animals from both collects (2006–2008).

**Table S3.** Presence of resistance to antibiotics and resistance score of *E. coli* issued from humans, human-associated animals and wild animals from both collects (2006–2008).

**Table S4.** Character mapping of the 96 *E. coli* strains belonging to the B2, D, E and F phylogroups and ordered as in the Fig. 1.