Evidence for a human-specific *Escherichia coli* clone

Olivier Clermont,1 Mathilde Lescat,1 Claire L. O’Brien,2 David M. Gordon,2 Olivier Tenaillon1 and Erick Denamur1*  
1INSERM U722 and Université Paris 7 Denis Diderot, France.  
2School of Botany and Zoology, The Australian National University, Canberra, ACT 0200, Australia.

**Summary**

*Escherichia coli* is a widespread commensal of the vertebrate intestinal tract. Until recently, no strong association between a particular clone and a given host species has been found. However, members of the B2 subgroup VIII clone with an O81 serotype appear to be human host specific. To determine the degree of host specificity exhibited by this clone, a PCR-based assay was used to screen 723 faecal and clinical isolates from humans, and 904 faecal isolates from animals. This clone was not detected among the animal isolates, but was discovered in people living in Africa, Europe and South America. The clone is rarely isolated from people suffering from intestinal or extraintestinal disease and is avirulent in a mouse model of extraintestinal infection. Fine-scale epidemiological analysis suggests that this clone is competitively dominant relative to other members of the B2 phylogenetic group and that it has increased in frequency over the past 20 years. This clone appears to be a good candidate for use as a probiotic, and may be suitable as an indicator of human faecal contamination in microbial source tracking studies.

**Introduction**

*Escherichia coli* is a widespread commensal of the lower intestinal tract of humans and other vertebrates that occasionally causes intestinal and extraintestinal diseases (Donnenberg, 2002). Understanding what makes *E. coli* an occasionally devastating pathogen requires a better knowledge of its ecology as a commensal. A first step in gaining this knowledge would be to determine if niche specialization occurs within the species. *E. coli* colonizes a wide variety of hosts with very diverse gut morphologies and digestive physiologies. *E. coli* also exhibits substantial genetic structure, with most faecal isolates belonging to one of four genetic groups (A, B1, B and B2) (Herzer et al., 1990; Wirth et al., 2006). However, epidemiological studies have only demonstrated weak associations between host species and the prevalence of strains of the four phylogenetic groups (Gordon and Cowling, 2003; Escobar-Páramo et al., 2006). Although the frequency of the four genetic groups varies to 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*. To date, no strong association between a particular clone and a given host species has been found, except perhaps for a hly B1 clone that appears to be restricted to animals (Escobar-Páramo et al., 2006).

The great majority of *E. coli* strains belonging to genetic group B2 are highly virulent in a mouse model of extraintestinal infection (Johnson et al., 2006). Recently, as part of a detailed determination of the genetic makeup of group B2 strains, we identified an unusual B2 strain, in that it is avirulent in the mouse model of extraintestinal infection (Le Gall et al., 2007). This strain has an O81 serotype and is a member of the B2 clonal subgroup VIII as determined by multilocus sequence typing (MLST) (Le Gall et al., 2007). Preliminary epidemiological evidence suggested that this strain might be human specific. Consequently, we undertook a large-scale epidemiological survey with the goal of determining the distribution of this strain in human and non-human vertebrates across four continents.

**Results**

**Global distribution of the B2 VIII/O81 clone**

A total of 1369 strains belonging to phylogenetic group B2, representing 438 faecal and 285 clinical isolates from humans, and 646 faecal isolates from wild, domesticated and zoo animals, were screened using the O81-specific primer. An additional 258 faecal isolates from non-domesticated animals and 15 O81 strains from septicemic chickens, all of unknown phylogenetic group, were also screened. All O81-positive strains were screened using the B2 clonal subgroup VIII-specific primers and, if unknown, their phylogenetic group membership was determined.

Despite surveying strains from over 60 animal species living on four continents, none of the animal isolates were...
found to possess the O81 locus and to be members of B2 clonal subgroup VIII. By contrast, 64 members of the B2 VIII/O81 clone were detected among the human isolates. Two of the B2 VIII/O81 strains were clinical isolates; one strain (IAI48) was responsible for a urinary tract infection (Picard et al., 1999) and the other (381A) was an entero-aggregative strain responsible for a case of diarrhoea (Escobar-Páramo et al., 2004a). B2 VIII/O81 strains were isolated from humans living in Africa, the Americas and Europe, but not from people living in Australia.

Seven B2 strains from both non-human vertebrates and humans yielded an O81 polymerase chain reaction (PCR) product but were not members of the B2 clonal subgroup VIII, indicating that the O81 serotype can be found in other B2 clonal groups. Further, three O81-positive strains belonging to phylogenetic groups A and B1 were detected from mammals living in the Americas (1 group A and 2 group B1 strains), as well as 15 O81 B1 strains from chickens sampled in Spain. All of the O81 strains that were not members of the B2 VIII/O81 clone exhibited a polymorphic pattern of virulence determinants, that was, in every case, distinct from the B2 VIII/O81 clone (data not shown).

To assess degree of sequence type diversity among the B2 VIII/O81 strains, we choose six isolates from different hosts and geographical regions, as well as the two disease causing isolates (Table 1) for further characterization. We first performed MLST by sequencing six essential genes (5921 bp) in these strains. The sequences were all identical except for one strain (Ben4d) that has a synonymous polymorphism in the putP gene.

We also performed an in silico search for strains of the B2 VIII/O81 clone in the Max Planck Institute MLST database (http://web.mpiib-berlin.mpg.de/mlst/dbs/Ecoli). This database contained, at the time it was searched, data for 1404 disease and faecal isolates largely recovered from humans and domesticated animals. For this search, the allele sequence for each of the seven genes used in this MLST scheme (adk, fumC, gyrB, icd, mdh, recA and purA) was determined for the B2 VIII/O81 strain ED1a. Only one strain (M716, ST 452), an isolate from an Australian mammal, matched the profile of ED1a. However, this isolate does not yield an O81 PCR product, and M716 differs from ED1a at seven bases over the 727 nucleotides of trpA locus, including the allele-specific site of the trpAVIII.f primer, as well as in its virulence profile.

**Virulence characteristics of B2 VIII/O81 strains**

The eight B2 VIII/O81 strains were tested in an experimental mouse model of extraintestinal virulence. Whereas most B2 strains exhibit a high virulence in such a model (Johnson et al., 2006), none of the tested strains

---

**Table 1. Principal characteristics of strains representing the O81 B2 subgroup VIII clone.**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Host origin</th>
<th>Condition of isolation</th>
<th>MLST profile</th>
<th>putP allele</th>
<th>putP product</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED1a</td>
<td>France</td>
<td>Faeces</td>
<td>381A</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>VD437</td>
<td>France</td>
<td>Faeces</td>
<td>Ben4d</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Ben4d</td>
<td>Benin</td>
<td>Faeces</td>
<td>M716</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Ben24a</td>
<td>Belgium</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Ben24b</td>
<td>Benin</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Ben27a</td>
<td>Benin</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Ben27b</td>
<td>Benin</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Colombia</td>
<td>Colombia</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Colombia</td>
<td>Colombia</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Central</td>
<td>African</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Africa</td>
<td>African</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Australia</td>
<td>Australia</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>France</td>
<td>France</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>France</td>
<td>France</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>France</td>
<td>France</td>
<td>Faeces</td>
<td>VD437</td>
<td>++</td>
<td>++</td>
</tr>
</tbody>
</table>

© 2008 The Authors
had any virulence. We then screened the isolates for the presence of 19 putative virulence determinants associated with extraintestinal disease. All the strains exhibited a similar virulence gene profile (Table 1). The iron metabolism genes ireA, aer and iha were present in all strains. However, the genes fyuA and ipr2, which are markers for the high-pathogenicity island, were absent from some strains. None of the genes classically associated with intraintestinal virulence was present in the ED1a reference strain.

Fine-scale analysis of the epidemiologic data
In a sample of faecal isolates collected post 2000 from 152 people living in Colombia, Benin, and France, and considering one randomly selected isolate per individual, the B2 VIII/O81 clone represented 4% of all isolates (3–7% according to the population) and 16% of the B2 strains recovered (11–28% according to the population). By contrast, in a sample of faecal isolates collected in the 1980s from 193 people living in the USA, Mali, Croatia, France and Sweden, the B2 VIII/O81 clone represented only 0.4% of all isolates tested and 4% of the B2 strains. This result suggests that the VIII/O81 clone has increased its representation in human faeces 10-fold in the past 20 years.

To determine if the B2 VIII/O81 clone exists at a high frequency within a host we analysed the data set cited above, but used all 10 isolates taken from each of the 152 subjects. O81 was detected in 12 of the 152 hosts and, in 42% of these people, the O81 clone represented at least 80% of the E. coli isolates recovered from that host. By contrast a non-VIII/O81 B2 strain was detected in 58 hosts, but in these cases the B2 strain represented < 80% of the E. coli in that host only 17% of the time. Put differently, 77% of the O81 strains detected were found in hosts in which they were the majority clone (> 80%), whereas only 37% of non-O81 B2 strains were found in a situation where they represented the majority (> 80%) of the E. coli in a host. This outcome suggests that the B2 VIII/O81 clone is able to achieve numerical dominance in a human more often than other group B2 strains. Furthermore, in one subject, all 30 isolates taken from faeces emitted at three different times (10 isolates per faeces) over a 1-year period as well as 10 isolates from stools obtained 6 years after the first sample were the B2 VIII/O81 clone. This result demonstrates that this clone can persist in the human gut for extend periods.

Survival in water
A human-specific strain of E. coli could be a useful marker in microbial source-tracking studies that attempt to determine if human faecal contamination of a water body is occurring. Consequently, we tested the survival of ED1a and compared it with the survival of E. coli K12 (Bogosian et al., 1996). In non-sterile river water at 20°C the decline in viable cell counts of both strains are identical, and by day 9 neither strain can be detected (Fig. 1A). This result is in accord with previously reported outcomes for K12 (Bogosian et al., 1996). In sterile river water at 20°C, viable cell counts of both strains declined by only 1–2 log in 35 days, with ED1a exhibiting poorer survival than K12 (Fig. 1B). These findings are consistent with earlier suggestions that the decline of E. coli populations in water is due to predation by protozoa, phages and exposure to heat-labile toxins.

Discussion
Host-specific E. coli pathogens are not unknown, for example it has long been thought that Shigella strains are primate specific (Donnenberg, 2002) and there are rabbit-specific enteropathogenic E. coli (Leyton et al., 2007). However, this is the first report of what is apparently a human host-specific commensal strain of E. coli, although it is not known if this strain can colonize other primates. This clone belongs to the B2 phylogenetic group and, although B2 strains are often responsible for extraintestinal infection (Picard et al., 1999), strains of this clone appear to rarely cause disease. Only two pathogenic strains belonging to this clone have been isolated so far, one strain causing a urinary tract infection and an enterogaugregative strain. The B2 VIII/O81 clone is absent from the Achtman E. coli MLST database and over half of the more than 1400 isolates in this database are listed as pathogens (Wirth et al., 2006). Based on its virulence determinant pattern and/or serotype, this clone is also absent from European and American collections of B2 strains responsible for newborn meningitis (n = 99, Bona-corso et al., 2003) and urosepsis (n = 108, Johnson and Stell, 2000; Bingen-Bidou et al., 2002). Finally, none of the strains of this clone is virulent in the mouse model of extraintestinal infection.

B2 VIII/O81 clone appears to be highly successful as it has been found on three continents and, in some populations, it can represent up to 7% of the isolates taken from human faeces. The reason for the absence of this clone in humans living in Australia is unknown. Geographic isolation is unlikely to be the answer, as over 60% of the MLST sequence types isolated from Australian humans have been detected in other parts of the world (D. Gordon unpubl. data). All isolates of the B2 VIII clone with an O81 serotype are also very homogeneous in that they possess a very similar suit of genes, while in other B2 strains these genes have a highly polymorphic distribution (this work and Le Gall et al., 2007). These observations
are consistent with the idea that the clone has undergone a recent expansion.

The existence of a human-specific *E. coli* clone has potential applied applications. The B2 VIII/O81 clone could be used as a probiotic, as it appears to be avirulent and good colonizer of the human gut. The avirulent nature of this strain means that it might be particularly appropriate for use in highly vulnerable patients, such as those in intensive care, the immunocomprised, or premature infants. By comparison, the currently used probiotic strains, Nissle O6 strain (Mutaflor®) isolated in 1917 from the faeces of a healthy soldier (Sun et al., 2005) and the A0 34/86 O83 strain isolated from porcine faeces (Hejnova et al., 2005), are highly virulent in our mouse model of extraintestinal virulence. Nissle O6 killed all 10 inoculated mice and A0 34/86 O83 killed nine of the 10 mice inoculated.

Microbial source tracking could be another application of the O81/B2 subgroup VIII clone. Coliforms and *E. coli*, in particular, have been used for many years, to assess both water quality and to determine the source of any faecal pollution that might occur in a water body (Scott et al., 2002). However, the use of *E. coli* in these efforts has been hampered by the apparent lack of any extensive host specificity in *E. coli* (Gordon, 2001) and by the ability of many *E. coli* strains to undergo significant cell division in the external environment (Barnes and Gordon, 2004; Power et al., 2005). The usefulness of the microbial indicators as a tool in risk assessment can be significantly enhanced by the identification of a human-specific clone. The fact that not all individuals carry the O81/B2 subgroup VIII clone is not restrictive, as human faecal pollution is generally due to release of sewage from wastewater treatment plants or septic tanks, and therefore represents...
faecal contamination not by individuals but by populations. The inability of the B2 subgroup VIII/O81 clone to undergo significant cell division in water makes it a good candidate marker of human faecal contamination of water. Furthermore, the PCR assay that we have developed can be used to identify the clone directly from water or soil samples.

We have identified an *E. coli* clone that is an excellent candidate for further molecular investigation aimed at understanding (i) what defines a commensal *E. coli* strain and (ii) what attributes of the strain are responsible for its host specificity. Consequently, we have initiated a project whose aim is to sequence the genome of strain ED1a (http://www.genoscope.cns.fr/externe/English/Projets/Projet_MW/organisme_MW.html) in conjunction with an in-depth metabolic characterization of this strain.

**Experimental procedures**

**Bacterial strains**

Several previously published collections were studied. A collection of 293 isolates representing all of the B2 strains among 1520 isolates taken from the faeces of 152 humans (10 strains per individual) (Escobar-Páramo et al., 2004b). These strains originated from Benin (68 isolates/24 subjects), Colombia (81 isolates/18 subjects) and France (144 isolates/25 subjects) and were obtained post 2000. Twenty-five human faecal B2 isolates from 25 subjects originating from Europe (France, Sweden, Croatia), Africa (Mali) and America (USA) and sampled in the 1980s (Ochman and Selander, 1984; Duriez et al., 2001) were also studied. A collection of 378 faecal isolates (310 isolates from 87 mammals, 68 isolates from 23 birds) representing all of the B2 strains among 1930 isolates from the faeces of 405 animals (one to 10 strains per individual). These strains were sampled between 1980 and 2000 in France, Africa (Ethiopia, Cameroon, Gabon) and the Americas (French Caribbean Islands, Venezuela, USA) from domesticated and wild (Escobar-Páramo et al., 2006), as well as zoo (Ochman and Selander, 1984) animals. A collection of 55 pathogenic B2 isolates from humans (one strain per patient) encompassing enteropathogenic, enterogroupaggregative, diffusely adhering and extraintestinal pathovars (Ochman and Selander, 1984; Picard et al., 1999; Escobar-Páramo et al., 2004a). These strains originated from France, Sweden, Brazil, USA and Central African Republic, and were collected between 1980 and 2000. The collection of strains isolated from humans living in the Canberra region of Australia was acquired by isolating a single colony from the faeces or an extraintestinal body site of each person sampled (Gordon et al., 2005; Gordon and O’Brien, 2006). A total of 120 B2 faecal isolates and 230 B2 strains taken from extraintestinal body sites were examined for this study. The collection of strains from native wild Australian vertebrates was acquired by selecting a single isolate from each faecal sample and a single faecal sample was taken from each host individual (Gordon and Cowling, 2003). The numbers of Australian B2 strains examined were: mammals (*n* = 234), birds (*n* = 30) and reptiles (*n* = 4). A collection described by Souza and colleagues (1999) and consisting of 258 *E. coli* faecal isolates from animals was analysed. For the Souza collection a single isolate per host was taken from faecal material of the following host types: mammals (*n* = 197), birds (*n* = 51) and reptiles (*n* = 10) living primarily in Central and South America. Lastly, 15 O81 *E. coli* strains isolated from the hearts of 15 chickens suffering from septicaemia in Spain during the 1990s (Bianco et al., 1997) were also studied.

**PCR screening of O81 and B2 subgroup VIII strains**

The serotype of the B2 clonal subgroup VIII strain ED1a was determined via traditional serology (Dr Chobi Deb Roy, *E. coli* reference center, Pennsylvania State University, USA). To design an O81-specific primer pair, the approach described by Clermont and colleagues (2007) was used. First, the 5’ extremity of the *rfb* cluster from ED1a was sequenced and using this sequence data an O81-specific primer was designed (rfoO81b: 5’-GAGCGATATATTACTGTTG-3’). This primer, together with the gndbI.f primer (5’-ATACC GACGACGGCATCTG-3’), yields a 383 bp fragment.

Primers to confirm that a strain was a member of the B2 clonal subgroup VIII were designed for the *trp* gene. The forward primer (trpAVIII.f: 5’-GGCCAAAGAAGGCGCATTCGA-3’) is specific to B2 subgroup VIII strains and together with a non-specific *trp*A reverse primer (trpA2.r: 5’-GCAAGCGCG CCTGCGGAAG-3’) produces a 394 bp fragment. The universal forward primer targeting the 3’ portion of *trpB* (trpBA.f: 5’-CGGCCATAAGACATCTTCA-3’) which, together with the trpA2.r primer, yields a 489 bp product from all *E. coli* strains was included in the reaction as a positive control.

Polymerase chain reactions were carried out in a 20 μl volume and contained 2 μl of 10× buffer (supplied with Taq polymerase), 20 pmol of each primer, 2 μM each dNTP, 1 μl of Taq polymerase (Ozyme, St Quentin en Yvelines, France), and 3 μl of bacterial lysate. Polymerase chain reactions were performed with an Eppendorf Mastercycler with MicroAm tubes using the following amplification conditions: 4 min at 94°C, 30 cycles of 5 s at 94°C and 10 s at 58°C, with a final extension step of 5 min at 72°C. Polymerase chain reaction products were loaded on 2% agarose gels containing SYBR® Safe DNA gel stain (Invitrogen, Cergy Pontoise, France) and visualized and photographed under UV light.

**MLST analysis**

The phylogenetic relationships among the strains were inferred using nucleotide sequence data from six essential genes [*trpA*, *trpB*, *paB*, *puIP*, *icd*, and *polB* (Gerdes et al., 2003)] which are thought to experience little recombination and produce a strong phylogenetic signal (Escobar-Páramo et al., 2004c).

**Virulence gene detection**

The following 19 genes, usually considered as extraintestinal ‘virulence determinants’ and involved in capsule synthesis (*neuC*), adhesion (*papC*, *papG*, *sfa/foc*), toxin production (*hly*, *cndT*), iron metabolism (*iron*, *ireA*, *aer*, *aer*, *fyrA*, *irp2*, *fimH*), pathogenicity (*attB*, *cs1*, *cvcA*, *fimH*, *hu*, *ipaH*, *irp2*, *neuC*), were examined in the 293 B2 strains using PCR. The primer sequences, annealing temperatures and amplification products lengths are shown in Table 1. PCR products were visualized by electrophoresis in 1% agarose gel stained with ethidium bromide, cloned and sequenced. Nucleotide sequences were compared with those of *E. coli* serotypes O157:H7 and O103:H2 (EMBL/GenBank Accession No. AJ270785 and AJ270786) using the BLAST program (Altschul et al., 1990).
iha) and other traits (hra, sat, usp, ompT, ibea, malX, traT), were detected by using established simplex or multiplex PCR assays (Johnson et al., 2006). An in silico search of the complete genome of B2 subgroup VIII strain ED1a (V. Barbe, C. Médiague, E. Denamur unpubl. data) was undertaken to determine if this strain harboured genes also found in the main pathovars of diarrhoeagenic E. coli, i.e. diffusely adhering E. coli (afaBC/daaC), enterohaemorrhagic E. coli (AA probe), enterobacterial E. coli and enteropathogenic E. coli (eae), enteropathogenic E. coli (bfpA), enterotoxigenic E. coli (Sta and LT-1), enterohaemorrhagic E. coli (stx1, stx2, and ehxA) and enteroinvasive E. coli (ipaB) (Escobar-Páramo et al., 2004a).

Survival of E. coli strains in water

The river water used was obtained from the Saint Martin canal in Paris at a site about 1 km upstream from the Seine. For the sterile water studies, the water was autoclaved for 45 min. About 100 ml aliquots of water were placed into 250 ml Erlenmeyer flasks capped with paper covers. After inoculation, the flasks were incubated at 20°C under a 12:12 light : dark cycle. Indigenous bacteria were present at a level of about 2 x 10<sup>7</sup> cfu ml<sup>-1</sup> of river water. Fresh cultures of the strains K12-MG1655 and ED1a that have been grown for 14 h in 10 ml of Luria–Bertani (LB) medium were washed with sterile 0.9% saline, centrifuged, re-suspended in 1 ml and added to the water. Samples for counting were removed directly from the water, diluted and plated on Drigalsky and LB (non-sterile water) or LB (sterile water) media. In the non-sterile water experiment, the differentiation between indigenous and inoculated bacteria was based on colony morphology and O81 PCR assay. All experiments were carried out in duplicate.

Mouse lethality assay

A mouse model of systemic infection was used to assess the intrinsic virulence of the O81 strains (Picard et al., 1999). For each strain, 10 outbred female Swiss OF1 mice (3–4 weeks old, 14–16 g) were challenged subcutaneously in the abdomen with a standardized bacterial inoculum (10<sup>5</sup> cfu ml<sup>-1</sup> of log-phase bacteria in 0.2 ml Ringer solution). Mortality was assessed over 7 days post challenge. The urosepsis strains CFT073 was used as a positive control and the falciparum-derived strain K12-MG1655 as a negative control. Data were available for three strains (ED1a, Ben4d, IAI48) (Le Gall et al., 2007). In this model system, lethality is a rather clear-cut parameter and strains were usually classified as non-killer (strains killing < 2 mice out of 10) or killer (strains killing > 8 mice) (Johnson et al., 2006).

Acknowledgements

We are grateful to Maryvonne Moulin-Schouleur and Jorge Blanco for providing the chicken O81 strains and to Valeria Souza for some animal strains. This work was partially funded by the ‘Fondation pour la Recherche Médicale’.

References


