Both Urinary and Rectal *Escherichia coli* Isolates Are Dominated by Strains of Phylogenetic Group B2

Lixin Zhang,* Betsy Foxman, and Carl Marrs

Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan

Received 27 March 2002/Returned for modification 25 May 2002/Accepted 6 August 2002

To compare the genetic structures of uropathogenic and commensal *Escherichia coli* populations, a total of 181 urinary and rectal *E. coli* isolates were classified into intraspecies phylogenetic groups by PCR amplifications of phylogenetic markers. The genetic variability of these isolates within phylogenetic groups was further assessed by enterobacterial repetitive intergenic consensus (ERIC) typing. The distributions of 10 known virulence factors were also examined. In contrast with most reports, phylogenetic group B2 not only accounted for the majority of urinary isolates from young women with urinary tract infections (69%) but also was the dominant group among the rectal isolates from healthy young women (48%). Such difference may be explained by geographic variation, difference in host population characteristics, or differences in sampling method, or a combination of the three. Strains with known virulence factors most frequently belonged to phylogenetic groups B2 and D. Additionally, group B2 and D rectal isolates were more heterogeneous than urinary isolates. Two subclusters existed within group B2 strains by ERIC typing. These subclusters were not evenly distributed between rectal and urine isolates and differed in virulence gene distribution.

*Escherichia coli* is a commensal organism of humans and other warm-blooded animals. It can also cause various diseases, both intestinal and extraintestinal, in these hosts. Variations in genetic backgrounds and the presence or absence of specialized virulence factors in the bacteria may contribute, in part, to the commensalism-versus-virulence duality of *E. coli*. *E. coli* populations have a clonal structure (12, 23), and indeed, various intestinal or extraintestinal *E. coli* infections have been linked to specific clones or groups of related clones (1, 17, 18, 24). Phylogenetic studies have shown that *E. coli* can be divided into four main phylogenetic groups, designated A, B1, B2, and D (12). Most *E. coli* strains responsible for urinary tract infections (UTI) and other extraintestinal infections belong to group B2 or, to a lesser extent, to group D (11, 13, 19). In addition, pathogenic *E. coli* strains are often marked by the presence of special virulence determinants. For example, uropathogenic *E. coli* strains are more likely to have P pili, S pili, afimbrial adhesin, and toxins such as hemolysin and cytotoxic necrotizing factor 1 (14, 16). Overall, strains of phylogenetic group B2 and D often carry virulence determinants that are lacking in group A and B1 strains (13, 19).

While many studies have examined the population structure of pathogenic isolates extensively, little attention has been paid to the structure of commensal strain populations. Since bowel flora is considered the natural reservoir of pathogenic strains in extraintestinal infections (7), the phylogenetic distribution of commensal *E. coli* isolates from healthy humans could provide an important comparison and insight on the spread of the potential pathogenic lineage. Limited reports have indicated that the group B2 *E. coli* strains that are responsible for most extraintestinal infections are rare in fecal samples (6, 20). This implies that acquiring the group B2 strain is important in the risk of infection. Duriez et al. recently examined phylogenetic groupings of commensal *E. coli* isolates from three geographic areas (6). They reported that commensal isolates are dominated by strains of A and B1 groups, with relatively fewer B2 strains. However, the observed difference in distribution of *E. coli* phylogenetic groups between pathogenic and commensal *E. coli* populations was based on comparing fecal and urine isolates from different host population cohorts (6). Such epidemiologically mismatched comparison groups may introduce the risk of sampling bias and potential confounding effects.

To further our understanding of the role of commensal enteric *E. coli* as the potential reservoir for UTI, and to better compare genetic relationships between pathogenic and commensal *E. coli* populations, we examined the population structure of both uropathogenic and commensal rectal isolates from the same human cohort of college-age women. We identified the phylogenetic groupings of the *E. coli* isolates and compared their distributions among collections. We further analyzed the *E. coli* isolates by enterobacterial repetitive intergenic consensus (ERIC) typing to compare the genetic diversity of strains within phylogenetic groups between urinary and rectal collections. The distributions of known virulence factors among phylogenetic groups were also studied.

**MATERIALS AND METHODS**

**Bacterial strains.** We randomly sampled *E. coli* isolates (*n* = 181) from two collections obtained from previous epidemiologic studies (8, 9, 16): urinary isolates from Michigan women aged 18 to 39 years with first UTI (93 of 315 women) and rectal isolates from healthy Michigan women aged 18 to 39 years without current UTI (88 of 408 women). All isolates had been previously screened for the presence of 10 virulence factors potentially associated with UTI: adhesins of the Dr family (drb), aerobactin (aer), group II capsule (kpsMT), group III capsule (capIII), cytotoxic necrotizing factor 1 (cnf1), alpha-hemolysin (hly), outer membrane protease T (ompT), members of the P fimbral family (gf), S fimbral adhesin (sfa), and type 1 pili (fim) (8, 16).

**Phylogenetic classification.** The phylogenetic grouping of the *E. coli* isolates was determined by a PCR-based method recently developed by Clermont et al. (5). This simple and rapid phylogenetic grouping method, which uses a combi-
nation of three DNA markers (chaA, yjaA, and DNA fragment tspE4C2), can accurately group E. coli isolates into one of four major phylogenetic lineages. As described by Clermont et al. (5'-GAGCTCCCAATTTACGTAAC GACCGTCGATTAG-3') and ChaA2 (5'-TGGCCTAGTACCAAAGACACG-3'), YjaA1 (5'-TGAATGTTTCGAGGAGCTGCAAC-3') and YjaA2 (5'-ATGGAG AATGCTTTCCTCAAAC-3'), and TspE4C2.1 (5'-GAGTAATGCTCGGCACT TCA-3') and TspE4C2.2 (5'-CGCGCCCAAGAATGTACGTA-3'), which generate 279-211-, and 152-bp fragments, respectively. We performed a two-step triple-PCR using the above six primers in a single reaction in a Perkin-Elmer GeneAmp 9700 thermal cycler under the following conditions: denaturation for 4 min at 94°C, 30 cycles of 5 s at 94°C and 10 s at 59°C, and a final extension step of 5 min at 72°C. The PCR products were analyzed on a 2% agarose gel. If the strain was positive for chaA, then the presence of yjaA was examined. The strain belonged to group B2 if it was positive for yjaA; otherwise, it was assigned to D group. If a strain was negative for chaA, then the presence of tspE4C2 was examined. The strain belonged to B1 group if it was positive for tspE4C2; otherwise, it was assigned to group A.

ERIC sequence typing. Identifying the ERIC sequences among bacterial genomes was used to examine the genetic diversity of E. coli strains. PCR amplifications of ERIC sequences were performed on all E. coli isolates using a modification of a protocol described previously (22). Each reaction mixture included 25 pmol of ERIC-2 primer (5'-AAGTAATGCTGGGATTGACG-3'), 0.4 mM concentration of each deoxynucleoside triphosphate, 5 mM MgCl2, 10 ng of DNA, 2 U of Platinum Taq DNA polymerase, and 1X PCR buffer (Gibco, BRL Life Technologies, Inc., Gaithersburg, Md.). A negative control without DNA was included in each batch of reactions. Instead of the original protocol's PCR conditions of 7 min of denaturation followed by 30 cycles of 30 s at 90°C, 1 min at 52°C, and 8 min at 65°C, our PCR was performed using the following conditions after a 2-min denaturation period at 94°C: 35 cycles of 30 s at 94°C, 1 min at 57°C, and 4.5 min at 72°C, with a final extension of 1 min at 72°C. Products were electrophoresed on a 2% agarose gel at 45 mA for 5 h, stained with Vistra Green (Amersham Life Science Products, Arlington Heights, Ill.) for 3 h, and visualized with a Storm PhosphorImager (Molecular Dynamics, Sunnyvale, Calif.). Three reference lanes were included per gel to facilitate measurement across gels.

Data analysis. Captured ERIC gel images were processed using BioNumerics software (Applied Maths, Ghent, Belgium). Similarity matrices were constructed on the basis of Pearson's correlation coefficient analysis of pairwise comparisons of ERIC patterns. We performed clustering analysis and constructed a dendrogram with the unweighted pair group method using arithmetic averages based on the similarity matrices (21). Differences in proportions among collections or phylogenetic groups were tested using χ² or Fisher's exact test. Differences in means were compared using Student's t test. All analyses were carried out using the SAS (Cary, N.C.) system, version 6.12.

Informed consent was obtained from all study participants. All study protocols were approved by the Institutional Review Board at the University of Michigan School of Public Health.

RESULTS

Distribution of urinary and rectal E. coli isolates among four phylogenetic groups. Among the total of 181 E. coli isolates analyzed, 106 (58.6%) belonged to phylogenetic group B2 (Table 1). Of the remaining isolates, 36 (19.9%), 25 (13.8%), and 14 (7.7%) isolates belonged to the D, A, and B1 groups, respectively. There was noticeable variation in the distribution of the phylogenetic groups between the first-UTI and rectal specimen collections. Although these two collections came from college-age women in the same geographic area, the first-UTI specimen collection had a higher proportion of group B2 strains (69 versus 48%; P = 0.004), a lower proportion of group A (8 versus 21%; P = 0.012) and B1 (3 versus 13%; P = 0.02) strains, and a similar proportion of group D strains as the rectal specimen collection. Group B2 strains were the most common and group B1 strains were the least common in both the UTI and rectal specimen collections. About 25% of the rectal isolates were from women without a history of UTI. However, there was no apparent difference in the distribution of the four phylogenetic groups between rectal isolates collected from women with or without a history of UTI.

Genetic comparison between urinary and rectal isolates of phylogenetic B2 and D lineage using ERIC. ERIC PCR analysis was used to further examine the genetic diversity of strains within each phylogenetic group from each collection. The average similarity of all pairwise Pearson's correlation coefficient analyses of ERIC patterns was calculated for each group. The average similarity gives a rough indicator of the diversity of the strains within a phylogenetic group. While the numbers of group A and B1 strains in each collection were too small to make meaningful comparisons, group B2 strains from the first-UTI specimen collection as a whole were more homogeneous than group B2 strains from the rectal specimen collection (56.8 versus 45.9%; P < 0.001), as were the group D strains in the rectal specimen collection (37.8 versus 49.6%; P < 0.001).

In the resulting ERIC dendrogram of all group B2 strains from both the rectal and the first-UTI specimen collections, these strains were clearly segregated into two large clusters (subclusters I and II) as indicated in Fig. 1. While strains in clusters I and II had less than 35% similarity, strains in cluster II were more homogeneous, with a minimal similarity between any two strains of more than 70%. Strains in cluster I were more heterogeneous but still were more similar to each other than to cluster II strains. These two clusters were not evenly distributed between UTI and rectal isolates. Out of the 43 cluster II strains, 30 (70%) were UTI isolates and only 13 (30%) were rectal isolates, while for the 63 cluster I strains, 34 (54%) were UTI isolates and 29 (46%) were rectal isolates. Furthermore, cluster II strains accounted for a larger proportion of the group B2 strains isolated from the first-UTI specimen than of those from the rectal specimen collection (47 versus 31%; P = 0.104).

Phylogenetic distribution of virulence genes among urinary and fecal isolates. The distribution of 10 previously screened virulence genes among four phylogenetic groups in all three collections was examined (Table 2). Most of these virulence genes were approved by the Institutional Review Board at the University of Michigan School of Public Health.

Informed consent was obtained from all study participants. All study protocols were approved by the Institutional Review Board at the University of Michigan School of Public Health.
FIG. 1. Dendrogram from ERIC analysis of 106 phylogenetic group B2 $E.\ coli$ isolates from both first-UTI and rectal specimen collections. It is constructed with the unweighted pair group method using arithmetic averages with similarity matrices generated based on Pearson’s correlation coefficient analysis of pairwise comparisons of ERIC patterns. Two major clusters are identified and labeled as I and II. The shaded triangles are collapsed branches in which all isolates are from the same collection (either urine or rectal). Numbers in parentheses are the number of isolates in each collapsed branch. The scale bar on the top indicates the similarity values. Abbreviations: R, isolates of rectal origin; U, isolates from urine.
TABLE 2. *E. coli* strains of the four phylogenetic groups carrying virulence genes

<table>
<thead>
<tr>
<th>Group (no. of isolates)</th>
<th>fim</th>
<th>pff</th>
<th>sfa</th>
<th>aer</th>
<th>kpsMT</th>
<th>ompT</th>
<th>hly</th>
<th>cnf1</th>
<th>drb</th>
<th>capIII</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First UTI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A (7)</td>
<td>7 (100)</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>3 (21)</td>
<td>3 (21)</td>
<td>1 (14)</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>B1 (3)</td>
<td>3 (100)</td>
<td>1 (33)</td>
<td>0 (0)</td>
<td>1 (33)</td>
<td>0 (0)</td>
<td>2 (67)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>D (19)</td>
<td>19 (100)</td>
<td>8 (42)</td>
<td>0 (0)</td>
<td>12 (63)</td>
<td>13 (63)</td>
<td>15 (79)</td>
<td>1 (5)</td>
<td>0 (0)</td>
<td>3 (16)</td>
<td>2 (11)</td>
</tr>
<tr>
<td>B2 (64)</td>
<td>64 (100)</td>
<td>34 (53)</td>
<td>26 (41)</td>
<td>30 (47)</td>
<td>60 (93)</td>
<td>63 (98)</td>
<td>29 (45)</td>
<td>26 (41)</td>
<td>14 (22)</td>
<td>3 (5)</td>
</tr>
<tr>
<td><strong>Rectal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A (18)</td>
<td>17 (94)</td>
<td>4 (22)</td>
<td>0 (0)</td>
<td>6 (33)</td>
<td>4 (22)</td>
<td>4 (22)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (11)</td>
</tr>
<tr>
<td>B1 (11)</td>
<td>11 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (18)</td>
<td>3 (27)</td>
<td>4 (36)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>D (17)</td>
<td>17 (100)</td>
<td>7 (41)</td>
<td>0 (0)</td>
<td>7 (41)</td>
<td>14 (82)</td>
<td>14 (82)</td>
<td>1 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (6)</td>
</tr>
<tr>
<td>B2 (42)</td>
<td>42 (100)</td>
<td>18 (43)</td>
<td>11 (26)</td>
<td>20 (48)</td>
<td>41 (98)</td>
<td>42 (100)</td>
<td>9 (21)</td>
<td>6 (14)</td>
<td>2 (5)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

DISCUSSION

We described the distributions of 93 uropathogenic strains and 88 commensal rectal strains from healthy women among the four main phylogenetic groups of *E. coli*. While group B2 strains dominated in the first-UTI specimen collection, they also accounted for nearly half of the rectal specimen collection. In addition, we examined genetic variability within each phylogenetic group using ERIC typing and showed that group B2 and D strains of UTI origin were genetically less diverse than those of rectal origin. Furthermore, two clusters were found among group B2 strains, and one of the clusters (cluster II) was more associated with those from urine. Known virulence genes were most often associated with group B2 and D strains.

Several recent studies suggest that extraintestinal pathogenic *E. coli* strains are mostly derived from phylogenetic group B2 (2, 3, 13, 19). In a study among 118 strains isolated from patients with meningitis and other miscellaneous extraintestinal infections, Duriez et al. found that 85% (72%) of them belonged to group B2 (6). Johnson et al. showed that 119 of the total 182 (65%) *E. coli* strains isolated from patients with bacteremia were of B2 origin (15). In our study, 69% of UTI isolates were B2 strains, which is very similar to the results of these other studies. Based on all of this information, it may be safe to estimate that group B2 strains account for about two-thirds of all extraintestinal *E. coli* infections.

Although group B2 strains were found less frequently in our rectal specimen collection than in the first-UTI specimen collection, they were still the most common group. This is in sharp contrast to other available reports, in which B2 group strains were usually the rarest. Picard et al. found that carboxylesterase B2 type (which corresponds to phylogenetic group B2) strains accounted for only 9% of examined commensal human strains (20). In a recent examination of commensal *E. coli* isolates in Mali, France, and Croatia using the same PCR-based phylogenetic grouping method, the frequencies of B2 isolates were found to be 2% (1 of 55), 10.5% (6 of 56), and 19% (11 of 57), respectively (6). The sharp contrast between our results and previous results may be due to geographic variation, differences in host population characteristics, differences in sampling methods, or a combination of the three. Our collection was obtained in a single city in the state of Michigan, while the above-cited studies were done in Europe. The first-UTI and rectal isolates studied here came from the same host population, college-age women. As this age group has the highest incidence of UTI (10), the proportion asymptptomatically carrying uropathogenic *E. coli* may be higher as well. Moreover, our commensal *E. coli* strains were isolated from rectal swabs rather than stool samples as in other studies. As the normal intestinal tract can carry a mixture of different *E. coli* strains, the dominant isolate from stool may be different from the dominant isolate at the rectal opening. How frequently the virulent B2 pathogens are routinely carried by healthy humans will affect the disease dynamic in the population. Further studies should investigate whether populations with higher rates of bowel carriage of group B2 strains have higher rates of UTI incidence.

The high percentage of B2- and D-group strains, the perceived pathogenic *E. coli* strains, in our rectal specimen collection also prompted us to ask whether B2- or D-group strains of rectal origin were similar to those from urine. ERIC analyses of B2- and D-group strains from first-UTI and fecal spec-
imen collections revealed that urine B2- and D-group isolates were not a random subset of those present rectally. B2 and D strains of first-UTI specimens were genetically more homogeneous than their fecal counterparts. This finding is similar to an analysis by Caugant et al. (4) of both commensal and UTI isolates by multilocus enzyme electrophoresis (MLEE). They demonstrated that UTI isolates had less genetic diversity than did fecal samples measured either at the level of individual structural genes or in terms of clonal diversity represented by MLEE types. Among MLEE types identified, certain MLEE types were more often represented in UTI samples and others were underrepresented in UTI samples. In our clustering analysis of the ERIC data, two subclusters, subclusters I and II, within group B2 strains were clearly present. Cluster II strains were more likely to be isolated from first-UTI specimens than from rectal specimens and were more likely to carry the most well known UTI virulence factors, sfa, hly, cnf1, and pff (class III). Such findings suggest that there is a genetic diversity within group B2 strains and that the strains are probably not equally pathogenic. It is possible that a healthy human population can have a high rate of B2 strain carriage but with less virulent B2 subclasses. Further studies are needed to examine the B2 strains in more detail in comparison to their commensal counterparts to further refine the definition of pathogenic E. coli. Studying the genetic relationships between pathogenic isolates within the genetic diversity of the commensal E. coli reservoir is essential for deciphering the molecular basis for the pathogenesis.

ACKNOWLEDGMENTS

We appreciate and thank Patricia Tallman for technical assistance and support.

This work was supported by NIH grants R01 DK35368 (B.F.) and R01 DK39434 (C.M.)

REFERENCES


