The Epidemic of Extended-Spectrum-β-Lactamase-Producing *Escherichia coli* ST131 Is Driven by a Single Highly Pathogenic Subclone, H30-Rx

Lance B. Price,‡ James R. Johnson,‡ Malika Aziz,‡ Constance Clabots,‡ Brian Johnston,‡ Veronika Tchesnokova,§ Lora Nordstrom,‡ Maria Billig,§ Sujay Chattopadhyay,‡ Marc Stegger,‡ Paul S. Andersen,‡ Paul R. Pearson,§ Kim Riddell,§ Peggy Rogers,§ Delia Scholes,§ Barbara Kahl,§ Paul Keim,‡,§ Evgeni V. Sokurenko‡

Division of Pathogen Genomics, the Translational Genomics Research Institute, Flagstaff, Arizona, USA; Department of Occupational and Environmental Health, George Washington University, Washington, DC, USA; Veterans Affairs Medical Center and University of Minnesota, Minneapolis, Minnesota, USA; Department of Microbiology, University of Washington School of Medicine, Seattle, Washington, USA; Microbiology and Infection Control, Statens Serum Institut, Copenhagen, Denmark; Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, Arizona, USA; Group Health Clinical Laboratory, Group Health Cooperative, Seattle, Washington, USA; Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA; Institute of Medical Microbiology, Universitätshospital Münster, Münster, Germany

L.B.P. and J.R.J. contributed equally to the project.

**ABSTRACT** The *Escherichia coli* sequence type 131 (ST131) clone is notorious for extraintestinal infections, fluoroquinolone resistance, and extended-spectrum beta-lactamase (ESBL) production, attributable to a CTX-M-15-encoding mobile element. Here, we applied pulsed-field gel electrophoresis (PFGE) and whole-genome sequencing to reconstruct the evolutionary history of the ST131 clone. PFGE-based cluster analyses suggested that both fluoroquinolone resistance and ESBL production had been acquired by multiple ST131 sublineages through independent genetic events. In contrast, the more robust whole-genome-sequence-based phylogenomic analysis revealed that fluoroquinolone resistance was confined almost entirely to a single, rapidly expanding ST131 subclone, designated H30-R. Strikingly, 91% of the CTX-M-15-producing isolates also belonged to a single, well-defined clade nested within H30-R, which was named H30-Rx due to its more extensive resistance. Despite its tight clonal relationship with H30Rx, the CTX-M-15 mobile element was inserted variably in plasmid and chromosomal locations within the H30-Rx genome. Screening of a large collection of recent clinical *E. coli* isolates both confirmed the global clonal expansion of H30-Rx and revealed its disproportionate association with sepsis (relative risk, 7.5; *P* < 0.001). Together, these results suggest that the high prevalence of CTX-M-15 production among ST131 isolates is due primarily to the expansion of a single, highly virulent subclone, H30-Rx.

**IMPORTANCE** We applied an advanced genomic approach to study the recent evolutionary history of one of the most important *Escherichia coli* strains in circulation today. This strain, called sequence type 131 (ST131), causes multiresistant bladder, kidney, and bloodstream infections around the world. The rising prevalence of antibiotic resistance in *E. coli* is making these infections more difficult to treat and is leading to increased mortality. Past studies suggested that many different ST131 strains gained resistance to extended-spectrum cephalosporins independently. In contrast, our research indicates that most extended-spectrum-cephalosporin-resistant ST131 strains belong to a single highly pathogenic subclone, called H30-Rx. The clonal nature of H30-Rx may provide opportunities for vaccine or transmission prevention-based control strategies, which could gain importance as H30-Rx and other extraintestinal pathogenic *E. coli* subclones become resistant to our best antibiotics.

**Horizontal gene transfer is one of the most powerful forces in bacterial evolution. The transformative potential of this process is perhaps best exemplified by the acquisition of antimicrobial resistance determinants: in a single genetic event, an antimicrobial-susceptible bacterium can acquire a complex suite of resistance determinants and become resistant to multiple antimicrobials. Thus, frequent horizontal transfer between different strains can potentially drive the spread of antibiotic resistance within the bacterial population, without any change in the distribution of strains. However, when virulent bacterial clones acquire such elements, they can emerge rapidly within the population through clonal expansion and thereby gain local or even global predominance (1–3). Quantifying the relative contribution of horizontal (gene transfer) and vertical (clonal expansion) mech-
FIG 1 PFGE dendrogram and whole-genome SNP-based phylogeny of E. coli ST131. (A) PFGE-based dendrogram of E. coli ST131 isolates (n = 524), as inferred within BioNumerics according to the unweighted pair group method based on Dice similarity coefficients. (B) Whole-genome SNP-based phylogeny of selected ST131 isolates (n = 105) and the NA114 reference genome. SNPs were identified from genomic regions equivalent to approximately 44.7% of the reference genome that was shared among all isolates and sequenced at ≥10× coverage. Analysis of these shared genomic regions revealed 2,531 parsimony-informative and 4,000 total SNPs from the core genome (excluding horizontally acquired regions) that were used to construct the phylogeny presented here. Homoplasy index (HI) = 0.012. The purple block highlights the H30 subclone.
multaneously, suggesting independent transfer events (33). This, together with the substantial diversity in transferable resistance elements in ST131, has led some to conclude that horizontal transfer must be the dominant mechanism whereby ESBLs have gained prominence among strains of the ST131 clone (7, 34, 35). However, other evidence suggests that clonal expansion contributes significantly to the spread of antimicrobial resistance within E. coli (36–39).

Until recently, our knowledge of the epidemiology and dispersal of bacterial strains, including of ST131 origin, has been based...
largely on multilocus sequence typing (MLST), which has limited resolution at the subclone level, and on pulsed-field gel electrophoresis (PFGE) analysis, which is highly vulnerable to distortions from horizontal gene transfer events and subjective interpretation. In the current study, we used whole-genome single-nucleotide polymorphism (SNP) analysis to reconstruct the ST131 phylogeny and then overlaid resistance determinants and phenotypic susceptibility on this phylogeny to elucidate the evolutionary history of fluoroquinolone resistance and ESBL production within this prominent pathogen.

RESULTS

PFGE analyses. A collection of 524 ST131 isolates cultured from humans and animals between 1967 and 2011 was analyzed using PFGE, which yielded a complex dendrogram (Fig. 1A). Within the PFGE-based dendrogram, the isolates that were fluoroquinolone resistant and/or blaCTX-M-15 positive, although exhibiting some clustering, were intermingled extensively with those that were fluoroquinolone susceptible and/or blaCTX-M-15 negative. As such, the PFGE analysis supported previous reports suggesting frequent horizontal acquisition of fluoroquinolone resistance determinants and blaCTX-M-15 among different ST131 subclones.

Whole-genome SNP-based phylogenetic reconstruction of ST131. From the total collection that underwent PFGE analysis, 105 ST131 isolates were systematically selected for genome sequencing according to prespecified criteria that emphasized diversity of genetic backgrounds according to PFGE. The 105 isolates, which derived from five countries and 23 states and provinces in Canada and the United States, included 22 CTX-M-15-producing isolates, which were widely distributed across the PFGE dendrogram (Fig. 1A).

Genomic comparisons identified SNP loci that were present in all isolates and, therefore, informative for phylogenetic reconstruction. The first phylogenetic tree included non-ST131 strain AA86 (group B2; ST176) (40) as an outgroup, to root the tree and to identify the basal clones within the ST131 phylogeny (see Fig. S1 in the supplemental material). Next, strain AA86 was excluded, and a new SNP matrix and phylogenetic tree were generated (see Fig. S2 in the supplemental material). Since (distant) strain AA86 lacks some of the genomic regions found within the ST131 clone, exclusion of AA86 increased the number of shared nucleotides in the sequence alignment and, therefore, increased the number of informative SNPs with which to resolve the ST131 phylogeny.

The homoplasy index (HI) for these two initial trees (see Fig. S1 and S2) was exceedingly high (>3.33), indicating substantial recombination. Phylogenetic reconstructions that include genomic regions acquired by horizontal gene transfer will not accurately represent the evolutionary history of clonal organisms. However, such phylogenies can be used to identify the regions acquired horizontally. This was accomplished here by mapping to the reference genome the HI values for individual SNPs, which revealed four large recombinant regions representing nearly 31% of the genome.

Exclusion of SNPs from the four horizontally acquired regions resulted in trees with minimal homoplasy (homoplasy index [HI] = 0.012) (see Fig. S3 in the supplemental material), suggestive of highly accurate phylogenies (41). Figure 1B shows the resultant whole-genome SNP phylogeny for the 105 ST131 isolates, plus the strain NA114 reference ST131 genome (42).

Whole-genome-based clustering of resistant subclones. The whole-genome SNP-based phylogeny showed distinct clustering of strains carrying specific fimH alleles (Fig. 1B), as well as gyrA and parC alleles and O type (see Dataset S1 and Table S1 in the supplemental material). In particular, strains carrying the fimH30 allele clustered as a single low-diversity clade, designated H30, which included 58 (95%) of the 61 fluoroquinolone-resistant isolates. Moreover, nearly all of the CTX-M-15-producing isolates, despite appearing to have diverse genetic backgrounds according to the PFGE-based dendrogram (Fig. 1A), collapsed into a distinct subclade within the H30 clade (Fig. 1B).

To further resolve the evolutionary history of the H30 subclone, genomic sequences from the 64 H30 isolates and their three nearest neighbors were analyzed separately from the rest of the isolates (Fig. 2). Aligning these sequences to the finished NA114 reference genome increased the number of shared nucleotides and revealed additional informative SNPs that were used to generate the high-resolution and highly accurate (HI = 0.000) phylogenetic tree shown in Fig. 2. This tree suggested that acquisition of the fimH30 allele preceded the acquisition of fluoroquinolone resistance by a single ancestor within the H30 subclone, which was followed by a large clonal expansion of fluoroquinolone-resistant H30 strains. To distinguish the clonally related fluoroquinolone-resistant H30 isolates from the ancestral fluoroquinolone-susceptible H30 isolates, the resistant subclone within H30 was designated H30-R.

In this high-resolution phylogeny, 20 (91%) of the 22 ST131 isolates that carried blaCTX-M-15—including isolates from Australia, South Korea, Portugal, Canada, and the United States—formed a distinct, single-ancestor subclone within H30-R. Because of its more extensive resistance characteristics, this blaCTX-M-15-associated subclone was designated H30-Rx (Fig. 2). Three canonical SNPs distinguished H30-Rx from the rest of H30-R with 100% fidelity.

Genomic location of the CTX-M-15 element. Multiple previous studies have reported that blaCTX-M-15 is positioned on a conjugal IncFII-type plasmid as part of a Tn3-like ISEcp1-blaCTX-M-15-orf477 mobile element. Here, we performed an in silico analysis to characterize the structure and genomic location of the CTX-M-15 mobile element among the 22 blaCTX-M-15-positive ST131 isolates. In each instance, blaCTX-M-15 was part of a typical Tn3-like ISEcp1-blaCTX-M-15-orf477 transposable element. While no SNPs were identified among the different CTX-M-15 elements, the regions flanking blaCTX-M-15 were frequently degraded by insertion/deletion (not shown). The Illumina short-read sequences were sufficient to reliably identify the element’s insertion site for all but three of the 22 CTX-M-15-positive isolates sequenced in the current study (Table 1). The insertion site varied among the isolates: 13 carried a single copy on an IncFII-type plasmid, four carried a single copy on the chromosome, and two carried one copy on the chromosome and another copy on an IncFII-type plasmid (Table 1). Moreover, among the strains with a chromosomally located element, five distinct chromosomal insertion sites were identified; only two strains, JJ1886 and JJ1887, which are the nearest neighbors in the SNP phylogeny (Fig. 2), carried the element in the same chromosomal location (Table 1).

Association of ESBL production and blaCTX-M-15 with the H30-Rx ST131 subclone. Because most of the isolates in the phylogenetic trees were of historic (i.e., pre-2009) origin, we assessed the generalizability of the observed association of the H30-Rx sub-
clone with ESBL production and \( \text{bla}_{\text{CTX-M-15}} \) by analyzing more recent clinical isolates, i.e., from 2011 to 2013. For this, a total of 261 ST131 isolates from Seattle, WA, Minneapolis, MN, and Münster, Germany, were assessed for \( \text{fimH} \) allele type, \( H30 \)-Rx subclone membership, ESBL production, and possession of \( \text{bla}_{\text{CTX-M-15}} \) (Table 2).

Among the 261 recent ST131 isolates, 174 (67%) belonged to the \( H30 \) subclone, whereas the remaining 87 (33%) carried one of several other \( \text{fimH} \)-associated \( \text{fimH} \) alleles, as described recently (15). Among the 174 \( H30 \) isolates, the 163 (94%) that were fluoroquinolone resistant were defined as \( \text{bla}_{\text{CTX-M-15}} \) carriers. Detection of \( H30 \)-Rx-specific SNPs showed that \( H30 \)-Rx comprised 44 (27%) of 163 \( H30 \)-R strains (Table 2).

Among the 44 \( H30 \)-Rx isolates, 34 (77%) were ESBL producing, and 33 of these carried \( \text{bla}_{\text{CTX-M-15}} \). This was very similar to the 74% carriage of \( \text{bla}_{\text{CTX-M-15}} \) observed among the genome-sequenced historic \( H30 \)-Rx isolates (Fig. 2) but significantly higher than the low prevalence of either ESBL production or \( \text{bla}_{\text{CTX-M-15}} \) carriage observed among the recent non-\( H30 \) ST131 isolates (3% for each trait), \( H30 \) but not \( H30 \)-R isolates (9% for each trait), and \( H30 \)-R but not \( H30 \)-Rx isolates (6% and 2% for the two traits, respectively) (Table 2). Thus, \( \text{bla}_{\text{CTX-M-15}} \) accounted for nearly all ESBL-producing isolates within the \( H30 \)-Rx subclone; conversely, within ST131 overall, the \( H30 \)-Rx subclone accounted for the vast majority of ESBL production and, especially, \( \text{bla}_{\text{CTX-M-15}} \) carriage. Moreover, this tight association between \( H30 \)-Rx and \( \text{bla}_{\text{CTX-M-15}} \) held true across the different laboratories that supplied the recent clinical isolates (data not shown).

**TABLE 1** CTX-M-15 element locations among the 22 *Escherichia coli* ST131 isolates

<table>
<thead>
<tr>
<th>Isolate name</th>
<th>Subclone</th>
<th>Genomic location(^a)</th>
<th>Chromosomai insertion site(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JJ2444</td>
<td>( H30 )-Rx</td>
<td>Chromosome</td>
<td>2,037,134</td>
</tr>
<tr>
<td>JJ2038</td>
<td>( H30 )-Rx</td>
<td>Chromosome</td>
<td>2,127,735</td>
</tr>
<tr>
<td>JJ1886</td>
<td>( H30 )-Rx</td>
<td>Chromosome</td>
<td>1,473,842</td>
</tr>
<tr>
<td>JJ1887</td>
<td>( H30 )-Rx</td>
<td>Chromosome</td>
<td>1,473,842</td>
</tr>
<tr>
<td>JJ2434</td>
<td>( H30 )-Rx</td>
<td>Plasmid and chromosome</td>
<td>4,493,369</td>
</tr>
<tr>
<td>MHS5800</td>
<td>( H30 )-Rx</td>
<td>Plasmid and chromosome</td>
<td>4,191,808</td>
</tr>
<tr>
<td>JJ2547</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>JJ2555</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>JJ2008</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>JJ1914</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>JJ2441</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>JJ2489</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>JJ2657</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>JJ2643</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>U004</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>CD358</td>
<td>( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>NA1114</td>
<td>( H30 )-Rx</td>
<td>Plasmid(^c)</td>
<td>NA</td>
</tr>
<tr>
<td>MH17102</td>
<td>( H30 )-Rx</td>
<td>Undetermined</td>
<td>NA</td>
</tr>
<tr>
<td>QUC02</td>
<td>( H30 )-Rx</td>
<td>Undetermined</td>
<td>NA</td>
</tr>
<tr>
<td>KN1604</td>
<td>( H30 )-Rx</td>
<td>Undetermined</td>
<td>NA</td>
</tr>
<tr>
<td>JJ2244</td>
<td>( H30 ), non-( H30 )-Rx</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
<tr>
<td>JJ2591</td>
<td>Non-( H30 )</td>
<td>Plasmid</td>
<td>NA</td>
</tr>
</tbody>
</table>

\(^a\) Marked as undetermined if the \( \text{m} \) \( \text{sil} \) analyses provided equivocal results.
\(^b\) Chromosomal location based on the \( JJJ1886 \) closed genome. NA, not applicable.
\(^c\) Reported previously.

**DISCUSSION**

The results of this study provide compelling evidence that clonal expansion is the dominant mechanism for the proliferation of...
both CTX-M-15 production and fluoroquinolone resistance in *E. coli* ST131. Past studies have shown that the determinants for both of these traits can be acquired through horizontal gene transfer or, in the case of fluoroquinolone resistance, independent mutation (15). However, the whole-genome SNP-based phylogenies presented here show that almost all of the fluoroquinolone-resistant ST131 isolates belong to a distinct subclone, H30-R, which was derived from a single common ancestor carrying the *fimH*30 allele (i.e., part of the H30 subclone). Likewise, 91% of the CTX-M-15-producing isolates form another distinct subclone, H30-Rx, which was derived from a single common ancestor within the H30-R subclone. These nested subclones form a Russian-doll-like configuration, within which each subsequent lineage is more extensively resistant than the former.

The nearly exclusive confinement of the CTX-M-15 element to the H30-Rx subclone was in striking contrast to this element’s promiscuity within the ST131 genome. We identified H30-Rx isolates with a copy of the element on the chromosome, on an IncFI1-type plasmid, and, in some instances, on both the chromosome and a plasmid. Moreover, among those isolates with a chromosomal CTX-M-15 element, the element was inserted in five different locations. Notably, the only two isolates with identical chromosomal insertion sites were recovered from epidemiologically linked adult siblings who both suffered from UTIs of varying severity and were suspected of sharing the same ST131 strain (43), as supported here by the close proximity of these isolates in the high-resolution H30 phylogenetic tree (Fig. 2). Among the 22 CTX-M-15-producing strains, the CTX-M-15 elements exhibited no SNPs and, therefore, no phylogenetic signal that could be compared with the host strain phylogeny.

One cannot exclude the possibility that chromosomal and even some plasmid-borne CTX-M-15 elements were acquired horizontally by H30-Rx on multiple occasions. However, three lines of evidence suggest that within H30-Rx, the chromosomally encoded CTX-M-15 is the result of repeated intragenomic mobilization of the plasmid-located Tn3-like *ISEcp1-blaCTX-M-15-orf477* element, rather than independent horizontal acquisition events. First, the sequence identity of the CTX-M-15 elements suggests that they represent copies of the same recently derived genetic element. Second, the plasmid location is the most common among H30-Rx isolates. Third, some strains with chromosomally encoded CTX-M-15 (e.g., JJ1886 and JJ1887) also maintain an IncFII-type plasmid, albeit missing the CTX-M-15 element, consistent with the CTX-M-15 element having moved from the plasmid to the chromosome.

The exact evolutionary history of CTX-M-15 acquisition in H30-Rx also remains to be elucidated. It is possible that H30-Rx was founded by an ancestor that acquired CTX-M-15 on an IncFII-type plasmid, which expanded and differentiated along with the subclone. Under this model, the CTX-M-15-negative H30-Rx isolates would represent independent gene loss events. An alternative explanation is that the H30-Rx subclone was founded by a CTX-M-15-negative ancestor and partially expanded before becoming extensively cephalosporin resistant later, through horizontal acquisition of the CTX-M-15 element. Indeed, this subclone may be under differential third-generation cephalosporin selection due to

### TABLE 3
Prevalence of ST131 subclones in relation to source population among 261 recent clinical isolates of *Escherichia coli* ST131 from the United States and Germany

<table>
<thead>
<tr>
<th>Source population</th>
<th>Total no. of ST131 isolates</th>
<th>Non-H30 (n = 87)</th>
<th>Non-H30-R (n = 11)</th>
<th>H30-R (n = 165)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States ambulatory</td>
<td>86</td>
<td>35 (41)</td>
<td>3 (4)</td>
<td>42 (49)</td>
</tr>
<tr>
<td>United States hospital</td>
<td>120</td>
<td>32 (27)</td>
<td>4 (3)</td>
<td>64 (53)</td>
</tr>
<tr>
<td>German hospital</td>
<td>55</td>
<td>20 (36)</td>
<td>4 (7)</td>
<td>13 (24)</td>
</tr>
</tbody>
</table>

*For prevalence of H30-Rx, United States ambulatory compared to United States hospital isolates, P = 0.054 (FET).*  
*For prevalence of H30-Rx, United States hospital compared to United States hospital isolates, P < 0.001 (FET).*  
*For prevalence of H30-Rx, United States hospital compared to German hospital isolates, P = 0.03 (FET).*  

### TABLE 4
Association of ST131 subclones with clinical sepsis among 162 recent clinical isolates of *Escherichia coli* ST131 from the United States

<table>
<thead>
<tr>
<th>Clinical presentation</th>
<th>Total ST131 isolates (n = 162)</th>
<th>Non-H30 (n = 36)</th>
<th>Non-H30-R (n = 6)</th>
<th>H30-R (n = 82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No sepsis</td>
<td>150 (92.6)</td>
<td>54 (96)</td>
<td>6 (100)</td>
<td>77 (94)</td>
</tr>
<tr>
<td>Sepsis</td>
<td>12 (7.4)</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>5 (6)</td>
</tr>
</tbody>
</table>

*For prevalence of sepsis, H30-Rx compared to non-H30, P = 0.008 (FET).*  
*For prevalence of sepsis, H30-Rx compared to non-H30-R (H30), P = 0.28 (FET).*  
*For prevalence of sepsis, H30-Rx compared to non-H30-R (H30-R), P = 0.016 (FET).*  
*For prevalence of sepsis, H30-Rx compared to all other ST131 (71/144, 5%), P = 0.003 (FET).*

---

**TABLE 3** Prevalence of ST131 subclones in relation to source population among 261 recent clinical isolates of *Escherichia coli* ST131 from the United States and Germany

<table>
<thead>
<tr>
<th>Source population</th>
<th>Total no. of ST131 isolates</th>
<th>Non-H30 (n = 87)</th>
<th>Non-H30-R (n = 11)</th>
<th>H30-R (n = 165)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States ambulatory</td>
<td>86</td>
<td>35 (41)</td>
<td>3 (4)</td>
<td>42 (49)</td>
</tr>
<tr>
<td>United States hospital</td>
<td>120</td>
<td>32 (27)</td>
<td>4 (3)</td>
<td>64 (53)</td>
</tr>
<tr>
<td>German hospital</td>
<td>55</td>
<td>20 (36)</td>
<td>4 (7)</td>
<td>13 (24)</td>
</tr>
</tbody>
</table>

*For prevalence of H30-Rx, United States ambulatory compared to United States hospital isolates, P = 0.054 (FET).*  
*For prevalence of H30-Rx, United States hospital compared to United States hospital isolates, P < 0.001 (FET).*  
*For prevalence of H30-Rx, United States hospital compared to German hospital isolates, P = 0.03 (FET).*  

---

**TABLE 4** Association of ST131 subclones with clinical sepsis among 162 recent clinical isolates of *Escherichia coli* ST131 from the United States

<table>
<thead>
<tr>
<th>Clinical presentation</th>
<th>Total ST131 isolates (n = 162)</th>
<th>Non-H30 (n = 36)</th>
<th>Non-H30-R (n = 6)</th>
<th>H30-R (n = 82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No sepsis</td>
<td>150 (92.6)</td>
<td>54 (96)</td>
<td>6 (100)</td>
<td>77 (94)</td>
</tr>
<tr>
<td>Sepsis</td>
<td>12 (7.4)</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>5 (6)</td>
</tr>
</tbody>
</table>

*For prevalence of sepsis, H30-Rx compared to non-H30, P = 0.008 (FET).*  
*For prevalence of sepsis, H30-Rx compared to non-H30-R (H30), P = 0.28 (FET).*  
*For prevalence of sepsis, H30-Rx compared to non-H30-R (H30-R), P = 0.016 (FET).*  
*For prevalence of sepsis, H30-Rx compared to all other ST131 (71/144, 5%), P = 0.003 (FET).*
its enhanced virulence, which might result in its being exposed to aggressive antimicrobial therapy more commonly than other E. coli strains. Further investigation, including total chromosome and plasmid closure, could clarify the most probable mechanism for the association between H30-Rx and CTX-M-15. Interestingly, however, the blaCTX-M-15-positive, FQ-resistant JJ2244 strain, which occupies a phylogenetic outgroup position relative to both H30-R and H30-Rx (see Fig. 2), likely represents an independently emerged multidrug-resistant clonal lineage within the H30 subclone. Indeed, this strain not only lacks all canonical H30-Rx SNPs but also has a distinct, recombinant FQ resistance-conferring gyrA-parC allele combination (i.e., 1AB/4A), compared with that present in H30-R (i.e., 1AB/1AAB).

The results of this whole-genome SNP-based analysis depict a considerably different evolutionary history for ST131 from that derived from PFGE analysis. Our use of an iterative approach to identify and exclude SNPs from recombinant regions elucidated an evolutionary path marked by clonal expansions rather than frequent lateral gene acquisitions. This underscores one of the major advantages of a whole-genome SNP-based approach relative to PFGE. Although PFGE likewise uses signatures from throughout the genome, it is highly vulnerable to phylogenetic distortions from horizontal gene transfer, which can lead to false assumptions about the evolutionary history of an organism (44), and from subjective interpretation of banding patterns and the (often invalid) presumption that similarly migrating bands represent the same chromosomal region.

The biological basis for the proliferation of H30-R and H30-Rx remains unclear. It is possible that antimicrobial resistance, the seemingly obvious explanation, is not the sole selective characteristic leading to the successive proliferation of H30-R and H30-Rx from within the H30 subclone. This is suggested by the fact that certain non-H30 ST131 isolates were identified that possessed the same phenotypic resistance traits without having reached a comparable level of success as H30-Rx. Increased virulence, as suggested by the significant association between H30-Rx and sepsis, could be a factor contributing to the success of this important subclone. Further investigations, including detailed comparative genomic, epidemiological, and functional studies, are needed to determine the basis for the success of H30-R and the strong association of H30-Rx with sepsis.

Regardless of mechanisms, the association of H30-Rx with sepsis, its broad multidrug resistance profile, and its rapid expansion and geographic dispersal warrant attention from the public health and clinical communities. Although continued accumulation of antibiotic resistance determinants may limit therapeutic options in the future, the clonal nature of H30-Rx may facilitate effective control strategies involving vaccines or transmission prevention.

**MATERIALS AND METHODS**

**Isolates and patients.** The molecular epidemiological analyses used a large collection (n = 1,908) of recent, consecutive, single-patient E. coli isolates from 6 clinical microbiology laboratories in the United States and Germany. The United States isolates (n = 1,518) were recovered in 2010 and 2011 from 5 locations, including Group Health Cooperative, Harborview Medical Center, Seattle Children’s Hospital, and University of Washington Medical Center (all in Seattle, WA) and the Veterans Affairs Medical Center in Minneapolis, MN, as described previously (15). The German isolates (n = 390) were recovered in 2012 at the University Hospital in Münster, Germany. All isolates underwent fumC-fimH (CH) clonotyping (45) to identify ST131 and its constituent CH clonotypes (i.e., fimH-specific subclones, including H30) and were assessed for ESBL production by disk diffusion as specified by the Clinical and Laboratory Standards Institute. Medical record data regarding presence of clinically diagnosed sepsis at the time of sample collection or during the subsequent 30 days were available for 1,133 (75%) of the 2010-2011 United States isolates. Each center’s institutional review board approved the study protocol.

**PFGE analysis.** The 524 historical and recent ST131 isolates were subjected to standardized XbaI PFGE analysis, as described previously (46). The dendrogram was inferred within BioNumerics (Applied Maths) according to the unweighted pair group method based on Dice similarity coefficients.

**Strain selection.** Selection of ST131 isolates for genome sequencing was done in successive phases. First, to sample the breadth of phylogenetic diversity within the ST (to the extent that this is reflected in PFGE profiles), 20 isolates were selected to represent widely distributed clusters within a PFGE profile dendrogram based on a published collection of 524 historical and recent ST131 isolates from diverse locales, years of isolation, and hosts (Fig. 1A). In selecting the representative isolate(s) for a given PFGE cluster, priority was given to (i) most recent year of isolation, (ii) human host, and (iii) fluoroquinolone resistance. Next, 28 additional isolates were selected from these same PFGE clusters based on (i) proximity in the dendrogram to the initially selected (index) isolate and (ii) differences from the index isolate with respect to host and/or fluoroquinolone phenotype. Subsequently, an additional 60 isolates were selected from both this initial collection and a large collection of recent human clinical ST131 isolates from Seattle, WA, and Minneapolis, MN, that had undergone sequence analysis of gyrA, parC, and fimH (to define subclones within ST131) and PFGE analysis. Here, selection criteria included (i) distinctive gyrA, parC, and/or fimH alleles, or combinations thereof, (ii) outliers with respect to fluoroquinolone phenotype, in comparison with other isolates sharing the same PFGE type or gyrA-parC-fimH allele combination, and (iii) distinctive host species, clinical presentations (e.g., published case report isolates), specimen types (e.g., food or environmental), or dates of isolation (e.g., oldest known and oldest published ST131 isolates). Of the 108 total selected isolates, four isolates were subsequently excluded due to questionable authenticity, leaving 104 isolates for genome sequencing.

**Genome sequencing.** DNA samples were prepared for multiplexed, paired-end sequencing on an Illumina Genome Analyzer Ix (Illumina, San Diego, CA). For each isolate, 1 to 5 μg DNA in 200 μl was sheared in a 96-well plate with the SonicMAN (part no. SCM1000-3; Matrical BioScience, Spokane, WA) to a size range of 200 to 1,000 bp, with the majority of material at ca. 600 bp, using the following parameters: prechill, 0°C for 75 s; cycles, 20; sonication, 10 s; power, 100%; lid chill, 0°C for 75 s; plate chill, 0°C for 10 s; postchill, 0°C for 75 s. The sheared DNA was purified using the QIAquick PCR purification kit (catalogue no. 28106; Qiagen, Valencia, CA). The enzymatic processing (end repair, phosphorylation, A tailing, and adaptor ligation) of the DNA followed the guidelines described in the Illumina protocol (“Preparing Samples for Multiplexed Paired-End Sequencing,” catalogue no. PE-930-1002, part no. 1005361). The enzymes for processing were obtained from New England Biolabs (catalogue no. E6000L; New England Biolabs, Ipswich, MA), and the oligonucleotides and adaptors were obtained from Illumina (catalogue no. PE-400-1001).

After ligation of the adaptors, the DNA was run on a 2% agarose gel for 2h, after which a gel slice containing 500- to 600-bp fragments of each DNA sample was isolated and purified using the QIAquick gel extraction kit (catalogue no. 28706; Qiagen, Valencia, CA). Individual libraries were quantified by quantitative PCR on an ABI 7900HT (part no. 4529001; Life Technologies Corporation, Carlsbad, CA) in triplicate at two concentrations, 1:1,000 and 1:2,000, using the Kapa library quantification kit (part no. KK4832 or KK4835; Kapa Biosystems, Woburn, MA). Based on the individual library concentrations, equimolar pools of no more than 12 indexed E. coli libraries were prepared at a concentration of at least 1 nM.
using 10 mM Tris-HCl (pH 8.0) and 0.05% Tween 20. To ensure accurate loading onto the flow cell, the same quantification method was used to quantify the final pools. The pooled paired-end libraries were sequenced on an Illumina Genome Analyzer IIX to a read length of at least 76 bp.

The genomes were sequenced at an average depth of 60.93× (standard deviation [SD] = 31.66, using the 4,971,461-base NA114 chromosome as a reference). An average of 4,654,457.54 bases (SD = 385,629.23) were sequenced at ≥10× coverage.

Identification of SNPs. Illumina whole-genome sequence data sets were aligned against the chromosome of a published ST131 reference genome (strain NA114; GenBank accession no. CP002797) (42) using the short-read alignment component of the Burrows-Wheeler Aligner. Each alignment was analyzed for SNPs using SolSNP, a Java-based DNA variant-calling tool for next-generation sequencing alignment data. SolSNP uses a modified Kolmogorov-Smirnov statistic and data filtering to call variants on high-coverage, aligned genomes (http://sourceforge.net/projects/solspan/). To avoid false calls due to sequencing errors, SNP loci were excluded if they did not meet a minimum coverage of 10× and if the variant was present in <90% of the base calls for that position. SNPs were combined for all of the sequenced genomes such that, for the locus to be included in the final SNP matrix, it had to be present in all of the genomes. SNPs falling in the duplicated regions on the reference genome were discarded.

Phylogenetic analysis. Phylogenetic trees were generated using the maximum-parsimony method in PAUP version 4.0b10. Using prior knowledge about near neighbors, a published E. coli strain belonging to the phylogenetic group B2 genome (strain AA86; GenBank accession no. AFET00000000) was selected as an outgroup to root the ST131 whole-genome sequence tree (45). ST131 isolates in the clade nearest to this bifurcation point were used to root subsequent trees.

The homoplasmy index (HI) was calculated in PAUP using the formula $HI = 1 - CI$, where CI is the consistency index. CI serves to measure the relative amount of homoplasgy in a cladogram, as assessed by the level of difficulty in fitting SNP alleles to a given tree. The CI is calculated using the formula $CI = m/s$, where $m$ is the total number of expected character changes and $s$ is the actual number of changes that occur in the tree.

Detection of H30-Rx-specific SNPs. Two SNPs that differentiate the CTX-M-15-associated subclones (H30-Rx) within the H30-R clade from the rest of the H30 subclade were interrogated using Sanger sequencing. SNP-200 was detected as a C-to-T transition at position 299 of the 462-bp PCR product generated using forward primer 5′-GACACCA TGGCTTTTGCCTC-3′ and reverse primer 5′-TGCTGCGAAACGCAAT TGAC-3′. SNP-264 was detected as a C-to-G transition at position 287 of the 462-bp PCR product generated using forward primer 5′-GTGGCGA TTTCGACGGTGT-3′ and reverse primer 5′-TATCCAGCACGTTCCA GTTG-3′. Isolates that tested positive for both SNPs were regarded as members of the H30-Rx subclone.

PCR-based detection of blaCTX-M-15. The CTX-M-15-encoding gene blaCTX-M-15 was detected by PCR using SNP-specific forward primer 5′-TAAACCCCGACCGGTTG-3′ and universal reverse primer 5′-GAATT TTGCGATCGGGG-3′ (47). PCR conditions were 10 min of denaturation at 95°C, 33 cycles of 30 s at 94°C, and 30 s at 67°C, followed by 7 min at 72°C elongation. The blaCTX-M-15-specific 483-bp PCR product was detected by agarose gel electrophoresis.

Location of CTX-M-15-encoding mobile element. Illumina short reads were aligned to the closely related chromosome of JJ1886 (CP006784, CP006785, CP006786, CP006787, CP006788, and CP006789) (54) and the pEC_L8 closed plasmid sequence using BWA-MEM (48) version 0.7.5a-r405 with the default settings for paired reads. Alignments were analyzed using the IGV tool (49, 50). The location of the conserved CTX-M-15 element (ISEcp1-blaCTX-M-15-orf477) was then inferred based on the sequence alignments. Close attention was paid to the boundaries of the element, since any unmapped or incorrectly mapped pairs or pairs with insertions or deletions provided clues as to the genomic location and possible rearrangements. Putative duplications were noted if the depth of coverage around the element was substantially higher than the flanking regions. Illumina short reads were also assembled with the Mira assembler (51). The contigs were aligned with Mauve (52) against JJ1886 and pEC_L8 to identify any chromosomal rearrangement in the contigs carrying the CTX-M-15 element.

**fmfl and gyrA-parC allelic assignments.** Sequenced isolates were assembled using VelvetOptimiser (version 2.2.2) and Velvet (53). All fmfl, gyrA, and parC sequences were compared to an in-house sequence library using nucleotide-nucleotide BLAST (version 2.2.25+). Sequence similarity matches were determined using thresholds of 100% nucleotide identity and 100% coverage of the query sequence length. Allele designations were assigned based on an in-house nomenclature for the gyrA-parC combination and fmfl.

**Statistical methods.** Comparisons of proportions were tested using Fisher’s exact test or a chi-square test (two tailed), with $P$ values of $<0.05$ as the criterion for significance.

**Accession number.** All Illumina sequences were deposited into the NCBI SRA (http://www.ncbi.nlm.nih.gov/sra), study accession number SRP027327.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at http://mbio.asm.orglookup/supp/doi:10.1128/mBio.00377-13/-/DCSupplemental.

Figure S1, PDF file, 0.1 MB.

Figure S2, PDF file, 0.1 MB.

Figure S3, PDF file, 0.1 MB.

Table S1, DOCX file, 0.1 MB.

Dataset S1, XLSX file, 0.1 MB.

Dataset S2, XLS file, 0.3 MB.

**ACKNOWLEDGMENTS**

This material is based upon work supported by Office of Research and Development, Medical Research Service, Department of Veterans Affairs, Merit Review grant 1H01 CX000192 01 (J.R.L.), the TGen Foundation (L.B.P.), NIH grant RC4 AI092828 (E.V.S.), and USAMRMC grant W81XWH-10-1-0753 (L.B.P.).

We thank Tania Conteste-Cuomo, Richard Lester, Michael Bork, Cindy M. Liu, and Laura Davis from TGen for their contributions to this project and the staff of the Clinical Microbiology Laboratories of Harbor-View Medical Center, Seattle, WA, University of Washington Medical Center, Seattle, WA, and Universitätsklinikum, Münster, Germany, for their excellent help in collecting isolates and associated clinical data. Providers of isolates (institutional affiliations) also include Javier Adachi (University of Texas M. D. Anderson Cancer Center, Houston, TX), Robert L. Bergbaken (Health Partners and Regions Medical Center, St. Paul, MN), Susan Butler-Wu (Department of Laboratory Medicine, Seattle Children’s Hospital, Seattle, WA), Mariana Castanheira (JMI Laboratories, North Liberty, IA), Tim Cleary (University of Miami, Miami, FL), Brad T. Cookson (Department of Microbiology and Department of Laboratory Medicine, University of Washington School of Medicine, Seattle, WA), Michael Cooperstock (University of Missouri School of Medicine, Columbia, MO), Chitrira DebRoy (Pennsylvania State University, University Park, PA), Paul Edelstein (Hospital of the University of Pennsylvania, Philadelphia, PA), Peter T. Ender (St. Luke’s University Hospital and Health Network, Bethlehem, PA), Ferric C. Fang (Department of Microbiology and Department of Laboratory Medicine, University of Washington School of Medicine, Seattle, WA), John Holter (Minneapolis VA Health Care System, Minneapolis, MN), Thomas M. Hooton (University of Miami, Miami, FL), Marie-Hélène Nicolas-Chanoine (Hôpital Beaujon, Clichy, France), Rob Owens (Cubist Pharmaceuticals, Falmouth, ME), Johann Pitout (Calgary Laboratory Services and University of Calgary, Calgary, AB, Canada), JoAnn Platell (University of Queensland, Brisbane, QLD, Australia), Xuan Qin (Department of Laboratory Medicine, Seattle Children’s Hospital, Seattle, WA), John Quinn (Atra Zeneca, Waltham, MA), James Rice (Scripps...
REFERENCES


