Role of Homologous Recombination in Adaptive Diversification of Extraintestinal *Escherichia coli*

Sandip Paul,^a Elena V. Linardopoulou,^a Mariya Billig,^a Veronika Tchesnokova,^a Lance B. Price,^b James R. Johnson,^c Sujay Chattopadhyay,^a Evgeni V. Sokurenko^a

Department of Microbiology, University of Washington, Seattle, Washington, USA^a; The Translational Genomics Research Institute (TGen), Flagstaff, Arizona, USA^b; Veterans Affairs Medical Center and University of Minnesota, Minneapolis, Minnesota, USA^c

The contribution of homologous exchange (recombination) of core genes in the adaptive evolution of bacterial pathogens is not well understood. To investigate this, we analyzed fully assembled genomes of two *Escherichia coli* strains from sequence type 131 (ST131), a clonal group that is both the leading cause of extraintestinal *E. coli* infections and the main source of fluoroquinolone-resistant *E. coli*. Although the sequences of each of the seven multilocus sequence typing genes were identical in the two ST131 isolates, the strains diverged from one another by homologous recombination that affected at least 9% of core genes. This was on a par with the contribution to genomic diversity of horizontal gene transfer and point gene mutation. The genomic positions of recombinant and mobile genetic regions were partially linked, suggesting their concurrent occurrence. One of the genes affected by homologous recombination was *fimH*, which encodes mannose-specific type 1 fimbrial adhesin, resulting in functionally distinct copies of the gene in ST131 strains. One strain, a uropathogenic isolate, had a pathoadaptive variant of *fimH* that was acquired by homologous replacement into the commensal strain background. Close examination of FimH structure and function in additional ST131 isolates revealed that recombination led to acquisition of several functionally distinct variants that, upon homologous exchange, were targeted by a variety of pathoadaptive mutations under strong positive selection. Different recombinant *fimH* strains also showed a strong clonal association with ST131 isolates that had distinct fluoroquinolone resistance profiles. Thus, homologous recombination of core genes plays a significant role in adaptive diversification of bacterial pathogens, especially at the level of clonally related groups of isolates.

Escherichia coli is a leading human bacterial pathogen, causing both intestinal and extraintestinal infections (1–3). It has been known for decades that pathogenic bacterial strains of *E. coli* and other species tend to be clonal in nature (i.e., to belong to a limited number of genetically related lineages) (4, 5). In the 1980s, a standardized set of *E. coli* strains was assembled as the *E. coli* reference collection (ECOR), leading to the discovery that most *E. coli* isolates could be classified into one of several major phylogenetic groups which have distinct pathotype associations (6). For example, most enterotoxigenic strains belong to so-called group A, B1, or AxB1, *Shigella* lineages belong to group B1 or ABD, and strains causing extraintestinal infections (extraintestinal pathogenic *E. coli* [ExPEC]) primarily belong to groups B2 and D. More recently, multilocus sequence typing (MLST), which is based on nucleotide sequence analysis of seven gene loci, has emerged as the major clonal typing scheme (7). MLST allowed separation of the strains to distinct clonal groups—sequence types (STs)—on the basis of the identity of the MLST loci. Furthermore, several STs have emerged as the leading clonal groups among ExPEC, including ST73, ST95, ST69, and, most recently, ST131 (8–10).

Although MLST has become widely accepted as a standardized method for typing of *E. coli*, it lacks sufficient clonal resolution for molecular epidemiological analysis applications, where methods like pulsed-field gel electrophoresis (PFGE) are still preferred (11). Moreover, detailed analysis of strains belonging to the same ST has revealed a high level of diversity in their pathogenic potential. For example, ST73 includes model pathogenic strains, such as urosepsis isolate CFT073, but it also includes probiotic strain Nissle1917 and asymptomatic bacteriuria strain 83972 (12, 13). Such strains with different pathogenicities within the same ST can be shown to belong to distinct genetic lineages by either PFGE or

mosaic gene mapping and, recently, partial- or whole-genome sequencing (14–17). These studies suggest that to understand the physiological significance of the clonal diversity of pathogenic bacterial species, it is necessary to examine the molecular basis of genetic variability among strains with identical MLST profiles.

The past decade has seen the rapid emergence and global spread of a new clonal group, ST131, that includes many fluoroquinolone (FQ)-resistant *E. coli* isolates (18–20). While ST131 has become the most prominent clonal group, very little is known about its within-clonal-group genomic diversity. Recently, a fully assembled genome was announced for ST131 strain NA114, isolated from the urine of a patient with prostatitis in India and characterized as serogroup O25 (21). Screening of the public databases revealed that the NA114 genome is very similar to the strain SE15 genome; the fully assembled genome of SE15 had been deposited earlier (22). SE15 was isolated from the feces of a healthy individual in Japan and was described as a serotype O150:H5 strain belonging to group B2 and containing a limited number of virulence factors found in ExPEC isolates (23). More recently, next-generation sequencing of several other genomes of ST131

Received 21 August 2012 Accepted 28 October 2012

Published ahead of print 2 November 2012

Address correspondence to Evgeni V. Sokurenko, evs@u.washington.edu.

S.P. and E.V.L. contributed equally to this article.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JB.01524-12>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JB.01524-12

FIG 1 Phylogram of concatenated sequences of seven housekeeping genes from SE15, NA114 (red font), and 23 other strains of *E. coli/Shigella*. The seven housekeeping genes used in MLST were selected for analysis. Pathotypes are mentioned for each strain (EAEC, EHEC, ETEC, EPEC, NMEC, and UPEC, enteroaggregative, enterohemorrhagic, enterotoxigenic, enteropathogenic, neonatal meningitis, and uropathogenic *E. coli*, respectively). Strains that cluster into different phylogroups (A, B1, B2, AxB1, ABD, and D) are also shown. The percentage of genes shared by different strains and strains SE15 and NA114 of *E. coli* are presented in the right panel. NA, not applicable.

strains was reported (24, 25), but the SE15 and NA114 genomes remain the only fully assembled ST131 genomes.

In order to obtain a deeper understanding of the evolution and diversity of ST131, we have performed a detailed comparative analysis of the genomes of strains NA114 and SE15 at the level of both mosaic and shared genes. We show that while NA114 and SE15 belong to the same ST131 clonal group, substantial genomic diversity has accumulated between the strains due to horizontal movement of genes, homologous recombination, and mutation. At least in part, all three genetic mechanisms have contributed to pathogenicity-adaptive evolution of major subclones within ST131.

MATERIALS AND METHODS

Strain selection. Fully assembled genomes of *E. coli* strains NA114 and SE15 were downloaded from NCBI GenBank. For comparison, 23 other completely annotated genomes of *E. coli* strains were selected and downloaded (Fig. 1). These strains were chosen because they are fully assembled and clonally distinct by MLST analysis (<http://mlst.ucc.ie>). Strains S88, UTI89, and IHE3034, being clonally identical, were the only exceptions. Although the current and other MLST schemes might be based in part on loci affected by recombination, different schemes could produce different STs within the same strain set.

ST131 and non-ST131 group-specific clinical isolates were subsets of the collection from a previous study (26).

Phylogenetic analysis and determining shared genes for comparative analysis. In order to compare the clonal relatedness of strains NA114 and SE15 with 23 fully annotated strains of *E. coli*, a set of seven housekeeping gene sequences was concatenated for each strain. These genes were *adk* (adenylate kinase), *fumC* (fumarate hydratase), *gyrB* (DNA gyrase), *icd* (isocitrate/isopropylmalate dehydrogenase), *mdh* (malate de-

hydrogenase), *purA* (adenylosuccinate dehydrogenase), and *recA* (ATP/GTP binding motif). The resulting concatenates were used to construct a maximum-likelihood (ML) phylogenetic tree with the help of a general time-reversible (GTR) substitution model. To depict an average neutral diversity among the strains, site-specific base frequencies were estimated by codon position distribution, as implemented in the PAUP* program (27).

The percentage of the genome shared by each of SE15 and NA114 with 23 other non-ST131 strains was determined using a stand-alone BLAST search. We employed an in-house software package, TimeZone, to perform BLAST searches of each of the annotated genes in SE15 and NA114 separately as a reference against each of 23 non-ST131 genomes and to extract homologs using a cutoff of $\geq 95\%$ for nucleotide sequence identity and sequence length coverage values. We excluded all annotated pseudogenes in reference genomes (SE15 or NA114). BLAST hits with either non-ACGT characters or internal stop codons were also excluded.

Detection of homologous gene recombination and mutations in the core genes of NA114 and SE15. In order to identify the core genes shared between strains NA114 and SE15, we used SE15 as the reference genome and extracted homologs from NA114 with nucleotide sequence identity and length coverage values of $\geq 95\%$ relative to the reference sequence, thereby identifying 3,997 core genes. Among these 3,997 genes, 2,747 genes were identical.

For each of the 1,250 polymorphic genes between NA114 and SE15 (ST131), we found the corresponding gene in five other non-ST131 strains representing the B2 phylogroup (ST127 strain 536, ST73 strain CFT073, ST15 strain E2348/69, ST452 strain ED1a, and ST95 strain UTI89). We used the SE15 gene sequence as a reference and performed BLAST searches with these five strains with $\geq 95\%$ sequence identity and length coverage cutoffs.

In each gene set, while we calculated the NA114-SE15 diversity (i.e.,

within-ST131 diversity) value, the between-ST diversity value was computed via two different approaches: (i) by use of the average nucleotide diversity of clonally different STs representing the B2 phylogroup and (ii) by use of the average diversity of the ST131 and non-ST131 sequence pairs. Finally, the mean of these values was considered the between-ST diversity value for any given gene. The combination of these two approaches, we believe, would minimize the impact on the between-ST diversity value due to the presence of any outlier (recombinant) sequence in the non-ST131 isolate(s).

A polymorphic gene is considered to be of mutated origin if the NA114-SE15 diversity value is below the lower 95% limit of between-ST diversity. In contrast, a gene with an NA114-SE15 diversity value above the lower 50% limit of between-ST diversity is designated a recombinant, while a gene with NA114-SE15 diversity leading to any intermediate value is defined to be of unknown origin. These criteria of detecting genes of putative recombinant or mutated origin were applied using the between-ST diversity values calculated via both approaches. In a conservative fashion, we assigned a gene polymorphism to be of mutated or recombinant origin only if the respective criteria were satisfied via both approaches. In other scenarios, we designated the gene polymorphism to be of uncertain (obscure) origin.

In order to test the validity of the diversity-based separation of the recombinant and mutated genes, 50 genes were randomly chosen from each of the mutated (out of 545 genes) and recombinant (out of 354 genes) groups. For each gene, the maximum-likelihood phylogenetic tree was generated on the aligned sequences of NA114, SE15, and 23 non-ST131 strains (Fig. 1) to determine the relative position of the NA114 and SE15 alleles within the 23 non-ST131 strains.

Sequencing of *fimH* and *fimA* allele in 43 different isolates from ST131 and cloning of *fimH* allele. We amplified the *fimH* and *fimA* gene sequences from each of the 43 *E. coli* isolates (see Table S1 in the supplemental material) using genomic DNA as a template by PCR. The primers for *fimH* amplification were *fimH* F-GGGGGTGCACCTCAGGAACCATTCAGGCA and *fimH* R-GGGGCATGCTTATTGATAACAAAAGTCAC (where F is forward and R is reverse), and the primers for *fimA* amplification were *fimA* F-GTTGTTCTGTCGGCTCTGTCC and *fimA* R-CTTGAAGTCCGATCCGCATTAG. PCR products were separated in a 1% agarose gel and purified on a QIAquick column (Qiagen). For sequencing, we used a BigDye Terminator (version 3.1) cycle sequencing kit (Applied Biosystems). The same PCR primers were used to amplify *fimH* in all ST131 strains, and *fimH30* and other *fimH* alleles were distinguished by Sanger sequencing of the PCR products.

We used strain DH5 α and *fim*-null K-12 derivative strain AAEC191A (provided by Ian Blomfield, University of Kent, United Kingdom) for cloning of *fimH* and expression of type 1 fimbriae for binding studies, as described previously (28).

Accession numbers of *fimH* and *fimA* recombinant allele. The sequence representatives of seven recombinant *fimH* alleles from GenBank were *fimH22* (GenBank accession number GQ487041, region 21 to 509), *fimH27* (GenBank accession number U00096, region 4546894 to 4547382), *fimH30* (GenBank accession number CP002797, region 4712714 to 4713202), *fimH31* (GenBank accession number FJ865843, region 64 to 552), *fimH35* (GenBank accession number CP000802, region 4553879 to 4554367), *fimH41* (GenBank accession number AP009378, region 4641973 to 4642461), and *fimH94* (GenBank accession number FJ865709, region 64 to 552). For five recombinant *fimA* alleles, the sequence representatives were *fimA3* (GenBank accession number AP009240, region 4792114 to 4792552), *fimA4* (GenBank accession number KC110846), *fimA5* (GenBank accession number KC110847), *fimA6* (GenBank accession number AF490890, region 40 to 478), and *fimA7* (GenBank accession number KC110848).

Bacterial binding. Yeast (*Saccharomyces cerevisiae*) aggregation of clinical isolates was carried out as follows: bacteria were grown overnight statically in super broth (SB) at 37°C and washed twice with 1× phosphate-buffered saline (PBS), and the optical density at 540 nm (OD₅₄₀) was adjusted to 4.0. Equal volumes of bacteria, baker's yeast (10 mg/ml in

1× PBS), and 0.2% bovine serum albumin (BSA)-PBS with or without 2% α -methyl-D-mannopyranoside were mixed on agglutination slides, and mannose-inhibited yeast aggregation was determined visually.

Static assays of bacterial adhesion to immobilized monomannose (1M) ligands (yeast mannan) and trimannose (3M) ligand (bovine RNase B) were carried out in 96-well plates as described previously (29). The isogenic bacteria expressing type 1 fimbriae with different *fimH* alleles were grown for 16 to 20 h under static conditions at 37°C in SB medium supplemented with antibiotics and 10 μ Ci/ml of [*methyl*-³H]thymidine (Perkin-Elmer), washed twice with 1× PBS, and adjusted to an OD₅₄₀ of 4.0. The flat-bottom 96-well plates were coated for 1 h at 37°C with 100 μ l of monomannosylated (1M) substrate yeast mannan (20 μ g/ml; Sigma), trimannosylated (3M) substrate RNase B (20 μ g/ml; Sigma), anti-FimH antibody (polyclonal antibody 280 [PAb280], 1:100), or anti-type 1 fimbrial antibody (PAb49, 1:100) in 0.2 M NaHCO₃ and 0.1% BSA-1× PBS as a control for nonspecific binding (30). After quenching with 0.1% BSA-PBS, 50 μ l of prewashed bacteria was mixed in the wells with 50 μ l of either 0.1% BSA-PBS or 100 mM inhibitor of binding of α -methyl-D-mannopyranoside (Sigma). Bacteria were allowed to bind for 40 min at 37°C, unbound bacteria were washed away, and bound bacteria were detected by the liquid scintillation method. The ratio of 1M/3M binding was calculated to determine the low- or high-binding profile of each isogenic strain. Binding of labeled bacteria to immobilized PAb280 (anti-FimH, 1:100) and PAb49 (anti-fimbriae, 1:200) in the presence of mannose was used to control the uniform level of expression of type 1 fimbriae and FimH.

Screening for several mosaic loci within 160 ST131 isolates and 84 non-ST131 isolates. Different loci which were either NA114 or SE15 specific or common in both were identified, and primers were designed (see Table S2 in the supplemental material) for PCR amplification of loci in 160 ST131 isolates and in a subset of loci in 84 non-ST131 isolates.

RESULTS

Gene mosaicism of the genomes of strains SE15 and NA114. We first compared the clonal relatedness of the genomes of NA114 and SE15 with 23 publicly available annotated genomes of *E. coli*, based on a seven-locus MLST typing scheme (<http://mlst.ucc.ie>) (7). The loci included the following housekeeping genes: *adh*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*. As expected, NA114 and SE15 grouped together with *E. coli* strains of phylogenetic group B2, which encompasses most ExPEC isolates and is distinct from phylogenetic groups carrying *Shigella* or other enteropathogenic strains (Fig. 1). Each of the seven loci was identical in NA114 and SE15, affirming that these strains are clonally related and display the same MLST profile distinctive for ST131.

Genome-wide comparisons showed that, on the basis of a threshold of 95% sequence identity and length coverage, the ST131 strains shared, on average, more genes with strains from the B2 group than with those from any other major phylogenetic group of *E. coli* (Fig. 1), affirming the B2 association of the ST131 isolates. NA114 shared, on average, 80% of the genome with B2 strains versus 69% of the genome with non-B2 strains, while for SE15, these numbers were 88% and 76%, respectively.

Not surprisingly, NA114 and SE15 shared more genes with each other than with other strains. A total of 3,997 genes were shared between the ST131 strains, and these accounted for 85% of the NA114 genome and 92% of the SE15 genome. The vast majority of these core (or backbone) genes were syntenic, exhibiting an identical gene order in both genomes. However, according to the published genome assembly, 9 gene clusters of at least 6 kb in size, which encompassed a total of 196 genes, were in different, translocated genome positions (not shown).

A total of 341 annotated open reading frames (ORFs) in SE15

FIG 2 Schematic representation of core and mosaic regions in *E. coli* strains SE15 and NA114. The core genomic region of strains SE15 and NA114 is represented by a gray rectangle. The mosaic regions (6 kb or larger) in the respective genomes are shown via two types of markers: genomic insertions or genomic replacements. The marker sizes correspond to the size of that region. MR numbers are mentioned for genomic replacements (1 to 6, blue) and genomic insertions (7 to 21, red). Positions of putative recombinant regions are marked by green bars within the core region.

did not have a homologue in NA114, whereas 729 ORFs in NA114 had no homologue in SE15; i.e., these genes were of a mosaic rather than a core nature. We evaluated in detail all mosaic regions (MRs) that were ≥ 6 kb, anticipating that these regions would include, on average, clusters of ≥ 5 genes. The two strains contained a total of 21 MRs. Six MRs were defined as genomic replacements, where different genes are inserted into the same genome position in different strains (Fig. 2, blue symbols), and 15 MRs were defined as genomic insertions, where at the same position in the core genome one strain has an insertion but the other has an uninterrupted sequence (Fig. 2, red symbols). Although gene deletions in one strain can appear as insertions in the respective genomic region of the other strain, the insertions defined as such here were of true insertional origin, on the basis of comparative genomic analysis of other B2 strains (not shown).

Thus, although most of the genes in ST131 strains SE15 and NA114 are of a core nature, 8% of the former strain's genome and 15% of the latter strain's genome are composed of mosaic genes which are positioned in dozens of locations in the genome backbone.

Functional nature of ST131 mosaic genes. Among 6 MRs that were genomic replacements, 4 were primarily comprised of bacteriophage and insertional element sequences, with the notable exception of MR-6 in strain NA114, which also contained a cluster of genes encoding antigen 43 (*sap*) and the IrgA homologue adhesin (*iha*). Both of these loci encode adhesive traits, are putative virulence factors of *E. coli*, and are absent from strain SE15. One of the genomic replacements (MR-4), located between the *galF* and *gnd* loci in both genomes, is involved in lipopolysaccharide (LPS; O-antigen) biosynthesis. In SE15, MR-4 contains 11 genes (11 kb) coding for the O150 antigen. In contrast, the NA114 MR-4 region has four genes (*rmlBDAC*) known to be involved in biosynthesis of L-rhamnose as part of the variable portion of LPS and then a 15-kb region encoding hypothetical proteins, none of which appears to be related to O-antigen synthesis. Another replacement region located between the *pyrBI* operon and the valyl-tRNA synthetase gene (MR-5) includes, in SE15, the arginine deaminase pathway genes *arcACBD*, involved in arginine degradation, but, in NA114, several genes of unknown function.

Among 15 genomic insertions, 4 (MRs 7 to 10) were found in SE15 and 11 (MRs 11 to 21) were found in NA114. Eight of these gene clusters (2 in SE15, 6 in NA114) were primarily of bacteriophage origin, whereas others were mostly composed of genes with defined or putative physiological and/or virulence significance. The two SE15-specific nonphage insertions were related to sero-

type-defining genes, including the group II capsule synthesis (*kps*) cluster (MR-9; 8 kb) and the LPS core biosynthesis locus (*rfa*) (MR-10; 8 kb). In contrast, of the five NA114-specific nonphage regions, one (MR-13) comprised genes coding for hypothetical, functionally undefined proteins and one (MR-21) coded for a type I restriction modification system (RMT1). The remaining three NA114-specific regions were composed to a great extent of genes encoding putative virulence factors. Specifically, MR-12 (37 kb) encodes a second lateral flagellar system, Flag2; MR-19 (14 kb) contains ferric citrate uptake genes (*fec*); and the largest insertion, MR-11 (160 kb), is a chromosomal island that includes virulence factors such as the secreted autotransporter toxin (*sat*), the aerobactin system (*iutA*, *iucABCD*), and a complete operon for digalactoside-specific P fimbriae with the *papA9* major subunit gene and *papGII* adhesive subunit gene.

Thus, most MRs in the ST131 strains are of phage origin in both strains, but a substantial number of NA114-specific genes code for putative virulence factors of extraintestinal *E. coli*, suggesting that their acquisition resulted in the increased pathogenic potential of NA114.

Homologous gene recombination versus mutation in the ST131 core genes. We next analyzed the diversity among the genes shared between the two ST131 strains. More than two-thirds of the core genes (2,747 of 3,997 genes, 68.7%) were identical. The nucleotide diversity of the 1,250 polymorphic genes varied from 0.01% to 4.94%, with a median diversity value of 0.43%, showing a bias toward the lower values. Because different genes might have different mutation rates, we normalized the diversity of each polymorphic gene to the diversity observed for the same genes across clonally unrelated *E. coli* strains. Thus, for each gene that was polymorphic between strains NA114 and SE15, we determined the gene diversity among strains representing six different STs of phylogroup B2 (Fig. 1): strain 536 for ST127, strain CFT073 for ST73, strain E2348/69 for ST15, strain ED1a for ST452, strain UTI89 for ST95, and strain SE15 for ST131. Then, the NA114-SE15 diversity value was divided by the between-ST diversity value for the corresponding genes. The distribution of the normalized diversity of the NA114-SE15 polymorphic gene was significantly positively skewed toward a higher level of diversity (skewness = 14.04, standard error of skew = 0.071; Fig. 3A), suggesting the possible presence of multiple subpopulations within the set.

We estimated in which of these genes the diversity was likely due to homologous recombination between ST131 and non-ST131 strains or, alternatively, accumulation of mutations within

FIG 3 (A) Distribution of the normalized diversity (ratio of NA114-SE15 diversity to between-ST diversity) of the 1,250 core polymorphic genes between NA114 and SE15. (B) Distributions of genes with a putative origin of mutation and recombination (as determined according to the strategy shown in panel C) are shown by red bars and blue bars, respectively. (C) Strategy used to define a possible source of genetic variability for each of the 1,250 polymorphic genes. The scale, which covers the 95% confidence interval of the mean between-ST diversity value (see Materials and Methods for details), shows that a polymorphic gene is considered mutant or recombinant within ST131 if the NA114-SE15 diversity value is below the lower 95% limit (red arrow) and above the lower 50% limit (blue arrow), respectively. The polymorphism in a gene with any intermediate value (i.e., between the 95% and 50% lower limits of the mean) is considered of obscure origin.

the ST131 strains themselves. Any of the polymorphic genes with an NA114-SE15 diversity value above the lower 50% confidence interval value for between-ST diversity was considered a putative recombinant (Fig. 3C). Genes with diversity values below the lower 95% confidence interval for between-ST diversity were regarded to have been derived by mutation (Fig. 3C). The remaining genes with intermediate NA114-SE15 diversity values (i.e., within the range of the lower 50% and 95% confidence intervals) were considered to be of uncertain origin. Of the 1,250 total polymorphic genes, 354 (28.3%) were defined as being affected by homologous recombination (Fig. 3B, blue bars), 545 (43.6%) were defined as being mutationally derived (Fig. 3B, red bars), and the remaining 351 (28.1%) were of uncertain origin. The average pairwise nucleotide diversity of the putative recombinant genes was 1.84%, while the average diversity of the putative mutated genes was 0.3% ($P < 0.001$).

In order to test the validity of the diversity-based assignment of the recombinant and mutational polymorphisms, phylogenetic congruency was determined for 50 genes randomly chosen from

each group. For this, we examined the relative position of the NA114 and SE15 alleles on the phylogenetic tree of the corresponding genes from 23 non-ST131 strains, as described above (Fig. 1; see detail in Materials and Methods). Among the polymorphic genes of likely recombinant origin, in 48 out of 50 genes the NA114 and SE15 alleles were incongruent; i.e., they were separate from each other on the tree and more closely related to an allele from a non-ST131 strain (see the representative gene tree in Fig. S1A in the supplemental material), strongly supporting their recombinant origin. In contrast, among polymorphic genes of likely mutational origin, the NA114 and SE15 alleles of all 50 genes were congruent on the tree; i.e., they were more closely related to one another than to any non-ST131 allele (see the representative gene tree in Fig. S1B in the supplemental material), supporting a mutational rather than recombinational origin. Thus, the original diversity-based assignment of the polymorphic genes shared between NA114 and SE15 into recombinant and mutated groups is valid. We avoided the use of the partially assembled genome because of the large number of existing gaps in them and the use of

FIG 4 (A) The genetic organization of the *fim* operon with flanking genes is depicted for *E. coli* strains NA114 and SE15. Genes are represented by arrows, the direction of which is according to the position of the genes in the sense or complementary strand. The total number of nucleotide changes and their distribution in nonsynonymous and synonymous positions are shown between any compared genes. Red rectangle, putative recombinant region encompassing 8 genes; dotted lines, deletions in NA114. (B) Nucleotide and amino acid (aa) differences (yellow-highlighted cells) between seven different recombinant alleles of *FimH* are presented along with the consensus sequence. Particular amino acid changes and their respective positions in the mature protein are also shown. SNP, single nucleotide polymorphism.

orthologous genes from another partially assembled ST131 strain would lead to the identification of additional recombinant alleles of few genes.

Approximately half (45%) of the putative recombinant genes could be combined into 22 regions, each composed of 3 or more consecutive recombinant genes (see Table S3 in the supplemental material). The recombinant gene blocks, which had an average length of 7.2 kb (Fig. 2, green bars), were widely distributed across both genomes. We analyzed the association between the positions of the recombinant gene blocks and the mosaic regions (i.e., insertions/deletions or replacements). The presence of a recombinant region within 10 kb on either side of any mosaic region was considered a linkage between the recombinant and mosaic regions. A total of 8 of 22 recombinant regions showed linkage with a mosaic region, significantly more than expected by chance alone ($P = 0.015$).

The putative mutation-affected genes likewise were distributed broadly across the genomes (not shown). The mutated loci contained a total of 2,055 nucleotide differences, with an average of 3.8 mutations per gene, which included 1,571 synonymous and 484 nonsynonymous mutations. Overall, the synonymous mutation rate (dS ; $dS = 0.01$) was 10-fold higher than the nonsynonymous mutation rate (dN ; $dN = 0.001$). There were 137 genes accumulating only nonsynonymous mutations. However, no gene was found to have a dN significantly higher than the dS owing to the low number of nonsynonymous mutations per gene.

Thus, about one-third of the genes shared between the two ST131 strains were polymorphic due to mutation or homologous recombination. Additionally, at least some homologous recombination events were positionally linked to the location of mosaic genes in the chromosome.

Homologous recombination in the type 1 fimbrial gene cluster. Both NA114 and SE15 possessed the type 1 fimbrial *fim* cluster

(Fig. 4A). While the 5' portion of the *fim* cluster (including *fimE*AIC) was nearly identical between NA114 and SE15, the remaining 3' portion (including *fimDFGH*) was highly diverse (Fig. 4A). Also, 4 genes immediately downstream of the cluster—*gntP* and *uxuABR*—were highly diverse. On the basis of the diversity thresholds used above, the *fimD*, *fimF*, *uxuB*, and *uxuR* genes of the *fimD-uxuR* region fell into the putative recombinant category; *fimH*, *gntP*, and *uxuA* were of obscure origin, and only *fimG* fell into the putative mutational category (see Table S4 in the supplemental material). However, on the basis of the congruency test, all of these genes were represented by incongruent alleles, supporting their recombinant origin (not shown). Therefore, we hypothesized that the entire *fimD-uxuR* region diverged between NA114 and SE15 by homologous recombination.

Furthermore, close examination of the regions flanking the recombinant regions identified deletion events. *fimB* (coding for the fimbrial switch recombinase) was partially deleted in NA114 (Fig. 4A), with a 35-bp-long remnant of an insertion sequence (IS) element copy being detected (see Fig. S2A in the supplemental material). This IS element was previously described in a subset of ST131 isolates (23). In contrast, the unaltered *fimB* copy in SE15 contained no IS element. However, downstream of the recombinant region, SE15 contained two IS elements (ECSF_4266 and ECSF_4267) and a hypothetical ORF (ECSF_4268) between *uxuR* and *iraD* that were fully or partially deleted in NA114 (Fig. 4A; see Fig. S2B in the supplemental material). We suggest that the *fimD-uxuR* recombination involved the insertion/deletion of the IS elements and nearby genes. Because of the more preserved nature of the flanking regions in SE15, the SE15 copy of the region is likely to be the evolutionary precursor to the copy in NA114.

Thus, at least one recombinational event in the core genes of ST131 strains affected the type 1 fimbrial gene cluster, with recombination into the NA114 ancestral lineage likely occurring.

FIG 5 Maximum-likelihood DNA phylograms of *fimH* and *fimA* alleles. The *fimH* tree was generated on the basis of an alignment of *fimH* sequences from 19 strains and 43 isolates. In the case of *fimA*, the tree was based on the alignment of sequences from 14 strains and 43 isolates. The internal fragments considered here for *fimH* and *fimA* were 64 to 552 bp and 40 to 478 bp, respectively. The recombinant alleles in the corresponding trees are shown by gray boxes, and the mutant variants are marked by the red font. The *fimH*-*fimA* allele pairing is marked by lines linking them. For each mutant derivative, the corresponding changes are marked within parentheses (capital letters for nonsynonymous changes with the respective amino acid position and lowercase letters for synonymous changes with the respective nucleotide position).

Functional effect of recombination and mutation in *fimH* of ST131 strains. The conserved, nonrecombinant region of the *fim* cluster contains *fimA*, encoding the major pilin subunit of type 1 fimbriae. The putative recombinant region contains *fimH*, encoding the minor, adhesive subunit that is located on the fimbrial tip. We determined the *fimH* and *fimA* allelic diversity within ST131 on a population level among 43 different ST131 isolates of diverse clinical origins (see Table S1 in the supplemental material). The pairwise nucleotide diversity (π) of *fimA* alleles ($\pi = 0.08 \pm 0.01$) was, on average, 4 times higher than that of *fimH* alleles ($\pi = 0.02 \pm 0.001$).

We then distinguished which *fimH* or *fimA* alleles were of a likely recombinant or mutational origin among the 43 ST131 isolates. On the basis of the phylogenetic congruence of the corresponding trees, there were 7 main recombinant alleles of *fimH* and 5 of *fimA*, from which minor variants evolved by point mutation (Fig. 5, boxed). In one instance, the same recombinant *fimH* allele was paired with two different *fimA* alleles, whereas in two instances, a given *fimA* allele was paired with multiple different *fimH* alleles (Fig. 5, cross bars). The recombinant alleles of *fimH* had 10 total putative mutational derivatives, whereas *fimA* had only 2. Furthermore, in *fimH* the 10 mutation-evolved alleles were derived by only 1 synonymous mutation but a total of 10 nonsynonymous mutations (Fig. 5), with dN/dS being equal to 3.44 ($P = 0.023$), strongly suggesting the action of positive selection for amino acid changes. We also detected the occurrence of hot-spot mutations (i.e., repeated phylogenetically unlinked changes in the same amino acid position) in mutational derivatives of the

fimH22 recombinant allele. There were different amino acid changes at position 106 for *fimH161* (A $\rightarrow$ V) and *fimH126* (A $\rightarrow$ T) and also different changes at position 41 for *fimH182* (Q $\rightarrow$ R) and *fimH162* (Q $\rightarrow$ H). The structural (amino acid) differences between different recombinant alleles of FimH are depicted in Fig. 4B. In contrast, the *fimA* alleles contained 1 synonymous and 1 nonsynonymous mutation, making it difficult to assess the action of positive selection for amino acid changes.

We determined the functional properties of all recombinant *fimH* alleles and their mutational derivatives in an isogenic background using radionuclide adhesion assay (RAA). The bacteria were bound to monomannose (1M) and trimannose (3M) substrates, and the 1M/3M ratios were calculated for all isogenic strains. Since in this type of assay the level of 3M binding is more or less uniform and much higher than that of 1M binding, the 1M/3M ratio serves as an indicator of a low- or high-binder phenotype (29). Among the primary recombinant alleles, the highest mannose binding was demonstrated by the NA114-like *fimH30* allele, with the binding exceeding, for example, 4-fold the binding of the SE15-like *fimH41* allele (Fig. 6). However, the mutational derivatives of recombinant *fimH* alleles had an increased mannose-binding ability compared to the ancestral allele, except for *fimH196*.

Thus, *fimH* and *fimA* genes within ST131 have diversified by both homologous recombination and mutation, with the FimH adhesin being under positive selection for point amino acid mutations with a distinct functional effect: increased mannose binding of the fimbria.

FIG 6 The 1M/3M binding ratios of seven ST131-*fimH* recombinant alleles (bold font) and their mutational derivatives are shown in the isogenic setting. Different colors represent different recombinant groups.

Clonal association of ST131 strains with NA114-like *fimH*. Considering the high allelic diversity of *fimH*, both recombinational and mutational, we examined the association of ST131 isolates carrying different *fimH* variants with the signature phenotype of the ST131 clonal group, FQ resistance. Among 160 ST131 isolates, the vast majority (92%) of resistant isolates contained either an NA114-like recombinant *fimH* allele (*fimH30*) or a mutational derivative thereof (Table 1). The strong association between *fimH30* and FQ resistance prompted us to determine what mosaic loci are clonally associated with the *fimH30* strains and, thus, could serve as genetic markers for them. For this purpose, we designed primers for several mosaic loci specific to NA114 (i.e., absent in SE15) and tested them with the set of 160 ST131 isolates (Table 1).

One set of primers targeted the Flag2 region, which is present in NA114 but not SE15. Though these primers recognized all strains from the *fimH30* isolates and none from the SE15-like

fimH41 isolates, they also yielded a band in most strains from other *fimH* subclones (Table 1). Another set of primers was designed to detect the presence of *papGII* and *papA9*, which are part of the P-fimbrial operon in the 160-kb pathogenicity island present in NA114 but not SE15. Only 8 (9.3%) of 86 isolates from the *fimH30* isolates were recognized by the *papGII* primers, and none was recognized by the *papA9* primers.

The mosaic loci tested as described above are positioned relatively far from the *fim* recombinant region. Thus, we also tested primers designed to be specific for the mosaic regions linked to the *fim* locus. The IS element in *fimB* was present in almost all *fimH30* isolates but was present in none of the non-*fimH30* ST131 isolates, showing a strong clonal association of this marker with the *fimH30* strains. Primers for another NA114-specific region linked to the *fim* cluster, encoding RMT1 and positioned 22.4 kb downstream of the *fim* operon, showed an even tighter, complete clonal association with the *fimH30* isolates. However, when a set of 84 non-ST131 strains was tested with the IS- and RMT1-specific primers, 6 (7.1%) of 84 were positive with both primer sets, indicating the limited utility of these regions as highly discriminatory genetic markers for ST131 strains with the *fimH30* allele.

Considering that primers for the two *fimH30*-linked loci also recognized some non-ST131 strains, we searched for an additional locus to discriminate ST131 and non-ST131 strains in general. Among the 3,997 genes shared by strains NA114 and SE15, 94 were absent from the genomes of the non-ST131 strains depicted in Fig. 1. Although the vast majority of these genes were of a hypothetical nature (not shown), at least two genes, which were contiguous (SE15 genes ECSF_0249 and ECSF_0250, spanning the coding DNA sequence [CDS] region from positions 279314 to 284040), had a putatively defined function. These loci were highly homologous to genes coding a type 1 restriction modification system which was genetically unrelated to the NA114-specific RMT1 described above. Thus, we designated the region of these two contiguous genes RMT1A. Primers specific for RMT1A detected 93.1% (149/160) of ST131 isolates but none of 84 non-ST131 strains, confirming the precise specificity of these genes for ST131 (Table 1). Furthermore, RMT1A-specific primers recognized 94% (81/86) of ST131 isolates with *fimH30*.

Thus, taken together, these results indicate that NA114-like

TABLE 1 Clone/ST-specific distribution of different markers

Trait ^a	No. (%) of strains with the associated trait		
	ST131 isolates		
	<i>fimH30</i> strains (<i>n</i> = 86)	Non- <i>fimH30</i> strains (<i>n</i> = 74)	Non-ST131 isolates (<i>n</i> = 84)
Fluoroquinolone resistance	82 (100)	7 (100)	9 (100)
Flag2 ^{pos}	86 (100)	49 (71)	ND ^c
Flag2 ^{neg}	0 (0)	25 (29)	ND
<i>papGII</i>	8 (10)	7 (0)	ND
<i>papA</i> F9	0 (0)	0 (0)	ND
<i>fimB</i> insertion	85 (100)	0 (0)	ND
RMT1 H30	86 (100)	0 (0)	6 (11)
RMT1A ST131	81 (94)	68 (100)	0 (0)

^a Flag2^{pos}, specific for a junction of Flag2 locus and its flanking sequence, *fimH30* linked; Flag2^{neg}, specific for sequences flanking the Flag2 locus; *papGII*, specific for *papG* allele II; *papA* F9, specific for *papA* allele F9 *fimB* insertion, specific for *fimB* insertion; RMT1 H30, specific for restriction modification system type 1, *fimH30* linked; RMT1A ST131, specific for restriction modification system type 1 (specific for ST131).

^b Percentage of the fluoroquinolone-resistant strains in that group.

^c ND, not determined.

fimH30 is clonally linked to the fluoroquinolone resistance of ST131 strains and to mosaic genes (e.g., RMT1) that are spatially linked to the *fim* cluster. In contrast, RMT1A mosaic genes are spatially linked to ST131 strains, and their presence is not significantly different in *fimH30* (94%) and non-*fimH30* (92%) strains.

DISCUSSION

This genome-wide comparison of two strains from the dominant antimicrobial-resistant ST131 clonal group of *E. coli*, with supplemental analysis of 160 additional ST131 strains and 84 non-ST131 strains, led to four main findings. First, the two ST131 genome strains are highly divergent, with one-third of the genes being different due to roughly equal contributions of horizontal gene acquisition, homologous recombination, and point mutation. Second, all three of these genetic mechanisms have contributed to pathoadaptive diversification within the clonal group. Third, the FimH tip adhesin of type 1 fimbriae is under strong positive selection for functional changes. Fourth, within ST131, fluoroquinolone resistance is manifested by a clonally linked group of strains.

Despite NA114 and SE15 both being from ST131 and by definition having identical MLST profiles, a sizable portion of their genomes is mosaic in nature: 8% of the SE15 genome and 15% of the NA114 genome have no orthologous regions in the other strain. In fact, the number of shared genes might not be a reliable measure of the phylogenetic relatedness of strains since, for example, in some instances as little as 10% of the SE15 genome and 18% of the NA114 genome were different (having genes with >5% nucleotide diversity) from the genomes of non-ST131 strains.

Most of the mosaic genes between NA114 and SE15 were of bacteriophage origin. However, among the nonphage mosaic genes, some were clearly of physiological and/or virulence significance. For example, the loci encoding lipopolysaccharide are very different in NA114 and SE15, corresponding with their different O antigens, i.e., O25 and O150, respectively. Other mosaic genes between NA114 and SE15 encode putative virulence-related traits of ExPEC. Interestingly, most of such genes are found only in NA114. They include genes coding for lateral flagella (second copy), iron uptake systems, secreted toxins, and adhesins. In addition, the presence of the multidrug efflux system in NA114 but not SE15 could result in differences in antimicrobial resistance between the strains. Indeed, although the susceptibility profile of SE15 has not been reported, NA114 was reported as being multidrug resistant (including resistance to amoxicillin, gentamicin, and nitrofurantoin) (31), with the efflux system likely contributing to resistance. Taken together, the mosaic gene differences are likely to result in a higher pathogenic potential for strain NA114 than strain SE15, consistent with the former being a urinary clinical isolate and the latter being a fecal commensal strain. However, a more refined analysis is needed to pinpoint the virulence differences between these two ST131 strains and to identify the directly contributory mosaic genes.

Another genetic mechanism of genome diversification is homologous recombination, which is the exchange of different alleles of orthologous genes between strains, while horizontal transfer is the movement of nonorthologous genes between the strains. Here, the presence of recombinant genes among the genes shared by NA114 and SE15 is supported by (i) the high diversity between NA115 and SE15 of certain genes that is comparable to the diversity of the same gene between different STs of *E. coli*; (ii) the incongruency of the NA114 and SE15 gene copies, i.e., their sep-

aration on different branches of the phylogenetic tree of the gene from diverse *E. coli* strains; (iii) the tendency of such genes to be clustered into contiguous regions on the NA114 and SE15 chromosomes; and (iv) positional linkage of the putative recombinant gene clusters with mosaic genes of a clearly mobile nature.

It is highly possible that many of the genes with polymorphisms of uncertain (obscure) origin, along with some of the ones defined to have a mutational origin, are actually recombinants. Also, we cannot exclude the possibility that some portion of the genes defined as recombinant between NA114 and SE15 are of mutational origin due to a potentially high mutation rate for some genes in one or both of these ST131 strains. Apart from these, the internal mutation events may also occur at the postrecombination level. Still, the assignment of polymorphic orthologous genes into those putatively derived by recombination versus those putatively derived by mutation allows comparisons of the different mechanisms to be made. First, approximately the same numbers of genes were polymorphic due to either recombination or mutation, so in this respect, the contribution of these processes to genomic diversity (at the gene level) is similar. However, considering that each separate recombinant gene or gene block is likely to represent a single homologous recombination event, we can estimate that approximately 112 recombination events into either NA114 or SE15 have occurred. This is 18 times lower than the number of mutations, which corresponds to previous estimates for a significantly higher rate of mutation than recombination in *E. coli* (32). However, in one study it was estimated that recombination accounts for more diversity than mutation in a subset of ECOR group A strains (33).

Interestingly, the location of many recombinant gene blocks is linked to the position of genetic insertions, deletions, or replacements, i.e., regions where mobile genetic elements are present. This indicates that horizontal gene movement and homologous recombination could be concurrent genomic events. One possibility is that acquisition of recombinant genes is a result of genetic drift, with horizontal movement of mobile genetic elements being the selective force. Alternatively, orthologous gene recombination itself may be under positive selection. In other words, recombinant alleles that move in could provide a fitness advantage relative to the fitness that occurs with the allele that is being replaced. Notably, many recombinant regions contain genes with potentially virulence-relevant functions, including transport proteins, transcriptional regulators, two-component system proteins, type II secretion proteins, etc. (see Table S3 in the supplemental material).

One of the NA114-SE15 shared regions affected by homologous recombination is the *fim* operon, encoding mannose-specific type 1 fimbriae, which is one of the most ubiquitous adhesive organelles of *E. coli*, being expressed by both intestinal commensals and extraintestinal pathogens. In intestinal isolates, type 1 fimbriae are proposed to be important for transient oropharyngeal colonization during fecal-oral transmission (34). In uropathogenic isolates, type 1 fimbriae were shown to be an important colonization factor in the urinary bladder and kidneys (34–36). The presence of IS elements (or their remnants) and gene deletions flanking the recombinant region supports the role of mobile genetic elements in homologous gene recombination. Also, because gene deletions are not reversible by nature, the fact that they are present in NA114 and the same regions are unaltered in SE15 indicates that *fim*

recombination into NA114 occurred, replacing the gene region in SE15.

Strain NA114 was not in our collection to test whether fimbriae are expressed when a large portion of *fimB* and most of the inserted IS element in it are deleted. However, unlike NA114, all other ST131 strains that are clonally related to the latter (based on the *fimH* allele) have an intact IS element inserted in *fimB*. Our study with these strains indicated that the presence of an IS element in *fimB* does not completely abrogate the fimbrial expression. This corresponds to the recently obtained results that have shown that type 1 fimbrial expression can be determined in ST131 strains with an *fimB* insertion (23). It was proposed that insertion in *fimB* leads to a slower switching on for type 1 fimbrial expression and involves another tyrosine recombinase in the process. Similarly, since we do not possess the SE15 strain in our collection, we tested the strains which are clonally related to SE15. We found that all these strains had intact *fimB* and express functional type 1 fimbriae.

Somewhat surprisingly, this recombination involving the *fim* region did not affect the gene encoding the major subunit, FimA, which is identical in both strains. FimA is a nonadhesive major structural subunit of the fimbrial shaft and is known to have the highest strain-to-strain diversity among all *fim* genes (37). FimA structural diversity, which reaches 8%, is likely to be driven by frequency-dependent selection for antigenic variation (10, 38).

The ST131 *fim* homologous recombination event instead replaced the operon region downstream from *fimA* that includes the gene encoding FimH, the 30-kDa adhesive subunit of type 1 fimbriae. Overall, allelic variation within *fimH* is much more limited than that within *fimA* and is on a par with the variation in the rest of the *fim* genes or in housekeeping core genes in general (39). However, naturally occurring single-amino-acid replacement mutations in FimH were shown to lead to an increased ability of the adhesin to bind mannoseylated glycoproteins on the surface of eukaryotic cells (10). This, in turn, leads to an increased shear-independent adhesion of the fimbriated bacteria to the target cells (40). The mutated forms of FimH were shown to be positively selected in uropathogenic *E. coli* strains and to result in increased bacterial tropism to urothelium and bladder colonization in a mouse model of urinary tract infection (41).

The recombinant copies of FimH in NA114 and SE15 differ in two amino acids, with the NA114 variant mediating significantly higher mannose-specific adherence, a typical phenotype for uropathogenic *E. coli* isolates. At the same time, the low static adherence of the SE15 FimH variant is more typical of intestinal commensal *E. coli* isolates. These phenotypes correspond well with the source of isolation of both strains. Thus, we hypothesize that the homologous recombination in the *fim* region of ST131 could have driven, at least in part, the clonal expansion of *fimH30*-carrying strains by selection for the highly binding *fimH* allele shown to be associated with the increased urovirulence of *E. coli*. However, the origin of *fimH30* remains to be determined, as we have still failed to identify *E. coli* strains carrying an allele that is the evolutionary precursor to *fimH30* (i.e., strains without the H166 mutation).

We also demonstrated here on the population level the action of positive selection in ST131 strains for increased FimH mannose binding by comparing *fimH* in multiple different ST131 isolates. This population genetics analysis showed that *fimH* is under much stronger positive selection for point structural changes than *fimA*, despite the greater overall diversity of *fimA*. Although, besides

fimH, other genes in the *fim* cluster or its flanking regions could be under positive selection, we believe that the *fimH* example demonstrates well that both recombination and mutation of orthologous genes could lead to functional and, thus, adaptive diversification of strains within the same ST.

It is likely that the high-mannose-binding *fimH* variant of NA114 could increase this strain's overall virulence. However, it is unclear to what extent other ST131 strains that carry the same *fimH* allele (*fimH30*) share the virulence factors specifically found in NA114. In fact, we demonstrated here that at least some of the putative virulence determinants of NA114 (such as the second lateral flagellum genes) either are not limited to or are not conserved among *fimH30*-carrying strains. Among the latter are genes coding for the P-fimbrial major subunit (*papA9*) and adhesin (*papGII*), which are present as part of the *pap* operon on the 160-kb pathogenicity island of NA114. Thus, it is possible that this island or at least parts of it are missing in most *fimH30* strains.

One of the most prominent characteristics of ST131 is its fluoroquinolone resistance, to which this clonal group is the major contributor worldwide (9, 42–47). We have recently demonstrated that this phenotype is predominantly associated with strains carrying an NA114-like *fimH30* allele (48). Here, we show that two mosaic gene regions—an IS insertion in *fimB* and one of the restriction modification systems (RMT1)—that are proximal to *fimH* are also clonally linked to *fimH30* strains within ST131. Interestingly, however, both the *fimB* insertion and RMT1 are found in some non-ST131 strains that also have the *fimH30* allele, in particular, those from another prominent ExPEC clonal group, ST73 (not shown). Thus, the entire region appears to move horizontally between different STs. The evolutionary pathway of the recombinant region transfer between ST131 and ST73 and this region's contribution to the pathogenic potential of *fimH30* strains are unknown. However, both the *fimB* insertion and RMT1 could serve as convenient genetic markers for molecular detection of fluoroquinolone-resistant variants within ST131 strains, especially in combination with an ST131-specific marker such as RMT1A, as shown here.

ACKNOWLEDGMENTS

We gratefully acknowledge Steve Moseley (University of Washington) for critical reading of the manuscript and helpful advice. We also acknowledge the anonymous donors for clinical isolates.

REFERENCES

- Clements A, Young JC, Constantinou N, Frankel G. 2012. Infection strategies of enteric pathogenic *Escherichia coli*. *Gut Microbes* 3:71–87.
- Pitout JD. 2012. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. *Frontiers Microbiol.* 3:9. doi:10.3389/fmicb.2012.00009.
- Eisenstein BI. 1989. Principles and practice of infectious disease, p 1658–1673. In Mandell GL, Dangles RG, Bennet JE (ed), vol 2. Churchill Livingstone, New York, NY.
- Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D, Donnelly P. 2011. Recombination and population structure in *Salmonella enterica*. *PLoS Genet.* 7:e1002191. doi:10.1371/journal.pgen.1002191.
- Ochman H, Selander RK. 1984. Evidence for clonal population structure in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 81:198–201.
- Chaudhuri RR, Henderson IR. 2012. The evolution of the *Escherichia coli* phylogeny. *Infect. Genet. Evol.* 12:214–226.
- Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* 60:1136–1151.

8. Blanco J, Mora A, Mamani R, Lopez C, Blanco M, Dahbi G, Herrera A, Blanco JE, Alonso MP, Garcia-Garrote F, Chaves F, Orellana MA, Martinez-Martinez L, Calvo J, Prats G, Larrosa MN, Gonzalez-Lopez JJ, Lopez-Cerero L, Rodriguez-Bano J, Pascual A. 2011. National survey of *Escherichia coli* causing extraintestinal infections reveals the spread of drug-resistant clonal groups O25b:H4-B2-ST131, O15:H1-D-ST393 and CGA-D-ST69 with high virulence gene content in Spain. *J. Antimicrob. Chemother.* 66:2011–2021.
9. Johnson JR, Porter SB, Zhanel G, Kuskowski MA, Denamur E. 2012. Virulence of *Escherichia coli* clinical isolates in a murine sepsis model in relation to sequence type ST131 status, fluoroquinolone resistance, and virulence genotype. *Infect. Immun.* 80:1554–1562.
10. Weissman SJ, Chattopadhyay S, Aprikan P, Obata-Yasuoka M, Yarova-Yarovaya Y, Stapleton A, Ba-Thin W, Dykhuizen D, Johnson JR, Sokurenko EV. 2006. Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. *Mol. Microbiol.* 59:975–988.
11. Noller AC, McEllistrem MC, Stine OC, Morris JG, Jr, Boxrud DJ, Dixon B, Harrison LH. 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J. Clin. Microbiol.* 41:675–679.
12. Hancock V, Vejborg RM, Klemm P. 2010. Functional genomics of probiotic *Escherichia coli* Nissle 1917 and 83972, and UPEC strain CFT073: comparison of transcriptomes, growth and biofilm formation. *Mol. Genet. Genomics* 284:437–454.
13. Vejborg RM, Friis C, Hancock V, Schembri MA, Klemm P. 2010. A virulent parent with probiotic progeny: comparative genomics of *Escherichia coli* strains CFT073, Nissle 1917 and ABU 83972. *Mol. Genet. Genomics* 283:469–484.
14. Cagnacci S, Gualco L, Debbia E, Schito GC, Marchese A. 2008. European emergence of ciprofloxacin-resistant *Escherichia coli* clonal groups O25:H4-ST 131 and O15:K52:H1 causing community-acquired uncomplicated cystitis. *J. Clin. Microbiol.* 46:2605–2612.
15. D'Astek BA, del Castillo LL, Miliwebsky E, Carbonari C, Palladino PM, Deza N, Chinen I, Manfredi E, Leotta GA, Masana MO, Rivas M. 2012. Subtyping of *Escherichia coli* O157:H7 strains isolated from human infections and healthy cattle in Argentina. *Foodborne Pathog. Dis.* 9:457–464.
16. Lee MY, Choi HJ, Choi JY, Song M, Song Y, Kim SW, Chang HH, Jung SI, Kim YS, Ki HK, Son JS, Kwon KT, Heo ST, Yeom JS, Shin SY, Chung DR, Peck KR, Song JH, Ko KS. 2010. Dissemination of ST131 and ST393 community-onset, ciprofloxacin-resistant *Escherichia coli* clones causing urinary tract infections in Korea. *J. Infect.* 60:146–153.
17. Platell JL, Trott DJ, Johnson JR, Heisig P, Heisig A, Clabots CR, Johnston B, Cobbold RN. 2012. Prominence of an O75 clonal group (clonal complex 14) among non-ST131 fluoroquinolone-resistant *Escherichia coli* causing extraintestinal infections in humans and dogs in Australia. *Antimicrob. Agents Chemother.* 56:3898–3904.
18. Johnson JR, Johnston B, Clabots C, Kuskowski MA, Castanheira M. 2010. *Escherichia coli* sequence type ST131 as the major cause of serious multidrug-resistant *E. coli* infections in the United States. *Clin. Infect. Dis.* 51:286–294.
19. Peirano G, Pitout JD. 2010. Molecular epidemiology of *Escherichia coli* producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4. *Int. J. Antimicrob. Agents* 35:316–321.
20. Rogers BA, Sidjabat HE, Paterson DL. 2011. *Escherichia coli* O25b-ST131: a pandemic, multiresistant, community-associated strain. *J. Antimicrob. Chemother.* 66:1–14.
21. Avasthi TS, Kumar N, Baddam R, Hussain A, Nandanwar N, Jadhav S, Ahmed N. 2011. Genome of multidrug-resistant uropathogenic *Escherichia coli* strain NA114 from India. *J. Bacteriol.* 193:4272–4273.
22. Toh H, Oshima K, Toyoda A, Ogura Y, Ooka T, Sasamoto H, Park SH, Iyoda S, Kurokawa K, Morita H, Itoh K, Taylor TD, Hayashi T, Hattori M. 2010. Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. *J. Bacteriol.* 192:1165–1166.
23. Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK, Bachmann N, Szubert M, Sidjabat HE, Paterson DL, Upton M, Schembri MA. 2011. Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One* 6:e26578. doi:10.1371/journal.pone.0026578.
24. Clark G, Paszkiewicz K, Hale J, Weston V, Constantinidou C, Penn C, Achtman M, McNally A. 2012. Genomic analysis uncovers a phenotypically diverse but genetically homogeneous *Escherichia coli* ST131 clone circulating in unrelated urinary tract infections. *J. Antimicrob. Chemother.* 67:868–877.
25. Yi H, Cho YJ, Hur HG, Chun J. 2011. Genome sequence of *Escherichia coli* AA86, isolated from cow feces. *J. Bacteriol.* 193:3681.
26. Weissman SJ, Johnson JR, Tchesnokova V, Billig M, Dykhuizen D, Riddell K, Rogers P, Qin X, Butler-Wu S, Cookson BT, Fang FC, Scholes D, Chattopadhyay S, Sokurenko E. 2012. High-resolution two-locus clonal typing of extraintestinal pathogenic *Escherichia coli*. *Appl. Environ. Microbiol.* 78:1353–1360.
27. Swofford DL. 2000. PAUP*: phylogenetic analysis using parsimony and other methods (software). Sinauer Associates, Sunderland, MA.
28. Sokurenko EV, Schembri MA, Trintchina E, Kjaergaard K, Hasty DL, Klemm P. 2001. Valency conversion in the type 1 fimbrial adhesin of *Escherichia coli*. *Mol. Microbiol.* 41:675–686.
29. Sokurenko EV, Chesnokova V, Doyle RJ, Hasty DL. 1997. Diversity of the *Escherichia coli* type 1 fimbrial lectin. Differential binding to mannose and uroepithelial cells. *J. Biol. Chem.* 272:17880–17886.
30. Tchesnokova V, Aprikan P, Kisiela D, Gowsy S, Korotkova N, Thomas W, Sokurenko E. 2011. Type 1 fimbrial adhesin FimH elicits an immune response that enhances cell adhesion of *Escherichia coli*. *Infect. Immun.* 79:3895–3904.
31. Jadhav S, Hussain A, Devi S, Kumar A, Parveen S, Gandham N, Wieler LH, Ewers C, Ahmed N. 2011. Virulence characteristics and genetic affinities of multiple drug resistant uropathogenic *Escherichia coli* from a semi urban locality in India. *PLoS One* 6:e18063. doi:10.1371/journal.pone.0018063.
32. Milkman R, Bridges MM. 1990. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. *Genetics* 126:505–517.
33. Guttman DS, Dykhuizen DE. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* 266:1380–1383.
34. Bloch CA, Stocker BA, Orndorff PE. 1992. A key role for type 1 pili in enterobacterial communicability. *Mol. Microbiol.* 6:697–701.
35. Bahrani-Mougeot FK, Buckles EL, Lockatell CV, Hebel JR, Johnson DE, Tang CM, Donnenberg MS. 2002. Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia coli* virulence determinants in the murine urinary tract. *Mol. Microbiol.* 45:1079–1093.
36. Bloch CA, Orndorff PE. 1990. Impaired colonization by and full invasiveness of *Escherichia coli* K1 bearing a site-directed mutation in the type 1 pilin gene. *Infect. Immun.* 58:275–278.
37. Boyd EF, Hartl DL. 1998. Diversifying selection governs sequence polymorphism in the major adhesin proteins fimA, papA, and sfaA of *Escherichia coli*. *J. Mol. Evol.* 47:258–267.
38. Peek AS, Souza V, Eguarte LE, Gaut BS. 2001. The interaction of protein structure, selection, and recombination on the evolution of the type-1 fimbrial major subunit (fimA) from *Escherichia coli*. *J. Mol. Evol.* 52:193–204.
39. Sokurenko EV, Chesnokova V, Dykhuizen DE, Ofek I, Wu XR, Krogfelt KA, Struve C, Schembri MA, Hasty DL. 1998. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc. Natl. Acad. Sci. U. S. A.* 95:8922–8926.
40. Thomas WE, Trintchina E, Forero M, Vogel V, Sokurenko EV. 2002. Bacterial adhesion to target cells enhanced by shear force. *Cell* 109:913–923.
41. Weissman SJ, Beskhlebnaya V, Chesnokova V, Chattopadhyay S, Stamm WE, Hooton TM, Sokurenko EV. 2007. Differential stability and trade-off effects of pathoadaptive mutations in the *Escherichia coli* FimH adhesin. *Infect. Immun.* 75:3548–3555.
42. Dhanji H, Doumith M, Rooney PJ, O'Leary MC, Loughrey AC, Hope R, Woodford N, Livermore DM. 2011. Molecular epidemiology of fluoroquinolone-resistant ST131 *Escherichia coli* producing CTX-M extended-spectrum beta-lactamases in nursing homes in Belfast, UK. *J. Antimicrob. Chemother.* 66:297–303.
43. Johnson JR, Johnston B, Clabots C, Kuskowski MA, Pendyala S, Debroy C, Nowicki B, Rice J. 2010. *Escherichia coli* sequence type ST131 as an emerging fluoroquinolone-resistant uropathogen among renal transplant recipients. *Antimicrob. Agents Chemother.* 54:546–550.
44. Platell JL, Cobbold RN, Johnson JR, Heisig A, Heisig P, Clabots C, Kuskowski MA, Trott DJ. 2011. Commonality among fluoroquinolone-resistant sequence type ST131 extraintestinal *Escherichia coli* isolates

- from humans and companion animals in Australia. *Antimicrob. Agents Chemother.* 55:3782–3787.
45. Simner PJ, Hoban DJ, Lagace-Wiens PR, Zhanel GG, Karlowsky JA. 2011. Fluoroquinolone resistance in *Escherichia coli* isolated from patients attending Canadian hospitals is associated with the ST131 clone. *Diagn. Microbiol. Infect. Dis.* 71:323–324.
 46. Williamson DA, Roberts SA, Paterson DL, Sidjabat H, Silvey A, Masters J, Rice M, Freeman JT. 2012. *Escherichia coli* bloodstream infection after transrectal ultrasound-guided prostate biopsy: implications of fluoroquinolone-resistant sequence type 131 as a major causative pathogen. *Clin. Infect. Dis.* 54:1406–1412.
 47. Yokota S, Sato T, Okubo T, Ohkoshi Y, Okabayashi T, Kuwahara O, Tamura Y, Fujii N. 2012. Prevalence of fluoroquinolone-resistant *Escherichia coli* O25:H4-ST131 (CTX-M-15-nonproducing) strains isolated in Japan. *Chemotherapy* 58:52–59.
 48. Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, Riddell K, Rogers P, Xuan Qin X, Butler-Wu S, Price LB, Maliha Aziz M, Nicolas-Chanoine MH, DebRoy C, Robicsek A, Hansen G, Urban C, Platell J, Trott D, Zhanel G, Weissman SJ, Cookson BT, Fang FC, Limaye A, Scholes D, Chattopadhyay S, Hooper DC, Sokurenko EV. Abrupt emergence of a single dominant multi-drug-resistant strain of *Escherichia coli*. *J. Infect. Dis.*, in press.