Molecular Analysis of Asymptomatic Bacteriuria *Escherichia coli* Strain VR50 Reveals Adaptation to the Urinary Tract by Gene Acquisition

Scott A. Beaton, a,b Nouri L. Ben Zakour, a,b Makrina Totsika, a,b,x Brian M. Forde, a,b Rebecca E. Watts, a,b,x Amanda N. Mabbett, a,b Jan M. Szubert, a,b Sohinee Sarkar, a,b Minh-Duy Phan, a,b Kate M. Peters, a,b Nicola K. Petty, a,b,x Nabil-Fareed Allikhan, a,b Mitchell J. Sullivan, a,b Jayde A. Gawthorne, a,b Mitchell Stanton-Cook, a,b Nguyet Thi Khanh Nhu, a,b, c Teik Min Chong, c Wai-Fong Yin, c Kok-Gan Chan, c Viktoria Hancock, d,e David W. Ussery, d,e Glen C. Ulett, b,e Mark A. Schembri a,b

Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, Queensland, Australia; School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia; Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia; Centre for Biological Sequence Analysis, DTU Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark; Griffith Health Institute and School of Medical Science, Griffith Health Centre, Griffith University, Gold Coast, Queensland, Australia

Urinary tract infections (UTIs) are among the most common infectious diseases of humans, with *Escherichia coli* responsible for 80% of all cases. One extreme of UTI is asymptomatic bacteriuria (ABU), which occurs as an asymptomatic carrier state that resembles commensalism. To understand the evolution and molecular mechanisms that underpin ABU, the genome of the ABU *E. coli* strain VR50 was sequenced. Analysis of the complete genome indicated that it most resembles *E. coli* K-12, with the addition of a 94-kb genomic island (GI-VR50-*pheV*), eight prophages, and multiple plasmids. GI-VR50-*pheV* contains genes encoding a number of UTI-associated virulence factors, namely, Afa (afimbrial adhesin), two autotransporter proteins (Ag43 and Sat), and aerobactin. We demonstrated that the presence of this island in VR50 confers its ability to colonize the murine bladder, as a VR50 mutant with GI-VR50-*pheV* deleted was attenuated in a mouse model of UTI in vivo. We established that Afa is the island-encoded factor responsible for this phenotype using two independent deletion (Afa operon and AfaE adhesin) mutants. *E. coli* VR50afa and VR50afaE displayed significantly decreased ability to adhere to human bladder epithelial cells. In the mouse model of UTI, VR50afa and VR50afaE displayed reduced bladder colonization compared to wild-type VR50, similar to the colonization level of the GI-VR50-*pheV* mutant. Our study suggests that *E. coli* VR50 is a commensal-like strain that has acquired fitness factors that facilitate colonization of the human bladder.

Received 20 October 2014 Returned for modification 14 November 2014 Accepted 9 January 2015


Editor: B. A. McCormick
Address correspondence to Scott A. Beaton, sc.beaton@uq.edu.au, or Mark A. Schembri, m.schembri@uq.edu.au.

* Present address: Makrina Totsika, Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology, Kelvin Grove, QLD, Australia; Rebecca E. Watts, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; Nicola K. Petty, The ithree Institute, University of Technology Sydney, Sydney, NSW, Australia; Viktoria Hancock, Therapeutic Fluid Research Group, Gamble Lundia AB, Lund, Sweden; David W. Ussery, Biosciences Division, Oak Ridge National Laboratories, Oak Ridge, Tennessee, USA.

S.A.B., N.L.B.Z., and M.T. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/IAI.02810-14.

Copyright © 2015, American Society for Microbiology. All Rights Reserved. doi:10.1128/IAI.02810-14.
ulence genes, such as those encoding type 1 fimbriae, P fimbriae, or hemolysin, but do not express the associated phenotype (9–12). Molecular and genomic analyses of the ABU prototype \( E. coli \) strain 83972 also support this observation. \( E. coli \) strain 83972 is a clinical isolate capable of long-term bladder colonization, as shown in human inoculation studies (13–15). The virulence-associated adhesion genes of \( E. coli \) strain 83972 have become attenuated through a series of independent mutational events. The type 1 fimbrial system has been inactivated by a major deletion encompassing 4.5 kb of the \( fim \) gene cluster affecting all genes except those encoding the minor components, namely, \( fimF, fimG, \) and \( fimH \) (16). The FIC system has been inactivated by point mutations in the fimbrial transport system affecting the \( focD \) gene encoding the usher protein (17). Finally, the \( P. fimbrial \) system has been inactivated by point mutations in the \( papG \) gene that have rendered the PapG adhesin nonfunctional (16). Thus, the ancestor of \( E. coli \) strain VR50, which belongs to the B2 clonal group, was most likely a virulent UPEC strain that has become attenuated (18). Recently, \( E. coli \) strain 83972 adaptation to the human urinary tract has also been linked to its ability to suppress RNA polymerase II–dependent host gene expression following human bladder colonization, thereby preventing the activation of host proinflammatory responses during infection (19).

A second group of \( E. coli \) strain VR50 is most closely related to commensal \( E. coli \) strains, particularly the K-12 strain MG1655, with the major difference being the presence of mobile genetic elements, including a large 94-kb genomic island (GI-VR50-\( pheV \)), prophage regions, and plasmids. GI-VR50 has a mosaic composition of genes, and we show that the genes encoding the Afa adhesin (which are located within GI-VR50-\( pheV \)) contribute significantly to VR50 adherence to bladder epithelial cells and colonization of the mouse bladder.

**MATERIALS AND METHODS**

**Bacterial strains and growth conditions.** \( E. coli \) strain VR50 (serotype OR:K1:K-1) was isolated from a 30-year-old otherwise healthy woman who had carried it for at least 1 year without any symptoms (20). \( E. coli \) strain 83972 and UPEC CFT073 have been described previously (21, 22). Bacteria were routinely grown at 37°C on solid or in liquid Luria-Bertani (LB) medium supplemented with appropriate antibiotics unless otherwise stated (23).

**DNA manipulations and genetic techniques.** Chromosomal DNA was purified using the UltraClean Microbial DNA isolation kit (Mo Bio Laboratories). PCR was performed using Taq polymerase according to the manufacturer’s instructions (Roche, Australia). Sequencing of PCR products was performed by the Australian Genome Research Facility. The \( E. coli \) strain VR50 phylogenetic group of origin was determined using a three-locus PCR-based method (24).

**Construction of mutants.** \( E. coli \) strain VR50 deletion mutants (VR50\( pheV \)-GI, VR50afaa, VR50afae, VR50fim, and VR50\( fimH \)) were constructed using the \( \lambda \)-Red recombinase gene replacement system (25). Briefly, the kanamycin gene from plasmid pKD4 was amplified using primers containing 50-nucleotide (nt) homology extensions to the beginnings and ends of the gene clusters to be deleted. The following primers were used: VR50\( pheV \)-GI, 1772 (5' - TCTGCTGAACACACTCAGCATTAGTTCTAGGCTGAATCAGGCTGCTGATTCAGTGGTGTGAGCTGCTTC) and 1673 (5' - CACCTAGGCGAGCTCTATGACGAAGCTGCTTC) for VR50afaa, 490 (5' - ATGGAGGAGGAGCTATGCTGATCCGACTACACGCGGAGGATGACGTTGTA) and 308 (5' - CACACACTCAGCATTAGTTCTAGGCTGAATCAGGCTGCTGATTCAGTGGTGTGAGCTGCTTC) for VR50fim. The monolayers were washed with PBS and inoculated with bacteria containing the homology arms and selection of kanamycin-resistant colonies. The kanamycin cassette was then removed using plasmid pCP20 (25). All deletions were confirmed by PCR and sequencing.

**Agglutination assays.** Afa-mediated mannose-resistant hemagglutination (MRHA) was assessed as previously described (26). Briefly, a 5% suspension (10 μl) of human type A red blood cells (RBCs) washed in phosphate-buffered saline (PBS) was mixed with a 10-μl bacterial suspension on glass slides in the presence and absence of α-mannose. The bacterial suspension was prepared by transferring cells from a freshly grown LB agar colony into 50 μl PBS.

**Adhesion assays.** \( E. coli \) binding to human HeLa and T24 bladder epithelial cells was assessed by a quantitative adhesion assay using confluent monolayers of epithelial cells (12). Epithelial cells were seeded into 24-well cell culture plates (Corning Inc., NY) at a concentration of 3 × 10⁵ cells per well and incubated overnight at 37°C in 5% CO₂ prior to infection. The monolayers were washed with PBS and inoculated with bacteria at a multiplicity of infection of 50 bacteria per epithelial cell for HeLa cell assays and 10 bacteria per epithelial cell for T24 cell assays. The bacteria were prepared by washing an overnight culture grown in LB once in PBS (3,500 × g; 15 min; 4°C) and resuspended to a concentration of 1.5 × 10⁹ CFU/ml in RPMI 1640 supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 2 mM glutamine (HeLa cell assays) or McCoy’s 5A medium (T24 cell assays). The infected monolayers were incubated at 37°C for 2 h (HeLa cells) or 1 h (T24 cells) in 5% CO₂ and then washed four times with PBS to remove nonadherent bacteria. The monolayers (n = 4) were lysed with 0.01% Triton X-100 in distilled water, and the lysates were diluted in PBS to obtain quantitative colony counts of bacteria on agar. Data pooled from a minimum of three independent experiments are presented as the mean number of CFU/ml plus the standard error of the mean (SEM).

**Genome sequencing and assembly.** The genome sequence of \( E. coli \) strain VR50 was determined using a combination of 454 (Roche) pyrosequencing (~24× coverage) and traditional Sanger sequencing (~1× coverage) (Australian Genome Research Facility). The following reads were included in the final assembly. (i) 1.2 million 454 GS20 shotgun reads (average read length, 101 bp); (ii) 85,155 454 GS-FLX 20-bp mate pair reads (average distance between reads, 2 kb); (iii) 7,680 Sanger mate pair reads from a PacBio clone library with an insert size range of 5 to 8 kb; and (iv) 384 Sanger end sequences from 192 gap-spanning PCR products. A hybrid assembly of 25 scaffolds (230 contigs) using GS-20, GS-FLX, and Sanger sequences was produced using the GS De Novo Assembler (version 1.1.02.15) (Roche). The physical map of the chromosome was also determined by OpGen Technologies, Inc. (Madison, WI), using the restriction enzyme NcoI and the optical-mapping technique. The order and orientations of

**References**
the scaffolds were confirmed by aligning the scaffolds on the optical map using Opgen Mapviewer. Gaps between linked contigs were closed either by walking on gap-spanning clones or with PCR products generated from genomic DNA. While the chromosome was finished to high quality, all attempts to unambiguously resolve the plasmid-related contigs were unsuccessful due to their highly repetitive nature. Therefore, two PacBio single-molecule real-time (SMRT) cells were carried out using the Pacific Biosciences RSII platform to close the plasmids and complete the genome.

PacBio sequencing was carried out by shearing two aliquots of ~4 μg genomic E. coli VR50 DNA, using g-Tube (Covaris), into fragments size targeted at 10 kb. The sheared samples were then purified and concentrated using washed Agencourt AMPure XP magnetic beads (Beckman Coulter Inc.) with size selection at 0.45-fold volume. Subsequently, SMRTbell template libraries were prepared using the commercial template preparation kit from Pacific Biosciences Inc. In brief, the DNA was end repaired, and adapters were ligated, followed by exonuclease digestion of incompletely ligated products. Using the provided P4-C2 DNA/polymerase binding kit from Pacific Biosciences, 0.83 nM libraries was then annealed with sequencing primers and bound to 50 nM P4 DNA polymerase. For enhanced loading efficiency into the sequencing zero-mode waveguides (ZMWs), 15 pM bound complexes was immobilized into Magbeads (Pacific Biosciences Inc.) according to the accompanying protocols. During sequencing, the duration of the sequence collection was set at 120 min with the stage start option. Upon acquisition of the sequencing data, short reads that were less than 50 bp were filtered off, and the minimum polymerase read quality was set at 0.75.

De novo genome assemblies of PacBio sequence reads were produced using SMRT Portal (v2.0.0) and the hierarchical genome assembly process (HGAP) (27), with default settings and a seed read cutoff length of 5,000 bp to ensure accurate assembly across E. coli rRNA operons. The hybrid 454/Sanger and PacBio assemblies were compared using MUMmer3 (28), Artemis Comparison Tool (29), and mauve (30), and discrepancies between the assemblies were investigated at the read level using BAMView (31).

Detection of DNA modifications was carried out using the RS_Modification_and_Motif_Analysis.1 tool from the SMRT analysis package version 1.2.0. In brief, PacBio reads were mapped to the complete VR50 genome. The polymerase kinetics, interpulse durations (IPDs) (32), were measured for each base, and methylated bases were identified by comparison of IPD ratios against an in silico kinetic reference model (details are available from Pacific Biosciences) and against the expected IPD signatures of the three bacterial methylation types: m⁵A, m⁴C, and m⁵C (33). Sequence motifs were identified using PacBio Motif finder v1 with a quality value (QV) cutoff of 30.

**Genome annotation and comparative analysis.** Annotation of the VR50 genome was performed by transferring the high-quality annotation of E. coli K-12 MG1655, used as a main reference, and E. coli EDL933, UTI89, CFT073, and 536 when relevant, using the Rapid Annotation Transfer Tool (RATT) (34). Complementary annotation data were provided by the Rapid Annotations Using Subsystems Technology (RAST) server (35). Manual curation was also performed to ensure the accuracy of the annotations transferred, with particular attention to regions of differences, prophage-related regions, and pseudogenes. BRIG (36), Artemis Comparison Tool (29), and mauve (30) were used to visually compare the VR50 assembly with various other genomes in order to identify regions of similarity and difference. High-quality single-nucleotide polymorphisms (SNPs) (bounded by 20 exact base pair matches on both sides) in pairwise genome comparisons were calculated using MUMmer3 (28).

An in silico multilocus sequence type (MLST) comparison of seven housekeeping genes from E. coli VR50 and representative complete E. coli genomes (Table 1) was carried out by retrieving the sequence fragments for adk, fumC, gyrB, icd, mdh, purA, and recA (http://mlst.warwick.ac.uk/mlst/). To more accurately determine the phylogenetic relatedness of E. coli VR50 to other E. coli strains, a SNP-based tree was constructed. In silico-simulated reads for five fully sequenced group A strains (HS, BW2952, DH10, W3110, and MG1655) and 12 additional E. coli strains (IAI1, EDL933, Sakai, UMNO26, IA139, SE15, S88, APEC O1, UT189, ED1a, CFT073, and S38) were mapped against the complete genome of VR50 using SHRIMP 2.0 (37). Simulated 150-bp paired reads with an insert length of 250 bp were generated at 150-fold coverage using a custom program (available on request). SNP calling and indel prediction were performed using the Nesoni package (Victorian Bioinformatics Consortium), and all conserved polymorphic positions in the data set were identified using the Nesoni n-way pairwise comparative-analysis tool. Polymorphic substitution sites were extracted and concatenated, and the resulting alignment was used for phylogenetic-tree construction. An unrooted maximum-likelihood (ML) phylogenetic tree based on the SNP alignment was constructed with RAxML 7.2.8 (38) using the GTR substitution model and 1,000 bootstrap replicates. The tree was plotted using FigTree 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

**Quantification of secreted IL-6.** The level of interleukin 6 (IL-6) secreted by HeLa epithelial cells was quantified by enzyme-linked immunosorbent assay (ELISA) as described by the manufacturer (R&D Systems). Culture supernatants from E. coli-stimulated epithelial cells were collected 2 h following infection.

**Mouse model of UTI.** The mouse model of UTI was used as described previously (39). Female C57BL/6 mice (8 to 10 weeks old) were purchased from the Animal Resources Center, Western Australia, and housed in sterile cages with ad libitum access to sterile water. An inoculum of 25 μl containing 5 × 10⁶ CFU of bacteria in PBS, was instilled directly into the bladder using a 1-ml tuberculosis syringe attached to the catheter. Urine was collected from each mouse 18 h after inoculation for quantitative colony counts. Groups of mice were euthanized by cervical dislocation 18 h after challenge; the bladders were then excised aseptically, weighed, and homogenized in PBS. The bladder homogenates were serially diluted in PBS and plated onto LB agar for colony counts. Data are expressed as the total number of CFU per gram of bladder tissue for each mouse. Experiments were performed with a minimum group size of eight.

**Nucleotide sequence accession numbers.** The complete chromosome (GenBank accession no. CP011134) and plasmid (GenBank accession no. CP011135 to CP011143) sequences have been deposited at DDBJ/EMBL/GenBank (Bioproject PRJEA61445). The version described in this paper is version 1.

**RESULTS**

**Phenotypic characteristics of E. coli VR50.** E. coli VR50 (OR: K1: H⁻) is an ABU isolate that belongs to phylogroup A. The strain was isolated from a 30-year-old otherwise healthy woman who had carried it for at least 1 year without any adverse effects. E. coli VR50 causes mannose-sensitive (MS) agglutination of yeast cells (indicative of type 1 fimbria expression) and mannose-resistant (MR) hemagglutination of human type A red blood cells (20). E. coli VR50 is also nonmotile and adheres strongly to HeLa epithelial cells, with the level of adhesion equal to that of the well-characterized pyelonephritis strain CFT073 and significantly greater than that of the ABU strain E. coli 83972 (Fig. 1A). Since E. coli VR50 is an ABU strain that was responsible for a prolonged infection in the absence of symptoms, we hypothesized that it might fail to trigger a strong proinflammatory response in comparison to other virulent UPEC strains. To test this, we compared the abilities of E. coli VR50 and CFT073 to stimulate IL-6 production following adhesion to epithelial cells. The level of secreted IL-6, which is an important proinflammatory cytokine associated with UTI, was quantified by ELISA. Despite equivalent adherence, the level of IL-6 secreted following adhesion of E. coli VR50 was significantly lower than the level of IL-6 secreted following adhesion of CFT073 (P < 0.001) (Fig. 1B). Taken together, these characteristics prompted us to perform an in-depth analysis of E. coli VR50, as described below.
The genome of VR50 had adenine methylated exclusively, with the majority (96.3%) of m^6A modifications occurring at 5'-Cm^6A-TGm^6A-3' deoxyadenosine methylase (Dam) motifs (Table 2). A total of 39,361 (98%) of 40,144 Dam sites contained m^6A. This frequency of Dam methylation in VR50 is consistent with previously characterized E. coli methyleases (40–42). In addition to Dam, VR50 encodes a single active type I restriction-modification (R-M) system in a typical type I operon (hisR-hisM-hsdS). The coding sequences in the operon share 100% amino acid sequence identity with the previously characterized type I R-M system EcoB from E. coli B; this methylase is predicted to recognize the motif 5'-TGm^6A-TGCT-3' (methylated base on opposite strand underlined) (Table 2). It should be noted that no ten-teen translocation (TET)-converted sample DNA was sequenced, and thus, m^5C methylated bases could not be discriminated from unmodified bases.

### TABLE 1 Summary of genome features of publicly available E. coli strains used in this study

<table>
<thead>
<tr>
<th>Organism</th>
<th>Pathotype* (host)</th>
<th>Size of chromosome (bp)</th>
<th>GC content (%)</th>
<th>No. of genes</th>
<th>No. of pseudogenes</th>
<th>No. of rRNAs</th>
<th>No. of tRNAs</th>
<th>No. of prophage regions</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli O157:H7</td>
<td>K-12 (human)</td>
<td>5,498,450</td>
<td>50.5</td>
<td>5,361</td>
<td>247</td>
<td>105</td>
<td>18</td>
<td>1</td>
<td>BA000007.2</td>
</tr>
<tr>
<td>E. coli O157:H7</td>
<td>TW14359 (human)</td>
<td>5,528,136</td>
<td>50.5</td>
<td>5,263</td>
<td>NA</td>
<td>22</td>
<td>106</td>
<td>NA</td>
<td>CP001368.1</td>
</tr>
<tr>
<td>E. coli O157:H7</td>
<td>EC4115 (human)</td>
<td>5,572,075</td>
<td>50.4</td>
<td>5,315</td>
<td>NA</td>
<td>22</td>
<td>110</td>
<td>2</td>
<td>CP001164.1</td>
</tr>
<tr>
<td>E. coli O127:H6</td>
<td>E2348/69 (human)</td>
<td>4,965,553</td>
<td>50.8</td>
<td>4,703</td>
<td>145</td>
<td>92</td>
<td>13</td>
<td>2</td>
<td>FM180568.1</td>
</tr>
<tr>
<td>E. coli O55:H7</td>
<td>CB9615 (EPEC)</td>
<td>5,386,352</td>
<td>50.5</td>
<td>5,028</td>
<td>69</td>
<td>22</td>
<td>101</td>
<td>19</td>
<td>CP001846.1</td>
</tr>
<tr>
<td>E. coli O139:H28</td>
<td>E24377A (ETEC)</td>
<td>4,979,619</td>
<td>50.6</td>
<td>4,873</td>
<td>118</td>
<td>22</td>
<td>97</td>
<td>13</td>
<td>FM180568.1</td>
</tr>
<tr>
<td>E. coli O178:H11</td>
<td>H10407 (ETEC)</td>
<td>5,153,435</td>
<td>51.2</td>
<td>4,746</td>
<td>94</td>
<td>22</td>
<td>87</td>
<td>9</td>
<td>FN649414</td>
</tr>
<tr>
<td>E. coli O81</td>
<td>ED1a (Commensal)</td>
<td>5,209,548</td>
<td>50.7</td>
<td>4,915</td>
<td>90</td>
<td>22</td>
<td>88</td>
<td>8</td>
<td>CP001846.1</td>
</tr>
<tr>
<td>E. coli O9:H4</td>
<td>HS (Commensal)</td>
<td>5,209,548</td>
<td>50.7</td>
<td>4,915</td>
<td>90</td>
<td>22</td>
<td>88</td>
<td>8</td>
<td>CP002166.2</td>
</tr>
<tr>
<td>E. coli O152:H28</td>
<td>SE11 (Commensal)</td>
<td>5,209,548</td>
<td>50.7</td>
<td>4,915</td>
<td>90</td>
<td>22</td>
<td>88</td>
<td>8</td>
<td>CP002166.2</td>
</tr>
<tr>
<td>E. coli O6:K15:H31</td>
<td>536 (UPEC)</td>
<td>4,938,920</td>
<td>50.5</td>
<td>4,685</td>
<td>NA</td>
<td>22</td>
<td>88</td>
<td>6</td>
<td>CP002431.1</td>
</tr>
<tr>
<td>E. coli O185:H1</td>
<td>ABU (human)</td>
<td>4,900,968</td>
<td>50.8</td>
<td>4,764</td>
<td>91</td>
<td>22</td>
<td>87</td>
<td>7</td>
<td>CP001396.1</td>
</tr>
<tr>
<td>E. coli O181:H51</td>
<td>MG1655 (human)</td>
<td>4,578,159</td>
<td>50.8</td>
<td>4,099</td>
<td>NA</td>
<td>22</td>
<td>88</td>
<td>NA</td>
<td>CP001846.1</td>
</tr>
<tr>
<td>E. coli ATCC 8739</td>
<td>W (human)</td>
<td>4,746,218</td>
<td>50.9</td>
<td>4,409</td>
<td>82</td>
<td>22</td>
<td>90</td>
<td>4</td>
<td>CP001396.1</td>
</tr>
<tr>
<td>E. coli O185:H1</td>
<td>MG1655 (human)</td>
<td>4,578,159</td>
<td>50.8</td>
<td>4,099</td>
<td>NA</td>
<td>22</td>
<td>88</td>
<td>NA</td>
<td>CP001846.1</td>
</tr>
</tbody>
</table>

*a* AIEC, adherent-invasive E. coli; APEC, avian pathogenic E. coli; EAEC, enteroaggregative E. coli; EHEC, enterohemorrhagic E. coli; EPEC, enteropathogenic E. coli; ETEC, enterotoxigenic E. coli; NMEC, neonatal meningitis E. coli; UPEC, uropathogenic E. coli.

*b* Values are given according to information retrieved from the sequence annotation combined with data from the literature, when available.

*c* NA, not available.
The *E. coli* VR50 genome is most similar to that of the K-12 strain MG1655. To further investigate the phylogenetic relationship of *E. coli* VR50 to other sequenced *E. coli* strains, an in silico MLST comparison was carried out. VR50 was found to belong to the sequence type (ST) 10 complex and was indistinguishable from the K-12 strains MG1655 and W3110. Additional housekeeping genes (i.e., *arcA*, *aroA*, *mtlD*, *pgi*, and *rpoS*) and the *fimH* adhesin gene (43) were also found to be identical between the K-12 and VR50 genomes. Whole-genome nucleotide comparisons confirmed that VR50 belongs to phylogroup A and is most closely related to K-12 strains (Fig. 2). Pairwise comparisons between complete genomes of phylogroup A *E. coli* strains indicated that the genomes of VR50 and K-12 MG1655 differ by 3,953 high-quality SNPs, with both strains differing from *E. coli* HS, the most divergent phylogroup A strain, by approximately 14,500 SNPs (Fig. 2). Although the genomes of VR50 and K-12 MG1655 are largely syntenic, VR50 exhibits two large inversions (12.4 kb and 54.2 kb) flanked by IS629 elements. The IS629 element has expanded to 19 chromosomal copies after acquisition by VR50, contributing to the larger number of pseudogenes observed in the genome (Table 1).

**E. coli** VR50 has acquired multiple mobile genetic elements. To identify genomic regions that may be associated with the ability to persist in the urinary tract, we compared the VR50 chromosome to other representative complete *E. coli* genomes using BLAST. We defined 15 regions of differences (RDs) (Table 3) and 8 prophages (Table 4) that were differentially distributed among 39 complete representative *E. coli* genomes (Fig. 3). RD-8 is the largest of these regions and carries a 94-kb *phe* tRNA-associated integrative island that we refer to as GI-VR50-*pheV*. VR50 also carries a 32-kb *asn* tRNA-associated island (RD-3) that shares 99% identity with the high-pathogenicity island from *Yersinia pestis* (44). While RD-3 is present in all UPEC strains and absent from most K-12 derivative strains and commensals, it is found in the laboratory strain BL21(DE3) and the B strain REL606.

Most of the prophage-related regions identified in the VR50 genome (VR50p1 to VR50p8) (Table 4) were found not in K-12 or other commensal strains. Of these, 3 are lambdoid prophages (VR50p2, VR50p4, and VR50p6) and have similarity to each other and to the lambdoid prophages in other *E. coli* genomes, with the exception of most K-12 derivative strains, from which they are absent. Of note, VR50p6 is inserted in the *mrlA* gene, which encodes a regulator required for curli production and extracellular matrix formation (45). VR50p3 is a P22-like phage, sharing similarities with the only P22-like phage found in UTI89, whereas prophages VR50p1 and VR50p8 are similar to the satellite phage P4 (Fig. 3). VR50p7 is related to the serotype-converting *Salmonella* phage epsilon 15 and to Eco39I-1 in UPEC strain IA39 but does not carry any of the genes involved in serotype conversion.

The plasmid pVR50A is an F-like conjugative plasmid, sharing nearly 90% of its backbone with pUTI89, with a nucleotide similarity of ~99%. A significant portion of pUTI89 is missing in pVR50A and is replaced by a multidrug resistance cluster that includes several antibiotic resistance genes, including *straB* (which confers resistance to streptomycin) and *folP* (which confers resistance to sulfonamide). Of note, a copy of the *blaTEM-1*

![FIG 1](http://iai.asm.org/) (A) Adhesion of *E. coli* VR50 and 83972 (ABU strains) and CFT073 (UPEC strain) to HeLa human epithelial cells. Epithelial cells were inoculated with bacteria and incubated for 1 h, with adherent bacteria enumerated by direct plating and colony counts. The results are the averages of two independent experiments plus standard errors of the mean. *E. coli* VR50 adhered to the cells in equal numbers to the UPEC strain CFT073. (B) ELISA demonstrating levels of IL-6 secreted by HeLa epithelial cells infected with *E. coli* VR50 and CFT073. The results are the averages of four replicates plus standard errors of the mean. *E. coli* VR50 stimulated a significantly reduced IL-6 response compared to CFT073, despite a similar level of adherence (*P* < 0.001).

![TABLE 2](http://iai.asm.org/) Active methyltransferase enzymes identified in VR50

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>Coordinates</th>
<th>Type</th>
<th>R-M/orphan</th>
<th>Target sequence</th>
<th>No. in genome</th>
<th>No. detected</th>
<th>% Methylated</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECVR50_3796</td>
<td>3831641–3832477</td>
<td>II</td>
<td>Orphan</td>
<td>G*&lt;sup&gt;ATC&lt;/sup&gt;</td>
<td>40,144</td>
<td>39,361</td>
<td>98</td>
<td>100% amino acid sequence identity to M.EcoKDam from <em>E. coli</em> K-12 MG1655</td>
</tr>
<tr>
<td>ECVR50 4827 to ECVR50 4829</td>
<td>4942889–4946401</td>
<td>II</td>
<td>R-M</td>
<td>TG*&lt;sup&gt;AN&lt;sub&gt;T&lt;/sub&gt;&lt;/sup&gt;GCT</td>
<td>775</td>
<td>755 (751)</td>
<td>97.4 (96.9)</td>
<td>100% amino acid sequence identity to M.EcoBl and S.EcoBl from <em>E. coli</em> B</td>
</tr>
</tbody>
</table>

*Underlined nucleotides reflect methylated bases on opposite DNA strand.*

*Figures for the complementary strand are in parentheses and suggest hemimethylation at 4 of 755 TG*<sup>AN<sub>T</sub></sup>GCT sites.*
gene, associated with ampicillin resistance, was also identified, but a truncation of its N terminus suggests its inactivation, as confirmed by antibiotic sensitivity phenotypic tests (data not shown). The other large plasmid found in VR50, pVR50B, possesses a complex mosaic structure comprising multiple insertion sequence (IS) elements, with regions of similarity to numerous UPEC-, neonatal meningitis *E. coli* (NMEC)-, and commensal-associated plasmids, including pCE10A (strain CE10; NMEC), pECOED (strain ED1a; commensal), pSE11-2 (SE11; commensal), and pUTI89 (UTI19; UPEC). Plasmid pVR50B carries a *tra* region that is almost identical to the corresponding region in pVR50A; the sequences of these two plasmids could only be properly resolved using PacBio sequencing. While pVR50B does not contain any antibiotic resistance genes, several putative virulence-associated cargo genes were identified, including *senB* (which encodes a putative enterotoxin), *imm* (which encodes a colicin Ib immunity protein), and the toxin-antitoxin *phd-doc* genes. Of note, VR50 also harbors several small plasmids that are very similar to cryptic plasmids often associated with other commensal strains; for example, pVR50C and pVR50D are similar to pColE1-like plasmids, and pVR50F is similar to pSE11-6.

Taken together, our data suggest *E. coli* VR50 is a commensal strain, remarkably similar to *E. coli* K-12, which has gained the ability to persist in the urinary tract through gene acquisition. This is in stark contrast to the prototypical ABU *E. coli* strain 83972, which appears to have evolved through mutation and gene loss from a virulent UPEC strain (9, 10, 15–17, 46).

The *pheV* genomic island in *E. coli* VR50 includes uropathogenicity genes. The most notable chromosomal variation between *E. coli* VR50 and MG1655 is a contiguous gene cluster in VR50 comprising the 94-kb genomic island GI-VR50-*pheV* (RD-9), a *kps* capsular biosynthesis gene cluster, and a *gsp* type II secretion system gene cluster (RD-9). GI-VR50-*pheV* is located between the *pheV* tRNA gene and the *ksp* gene cluster (Fig. 4). It is delimited by direct repeats (DR) that encompass the last 19 bp of the *pheV* tRNA gene and a truncated tRNA gene (*phe*V = V), equivalent to *attL* and *attR*, left and right junction sites generated by phage insertion (47). Consistent with the features of many other integrative islands, a P4-like integrase gene is located immediately adjacent to the 5′ end of GI-VR50-*pheV*. The GC content of GI-VR50-*pheV* is 48.11% compared with the *E. coli* VR50 genomic average of 50.78%. GI-VR50-*pheV* contains several genes associated with UPEC virulence, including (i) remnants of the *pap* cluster; (ii) the *iuc* aerobactin biosynthesis locus and the *iha* siderophore receptor; (iii) two autotransporter (AT)-encoding genes, *flu* (encoding antigen 43 [Ag43]) and *sat* (encoding a secreted AT toxin); and (iv) the *afaABCDE* genes, the Afa/Dr adhesin family chaperone-usher fimbrial genes.

The gene content of GI-VR50-*pheV* is most similar to those of *pheV*-associated genomic islands in the probiotic strain Nissle 1917, as well as UPEC strains CFT073, UMN026, 536, and IA139 (Table 5 and Fig. 5). Like other *pheV* islands, GI-VR50-*pheV* has a mosaic structure of defined sequence modules separated by a variety of insertion sequences (Fig. 5). Modules encoding virulence factors are highly conserved at the nucleotide level (i.e., typically >95% nucleotide identity), although not necessarily in the same order or orientation or at the same level of completeness. Intervening regions typically contain strain-specific sequences or insertion sequence elements that presumably facilitate the many rearrangements observed between islands. For example, the 28,718-nt region containing the *sat, iuc, and iha* modules in VR50 is 99% identical to equivalent modules in the Nissle 1917, CFT073, and
TABLE 3 Major regions of differences identified in VR50 compared to other E. coli strains

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Start nucleotide</th>
<th>End nucleotide</th>
<th>Size (bp)</th>
<th>Insertion site</th>
<th>Virulence-related or other notable gene(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD-1</td>
<td>244012</td>
<td>277704</td>
<td>33,692</td>
<td>aspV RNA</td>
<td>yafT</td>
<td>Lipoprotein</td>
</tr>
<tr>
<td>RD-2</td>
<td>1527896</td>
<td>1547892</td>
<td>19,996</td>
<td>ydbL, yebG</td>
<td>fesRb, paaABCDCEFHIJKXY</td>
<td>Phenylacetic acid degradation</td>
</tr>
<tr>
<td>RD-3</td>
<td>2125253</td>
<td>2160033</td>
<td>34,780</td>
<td>asntB RNA</td>
<td>Yersiniabactin-biosynthetic cluster</td>
<td>Yersinia pestis</td>
</tr>
<tr>
<td>RD-4</td>
<td>2127863</td>
<td>2192851</td>
<td>19,988</td>
<td>yehF</td>
<td>flu1</td>
<td>Antigen 43 phase-variable biofilm formation autotransporter</td>
</tr>
<tr>
<td>RD-5</td>
<td>2216480</td>
<td>2225524</td>
<td>9,044</td>
<td>gnd, gfdL</td>
<td>O-antigen region</td>
<td></td>
</tr>
<tr>
<td>RD-6</td>
<td>3037592</td>
<td>3047232</td>
<td>9,640</td>
<td>yap, cysH</td>
<td>CRISPR locus</td>
<td></td>
</tr>
<tr>
<td>RD-7</td>
<td>3157781</td>
<td>3169951</td>
<td>12,170</td>
<td>glyU</td>
<td>ETT2 cluster remnant</td>
<td></td>
</tr>
<tr>
<td>RD-8</td>
<td>3295104</td>
<td>3407870</td>
<td>112,766</td>
<td>pheV RNA</td>
<td>pap cluster remnant; iuc, iha, flu2, sat, and afaABCDE kps cluster</td>
<td>GI-VR50 pheV capsular biosynthesis gene cluster</td>
</tr>
<tr>
<td>RD-9</td>
<td>3407872</td>
<td>3442920</td>
<td>35,048</td>
<td>kpsM, ygbH</td>
<td>gsp1, gcaB, gfeDC</td>
<td>T2SS*-related genes (general secretory pathway); glycolate metabolism genes</td>
</tr>
<tr>
<td>RD-10</td>
<td>3770072</td>
<td>3782784</td>
<td>12,712</td>
<td>rpsL, bfr</td>
<td>gsp2 cryptic cluster</td>
<td>T2SS*-related genes (general secretory pathway; cryptic)</td>
</tr>
<tr>
<td>RD-11</td>
<td>3952087</td>
<td>3965106</td>
<td>13,019</td>
<td>yhiM, yhiN</td>
<td>Tn7-related genes</td>
<td>Sugar catabolism</td>
</tr>
<tr>
<td>RD-12</td>
<td>4403223</td>
<td>4419794</td>
<td>16,571</td>
<td>typA, altd</td>
<td>yhiLMN, ompL, and yhiYOPQRSTUVWXY</td>
<td></td>
</tr>
<tr>
<td>RD-13</td>
<td>4704726</td>
<td>4718639</td>
<td>13,913</td>
<td>pheU RNA</td>
<td>fecURABCDE regulon</td>
<td>Iron dicitrate transport</td>
</tr>
<tr>
<td>RD-14</td>
<td>4859525</td>
<td>4874849</td>
<td>15,324</td>
<td>leuX RNA</td>
<td>Putative oxidoreductase putative transporters</td>
<td></td>
</tr>
<tr>
<td>RD-15</td>
<td>4912259</td>
<td>4947503</td>
<td>35,244</td>
<td>nanS, yjiA</td>
<td>hsdRMS, mrv, and mcrBC</td>
<td>Restriction-modification enzymes; pentose and pentitol sugar breakdown</td>
</tr>
</tbody>
</table>

UMN026 pheV islands; however, sat and iuc are inverted in Nissle 1917 and iha is inverted in VR50 (Fig. 5) (48). Although in Nissle 1917 this region is also flanked by IS2-like sequences, the insertion sequences found between these modules are different in all four strains, reflecting the dynamic evolution of genomic islands (Fig. 5).

The archetypal pheV island is from CFT073 (PAI-1CFT073). PAI-1CFT073 contains the sat, iha, iut, and flu gene modules with a high degree of identity to islands from Nissle 1917 and VR50; in addition, it includes the pap operon and a hylCABD operon. The mosaic structure of the pheV island is further illustrated by the fact that, although it is not present in the well-characterized UPEC strain UTI89, several modules similar to those on GI-VR50-pheV are found on other UTI89 islands (e.g., leuX). Although the K-12 and HS commensal strains do not carry a pheV island, it is found in the sequenced ED1a commensal strain (49). GI-ED1a-pheV contains fragments of modules typically associated with UPEC (i.e., iha and sat), with full-length iha and iuc loci on the nearby pheU island. The genomes of the Shiga toxin-producing E. coli (STEC) strains Sakai, EDL933, and...

TABLE 4 Prophage-related regions identified in VR50

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Start nucleotide</th>
<th>End nucleotide</th>
<th>Size (bp)</th>
<th>Insertion site</th>
<th>Direct repeat (bp)</th>
<th>Type</th>
<th>Virulence-related or other notable gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VR50p1</td>
<td>300479</td>
<td>312346</td>
<td>11,868</td>
<td>thrW tRNA</td>
<td>16</td>
<td>P4-like</td>
<td>Outer-membrane porin proteins; Lom-like outer membrane proteins</td>
</tr>
<tr>
<td>VR50p2</td>
<td>800678</td>
<td>832594</td>
<td>31,917</td>
<td>ybfB</td>
<td>ND</td>
<td>P4-like</td>
<td>Lambda, remnant</td>
</tr>
<tr>
<td>VR50p3</td>
<td>1082316</td>
<td>1125788</td>
<td>43,473</td>
<td>torS, torT</td>
<td>ND</td>
<td>P22-like</td>
<td>Lambda</td>
</tr>
<tr>
<td>VR50p4</td>
<td>1369562</td>
<td>1402434</td>
<td>50,873</td>
<td>ompW</td>
<td>39</td>
<td>P22-like</td>
<td>Lambda</td>
</tr>
<tr>
<td>VR50p5</td>
<td>1678552</td>
<td>1691947</td>
<td>13,396</td>
<td>yntL, hipA</td>
<td>ND</td>
<td>Unitypeable; remnant</td>
<td>Lambda</td>
</tr>
<tr>
<td>VR50p6</td>
<td>2324694</td>
<td>2371668</td>
<td>46,705</td>
<td>mrlA</td>
<td>18</td>
<td>P4-like</td>
<td>Lambda</td>
</tr>
</tbody>
</table>

**a** Prophage genomes appear to have all the genes necessary to produce fully functional phages.

**b** Pseudogene.

**ND** Not determined.
EC4115 carry a large deletion encompassing pheV and the adjacent kps capsular and gsp type II secretion loci. However, the pheV tRNA gene is included in a nonsyntenic locus and is known to be a hot spot for insertion of the locus for enterocyte effacement (LEE) (50, 51). The afa-8 locus on a pheV genomic island has been described previously (52), although the associated int gene is like intP4-2, not intP4-1, and the VR50 afa locus is quite divergent from afa-8 (see below).

GI-VR50-pheV is formed from two independent integration events. GI-VR50-pheV is a composite structure with evidence of two independent pheV island insertions mediated by different IntP4-like integrase genes (Fig. 5). As observed in the well-characterized GI-536-pheV pathogenicity island (PAI V536), GI-VR50-pheV is bound by direct repeats corresponding to the 3' end of the pheV tRNA gene (Table 6). Like pheV islands in CFT073, Nissle 1917, and UMN026, the right-hand direct repeat (pheV) is 19 nt
and contains a substitution (Table 6). Strains 536 and APECO1 share a different pheV sequence, corresponding to a duplication of the 23 nt at the 3' end of pheV and a single-nucleotide deletion (Table 6) (4, 53), whereas, GI-SMS-3-5-pheV is bounded by a set of perfect 19-nt direct repeats. Intriguingly, within GI-VR50-pheV there is a perfect 19-nt direct repeat of the pheV 3' region (pheV-2) immediately upstream of a different P4-like integrase gene (intP4-2) (Fig. 5 and Table 6). IntP4-1 and IntP4-2 share only 43% amino acid identity, making it certain that they are not the result of an internal duplication. A comparison of integrase sequences associated with pheV islands indicated that VR50 intP4-1 shares 94 to 95% nucleotide identity with the pheV-associated integrase genes of SMS-3-5 and E24377A. VR50 intP4-2 shares at least 94% nucleotide identity with the integrase genes of most other pheV islands from E. coli genomes (e.g., Int PAI V536) and 98 to 99% identity with int genes from Nissle 1917, CFT073, and UMN026 pheV islands.

The PAI-536-pheV (PAI V536) island has been extensively

### TABLE 5

<table>
<thead>
<tr>
<th>Genomic island</th>
<th>Alternate name</th>
<th>Size (nt)</th>
<th>Genome position (nt)</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI-VR50-pheV</td>
<td>NA</td>
<td>93,983</td>
<td>3295095–3389077</td>
<td>This study</td>
</tr>
<tr>
<td>GI-Nissle1917-pheV</td>
<td>GEI-II1ISSLE1917</td>
<td>77,685</td>
<td>NA</td>
<td>(AJ586888)</td>
</tr>
<tr>
<td>GI-CFT073-pheV</td>
<td>PAI-I_CFT073</td>
<td>104,676</td>
<td>3406225–3510900</td>
<td>AE014075 (AF081285)</td>
</tr>
<tr>
<td>GI-UMN026-pheV</td>
<td>NA</td>
<td>82,784</td>
<td>3445924–3528707</td>
<td>CU928163</td>
</tr>
<tr>
<td>GI-536-pheV</td>
<td>PAI-V_536</td>
<td>48,805</td>
<td>3128084–3176888</td>
<td>CP000247 (AJ617685)</td>
</tr>
<tr>
<td>GI-APECO1-pheV</td>
<td>NA</td>
<td>55,888</td>
<td>3305017–3360904</td>
<td>CP000468 (DQ095216)</td>
</tr>
<tr>
<td>GI-SMS-3-5-pheV</td>
<td>NA</td>
<td>91,940</td>
<td>3197159–3289098</td>
<td>CP000970</td>
</tr>
<tr>
<td>GI-S88-pheV</td>
<td>NA</td>
<td>55,896</td>
<td>3219801–3275696</td>
<td>CU928161</td>
</tr>
<tr>
<td>GI-ED1a-pheV</td>
<td>NA</td>
<td>135,753</td>
<td>3374807–3510560</td>
<td>CU928162</td>
</tr>
<tr>
<td>GI-S89-pheV</td>
<td>NA</td>
<td>110,593</td>
<td>3339841–3450433</td>
<td>CU928145</td>
</tr>
<tr>
<td>GI-IAI39-pheV</td>
<td>NA</td>
<td>38,534</td>
<td>3545466–3583999</td>
<td>CU928164</td>
</tr>
<tr>
<td>GI-E24377A-pheV</td>
<td>NA</td>
<td>101,766</td>
<td>3325510–3427276</td>
<td>CP0008000</td>
</tr>
</tbody>
</table>

a The APECO1 and S88 pheV islands are virtually identical (99% at the nucleotide level); the transposition of a 2,581 nt insertion sequence is the only substantial difference (data not shown).

b The size was calculated from the first nucleotide of pheV to the last nucleotide of the pheV direct repeat. No repeat was identified in GI-ED1a-pheV or GI-IAI39-pheV, and the right-hand junction was taken as the breakpoint adjacent to the kpsF locus. E24377A contains a composite arrangement of two islands, similar to VR50 (data not shown).

c The GenBank accession number for the complete genome sequence, where available. Accession numbers in parentheses refer to GenBank accession numbers for genomic islands, if published separately prior to the genome sequence. AF081285 covers only 13,710 bp of the CFT073 genomic island.

d NA, not applicable.
studied and found to excise at lower frequencies than other *E. coli* 536 pathogenicity islands, possibly due to a single-nucleotide deletion in the right-hand DR (54). Interestingly, the left-hand DR of *pheV* islands from VR50, Nissle 1917, CFT073, UMNO26, 536, APECO1, and SMS-3-5. Conserved modules are colored the same and named for the best-known locus or gene. Strain-specific modules are colored dark gray. Insertion sequences (often comprising more than one CDS) are represented as unfilled arrowheads. The P4-like integrase sequences are represented as black arrowheads. The *pheV* and *pheV* direct repeats denote the boundaries of each island. The *kpsF* and *kpsM* modules, which are not considered part of the island, are colored light gray; variation in the gene content of the *kpsM* module is not shown. Inversions of individual modules relative to the VR50 genome are shown with an asterisk adjacent to the module name. The diagram is drawn to scale.

GI-VR50-*pheV* contributes to colonization of the mouse urinary tract. To assess the contribution of GI-VR50-*pheV* to the ability of *E. coli* VR50 to colonize the mouse urinary tract, we deleted the island using λ-Red-mediated homologous recombination. This resulted in the replacement of the entire 94-kb island with a kanamycin resistance gene cassette (referred to as VR50-*pheV-GI*). *E. coli* VR50 colonized the mouse bladder in high numbers, demonstrating its suitability in this infection model. However, we observed significantly reduced colonization of the bladder by VR50-*pheV-GI* compared to the wild-type VR50 strain (*P* = 0.002) (Fig. 6). Thus, GI-VR50-*pheV* contributes to the ability of *E. coli* VR50 to colonize the mouse bladder. Based on our bioinformatic analysis of GI-VR50-*pheV*, we predicted that this
phenotype was directly associated with production of the Afa adhesin and tested this hypothesis (see below).

**Phylogenetic analysis of the AfaE adhesin.** We performed a phylogenetic analysis of the *E. coli* VR50 adhesin-encoding *afaE* gene. The closest homologues of AfaE were the group 254 Dr family adhesins, which were sequenced as part of a larger study investigating the diversity of Dr family adhesin-encoding genes from 100 *E. coli* isolates (55). The authors divided the Dr family into 8 distinct groups (afaE-1, afaE-2, afaE-5, afaE-111, daae, dbre, draE, and 254). A multiple-sequence alignment of VR50 afaE with the group 254 adhesin gene sequences showed only between five and nine nucleotide substitutions relative to other group members (>98% nucleotide sequence identity). Four of these substitutions are unique to VR50 and are found in the 3’ region; two VR50 substitutions cause nonsynonymous replacements of the terminal residues (GGYWAK to GGYWTN). Interestingly, the same substitutions are found in another subgroup (DaaE). Although no binding function has been described for the C-terminal residues of the Dr adhesin, the observation that positive selection of point mutations drives changes for adhe-

**TABLE 6 Alignment of pheV and phe’V DR at genomic-island boundaries**

<table>
<thead>
<tr>
<th>pheV island</th>
<th>Direct-repeat sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>VR50 pheV</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>54–76</td>
</tr>
<tr>
<td>VR50 phe’V</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>19/19-nt DR</td>
</tr>
<tr>
<td>VR50 phe’V2</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>19/19-nt DR</td>
</tr>
<tr>
<td>Nissle pheV</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>54–76</td>
</tr>
<tr>
<td>Nissle phe’V</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>19/19-nt DR</td>
</tr>
<tr>
<td>CFT073 pheV</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>54–76</td>
</tr>
<tr>
<td>CFT073 phe’V</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>19/19-nt DR</td>
</tr>
<tr>
<td>UMN026 pheV</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>54–76</td>
</tr>
<tr>
<td>UMN026 phe’V</td>
<td>TTCGATTCCGAAGCCGGACACCA</td>
<td>19/19-nt DR</td>
</tr>
</tbody>
</table>

<sup>a</sup>The columns in boldface contain differences (underlined) in phe’V (the direct repeat at the 3’ end of island) relative to the pheV tRNA gene. The nucleotides in italics are not considered part of the phe’V repeat. The complete 76-nt pheV tRNA gene is 100% identical in all the *E. coli* strains shown here.

**FIG 7** (A) VR50 expresses functional Afa. Shown is hemagglutination of human RBCs by *E. coli* VR50 (i) in contrast to *E. coli* VR50afa (ii), which was unable to mediate hemagglutination of human RBCs under the same conditions. The images were viewed under ×63 magnification. (B) Adhesion of *E. coli* VR50, VR50pheV-GI, VR50afa, VR50afaE, VR50fim, and VR50fimH to T24 human bladder epithelial cells. The epithelial cells were inoculated with bacteria and incubated for 1 h, with adherent bacteria enumerated by direct plating and colony counts. The results are the averages of two independent experiments plus standard errors of the mean. Mutant strains VR50pheV-GI, VR50afa, and VR50afaE adhered significantly less to T24 cells than wild-type VR50 and the two type 1 fimbria-deficient mutants, VR50fim and VR50fimH (*P < 0.001; analysis of variance [ANOVA]).

**FIG 6** Persistence of *E. coli* VR50, VR50pheV-GI, VR50afa, and VR50afaE in the bladders of C57BL/6 mice 18 h after intraurethral challenge. The results represent log<sub>10</sub> CFU/g bladder tissue of individual mice, and the horizontal bars mark group medians. A minimum of 8 mice were assessed per group. *E. coli* VR50 was recovered from the bladders of infected mice in significantly higher numbers than each of the mutant strains (*P = 0.002; Kruskal-Wallis test).
ingly, mutants of VR50 with the entire type 1 fimbrial gene cluster (VR50fim) or the fimH gene (VR50fimH) deleted did not show any significant difference in adherence to T24 bladder epithelial cells compared to the parent VR50 strain. Taken together, the data support a role for the AfaE adhesin of \textit{E. coli} VR50 in adhesion to human bladder epithelial cells.

\textbf{Afa contributes to colonization of the mouse bladder by \textit{E. coli} VR50.} In light of the above in vitro adherence data and the in vivo mouse colonization phenotype of VR50\textit{phev-GI}, we tested \textit{E. coli} VR50afa and VR50afaE for the ability to colonize and survive in the mouse urinary tract following transurethral infection (Fig. \textit{6}). \textit{E. coli} VR50afa and VR50afaE were both significantly attenuated for colonization of the mouse bladder. The degree of attenuation was identical to that observed for the AfaE adhesin of VR50 in adhesion to human bladder epithelial cells.

\textbf{DISCUSSION} \textit{ABU} \textit{E. coli} strains, which colonize the human urinary tract without causing symptoms, comprise strains that have become attenuated through gene loss or deletion (e.g., \textit{E. coli} 83972), as well as commensal-like strains that have acquired fitness factors that facilitate colonization of the human bladder. Here, we demonstrate that \textit{E. coli} VR50 is an example of a commensal \textit{E. coli} strain that has gained important genes to enable it to survive in the urinary bladder.

The complete genome sequence of VR50 revealed multiple accessory genes that may contribute to its survival in the human urinary tract. In general, these genes were associated with a range of mobile genetic elements, including plasmids, prophages, and genomic islands. We identified 8 prophage regions in the VR50 chromosome, 4 of which are likely to be functional, with another 2 encoding satellite phage. Several putative cargo gene products are encoded within the prophage regions, most of which have unknown functions, but some have similarity to known proteins, e.g., DinI-like damage-inducible proteins (VR50p6), outer-membrane porin proteins (VR50p2 and VR50p6), Lom-like outer-membrane proteins (VR50p2, VR50p4, and VR50p6), and a DNA adenine methylase (VR50p7). \textit{E. coli} VR50 also carries two large plasmids, one of which contains several antibiotic resistance genes. Prophages, genomic islands, and plasmids are themselves comprised of smaller mobile genetic elements, some of which appear to have undergone expansion in VR50 since its divergence from other \textit{A. E. coli} strains. For example, the IS629 element is not found in \textit{E. coli} K-12 MG1655, but there are 19 copies distributed throughout the \textit{E. coli} VR50 chromosome, including disruptive insertions within genes encoding putative outer-membrane proteins and chaperone-usher fimbriae (see Data Set S1 in the supplemental material).

We focused a major part of our analysis on the GI-VR50-\textit{pheV} island, which contains multiple virulence-associated genes, including those encoding Afa, Ag43, Sat, and aerobactin. Lloyd et al. recommended that GIs are differentiated from PAIs by the fact that the latter contain established or putative virulence genes whereas the former contain genes with unknown functions (58). The original criteria for classifying pathogenicity islands suggested that the term should be reserved for islands preferentially associated with pathogenic strains (59). It is clear from a simple comparison of \textit{pheV}-associated islands in commensal, pathogenic, and asymptomatic strains that there are no apparent qualitative differences between GIs and PAIs, so use of the more neutral GI terminology may be more appropriate to describe integrated islands in the VR50 genome. Direct repeats can often be used to distinguish between integrative islands that appear to play such a dynamic role in the adaptation to new environments (e.g., PAI-536-\textit{pheV} and PAI-CFT073-\textit{pheV}) and lineage-specific indels. Delineating these boundaries is not always trivial, as the GI-\textit{pheV} case demonstrates. For example, the \textit{kps} capsule biosynthesis locus is included within the island boundaries in the original report of the Nissle 1917 \textit{pheV} island (48). Similarly, a comparison of several \textit{pheV} islands does not distinguish between the integrative island at the \textit{pheV} gene and downstream variability in the \textit{kps} and \textit{gsp} loci, even though much of the \textit{gsp} secretion locus is present in a syntenic location in the \textit{Escherichia fergusonii} genome (49). Recent studies of the \textit{E. coli} ST131 lineage by our group have begun to reveal the extent of recombination associated with GI-\textit{pheV} and other mobile genetic elements (60).

The GI-VR50-\textit{pheV} island is very similar to \textit{pheV}-associated islands from a wide variety of sequenced strains, e.g., the probiotic \textit{E. coli} Nissle 1917, extraintestinal pathogenic \textit{E. coli} (ExPEC)/UPEC (CFT073, UMNO26, IA39, S88, 536, and APEC01), enterotoxigenic \textit{E. coli} (ETEC) (E24377A), enteroaggregative \textit{E. coli} (EAEC) (55989), and commensal (ED1a) and environmental (SMS-3-5) strains. Remarkably, the LEE has been associated with \textit{pheV} islands in rabbit enteropathogenic (EPEC) (51) and STEC (61). It appears that these integrative islands are modular, and this mosaicism could be the key factor in their success, allowing rapid insertion and deletion of new elements that provide a selective advantage in different environments. In the case of \textit{E. coli} VR50, our molecular analyses suggest that a selective advantage is provided by the acquisition of the Afa/Dr adhesin locus. Another interesting consequence of this modularity is that there is extensive exchange between islands via homologous recombination. For example, although \textit{E. coli} UTI189 lacks the \textit{pheV} island, many of the core modules found on GI-VR50-\textit{pheV} (e.g., sat, iut, and \textit{flu}) are found on other UTI89 islands. As the number of complete genomes increases, it should be possible to obtain a clearer picture of the role that integrative islands play in pathogen evolution.

It is an open question as to what role the other \textit{pheV} island-encoded “pathogenicity factors,” such as the \textit{sat} gene product and aerobactin, play in the life styles of VR50 and other intestinal commensal organisms. Given the ease with which elements or the entire island are lost from these loci in some strains (54), it appears that these elements are under strong selective pressure to be retained. In the case of VR50, the presence of a complete aerobactin gene locus and the gene encoding the Iha siderophore receptor within GI-VR50-\textit{pheV} suggests these genes might contribute to its overall fitness in the bladder. Iron is a limiting nutrient in the human urinary tract, and UTI \textit{E. coli} strains often produce several siderophores to acquire ferric iron (Fe$^{3+}$) from the host. Four siderophore systems have been described from UTI \textit{E. coli}, i.e., enterobactin (which is common to all strains), salmochelin, yersiniabactin, and aerobactin (62, 63). Indeed, UPEC mutants with genes encoding the production of siderophores deleted are less virulent in the mouse urinary tract (64). An enhanced ability of \textit{ABU} \textit{E. coli} strains to acquire iron has also been proposed as a mechanism by which they outcompete UPEC during growth in human urine (39, 63, 65). We did not observe any significant
difference in the overall level of siderophores produced by *E. coli* strains MG1655 and VR50 as measured by the chrome azurol S (CAS) assay (reference 66 and data not shown); however, this is most likely due to the common production of enterobactin by both strains.

GI-VR50-pheV was shown to contribute to the ability of *E. coli* VR50 to colonize the mouse urinary tract, and we further defined Afa as a GI-encoded factor that mediates adherence of VR50 to T24 bladder epithelial cells. The Afa (and Dr) adhesins comprise a family of surface-located factors encoded by the *afa* (67–71), *dra* (72, 73), and *daa* (74, 75) operons. These operons possess similar genetic structures and have been associated with *E. coli* strains of diarrheagenic and UTI origin (76, 77). Many Afa adhesins recognize as a receptor the decay-accelerating factor (DAF), a complement-regulatory protein present on the surfaces of a range of human epithelial cells (including epithelial cells of the urinary tract) (78, 79). Some Afa variants also mediate binding to type IV collagen and carboxyembryonic antigen-related cell adhesion molecules (80, 81). We demonstrated a role for Afa (but not type 1 fimbiae) in the adhesion of *E. coli* VR50 to T24 bladder epithelial cells. Afa has been shown to mediate adherence and invasion of *E. coli* in HeLa and Chinese hamster ovary cells (82, 83). In the case of VR50, we previously reported that the strain does not readily invade T24 cells, with low numbers observed intracellularly (84). However, it is possible that the *E. coli* VR50 cells recovered from HeLa cells in our assay may represent both adhered and intracellular bacteria. In the mouse UTI model, *E. coli* VR50 colonized the bladder with high efficiency, and the overall cell numbers were very similar to the level of colonization we routinely observe for CFT073 (56, 57, 85). In contrast, VR50pheV-GI, VR50afa, and VR50afaE all displayed significantly attenuated colonization of the mouse bladder, strongly suggesting that the *afa* genes within GI-VR50-pheV enhance the fitness of VR50 in the urinary tract.

The innate immune response is essential for defense against UTI, and the associated tissue inflammation is a major cause of symptoms, leading to tissue damage and contributing to the severity of disease. The innate response is affected by the virulence of the infecting strain, and many well-characterized virulence factors act by triggering inflammation. UPEC triggers the secretion of epithelial cytokines and chemokines that contribute to inflammation and pathology. For example, IL-6 levels in the urine correlate with disease severity, and IL-8 results in recruitment of neutrophils to the site of infection (86–88). Despite equivalent adherence to epithelial cells, *E. coli* CFT073 triggered a significantly stronger IL-6 cytokine response than *E. coli* VR50. This property may be associated with the ability of *E. coli* VR50 to colonize the bladder in high numbers without provoking a host inflammatory response and is consistent with the low IL-6 response observed in individuals deliberately colonized with *E. coli* 83972 (89). It remains to be determined if VR50 is able to suppress RNA polymerase II-dependent host gene expression, as has been reported recently for *E. coli* 83972 (19). Furthermore, host genetic variation that leads to suppression of the innate immune response to UTI (e.g., via Toll-like receptor 4 [TLR4] promoter polymorphisms) also contributes to protection against symptomatic disease in some ABU patients (90–93).

ABU occurs in up to 6% of healthy individuals, 18% of diabetics (mostly women), and 20% of elderly individuals (women more often than men). *E. coli* is also the most frequent cause of ABU. An understanding of the molecular mechanisms that underpin ABU would have a major impact on current approaches to treating and preventing symptomatic UTI for several reasons. First, ABU strains provide a unique model to identify factors that enable UPEC to colonize the urinary tract. Second, the outcomes of ABU vary; in some patients, ABU can predispose to kidney infection, while in other patients, ABU can prevent colonization by more virulent strains. A greater understanding of the risk factors associated with ABU is therefore required to guide treatment decisions. Third, the unnecessary treatment of ABU in some patients may be associated with the rapid spread of genes coding for antibiotic resistance, particularly in health care settings. In this respect, it is interesting that *E. coli* VR50 contains a truncated copy of the *blaTEM-1* gene. Although not functional, a significant association between ampicillin resistance and the presence of the *afa* and *dra* genes has been reported previously (94). Finally, *E. coli* represents a viable prophylactic approach for the prevention and treatment of chronic and recurrent UTIs. The use of *E. coli* as a prophylactic agent could address important concerns about antibiotic resistance in patients with UTI recalcitrant to treatment with currently available antibiotics. In the case of VR50, its capacity to express Afa, which contributes to chronic pyelonephritis (77, 95) and recurrent cystitis (96), suggests that it would not constitute a suitable prophylactic ABU strain for human use.

In summary, our study has determined the complete genome sequence of *E. coli* VR50 and demonstrated that it is a commensal-like strain that has acquired a number of UPEC-associated virulence factors. One of these factors, Afa, contributes specifically to VR50 colonization of the mouse bladder.

ACKNOWLEDGMENTS

This work was supported by grants from the National Health and Medical Research Council (NHMRC) of Australia (511224, 569676, 631654, and APP1033799). M.A.S. is supported by an Australian Research Council (ARC) Future Fellowship (FT100100662), S.A.B. is supported by an NHMRC Career Development Fellowship (APP1090456), and M.T. is supported by an ARC Discovery Early Career Researcher Award (DE130101169). PacBio sequencing was carried out with the support of UM-MOHE HIR grant no. H-50001-A000027.

We thank John Davis and Annette McGrath from the Australian Genome Research Facility for assistance in library preparation and 454/Sanger hybrid assembly. We are indebted to the contribution of Per Klemm, who sadly passed away during the course of this study.

REFERENCES


4289.
76. Servin AL. 2005. Pathogenesis of Afa/Dr diffusely adhering Escherichia
77. Servin AL. 2014. Pathogenesis of human diffusely adhering Escherichia
coli expressing Afa/Dr adhesins (Afa/Dr DAEC): current insights and fu-
1128/CMR.00036-14.
consensus repeat-3 domain of recombinant decay-accelerating factor is
recognized by Escherichia coli recombinant Dr adhesin in a model of a
1084/jem.176.6.2115.
Identification of the complement decay-accelerating factor (DAF) on ep-
http://dx.doi.org/10.1084/jem.165.3.848.
80. Korotkova N, Cota E, Lebedin Y, Monpouet S, Guignot J, Servin AL,
Matthews S, Moseley SL. 2006. A subfamily of Dr adhesins of Escherichia
coli bind independently to decay-accelerating factor and the N-domain of
org/10.1074/jbc.M606568200.
82. Rana T, Hasan RJ, Nowicki S, Venkatarajan MS, Singh R, Urvil PT,
Popov V, Braun WA, Popik W, Goodwin JS, Nowicki BJ. 2014. Com-
plement protective epitopes and CD55-microtubule complexes facilitate
the invasion and intracellular persistence of uropathogenic Escherichia
83. Goluszko P, Popov V, Selvarangan R, Nowicki S, Pham T, Nowicki BJ.
1997. Dr fimbiae operon of uropathogenic Escherichia coli mediates mi-
crotubule-dependent invasion to the HeLa epithelial cell line. J Infect Dis
84. Bokil NJ, Totsika M, Carey AJ, Stacey KJ, Hancock V, Saunders BM,
survival of uropathogenic Escherichia coli: differences between diverse
clinical isolates and between mouse and human macrophages. Immuno-
Trade-off between iron uptake and protection against oxidative stress:
deletion of cuoA promotes uropathogenic Escherichia coli virulence in a
http://dx.doi.org/10.1128/JB.00451-08.
86. Wullt B, Bergsten G, Connell H, Rollano P, Gebratsedik N, Hang L,
Svanborg C. 2001. P-fimbriae trigger mucosal responses to Escherichia coli
10.1046/j.1462-5822.2001.00111.x.
2001. Toll-like receptor signaling and chemokine receptor expression in-
fluence the severity of urinary tract infection. J Infect Dis 183(Suppl 1):
Do type 1 fimbiae promote inflammation in the human urinary tract?
00912.x.
2011. Genetic control of the variable innate immune response to asympto-
1371/journal.pone.0028289.
90. Ragnarsson B, Fischer H, Godaly G, Gronberg-Hernandez J, Gustaf-
sson M, Karpman D, Lundstedt AC, Latay N, Ramisch S, Svensson ML,
Wullt B, Yadvad M, Svanborg C. 2008. TLR- and CXCR1-dependent
innate immunity: insights into the genetics of urinary tract infections.
2008.02004.x.
N, Tammi M, Gustafsson M, Lundstedt AC, Latay N, Ramisch S, Svensson ML,
93. Ragnarsson B, Samuelsson M, Gustafsson MC, Leijonhufvud I, Karp-
man D, Svanborg C. 2007. Reduced Toll-like receptor 4 expression in
dx.doi.org/10.1086/518893.
94. Hart A, Nowicki BJ, Reisner B, Pawelczyk E, Goluszko P, Urivl P,
Anderson G, Nowicki S. 2001. Ampicillin-resistant Escherichia coli in
gestational pyleonephritis: increased occurrence and association with the
org/10.1086/320196.
chronic pyelonephritis with Escherichia coli O75:K5:H-bearing Dr fim-
briae: mutation in the dra region prevented tubulointerstitial nephritis. J
Marrs CF. 1995. Virulence characteristics of Escherichia coli causing first
urinary tract infection predict risk of second infection. J Infect Dis 172: