Microbiology (2011), **157**, 1088–1102 **DOI** 10.1099/mic.0.043018-0

Genomic characterization of asymptomatic *Escherichia coli* isolated from the neobladder

Jason W. Sahl,[1]† Amanda L. Lloyd,[2]† Julia C. Redman,[1] Thomas A. Cebula,[3] David P. Wood,[4] Harry L. T. Mobley[2] and David A. Rasko[1,5]

Correspondence
David A. Rasko
drasko@som.umaryland.edu

[1]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA

[2]Department of Microbiology and Immunology, University of Michigan Medical School, 1150 West Medical Center Drive, 5641 Medical Science II, Ann Arbor, MI 48109, USA

[3]Johns Hopkins University, Department of Biology, 3400 North Charles Street, Baltimore, MD 21218, USA

[4]University of Michigan Medical School, Department of Urology, 1500 East Medical Center Drive, Ann Arbor, MI 48109, USA

[5]Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

The replacement of the bladder with a neobladder made from ileal tissue is the prescribed treatment in some cases of bladder cancer or trauma. Studies have demonstrated that individuals with an ileal neobladder have recurrent colonization by *Escherichia coli* and other species that are commonly associated with urinary tract infections; however, pyelonephritis and complicated symptomatic infections with ileal neobladders are relatively rare. This study examines the genomic content of two *E. coli* isolates from individuals with neobladders using comparative genomic hybridization (CGH) with a pan-*E. coli*/*Shigella* microarray. Comparisons of the neobladder genome hybridization patterns with reference genomes demonstrate that the neobladder isolates are more similar to the commensal, laboratory-adapted *E. coli* and a subset of enteroaggregative *E. coli* than they are to uropathogenic *E. coli* isolates. Genes identified by CGH as exclusively present in the neobladder isolates among the 30 examined isolates were primarily from large enteric isolate plasmids. Isolations identified a large plasmid in each isolate, and sequencing confirmed similarity to previously identified plasmids of enteric species. Screening, via PCR, of more than 100 isolates of *E. coli* from environmental, diarrhoeagenic and urinary tract sources did not identify neobladder-specific genes that were widely distributed in these populations. These results taken together demonstrate that the neobladder isolates, while distinct, are genomically more similar to gastrointestinal or commensal *E. coli*, suggesting why they can colonize the transplanted intestinal tissue but rarely progress to acute pyelonephritis or more severe disease.

Received 26 June 2010
Revised 20 December 2010
Accepted 14 January 2011

INTRODUCTION

The surgical reconstruction of the lower urinary tract is required in cases of bladder cancer, neurological disorder of the bladder (neuropathic bladder), trauma or other clinical indications for the removal of bladder (cystectomy) (Davidsson *et al.*, 2000; Falagas & Vergidis, 2005; Wullt *et al.*, 2004). The bladder can be replaced via a number of surgical techniques and with a variety of tissues, the most common being the construction of an orthotopic neobladder with ileal or colonic tissue (Falagas & Vergidis, 2005). In the construction of a neobladder, a section of the ileal or colonic tissue is excised and formed into a receptacle for the draining ureters (Falagas & Vergidis,

†These authors contributed equally to this work.

Abbreviations: CGH, comparative genomic hybridization; EAEC, enteroaggregative *E. coli*; EHEC, enterohaemorrhagic *E. coli*; EIEC, enteroinvasive *E. coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*; gDNA, genomic DNA; MEV, Multiple Experiment Viewer; MLST, multi-locus sequence typing; SPATE, serine protease autotransporter of *Enterobacteriaceae*; UPEC, uropathogenic *E. coli*; UTI, urinary tract infection.

Comparative genomic hybridization data can be located at www.ncbi.nlm.nih.gov/geo/ under series number GSE27326.

A supplementary figure and five supplementary tables are available with the online version of this paper.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

2005). The connection of the intestinal tissue with the ureter is a key feature of the surgery to prevent reflux of the urine into the kidney, potentially leading to pyelonephritis, kidney scarring and sepsis (Kristjánsson *et al.*, 1995a, b).

Bacteriuria is common among individuals who have had neobladder surgery, with *Escherichia coli*, *Proteus* spp., *Klebsiella* spp. and *Enterobacter faecalis* being the most common bacterial species present (Iwakiri *et al.*, 2002; Wullt *et al.*, 2004). The physical processes associated with the construction of a neobladder may add confounding factors to the attribution of the origin of the bacteriuria; compared with the bladder, the neobladder has altered innervation, which subsequently alters the voiding pattern and bladder retention volume (Iwakiri & Freiha, 1993; Iwakiri *et al.*, 2002; Wullt *et al.*, 2004). Additionally, any innate immunity that was associated with bladder epithelium and associated lower urinary tract is no longer present, although some innate immune attributes of the ileal tissue may contribute to protection (Iwakiri *et al.*, 1993, 2002; Wullt *et al.*, 2004). Studies have postulated that the reason for the lack of widespread upper urinary tract involvement is that the neobladder is a low-pressure environment and thus there is not sufficient pressure to cause vesicouretal reflux of the bacteria into the upper urinary tract (Shaaban *et al.*, 1992a, b).

Long-term studies have demonstrated that many patients with neobladders are colonized continually with the aforementioned bacterial species commonly associated with urinary tract infection (Wullt *et al.*, 2004); however, the rates of symptomatic infection are relatively low, as measured by pyelonephritis and systemic infection associated with colonization (Kristjánsson *et al.*, 1995a, b). The reasons for the lack of symptomatic presentation are not well understood. Åkerlund *et al.* (1994) suggested that aggressive treatment with antibiotics is not required and the observed bacteriuria should be considered part of colonization by normal microbiota (Åkerlund *et al.*, 1994). Conflicting evidence related to the pathogenic potential of *E. coli* colonizing the urinary tract has been presented in the literature. Studies by Schlager *et al.* (2000, 2008) demonstrate that the *E. coli* isolates from neobladder and neurogenic cases are more similar to isolates from individuals with uncomplicated or asymptomatic urinary tract infection using multi-locus sequence typing (MLST) and directed PCR. Conversely, Keegan *et al.* (2000, 2003) used pulsed-field gel electrophoresis and focused examination of uropathogenic virulence factors to suggest that neobladder isolates have a greater pathogenic potential than faecal isolates. In contrast to these previous studies, the current study takes a complete genomic approach to the examination of *E. coli* isolates from neobladder patients.

The *E. coli* and *Shigella* species represent one of the most studied organisms on a genomic scale. As of December 2010, more than 150 genome projects have been completed, yielding either a finished genome sequence (closed to a high standard) or a genome sequence in draft form (multiple high quality contigs) (see http://genomesonline.org/ and http://www.ncbi.nlm.nih.gov/sites/genome). With new genomic technologies becoming more widespread, there are multiple projects that will soon add hundreds of draft genomes from this taxonomic group. Pan-genome studies from publicly available genomes have identified approximately 2200 genes that are common to all *E. coli*; furthermore, the pan-genome of this organism is open, suggesting that there is a constant influx of genetic material into the species (Rasko *et al.*, 2008; Touchon *et al.*, 2009). Molecular models (Tettelin *et al.*, 2008) of the conserved and unique genes in each of the genomes have suggested that ~300 new genes could be identified in each new sequenced genome (Rasko *et al.*, 2008).

Comparative genomic hybridization has been used previously to determine genetic composition in closely related isolates of *E. coli* and *Shigella* (Fukiya *et al.*, 2004; Nie *et al.*, 2006). In the current study, a pan-genome microarray for *E. coli* and *Shigella* containing the genomes of 32 diverse *E. coli* and *Shigella* as well as 46 related plasmid sequences (Fang *et al.*, 2010) was utilized. This pan-genome microarray design is a significant advance over previous array designs, as it contains multiple genomes and allows for the examination of horizontal gene flow between the pathogenic variants (pathovars) of *E. coli* and *Shigella* species included on the array.

The primary goal of this study was to determine whether the environment (urine-filled bladder) or the tissue (gastrointestinal origin) is the defining feature in the colonization of the neobladder by *E. coli*. Overall, these genomic studies demonstrate that isolates from neobladder sources are more similar to commensal and laboratory-adapted isolates, and to one heterogeneous pathovar of diarrhoeagenic *E. coli*, than they are to isolates from uropathogenic sources. This suggests that tissue tropism for the gastrointestinal tract overrides the selective forces of the urine in the colonization processes of the neobladder by *E. coli*.

METHODS

Strain isolation. Isolates were obtained from the urine of asymptomatic individuals who had undergone bladder removal (cystectomy) due to cancer and then received neobladder reconstruction utilizing ileal tissue. Clinical parameters for each of these subjects are included in Table 1. Samples were collected in accordance with the IRB protocol number HUM00004950 at the University of Michigan Medical School. Bacterial isolates were selected by growth on selective media and presumptively identified as *E. coli*. The identity of the isolates as *E. coli* were confirmed by amplification and sequencing of the 16S rRNA gene by conserved universal primers (Lane *et al.*, 1985). The isolates were also sent to the Pennsylvania State University *E. coli* Reference Center for the determination of O- and H-antigens as well as toxin and adhesin typing (http://ecoli.cas.psu.edu/, Table 1).

Genomic DNA isolation for hybridization. Efforts were made to minimize the number of passages of all isolates. Bacterial cultures were grown overnight from a population in 50 ml Luria broth (LB) and genomic DNA was isolated according to standard methods (Ge &

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

Table 1. Clinical and microbiological features associated with neobladder isolates

Sample	Symptom status	Sex of patient	Age of patient	Time since last symptomatic UTI	Time since neobladder	Voiding status	Sero group	H-antigen
NB001	Asymptomatic	Male	70	Not in post-surgical time	<1 year*	Self	NT†	4
NB002	Asymptomatic	Male	69	6 months‡	~10 years	Self	NT	9

*Date of neobladder January 2008; date of sample 29 September 2008.
†NT, non-typable.
‡Last UTI April 2008.

Taylor, 1992). Briefly, bacterial cells were concentrated by centrifugation, washed and suspended in isolation buffer (0.15 M Tris, 0.1 M EDTA, pH 8.0). SDS was added to 1 % (v/v) final concentration and allowed to incubate for 1 h at 55 °C or until the solution cleared. Two volumes of phenol/chloroform/isoamyl alcohol (25 : 24 : 1) were added and briefly mixed by vortexing. The resulting solution was separated by centrifugation at 12 000 ***g*** for 15 min at 4 °C. The aqueous layer (top) was removed to a new tube and mixed with 2 volumes of chloroform. The mixture was separated by centrifugation at 12 000 ***g*** for 15 min at 4 °C and the aqueous layer (top) was transferred to a clean tube. The aqueous layer was then extracted with at least 10 volumes of ice-cold ethanol. The precipitated DNA was spooled out of the mixture and suspended in TE (10 mM Tris/HCl, 1 mM EDTA, pH 8.0). The purified mixture was further digested with RNase overnight at 37 °C and reprecipitated the following day with 0.1 volumes of 3 M sodium acetate and 10 volumes of ice-cold ethanol. The pellet was allowed to air dry and then dissolved in a minimal volume of nuclease-free water (Ambion). The quantity and quality of genomic DNA (gDNA) was verified by gel electrophoresis, nanodrop and picogreen assay.

Comparative genome hybridization (CGH). gDNA was prepared for CGH as previously described (http://www.affymetrix.com/support) with the following modifications. Approximately 7.5 µg purified gDNA isolated as described above was digested with the DNAfree DNase kit (ABI) using timed samples to provide a sample with a size range of between 20 and 200 bp. DNase was inactivated with the Inactivation Reagent provided with the kit. Sample ranges were determined by gel electrophoresis on a 4–20 % gradient gel (Bio-Rad) with Low MW Ladder (Invitrogen). Digested DNA was labelled with biotin by the addition of biotin-11-ddATP (Perkin-Elmer), 2 µl terminal transferase enzyme (Promega) and appropriate buffers. The digests were incubated for 2 h at 37 °C and the labelling was quenched with 0.5 M EDTA. Verification of label incorporation was accomplished by mixing the biotin-labelled DNA with Immunopure NeutrAvidin (Pierce), incubation at room temperature for 5 min, electrophoresis on a 4–20 % gradient gel (Bio-Rad) and visualization. A clear shift in size distribution from unlabelled to labelled DNA was verified prior to hybridization.

Hybridization of the digested probe to the pan-genome array. The pan-genome microarrays used in this study were designed by FDA-Laurel and contain the genomes of 32 diverse *E. coli* and *Shigella* spp. as well as 46 related plasmid sequences (a description and distribution of the microarray are available via the website at the Pathogen Functional Genomics Resources Center, http://pfgrc.jcvi.org/index.php/home/full_news/current/2009_1_9.html; Fang *et al.*, 2010). The microarrays were stored and processed according to the manufacturer's suggestions. Briefly, microarrays were allowed to warm to room temperature for 30 min. The OligoB2 mixture (Affymetrix) was heated to 65 °C for 5 min then mixed with the labelled and verified gDNA, DMSO and hybridization buffer included in the Affymetrix Hybridization, Wash and Stain kit. This solution was further denatured at 94 °C for 5 min. During this incubation the microarrays were equilibrated with prehybridization buffer and placed in the Affymetrix hybridization oven at 45 °C for 10 min, rotating at 60 r.p.m. The labelled DNA was then hybridized to the microarray for 16 h at 45 °C, rotating at 60 r.p.m.

Microarray washing and staining. Microarrays were washed and stained according to manufacturer's specifications using two stain cycles and the prokaryotic washing protocol (http://www.affymetrix.com).

Data analysis. Initial data analysis was performed with the Gene Chip Operating System (GCOS) suite of tools provided by Affymetrix. Additional analysis utilized the Affymetrix power tools (APT) software (http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx). The data analysis parameters have been established empirically using known reference genomes included on the array. Within the APT, the MAS5 algorithm (Hubbell *et al.*, 2002; Liu *et al.*, 2002) was utilized with the perfect match and mismatch calculations and a Tau of 0.150 to detect which probes were present or absent (command=apt-probeset-summarize -a pm-mm,mas5-detect.calls=1.pairs=1.Tau=0.150 -d FDA_ECSGa520423F.cdf -o mas5_Analysis.dir --cel-files cel_file_list.txt). Features that were present or absent in all samples were removed from further analysis. The resulting features, known as the variable gene set, were analysed using Multiple Experiment Viewer v4.5 (MEV; http://www.tm4.org/mev/). Custom Perl scripts written by the authors were used to complete further genomic analyses.

Plasmid preparation. Each of the neobladder isolates was grown in 150 ml LB overnight and subjected to plasmid isolation using the Phase Prep BAC DNA kit (Sigma NA0100-1KT). Each of the plasmid preparations was submitted for sequencing at the Institute for Genome Sciences Genomics Resource Center. Each sample was part of a multiplexed lane of sequencing on the Solexa/Illumina platform. A total of 148 000 50 bp paired end sequences were obtained from plasmid sequences from both neobladder plasmid preparations. *De novo* assembly was performed using Velvet (Zerbino & Birney, 2008) with parameters optimized by the VelvetOptimizer http://www.bioinformatics.net.au/software.velvetoptimiser.shtml. The resulting contigs and scaffolds were compared by BLASTX (Altschul *et al.*, 1990) to the nr database at NCBI, containing all nucleotide data in GenBank, RefSeq Nucleotides, EMBL (European nucleotide database), DDBJ (Japanese nucleotide database) and PDB (Protein Data Bank) sequences, but not expressed sequence tags (EST), Sequence Tagged Sites (STS), Genome Survey Sequences (GSS), or incomplete high-throughput genome sequencing projects, to verify homology to previously annotated plasmid sequence. Contigs were mapped to large (>50 kb) plasmids from *E. coli*, *Shigella* and *Salmonella* using MUMmer (Delcher *et al.*, 2002, 2003; Kurtz *et al.*, 2004). Plasmid sequences from the two isolates were aligned to one another

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

with Mugsy (Angiuoli & Salzberg, 2011) to identify regions of conservation.

***E. coli* genomic isolation for PCR screening.** Genomic DNA was obtained from 133 isolates belonging to three separate sample origins: 21 isolates from the diarrhoeagenic *E. coli* set (DECA), the ECOR set of 72 isolates from the STEC Center (http://www.shigatox.net/stec/cgi-bin/index), which includes 11 isolates of uropathogenic *E. coli*, and an additional 40 *E. coli* isolates from urinary tract infection (UTI) sources (Mobley *et al.*, 1990). Cultures were grown overnight in LB and genomic DNA was isolated using the QIAamp DNA Mini kit (Qiagen) and quantified using Nanodrop, and standard quantities (50 ng) were added to PCRs.

Amplification and analysis of amplicon-based screening of diarrhoeagenic and uropathogenic *E. coli*. PCR primer sequences and targets are listed in Supplementary Table S1, available with the online version of this paper. Primers were designed with a melting point between 68 °C and 70 °C and to amplify products using a touchdown reaction strategy consisting of five cycles of 95 °C for 30 s, 70 °C for 1 min, five cycles of 95 °C for 30 s, 68 °C for 1 min, and 30 cycles of 95 °C for 30 s, 66 °C for 1 min and 72 °C for 5 min using AmpliTaq Gold 360 (Applied Biosystems) reagents. Each of the amplified products was designed to a different size so that the reactions could be multiplexed into a single reaction per sample. Reference genomic and sample DNAs were assayed under recommended conditions with a 2 mM $MgCl_2$ concentration. Each neobladder group (NB001-specific, NB002-specific, and NB001 and NB002) was multiplexed and assayed in a single tube, increasing magnesium concentration to 2.5 mM, dNTP to 0.3 μM each and doubling *Taq* concentration. Amplicons were electrophoresed on 1.5 % agarose gels with ethidium bromide and visualized on a Bio-Rad GelDoc system.

RESULTS

Clinical features of the subjects with neobladders

Two bacterial isolates were cultured from the urine of individuals with neobladders. Both subjects were male, >69 years of age, had received neobladders due to cancer and could void urine from the neobladder independent of catheterization. Both were asymptomatic at the time of isolate collection and had not had a symptomatic urinary tract infection in the 6 months prior to obtaining the culture. Interestingly, the NB001 isolate was provided from a subject who had received a neobladder approximately 8 months earlier, whereas the NB002 isolate was from a subject who had had a neobladder for ~10 years. These isolates provide a baseline for the genomic analysis of *E. coli* isolates from the neobladder.

Microbiological features of isolates from neobladder samples

The bacterial isolates from the two cultures presumptively identified to be *E. coli* were confirmed using comparative analysis of 16S rRNA gene sequences against public databases (data not shown). Serotyping of the O- and H-antigens demonstrated that each of these isolates was non-typable for the O-antigen, but NB001 and NB002 were H-antigen 4 and 9 positive, respectively (Table 1). Additional molecular typing at the *E. coli* Reference Center at the Pennsylvania State University (http://ecoli.cas.psu.edu/) for common diarrhoeagenic *E. coli* toxins [including enterotoxigenic *E. coli* (ETEC) *eltA* (LT), *estA* (STa), *estb* (STb); enterohaemorrhagic *E. coli* (EHEC) Shiga toxin *stxA*, *stxb* and cytotoxic necrotizing factors (*cnf1*, *cnf2*)] as well as enteropathogenic *E. coli* (EPEC) bundle-forming pilus (*bfp*) and EPEC/EHEC intimin (*eae*) did not identify any of these classic virulence factors.

Comparative genomic hybridization of reference and neobladder isolates

Comparative genomic hybridization to the pan-genome *E. coli* and *Shigella* microarray was used to determine the genetic relatedness of the neobladder isolates to a panel of reference *E. coli* isolates (Table 2). This reference isolate set contains representatives of each of the diarrhoeagenic *E. coli* pathovars, a representative of each *Shigella* species and laboratory-adapted as well as human commensal *E. coli* isolates. Through analysis of all 23 090 features on the

Table 2. Isolates hybridized to the pan-genome array

Strain	Pathovar	Genomic reference
EDL933	EHEC	Perna *et al.* (2001)
Sakai	EHEC	Hayashi *et al.* (2001)
8624	EHEC	D. A. Rasko, unpublished
OK1114	EHEC	This work
TX1999	EHEC	This work
B171	EPEC	Rasko *et al.* (2008)
E110019	EPEC	Rasko *et al.* (2008)
E22	EPEC	Rasko *et al.* (2008)
Ec042	EAEC	Unpublished
101-1	EAEC	Rasko *et al.* (2008)
55989	EAEC	Touchon *et al.* (2009)
17-2	EAEC	This work
JM221	EAEC	This work
B7A	ETEC	Rasko *et al.* (2008)
E24377A	ETEC	Rasko *et al.* (2008)
HS	Commensal	Rasko *et al.* (2008)
K-12	Lab	Blattner *et al.* (1997)
CFT073	UPEC	Welch *et al.* (2002)
F11	UPEC	Rasko *et al.* (2008)
UTI536	UPEC	Hochhut *et al.* (2006)
F24	UPEC	Mobley *et al.* (1990)
F38	UPEC	Mobley *et al.* (1990)
CFT01	UPEC	Mobley *et al.* (1990)
NB001	Neobladder	This work
NB002	Neobladder	This work
EIEC_53638	EIEC*	D. A. Rasko, unpublished
SD1_1617	*Shigella dysenteriae*	Hale *et al.* (1984)
SF_2457T	*Shigella flexneri*	Wei *et al.* (2003)
SS_53G	*Shigella sonnei*	Unpublished
SB_BS512	*Shigella boydii*	D. A. Rasko, unpublished

*EIEC, enteroinvasive *E. coli*.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

microarray, 4471 features (19.36 %) were identified that were shared by all 30 isolates hybridized in this study and 1305 features (5.65 %) that were absent in all 30 isolates examined. The remaining 17 314 features displayed a variable pattern of presence or absence. The variable gene set of the 30 isolates included in this study was clustered with an untrained hierarchical clustering method with Pearson correlation and 1000 iterations in MEV (http://www.tm4.org/mev/). This analysis separated isolates into four distinct clusters that are similar to the pathovar evolutionary histories and phylogenetic groups as previously described (Donnenberg & Whittam, 2001; Whittam, 1993; Whittam *et al.*, 1983). In Fig. 1, the four clusters are uropathogenic *E. coli* (UPEC) (coloured in gold; upper right), *Shigella* and O157 : H7 EHEC (coloured in burgundy and brown, respectively; upper left), non-O157 : H7 EHEC, EPEC and ETEC (coloured brown, blue and red respectively; lower right) and laboratory-adapted, commensal, neobladder isolates and three out of five enteroaggregative *E. coli* (EAEC) (coloured black, red and green; lower left). EAEC isolates are represented in three of the four clusters, but not in the cluster that contains the UPEC isolates. This is most probably due to the diverse nature of the EAEC pathovar as previously determined by multi-locus sequence typing and phenotypic methods (Czeczulin *et al.*, 1999). Notably, both neobladder isolates examined are within the commensal/laboratory-adapted cluster (Fig. 1, arrows) and not within the UPEC cluster. These data demonstrate that the neobladder isolates are more similar to commensal and/or diarrhoeagenic *E. coli* than they are to UPEC isolates on a genomic level.

Identification of UPEC virulence factors in the CGH data

On a whole-genome scale the neobladder isolates did not appear similar to UPEC (Fig. 1). To determine if the neobladder isolate diversity extended to UPEC virulence factors, a defined set of UPEC virulence factors previously identified (Lloyd *et al.*, 2007, 2009) was examined. In total, 172 UPEC virulence factor gene features were identified, including serine protease autotransporters of *Enterobacteriaceae* (SPATEs), genes involved in the acquisition or sequestration of iron, adhesins and pathogenicity islands. Fig. 2 demonstrates that the UPEC-associated virulence factors are conserved in the UPEC isolates. Conversely, these UPEC virulence factors are not consistently identified in non-UPEC isolates. This finding could also be extended to the various pathovars represented in the hybridization data (see Supplementary Table S2). The difference between the UPEC virulence factor profile in each of the other pathovars was statistically significant, as determined by a Student's *t*-test ($P<0.05$) with the exception of the neobladder isolates ($P=0.058$). This borderline *P*-value is most probably due to the relatively low number of isolates in the neobladder group.

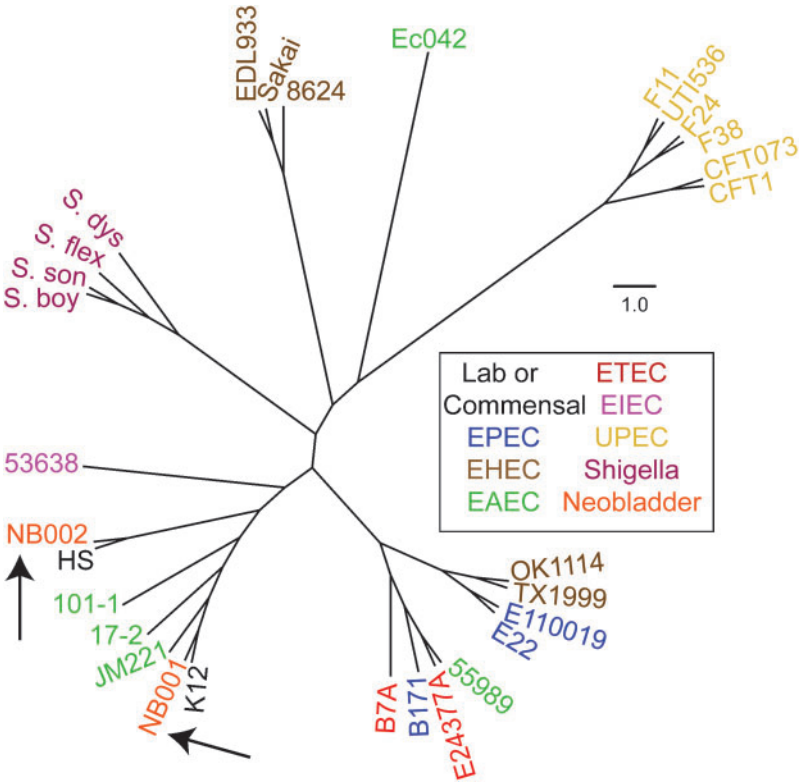

Fig. 1. Hierarchical cluster diagram displaying the relatedness of the reference and neobladder isolates. Using only the variable regions from the CGH studies the samples were analysed by sample with hierarchical clustering including support trees with Pearson correlation including 1000 iterations. Isolates of *E. coli* and *Shigella* are coloured according to pathovar. The arrows indicate the location of the neobladder isolates relative to the other isolates screened. The neobladder isolates cluster in a group that includes EAEC, commensal and laboratory isolates but no UPEC isolates.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

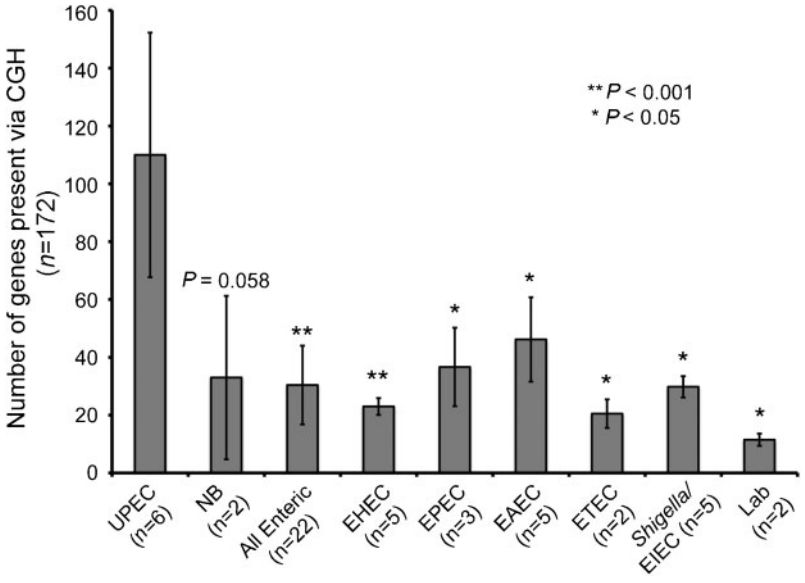

Fig. 2. Identification of UPEC virulence factors in the comparative genomic hybridization (CGH) data. Gene features previously identified as UPEC virulence factors (Supplementary Table S3) from the CGH data were identified and examined. Groups of isolates from each of the pathovars and species were examined as labelled in Table 2. The number of virulence genes present in each of the pathovars is represented by the mean and standard deviation (or standard error). When compared with the genes present in UPEC isolates each of the other pathovars has significantly fewer UPEC virulence genes, as measured with a Student's *t*-test, with the exception of the neobladder isolates (P=0.058).

Distribution of the features among the 30 isolates

Examination of the distribution of the number of features present or absent in each genome identifies the open pan-genome nature of *E. coli* (Rasko *et al.*, 2008) (Fig. 1). As mentioned previously, of the 23 090 features on the array, 4471 features were present in all and 1305 features absent in all the 30 genomes examined. The remaining features are distributed among the other isolates in a pattern that suggests uninhibited lateral gene transfer between all of the pathovars and isolates. Detailed examination of each of the cluster-specific features provides genomic content that is exclusive to each of the pathogenic clusters. There are 273 features that are exclusively present among the UPEC isolates screened and not present in any of the diarrhoeagenic isolates or the neobladder isolates (see Supplementary Table S4). Many of these features are from the *E. coli* CFT073 genome (Welch *et al.*, 2002), a prototype UPEC isolate (Mobley *et al.*, 1990). Interestingly, there are no features that are exclusively shared between the neobladder isolates and the UPEC isolates; however, there are 154 features that are shared exclusively between the neobladder isolates and the non-UPEC isolates included in this study (see Supplementary Table S5). These data further support the hypothesis that the neobladder isolates are more similar to diarrhoeagenic isolates than to UPEC isolates.

Hybridization features exclusive to the neobladder isolates

The identification of features that were exclusive to the neobladder isolates provides insight into what may facilitate colonization of the neobladder. Genes that were exclusive to each of the neobladder strains included 100 genes for NB001 and 20 genes for NB002 (Tables 3 and 4 respectively). Nine features were identified that were exclusively present in both neobladder isolates (Table 5). All nine of these features are similar to those encoded on the ETEC plasmid pCOO (Froehlich *et al.*, 2005). These genes are all from a type IV secretion apparatus known as the R64 thin pilus, which is utilized in liquid-based conjugations (Kim & Komano, 1997). Based on these hybridization data, it was postulated that each isolate contains one or more plasmids.

Plasmid isolation and sequencing from neobladder isolates

The plasmids from each neobladder isolate were purified and their size examined based on gel electrophoresis (Fig. 3a). Each of the isolates contained a large (>50 kb) plasmid and NB001 also contained a much smaller plasmid (~ 3 kb). The plasmid preparations were then sequenced to identify if the plasmids were similar to other plasmids that have been previously identified from enteric sources. The assembled contigs for each plasmid preparation were compared with a local database containing the complete plasmids in GenBank from *E. coli*, *Shigella* and *Salmonella* isolates. Each isolate contained a plasmid similar to a different large plasmid. The 87 plasmid contigs from the isolate NB001, totalling ~130 kb, were similar to the pAPEC-O2-ColV plasmid (GenBank accession number AY545598; Skyberg *et al.*, 2006). The 50 plasmid contigs, totalling ~72 kb, obtained from the NB002 isolate were most similar to the pSL476_91 plasmid from *Salmonella enterica* subsp. *enterica* serovar Heidelberg str. SL476 (GenBank accession number CP001118). For each large plasmid, the sequences do not appear to be identical to the closest relative. However, between neobladder plasmids, a conserved backbone of plasmid sequence (41 kb) was identified by the Mugsy alignment (Fig. 3b, c). This region contained conserved plasmid functions including replication and transfer.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

Table 3. Genes specific to isolate NB001

Affy_ID	Origin	Gene designation	Annotation
APECO1_3532_at	*E. coli* APECO1	APECO1_3532	Putative superfamily I DNA helicase
APECO1_3530_at	*E. coli* APECO1	APECO1_3530	*tia*; Tia invasion determinant
APECO1_3484_at	*E. coli* APECO1	APECO1_3484	Hypothetical protein
APECO1_3483_at	*E. coli* APECO1	APECO1_3483	Conserved hypothetical protein; putative EAL domain
N5007_x_at	*E. coli* E2348/69	N5007	*traX*; conjugal transfer pilus acetylation protein TraX
N5010_at	*E. coli* E2348/69	N5010	*traD*; conjugal transfer protein TraD
N5020_s_at	*E. coli* E2348/69	N5020	*artA*; hypothetical protein
N5021_s_at	*E. coli* E2348/69	N5021	Hypothetical protein
V2175_at	*E. coli* 7122 O78	V2175	Hypothetical protein
V2176_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV64	*cvaB*; colicin V secretion/processing ATP-binding protein CvaB [EC 3.4.22.–]
V2177_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV63	*cvaA*; colicin V secretion protein CvaA; K02022
V2182_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV59	Phospho-2-dehydro-3-deoxyheptonate aldolase [EC 2.5.1.54]; K01626 3-deoxy-7-phosphoheptulonate synthase [EC 2.5.1.54]
V2189_s_at	*E. coli* 7122 O78	V2189	Hypothetical protein
V2198_at	*E. coli* 7122 O78	V2198	Hypothetical protein
V2200_at	*E. coli* 7122 O78	V2200	Hypothetical protein
V2202_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV34	*etsB*; EtsB; K05685 macrolide transport system permease protein; K05686 macrolide transport system ATP-binding protein
O2ColV33_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV32	*etsA*; putative type I secretion membrane-fusion protein EtsA
V2205_at	*E. coli* 7122 O78	V2205	Hypothetical protein
V2206_at	*E. coli* 7122 O78	V2206	Hypothetical protein
V2213_s_at	*E. coli* plasmid p1658/97	p166897_107	Mig-14 family protein
V2224_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV4	*sitD*; SitD protein, iron transport protein, inner membrane component
V2227_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV11	*iucA*; aerobactin siderophore biosynthesis protein IucA
V2233_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV14	*iucD*; L-lysine 6-monooxigenase IucD [EC 1.14.13.59]
V2234_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV15	*iutA*; ferric aerobactin receptor precursor
IS2_03_s_at	*E. coli* K-12 MG1655	IS2_3	*aph*, kan; aminoglycoside 3′-phosphotransferase; K00897 aminoglycoside 3′-phosphotransferase [EC 2.7.1.95]
p166897_027_at	*E. coli* plasmid p1658/97	p166897_027	*artA*; hypothetical protein
p166897_029_x_at	*E. coli* plasmid p1658/97	p166897_029	Putative conjugal transfer protein TrbB
p166897_031_s_at	*E. coli* plasmid p1658/97	p166897_031	*trbF*; conjugal transfer protein TrbF
p166897_106_at	*E. coli* plasmid p1658/97	p166897_106	*hlyF*; haemolysin HlyF
pIGAL1_01_s_at	*E. coli* A2363 plasmid pAPEC-O2-R	O2R_122	Transposase
pIGAL1_02_s_at	*E. coli* A2363 plasmid pAPEC-O2-R	O2R_123	Transposon Tn*3* resolvase
O2R_06_at	*E. coli* A2363 plasmid pAPEC-O2-R	O2R_6	Hypothetical protein
O2R_24_x_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV118	*traQ*; conjugal transfer pilin chaperone TraQ
LH0096_x_at	*E. coli* B171	LH0096	Conserved predicted protein
LH0129_at	*E. coli* B171	LH0129	Putative immunity protein for colicin IA
pO86A1_p048_x_at	*E. coli* plasmid pO861A1	pO86A1_p048	*imm*; colicin-Ia immunity protein
O2ColV164_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV164	Hypothetical protein
O2ColV168_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV168	Hypothetical protein
O2ColV169_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV169	Hypothetical protein

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

Table 3. cont.

Affy_ID	Origin	Gene designation	Annotation
O2ColV170_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV170	Carbamate kinase-like protein YahI; K00926 carbamate kinase [EC 2.7.2.2]
O2ColV171_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV171	Hypothetical protein
O2ColV172_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV172	Putative permease
O2ColV173_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV173	Hypothetical protein
O2ColV174_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV174	Hypothetical protein
O2ColV175_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV175	Conserved predicted protein
O2ColV178_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV178	ABC-type transporter, ATPase and permease components [EC 3.6.3.–]
O2ColV179_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV179	Hypothetical protein
O2ColV180_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV180	Hypothetical protein
O2ColV183_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV183	Putative IS*602* transposase OrfA
O2ColV184_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV184	Hypothetical protein
O2ColV185_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV185	*yahB*; putative DNA-binding transcriptional regulator
O2ColV186_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV186	Ankyrin; K06867
O2ColV187_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV187	*yahE*; hypothetical protein
O2ColV188_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV188	FdrA family protein
O2ColV189_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV189	Hypothetical protein
O2ColV190_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV190	*yahG*; hypothetical protein
O2ColV191_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV191	Xanthine/uracil/vitamin C permease; K06901 putative MFS transporter, AGZA family, xanthine/uracil permease
O2ColV192_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV192	*yahI*; carbamate kinase-like protein YahI; K00926 carbamate kinase [EC 2.7.2.2]
O2ColV193_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV193	Hypothetical protein
O2ColV194_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV194	Putative transposase
O2ColV5_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV5	Putative enolase
O2ColV6_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV6	Putative enolase
O2ColV7_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV7	Hypothetical protein
O2ColV8_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV8	Putative membrane protein; CrcB-like protein; K06199 CrcB protein
O2ColV9_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV9	*shiF*; putative membrane transport protein ShiF
O2ColV15_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV15	*iutA*; ferric aerobactin receptor precursor IutA

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

Table 3. cont.

Affy_ID	Origin	Gene designation	Annotation
O2ColV16_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV16	Hypothetical protein
O2ColV21_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV21	Hypothetical protein
O2ColV25_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV25	Hypothetical protein
O2ColV26_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV26	*ompT*; outer membrane protease [EC 3.4.23.49]
O2ColV27_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV27	Hypothetical protein
O2ColV28_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV28	Hypothetical protein
O2ColV29_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV29	Transposase B; K07497 putative transposase
O2ColV46_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV46	Putative transposase ORF B (fragment), IS*2*
O2ColV47_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV47	Conserved hypothetical protein; putative GTPase
O2ColV49_x_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV49	*borD*; putative lipoprotein; DLP12 prophage
O2ColV49.2_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV49.2	Hypothetical protein
O2ColV51_x_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV51	Predicted transposase
O2ColV58_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV58	Hypothetical protein
O2ColV60_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV60	Hypothetical protein
O2ColV65_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV65	*cvaC*; colicin V precursor (microcin V bacteriocin)
O2ColV66_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV66	*cvi*; colicin V immunity protein
O2ColV67_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV67	Hypothetical protein
O2ColV68_s_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV68	Hypothetical protein
O2ColV75_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV75	Resolvase domain-containing protein
O2ColV76_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV76	Hypothetical protein
O2ColV78_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV78	DNA-binding protein
O2ColV84_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV84	Resolvase
O2ColV85_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV85	Resolvase
O2ColV91_at	*E. coli* A2363 plasmid pAPEC-O2-Col-ColV	O2ColV91	*kda*; putative klebicin D activity protein
pCoo111_at	*E. coli* plasmid pCoo	pCoo111	Prepilin peptidase
pCoo112_at	*E. coli* plasmid pCoo	pCoo112	Lytic transglycosylase PilT
pCoo113_at	*E. coli* plasmid pCoo	pCoo113	Type IV prepilin
pCoo114_at	*E. coli* plasmid pCoo	pCoo114	Integral membrane protein
pCoo116_at	*E. coli* plasmid pCoo	pCoo116	*pilP*; type IV pilus biogenesis protein PilP
pCoo118_at	*E. coli* plasmid pCoo	pCoo118	Type IVB pilus formation outer-membrane protein, R64 PilN family

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

Table 3. cont.

Affy_ID	Origin	Gene designation	Annotation
pCoo121_at	*E. coli* plasmid pCoo	pCoo121	PilK protein
pCoo122_at	*E. coli* plasmid pCoo	pCoo122	PilJ protein
pCoo123_at	*E. coli* plasmid pCoo	pCoo123	PilI protein
pCoo125_at	*E. coli* plasmid pCoo	pCoo125	PilK protein

Screening of environmental, diarrhoeagenic and uropathogenic *E. coli* for features identified in neobladder isolates

Gene features were selected for PCR screening from the CGH data that were present in both of the neobladder isolates and not in any other group (three features), exclusive to each of the neobladder isolates (three features for NB001, two features for NB002) or more broadly conserved in UPEC (two features) or enteric *E. coli* isolates (three features) (Supplementary Fig. S1). Primer sequences and gene features are listed in Supplementary Table S1. We interrogated 133 *E. coli* isolates: 61 isolates from environmental sources or non-human animals (http://www.shigatox.net/stec/cgi-bin/index), 21 diarrhoeagenic isolates (http://www.shigatox.net/stec/cgi-bin/index) and 51 isolates from human urinary tracts (Mobley *et al.*, 1990). Gene features that were demonstrated in the CGH data to be more prevalent in the UPEC or enteric genomes were more prevalent in each respective sample population when examined by PCR screening in the larger study group. For example, genes identified as more common in UPEC isolates by CGH were more prevalent in the examined UPEC isolates than in the enteric or environmental isolates (Fig. 4). The features that were identified in one or more of the neobladder isolates were not widely distributed among

Table 4. Genes specific to isolate NB002

Affy_ID	Origin	GenBank accession	Gene designation	Annotation
SBO_0869_x_at	*Shigella boydii* Sb227	NC_007613	SBO_0869	*galF*; UTP–glucose-1-phosphate uridylyltransferase subunit GalF; K00963 UTP–glucose-1-phosphate uridylyltransferase [EC 2.7.7.9]
SFV_0202_at	*Shigella flexneri* 5 str. 8401	NC_008258	SFV_0202	Hypothetical protein; K11895 type VI secretion system protein ImpH
SFV_0630_at	*Shigella flexneri* 5 str. 8401	NC_008258	SFV_0630	Putative *rhs*-family protein
SD1_0080_x_at	*Shigella* sp.	AAMJ01000001.1	SD1_0080	Putative L-xylulose-5-phosphate 3-epimerase; K03082 hexulose-6-phosphate isomerase [EC 5.–.–.–]
SD1_0822_at	*Shigella* sp.	AAMJ01000002.1	SD1_0822	Hypothetical protein
SD1_0823_at	*Shigella* sp.	AAMJ01000002.1	SD1_0823	Transcriptional regulator
SD1_0824_at	*Shigella* sp.	AAMJ01000002.1	SD1_0824	Galactitol-specific enzyme IIA of phosphotransferase system; K02773 PTS system, galactitol-specific IIA component [EC 2.7.1.69]
SD1_0825_at	*Shigella* sp.	AAMJ01000002.1	SD1_0825	PTS system, galactitol-specific IIB component, putative; K02774 PTS system, galactitol-specific IIB component [EC 2.7.1.69]
SD1_0827_at	*Shigella* sp.	AAMJ01000002.1	SD1_0827	PTS system, galactitol-specific IIC component
SD1_3295_at	*Shigella* sp.	AAMJ01000019.1	SD1_3295	Hypothetical protein; K11891 type VI secretion system protein ImpL
N0356_x_at	*E. coli* E2348/69	E2348_69000001.1	N0356	Hypothetical protein
N0946_at	*E. coli* E2348/69	E2348_69000001.1	N0946	Hypothetical protein
N0948_at	*E. coli* E2348/69	E2348_69000001.1	N0948	Hypothetical protein
N0950_at	*E. coli* E2348/69	E2348_69000001.1	N0950	Hypothetical protein
N0953_at	*E. coli* E2348/69	E2348_69000001.1	N0953	Hypothetical protein
N1524_at	*E. coli* E2348/69	E2348_69000001.1	N1524	*pduF*; propanediol diffusion facilitator
N1528_at	*E. coli* E2348/69	E2348_69000001.1	N1528	Hypothetical protein
N1529_at	*E. coli* E2348/69	E2348_69000001.1	N1529	Hypothetical protein
V1141_at	*E. coli* 7122 O78	EC7122000072	V1141	Hypothetical protein
EC869_7014_s_at	*E. coli* APECO1	NC_008563	APECO1_4474	Putative protein LysA from bacteriophage P2

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

Table 5. Genes identified by CGH as only present in the neobladder isolates

Affy_ID	Origin	GenBank accession	Gene number	Gene name	Annotation
pCoo111_at	*E. coli* plasmid pCoo	NC_007635	pCoo111	*pilU*	Prepilin peptidase
pCoo112_at	*E. coli* plasmid pCoo	NC_007635	pCoo112	*pilT*	Putative transglycosylase PilT
pCoo113_at	*E. coli* plasmid pCoo	NC_007635	pCoo113	*pilS*	Type IV prepilin
pCoo114_at	*E. coli* plasmid pCoo	NC_007635	pCoo114	*pilR*	Integral membrane protein
pCoo116_at	*E. coli* plasmid pCoo	NC_007635	pCoo116	*pilP*	PilP protein
pCoo118_at	*E. coli* plasmid pCoo	NC_007635	pCoo118	*pilN*	PilN protein
pCoo121_at	*E. coli* plasmid pCoo	NC_007635	pCoo121	*pilK*	PilK protein
pCoo122_at	*E. coli* plasmid pCoo	NC_007635	pCoo122	*pilJ*	PilJ protein
pCoo123_at	*E. coli* plasmid pCoo	NC_007635	pCoo123	*pilI*	PilI protein

the isolates of the three populations examined and did not appear to correlate based on the origin of the isolates.

DISCUSSION

Previous studies of neobladder-associated bacteria have focused on a limited number of virulence factors (Keegan *et al.*, 2000, 2003), but to our knowledge this is the first study of neobladder isolates on a genomic scale. In this study, the genomic content of two *E. coli* isolates obtained from individuals who had had neobladder surgery were examined. Comparative genomic hybridization demonstrated that the neobladder isolates were not in the same cluster as UPEC isolates, based on presence/absence in the variable gene content, but were more similar to the diarrhoeagenic, commensal or laboratory *E. coli* isolates. The distribution of gene presence and absence among the collection of 30 isolates examined using a pan-*E. coli/Shigella* genome microarray suggested that there was no barrier to horizontal gene transfer. These findings were extended by examining the hybridization patterns of known UPEC virulence factors, and these data indicate that the neobladder isolates are selected for by the environment from which they were isolated much in the same way that UPEC isolates are selected by the environment of the urinary tract (Chen *et al.*, 2006, 2009). The five UPEC isolates examined in this study exclusively shared 273 features (see Supplementary Table S4), many originating in UPEC isolates or the closely related avian *E. coli* (Johnson *et al.*, 2007). This suggests that there is a conserved gene repertoire selected among the UPEC isolates for survival and/or colonization of the urinary tract. Interestingly, no array feature was shared exclusively between the UPEC and the neobladder isolates, whereas 154 features were shared with the neobladder, diarrhoeagenic, commensal or laboratory-adapted isolates and not the UPECs (Supplementary Table S5). Further screening by PCR of a collection of 133 *E. coli* isolates indicated that the genes unique to the neobladder isolates were not common among the environmental, diarrhoeagenic or UPEC isolates screened. This suggests that while on a genomic scale the neobladder isolates are similar to enteric isolates, there does appear to be some unknown pressure to retain the neobladder-specific genes that are not observed in the other populations. These analyses indicate that the tropism of the bacterium to the gastrointestinal tissue in a neobladder is stronger than the selective pressure of the UPEC genes to survive or colonize in the presence of urine.

One question that remains unanswered by this study is whether the neobladder isolates obtained at 8 months post surgery and over 10 years post surgery were present in

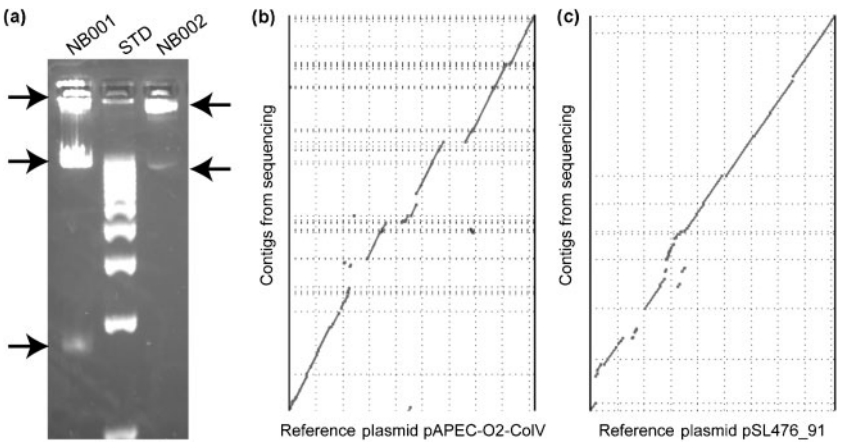

Fig. 3. Neobladder isolates contain plasmids. (a) Plasmids were isolated from overnight cultures of NB001 and NB002 using the Phase prep BAC DNA kit (Sigma NA0100-1KT) plasmid preparation kit. Plasmids were visualized on a 1.0 % agarose gel stained with ethidium bromide (STD, Invitrogen 1 kb Plus DNA Ladder 8). (b, c) MUMmer plots demonstrating that in selected sequencing the plasmids are most similar to pAPEC-O2-ColV and pSL476-91 in NB001 and NB002 respectively. The compared plasmids demonstrate coverage of the entire plasmid and are >180 and 90 kb respectively.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

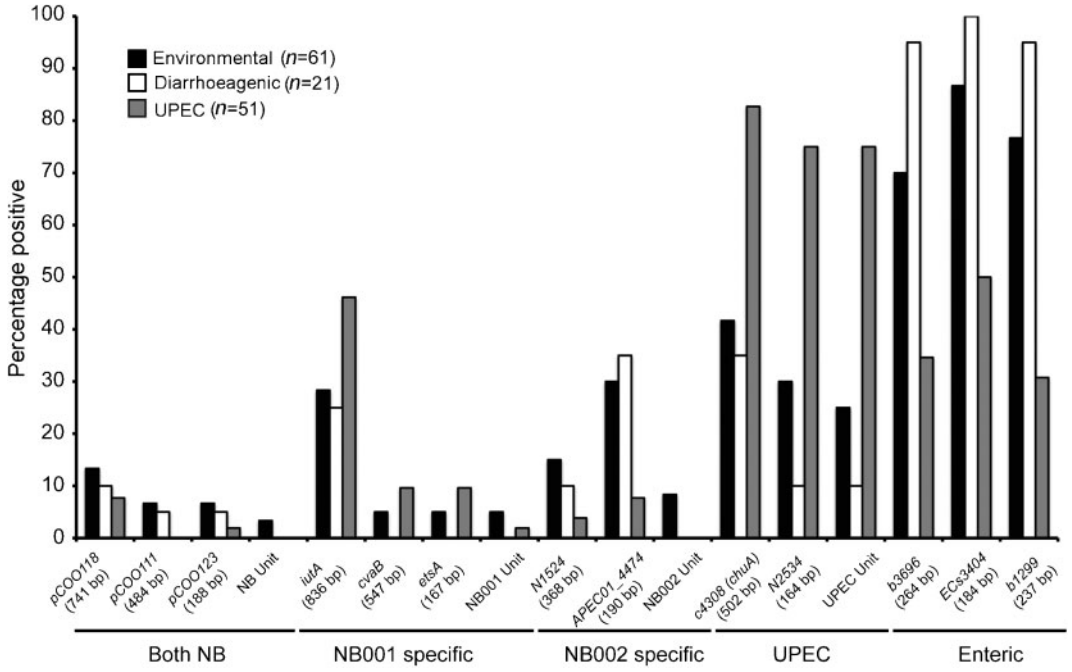

Fig. 4. PCR screening of environmental, diarrhoeagenic and uropathogenic *E. coli*. A total of 133 isolates were selected for PCR-based screening representing populations from environmental (black bars), diarrhoeagenic (white bars) and uropathogenic (grey bars) sources. Assays were developed which represented genes that, according to CGH, were conserved in both neobladder isolates, either neobladder isolate, UPEC or enteric isolates. The neobladder features were present only in few isolates from any of the populations (first three groups of assays). When taken as a unit, these genes were present in even fewer isolates. UPEC and enteric gene features were present in the UPEC or diarrhoeagenic/environmental isolates, respectively. These data suggest that the neobladder-specific genes are not widely distributed in these populations.

association with the tissue as it was transported from the ileum to the urinary tract. Antibiotics are frequently employed prior to and after surgery to decrease the bacterial load. Additionally, there is a process of physical cleaning of the ileal tissue to be used for neobladder construction; however, this tissue can never be made sterile and there is the likely carryover of microbiota that includes *E. coli* (Studer *et al.*, 1991, 1996). It is possible that the microbiota, and *E. coli* as a member of the gastrointestinal flora, were transferred from the gastrointestinal tract to the urinary tract with the tissue. The *E. coli* isolate obtained over 10 years post surgery is unlikely to be identical to an original isolate from the gastrointestinal tract, but the isolate obtained less than 1 year post surgery is a possible remnant of the gastrointestinal tract flora. Another remaining question is whether the isolates obtained are representative of the *E. coli* in the neobladder community over time; for example, if isolates were examined longitudinally would we find the same isolate multiple times or are the isolates in flux over time? This study represents a preliminary foray into the genomics of *E. coli* neobladder isolates and does not suggest that this limited number of isolates represents the complete genetic diversity of neobladder isolates. Further study of larger sample collections, including additional neobladder isolates, will be required to determine if these findings are generally applicable to a large collection of neobladder *E. coli*.

The comparison of multiple isolates from each of the major pathovars helps to identify how each of these isolates are related on a genomic scale (Fig. 1). Recent genomic sequencing studies have examined multiple genomes and demonstrated that there is a significant amount of variability within each of the *E. coli* pathovars (Rasko *et al.*, 2008; Touchon *et al.*, 2009). Comparative genomic hybridization using a pan-genome array allows for rapid and accurate examination of the genomic content and provides a limited view of the genetic variability and horizontal gene transfer between the isolates in relation to the data that are contained on the microarray. Interestingly, the relative groupings in Fig. 1 are similar to the phylogenetic relationships that have been previously determined for a number of these isolates (Donnenberg & Whittam, 2001; Whittam *et al.*, 1983, 1993). In this study, a number of interesting phylogenetic trends are observed that must be confirmed with larger strain collections. For example, the EAEC isolates examined do not consistently fall into any one of the clusters observed in this study. This is congruent with MLST studies suggesting that this pathovar was more variable than other pathovars in terms of phylogeny (Czeczulin *et al.*, 1999; Valentiner-Branth *et al.*, 2003) and in the ability to cause disease in humans (Nataro *et al.*, 1995). EHEC was the one other pathovar that was represented in multiple clusters; however, this observation

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

is consistent with the literature on EHEC evolution based on MLST and virulence markers (Donnenberg & Whittam, 2001). The O157 : H7 isolates EDL933, 86-24 and Sakai are members of the EHEC1 group whereas OK1114 and TX1999 are recent O111 : H8 isolates and members of the EHEC2 phylogenetic group (Donnenberg & Whittam, 2001). The comparative genomic hybridization and analysis confirm general MLST and virulence gene typing but provide a much more robust method to examine these isolates on a genomic scale. By continuing to expand our bank of genomic hybridizations, we can rapidly and accurately examine and include new isolates in these comparative studies.

In addition to comparing the hybridization profiles to a bank of reference genome hybridizations, the information from the hybridization of these neobladder isolates was also used to focus further analysis on the plasmid content. The hybridization patterns of the unique features of the neobladder isolates (Table 5) suggested that each isolate contained at least one plasmid. The genes shared only between the neobladder isolates are part of a type IV secretion system from the pCOO plasmid in ETEC (Froehlich *et al.*, 2005). Other genes that are specific to each of the isolates (Tables 3 and 4) are also part of plasmid transfer or replication genes, suggesting that the core of the type IV system is similar but not identical between these two plasmids. Further isolation and sequencing of the plasmids from each isolate revealed that each isolate harboured a large plasmid (>50 kb, Fig. 3). The NB001 plasmid was most similar to the avian pathogenic *E. coli* (APEC) plasmid (Johnson *et al.*, 2007) and the NB002 plasmid was most similar to the *Salmonella enterica* subsp. *enterica* serovar Heidelberg str. SL476 pSL476_91 plasmid. Each of these plasmids are common mobile plasmids in the enteric pathogens (Johnson & Nolan, 2009) and further represent the horizontal gene transfer of these isolates and their similarity to the *E. coli* from the gastrointestinal tract.

Patients with neobladders often indicate that it is difficult to determine if they are experiencing a urinary tract infection, as the symptomology does not present in the same manner as a typical UTI, even though the bacterial counts may be clinically significant (Suriano *et al.*, 2008; Wullt *et al.*, 2004). There is often an absence of pain, and the incidence of upper urinary tract involvement is less common than in traditional UTI (Kristjánsson *et al.*, 1995a, b). This lack of ascending infection is often attributed to the fact that the neobladder is a low-pressure vesicle and thus there is a lack of pressure to aid the bacterium in ascending the ureters. However, genomic examination of neobladder isolates indicates that these isolates are more similar to the diarrhoeagenic isolates and not similar to UPEC isolates. It is possible that the *E. coli* isolates within a neobladder, being more similar to diarrhoeagenic, commensal and laboratory-adapted isolates, do not have the genetic capability to ascend the ureters, adhere to proximal tubule epithelium and cause pyelonephritis or subsequent systemic infection. Additionally, it may be that colonization by a commensal-like *E. coli* that causes no disease ends up creating a 'probiotic effect' that inhibits colonization by UPEC-like strains, thus preventing the secondary disease often observed. The ability of the gastrointestinal tissue to tolerate interactions with bacterial communities may prove beneficial, as colonized tissue does not result in a symptomatic presentation. This study allows us to postulate that the 'colonization' of the neobladder could, and most likely should, be considered a commensal colonization of gastrointestinal tissue and not an infectious process requiring treatment.

ACKNOWLEDGEMENTS

This work was supported from startup funds from the State of Maryland to the Rasko laboratory.

REFERENCES

Åkerlund, S., Campanello, M., Kaijser, B. & Jonsson, O. (1994). Bacteriuria in patients with a continent ileal reservoir for urinary diversion does not regularly require antibiotic treatment. *Br J Urol* **74**, 177–181.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Angiuoli, S. V. & Salzberg, S. L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* **27**, 334–342.

Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K. & other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462.

Chen, S. L., Hung, C. S., Xu, J., Reigstad, C. S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R. R. & other authors (2006). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proc Natl Acad Sci U S A* **103**, 5977–5982.

Chen, S. L., Hung, C. S., Pinkner, J. S., Walker, J. N., Cusumano, C. K., Li, Z., Bouckaert, J., Gordon, J. I. & Hultgren, S. J. (2009). Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding. *Proc Natl Acad Sci U S A* **106**, 22439–22444.

Czeczulin, J. R., Whittam, T. S., Henderson, I. R., Navarro-Garcia, F. & Nataro, J. P. (1999). Phylogenetic analysis of enteroaggregative and diffusely adherent *Escherichia coli*. *Infect Immun* **67**, 2692–2699.

Davidsson, T., Wullt, B., Könyves, J., Månsson, A. & Månsson, W. (2000). Urinary diversion and bladder substitution in patients with bladder cancer. *Urol Oncol* **5**, 224–231.

Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* **30**, 2478–2483.

Delcher, A. L., Salzberg, S. L. & Phillippy, A. M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics* Chapter 10, 10.3.

Donnenberg, M. S. & Whittam, T. S. (2001). Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. *J Clin Invest* **107**, 539–548.

Falagas, M. E. & Vergidis, P. I. (2005). Urinary tract infections in patients with urinary diversion. *Am J Kidney Dis* **46**, 1030–1037.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

Fang, H., Xu, J., Ding, D., Jackson, S. A., Patel, I. R., Frye, J. G., Zou, W., Nayak, R., Foley, S. & other authors (2010). An FDA bioinformatics tool for microbial genomics research on molecular characterization of bacterial foodborne pathogens using microarrays. *BMC Bioinformatics* **11** (Suppl 6), S4.

Froehlich, B., Parkhill, J., Sanders, M., Quail, M. A. & Scott, J. R. (2005). The pCoo plasmid of enterotoxigenic *Escherichia coli* is a mosaic cointegrate. *J Bacteriol* **187**, 6509–6516.

Fukiya, S., Mizoguchi, H., Tobe, T. & Mori, H. (2004). Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J Bacteriol* **186**, 3911–3921.

Ge, Z. & Taylor, D. E. (1992). *H. pylori* DNA transformation by natural competence and electroporation. In *Helicobacter pylori Protocols*, pp. 145–152. Edited by C. L. Clayton & H. L. T. Mobley. Totowa: Humana Press.

Hale, T. L., Guerry, P., Seid, R. C., Jr, Kapfer, C., Wingfield, M. E., Reaves, C. B., Baron, L. S. & Formal, S. B. (1984). Expression of lipopolysaccharide O antigen in *Escherichia coli* K-12 hybrids containing plasmid and chromosomal genes from *Shigella dysenteriae* 1. *Infect Immun* **46**, 470–475.

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C. G., Ohtsubo, E., Nakayama, K. & other authors (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157 : H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**, 11–22.

Hochhut, B., Wilde, C., Balling, G., Middendorf, B., Dobrindt, U., Brzuszkiewicz, E., Gottschalk, G., Carniel, E. & Hacker, J. (2006). Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. *Mol Microbiol* **61**, 584–595.

Hubbell, E., Liu, W. M. & Mei, R. (2002). Robust estimators for expression analysis. *Bioinformatics* **18**, 1585–1592.

Iwakiri, J. & Freiha, F. (1993). Stanford pouch ileal neobladder: clinical, radiologic, and urodynamic follow-up. *Urology* **41**, 517–522.

Iwakiri, J., Gill, H., Anderson, R. & Freiha, F. (1993). Functional and urodynamic characteristics of an ileal neobladder. *J Urol* **149**, 1072–1076.

Iwakiri, J., Freiha, F. S. & Shortliffe, L. M. (2002). Prospective study of urinary tract infections and urinary antibodies after radical prostatectomy and cystoprostatectomy. *Urol Clin North Am* **29**, 251–258, xii.

Johnson, T. J. & Nolan, L. K. (2009). Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol Mol Biol Rev* **73**, 750–774.

Johnson, T. J., Kariyawasam, S., Wannemuehler, Y., Mangiamele, P., Johnson, S. J., Doetkott, C., Skyberg, J. A., Lynne, A. M., Johnson, J. R. & Nolan, L. K. (2007). The genome sequence of avian pathogenic *Escherichia coli* strain O1 : K1 : H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J Bacteriol* **189**, 3228–3236.

Keegan, S. J., Hormaeche, C. E., Pearson, J. P. & Gally, D. L. (2000). Characterisation and adherence mechanisms of *Escherichia coli* strains causing infections in patients with a reconstructed bladder. *Adv Exp Med Biol* **485**, 263–271.

Keegan, S. J., Graham, C., Neal, D. E., Blum-Oehler, G., N'Dow, J., Pearson, J. P. & Gally, D. L. (2003). Characterization of *Escherichia coli* strains causing urinary tract infections in patients with transposed intestinal segments. *J Urol* **169**, 2382–2387.

Kim, S. R. & Komano, T. (1997). The plasmid R64 thin pilus identified as a type IV pilus. *J Bacteriol* **179**, 3594–3603.

Kristjánsson, A., Bajc, M., Wallin, L., Willner, J. & Månsson, W. (1995a). Renal function up to 16 years after conduit (refluxing or anti-reflux anastomosis) or continent urinary diversion. 2. Renal scarring and location of bacteriuria. *Br J Urol* **76**, 546–550.

Kristjánsson, A., Wallin, L. & Månsson, W. (1995b). Renal function up to 16 years after conduit (refluxing or anti-reflux anastomosis) or continent urinary diversion. 1. Glomerular filtration rate and patency of uretero-intestinal anastomosis. *Br J Urol* **76**, 539–545.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12.

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L. & Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* **82**, 6955–6959.

Liu, W. M., Mei, R., Di, X., Ryder, T. B., Hubbell, E., Dee, S., Webster, T. A., Harrington, C. A., Ho, M. H. & other authors (2002). Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics* **18**, 1593–1599.

Lloyd, A. L., Rasko, D. A. & Mobley, H. L. (2007). Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. *J Bacteriol* **189**, 3532–3546.

Lloyd, A. L., Henderson, T. A., Vigil, P. D. & Mobley, H. L. (2009). Genomic islands of uropathogenic *Escherichia coli* contribute to virulence. *J Bacteriol* **191**, 3469–3481.

Mobley, H. L., Green, D. M., Trifillis, A. L., Johnson, D. E., Chippendale, G. R., Lockatell, C. V., Jones, B. D. & Warren, J. W. (1990). Pyelonephritogenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. *Infect Immun* **58**, 1281–1289.

Nataro, J. P., Deng, Y., Cookson, S., Cravioto, A., Savarino, S. J., Guers, L. D., Levine, M. M. & Tacket, C. O. (1995). Heterogeneity of enteroaggregative *Escherichia coli* virulence demonstrated in volunteers. *J Infect Dis* **171**, 465–468.

Nie, H., Yang, F., Zhang, X., Yang, J., Chen, L., Wang, J., Xiong, Z., Peng, J., Sun, L. & other authors (2006). Complete genome sequence of *Shigella flexneri* 5b and comparison with *Shigella flexneri* 2a. *BMC Genomics* **7**, 173.

Perna, N. T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J. & other authors (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157 : H7. *Nature* **409**, 529–533.

Rasko, D. A., Rosovitz, M. J., Myers, G. S., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R. & other authors (2008). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* **190**, 6881–6893.

Schlager, T. A., Whittam, T. S., Hendley, J. O., Wilson, R. A., Bhang, J., Grady, R. & Stapleton, A. (2000). Expression of virulence factors among *Escherichia coli* isolated from the periurethra and urine of children with neurogenic bladder on intermittent catheterization. *Pediatr Infect Dis J* **19**, 37–41.

Schlager, T. A., Johnson, J. R., Ouellette, L. M. & Whittam, T. S. (2008). *Escherichia coli* colonizing the neurogenic bladder are similar to widespread clones causing disease in patients with normal bladder function. *Spinal Cord* **46**, 633–638.

Shaaban, A. A., el-Nono, I. H., Abdel-Rahman, M. & Ghoneim, M. A. (1992a). The urodynamic characteristics of different ileal reservoirs: an experimental study in dogs. *J Urol* **147**, 197–200.

Shaaban, A. A., Gaballah, M. A., el-Diasty, T. A. & Ghoneim, M. A. (1992b). Urethral controlled bladder substitution: a comparison between the intussuscepted nipple valve and the technique of Le Duc as antireflux procedures. *J Urol* **148**, 1156–1161.

Skyberg, J. A., Johnson, T. J., Johnson, J. R., Clabots, C., Logue, C. M. & Nolan, L. K. (2006). Acquisition of avian pathogenic *Escherichia coli*

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02

plasmids by a commensal *E. coli* isolate enhances its abilities to kill chicken embryos, grow in human urine, and colonize the murine kidney. *Infect Immun* **74**, 6287–6292.

Studer, U. E., Casanova, G. A. & Zingg, E. J. (1991). Continent urinary diversion by an ileal bladder substitute. *Prog Clin Biol Res* **370**, 59–66.

Studer, U. E., Danuser, H., Hochreiter, W., Springer, J. P., Turner, W. H. & Zingg, E. J. (1996). Summary of 10 years' experience with an ileal low-pressure bladder substitute combined with an afferent tubular isoperistaltic segment. *World J Urol* **14**, 29–39.

Suriano, F., Gallucci, M., Flammia, G. P., Musco, S., Alcini, A., Imbalzano, G. & Dicuonzo, G. (2008). Bacteriuria in patients with an orthotopic ileal neobladder: urinary tract infection or asymptomatic bacteriuria? *BJU Int* **101**, 1576–1579.

Tettelin, H., Riley, D., Cattuto, C. & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* **11**, 472–477.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C. & other authors (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* **5**, e1000344.

Valentiner-Branth, P., Steinsland, H., Fischer, T. K., Perch, M., Scheutz, F., Dias, F., Aaby, P., Mølbak, K. & Sommerfelt, H. (2003). Cohort study of Guinean children: incidence, pathogenicity, conferred protection, and attributable risk for enteropathogens during the first 2 years of life. *J Clin Microbiol* **41**, 4238–4245.

Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G., III, Rose, D. J. & other authors (2003). Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* **71**, 2775–2786.

Welch, R. A., Burland, V., Plunkett, G., III, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A. & other authors (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**, 17020–17024.

Whittam, T. S., Ochman, H. & Selander, R. K. (1983). Geographic components of linkage disequilibrium in natural populations of *Escherichia coli*. *Mol Biol Evol* **1**, 67–83.

Whittam, T. S., Wolfe, M. L., Wachsmuth, I. K., Orskov, F., Orskov, I. & Wilson, R. A. (1993). Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect Immun* **61**, 1619–1629.

Wullt, B., Agace, W. & Mansson, W. (2004). Bladder, bowel and bugs – bacteriuria in patients with intestinal urinary diversion. *World J Urol* **22**, 186–195.

Zerbino, D. R. & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829.

Edited by: S. D. Bentley

Downloaded from www.sgmjournals.org by
IP: 93.91.26.97
On: Sat, 29 Aug 2015 07:19:02