



In Vivo mRNA Profiling of Uropathogenic *Escherichia coli* from Diverse Phylogroups Reveals Common and Group-Specific Gene Expression Profiles

Piotr Bielecki, Uthayakumar Muthukumarasamy, Denitsa Eckweiler, Agata Bielecka, Sarah Pohl, Ansgar Schanz, Ute Niemeyer, Tonio Oumeraci, Nils von Neuhoff, Jean-Marc Ghigo, et al.

► To cite this version:

Piotr Bielecki, Uthayakumar Muthukumarasamy, Denitsa Eckweiler, Agata Bielecka, Sarah Pohl, et al.. In Vivo mRNA Profiling of Uropathogenic *Escherichia coli* from Diverse Phylogroups Reveals Common and Group-Specific Gene Expression Profiles. *mBio*, 2014, 5 (4), pp.e01075-14. 10.1128/mBio.01075-14 . pasteur-02017377

HAL Id: pasteur-02017377

<https://pasteur.hal.science/pasteur-02017377>

Submitted on 13 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

In Vivo mRNA Profiling of Uropathogenic *Escherichia coli* from Diverse Phylogroups Reveals Common and Group-Specific Gene Expression Profiles

Piotr Bielecki,^{a,*} Uthayakumar Muthukumarasamy,^{a,b} Denitsa Eckweiler,^{a,b} Agata Bielecka,^{a,b} Sarah Pohl,^{a,b} Ansgar Schanz,^c Ute Niemeyer,^d Tonio Oumeraci,^e Nils von Neuhoff,^e Jean-Marc Ghigo,^f Susanne Häussler^{a,b}

Institute for Molecular Bacteriology, Twincore, Centre for Clinical and Experimental Infection Research, A Joint Venture of the Helmholtz Centre for Infection Research and the Hannover Medical School, Hannover, Germany^a; Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany^b; Private Practice, Salzgitter, Germany^c; Private Practice, Wolfenbuettel, Germany^d; Institute of Cell and Molecular Pathology, Hannover Medical School (MHH), Hannover, Germany^e; Institute Pasteur, Unité de Génétique des Biofilms, Paris, France^f

* Present address: Piotr Bielecki, Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut, USA.

P.B. and U.M. contributed equally to this work.

ABSTRACT mRNA profiling of pathogens during the course of human infections gives detailed information on the expression levels of relevant genes that drive pathogenicity and adaptation and at the same time allows for the delineation of phylogenetic relatedness of pathogens that cause specific diseases. In this study, we used mRNA sequencing to acquire information on the expression of *Escherichia coli* pathogenicity genes during urinary tract infections (UTI) in humans and to assign the UTI-associated *E. coli* isolates to different phylogenetic groups. Whereas the *in vivo* gene expression profiles of the majority of genes were conserved among 21 *E. coli* strains in the urine of elderly patients suffering from an acute UTI, the specific gene expression profiles of the flexible genomes was diverse and reflected phylogenetic relationships. Furthermore, genes transcribed *in vivo* relative to laboratory media included well-described virulence factors, small regulatory RNAs, as well as genes not previously linked to bacterial virulence. Knowledge on relevant transcriptional responses that drive pathogenicity and adaptation of isolates to the human host might lead to the introduction of a virulence typing strategy into clinical microbiology, potentially facilitating management and prevention of the disease.

IMPORTANCE Urinary tract infections (UTI) are very common; at least half of all women experience UTI, most of which are caused by pathogenic *Escherichia coli* strains. In this study, we applied massive parallel cDNA sequencing (RNA-seq) to provide unbiased, deep, and accurate insight into the nature and the dimension of the uropathogenic *E. coli* gene expression profile during an acute UTI within the human host. This work was undertaken to identify key players in physiological adaptation processes and, hence, potential targets for new infection prevention and therapy interventions specifically aimed at sabotaging bacterial adaptation to the human host.

Received 17 March 2014 Accepted 10 June 2014 Published 5 August 2014

Citation Bielecki P, Muthukumarasamy U, Eckweiler D, Bielecka A, Pohl S, Schanz A, Niemeyer U, Oumeraci T, von Neuhoff N, Ghigo J-M, Häussler S. 2014. *In vivo* mRNA profiling of uropathogenic *Escherichia coli* from diverse phylogroups reveals common and group-specific gene expression profiles. mBio 5(4):e01075-14. doi:10.1128/mBio.01075-14.

Editor Scott Hultgren, Washington University School of Medicine

Copyright © 2014 Bielecki et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Susanne Häussler, susanne.haeussler@twincore.de.

To successfully thrive in the host environment during the course of an infection, pathogens have to rapidly adapt to the specific conditions encountered. Thereby, a key to understanding microbial pathogenesis lies in knowledge of which genes are expressed to initiate and maintain the infection and of the global impact of the host environment on the transcriptional profile of the pathogen (1). Urinary tract infections (UTI) are one of the most common bacterial infections worldwide, and most of them (over 80%) are caused by uropathogenic *Escherichia coli* (UPEC) (2). It is widely accepted that UPEC strains originate from the distal gut microbiota where they mostly behave as commensals (3), although UPEC strains are armed with extra virulence genes (4). Those virulence genes are often present on strain-specific pathogenicity islands (PAIs), which are clusters of virulence-

related genes (5–7). PAIs are diverse in content and genome location and, as more sequence information of more examples of the islands accumulates, greater insights into their role in disease can be expected (8, 9).

UTI is recognized as presence of the bacteria in urine (bacteriuria). During the course of infection, bacterial cells are attaching to human epithelial cells, utilizing chaperone usher (CU) fimbriae that contain adhesins on their tips (10). The prototypical CU type I fimbriae adhesion can lead to intracellular invasion of bladder epithelial cells (11). UPEC strains are known to enter the cytoplasm and form biofilm-like structures called intracellular bacterial communities (IBC) (12). After maturation of IBC, the UPEC cells can disperse into urine, or as part of the host response the infected epithelial cells may be exfoliated and released into urine.

Exfoliated cells are replaced with transition epithelial cells, which may be as well invaded by UPEC, where it forms quiescent intracellular reservoirs (QIR) characterized by their persistence and antibiotic resistance (13).

In vitro studies and various animal models have been valuable for exploring UPEC pathogenesis (14, 15) and have led to significant advances in understanding key pathogenicity mechanisms (16–22). Knowledge of UPEC gene expression during naturally occurring UTI will further add to the full understanding of microbial pathogenesis of this widespread bacterial pathogen. Indeed, investigation of complex transcriptional adaptation processes of UPECs to the human host is expected to uncover key regulatory components and to provide unique insight into bacterial pathogenicity (23). Furthermore, the identification of *E. coli* virulence genes associated with UTIs is potentially valuable in differentiating UPEC from nonuropathogenic *E. coli* and might lead to the introduction of virulence typing strategies into clinical microbiology.

Today, the significant advances in next-generation sequencing technologies enable unbiased and very accurate quantitative annotation-independent detection of transcripts at high resolution (24). Furthermore, RNA sequencing (RNA-seq) can be used to extract genotype information from the cDNA on a single-nucleotide resolution level, providing profound insights into phylogenetic relatedness. Although RNA sequencing studies have been widely used for quantitative and qualitative transcriptional profiling of various bacterial pathogens (24–30), the application of RNA-seq to determine global transcriptional profiles during the infection of the human host has remained very limited.

In this study, we used strand-specific RNA-Seq to generate comprehensive *in vivo* transcriptional profiles of 21 UPEC stains causing symptomatic UTI in a cohort of elderly patients and gained profound insights into the conservation/variation of transcription patterns across UPEC isolates that exhibited a broad phylogenetic distribution. While most known UPEC virulence factors could be identified, comparison of the *in vivo* transcriptional profiles uncovered a set of genes that is specifically transcribed during the course of an infection and which cannot be inferred from analyzing genomes or from transcriptional profiles of UPEC isolates recorded under laboratory culture conditions.

RESULTS

Broad phylogenetic distribution of *E. coli* UTI isolates isolated from elderly patients. With the aim to record *in vivo* transcriptional profiles of UPEC stains, urine samples were collected from outpatients with symptomatic UTI prior to antibiotic treatment. Overall, 21 urine samples were included in this study. All of them were culture positive on MacConkey agar plates, with more than 10^6 *E. coli* CFU/ml urine in pure cultures, and microscopic inspection of urine sediments revealed the presence of massive numbers of neutrophils ($>100/\mu\text{l}$). The 21 patients were mainly elderly (mean age above 60 years, with only 4 patients being younger than 60 years), 8 were male, and 13 were female. RNA isolation procedures and strand-specific Illumina-based RNA sequencing of bacterial mRNA were performed, and the raw sequence output after the removal of reads that mapped to the human genome consisted of 61.01 million reads. Thus, on average, 2.9 million reads were retrieved from each of the 21 samples. In accordance with the finding that the gene content between pairs of *E. coli* genomes may

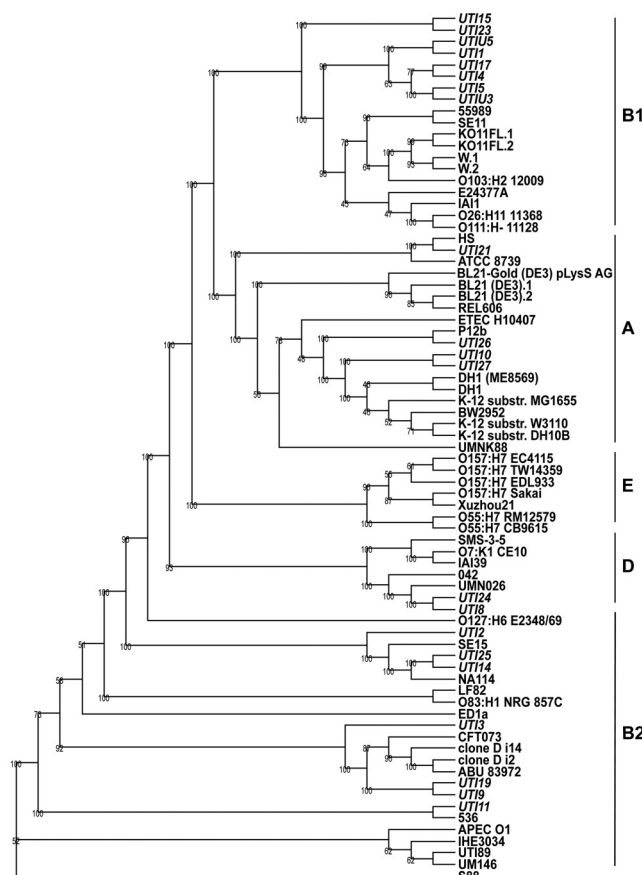


FIG 1 Phylogenetic tree of 54 previously sequenced strains and the 21 clinical isolates from this (in italic) work based on sequence variation within 336 genes. Phylogenetic groups are indicated based on previous reports (34, 35). The numbers show the bootstrapping values as provided by RaxML.

diverge by more than 30%, the range of gene numbers to which those reads mapped was between 3,848 and 4,972.

In *E. coli*, $<3\%$ of nucleotide divergence is found among conserved genes in the various genomes (6). This high degree of homogeneity allows the establishment of phylograms that are built upon sequence variations. Previous studies have identified five major phylogenetic groups, (B2, B1, D, A, and E), corresponding to *E. coli* strains with distinct capability to cause disease and to inhabit various ecological niches (31–36). Figure 1 depicts the phylogenetic distribution of previously sequenced *E. coli* isolates that have been grouped into the five phylogenetic *E. coli* groups. This tree is based on sequence variations of 336 genes (for those genes, at least 80% sequencing coverage across the 21 UTI isolates was detected), which allowed us to use the genotype information from the RNA-seq data of the *E. coli* genomes to assign the 21 UTI isolates of this study to the clusters within the phylogenetic tree (Fig. 1). Reflecting the fact that our study group consisted mostly of elderly patients, we found a broad distribution of the 21 UTI-associated isolates between the phylogenetic groups. A total of 43% of the 21 isolates belong to the virulent *E. coli* strain phylogroups B2 and D (B2, 33%; D, 10%), whereas the others are distributed in the B1 (38%) and A (19%) phylogroups.

Commonly transcribed genes of the *E. coli* UTI isolates exhibit a conserved expression profile. With the aim to uncover the

full extent of the *in vivo* gene expression profile of the 21 clinical *E. coli* isolates, we mapped all obtained Illumina sequencing reads to a list of 12,331 nonredundant *E. coli* genes. This list of genes was generated by the comparative genomic analysis of 54 previously fully sequenced *E. coli* genomes (see Materials and Methods). The entire list, including ortholog identifiers (IDs) as well as the expression values of the 21 UTI samples, is provided in Data Set S1 in the supplemental material. This list includes 2,129 genes shared by all 54 strains and 10,202 genes that are absent in at least one of the 54 strains. Among the latter, 3,257 genes were found in only one of the 54 published genomes as singletons. Only very few genes having homologs in all 54 sequenced *E. coli* isolates were not transcribed in any of the 21 isolates under *in vivo* conditions, indicating that expression of most of the core genome is relevant for bacterial replication in the human urinary tract. Furthermore, we found a large set of overall 2,589 genes that were commonly transcribed in all isolates during *in vivo* conditions, which—depending on the genome size of the isolates—accounts for 52% to 67% of all transcribed genes within one isolate. As depicted in Fig. S1 in the supplemental material, those commonly expressed 2,589 genes appear to be unregulated or constitutively expressed, as the overall variation of the expression profiles among the isolates was low and the genes were expressed at a generally high level independently of their phylogenetic group specificity. As expected, many of these genes correspond to genes required for the maintenance of basic cellular functions, such as DNA repair, ATP synthesis, aminosugar metabolism, and protein transport (see Table S1 in the supplemental material).

Since we found only a low variation in the expression levels of the genes commonly transcribed in all 21 *E. coli* isolates at the time of mRNA sampling, hierarchical clustering based on their transcriptional profiles did not reveal specific and distinct clusters. We also performed matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry biotyping to elucidate whether protein fingerprints might uncover clusters that serve for the identification of phylogenetic relatedness. MALDI-TOF mass spectrometry (see Fig. S2) correctly classified our UTI *E. coli* isolates on the species level. However, a dendrogram based on Minkowski distances and group averages did not reveal distinct subgroups within our isolates that would correlate to the previously identified phylogenetic groups B2, B1, A, and D. This may reflect the fact that MALDI-TOF mass spectrometry covers mostly housekeeping proteins, e.g., the ribosomal proteins, and therefore is ill suited to discriminate phylogenetic relationships.

The *in vivo* gene expression profile of the *E. coli* UTI isolates correlates with phylogenetic group clustering. Mapping of all obtained Illumina sequencing reads to the list of 12,331 nonredundant *E. coli* genes revealed—apart from the 2,589 commonly transcribed genes (see above)—a large fraction of genes (6,305 genes) that were expressed in at least one of the 21 UTI strains (see Data Set S1).

Remarkably, clustering of the *in vivo* transcripts based on principal component analysis (PCA) of the 21 UTI isolates (Fig. 2), including commonly transcribed genes as well as those of the flexible genome, compared very well to that of phylogenetic clustering based on the single nucleotide polymorphism (SNP) profile (Fig. 1). The expression profile of the 21 UTI samples clustered into three main groups that represented the B2, D, and A/B1 phylogenetic groups. Of note, clustering became even more accurate and well separated when only the expression of genes of the flex-

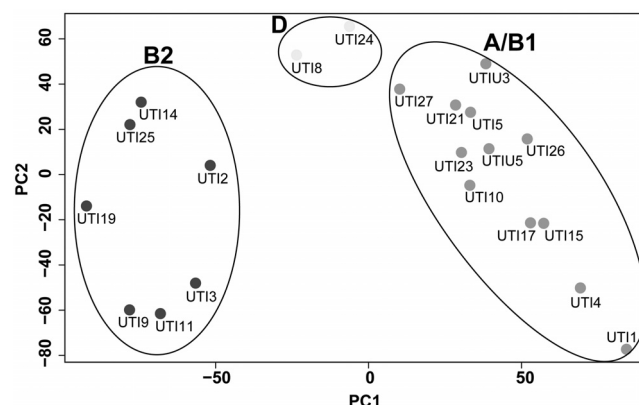


FIG 2 Clustering of the *in vivo* transcripts of the 21 UTI isolates based on principal component analysis (PCA). Clustering clearly reflects phylogenetic relatedness as the clinical isolates grouped according to their affiliation to the B2, D, and A/B1 phylogroups.

ible genomes was included in the analysis (data not shown). These results are in agreement with previous reports (37, 38) and clearly demonstrate that the presence of group-specific gene repertoires, and not a difference in overall gene expression profiles, impacts on clustering of the UTI isolates into the phylogroups.

We also performed a *de novo* assembly of reads from the 21 isolates that did not match any of the 54 sequenced genomes, which resulted in the identification of 158 potential genes, 48 of which are organized in operon structures. A total of 105 of the genes have homologs in *E. coli*, and 53 have homologs in other *Enterobacteriaceae* (see Table S2).

***In vivo* mRNA expression profiling of known UPEC virulence factors.** Many of the genes found to be expressed *in vivo* in the 21 UTI isolates included known key *E. coli* virulence factors. Although we sampled voided bacteria, which are clearly distinct from attached and biofilm-grown bacteria, genes responsible for adhesion to the uroepithelium, e.g., type I fimbriae (*fim*) (16), P fimbriae (*pap*), F1C/S fimbriae (*foc* and *sfa*), were found (Table 1). However, we did not find a uniform expression of any of those common adhesion-related genes. Whereas no or only very low expression of *fimA*, whose expression has been demonstrated to enhance *E. coli* virulence in the urinary tract (16), could be detected in 13 UPEC isolates in this study, the *fimA* gene and the subsequent operon was highly expressed in 8 isolates. Additionally, P fimbriae and F1C/S fimbriae-encoding genes were expressed in a subset of isolates (5 and 3 isolates, respectively). Interestingly, F1C/S fimbriae gene expression was exclusively found in isolates which grouped to the phylogenetic B2 cluster.

Genes encoding iron acquisition systems were found to be widely expressed *in vivo*. The enterobactin and its transport system-encoding genes (*ent* and *fep*) were expressed in all UTI isolates without exception, whereas expression of aerobactin (*iuc*), yersiniabactin (*irp*), and salmochelin (*iro*) genes was less uniform. Expression of the heme-mediated iron acquisition system (*chu*) was present in 100% of isolates clustering with the D and B2 phylogenetic groups and in 75% of the isolates clustering with group A but absent in those clustering with group B1. Capsular polysaccharide expression was observed in 100% of isolates clustering with group D and B2, in 50% clustering with group A, and only partially in one isolate of group B1. The expression of extracellular

TABLE 1 UPEC virulence genes present in the 21 clinical isolates

	Result by phylogenetic group and UTI isolate no. ^c																				
	D		A				B1							B2							
	24	8	10	27	26	21	23	15	1	5	U3	U5	4	17	11	3	9	19	14	25	2
Adhesion																					
<i>fim</i>	++	++	++	++	—	—	+	+	+	+	++	++	++	+	++	+	+	+	+	+	+
<i>pap</i>	+	—	—	+	—	—	—	—	—	—	—	—	—	—	+	++	—	++	—	—	—
<i>FIC/S</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	+	—	++	+	—	—	—
Iron acquisition																					
<i>ent</i>	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
<i>fep</i>	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
<i>iuc</i>	+	—	—	—	++	++	++	++	+	++	++	—	++	—	++	++	—	++	++	++	—
<i>iro</i>	—	—	+	—	—	++	++	++	++	++	++	—	—	—	++	—	++	++	—	—	—
<i>irp</i>	++	—	—	—	—	++	+	—	—	++	++	—	—	++	++	+	++	++	+	++	++
<i>chu</i>	++	++	+	+	—	++	—	—	—	—	—	—	—	—	++	++	++	++	++	++	+
Capsule																					
<i>kps</i>	++	++	—	++	—	+	—	—	—	—	—	—	—	+	++	++	++	++	++	++	++
Toxins																					
<i>cnf1</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	++	—	++	—	—	—	—
<i>hlyA</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	++	—	++	++	—	—	—
<i>picU</i>	—	—	—	—	—	—	—	—	—	—	—	—	++	—	—	—	++	++	+	++	—
<i>sat</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	++	—	++	++	++	—
<i>vat</i>	—	—	+	—	—	++	—	—	—	—	++	—	—	—	++	++	++	++	—	—	++
<i>usp</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	++	++	++	++	++	++	++
<i>pks^a</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	++	—	++	++	—	—	—
<i>colV^b</i>	—	—	—	—	—	++	—	++	—	++	++	—	—	—	++	—	—	—	—	—	—

^a *pks* island genes *c2471* to *c2451*.
^b Colicin V transport genes *cvaAB* deduced from *de novo* analysis (see Table S2 in the supplemental material).
^c The presence or absence of the gene/operon is indicated as follows: -, no reads detected (nRPK value from 0 to 1.5); +, reads detected with low values (nRPK from 1.5 to 2.0) or partial operons were detected; ++, genes with nRPK values of >2.

toxin-encoding genes *in vivo* was less frequent. Overall extracellular toxin expression was most frequent in those UTI isolates that clustered with strains from the B2 phylogenetic group. *cnf1* and *hlyA* expression was observed in only two and three isolates, respectively. Expression of genes encoding serine protease auto-transporter PicU was present in five isolates, whereas the gene encoding the serine protease autotransporter Sat was expressed in four isolates, all of them clustering with the B2 phylogenetic group. The vacuolating autotransporter toxin-encoding gene *vat* was expressed in 8 isolates, and *usp* gene expression was found exclusively in isolates that clustered with group B2 isolates. The genes encoding the transport system of colicin V were detected in 5 isolates, and the *clb* operon encoded on the *pks* island (39) was observed to be expressed in 3 isolates, again clustering with strains of the B2 group.

In vivo expression profiling of small regulatory RNAs. RNA-seq profiling enabled us to investigate expression of small regulatory RNAs (sRNAs), which have been assigned central roles in virulence and environmental fitness (40, 41). Eleven sRNAs were identified that exhibited high *in vivo* expression levels in all or most of the 21 clinical UTI isolates (Fig. 3). Among them, *ryiA* (*glmZ*) and *csrB* exhibited the highest *in vivo* expression levels. The sRNA RyiA (*GlmZ*) activates *glmS* expression. *GlmS* synthesizes glucosamine-6-phosphate (GlcN-6-P) and thus delivers precursor molecules for the biosynthesis of peptidoglycan and lipopolysaccharides (LPS), which are essential elements of the Gram-negative bacterial cell wall (42, 43). Another sRNA that was found to be highly expressed during UTI was *csrB*. Both sRNAs *csrB* and *csrC* are modulatory components of the carbon storage regulatory (Csr) network. They contain multiple

CsrA binding sites, which permit them to sequester and antagonize CsrA, a pleiotropic regulator of carbon metabolism (44–46). Transcription of these two small RNAs is regulated by the BarA/UvrY two-component signal transduction system (TCS) in *E. coli* or by homologous systems such as GacS/GacA in other bacteria (47). The Csr system (or the homolog RsmA/RsmZ) is present in many eubacteria and is known to be involved in mediating adaptive physiology, timed virulence trait expression in animal pathogens (48, 49), and biofilm formation (50, 51). Recently, it was shown to interact with the stringent response regulatory system (52). Although other sRNA were also identified to be expressed during *in vivo* growth, their overall expression levels were often lower than those observed in four representatives of our clinical isolates that were cultivated *in vitro* under rich medium conditions until late exponential growth phase. Apparently, a large number of those highly *in vitro*-expressed sRNAs serve the adaptation to stationary phase of growth (Fig. 3). Among those, we found *micA*, a negative regulator of *ompA* (53), *ryhA* (*arcZ*), and *rprA*, encoding sRNAs that increase the translation of the stationary sigma factor RpoS (54, 55). Their expression, as well as the expression of *sroC* and *ryeB*, has been associated with stationary phase of growth (56, 57). Additionally, the products of *rprA*, *isrA* (*mcsA*), and *omrA* were strongly expressed. Those sRNAs have been shown to negatively regulate the translation of CsgD, the major transcriptional regulator of *E. coli* curli biosynthesis (58, 59).

Identification of infection-relevant gene expression profiles in the *E. coli* UTI isolates. In addition to known virulence factors, we aimed at identifying infection-relevant genes that are commonly expressed in UPEC isolates *in vivo*. We therefore cultivated

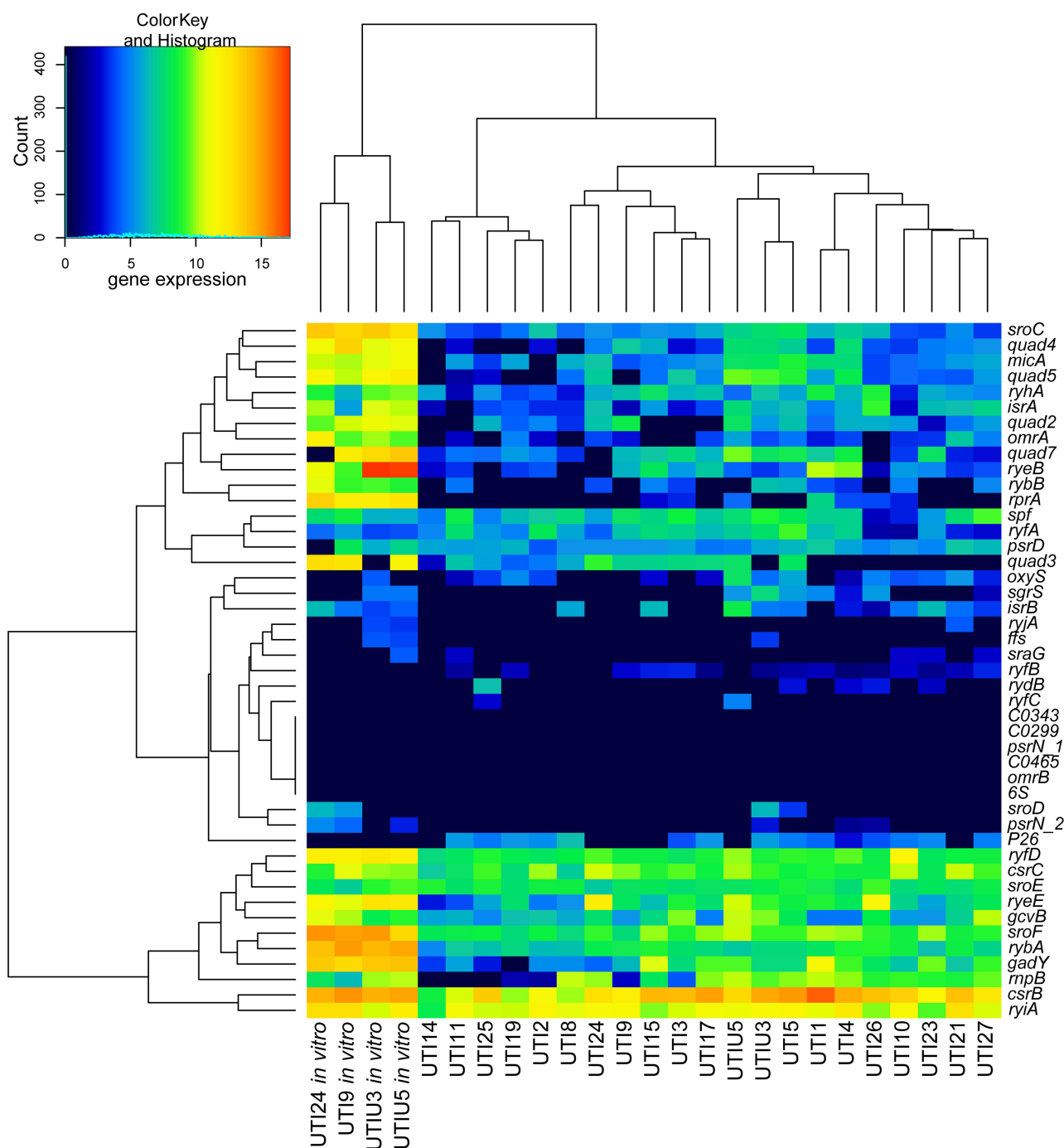


FIG 3 Expression profile of the small regulatory RNAs in the 21 UTI isolates and the 4 UTI isolates cultivated *in vitro*. The genes (vertical) are hierarchically clustered using Pearson distances, and the samples (horizontal) are clustered according to Spearman rank correlation. The histogram describes the correlation of the color to the nRPK value of absolute expression.

4 of the UTI isolates (isolates UT103 and UT105 clustering with phylogroup B1, UT124 clustering with group D, and UT19 clustering with group B2) *in vitro* under rich medium conditions and recorded the transcriptional profiles. A total of 202 genes were found to be upregulated under *in vivo* conditions in the 4 strains, and all of those genes have been demonstrated to be expressed in all 21 UTI isolates in this study under *in vivo* conditions. A detailed

list of the 202 commonly and exclusively *in vivo* expressed genes is provided in Table S3 in the supplemental material (detailed data on all differentially expressed genes is given in Data Set S2A [up-regulated genes] and S2B [down-regulated genes] in the supplemental material). Whereas only 23 hypothetical or conserved hypothetical genes were found, use of the systematic functional annotation provided by Gene Ontology revealed that 20% of the

genes belonged to functional groups involved in general biological processes such as ATP synthesis and catabolic processes as well as transcription, translation, and DNA replication and repair. Furthermore, many genes were found to be involved in rRNA and tRNA processing, indicating that the bacteria rapidly grow in the human urinary tract. Consistent with the fact that main carbon sources for *E. coli* during UTI are peptides and amino acids, genes that belong to biological processes of proteolysis, protein transporters, carbohydrate metabolism, and fatty acid biosynthesis were represented. We also found that genes encoding enzymes of the pyruvate dehydrogenase complex were highly expressed. Another large group of genes commonly expressed *in vivo* were genes involved in the regulation of bacterial cell shape and in bacterial stress responses, such as responses to toxic substances, including antibiotics. Furthermore, we found an *in vivo* overexpression of *ampG* encoding a peptidoglycan permease, which was shown to be involved in evasion of the host innate immune system during UTI (60). We also found *gidA* encoding the tRNA uridine 5-carboxymethylaminomethyl modification enzyme (61) among the commonly *in vivo* expressed genes. *gidA* is known to impact on the posttranscriptional level on a number of virulence factors in *Pseudomonas syringae* (62), *Aeromonas hydrophila* (63), *Shigella flexneri* (64), *Streptococcus suis* (65), *Streptococcus pyogenes* (66), *Salmonella enterica* serovar Typhimurium (67), and *E. coli* (68). Of note, the *gidA* gene was also shown to be upregulated in the majority of patients' samples from a previous study on UPEC transcriptomics (69) and in an earlier murine UTI *in vivo* gene expression study (14). Overall, our transcriptional data are remarkably consistent with those previous reports (14, 69) and with the transcriptional profile of *E. coli* isolated from patients with asymptomatic bacteriuria (ABU) (70). Expression of genes involved in nitrate/nitrite metabolism and nitric oxide (NO) protection, upregulation of iron acquisition systems, and genes involved in carbohydrate and amino acid metabolism were commonly observed (14, 69, 70), reflecting bacterial adaptation to the growth conditions encountered in the environment of the urinary tract. Interestingly, for 3 (*carA*, *carB*, and *argC*) out of the 202 commonly highly expressed *in vivo* genes in our study, it was shown that their inactivation poses a competitive disadvantage to the respective mutants in the mouse urinary tract (71). These results clearly suggest that their expression is crucial for growth in the urinary tract.

Identification of genes that are exclusively expressed in the *E. coli* B1 and A phylogenetic groups. To evaluate whether the UTI-associated isolates that group with B1 and A express a distinct set of genes potentially relevant for the infection process, we extracted from the list of genes that were found to be differentially regulated among the 21 UPEC isolates those that were specifically expressed in the 12 phylogroup A/B1 isolates. We identified 142 genes that were expressed at a significantly higher level in the 12 phylogroup A/B1 isolates (see Fig. S3 and Table S4A in the supplemental material), compared to all other 9 isolates. Interestingly, 27 (19%) of these genes were associated with utilization of alternative carbon sources, with in particular the complete set of the 12 genes required for phenylalanine degradation into succinyl coenzyme A (CoA) (*tynA*, *feaB*, *paaKEACBGZJFH*), indicating that those isolates have access to sufficient amounts of phenylalanine in the urine. Of note, mutations in *aroA* have been used to construct attenuated strains of various Gram-negative bacteria, including *E. coli* (72). Thereby, the attenuation is due to the in-

ability of *aroA* mutants to synthesize chorismate, which is a precursor of important biochemical intermediates such as indole and aromatic amino acids, many alkaloids, and other aromatic metabolites, as well as folate and 2,3-dihydroxybenzoic acid used for enterobactin biosynthesis. The availability of aromatic amino acids in the urine may not only enable *E. coli* growth on 2-phenylalanine but also may save chorismate for iron chelator biosynthesis as a crucial virulence trait. Interestingly, among the *in vivo* expressed genes that were found to be enriched in the 12 phylogroup A/B1 isolates, we also found *iroC* and *iroD* involved in transport and procession of the siderophore salmochelin. *iroC* was also upregulated *in vivo* compared to LB cultures in one of two isolates, clustering with group B1, for which an *in vitro* transcriptional profile was recorded. These results indicate that the siderophore may play an important role in iron acquisition within the subgroup of UPEC isolates that cluster in the A/B1 phylogroup and that lack the common UPEC-associated virulence gene expression.

We could also detect 13 (9%) genes encoding fimbrial adhesins mostly described as functional but cryptic, including the *ycb* operon (*ycbRSTUVF*), part of the *yra* operon (*yraH*, *yraJ*, and *yraK*) (73, 74), and genes encoding a CS1-type fimbrial structure (10CE_3624, -25, -26, and -27) that is usually associated with enterotoxigenic *E. coli* (75). The enrichment in fimbriae genes in the group of the 12 studied A/B1 isolates could reflect a characteristic increased adhesion capability (76).

We also observed expression of Rhs element genes. Many bacteria contain all or part of 5 Rhs elements: RhsA, -B, -C, -D, and -E, scattered around the chromosome. Each Rhs region contains a 3.7-kb GC-rich DNA sequence that is 99% identical from one element to another. These high-identity levels between Rhs proteins was proposed to mediate major intraspecies chromosomal rearrangements, hence their name (which stands for "recombination hot spot") (77). However, high conservation of intact *rhs* main genes (*rhsA*, -B, -C, -D, -E) also suggested that they could contribute to a function subjected to selective pressure (78). Intriguingly, *rhs* genes are not expressed to a detectable extent during routine cultivation, and the conditions leading to Rhs expression have not yet been elucidated (79). Our mRNA expression analysis demonstrates that some Rhs elements are specifically expressed *in vivo* in all 12 UPEC isolates belonging to the B1/A phylogroup. Of note, recent studies suggested that expression of Rhs elements are associated with bacterium-host or bacterium-bacterium interactions, suggesting that such functions could contribute to UTI (80, 81). Their expression has furthermore been associated with toxin-antitoxin (TA) activity and to be potentially delivered through a type 6 secretion apparatus delivering effectors both in prokaryotic and eukaryotic prey cells (82, 83). Some TA systems have recently been shown to be important for colonization of the bladder (*yefM-yoeB* and *tomB-hha*) and survival within the kidneys (*pasTI*, previously named *yjfGF*) in a murine UTI model (84). In this study, the *chpAR*, *yafQ-dinJ*, and *hicBA* TA systems were found to be highly expressed *in vivo*, specifically in the isolates clustering with phylogroups A/B1.

Identification of genes that are exclusively expressed in the *E. coli* from the B2 phylogenetic group. Besides identification of genes specifically expressed in the isolates clustering with the B1 and A phylogroups, we also identified 389 genes that were specifically expressed in strains clustering with the B2 phylogroup. A total of 208 out of the 389 genes encode hypothetical or conserved

hypothetical proteins, and 102 are annotated as encoding putative proteins (see Table S4B in the supplemental material). Apart from the well-described virulence genes, such as *sat*, encoding the secreted autotransporter toxin (85), or *usp*, encoding the uropathogenic-specific protein (86), as well as *yadC*, *yadN*, and *yfcPQU*, encoding putative fimbria-like proteins (73, 87), we found a large number of genes encoding transporter and secretion systems. We found genes encoding components of type II general secretion pathways *yheBDK* and *hofDFGHIK*, also annotated as *gsp* genes in the *gspC-O* operon involved in secretion of endochitinase *yheB* (*chiA*) (88). Secreted chitinase is increasingly recognized as a virulence factor of pathogenic bacteria infecting mammal host (89). We also found that components of the hypothetical type VI secretion pathway (encoded by APECO1_3694, 3695, 3696, 3698, 3702, 3705, 3711, and 3712, *E. coli* APEC O1:K1:K7 gene IDs) were expressed *in vivo*. Furthermore, a large group of genes encoding various transport systems, like *yjcTU* encoding a D-allose ABC transporter and a putative iron compound ABC transporter encoded by APECO1_3384 to APECO1_3389, as well as a B2 phylogroup-specific expression of phosphotransferase systems (PTS) responsible for transport of sugars into the bacterial cell, were identified. In contrast to the isolates clustering with the A/B1 phylogroups that exhibited extensive upregulation of the phenylalanine degradation pathway, isolates clustering with the B2 group seem to use various sugars as main carbon and energy sources.

DISCUSSION

A key to understand microbial pathogenesis is to unravel how the host environment impacts on the global gene expression pattern of a pathogen and to identify the gene repertoire whose expression is essential for the initiation and maintenance of an infection. In this study, we applied massive parallel cDNA sequencing (RNA-seq) to provide unbiased, deep, and accurate insight into the nature and the dimension of the uropathogenic *E. coli* gene expression profile during an acute infection within the human host measured on bacteria present in voided urine. It is essential to indicate here that complex bacterial communities are present in the course of infection. In the current sampling procedure, we analyzed mainly planktonic bacteria, probably mixed with IBC from exfoliated epithelial bladder cells. It is possible that transcription profiling of selected adhesive cell population or IBC only would result in different gene expression results.

With a total of 21 *in vivo* transcriptomes, this study includes a large number of bacterial strains studied in respect to pathogenic *E. coli* gene expression following naturally occurring symptomatic human UTI. We applied RNA-seq to detect global transcriptional profiles independent of genome annotations and analyzed the *in vivo* transcriptomes to their full extent, including flexible genomic elements and expression of small regulatory RNAs. Furthermore, we identified single nucleotide polymorphisms (SNPs) in the bacterial isolates and used their cumulative differences to provide a large number of discriminators. These discriminators represent typing markers to distinguish bacterial isolates and to group the 21 UPEC isolates to one of the four main phylogenetic groups, A, B1, B2, and D.

Our findings on gene expression profiles in the urine of patients suffering from a UTI are generally consistent with data generated using murine models and a previous array-based transcriptome study of gene expression during a human UTI (14, 69, 70).

When comparing the *in vivo* gene expression profiles to those recorded under laboratory medium conditions, we found that *E. coli* adapts to the conditions encountered within the human host by expressing genes required for rapid replication, acquisition of iron, attachment to the uroepithel, and evasion of the immune system, while variably expressing virulence genes. Analysis of sRNA expression revealed consistent expression of sRNA involved in cell wall biosynthesis and integration of membrane proteins (*glmZY*) (42, 43) and in mediating adaptive physiology and timed virulence (*csrBC*) (44–46, 48, 49), underpinning the role of sRNAs in bacterial adaptation processes.

Although it is widely accepted that UPEC strains originate from the distal gut microbiota, they seem to be capable of colonizing the urinary tract and to cause symptomatic infections of cystitis and pyelonephritis, because they are armed with extra virulence genes that distinguish them from *E. coli* commensals (4). Several studies have demonstrated that the phylogroups differ in respect to the presence of virulence factors and ecological niches, and UPEC isolates have previously been found to be more prevalent in group B2. In line with this, we found 7 UPEC isolates that grouped with the B2 phylogenetic group, and they expressed several virulence genes *in vivo* that have been associated with UPEC strains exhibiting full-pathogenic potential. Nevertheless, and in accordance with previous studies on atypical UTI patient populations (90–92), in our study, which was performed on samples collected mainly from elderly patients, as many of 12 out of the 21 UPEC isolates analyzed were assigned to the A and B1 phylogenetic groups, which predominate among commensal *E. coli*.

We found that *E. coli* isolates that have been assigned to the four phylogroups share a large general gene expression profile, overall 2,589 genes were commonly transcribed in all isolates during the *in vivo* conditions, which—depending on the genome size of the isolates—accounts for 52% to 67% of the transcribed genome of the individual isolates. This conservation of a large part of the genome expression might account also for the finding that MALDI-TOF mass spectrometry, which probably corresponds to more- or less-conserved housekeeping proteins, does not allow a robust discrimination into the previously identified phylogenetic groups B2B1, A, and D. Although the 21 isolates share a large general gene expression profile, they do express clearly distinct flexible genomes. We found a strong correlation between the *E. coli in vivo* expression of the flexible genome and the genetic background of the isolate. However, as has been described before (37, 38), this correlation was dependent on the acquisition of group-specific gene repertoires in the flexible genomes rather than on a difference in their expression profile, possibly reflecting their evolution in distinct niches.

Not only did our study identify previously described virulence-associated genes that were exclusively expressed in the 7 UPEC isolates clustering with group B2, but we also identified a novel set of genes overrepresented in those isolates. Among those, we found a large number of genes encoding transporter and secretion systems, indicating that they play a role in pathogenicity of B2 group isolates. Furthermore, we identified a set of 142 genes whose expression was demonstrated to be specifically enriched in the 12 isolates that clustered with the A/B1 phylogroups, including genes encoding phenylalanine degradation pathway, a siderophore, fimbrial adhesins, and Rhs elements.

As more examples of *in vivo* transcriptional profiles accumulate, greater insights into the role of new genes involved in micro-

bial pathogenicity can be expected. However, further investigations are required to unravel the specific impact of novel virulence-determining factors in the establishment and maintenance of the disease. Thereby, the application of *in vivo* RNA-seq seems to be particularly appropriate, as it affords detailed quantitative and qualitative sequence information that is independent of genome annotations and thus allows the establishment of full transcriptional profiles, including flexible genomic elements and expression of small regulatory RNAs. Furthermore, knowledge of SNPs as identified by the use of RNA-seq enables highly resolving phylogenetic grouping of clinical isolates and thus provides a basis for further global phenotypic-genotypic correlation studies.

MATERIALS AND METHODS

Ethical statement. Urine samples were collected from 21 outpatients with symptomatic urinary tract infections and subjected to bacterial RNA extraction procedures. Samples were collected according to the standards of the Declaration of Helsinki. The sample provided for this research was subtracted from the samples collected for routine microbiological tests, which are made on a regular basis; therefore, no additional procedures were carried out on the patients. Samples were analyzed upon informed consent from the patients.

Bacterial RNA extraction and Illumina-based RNA sequencing. Urine samples (approximately 20 ml) were mixed with RNAprotect reagent (Qiagen), incubated for 15 to 30 min at room temperature, and centrifuged for 15 min at $4,000 \times g$ at 4°C , and the pellet was frozen at -70°C . RNA isolation was performed using the RNeasy minikit (Qiagen) according to the manufacturer's instruction with some modifications, and the DNA was removed by the use of a DNA-free kit (Ambion). Enrichment for bacterial RNA was achieved by using the MicroBEnrich kit (Ambion) according to the manufacturer's instructions.

Four UTI-associated isolates were also cultured *in vitro* in LB medium. RNAprotect reagent (Qiagen) was added to 3 ml of LB culture following growth to late exponential phase. All of the samples were treated for bacterial RNA enrichment. After depletion of rRNA from the samples, total RNA was subjected to a commercial capture and depletion system (MICROBExpress bacterial RNA enrichment kit; Ambion), strand-specific bar-coded cDNA libraries were generated as described (51), and all samples were single-end sequenced on an Illumina GenomeAnalyzer-IIx at a 36-bp read length. Traces of human reads were removed from the raw sequence output by mapping all reads to the latest human genome release, GRCh37. Mapping was performed with Bowtie (93), allowing for maximal 2 mismatches per read. The sequence output after human read removal consisted of 61.01 million reads.

***E. coli* reference sequences.** We used the genomic sequences of 54 *E. coli* isolates that were available for download from GenBank/EMBL (September 2012) as a reference to map all Illumina reads obtained in this study. The 54 *E. coli* genomes contain 252,623 genes, which give an average of 4,678 genes per strain (more details are presented in Table S5 in the supplemental material). With the aim to collapse those genes into gene families and define the genes present in all genomes, we first extracted all coding sequences (CDS) from the corresponding genomes. We then blasted the protein sequences found in all genomes against each other using BLASTP (94), discarding hits with $<90\%$ length and 50% sequence identity. Only if a gene product had a maximal reciprocal set of homologs in all other strains, 54 in total, the corresponding gene was considered "core"; otherwise, it was considered "flexible." Flexible CDS that had homologs in 53 or 52 of the 54 *E. coli* genomes were reevaluated. The set of core genes detected in the reciprocal Blast search comprised 1,719 CDS, while there were an additional 363 CDS assigned to the core genome, summing to 2,082 core CDS (see Data Set S1 in the supplemental material). Apart from the 2,082 core CDS, we also identified 10,202 flexible CDS, including 3,257 singletons. Among the 54 completed *E. coli* genomes considered here, O26:H11/AP010953, O111:H/AP010960, and O103:H2/

AP010958 exhibited the highest number of annotated small RNAs. We extracted the genomic sequences of 70 noncoding RNAs (ncRNAs) from the O26:H11 genome and performed BLASTN searches against each of the 54 genomes in order to define how many of these ncRNAs are present in all *E. coli* genomes. A total of 47 ncRNAs (45 small RNAs and 2 ncRNA, *rne5*, an RNase 5' untranslated region [UTR] element, and Alpha_RBS, a ribosomal binding site of alpha operon) were found in all 54 genomes, and 41 of them were expressed at least in one of our clinical isolates. These ncRNAs were included in the core genome that consisted of 2,129 (2,082 CDS and 47 ncRNAs) genes. Finally, the sum of core (2,129) and flexible (10,202) genes amounted to 12,331 genes. The data representing the compiled 12,331-gene list, the orthologous gene IDs (with sequence length composition and percentage of identity), and the gene expression levels of each of 21 samples are presented as Data Sets S1A and S1B in the supplemental material.

Mapping and gene expression profiling. The raw Illumina sequence reads (36-bp single end) were first split according to their bar codes using the *fastq-mcf* script of the *ea-utils* package (95), and then the bar code sequences were removed. We used the *bowtie-build* module in the Bowtie package (93) to build an indexed reference based on the 12,331 *E. coli* genes found in the 54 reference genomes as defined in the previous step. Mapping to the reference was performed using Bowtie with options "-m 1 -best -strata" to allow only uniquely mapping hits and avoid uncertainties regarding repeat regions and ribosomal genes. Finally, the read counts per gene (RPG) were recorded for each annotated gene and were used as an input for differential gene expression calculations with the R package *DESeq* (96). Briefly, the RPG data were normalized for variation in library size/sequencing depth by using the *estimateSizeFactor* function of *DESeq*. Differentially expressed genes were identified using the *nbinomTest* function based on the negative binomial model. Genes were considered to be differentially regulated only if their absolute logarithmic fold change over the control was higher than 1 at a false discovery rate of a maximum 5% (Benjamini and Hochberg *P* value correction provided in *DESeq*). In those clinical samples where no technical replica was sequenced, the uncorrected *P* values at 5% cutoff were used instead of the corrected ones.

***De novo* assembly.** All reads that did not map to the 12,331 *E. coli* genes were used as input for *de novo* transcriptome assembly with Velvet (97). We used a wide range of k-mers, 27 to 37, and a minimal transcript length of 100 bp. The assembled transcripts were blasted against all microbial genes downloaded from the MBGD Database (98) using a minimal hit length of 100 bp and sequence similarity higher than 90%. After removing the ribosomal gene hits, we identified 156 additional nonredundant genes.

Phylogenetic tree. A consensus sequence for overall 336 genes (that had at least 80% sequencing coverage across the 21 UTI isolates) was generated by the use of the *mpileup* option in the SAMtools package (99). The corresponding orthologous gene sequences extracted from the 54 *E. coli* genomes were subsequently included. The sequence redundancies and gaps in sequence coverage were removed, resulting in a 2.3-Mb multi-Fasta file used for multiple alignment with Clustal Omega (100). The alignment was the subject of further refinement with RaxML (101), performing 500 bootstrapping steps and testing 50 trees. The consensus tree was drawn with Dendroscope (102).

Gene ontology terms. We downloaded the current UniProt Gene Ontology (GO) knowledgebase (103). Using custom Perl scripts, we mapped the gene locus IDs (in KEGG format) to their UniProt identifiers and extracted the relevant GO IDs. The GO ID lists were summarized using the QuickGO browser (104).

MALDI-TOF mass spectrometry biotyping. Intact cell smears of 19 *E. coli* isolates (for two patient samples, no bacterial cultures were preserved) were prepared in 10 biological replicates on MALDI target plates (MSP 96 polished steel target; Bruker Daltonics, Bremen, Germany) by following standard procedures. The air-dried smears were overlaid with 1 μl of saturated alpha-cyano-4-hydroxycinnamic acid matrix solution. *E. coli* DH5 α bacterial test standard (Bruker Daltonics) was used for ex-

ternal calibration. Bacterial profile spectra were acquired in duplicates using a MicroflexLT MALDI-TOF device (Bruker Daltonics) for analysis in the mass range between 3 and 15,000 *m/z* with the Biotyper 3.1 software (Bruker Daltonics). In a quality-control step, spectra characterized by excessive noise and/or Biotyper scores indicating unreliable identification (<1.7) were excluded from our profile spectra library. We then generated reference spectra of each strain from the remaining 322 profile spectra using Biotyper MSP generation standard settings (105), yielding reference spectra for classification of our closely related *E. coli* strains. In a further quality-control step, we validated that our *E. coli* strains clustered together with the 11 *E. coli* strains among the more than 4,000 strains in the Biotyper database. The 19 strain reference spectra were clustered based on Minkowski distances and group averages.

Nucleotide sequence accession number. The sequencing data have been submitted to SRA under the project accession no. SRP029244.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01075-14/-/DCSupplemental>.

Data Set S1, XLS file, 0.1 MB.
Data Set S2, XLS file, 0.1 MB.
Figure S1, TIF file, 0.8 MB.
Figure S2, TIF file, 0.1 MB.
Figure S3, TIF file, 0.5 MB.
Table S1, PDF file, 0.3 MB.
Table S2, PDF file, 0.1 MB.
Table S3, PDF file, 0.1 MB.
Table S4, PDF file, 0.3 MB.
Table S5, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We acknowledge Klaus Hornischer for his assistance in database management, Amelie Stalke for assistance in Biotyper analysis, and Doris Hopert, Nadine Kubatzki, Michelle Reichel, and Marie-Louise Heydecke for professional assistance in patient sample collection and handling.

This work was supported by an ERC starter grant (RESISTOME 260276) and by the President's Initiative and Networking Fund of the Helmholtz Association of German Research Centers (HGF) under contract number VH-GS-202.

REFERENCES

- Cotter PA, Miller JF. 1998. *In vivo* and *ex vivo* regulation of bacterial virulence gene expression. *Curr. Opin. Microbiol.* 1:17–26. [http://dx.doi.org/10.1016/S1369-5274\(98\)80138-0](http://dx.doi.org/10.1016/S1369-5274(98)80138-0).
- Stamm WE, Hooton TM. 1993. Management of urinary tract infections in adults. *N. Engl. J. Med.* 329:1328–1334. <http://dx.doi.org/10.1056/NEJM199310283291808>.
- Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2:123–140. <http://dx.doi.org/10.1038/nrmicro818>.
- Köhler CD, Dobrindt U. 2011. What defines extraintestinal pathogenic *Escherichia coli*? *Int. J. Med. Microbiol.* 301:642–647. <http://dx.doi.org/10.1016/j.ijmm.2011.09.006>.
- Janke B, Dobrindt U, Hacker J, Blum-Oehler G. 2001. A subtractive hybridisation analysis of genomic differences between the uropathogenic *E. coli* strain 536 and the *E. coli* K-12 strain MG1655. *FEMS Microbiol. Lett.* 199:61–66. <http://dx.doi.org/10.1111/j.1574-6968.2001.tb10651.x>.
- Welch RA, Burland V, Plunkett G, III, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HL, Donnenberg MS, Blattner FR. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 99:17020–17024. <http://dx.doi.org/10.1073/pnas.252529799>.
- Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, Sabo A, Blasiar D, Bieri T, Meyer RR, Ozersky P, Armstrong JR, Fulton RS, Latreille JP, Spieth J, Hooton TM, Mardis ER, Hultgren SJ, Gordon JL. 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. <http://dx.doi.org/10.1073/pnas.0600938103>.
- Brzuszkiewicz E, Brüggemann H, Liesegang H, Emmerth M, Olschläger T, Nagy G, Albersmann K, Wagner C, Buchrieser C, Emody L, Gottschalk G, Hacker J, Dobrindt U. 2006. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc. Natl. Acad. Sci. U. S. A.* 103:12879–12884. <http://dx.doi.org/10.1073/pnas.0603038103>.
- Lloyd AL, Henderson TA, Vigil PD, Mobley HL. 2009. Genomic islands of uropathogenic *Escherichia coli* contribute to virulence. *J. Bacteriol.* 191:3469–3481. <http://dx.doi.org/10.1128/JB.01717-08>.
- Wurpel DJ, Beatson SA, Totsika M, Petty NK, Schembri MA. 2013. Chaperone-usher fimbriae of *Escherichia coli*. *PLoS One* 8:e52835. <http://dx.doi.org/10.1371/journal.pone.0052835>.
- Martinez JJ, Mulvey MA, Schilling JD, Pinkner JS, Hultgren SJ. 2000. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *EMBO J.* 19:2803–2812. <http://dx.doi.org/10.1093/emboj/19.12.2803>.
- Anderson GG, Palermo JJ, Schilling JD, Roth R, Heuser J, Hultgren SJ. 2003. Intracellular bacterial biofilm-like pods in urinary tract infections. *Science* 301:105–107. <http://dx.doi.org/10.1126/science.1084550>.
- Mulvey MA, Schilling JD, Hultgren SJ. 2001. Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection. *Infect. Immun.* 69:4572–4579. <http://dx.doi.org/10.1128/IAI.69.7.4572-4579.2001>.
- Snyder JA, Haugen BJ, Buckles EL, Lockett CV, Johnson DE, Donnenberg MS, Welch RA, Mobley HL. 2004. Transcriptome of uropathogenic *Escherichia coli* during urinary tract infection. *Infect. Immun.* 72:6373–6381. <http://dx.doi.org/10.1128/IAI.72.11.6373-6381.2004>.
- Haugen BJ, Pellett S, Redford P, Hamilton HL, Roesch PL, Welch RA. 2007. *In vivo* gene expression analysis identifies genes required for enhanced colonization of the mouse urinary tract by uropathogenic *Escherichia coli* strain CFT073 *dsdA*. *Infect. Immun.* 75:278–289. <http://dx.doi.org/10.1128/IAI.01319-06>.
- Connell I, Agace W, Klemm P, Schembri M, Märdil S, Svanborg C. 1996. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proc. Natl. Acad. Sci. U. S. A.* 93:9827–9832. <http://dx.doi.org/10.1073/pnas.93.18.9827>.
- Mulvey MA, Lopez-Boado YS, Wilson CL, Roth R, Parks WC, Heuser J, Hultgren SJ. 1998. Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* 282:1494–1497. <http://dx.doi.org/10.1126/science.282.5393.1494>.
- Mills M, Meysick KC, O'Brien AD. 2000. Cytotoxic necrotizing factor type 1 of uropathogenic *Escherichia coli* kills cultured human uroepithelial 5637 cells by an apoptotic mechanism. *Infect. Immun.* 68:5869–5880. <http://dx.doi.org/10.1128/IAI.68.10.5869-5880.2000>.
- Torres AG, Redford P, Welch RA, Payne SM. 2001. TonB-dependent systems of uropathogenic *Escherichia coli*: aerobactin and heme transport and TonB are required for virulence in the mouse. *Infect. Immun.* 69:6179–6185. <http://dx.doi.org/10.1128/IAI.69.10.6179-6185.2001>.
- Bahrani-Mougeot FK, Buckles EL, Lockett CV, Hebel JR, Johnson DE, Tang CM, Donnenberg MS. 2002. Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia coli* virulence determinants in the murine urinary tract. *Mol. Microbiol.* 45:1079–1093. <http://dx.doi.org/10.1046/j.1365-2958.2002.03078.x>.
- Mysorekar IU, Mulvey MA, Hultgren SJ, Gordon JL. 2002. Molecular regulation of urothelial renewal and host defenses during infection with uropathogenic *Escherichia coli*. *J. Biol. Chem.* 277:7412–7419. <http://dx.doi.org/10.1074/jbc.M110560200>.
- Lane MC, Alteri CJ, Smith SN, Mobley HL. 2007. Expression of flagella is coincident with uropathogenic *Escherichia coli* ascension to the upper urinary tract. *Proc. Natl. Acad. Sci. U. S. A.* 104:16669–16674. <http://dx.doi.org/10.1073/pnas.0607898104>.
- Hannan TJ, Totsika M, Mansfield KJ, Moore KH, Schembri MA, Hultgren SJ. 2012. Host-pathogen checkpoints and population bottlenecks in persistent and intracellular uropathogenic *Escherichia coli* bladder infection. *FEMS Microbiol. Rev.* 36:616–648. <http://dx.doi.org/10.1111/j.1574-6976.2012.00339.x>.
- Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, Waldor MK. 2011. RNA-Seq-based monitoring of infection-linked changes in *Vibrio cholerae* gene expression. *Cell Host Microbe* 10:165–174. <http://dx.doi.org/10.1016/j.chom.2011.07.007>.
- Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR, Dougan G. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella Ty-*

- phi. PLoS Genet. 5:e1000569. <http://dx.doi.org/10.1371/journal.pgen.1000569>.
26. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464:250–255. <http://dx.doi.org/10.1038/nature08756>.
 27. Sahl JW, Rasko DA. 2012. Analysis of global transcriptional profiles of enterotoxigenic *Escherichia coli* isolate E24377A. Infect. Immun. 80:1232–1242. <http://dx.doi.org/10.1128/IAI.06138-11>.
 28. Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, Greenberg EP, Sorek R, Lory S. 2012. The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. PLoS Pathog. 8:e1002945. <http://dx.doi.org/10.1371/journal.ppat.1002945>.
 29. Jorth P, Trivedi U, Rumbaugh K, Whiteley M. 2013. Probing bacterial metabolism during infection using high-resolution transcriptomics. J. Bacteriol. 195:4991–4998. <http://dx.doi.org/10.1128/JB.00875-13>.
 30. Dugar G, Herbig A, Förstner KU, Heidrich N, Reinhardt R, Nieselt K, Sharma CM. 2013. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. PLoS Genet. 9:e1003495. <http://dx.doi.org/10.1371/journal.pgen.1003495>.
 31. Ochman H, Selander RK. 1984. Evidence for clonal population structure in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 81:198–201. <http://dx.doi.org/10.1073/pnas.81.1.198>.
 32. Ochman H, Selander RK. 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. 157:690–693.
 33. Escobar-Páramo P, Grenet K, Le Menach A, Rode L, Salgado E, Amorin C, Gouriou S, Picard B, Rahimy MC, Andreumont A, Denamur E, Ruimy R. 2004. Large-scale population structure of human commensal *Escherichia coli* isolates. Appl. Environ. Microbiol. 70:5698–5700. <http://dx.doi.org/10.1128/AEM.70.9.5698-5700.2004>.
 34. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiappello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguénec C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Médigue C, Rocha EP, Denamur E. 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet. 5:e1000344. <http://dx.doi.org/10.1371/journal.pgen.1000344>.
 35. Leimbach A, Hacker J, Dobrindt U. 2013. *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. Curr. Top. Microbiol. Immunol. 358:3–32. http://dx.doi.org/10.1007/82_2012_303.
 36. Chaudhuri RR, Henderson IR. 2012. The evolution of the *Escherichia coli* phylogeny. Infect. Genet. Evol. 12:214–226. <http://dx.doi.org/10.1016/j.meegid.2012.01.005>.
 37. 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. <http://dx.doi.org/10.1128/JB.186.12.3911-3921.2004>.
 38. Zhang Y, Lin K. 2012. A phylogenomic analysis of *Escherichia coli* *Shigella* group: implications of genomic features associated with pathogenicity and ecological adaptation. BMC Evol. Biol. 12:174. <http://dx.doi.org/10.1186/1471-2148-12-174>.
 39. Nougayrède JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, Oswald E. 2006. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. Science 313:848–851. <http://dx.doi.org/10.1126/science.1127059>.
 40. Storz G, Vogel J, Wassarman KM. 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell 43:880–891. <http://dx.doi.org/10.1016/j.molcel.2011.08.022>.
 41. Waters LS, Storz G. 2009. Regulatory RNAs in bacteria. Cell 136:615–628. <http://dx.doi.org/10.1016/j.cell.2009.01.043>.
 42. Urban JH, Papenfort K, Thomsen J, Schmitz RA, Vogel J. 2007. A conserved small RNA promotes discoordinate expression of the *glmUS* operon mRNA to activate GlmS synthesis. J. Mol. Biol. 373:521–528. <http://dx.doi.org/10.1016/j.jmb.2007.07.035>.
 43. Kalamorz F, Reichenbach B, März W, Rak B, Görke B. 2007. Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in *Escherichia coli*. Mol. Microbiol. 65:1518–1533. <http://dx.doi.org/10.1111/j.1365-2958.2007.05888.x>.
 44. Liu MY, Gui G, Wei B, Preston JF, III, Oakford L, Yüksel U, Giedroc DP, Romeo T. 1997. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. J. Biol. Chem. 272:17502–17510. <http://dx.doi.org/10.1074/jbc.272.28.17502>.
 45. Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, Baker CS, Georgellis D, Babitzke P, Romeo T. 2003. A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. Mol. Microbiol. 48:657–670. <http://dx.doi.org/10.1046/j.1365-2958.2003.03459.x>.
 46. Romeo T, Gong M, Liu MY, Brun-Zinkernagel AM. 1993. Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. J. Bacteriol. 175:4744–4755.
 47. Lucchetti-Miganeh C, Burrows E, Baysse C, Ermel G. 2008. The post-transcriptional regulator CsrA plays a central role in the adaptation of bacterial pathogens to different stages of infection in animal hosts. Microbiology 154:16–29. <http://dx.doi.org/10.1099/mic.0.2007/012286-0>.
 48. Fortune DR, Suyemoto M, Altier C. 2006. Identification of CsrC and characterization of its role in epithelial cell invasion in *Salmonella enterica* serovar Typhimurium. Infect. Immun. 74:331–339. <http://dx.doi.org/10.1128/IAI.74.1.331-339.2006>.
 49. Bhatt S, Edwards AN, Nguyen HT, Merlin D, Romeo T, Kalman D. 2009. The RNA binding protein CsrA is a pleiotropic regulator of the locus of enterocyte effacement pathogenicity island of enteropathogenic *Escherichia coli*. Infect. Immun. 77:3552–3568. <http://dx.doi.org/10.1128/IAI.00418-09>.
 50. Wang X, Dubey AK, Suzuki K, Baker CS, Babitzke P, Romeo T. 2005. CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. Mol. Microbiol. 56:1648–1663. <http://dx.doi.org/10.1111/j.1365-2958.2005.04648.x>.
 51. Dötsch A, Eckweiler D, Schniederjans M, Zimmermann A, Jensen V, Scharfe M, Geffers R, Häussler S. 2012. The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. PLoS One 7:e31092. <http://dx.doi.org/10.1371/journal.pone.0031092>.
 52. Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, Cashel M, Babitzke P, Romeo T. 2011. Circuitry linking the Csr and stringent response global regulatory systems. Mol. Microbiol. 80:1561–1580. <http://dx.doi.org/10.1111/j.1365-2958.2011.07663.x>.
 53. Udekwi KI, Darfeuille F, Vogel J, Reimegård J, Holmqvist E, Wagner EG. 2005. Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. Genes Dev. 19:2355–2366. <http://dx.doi.org/10.1101/gad.354405>.
 54. Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA. 2010. Positive regulation by small RNAs and the role of Hfq. Proc. Natl. Acad. Sci. U. S. A. 107:9602–9607. <http://dx.doi.org/10.1073/pnas.1004435107>.
 55. Majdalani N, Chen S, Murrow J, St John K, Gottesman S. 2001. Regulation of RpoS by a novel small RNA: the characterization of RprA. Mol. Microbiol. 39:1382–1394. <http://dx.doi.org/10.1111/j.1365-2958.2001.02329.x>.
 56. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. 2001. Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. 15:1637–1651. <http://dx.doi.org/10.1101/gad.901001>.
 57. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr. Biol. 11:941–950. [http://dx.doi.org/10.1016/S0960-9822\(01\)00270-6](http://dx.doi.org/10.1016/S0960-9822(01)00270-6).
 58. Jørgensen MG, Nielsen JS, Boysen A, Franch T, Møller-Jensen J, Valentin-Hansen P. 2012. Small regulatory RNAs control the multicellular adhesive lifestyle of *Escherichia coli*. Mol. Microbiol. 84:36–50. <http://dx.doi.org/10.1111/j.1365-2958.2012.07976.x>.
 59. Holmqvist E, Reimegård J, Sterk M, Grantcharova N, Römmling U, Wagner EG. 2010. Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. EMBO J. 29:1840–1850. <http://dx.doi.org/10.1038/emboj.2010.73>.
 60. Billips BK, Schaeffer AJ, Klumpp DJ. 2008. Molecular basis of uro-

- pathogenic *Escherichia coli* evasion of the innate immune response in the bladder. *Infect. Immun.* 76:3891–3900. <http://dx.doi.org/10.1128/IAI.00069-08>.
61. Yim L, Moukadiri I, Björk GR, Armengod ME. 2006. Further insights into the tRNA modification process controlled by proteins MnmE and GidA of *Escherichia coli*. *Nucleic Acids Res.* 34:5892–5905. <http://dx.doi.org/10.1093/nar/gkl752>.
 62. Kinscherf TG, Willis DK. 2002. Global regulation by *gidA* in *Pseudomonas syringae*. *J. Bacteriol.* 184:2281–2286. <http://dx.doi.org/10.1128/JB.184.8.2281-2286.2002>.
 63. Sha J, Kozlova EV, Fadl AA, Olano JP, Houston CW, Peterson JW, Chopra AK. 2004. Molecular characterization of a glucose-inhibited division gene, *gidA*, that regulates cytotoxic enterotoxin of *Aeromonas hydrophila*. *Infect. Immun.* 72:1084–1095. <http://dx.doi.org/10.1128/IAI.72.2.1084-1095.2004>.
 64. Durand JM, Dagberg B, Uhlin BE, Björk GR. 2000. Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the *virF* gene. *Mol. Microbiol.* 35:924–935. <http://dx.doi.org/10.1046/j.1365-2958.2000.01767.x>.
 65. Li W, Liu L, Qiu D, Chen H, Zhou R. 2010. Identification of *Streptococcus suis* serotype 2 genes preferentially expressed in the natural host. *Int. J. Med. Microbiol.* 300:482–488. <http://dx.doi.org/10.1016/j.ijmm.2010.04.018>.
 66. Cho KH, Caparon MG. 2008. tRNA modification by GidA/MnmE is necessary for *Streptococcus pyogenes* virulence: a new strategy to make live attenuated strains. *Infect. Immun.* 76:3176–3186. <http://dx.doi.org/10.1128/IAI.01721-07>.
 67. Shippy DC, Eakley NM, Lauhon CT, Bochsler PN, Fadl AA. 2013. Virulence characteristics of *Salmonella* following deletion of genes encoding the tRNA modification enzymes GidA and MnmE. *Microb. Pathog.* 57:1–9. <http://dx.doi.org/10.1016/j.micpath.2013.01.004>.
 68. Yu H, Kim KS. 2012. mRNA context dependent regulation of cytotoxic necrotizing factor 1 translation by GidA, a tRNA modification enzyme in *Escherichia coli*. *Gene* 491:116–122. <http://dx.doi.org/10.1016/j.gene.2011.10.013>.
 69. Hagan EC, Lloyd AL, Rasko DA, Faerber GJ, Mobley HL. 2010. *Escherichia coli* global gene expression in urine from women with urinary tract infection. *PLoS Pathog.* 6:e1001187. <http://dx.doi.org/10.1371/journal.ppat.1001187>.
 70. Roos V, Klemm P. 2006. Global gene expression profiling of the asymptomatic bacteriuria *Escherichia coli* strain 83972 in the human urinary tract. *Infect. Immun.* 74:3565–3575. <http://dx.doi.org/10.1128/IAI.01959-05>.
 71. Vejborg RM, de Evgrafov MR, Phan MD, Totsika M, Schembri MA, Hancock V. 2012. Identification of genes important for growth of asymptomatic bacteriuria *Escherichia coli* in urine. *Infect. Immun.* 80:3179–3188. <http://dx.doi.org/10.1128/IAI.00473-12>.
 72. Kariyawasam S, Wilkie BN, Gyles CL. 2004. Construction, characterization, and evaluation of the vaccine potential of three genetically defined mutants of avian pathogenic *Escherichia coli*. *Avian Dis.* 48:287–299. <http://dx.doi.org/10.1637/7093>.
 73. Korea CG, Badouraly R, Prevost MC, Ghigo JM, Beloin C. 2010. *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environ. Microbiol.* 12:1957–1977. <http://dx.doi.org/10.1111/j.1462-2920.2010.02202.x>.
 74. Ledebore NA, Frye JG, McClelland M, Jones BD. 2006. *Salmonella enterica* serovar Typhimurium requires the Lpf, Pef, and Tafi fimbriae for biofilm formation on HEp-2 tissue culture cells and chicken intestinal epithelium. *Infect. Immun.* 74:3156–3169. <http://dx.doi.org/10.1128/IAI.01428-05>.
 75. Sakellaris H, Balding DP, Scott JR. 1996. Assembly proteins of CS1 pili of enterotoxigenic *Escherichia coli*. *Mol. Microbiol.* 21:529–541. <http://dx.doi.org/10.1111/j.1365-2958.1996.tb02562.x>.
 76. Korea CG, Ghigo JM, Beloin C. 2011. The sweet connection: solving the riddle of multiple sugar-binding fimbrial adhesins in *Escherichia coli*: multiple *E. coli* fimbriae form a versatile arsenal of sugar-binding lectins potentially involved in surface-colonisation and tissue tropism. *BioEssays* 33:300–311. <http://dx.doi.org/10.1002/bies.201000121>.
 77. Hill CW. 1999. Large genomic sequence repetitions in bacteria: lessons from rRNA operons and Rhs elements. *Res. Microbiol.* 150:665–674. [http://dx.doi.org/10.1016/S0923-2508\(99\)00125-4](http://dx.doi.org/10.1016/S0923-2508(99)00125-4).
 78. Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. 2007. Genes under positive selection in *Escherichia coli*. *Genome Res.* 17:1336–1343. <http://dx.doi.org/10.1101/gr.6254707>.
 79. Hill CW, Feulner G, Brody MS, Zhao S, Sadosky AB, Sandt CH. 1995. Correlation of Rhs elements with *Escherichia coli* population structure. *Genetics* 141:15–24.
 80. Kung VL, Khare S, Stehlik C, Bacon EM, Hughes AJ, Hauser AR. 2012. An *rhs* gene of *Pseudomonas aeruginosa* encodes a virulence protein that activates the inflammasome. *Proc. Natl. Acad. Sci. U. S. A.* 109:1275–1280. <http://dx.doi.org/10.1073/pnas.1109285109>.
 81. Koskiniemi S, Lamoureux JG, Nikolakis KC, t'Kint de Roodenbeke C, Kaplan MD, Low DA, Hayes CS. 2013. Rhs proteins from diverse bacteria mediate intercellular competition. *Proc. Natl. Acad. Sci. U. S. A.* 110:7032–7037. <http://dx.doi.org/10.1073/pnas.1300627110>.
 82. Poole SJ, Diner EJ, Aoki SK, Braaten BA, t'Kint de Roodenbeke C, Low DA, Hayes CS. 2011. Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS Genet.* 7:e1002217. <http://dx.doi.org/10.1371/journal.pgen.1002217>.
 83. Shneider MM, Buth SA, Ho BT, Basler M, Mekalanos JJ, Leiman PG. 2013. PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* 500:350–353. <http://dx.doi.org/10.1038/nature12453>.
 84. Norton JP, Mulvey MA. 2012. Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic *Escherichia coli*. *PLoS Pathog.* 8:e1002954. <http://dx.doi.org/10.1371/journal.ppat.1002954>.
 85. Guyer DM, Radulovic S, Jones FE, Mobley HL. 2002. Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuulating cytotoxin for bladder and kidney epithelial cells. *Infect. Immun.* 70:4539–4546. <http://dx.doi.org/10.1128/IAI.70.8.4539-4546.2002>.
 86. Bauer RJ, Zhang L, Foxman B, Siitonen A, Jantunen ME, Saxen H, Marrs CF. 2002. Molecular epidemiology of 3 putative virulence genes for *Escherichia coli* urinary tract infection---*usp*, *iha*, and *iroN* (*E. coli*). *J. Infect. Dis.* 185:1521–1524. <http://dx.doi.org/10.1086/340206>.
 87. Spurbeck RR, Stapleton AE, Johnson JR, Walk ST, Hooton TM, Mobley HL. 2011. Fimbrial profiles predict virulence of uropathogenic *Escherichia coli* strains: contribution of *ygi* and *yad* fimbriae. *Infect. Immun.* 79:4753–4763. <http://dx.doi.org/10.1128/IAI.05621-11>.
 88. Francetic O, Belin D, Badaut C, Pugsley AP. 2000. Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion. *EMBO J.* 19:6697–6703. <http://dx.doi.org/10.1093/emboj/19.24.6697>.
 89. Frederiksen RF, Paspaliari DK, Larsen T, Storgaard BG, Larsen MH, Ingmer H, Pálci MM, Leisner JJ. 2013. Bacterial chitinases and chitin-binding proteins as virulence factors. *Microbiology* 159:833–847. <http://dx.doi.org/10.1099/mic.0.051839-0>.
 90. Narciso A, Nunes F, Amores T, Lito L, Melo-Cristino J, Duarte A. 2012. Persistence of uropathogenic *Escherichia coli* strains in the host for long periods of time: relationship between phylogenetic groups and virulence factors. *Eur. J. Clin. Microbiol. Infect. Dis.* 31:1211–1217. <http://dx.doi.org/10.1007/s10096-011-1431-7>.
 91. Sabaté M, Moreno E, Pérez T, Andreu A, Prats G. 2006. Pathogenicity island markers in commensal and uropathogenic *Escherichia coli* isolates. *Clin. Microbiol. Infect.* 12:880–886. <http://dx.doi.org/10.1111/j.1469-0691.2006.01461.x>.
 92. Ramos NL, Sekikubo M, Dzung DT, Kosnopfel C, Kironde F, Mirembe F, Brauner A. 2012. Uropathogenic *Escherichia coli* isolates from pregnant women in different countries. *J. Clin. Microbiol.* 50:3569–3574. <http://dx.doi.org/10.1128/JCM.01647-12>.
 93. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. <http://dx.doi.org/10.1186/gb-2009-10-3-r25>.
 94. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
 95. Aronesty E. 2013. Comparison of sequencing utility programs. *Open Bioinform. J.* 7:1–8. <http://dx.doi.org/10.2174/1875036201307010001>.
 96. Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. <http://dx.doi.org/10.1186/gb-2010-11-10-r106>.
 97. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read

- assembly using de Bruijn graphs. *Genome Res.* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
98. Uchiyama I, Mihara M, Nishide H, Chiba H. 2013. MBGD update 2013: the microbial genome database for exploring the diversity of microbial world. *Nucleic Acids Res.* 41:D631–D635. <http://dx.doi.org/10.1093/nar/gks1006>.
 99. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>.
 100. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. <http://dx.doi.org/10.1038/msb.2011.75>.
 101. Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <http://dx.doi.org/10.1093/bioinformatics/btl446>.
 102. Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61:1061–1067. <http://dx.doi.org/10.1093/sysbio/sys062>.
 103. Dimmer EC, Huntley RP, Alam-Faruque Y, Sawford T, O'Donovan C, Martin MJ, Bely B, Browne P, Mun Chan W, Eberhardt R, Gardner M, Laiho K, Legge D, Magrane M, Pichler K, Poggioli D, Sehra H, Auchincloss A, Axelsen K, Blatter MC, Boutet E, Braconi-Quintaje S, Breuza L, Bridge A, Coudert E, Estreicher A, Famiglietti L, Ferro-Rojas S, Feuermann M, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, James J, Jimenez S, Jungo F, Keller G, Lemercier P, Lieberherr D, Masson P, Moinat M, Pedruzzi I, Poux S, Rivoire C, Roechert B, Schneider M, Stutz A, Sundaram S, Tognolli M, Bougueleret L, Argoud-Puy G, Cusin I, Duek-Roggli P, Xenarios I, Apweiler R. 2012. The UniProt-GO annotation database in 2011. *Nucleic Acids Res.* 40: D565–D570. <http://dx.doi.org/10.1093/nar/gkr1048>.
 104. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. 2009. QuickGO: a Web-based tool for Gene Ontology searching. *Bioinformatics* 25:3045–3046. <http://dx.doi.org/10.1093/bioinformatics/btp536>.
 105. Karger A, Stock R, Ziller M, Elschner MC, Bettin B, Melzer F, Maier T, Kostrzewa M, Scholz HC, Neubauer H, Tomaso H. 2012. Rapid identification of *Burkholderia mallei* and *Burkholderia pseudomallei* by intact cell matrix-assisted laser desorption/ionisation mass spectrometric typing. *BMC Microbiol.* 12:229. <http://dx.doi.org/10.1186/1471-2180-12-229>.