

Transcriptomics and adaptive genomics of the asymptomatic bacteriuria *Escherichia coli* strain 83972

Viktoria Hancock · Aswin S. Seshasayee ·
David W. Ussery · Nicholas M. Luscombe · Per Klemm

Received: 20 December 2007 / Accepted: 31 January 2008 / Published online: 4 March 2008
© The Author(s) 2008

Abstract *Escherichia coli* strains are the major cause of urinary tract infections in humans. Such strains can be divided into virulent, UPEC strains causing symptomatic infections, and asymptomatic, commensal-like strains causing asymptomatic bacteriuria, ABU. The best-characterized ABU strain is strain 83972. Global gene expression profiling of strain 83972 has been carried out under seven different sets of environmental conditions ranging from laboratory minimal medium to human bladders. The data reveal highly specific gene expression responses to different conditions. A number of potential fitness factors for the human urinary tract could be identified. Also, presence/absence data of the gene expression was used as an adaptive genomics tool to model the gene pool of 83972 using primarily UPEC strain CFT073 as a scaffold. In our analysis, 96% of the transcripts filtered present in strain 83972 can be found in CFT073, and genes on six of the seven pathogenicity islands were expressed in 83972. Despite the very different patient symptom profiles, the two strains

seem to be very similar. Genes expressed in CFT073 but not in 83972 were identified and can be considered as virulence factor candidates. Strain 83972 is a deconstructed pathogen rather than a commensal strain that has acquired fitness properties.

Keywords Asymptomatic bacteriuria · Global gene expression · Microarray · Urinary tract infections · Virulence factors

Introduction

Urinary tract infection (UTI) is one of the most common infectious diseases in humans and a major cause of morbidity. It is estimated that 40–50% of adult healthy women have experienced at least one UTI episode (Foxman 2002). UTI can be caused either by pathogenic strains leading to symptomatic UTI or by asymptomatic bacteriuria (ABU) strains resulting in a symptom-free carriage resembling commensalism. *Escherichia coli* is responsible for more than 80% of all UTIs. Acute pyelonephritis is a severe acute systemic infection caused by uropathogenic *E. coli* (UPEC) clones with virulence genes clustered on “pathogenicity islands” (PAIs) (Eden et al. 1976; Funfstuck et al. 1986; Stenqvist et al. 1987; Orskov et al. 1988; Johnson 1991; Welch et al. 2002). Paradoxically, a large proportion of UTIs are caused by ABU *E. coli*. Individuals infected with ABU-class *E. coli* may carry high urine titres of a single *E. coli* strain for months or years without provoking a host response.

Escherichia coli 83972 is a prototype ABU strain and undoubtedly the best-characterised ABU-class *E. coli* to date. Strain 83972 was originally isolated in the 1970s from a young girl who had carried it for at least 3 years without

Communicated by D. Andersson.

V. Hancock · P. Klemm (✉)
Microbial Adhesion Group, Risø DTU,
Technical University of Denmark, Building 301,
2800 Kgs Lyngby, Denmark
e-mail: pk1@bio.dtu.dk

A. S. Seshasayee · N. M. Luscombe
EMBL-European Bioinformatics Institute,
Wellcome Trust Genome Campus,
Cambridge CB10 1SD, UK

D. W. Ussery
Centre for Biological Sequence Analysis,
DTU Department of Systems Biology,
Technical University of Denmark,
2800 Kgs Lyngby, Denmark

symptoms (Lindberg et al. 1975; Andersson et al. 1991). The strain is well adapted for growth in the human urinary tract (UT) where it establishes long-term bacteriuria (Hull et al. 2000). It has been used for prophylactic purposes in numerous studies; as such it has been used as an alternative treatment in patients with recurrent UTI who are refractory to conventional therapy (Hull et al. 2000). An ongoing study on patients infected with strain 83972 has so far reported over 50 patient years with no serious side effects (Sundén et al. 2006).

ABU patients may carry a single strain for months or years, creating a condition that resembles commensalism, but with a strain that may have evolved from a pathogenic ancestor. Several lines of evidence support the notion that the ancestor of strain 83972 was a pyelonephritic UPEC strain; it belongs to the B2 clonal group, a group associated with pyelonephritis and other extra-intestinal invasive clinical syndromes such as bacteremia, prostatitis and meningitis; the strain also contains gene clusters in various stages of erosion encoding the three UPEC-class fimbriae, i.e. the *fim*, *pap* and *sfalfoc* clusters (Klemm et al. 2006; Roos et al. 2006a).

Several studies have investigated the virulence characteristics of uropathogenic *E. coli* (UPEC) isolates and ABU isolates, in order to get better understanding of how some UTI strains can cause severe disease, while others can be used prophylactically to prevent the same (Blanco et al. 1996; Dobrindt et al. 2003; Vranes et al. 2003; Johnson et al. 2005; Marrs et al. 2005). ABU strains have been shown to lack many of the virulence-associated phenotypes; many of them are nonhaemolytic, nonadherent and lack haemagglutination ability (Vranes et al. 2003). Strain 83972 lacks many of the virulence-associated phenotypes but has been shown to carry many of the virulence-associated genes, such as *kps*, *iutA*, *fyuA* and *malX* (Dobrindt et al. 2003). However, apart from the fimbrial clusters, the strain has not been sequenced and it is not known which genes it shares with other *E. coli* isolates, which of the genes in the *E. coli* “core genome” it carries and which genes it shares with other UTI isolates.

Thus far, although a number of UPEC isolates (i.e. CFT073, 536, UTI89 and F11) have been completely sequenced, no genomic sequencing of any ABU strain has been reported. Comparative genomics profiling using microarray chips designed to cover entire genomes is one strategy to obtain information about variability between different strains of the same species and indication of horizontal gene transfer (Willenbrock et al. 2006). DNA microarray-assisted functional genomics provides the global expression profile of a strain, revealing which genes are expressed under certain conditions. Global gene expression profiling of ABU strain 83972 employing the GeneChip *E. coli* Genome 2.0 Array (Affymetrix), containing four

E. coli genomes including that of the UPEC isolate CFT073, will not only provide information regarding the up- and down-regulation of genes comparing different conditions, but will also reveal which genes are actually present (expressed) in the genome of 83972. Bacterial pathogens differ from commensals by expression of specific virulence factors such as those that mediate histological damage. Commensals, in contrast, have generally been regarded as bacteria lacking such virulence factors or other specific mechanisms for interaction with host tissues. Here we compare the global expression profiles of *E. coli* ABU strain 83972 grown under a number of different in vitro conditions and in three patients in order to get a representative picture of which genes are present/expressed in the genome of this asymptomatic UTI strain.

Materials and methods

Bacterial strain

Escherichia coli 83972 is a prototype ABU strain and lacks defined O and K surface antigens (Lindberg et al. 1975). It belongs to the ECOR group B2 together with many other UTI strains such as the well-characterized and virulent *E. coli* isolates CFT073, 536 and J96.

Growth conditions and stabilisation of RNA for microarray experiments of *E. coli* 83972 grown on urine agar plates

Human urine was collected from four healthy men and women volunteers who had no history of UTI or antibiotic use in the prior 2 months. The urine was pooled, filter sterilised, stored at 4°C, and used within the following day. *E. coli* 83972 was grown aerated in triplicates in 10 ml of human urine for 6 h. Thereafter, 100 µl of each culture was spread on urine plates (1:1 ratio of human urine and 0.9% NaCl) containing 1.5% agar. The plates were incubated at 37°C for 16 h. Subsequently, 600 µl of a 1:2 mixture of PBS and RNeasy Protect™ Bacteria Reagent (QIAGEN AG) was poured on the plates, mixed with the lawn of cells and incubated for 5 min at room temperature to stabilise RNA. The stabilised mixture was then centrifuged and pellets were stored at –80°C. The samples from 83972, grown exponentially in MOPS and urine, in urine biofilms and in patients (>10⁸ CFU/ml) were all treated identically with RNeasy Protect Bacteria Reagent and have been described previously (Roos and Klemm 2006; Hancock and Klemm 2007).

RNA isolation and microarray hybridisation

Total RNA was isolated using the RNeasy® Mini Kit (QIAGEN AG) and on-column DNase digestion was

performed using RNase-Free DNase Set (QIAGEN AG). The quality of the total RNA was examined by agarose gel electrophoresis and by measuring the absorbance at 260 and 280 nm to ensure intact high-quality RNA. Purified RNA was precipitated with ethanol and stored at -80°C until further use. Conversion of RNA (10 μg per sample) to cDNA, labelling and microarray hybridisation were performed according to the GeneChip Expression Analysis Technical Manual 701023 Rev. 4 (Affymetrix, Inc., Santa Clara, CA). GeneChip *E. coli* Genome 2.0 Arrays (Affymetrix) were used for hybridisation of the labelled cDNA. The microarrays were scanned using the GeneChip Scanner 3000.

Data analysis

The raw intensities from the microarray experiments were background corrected and quantile-normalised. All microarray data in the study were obtained from mRNA being converted to cDNA, i.e. no genomic DNA was used for hybridisation. Probe intensities were summarised to yield expression values for each probe set or gene. These calculations were performed using the implementation of GCRMA (Wu et al. 2004) in Bioconductor (Gentleman et al. 2004) (<http://www.bioconductor.org>, <http://www.r-project.org>). In order to derive a cut-off expression value for making presence/absence calls, we made use of intensities due to control probe sets with IDs beginning with AFFY. There were 96 such probes. The cut-off value was set so that only the top 1/16th of these control probes would be flagged as present. As a result 4,109 genes in the array were marked as present; the remaining genes are referred to as “absent” throughout this report, i.e. these genes could be truly absent, non-homologous or not expressed during any of the seven different growth conditions. Orthologs of all the genes in the array across *E. coli* K12 MG1655, *E. coli* O157:H7 Sakai, *E. coli* O157:H7 EDL933 and *E. coli* CFT073 were identified using bidirectional best hit BLAST.

Microarray data accession number

The supporting microarray data have been deposited in ArrayExpress (<http://www.ebi.ac.uk/arrayexpress>) with accession numbers E-MEXP-584 (MOPS, urine and patient arrays), E-MEXP-926 (biofilm arrays) and E-MEXP-1453 (urine-agar plate arrays).

Results

Genes expressed in ABU *E. coli* 83972

The bacterial transcriptome is a dynamic entity that reflects the organism’s immediate, ongoing response to its environ-

ment. DNA microarray-assisted functional genomics provides the global expression profile of the genome. The genomic expression profiles of the urinary tract infectious *E. coli* isolate 83972 were analysed under several different growth conditions and in different media using the GeneChip *E. coli* Genome 2.0 Array (Affymetrix). This array contains approximately 10,000 probe sets for all 20,366 genes present in *E. coli* strains MG1655 (K-12), CFT073 (UPEC), EDL933 (EHEC) and O157:H7-Sakai (EHEC). Due to the high degree of similarity between the *E. coli* strains, whenever possible, a single probe set is tiled to represent the equivalent ortholog in all the four strains.

In total, 21 microarrays were included in the study; arrays in triplicates were hybridised with RNA of the ABU strain 83972 cultured (1) aerobically to exponential phase in MOPS minimal medium, (2) aerobically to exponential phase in pooled human urine, (3) on urine agar plates, (4) statically in urine biofilm on Petri dishes and finally, (5–7) in three patients (Pat1, Pat2 and Pat3) in vivo. Figure 1 shows the expression levels of all CFT073 genes in strain 83972 during growth in the different environments; many genes were similarly expressed during all seven conditions. However, some genes were expressed only during one or a few of the conditions. For example, the genes encoding yersiniabactin in the high pathogenicity island (HPI), i.e. PAI-*asnT*, were highly expressed in Pat2 (and in biofilm), but much lower during the other conditions. The *c2557–c2563* genes (around 2.4 M in Fig. 1), involved in nucleotide sugar and mannose metabolism and encoding hypothetical proteins, were highly expressed in Pat3 but not under any other condition. Another example is the *c1968–c1971* genes (around 1.8 M), i.e. *ydfI* encoding a D-mannonate oxidoreductase, *ydfJ* encoding a metabolite transport protein and *rspAB* involved in the starvation response, which also were highly expressed only in Pat3.

In total, there were 108 genes that were significantly changed in all six urine environments compared with MOPS. Twenty of these genes were up-regulated in all six urine conditions whereof half were related to different iron systems, i.e. *iroN*, *fepA*, *fecI*, *iucBC*, *fhuA* and *exbD*, as well as *b3337* and *b1995* involved in iron storage and encoding a putative haemin receptor, respectively. The other urine up-regulated genes were *marA*, a multiple antibiotic resistance gene, *sodA*, encoding superoxide dismutase, *ahpC*, encoding hydroperoxide reductase, *b1452*, *c1220*, *c4210*, *lysA*, *rrsG*, *rrsH* and *yrbL*. Most iron acquisition systems were expressed in all the six urine environments; the enterobactin, salmochelin, aerobactin, haem and *sitABCD* systems were all expressed in all the six urine conditions (although weaker in the urine plates). Interestingly, the *fec* system, which is a citrate-dependent iron uptake system found in K-12 but missing in CFT073 and other UPEC strains, was highly expressed in Pat3. Up-regulation of all

these iron-uptake systems revealed that the strain has an impressive array of iron acquisition systems and all of these are active in the human bladder.

Nineteen of the top 31 highest expressed genes overall were genes involved in ribosomal synthesis. The high expression of ribosomal genes in *E. coli* 83972 suggests a rapid growth rate; the highest expression values were obtained in Pat1 followed by MOPS, urine and Pat2, indicating a growth rate just as fast in the patients *in vivo* as in exponential growth phase in a shake flask. This supports our hypothesis that the strain's optimized growth properties in human urine explain its ability to successfully colonize the human urinary tract in the absence of functional fimbriae (Roos et al. 2006b).

Figure 1 reveals that strain 83972 almost exclusively expresses the iron uptake and transport systems in the seven CFT073 PAIs, almost none of the other genes in these islands are expressed. There are only two exceptions; *c0300*, located in PAI-*aspV* encoding a hypothetical protein, and *c3686–3690*, located in PAI-*pheV* encoding YrbH and KpsEDC. The *yrbH* gene belongs to the 131 genes that were recently identified as UPEC specific and it was the second highest expressed UPEC-specific gene in mice (Lloyd et al. 2007); in our samples the highest expression

was found in the three patients and in MOPS. Outside the PAIs there are a few genes/gene clusters that are highly expressed in all urine samples or only in the patients. The enterobactin system was up-regulated during all urine conditions and the *chu* cluster (involved in haem uptake and transport) was highest up-regulated in the patients followed by *in vitro* urine growth. The *ycdO* and *ycdB* genes were highly expressed in the three patients; these have recently been identified to encode haemoproteins, probably involved in iron transport, induced at acidic conditions (Sturm et al. 2006).

Looking at the significantly changed genes for all six urine conditions compared with MOPS (in total 1,897 genes) revealed that Pat2 and Pat3 shared the largest number of similarly changed genes; 75% of all changed genes in Pat2 are regulated in the same way (i.e. up or down-regulated) in Pat3 (Fig. 2). Interestingly, Pat1 shared the largest number similarly regulated genes with the biofilm growth mode; also for Pat2 and Pat3, the biofilm growth mode showed a larger number of similarly regulated genes than Pat1 or any other condition. This could indicate that the expression profile of strain 83972 during *in vivo* growth is closer related to biofilm growth than to growth in shake flasks or plates.

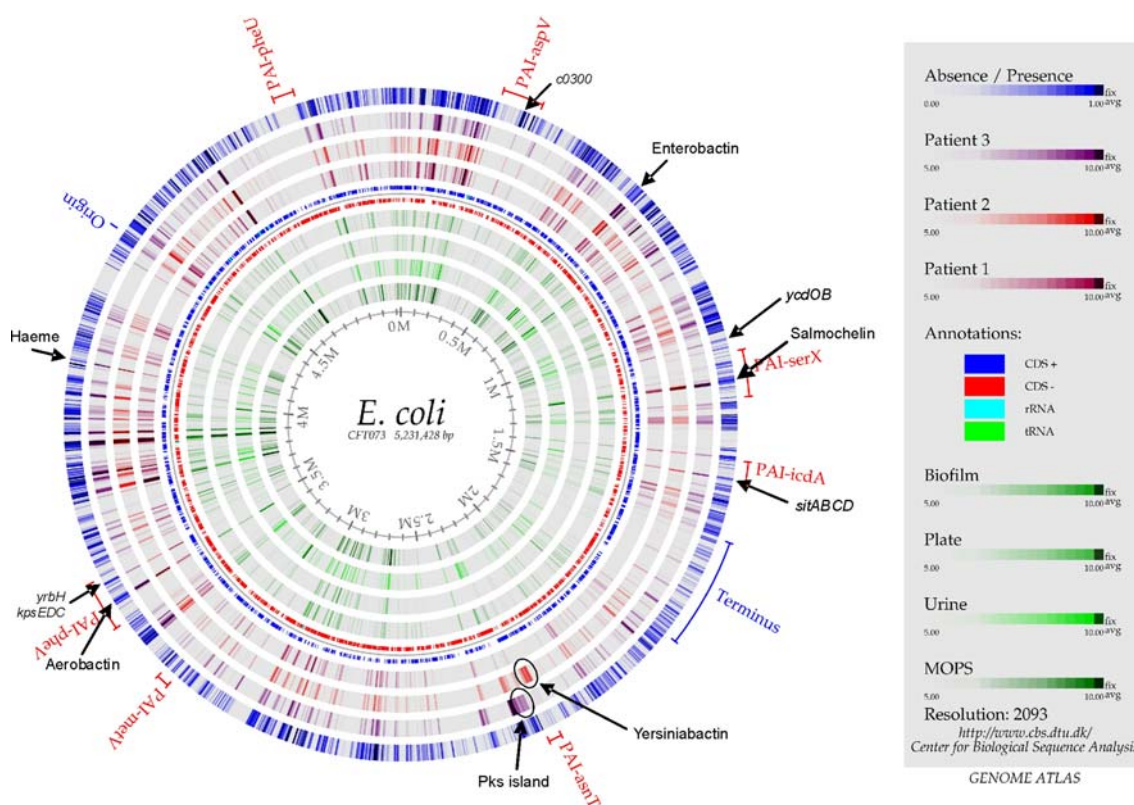


Fig. 1 The expression levels of CFT073 genes in strain 83972 during seven different growth conditions. The outer blue circle shows the calculated absence (0.0) and presence (1.0) of CFT073 genes in ABU strain 83972. The seven PAIs of CFT073 are indicated in red

UP+DOWN	Urine	Plate	Biofilm	Pat1	Pat2	Pat3
Urine	664	311	348	221	297	363
Plate		938	563	285	269	347
Biofilm			1082	369	343	408
Pat1				648	339	304
Pat2					594	446
Pat3						818

Fig. 2 Number of significantly up- and down-regulated genes in strain 83972 during the different growth conditions (i.e. exponential growth in urine, on urine-agar plates, in urine biofilm, in vivo in three patients) compared with exponential growth in MOPS minimal lab medium. The *diagonal boxes* (dark blue colour) show the number of significantly changed genes during cultivation in that specific condition compared with MOPS (e.g. 664 genes were up- or down-regulated in urine compared with MOPS and 938 genes were changed in plates compared with MOPS) and the other *boxes* show the number of significantly changed genes shared between two conditions (e.g. 311 of the 664 and 938 significantly changed genes in urine and plates compared with MOPS were shared between these two conditions, i.e. up- or down-regulated in both urine and plates compared with MOPS). Stronger blue colour indicates larger number of significantly changed genes shared between two conditions

Closeness to CFT073

Given the different growth conditions analysed, it is not unrealistic to assume that most genes present in strain 83972 would be expressed, to some extent, under at least one of these seven different conditions/environments, i.e. growth in liquid and on solid media; during exponential phase, in biofilm and during colony-forming conditions; in different growth media (human urine and minimal lab medium); as well as in vivo in three different individuals.

Data analysis of the 21 microarrays revealed that of the 8,716 *E. coli* transcripts on the microarray (not including probes representing intergenic regions and controls), 4,109 transcripts (47%) showed expression levels above detection limit during at least one of the growth conditions investigated (referred to as “present”, see blue, outer circle in Fig. 1). Figure 3 shows the distribution among the four *E. coli* genomes represented on the microarray of these 4,109 transcripts expressed in *E. coli* 83972. Not surprisingly, the UTI strain 83972 shows highest similarity with the UPEC isolate CFT073 of the four genomes on the array; the large majority of the 4,109 transcripts found present (96.3%) can be found in CFT073, corresponding to 71% of the CFT073 genome. *E. coli* 83972 expressed 150 genes that do not exist in CFT073; 85 of these can be found in MG1655 and the remaining 65 genes can be found exclusively in one or both of the two EHEC strains present on the array (Fig. 3). Thirty of the 65 genes homologous to EHEC genes are encoding proteins of cryptic prophages, whereas the large majority of the remaining 35 genes encode unknown or

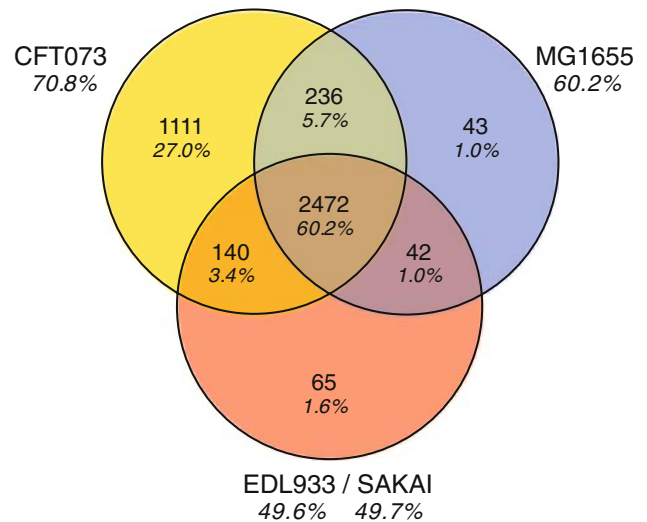


Fig. 3 Venn diagrams showing the distribution of the 4,109 genes filtered present in strain 83972. The percentages indicated below each strain show how large part of the genome of the corresponding strain was filtered present in strain 83972

hypothetical proteins. The 85 genes that can be found in MG1655 but not in CFT073 includes the *fec* cluster encoding an iron citrate transport system (*fecABCDEIR*). In total, 3,959 CFT073 genes were expressed in strain 83972; this could be compared with 4,162 CFT073 genes present in the UPEC (cystitis) isolate F11 (Lloyd et al. 2007).

E. coli core genome

There is a large diversity in size of the chromosome of *E. coli*; in all 32 *E. coli* (and *Shigella*) genomes that have been fully sequenced, or at least with an expected coverage of greater than 99%, the size of the chromosome ranges from 4.5 to 5.6 Mbp. The genomes show a considerable amount of diversity, and the estimated size of the current pan-genome was estimated to contain 9,433 different genes (Willenbrock et al. 2008). Several studies have identified sets of “core genes” found in most *E. coli* genomes. However, the number of these core genes tends to decrease as the full genomic sequences of new *E. coli* strains become available. The size of the *E. coli* core genome has recently been predicted to contain 1,563 genes for an infinite number of *E. coli* strains, and the number of new genes predicted from each new *E. coli* genome that is sequenced is ~79 (Willenbrock et al. 2008). In our analysis, 2,472 (60%) of the genes found present in strain 83972 were common in all the four *E. coli* genomes on the array (Fig. 3), which is well above the estimated *E. coli* core genome and also above the 2,241 common genes conserved among the 32 sequenced *E. coli* strains (Willenbrock et al. 2008). Furthermore, considering the fact that the microarray contains only four *E. coli* genomes, the total number of genes

detected present (4,109 genes) in 83972 seems reasonable comparing the size of other sequenced UTI *E. coli* genomes. The genome size of strain 83972 has been reported to be 4.9 ± 0.2 Mbp (Zdziarski et al. 2007), indicating that the strain contains roughly an additional 800 genes, not identified in the present analysis.

Of the 2,734 transcripts on the chip that are present in all the four strains represented on the microarray, 393 transcripts were below detection limit on all 21 microarrays and filtered as “absent” in strain 83972. These included 81 genes encoding hypothetical proteins. Several of the absent genes were found in clusters, many of which are involved in surface structure elements and chemotaxis. These included genes involved in flagellar biosynthesis (*flgABCDEF*, *flhABE*, *fliACDEFGHIJKLMNOPQRSTZ* and *motAB*), curli production (*csxABCEFG*), colanic acid synthesis (*wcaABCDEF* and *wza*) and chemotaxis (*cheBRWYZ* and *tap*). Other whole cluster of genes that were not expressed in the ABU strain but found in all the four *E. coli* present on the chip were *hyaBCDEF* (hydrogenase I), *hycACD* (hydrogenase 3), *tauABCD* (responsible for taurine uptake in *E. coli*) and *b1500–1505* (containing the fimbrial-like genes *ydeQRST*), as well as the *fimEAIC* genes which previously have been shown to be absent in strain 83972 (Klemm et al. 2006).

UPEC-associated genes present in strain 83972

The four UPEC isolates that have been sequenced, CFT073, UTI89, 536 and F11, contain 5,379, 5,154, 4,766 and 4,467 genes, respectively, on the chromosome. CFT073 and 536 are both O6 strains and yet show a large diversity; the genome of 536 is almost 300 kb smaller than that of CFT073 (Brzuszkiewicz et al. 2006). The genomic differences are mainly restricted to large pathogenicity islands, the additional DNA in CFT073 are genes of five cryptic prophages, which are absent in 536 (Brzuszkiewicz et al. 2006). The 427 genes that are present only in the strain 536, and the 432 genes present only in the two UPEC (compared with other sequenced *E. coli*) are scattered all over the genome (Brzuszkiewicz et al. 2006). Over 70% of the CFT073 transcripts were present in strain 83972 compared with 89% of the CFT073 transcripts found in strain 536. Figure 4 shows the homology of 16 sequenced *E. coli* and *Shigella* isolates including the three sequenced UPEC strains (UTI89, 536 and F11) pasted on the CFT073 genome; the outer, red circle in the figure shows the results from the presence/absence analysis on strain 83972. Many virulence-associated genes are located on the large pathogenicity islands (PAIs) found in different UPEC strains. The large pathogenicity island at *pheV* in CFT073 (also called PAI I_{CFT073}) encodes haemolysin (*hlyCABD*), aerobactin biosynthesis proteins (*iutA* and *iucABCD*), antigen 43

(*c3655*) and the secreted autotransporter toxin (*sat*); these were all filtered present in our analysis, suggesting that strain 83972 harbours a similar island on its chromosome. Interestingly, the aerobactin system is missing in the other three UPEC isolates. Furthermore, this PAI contains genes encoding the uropathogenic-associated P fimbriae (*pap-IBAHCDJKEFG*). The *pap* gene cluster of 83972 has been sequenced (Klemm et al. 2006); the *pap* genes are all present and show 72–100% sequence homology with the corresponding genes in CFT073. The results of the microarray analysis corresponded very well to the observed sequence homology of the different genes in the cluster (i.e. if a specific gene on the microarray is represented with probes that contain a non-homologous region compared with the corresponding gene in the hybridised sample, that gene will not hybridise and will be filtered absent); the six genes with highest sequence homology were filtered present (i.e. *pap-HCDJKF* with 98, 100, 100, 98, 99 and 95% homology, respectively) and the four with least sequence homology were filtered absent (i.e. *papIAEG* with 94, 83, 77 and 72% homology).

The employed microarray contains probes for all ten known and putative fimbriae-encoding gene clusters in CFT073. Together with the *pap* cluster, two other fimbrial clusters that have been associated with UPEC virulence are known to be present in strain 83972 and have been sequenced, i.e. the *fim* and *sfalfoc* clusters. As for the *pap* cluster, the filtering of absent genes corresponded very well to the actual presence and sequence homology of the genes; strain 83972 contains a large deletion in the *fim* cluster but shows high sequence homology with the present genes, and all the genes in the deleted part of the cluster, i.e. *fimEAIC*, were filtered absent (see blow-up in Fig. 4). Also, the *sfal/foc* cluster in 83972 shows high homology with that in CFT073 (98–100%), and eight of nine genes were filtered present; the putative regulatory gene, *sfaC*, was filtered absent. Regarding the other fimbrial clusters present on the microarray, none of the genes encoding F9 fimbriae, which appear to be common in UPEC and plays a role in biofilm formation (Ulett et al. 2007), and another putative fimbriae (*yehABCD*) were expressed and might be absent in strain 83972 (Table 1).

Presence of other pathogenicity islands in 83972

Strain 83972 seems to carry most of the pathogenicity islands of CFT073 (or PAIs similar to the ones in CFT073) according to our present/absent analysis (Table 2). The only PAI of CFT073 in which most genes (i.e. 93%) were filtered absent in strain 83972 is PAI-*pheU* (PAI II_{CFT073}), the island that contains a second *pap* cluster. The three genes filtered present in this PAI are present in several of the other sequenced *E. coli* and *Shigella* strains indicating

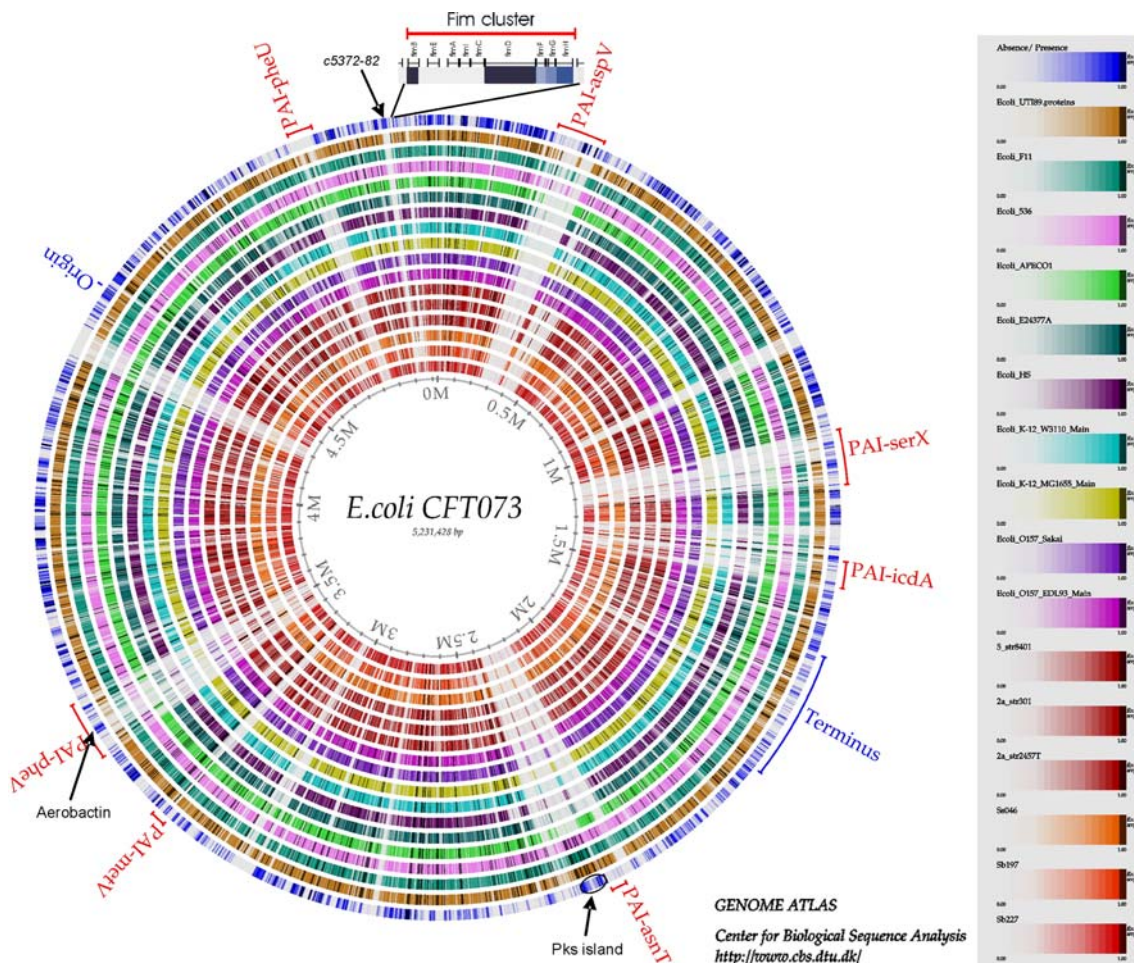


Fig. 4 BLAST atlas comparing the absent (0.0) and present (1.0) CFT073 genes in strain 83972 with other sequenced *E. coli* and *Shigella* strains, including the three sequenced UPEC isolates 536, UT189 and F11. The UPEC CFT073 genome is used as reference. The outer blue circle represents the calculated absence/presence in 83972

that these three genes not are unique/characteristic for this island wherefore this PAI is most probably absent in strain 83972.

Insertion of the high pathogenicity island (HPI) of *Yersinia pestis* has been suggested to be one of the earliest events in the evolution of extraintestinal *E. coli* strains (Welch et al. 2002). The genes of HPI encoding yersiniabactin (Ybt) were all expressed in strain 83972. The HPI genes have been found up-regulated during urine biofilm growth of 83972 indicating that Ybt-mediated iron-uptake might play an important role in biofilm growth (Hancock and Klemm 2007) and a deletion mutant in the Ybt uptake receptor (FyuA) exhibits reduced biofilm formation (Hancock et al. 2008). The HPI genes have also been found up-regulated in vivo in two of the three patients (particularly in Pat2, see Fig. 1) infected with this strain (Roos and Klemm 2006).

The *pks* island, a recently characterised and widely spread genomic island found in, for example, meningitis

strains and the uropathogenic strain CFT073, encodes a machinery for the synthesis of peptide–polyketides hybrid compounds (Nougayrede et al. 2006). The presence of the island is associated with the accumulation of double-strand DNA breaks in host cells and has genotoxic activity (Nougayrede et al. 2006). This island was expressed in strain 83972 and up-regulated in urine and in vivo (Table 4; Fig. 1). The *pks* island is widely distributed within *E. coli* phylogenetic group B2, and has been found in both pathogenic and commensal isolates; in commensal strains the cell-cycle-blocking activity might slow the turnover of the intestinal epithelium, and therefore prolong colonisation.

followed by the three UPEC isolates; the six inner circles represent *Shigella* strains. The seven PAIs of CFT073 are indicated in red. The blow-up shows the presence/absence of the *fim* cluster (c5391–5400) in strain 83972

Presence of positively selected UPEC genes

A recent paper comparing the UPEC isolates CFT073 and UT189 with six other finished *E. coli* genome sequences

Table 1 Analysis of fimbriae-encoding genes in strain 83972

Description	c number	Genes	No of genes	No (%) of absent genes	Absent
Putative chaperone-usher fimbrial operon F1C ^a	<i>c0166–0172</i>	<i>yadN-ecpD-htrE-yadMLKC</i>	7	3 (43)	<i>ecpD, yadMK</i>
F9	<i>c1237–1245</i>	<i>sfaCB-focAICDFGH</i>	9	1 (11)	<i>sfaC</i>
Putative chaperone-usher fimbrial operon	<i>c1931–1936</i>	<i>c1936-34-ydeSRQ</i>	6	6 (100)	All
Putative chaperone-usher fimbrial operon	<i>c2635–2638</i>	<i>yehABCD</i>	4	4 (100)	All
Putative chaperone-usher fimbrial operon	<i>c2878–2884</i>	<i>yfcOPQRSUV</i>	7	5 (71)	<i>yfcQRSUV</i>
P fimbriae ^a	<i>c3583–3593</i>	<i>papIBAHCDJKEFG</i>	11	4 (36)	<i>papIAEG^b</i>
Putative chaperone-usher fimbrial operon	<i>c3791–3794</i>	<i>ygiLGH-c3794</i>	4	1 (25)	<i>ygiL</i>
Auf fimbriae	<i>c4207–4214</i>	<i>aufABCDEFGF</i>	8	7 (88)	<i>aufBCDEFG</i>
P fimbriae (2) ^a	<i>c5179–5189</i>	<i>papIBAHCDJKEFG</i>	2 (<i>papAD</i>)	1 (50)	<i>papA_2</i>
Type 1 fimbriae	<i>c5391–5399</i>	<i>fimBEAICDFGH</i>	9	4 (44)	<i>fimEAIC</i>

^a Fimbrial operons present on PAIs. The two *pap* clusters share the same probes on the array with exception for *papA* and *papD*, which are represented by two separate probe sets each

^b None of these *pap* genes are absent in strain 83972, but *papIAEG* were filtered absent in the microarray analysis due to non-homologous sequence regions compared with the CFT073 *pap* probes present on the array

Table 2 Analysis of presence of pathogenicity islands in strain 83972

Island name	Common name	c number	No of genes ^a	Absent (%)	Virulence-associated genes ^b
PAI-CFT073-aspV	PAI III CFT073	<i>c0253–c0368</i>	96	40 (42)	<i>cdiA</i> (<i>c0345</i>), <i>picU</i> (<i>c0350</i>)
PAI-CFT073-serX		<i>c1165–c1293</i>	92	33 (36)	<i>mchBCDEF</i> (<i>c1227, c1229–1232</i>), <i>sfalfoc</i> (<i>c1237–c1247</i>), <i>iroNEDCB</i> (<i>c1250–c1254</i>), <i>ag43</i> (<i>c1273</i>)
PAI-CFT073-icdA		<i>c1518–c1601</i>	42	5 (12)	<i>sitDCBA</i> (<i>c1597–1600</i>)
PAI-CFT073-asnT	HPI CFT073	<i>c2418–c2437</i>	19	3 (16)	<i>fyuA</i> (<i>1246</i>)
PAI-CFT073-metV		<i>c3385–c3410</i>	26	17 (65)	<i>hcp</i> (<i>c3391</i>), <i>clpB</i> (<i>c3392</i>)
PAI-CFT073-pheV	PAI I CFT073	<i>c3556–c3698</i>	119	51 (43)	<i>hlyA</i> (<i>c3570</i>), <i>pap</i> (<i>c3582–c3593</i>), <i>iha</i> (<i>c3610</i>), <i>sat</i> (<i>c3619</i>), <i>iutA</i> , <i>iucDCBA</i> (<i>c3623–3628</i>), <i>ag43</i> (<i>c3655</i>), <i>kpsTM</i> (<i>c3697–c3698</i>)
PAI-CFT073-pheU	PAI II CFT073	<i>c5143–c5216</i>	46	43 (93)	<i>pap2</i> (<i>c5179–c5189</i>)

^a No of genes in the PAI that were present on the array with unique probes (i.e. genes that are not orthologues to any other *E. coli* transcripts present on the array)

^b Boldface indicates genes filtered present in strain 83972

presented 29 genes that are under positive selection only in UPEC strains (Chen et al. 2006). These 29 genes are involved in various aspects of cell surface structure, DNA metabolism, nutrient acquisition and UTI. Of these 29 genes, 25 were filtered present in our ABU strain 83972; many of these genes are represented by more than one transcript on the array due to sequence differences among the four strains present on the array, in all cases the gene filtered present in 83972 corresponded to the CFT073 transcript. Four genes were filtered absent, *agal*, *yjiL*, *recC* and *yegO*; they encode a putative galactosamine-6-phosphate isomerase, a hypothetical protein, exodeoxyribonuclease V gamma subunit and a hypothetical transport protein,

respectively. The genes in the two COG categories that were significantly enriched in the two UPEC strains, i.e. “cell wall/membrane biogenesis” (*amiA*, *cutE*, *fepE*, *ompC*, *ompF* and *yfaL*) and “secondary metabolites biosynthesis, transport and metabolism” (*entD*, *entF* and *yojI*) (Chen et al. 2006), were all present in strain 83972.

Functional analysis of MG1655 transcripts of ABU *E. coli* 83972

To gain more information concerning what type of genes were absent, the MG1655 genes were grouped into functional categories defined by the clusters of orthologous

groups (COGs) of proteins (Tatusov et al. 1997). Previous studies have, in attempts to identify essential genes and the *E. coli* core genome, found that groups with genes involved in metabolism and various cellular processes (excluding cell motility) contain a substantially higher percentage of conserved and essential genes, while COGs with genes of unknown function and external origin as well as genes involved in signalling and motility contain fewer essential genes (Anjum et al. 2003; Gerdes et al. 2003). Classification of the absent genes of strain 83972 revealed that the groups “cell motility”, “defence mechanisms” and “not in COGs” had a significant overrepresentation of absent genes (Table 3). A significantly lower proportion of absent genes were found in the groups: “cell cycle control”, “posttranslational modification” and “translation”. This is in agreement with a previously published study of pathogenic *E. coli*; Anjum et al. (2003) studied 26 strains of *E. coli* and found that the two groups with largest proportion of absent genes were “not in COGs” and “cell motility”, while the six groups with the lowest proportion of absent genes were “translation”, “cell division”, “posttranslational modification”, “coenzyme metabolism”, “nucleotide transport and metabolism” and “energy production and conversion”, which all, with exception for the last group, contained

significantly fewer absent genes in strain 83972 (Table 3). This suggests that strain 83972 utilises a similar set of core genes as other *E. coli* strains.

CFT073 genes absent in strain 83972

There were 1,636 CFT073 genes that could not be detected according to our expression profiling in ABU strain 83972; 961 of these genes are exclusively found in CFT073, i.e. not present in the other three strains represented on the array. The majority, 645 genes, corresponded to hypothetical, putative or unknown proteins. Considering the very different patient symptom profiles of strains CFT073 and 83972 (one being a true pathogen, while the latter is a commensal-like strain), genes that are present in UPEC isolate CFT073 but not expressed in ABU strain 83972 can be considered as virulence factor candidates. However, most genes associated with UPEC pathogenesis were expressed in strain 83972 and up-regulated during growth in urine, e.g. all iron-related genes encoding uptake and transport of aerobactin, salmochelin, yersiniabactin and haem/haemoglobin (Table 4). Two exceptions were the *ireA* gene encoding an iron-regulated outer-membrane protein that was filtered

Table 3 Distribution of absent genes in functional categories

Functional category	Absent		Total	Z test P value
	No.	%		
Amino acid transport and metabolism	98	33.1	296	0.636
Carbohydrate transport and metabolism	126	42.3	298	0.025
Cell cycle control, cell division and chromosome partitioning	5	16.1	31	0.000
Cell motility	77	84.6	91	0.000
Cell wall/membrane/envelope biogenesis	81	40.5	200	0.087
Coenzyme transport and metabolism	25	23.1	108	0.001
Defense mechanisms	17	48.6	35	0.000
Energy production and conversion	88	36.7	240	0.563
Function unknown	61	24.7	247	0.003
General function prediction only	83	30.9	269	0.255
Inorganic ion transport and metabolism	55	34.4	160	0.921
Intracellular trafficking, secretion and vesicular transport	8	22.2	36	0.000
Lipid transport and metabolism	21	29.6	71	0.129
Nucleotide transport and metabolism	19	24.4	78	0.002
Posttranslational modification, protein turnover, chaperones	24	20.2	119	0.000
Replication, recombination and repair	70	42.4	165	0.023
Secondary metabolites biosynthesis, transport and catabolism	22	40.7	54	0.075
Signal transduction mechanisms	39	33.6	116	0.748
Transcription	78	33.2	235	0.654
Translation, ribosomal structure and biogenesis	21	13.5	156	0.000
Not in COGs	577	54.2	1065	0.000
	1,595	39.2	4,070	

Table 4 Characteristics of ABU isolate 83972 compared with UPEC isolates CFT073, UTI89 and 536

Characteristic ^a	CFT073	UTI89	536	83972	Expression in 83972 ^b
Serotype	O6	O18	O6	O? ^c	
Capsule	K2	K1	K15	K? ^c	
Chu	+	+	+	+	U, BF, Pat
Ent	+	+	+	+	U, BF, Pat
Fep	+	+	+	+	PI, U, BF, Pat
Feo	+	+	+	+	BF, Pat
Fhu	+	+	+	+	PI, U, BF, Pat
Iro	+	+	+	+	U, Pat
Iuc	+	–	–	+	PI, U, BF, Pat
IutA	+	–	–	+	U, BF, Pat
Sit	+	+	+	+	U, BF, Pat
FyuA	+	+	+	+	U, BF, Pat
Iha	+	–	–	+	U, Pat
IreA	+	–	–	–	
Pks island	+	+	+	+	U, Pat
RfaH	+	+	+	+	BF
D-serine	+	+	+	+	Pat
Pap	+	+	+	–	
Fim	+	+	+	–	
Foc/sfa	+	+	+	–	
Vat	+	+	+	+	BF, Pat
Sat	+	–	–	+	U
Tsx	+	+	+	–	
Biofilm formation	1.0		1.3	14.4	

^a Boldface indicates genes that were filtered absent in strain 83972

^b Up-regulation in urine (U), biofilm (BF), plates (PI) and patients (Pat) compared with MOPS minimal medium

^c Undefined. Extensive electron microscopy analysis of the strain has never reported any capsule

absent as well as the *tsx* gene encoding a nucleoside-binding outer-membrane protein. Although the *tsx* gene has not previously been associated with UPEC virulence, it has just recently been identified together with more well-known UPEC genes as involved in movement from the intestinal tract to the bladder and vagina (i.e. occurred significantly more often in multiple-site isolates than in rectal site-only isolates) (Xie et al. 2006); furthermore, *Tsx* was also recently identified together with 22 other outer-membrane proteins from CFT073 cells grown under conditions mimicking the urinary tract (Hagan and Mobley 2007).

Type IV fimbriae are assembled by the type II general secretory pathway. They occur in a wide range of species and frequently are associated with diseases. The *ppdD* and *hofBC* genes (*b0106–0108*), which encode type IV prepilin

and are present in CFT073, EDL933 and MG1655, were filtered absent in strain 83972.

CFT073 genes present in strain 83972 but not found in other UPEC strains

The majority of the genes that are absent in the other three UPEC isolates (i.e. 536, UTI89 and F11) were filtered absent in strain 83972 as well (gaps in Fig. 4). However, there are a few exceptions where a gene that is not found in any of the other UPEC strains is filtered present in strain 83972. The aerobactin system belongs to one of the exceptions, indicating that strain 83972 is particularly well equipped with iron uptake systems. The other exceptions are all but one located on PAIs and they all encode hypothetical proteins: *c1194–c1204* (on PAI-*serX*), *c1522–c1528* (on PAI-*icdA*), *c3394–c3396* (on PAI-*metV*), *c3681–c3682* (on PAI-*pheV* where the aerobactin genes also are found) and *c5372–c5382*. *c3394–c3396* and *c5372–c5382* are not present in any of the 16 sequenced *E. coli* and *Shigella* strains represented in Fig. 4, indicating that some genes unique to CFT073 can be found in strain 83972 as well.

Discussion

Bacterial genomes are under constant change. New genes are acquired by horizontal transfer and old ones are lost by mutations. It is generally believed that commensal *E. coli* can become pathogenic through the acquisition of novel genes encoding virulence factors and niche-adaptation factors (Kaper et al. 2004). In contrast to organisms that have acquired genes for pathogenesis, *E. coli* 83972 is an example of an organism that has adapted to a commensal-like existence through gene deletions and point mutations. Using primarily the CFT073 as a scaffold, we used presence/absence data from seven sets of different gene expression profiles (in total 21 microarrays) to model the gene pool of strain 83972. Given the limitations of the approach, i.e. genes not present on the employed chip have been ignored, a substantial body of information was gathered concerning the genomic content of the strain. As it turned out the strain was highly similar to CFT073; 96% (3,959) of the genes found to be expressed on the employed microarray by 83972 are also found in CFT073, and genes on six of the seven pathogenicity islands of CFT073 were expressed by 83972; furthermore, CFT073 genes not found in any other UPEC isolate were expressed by 83972. An estimated ~900 CFT073 genes are not expressed by 83972. Arguably, in the light of the difference in patient symptoms invoked by encounters with the two strains, this list represents virulence gene candidates.

Although strain 83972 seems to be a deconstructed uropathogen and does not provoke symptoms in the human host it grows fast in urine and is an excellent colonizer of the human bladder (Roos and Klemm 2006; Roos et al. 2006b; Klemm et al. 2007). It can do so because it has kept a large assortment of fitness factors required for this particular ecological niche. Among the genes expressed under realistic environmental conditions such as in the human bladder are candidates for fitness factor genes, e.g. the many iron acquisitions systems expressed by the strain and many genes involved in sugar acid and amino acid metabolism. Interestingly, many of the known and putative virulence factors of the urinary tract are expressed by strain 83972 and might therefore be considered as fitness factors rather than virulence factors; these include 25 of 29 positively selected UPEC genes as well as the newly characterised *pks* island inducing breaks in double-stranded DNA in host cells. Also, virulence-associated genes such as *cdiA*, *mchBCDEF*, *flu*, *hcp*, *rfaH*, *sat*, *picU* and *vat* were all expressed by strain 83972. Very few of the known or putative virulence factors were absent in (or not expressed by) strain 83972. The *pap*, *fim* and *foc/sfa* clusters encoding UPEC-class fimbriae are dysfunctional in strain 83972 and the *clpB*, *ireA* and *tsx* genes were not expressed in the ABU strain. These stand out as potential virulence candidates together with a number of uncharacterised genes encoding hypothetical proteins.

Thus from the analyses performed here we can make predictions about several gene categories such as potential virulence genes, fitness genes and “household-class” genes. It is also noteworthy that the information reported herein complements a potential genome sequence of strain 83972. Whole genome sequencing can identify the presence of genes but is unable to reveal if they are transcribed. Genes can be silenced not only due to lesions in the actual gene and its promoter but also due to mutations of genes encoding regulatory factors. The methodology employed in the present work reveals the active genome of strain 83972.

ABU strain 83972 is closely related to fully virulent uropathogenic strains. All evidences suggest that the strain is a deconstructed pathogen. This study dispels the commonly held idea that ABU strains are commensals that have picked up niche-adaptation genes by horizontal gene transfer. Rather, strain 83972 was originally a true pathogenic strain that has lost whole or part of operons that contribute to virulence.

Acknowledgments This work was supported by grants from the Danish Medical Research Council (271-06-0555), Lundbeckfonden and Inlaks Foundation, India.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Andersson P, Engberg I, Lidin-Janson G, Lincoln K, Hull R, Hull S, Svanborg C (1991) Persistence of *Escherichia coli* bacteriuria is not determined by bacterial adherence. *Infect Immun* 59:2915–2921
- Anjum MF, Lucchini S, Thompson A, Hinton JC, Woodward MJ (2003) Comparative genomic indexing reveals the phylogenomics of *Escherichia coli* pathogens. *Infect Immun* 71:4674–4683
- Blanco M, Blanco JE, Alonso MP, Blanco J (1996) Virulence factors and O groups of *Escherichia coli* isolates from patients with acute pyelonephritis, cystitis and asymptomatic bacteriuria. *Eur J Epidemiol* 12:191–198
- Brzuszkiewicz E, Brüggemann H, Liesegang H, Emmerth M, Ölschläger T, Nagy G, Albermann 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 USA* 103:12879–12884
- 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 JI (2006) Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proc Natl Acad Sci USA* 103:5977–5982
- Dobrindt U, Agerer F, Michaelis K, Janka A, Buchrieser C, Samuelson M, Svanborg C, Gottschalk G, Karch H, Hacker J (2003) Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J Bacteriol* 185:1831–1840
- Eden CS, Hanson LA, Jodal U, Lindberg U, Akerlund AS (1976) Variable adherence to normal human urinary-tract epithelial cells of *Escherichia coli* strains associated with various forms of urinary-tract infection. *Lancet* 1:490–492
- Foxman B (2002) Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am J Med* 113(Suppl 1A):5S–13S
- Funfstuck R, Tschape H, Stein G, Kunath H, Bergner M, Wessel G (1986) Virulence properties of *Escherichia coli* strains in patients with chronic pyelonephritis. *Infection* 14:145–150
- Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini A, Sawitzki G, Smith C, Smyth G, Tierney L, Yang J, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80
- Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D’Souza M, Baev MV, Grechkin Y, Meeh F, Fonstein MY, Overbeek R, Barabasi AL, Oltvai ZN, Osterman AL (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* 185:5673–5684
- Hagan EC, Mobley HL (2007) Uropathogenic *Escherichia coli* outer membrane antigens expressed during urinary tract infection. *Infect Immun* 75:3941–3949
- Hancock V, Ferrières L, Klemm P (2008) The ferric yersiniabactin uptake receptor FyuA is required for efficient biofilm formation by urinary tract infectious *Escherichia coli* in human urine. *Microbiology* 154:167–175
- Hancock V, Klemm P (2007) Global gene expression profiling of asymptomatic bacteriuria *Escherichia coli* during biofilm growth in human urine. *Infect Immun* 75:966–976
- Hull R, Rudy D, Donovan W, Svanborg C, Wieser I, Stewart C, Darouiche R (2000) Urinary tract infection prophylaxis using *Escherichia coli* 83972 in spinal cord injured patients. *J Urol* 163:872–877

- Johnson JR (1991) Virulence factors in *Escherichia coli* urinary tract infection. *Clin Microbiol Rev* 4:80–128
- Johnson JR, Kuskowski MA, Gajewski A, Soto S, Horcajada JP, Jimenes de Anta MT, Vila J (2005) Extended virulence genotypes and phylogenetic background of *Escherichia coli* isolates from patients with cystitis, pyelonephritis, or prostatitis. *J Infect Dis* 191:46–50
- Kaper JB, Nataro JP, Mobley HL (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2:123–140
- Klemm P, Roos V, Ulett GC, Svanborg C, Schembri MA (2006) Molecular characterization of the *Escherichia coli* asymptomatic bacteriuria strain 83972: the taming of a pathogen. *Infect Immun* 74:781–785
- Klemm P, Hancock V, Schembri MA (2007) Mellowing out: adaptation to commensalism by *Escherichia coli* asymptomatic bacteriuria strain 83972. *Infect Immun* 75:3688–3695
- Lindberg U, Hanson LA, Jodal U, Lidin-Janson G, Lincoln K, Olling S (1975) Asymptomatic bacteriuria in schoolgirls. II. Differences in *Escherichia coli* causing asymptomatic bacteriuria. *Acta Paediatr Scand* 64:432–436
- Lloyd AL, Rasko DA, Mobley HL (2007) Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. *J Bacteriol* 189:3532–3546
- Marrs CF, Zhang L, Foxman B (2005) *Escherichia coli* mediated urinary tract infections: are there distinct uropathogenic *E. coli* (UPEC) pathotypes? *FEMS Microbiol Lett* 252:183–190
- Nougayrede J-P, 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
- Orskov I, Svanborg Eden C, Orskov F (1988) Aerobactin production of serotyped *Escherichia coli* from urinary tract infections. *Med Microbiol Immunol (Berl)* 177:9–14
- 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
- Roos V, Schembri MA, Ulett GC, Klemm P (2006a) Asymptomatic bacteriuria *Escherichia coli* strain 83972 carries mutations in the *foc* locus and is unable to express F1C fimbriae. *Microbiology* 152:1799–1806
- Roos V, Ulett GC, Schembri MA, Klemm P (2006b) The asymptomatic bacteriuria *Escherichia coli* strain 83972 out-competes UPEC strains in human urine. *Infect Immun* 74:615–624
- Stenqvist K, Sandberg T, Lidin-Janson G, Orskov F, Orskov I, Svanborg-Eden C (1987) Virulence factors of *Escherichia coli* in urinary isolates from pregnant women. *J Infect Dis* 156:870–877
- Sturm A, Schierhorn A, Lindenstrauss U, Lillie H, Bruser T (2006) YcdB from *Escherichia coli* reveals a novel class of Tat-dependently translocated hemoproteins. *J Biol Chem* 281:13972–13978
- Sundén F, Håkansson L, Ljunggren E, Wullt B (2006) Bacterial interference—is deliberate colonization with *Escherichia coli* 83972 an alternative treatment for patients with recurrent urinary tract infection? *Int J Antimicrob Agents* 28S:S26–S29
- Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631–637
- Ulett GC, Mabbett AN, Fung KC, Webb RI, Schembri MA (2007) The role of F9 fimbriae of uropathogenic *Escherichia coli* in biofilm formation. *Microbiology* 153:2321–2331
- Welch RA, Burland V, Plunkett G 3rd, 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 USA* 99:17020–17024
- Willenbrock H, Petersen A, Sekse C, Kiil K, Wasteson Y, Ussery DW (2006) Design of a seven-genome *Escherichia coli* microarray for comparative genomic profiling. *J Bacteriol* 188:7713–7721
- Willenbrock H, Hallin PF, Wassanar TM, Ussery DW (2008) Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray. *Genome Biol* 8:R267
- Vranes J, Kruzic V, Sterk-Kuzmanovic N, Schonwald S (2003) Virulence characteristics of *Escherichia coli* strains causing asymptomatic bacteriuria. *Infection* 31:216–220
- Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F (2004) A model based background adjustment for oligonucleotide expression arrays. *JASA* 99:909–917
- Xie J, Foxman B, Zhang L, Marrs CF (2006) Molecular epidemiologic identification of *Escherichia coli* genes that are potentially involved in movement of the organism from the intestinal tract to the vagina and bladder. *J Clin Microbiol* 44:2434–2441
- Zdziarski J, Svanborg C, Wullt B, Hacker J, Dobrindt U (2007) Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? *Infect Immun* 76:695–703