Characterization of genomic diversity in extraintestinal pathogenic *Escherichia coli* (ExPEC) and development of a diagnostic DNA microarray for the differentiation of ExPEC isolates causing urinary tract infections

Johanna Sorsa

University of Helsinki

Faculty of Biosciences

Department of Biological and Environmental Sciences

General Microbiology

and

University of Munich

Faculty of Medicine

Max von Pettenkofer-Institut

für Hygiene und medizinische Mikrobiologie

Department of Bacteriology

Academic Dissertation in General Microbiology

To be presented, with the permission of the Faculty of Biosciences, Department of Biological and Environmental Sciences, University of Helsinki, for public criticism in the Auditorium 1041, Viikinkaari 5, on the 13th of April, at 12 o´clock noon.

Helsinki 2007

Superviced by Assistant Professor Sören Schubert, PhD
Max on Pettenkofer-Instiut für Hygiene und medizinische Mikrobiologie
University of Munich
Germany

Reviewed by University lecturer Benita Westerlund-Wikström, PhD
Department of Biological and Environmental Sciences
University of Helsinki
Finland

Professor Anja Siitonen, PhD
Department of Bacterial and Inflammatory Diseases
National Public Health Institute
Helsinki, Finland

Opponent Assistant Professor Ulrich Dobrindt, PhD
Institut für Molekulare Infektionsforschung
University of Würzburg
Germany

ISBN 978-952-92-1787-8 (paperback)
ISBN 978-952-10-3809-9 (pdf)
http://ethesis.helsinki.fi

Yliopistopaino, Helsinki 2007

Cover: Bavaria, the patron saint of Bavaria.

There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.

Charles Darwin, The Origin of Species

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This project is composed of four separate studies (Studies I to IV), which are reviewed in the original publications listed below (Reprinted with kind permission from publishers). Some unpublished data is also included.

Study I **Sorsa, L. J.**, Dufke, S., Heesemann, J., Schubert, S.
Characterization of an *iroBCDEN* gene cluster on a transmissible plasmid of uropathogenic *Escherichia coli*: evidence for horizontal transfer of a chromosomal virulence factor.
Infection and Immunity 2003; 71:3285-3293.

Study II **Sorsa, L. J**., Dufke, S., Schubert, S.
Identification of novel virulence-associated loci in uropathogenic *Escherichia coli* by suppression subtractive hybridization.
FEMS Microbiological Letters 2004; 230:203-208.

Study III **Sorsa, L. J.**, Feldmann, F., Hildinger, K., Dufke, S., Schubert, S. Characterization of four novel genomic regions of uropathogenic *Escherichia coli* highly associated with the extraintestinal virulent phenotype: a jigsaw puzzle of genetic elements.
International Journal of Medical Microbiology, February 2007.

Study IV **Sorsa, L. J.**, Müller, T., Feldmann, F., Hildinger, K., Schubert, S.
Development and validation of a diagnostic DNA microarray specific for extraintestinal pathogenic *Escherichia coli* (ExPEC).
Manuscript in preparation.

ABBREVIATIONS

AA	Amino acid
Afa	AfaDE family of adhesins
Afa/Dra	Afa/Dr family of adhesins
APEC	Avian pathogenic *E. coli*
Auf	Auf fimbriae
cDNA	Complementary-DNA
CDT	Cytolethal distending toxin
ChuA	*E. coli* heme-utilization, outer membrane hemin receptor
CNF	Cytotoxic necrotizing factor
Cva	Colicin
Dra	Dr fimbriae
EBI	the European Bioinformatics Institute
EHEC	Enterohemorrhagic *E. coli*
Ent	Enterobactin siderophore system
EPEC	Enteropathogenic *E. coli*
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extraintestinal pathogenic *E. coli*
FepA	Siderophore enterobactin receptor
Fig.	Figure
Fim	Type I fimbriae
Foc	F1C fimbriae
FyuA	Siderophore yersiniabactin receptor
GI	Genomic island
Hbp	Hemoglobin-protease
Hek	Hemagglutinin from *E. coli* K1
Hly	Hemolysin
Hra	Heat-resistant agglutinin
IC	Intermediate cell
Ig	Immunoglobulin
Iha	IrgA-homologue adhesin
IL	Interleukin
IrgA	Putative protein encoded by an iron -egulated gene
IreA	Iron-responsive element
Iro	Salmochelin siderophore system
IroN	Siderophore salmochelin receptor
Irp	Yersiniabactin siderophore system
IS	Insertion sequence
Iss	Protein involved in increased serum resistence and survival
Iuc	Aerobactin s siderophore systen
IutA	Siderophore aerobactin receptor
Kb	Kilo base pair
LPS	Lipopolysaccharide
Mb	Mega base pair
Mch	Microcin
ModD	Putative iron transport protein

n	Number of sequences
NCBI	the National Center for Biotechnological Information
NMEC	Newborn meningitis associated *E. coli*
nt	Nucleotide
OmpT	Outer membrane protein T
ORF	Open reading frame
PAI	Pathogenicity island
Pap	P fimbriae
PCR	Polymerase chain reaction
PG	Phylogenetic group
Pic	Putative serine protease autotransporter protein
PMN	Polymorphonuclear leukocyte
PrrA	Putative siderophore receptor
Uko	Unknown ORF
UPEC	Uropathogenic *E. coli*
Sat	Serine auto-transporter toxin
Scr	Sucrose specific phosphotransferase system
Sfa	S fimbriae
sp.	Species
spp.	Plural form of sp.
TLR	Toll-like receptor
TonB	Ferri-siderophore complex transport protein
tp^r	Trimethoprim resistance cassette
tRNA	Transfer-RNA
Tsh	Putative heme-binding serine protease
SFC	Superficial facet cell
Usp	Uropathogen specific protein
UTI	Urinary tract infection
VGA	Virulence gene algorithm
VP	Extraintestinal virulence potential
Yc73	Putative protein involved in iron transport

SUMMARY

Extraintestinal pathogenic *Escherichia coli* (ExPEC) represent a diverse group of strains of *E. coli*, which infect extraintestinal sites, such as the urinary tract, the bloodstream, the meninges, the peritoneal cavity, and the lungs (Russo and Johnson, 2003). Urinary tract infections (UTIs) caused by uropathogenic *E. coli* (UPEC), the major subgroup of ExPEC, are among the most prevalent microbial diseases world wide and a substantial burden for public health care systems (Foxman and Brown, 2003; Russo and Johnson, 2003). UTIs are responsible for serious morbidity and mortality in the elderly, in young children, and in immune-compromised and hospitalized patients (Bagshaw and Laupland, 2006; Foxman and Brown, 2003; Scholes *et al.*, 2005; Tambyah *et al.*, 2002).

ExPEC strains are different, both from genetic and clinical perspectives, from commensal *E. coli* strains belonging to the normal intestinal flora and from intestinal pathogenic *E. coli* strains causing diarrhea (Caprioli *et al.*, 2005; Johnson and Russo, 2005; Nataro, 2005). ExPEC strains are characterized by a broad range of alternate virulence factors, such as adhesins, toxins, and iron accumulation systems (Hochhut *et al.*, 2005; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Russo and Johnson, 2003). Unlike diarrheagenic *E. coli*, whose distinctive virulence determinants evoke characteristic diarrheagenic symptoms and signs, ExPEC strains are exceedingly heterogeneous and are known to possess no specific virulence factors or a set of factors, which are obligatory for the infection of a certain extraintestinal site (e. g. the urinary tract; Caprioli *et al.*, 2005; Clarke *et al.*, 2003; Johnson *et al.*, 2005b; Johnson *et al.*, 2005d; Nataro, 2005; Welch *et al.*, 2002).

The ExPEC genomes are highly diverse mosaic structures in permanent flux (Dobrindt *et al.*, 2003). These strains have obtained a significant amount of DNA (predictably up to 25% of the genomes) through acquisition of foreign DNA from diverse related or non-related donor species by lateral transfer of mobile genetic elements, including pathogenicity islands (PAIs), plasmids, phages, transposons, and insertion elements (IS; Dobrindt *et al.*, 2004; Dobrindt and Hacker, 2001; Fischer and Eisenberg, 1999; Ochman *et al.*, 2005; Pal *et al.*, 2005). The ability of ExPEC strains to cause disease is mainly derived from this horizontally acquired gene pool; the extragenous DNA facilitates rapid adaptation of the pathogen to changing conditions and hence the extent of the spectrum of sites that can be infected (Escobar-Paramo *et al.*, 2004; Hacker and Carniel, 2001; Lerat *et al.*, 2005; Ochman *et al.*, 2005). However, neither the amount of unique DNA in different ExPEC strains (or UPEC strains) nor the mechanisms lying behind the observed genomic mobility are known.

Due to this extreme heterogeneity of the UPEC and ExPEC populations in general, the routine surveillance of ExPEC is exceedingly difficult. In this project, we presented a novel virulence gene algorithm (VGA) for the estimation of the extraintestinal virulence potential (VP, pathogenicity risk) of clinically relevant ExPECs and fecal *E. coli* isolates. The VGA was based on a DNA microarray specific for the ExPEC phenotype (ExPEC pathoarray). This array contained 77 DNA probes homologous with known (e.g. adhesion factors, iron accumulation systems, and toxins) and putative (e.g. genes predictably involved in adhesion, iron uptake, or in metabolic functions) ExPEC virulence determinants. In total, 25 of DNA probes homologous with known virulence factors and 36 of DNA probes representing putative extraintestinal virulence determinants were found at significantly higher frequency in virulent ExPEC isolates than in commensal *E. coli* strains. We showed that the ExPEC pathoarray and the VGA could be readily used for the differentiation of highly virulent ExPECs both from less virulent ExPEC clones and from commensal *E. coli* strains as well. Implementing the VGA in a group of unknown ExPECs (n=53) and fecal *E. coli* isolates (n=37), 83% of strains were correctly identified as extraintestinal virulent or commensal *E. coli*. Conversely, 15% of clinical ExPECs and 19% of fecal *E. coli* strains failed to raster into their respective pathogenic and non-pathogenic groups. Clinical data and virulence gene profiles of these strains warranted the estimated VPs; UPEC strains

with atypically low risk-ratios were largely isolated from patients with certain medical history, including *diabetes mellitus* or catheterization, or from elderly patients. In addition, fecal *E. coli* strains with VPs characteristic for ExPEC were shown to represent the diagnostically important fraction of resident strains of the gut flora with a high potential of causing extraintestinal infections.

Interestingly, a large fraction of DNA probes associated with the ExPEC phenotype corresponded to novel DNA sequences without any known function in UTIs and thus represented new genetic markers for the extraintestinal virulence. These DNA probes included unknown DNA sequences originating from the genomic subtractions of four clinical ExPEC isolates as well as from five novel cosmid sequences identified in the UPEC strains HE300 and JS299. The characterized cosmid sequences (pJS332, pJS448, pJS666, pJS700, and pJS706) revealed complex modular DNA structures with known and unknown DNA fragments arranged in a puzzle-like manner and integrated into the common *E. coli* genomic backbone. Furthermore, cosmid pJS332 of the UPEC strain HE300, which carried a chromosomal virulence gene cluster (*iroBCDEN*) encoding the salmochelin siderophore system, was shown to be part of a transmissible plasmid of *Salmonella enterica*. Taken together, the results of this project pointed towards the assumptions that first, (i) homologous recombination, even within coding genes, contributes to the observed mosaicism of ExPEC genomes and secondly, (ii) besides en block transfer of large DNA regions (e.g. chromosomal PAIs) also rearrangements of small DNA modules provide a means of genomic plasticity. The data presented in this project supplemented previous whole genome sequencing projects of *E. coli* (Blattner *et al.*, 1997; Chen *et al.*, 2006; Hayashi *et al.*, 2006; Perna *et al.*, 2001; Welch *et al.*, 2002) and indicated that each *E. coli* genome displays a unique assemblage of individual mosaic structures, which enable these strains to successfully colonize and infect different anatomical sites.

1. INTRODUCTION

1.1 The clinical background of urinary tract infections (UTIs) caused by extraintestinal pathogenic *Escherichia coli* (ExPEC)

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are a diverse group of strains infecting extraintestinal locations, including the urinary tract (infected by uropathogenic *E. coli* [UPEC]), the bloodstream (infected by sepsis-associated pathogenic *E. coli*), and the meninges of the neonates (infected by newborn meningitis-associated *E. coli* [NMEC]; for recent review see Russo and Johnson, 2003). Urinary tract infections (UTIs) are one of the most frequently acquired bacterial infections world wide and UPEC, the major subgroup of ExPEC, accounts for around 90% of all ambulatory UTIs (Foxman and Brown, 2003). Approximately half of the adult women will develop UTI during their lifetime, one fourth of the women will experience one or more recurrent infections after the first UTI episode, and about 5% of the women suffer of persistent infections, which significantly interfere with the daily life. UTIs are responsible for serious morbidity and mortality in the elderly, in young children, and in immune-compromised patients (Foxman and Brown, 2003; Scholes *et al.*, 2005). Taken together, UTI episodes greatly burden the public health care systems and the national economies due to lost hours of work (Russo and Johnson, 2003).

A UTI is clinically defined as the presence of a significant number of pathogenic organisms in the urine (most often 10^5 colony forming units/ml urine). The symptom palette extends from asymptomatic bacteriuria to painful and frequent urination, bloody urine, abdominal pain, nausea, vomiting, and fever (Stamm, 1982). Symptomatic UTI can be divided into cystitis and pyelonephritis. Cystitis is an infection of the lower urinary tract and the urinary bladder (Fig. 1). Pyelonephritis is a life threatening infection affecting the kidneys (Fig. 1). If pyelonephritis is untreated, it often leads to renal failure and systemic spread of the pathogen into the blood stream (i.e. urosepsis; Fig. 1; Faro and Fenner, 1998; Mehnert-Kay, 2005).

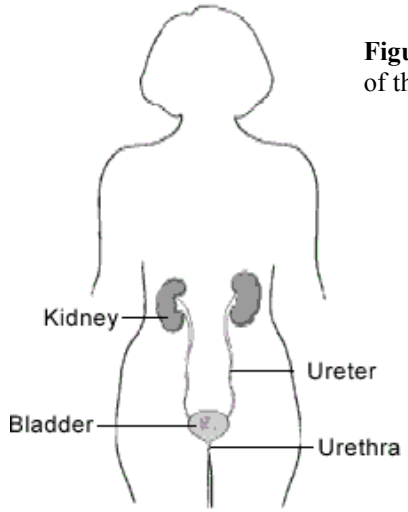

Figure 1. The urinary tract. The urinary system consists of the kidneys, the ureter, the bladder, and the urethra.

ExPEC strains are different regarding both their genetic backgrounds and clinical outcomes from commensal *E. coli* strains of the normal intestinal flora and from intestinal pathogenic *E. coli* strains causing intestinal disease (Dobrindt *et al.*, 2003; Johnson and Russo, 2005). Intestinal pathogenic *E. coli* strains possess distinctive virulence factors characteristic of their particular pathomechanisms and pathogroup (e.g. shiga-toxins of enterohemorrhagic strains [EHEC], intimin of

enteropathogenic strains [EPEC] and heat-labile and –stable toxins of enterotoxigenic [ETEC] strains). These determinants are also responsible for the resulting diarrheagenic symptoms (Caprioli *et al.*, 2005; Clarke *et al.*, 2003; Farthing, 2004). In contrast, ExPEC exhibit a broad range of virulence factors that enable these strains to colonize mucosal surfaces, avoid or subvert local and systemic host defense mechanisms, scavenge essential nutrients (e.g. iron), injure and invade the host, and stimulate a noxious inflammatory response (Bower *et al.*, 2005; Russo and Johnson, 2003). ExPEC strains colonize effectively the gastrointestinal tract; they constitute a predominant fraction (20%) of fecal *E. coli* but do not cause any gastrointestinal disease (Johnson *et al.*, 1998; Murray *et al.*, 2004; Nowrouzian *et al.*, 2005; Siitonen, 1992). To date, no single virulence factor has been shown to be unique or restricted to the ExPEC phenotype. Rather, complementary sets of virulence factors interplay to direct bacteria into infectious pathways that result in disease in a susceptible host (Bower *et al.*, 2005; Seed and Hultgren, 2005). It is still largely uncertain, whether ExPECs are a group of pathogens capable of colonizing various anatomical sites, or whether each member is equipped with a particular set of virulence determinants allowing it to infect distinct body sites by definite pathomechanisms (e.g. the UPEC pathotype; Johnson and Russo, 2002; Marrs *et al.*, 2005).

1.2 Bacterial factors contributing to the development and establishment of UTIs caused by ExPEC

1.2.1 Known fimbrial adhesins

The bladder is a difficult organ to colonize and infect. The intruding pathogen must first withstand the powerful clearing actions of the urine flow and secondly, the active exfoliation of the epithelial cells following bacterial attachment (Williams and Schaeffer, 2004). In addition, the pathogen has to avoid or resist rapid and highly effective inflammatory responses that are evoked locally (Chowdhury *et al.*, 2004). Up to date, a number of bacterial virulence factors have been characterized (e.g. adhesins, iron accumulation systems, and toxins), which allow ExPEC to infect distinct body sites (Table 1; Johnson and Russo, 2005; Marrs *et al.*, 2005).

Despite of the rapidly accumulating epidemiological and experimental data, the ultimate factor(s) required by UPEC and other ExPEC, as well as the underlying molecular mechanisms required for initiation and establishment of an extraintestinal infection are still unclear. As virulence determinants in *E. coli* are controlled by global regulators (e.g. transcriptional regulators RfaH, DsdC, DegS, and YbtA), which act in concert under environmental stimuli (e.g. iron deficiency), the causal relationship between the genotype and the pathotype is often ambiguous (Anisimov *et al.*, 2005; Nagy *et al.*, 2002; Redford *et al.*, 2003; Roesch *et al.*, 2003). UPEC strains exhibit a number of adhesins, which allow these pathogens to attach to the urinary tract tissues. Many of the UPEC adhesins have multiple roles; they are involved in adhesion, invasion, cell cycle control, and modulation of the inflammatory reactions. Adhesins can be classified as either fimbrial or afimbrial depending upon whether or not the adhesin is displayed as part of the pilus or the fimbriae. Known fimbrial adhesins implicated in the urinary tract pathogenesis include P, S, Type I, and Dr fimbriae as well as novel fimbriae encoded by the *auf* gene cluster (Bergsten *et al.*, 2004; Buckles *et al.*, 2004; Johnson *et al.*, 2006; Klumpp *et al.*, 2006; Servin, 2005). Vaccination with purified proteins of P, Type I, and Dr fimbriae have been shown to protect against infection with UPEC *in vivo* (Goluszko *et al.*, 2005; Langermann *et al.*, 2000; Roberts *et al.*, 2004).

In UPEC, Type I fimbriae (encoded by the *fim* gene cluster) are important virulence factors (Bahrani-Mougeot *et al.*, 2002; Connell *et al.*, 2000; Hull *et al.*, 2002; Langermann *et al.*, 2000). Several members of the family of *Enterobacteriaceae* (Boddicker *et al.*, 2002; Gerlach *et al.*, 1989) produce these fimbriae. Type I fimbriae are expressed by over 90% of *E. coli* strains and they are uniformly distributed among commensal and pathogenic strains (Table 1; Schlager *et al.*, 2003). FimH, the adhesin subunit of Type I fimbria, is heterogeneous and has different isoforms, which bind selectively to either monomannose or trimannose residues on a variety of host cell types, mucus, and abiotic surfaces (Sokurenko *et al.*, 2001; Sokurenko *et al.*, 2004; Vandemaele *et al.*, 2004). All naturally occurring FimH variants are capable of mediating strong interactions with trimannose, but the ability to bind a single monomannose residue is the highly variable property of FimH isoforms (Sokurenko *et al.*, 2004). UPECs express mainly FimH variants with high affinity to monomannose present only on the receptors of the vaginal, urethral, and bladder epithelium (primarily uroplakins Ia and Ib; Chapter 1.3.1; Ishikawa *et al.*, 2004). Nevertheless, recent evidence indicates that factors other than differences in the mannose binding capacity of FimH may explain the distinct FimH adhesion phenotypes. Bouckaert and colleagues studied a number of FimH variants originating from fecal, uropathogenic and enterohemorrhagic *E. coli* and could not find a correlation between the FimH variation and the affinities or specificities of the variant FimH receptor-binding domains for oligomannosides. This suggests that varying FimH binding strengths may be more dependent upon the

receptor (i.e. carbonhydrate expression of the targeted host tissues or the urinary inhibitors), rather than the adhesin itself (Bouckaert *et al.*, 2006).

P fimbriae (pyelonephritis-associated pilus encoded by the *pap* gene cluster) are present in more than 70% of the pyelonephritogenic *E. coli* (Table 1; Johnson *et al.*, 2005b; Johnson *et al.* 2005d). PapG, the adhesin subunit of P fimbriae, selectively recognizes and binds to Gal-α(1-4)Gal – containing glycosphingolipids present only on the renal tissue epithelium (Dodson *et al.*, 2001; Ishikawa *et al.*, 2004; Larsson *et al.*, 2003). PapG has three allelic variants (PapGI, PapGII, and PapGIII), with class II being the most prevalent allele among ExPEC (34%; Table 1; Bingen-Bidois *et al.*, 2002; Johnson *et al.*, 1999).

Occurrence of the closely related S (the *sfa* gene cluster) and F1C fimbriae (the *foc* gene cluster) is associated to *E. coli* infecting the urinary tract and the meninges (Table 1; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Korczak *et al.*, 2005). The S and F1C fimbrial adhesins (SfaS and FocA, respectively) recognize and attach to the α-sialic acid residues present on diverse epithelial surfaces (Hacker *et al.*, 1993; Morschhauser *et al.*, 1993; Ott *et al.*, 1988). These fimbriae are encoded by around 35% of ExPEC strains (Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Korczak *et al.*, 2005).

Commonly encountered virulence factors in UPEC are Dr fimbriae (encoded by the *dra* gene cluster; Table 1; Goluszko *et al.*, 2005). Dr fimbriae belong to the family of Afa/Dr adhesins (the hallmark of diffuse adherent *E. coli*), which bind to the complement regulatory protein decay accelerating factor expressed on the epithelial surfaces (Fang *et al.*, 2004; Hasan *et al.*, 2002; Nowicki *et al.*, 2001). The afimbrial members of the Afa/Dr family of adhesins are referred to more closely in the following Chapter 1.2.2.

The fimbrial "newcomers" (Auf fimbriae) of ExPEC are encoded by the *aufABCDEFG* gene cluster, which was recently described in the archetypal UPEC strain CFT073 by Buckles and co-workers (Buckles *et al.*, 2004). Auf fimbriae are significantly associated with UPEC when compared to commensal *E. coli* belonging to the intestinal normal flora (Table 1). AufA, the major structural unit of Auf fimbriae, is expressed *in vivo* in the murine urinary tract (Buckles *et al.*, 2004). So far, no hemagglutination or cellular adherence properties have been described for Auf fimbriae and hence the role of these fimbriae in the ExPEC pathogenesis is unclear.

During an UTI, the ExPEC population alternates between fimbriated and non-fimbriated states. Most fimbrial gene clusters are phase-variable, and the ability to control the ON/OFF- switch of fimbrial biogenesis strongly influences the virulence of ExPEC (Gunther IV *et al.*, 2002; Hernday *et al.*, 2004; Lee *et al.*, 2004; Schlager *et al.*, 2003). Recent studies show that spatially separated fimbrial gene clusters "communicate" with each other; the regulatory genes of one cluster influence the expression of other fimbriae (Pourbakhsh *et al.*, 1997; Roesch *et al.*, 2003; Snyder *et al.*, 2004; Snyder *et al.*, 2005). Nonetheless, neither the transient fimbriation status in an infecting population, nor the environmental stimuli required for the full expression of fimbriae during an infection are known. Even if the roles of different fimbriae in the ExPEC pathogenesis are largely unresolved, there is strong evidence suggesting that the ExPEC fimbriae are involved in the various stages of UTI; they promote bacterial adhesion and invasion as well as cytokine production, inflammation, and apoptosis (i.e. the programmed cell death) in the host's phagocytic and epithelial cells (Mulvey *et al.*, 2000; Mysorekar *et al.*, 2002; Schilling *et al*., 2003; Svanborg *et al.*, 2001).

1.2.2 Known afimbrial adhesins

The relation between bacterial adherence and urovirulence become evident in the 1970s (Duguid *et al.*, 1955; Eden *et al.*, 1976; Eden and Hansson, 1978). The first afimbrial adhesins in UPEC were described in the 1980s and referred to as "X adhesins" (Vaisanen *et al.*, 1981). To date, a number of afimbrial virulence traits have been characterized, which allow UPEC adhere to and, in some cases, invade tissues of the urinary tract (e.g. the Afa/Dr adhesins; Doye *et al.*, 2002; Guignot *et al.*, 2001).

The *E. coli* strains bearing AfaDE adhesins (the *afa* gene clusters) of the Afa/Dr adhesin family are associated with both urinary tract and intestinal infections (Table 1; Servin, 2005). The afimbrial AfaDE adhesin is composed of two proteins with different roles in the bacterium-cell interactions, i.e. the adhesin subunit AfaE and the invasin subunit AfaD (Anderson *et al.*, 2004; Garcia *et al.*, 1996). The adhesive subunit AfaE is highly variable; the 13 known subtypes of AfaE differ in amino acid (AA) sequence, receptor repertoire, and host specificity (Nowicki *et al.*, 2001; Servin, 2005). The fimbrial and afimbrial Afa/Dr adhesins of UPEC (e.g. Dr fimbriae [Chapter 1.2.1.] and AfaE subtypes -1, 3, and -8) primarily elicit the recruitment of decay accelerating factor (Berger *et al.*, 2004; Garcia *et al.*, 2000; Le Bouguenec *et al.*, 2001). However, internalization promoted by the AfaD invasin is mediated through interaction with the epithelial integrin receptors (Guignot *et al.*, 2001; Plancon *et al.*, 2003).

The putative adhesin Hek (encoded by the *hek* gene; hemagglutinin from *E. coli* K1) was originally identified in a chromosomal gene island termed the Pathogenicity Island (PAI) II of the archetypal UPEC strain 536 (Dobrindt *et al.*, 2002). The protein sequence of Hek is highly similar (90% sequence identity) to heat-resistant agglutinin 1 (Hra) of EPEC (Lutwyche *et al.*, 1994). The *hek* homologue *hra* is significantly associated with UPEC when compared to *E. coli* of fecal origin (Table 1; Srinivasan *et al.*, 2003). While the functional role of Hek in UTI is unknown, Hra in intestinal pathogenic *E. coli* is a mannose-resistant hemagglutinin, which promotes agglutination of human erythrocytes and colonic cells (Fleckenstein *et al.*, 2000).

Uropathogenic variant of the outer membrane protein Iha (encoded by the *iha* gene; IrgA-homologue adhesin) was originally identified in the UPEC strain CFT073 (Redford *et al.*, 2003). This protein is homologous with the IrgA (iron-regulated gene A) -like adhesins of EHEC and *Vibrio cholerae* (Goldberg *et al.*, 1992; Tarr *et al.*, 2000). In *V. cholerae* IrgA has a dual function; it serves both as an adhesin and as a receptor for the siderophore enterobactin (Mey *et al.*, 2002). Recently, Johnson and colleagues showed that *iha* is significantly associated with the ExPEC phenotype (Table 1; Johnson *et al.*, 2005a; Kanamaru *et al.*, 2003). In addition, Iha conferred upon a non-adherent *E. coli* the ability to adhere to uroepithelial cells *in vitro* and *in vivo* (Johnson *et al.*, 2005a). However, whether Iha contributes to the adherence directly or indirectly in concert with other bacterial components is not known.

1.2.3 Iron accumulation systems

Iron is essential for the growth of *E. coli*. Even though iron is abundant in nature, most of it forms insoluble hydroxide-complexes under oxygen and at neutral pH. In mammalian organisms, the concentration of free iron (Fe^{3+}) is extremely low (10^{-18} M); the remaining iron concentration is further reduced by binding of iron to the host's iron-chelating proteins, like albumin, heme, hemoglobin, ferritin, transferrin, and lactoferrin (Bullen *et al.*, 2005). Therefore, pathogenic *E. coli* have developed several effective iron uptake systems and the ability of iron accumulation significantly contributes to the virulence of these strains (Braun, 2003; Bullen *et al.*, 2005).

Lankford *et al.* introduced the term "siderophore" in 1973 (from the Greek "iron bearer") - 20 years after discovery of the first siderophore ferrichrome - to describe low molecular weight ferric iron chelators synthesized by bacteria, yeasts, fungi, and plants (Haas, 2003; Lankford, 2005; Neilands, 1981; Staiger, 2002). Most of the *E. co*li strains synthesize siderophores and their receptors in response to iron stress. In addition to their own siderophores, these strains are also able to utilize several different siderophores synthesized by other microorganisms (Braun, 2003). Additionally, a number of iron uptake systems of *E. coli* are siderophore independent (e.g. hemin receptors). These receptors act in concert with cytolytic hemolysins (see Chapter 1.2.4.), and thus account for the uptake of iron bound to the intracellular heme and hemoglobin (Andrews *et al.*, 2003; Braun, 2003).

ExPEC strains possess several sophisticated siderophore systems; 11 functional and putative iron uptake systems have been identified in the archetypal UPEC strain CFT073 (Welch *et al.*, 2002). The distinct siderophores ensure a constant supply of iron in different iron-deficient environments (e.g. due to varying water solubility; Konopka and Neilands, 1984). Several siderophores favored by ExPEC (e.g. aerobactin, yersiniabactin, and salmochelin) are largely unaffected by the siderophore-binding serum proteins, such as albumin (Bister *et al.*, 2004; Konopka and Neilands, 1984). The known siderophore and siderophore receptor systems in ExPEC include enterobactin, aerobactin, yersiniabactin, and salmochelin uptake systems (Hantke *et al.*, 2003; Pettis and McIntosh, 1987; Schubert *et al.*, 2004; Torres *et al.*, 2001). Ferri-siderophore complexes are internalized through a complex network of proteins, composed of the siderophore specific receptors and of proteins involved in the energy transfer and trans-membrane passage of these complexes (e.g. TonB; Braun, 2003). The expression of siderophore systems is up-regulated during bacterial growth in the urinary tract, which indicates the importance of effective iron assimilation for the UPEC virulence (Snyder *et al.*, 2004).

The siderophore system specific for the hydroxamate siderophore aerobactin significantly contributes to the virulence of ExPEC (Torres *et al.*, 2001). The aerobactin biosynthetic gene cluster (*iuc*) and the ferri-aerobactin receptor (*iutA*) were first identified in enteroinvasive *E. coli* as part of the virulence plasmid pColV (Williams, 1979). The *iuc-iutA* system is distributed throughout intestinal and extraintestinal pathogenic strains of *E. coli*, *Shigella* spp. *Salmonella* sp., *Klebsiella* sp. and *Vibrio* sp., and its expression influences the virulence of these strains (Table 1; Demir and Kaleli, 2004; Kingsley *et al.*, 1995; Koczura and Kaznowski, 2003; Moon *et al.*, 2004; Purdy and Payne, 2001).

The catecholate siderophore yersiniabactin was first described in *Yersinia pestis*, the infectious agent of bubonic plague (Fetherston and Perry, 1994). In most of the characterized ExPEC strains, the yersiniabactin biosynthetic gene cluster (*irp*) and the yersiniabactin receptor FyuA are encoded by a conserved chromosomal gene island denoted as the High Pathogenicity Island (Carniel *et al.*, 1996). To date, this island has also been identified in several other pathogenic members of the family *Enterobacteriaceae* (Schubert *et al.*, 2004). The *irp-fyuA* gene cluster is associated with the ExPEC phenotype and its expression contributes to the virulence of these strains *in vivo* (Table 1; Koczura and Kaznowski, 2003; Schubert *et al.*, 1998; Schubert *et al.*, 2002).

The siderophore receptor IroN was first described in *S. enterica* and later in the archetypal ExPEC strain CP9 (Baumler *et al.*, 1998; Russo *et al.*, 1999). The receptor IroN and the enterochelin glycosylase IroB, which is required for the synthesis of the novel siderophore salmochelin from the siderophore enterobactin, are encoded by the *iro* gene cluster (Hantke *et al.*, 2003). IroN mediates the uptake of salmochelin and other catecholate siderophores (e.g. enterobactin and dihydrobenzoic acid) and contributes significantly to the virulence of UPEC and NMEC in the infection models of ascending UTI and neonatal meningitis, respectively (Table 1; Hantke *et al.*, 2003; Negre *et al.*, 2004; Russo *et al.*, 2002).

The siderophore enterobactin (encoded by the *ent* gene cluster) and its receptor FepA are produced by pathogenic and non-pathogenic *E. coli* as well as by a number of other Gram negative and positive bacteria (Table 1; O'Brien and Gibson, 1970; Raymond *et al.*, 2003). However, although enterobactin is an effective iron chelator, it is bound strongly by the serum albumin and hence inactivated in the human serum (Konopka and Neilands, 1984). As enterobactin and its receptor are produced equally by pathogenic and non-pathogenic *E. coli* strains, their role in UTIs is probably only diminutive (Demir and Kaleli, 2004).

A number of gene clusters with putative function in the iron metabolism have been described in ExPEC. For example, the *prrA-modD-yc73-fepC* gene cluster (TonB-dependent outer membrane receptor; molybdenum transport protein; protein similar to *Haemophilus influenzae* protein yc73; ferric enterobactin transport ATP-binding protein) was originally identified as part of a PAI of the UPEC strain CFT073 and it is predominantly found in clinical ExPEC isolates (Table 1; Guyer *et al.*, 1998). The genes encoding outer membrane proteins ChuA (*chuA*; *E. coli* heme-utilization) and IreA (*ireA*; iron-responsive element) are found in several ExPEC strains (Table 1; Russo *et al.*, 2001; Torres and Payne, 1997). By the virtue of sequence homology, these proteins are involved in the iron uptake in TonB-dependent manner (Guyer *et al.*, 1998). However, their function and impact in the ExPEC virulence must be elucidated.

1.2.4 Toxins and proteases

Several bacterial toxins (endo- and exotoxins), proteases, and other effector proteins have established or putative functions in the ExPEC virulence. Their various biological roles and activities include cell adhesion, iron accumulation, and cell invasion through cell lysis and disruption of the mucin layer and the epithelium, as well as modulation and induction of the cell cycle, inflammatory reactions, and apoptosis (Goni and Ostolaza, 1998; Nougayrede *et al.*, 2005).

The *E. coli* α-hemolysin (encoded by the *hlyA* gene) was identified as an urovirulence factor in the early 1920s (Table 1; Cooke and Ewins, 1975; Dudgeon, 1921). The *hly* operon encodes the secreted cytolytic toxin HlyA and accessory proteins required for its modification and secretion (Cross *et al.*, 1990; Welch and Pellett, 1988). The UPEC HylA has a dual physiological function; at high concentrations, this toxin is cytolytic due to its pore-forming activity in the plasma membrane (Goni and Ostolaza, 1998; Soloaga *et al.*, 1999). At sublytical concentrations, HlyA interacts with the cell membrane and induces intracellular Ca^{2+}-oscillations. In this way, it interferes with the Ca^{2+} -dependent signaling pathways involved in the modulation of inflammatory responses of the target cells (Oxhamre *et al.*, 2005; Soderblom *et al.*, 2002; Uhlen *et al.*, 2000).

Cytotoxic necrotizing factors 1 and 2 (encoded by *cnf* genes; CNFs) are lethal for a wide variety of eukaryotic cells and produced by intestinal and extraintestinal pathogenic *E. coli* (Table 1; Hoffmann and Schmidt, 2004; Landraud *et al.*, 2004; Mills *et al.*, 2000). The CNF toxins are composed of two domains: the cell-binding domain and the catalytic domain (Lemichez *et al.*, 1997). These toxins activate the host's Rho GTPases and therefore influence the regulatory pathways of the actin cytoskeleton, transcription, cell transformation, cell proliferation, and apoptosis (Chung *et al.*, 2003; Falzano *et al.*, 2003; Mills *et al.*, 2000; Oswald *et al.*, 1994; Peres *et al.*, 1997). Consequently, the phagocytic activity of the epithelial cells increases, which enables the crossing of the epithelial barrier by the CNF-producing *E. coli* (Fiorentini *et al.*, 2001; Landraud *et al.*, 2004). In addition, similarly to several other virulence factors described for UPEC (see chapters 1.2.1. and 1.2.2.), also CNFs modulate and impair immune responses by influencing the secretion of inflammatory mediators (Fiorentini *et al.*, 2003; Malorni *et al.*, 2003; Munro *et al.*, 2004).

UPEC has four outer membrane serine proteases (Sat, PicU, Hbp, and OmpT) with different substrate specificities and protease functions (Table 1; Guyer *et al.*, 2002; Kukkonen and Korhonen, 2004; Otto *et al.*, 1998; Parham *et al.*, 2004). Sat (*sat*) is a serine auto-transporter toxin, which displays cytopathic activities on the kidney and bladder epithelium *in vitro* and *in vivo* (Guyer *et al.*, 2002). Hemoglobin-protease Hbp (*hbp*) is a homologue of the Tsh hemagglutinin of avian pathogenic *E. coli* (APEC). Hbp and Tsh are heme-binding proteins with serine protease activity targeted at hemoglobin (Tivendale *et al.*, 2004). The UPEC PicU (*picU*) is homologous with Pic (protein involved in intestinal colonization) of *Shigella* spp. and enteroaggregative *E. coli*. Pic proteases disrupt mucosal barriers by degrading mucosal pepsin, spectrin, and mucin (Bellini *et al.*, 2005; Henderson *et al.*, 1999; Parham *et al.*, 2004). OmpT (*ompT*), the fourth outer membrane endopeptidase of ExPEC, is structurally similar to plasminogen activator Pla of Y. pestis and degrades antimicrobial peptides secreted by the epithelial cells and macrophages (Grodberg *et al.*, 1988; Kukkonen and Korhonen, 2004).

Cytolethal distending toxins (encoded by *cdt* genes; CDTs) are bacterial effector molecules able to manipulate the eukaryotic cell cycle (Nougayrede *et al.*, 2005). These toxins consist of the catalytic domain (CdtB) and carrier subunits required for the receptor recognition and delivery of the catalytical domain into nucleus of the target cells (Lara-Tejero and Galan, 2001). CdtB causes chromatin damage, which activates the cellular DNA damage signaling pathways (Frisan *et al.*, 2002). The most commonly reported response to the CDT intoxication is an irreversible blockade of the cell cycle, but the effects of CDTs vary depending on the cell type (e.g. necrosis *versus* apoptosis; Bielaszewska *et al.*, 2005; Yamamoto *et al.*, 2004). Although a number of ExPEC strains are positive for CDTs, their function in the pathogenesis of these strains is not known (Table 1).

Microcins (*mch*) and colicins (*cva*) are antimicrobial peptides secreted by various members of the *Enterobacteriaceae* family (Table 1; Gillor *et al.*, 2004; Patzer *et al.*, 2003). They are produced under conditions of nutrient depletion and are active against phylogenetically related microbial strains. Microcins and colicins are considered to play important roles in the microbial competitions within the infecting populations and the intestinal flora (Azpiroz *et al.*, 2001; Snyder *et al.*, 2004).

Uropathogen specific protein (*usp*) was first identified as part of a novel PAI of the UPEC strain Z42 by Kurazono and colleagues in Japan in 1999 (Kurazono *et al.*, 2000; Yamamoto *et al.*, 2001). Usp is similar to the *Pseudomonas aeruginosa* toxin pyocin and the *E. coli* colicins. The *usp* gene is significantly associated with ExPEC and its expression increases the infectivity of non-pathogenic *E. coli* laboratory strains in the mouse model of ascending pyelonephritis (Table 1; Yamamoto *et al.*, 2001). However, the biological function of Usp in UTIs is still unknown.

The gram-negative endotoxin lipopolysaccharide (LPS) is the main bacterial factor in the development of endotoxemia, the pathogen-induced systemic immune response leading to the lethal shock-condition of the host (Zimecki *et al.*, 2004). In the ascending UTI, the O-polysaccharide moiety of the UPEC LPS has an important regulatory function (Fischer *et al.*, 2006; Svanborg *et al.*, 2001). LPS signals through Toll-like receptor -4 (TLR4) of the urothelial and inflammatory cells and hence induces production of pro-inflammatory mediators (e.g. cytokines and chemokines) and cellular nitric oxide synthases (Austin *et al.*, 2003; Backhed *et al.*, 2003; Cross *et al.*, 1995). The LPS-induced systemic inflammation breaks the inertia of mucosal barrier and allows UPEC to gain the access into the underlying tissues. On the mucosal surfaces of the urinary tract, the LPS-dependent activation of host's TLR4-signaling is dependent of the adhesins of Type I or P fimbriae (Chapter 1.2.1.; Blomgran *et al.*, 2004; Fischer *et al.*, 2006; Hedlund *et al.*, 2001).

1.2.5 Surface coatings

Bacterial surface coatings, i.e. capsular polysaccharides (K-antigens), LPS, and O-polysaccharide moieties of LPS (O-antigens) have been implicated as possible virulence determinants in extraintestinal infections (Kusecek *et al.*, 1984; Russo *et al.*, 1996). There are 176 different O-serogroups and more than 80 K-antigens in *E. coli*. In addition, many *E. coli* strains are able to synthesize colonic acid, the main component of the polysaccharide coatings of biofilms (Reisner *et al.*, 2003). LPS and its impact in UTIs were described in more detail in chapter 1.2.4.

E. coli produces four types of capsular polysaccharides (groups 1 to 4), each of which comprises of a large number of distinct types (Keenleyside *et al.*, 1993). Most ExPEC strains have capsules and certain capsule types (e.g. K1, K5, K10, and K54) have been shown to contribute to the pathogenesis of these strains (Horwitz and Silverstein, 1980; Pluschke *et al.*, 1983; Schneider *et al.*, 2004; Suerbaum *et al.*, 1994; Whitfield and Roberts, 1999). Similarly, UPEC strains are more likely to belong to the O-antigen serotypes O1, O2, O4, O6, O16, O18, O22, O25, and O75 (Johnson *et al.*, 1994; Korhonen *et al.*, 1985; Kusecek *et al.*, 1984; Russo *et al.*, 1996). The genes responsible for the capsular polysaccharide and LPS synthesis in the UPEC strain CFT073 are highly expressed during experimental UTI, which indicates the importance of surface coatings *in vivo* (Snyder *et al.*, 2004). However, whether certain O-polysaccharide or capsular antigens contribute to the ascending UTI is not clear. Some studies implicate that bacterial coatings may account for the increased serum resistance and thus impede the clearance of ExPEC by phagocytosis (Horwitz and Silverstein, 1980; Leying *et al.*, 1990; Pluschke *et al.*, 1983; Rizvi and Kumar, 2003; Suerbaum *et al.*, 1994).

1.2.6 Flagellar motility

At present, 53 flagellar antigens (H-antigens) are described for *E. coli* (Wang *et al.*, 2003). The virulence of diarrheagenic *E. coli* has been shown to be dependent of the expression of functional flagella (Best *et al.*, 2005; Nougayrede *et al.*, 2003; Sircili *et al.*, 2004). Flagella are supposed to be essential adhesive organs in establishing *E. coli* extra- and intracellular biofilms (Justice *et al.*, 2004; Reisner *et al.*, 2003; Tenorio *et al.*, 2003; Wright *et al.*, 2005). However, the impact of flagellar motility in the colonization of the urinary tract is still subtle. Flagellar and chemotaxis genes have shown to be down regulated during an experimental UTI (Snyder *et al.*, 2004). But while the non-motile UPEC can advance through the stages of UTI in both the bladder and the kidney, the functional flagella have shown to confer a fitness advantage in persisting throughout the urinary tract and may therefore enhance the UPEC pathogenesis (Anderson *et al.*, 2004; Justice *et al.*, 2004; Lane *et al.*, 2005; Snyder *et al.*, 2004; Wright *et al.*, 2005).

Table 1. Prevalence of genes encoding known ExPEC virulence factors among ExPECs and commensal *E. coli* isolates.

Gene	Prevalence (%)[a]		Reference
	ExPEC	Commensal *E. coli*	
Fimbrial adhesins			
auf	60	31	Buckles *et al.*, 2004
fimH	67-100	58-92	Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Moreno *et al.*, 2005
papC	60	37	Johnson *et al.*, 2005b; Johnson *et al.* 2005d
papGI	0-2	0	Bingen-Bidois *et al.*, 2002; Johnson *et al.*, 2005
papGII	40-62	3-25 (45[b])	Johnson *et al.*, 2005; Moreno *et al.*, 2005; Nowrouzian *et al.*, 2003
papGIII	5-31	3 (34[b])	Bingen-Bidois *et al.*, 2002; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Nowrouzian *et al.*, 2003
sfa/foc	14-24 (59[c])	0 (26[b])	Bingen-Bidois *et al.*, 2002; Duriez *et al.*, 2001; Johnson and Stell, 2000; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Nowrouzian *et al.*, 2003; Watt *et al.*, 2003
Afimbrial adhesins			
afa/dra	3-14 (39[c])	0-5	Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Johnson and Stell, 2000; Moreno *et al.*, 2005; Watt *et al.*, 2003
hra	43-66	28	Srinivasan *et al.*, 2003
iha	33-74	12-25	Kanamaru *et al.*, 2003; Moreno *et al.*, 2005; Johnson *et al.*, 2005b; Johnson *et al.* 2005d
Iron accumulation systems			
aer	50-80	16-50	Bingen-Bidois *et al.*, 2002; Duriez *et al.*, 2001; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Korczak *et al.*, 2005; Moreno *et al.*, 2005; Nowrouzian *et al.*, 2003; Ruiz *et al.*, 2002
chuA	70-95	30	Hoffmann *et al.*, 2001; Korczak *et al.*, 2005
ent	44-86	63-90	Demir and Kaleli, 2004; Koczura and Kaznowski, 2003b

fep	90	22	Guyer *et al.*, 1998
ireA	26	0	Russo *et al.*, 2001
iro	51-78	7-44	Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Kanamaru *et al.*, 2003
mod	93	33	Guyer *et al.*, 1998
prr	87	26	Guyer *et al.*, 1998
y73	90	19	Guyer *et al.*, 1998
irp	81-92	58-67	Bingen-Bidois *et al.*, 2002; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Johnson and Stell, 2000 ; Moreno *et al.*, 2005
Toxins and proteases			
cdtB	6-9	5	Johnson and Stell, 2000; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Moreno *et al.*, 2005
cnf	15–61	0-10	Duriez *et al.*, 2001; Korczak *et al.*, 2005; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Johnson and Stell, 2000
hbp	63	33	Heimer *et al.*, 2004
mch	36	3	Grozdanov *et al.*, 2004
hly	21-79	4-20 (23[b])	Duriez *et al.*, 2001; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Johnson and Stell, 2000; Moreno *et al.*, 2005; Nowrouzian *et al.*, 2003
ompT	83-94	51-68	Kanamaru *et al.*, 2003
pic	31	7	Heimer *et al.*, 2004
sat	30-55	22	Guyer *et al.*, 2000; Ruiz *et al.*, 2002
usp	85	24	Kanamaru *et al.*, 2003

a The estimated prevalence of virulence factors varies between diverse studies and is strongly dependent upon the strain collection used
b Healthy infants of 3 to 12 months of age
c Pregnant women

1.3 Impact of host factors contributing to the development and recurrence of UTIs caused by ExPEC

Development and severity of UTIs are influenced not only by the invading pathogen but also by the host. A variety of host factors, such as age, gender, pregnancy, or immunological status may predispose towards UTI and allow less virulent pathogens to cause the disease (Gordon *et al.*, 2005).

Structural disorders of the urinary tract or the presence of foreign bodies (e.g. the urinary catheter) impair the urine flux and are common risk factors for UTI (Honkinen *et al.*, 1999; Nicolle, 2005). In addition, metabolic disorders and autoimmunity diseases (e.g. *diabetes mellitus* and *lupus erythematosus*) have been shown to contribute first, to the enhanced colonization of the otherwise sterile urinary tract by pathogenic bacteria and secondly, to the impaired clearance of an infection by the host (Hidalgo-Tenorio *et al.*, 2004; Hoepelman *et al.*, 2003).

UTIs caused by UPEC are the most common bacterial infections during pregnancy. These infections are often accompanied by serious complications, such as preterm labor or early onset of neonatal sepsis, which may lead to significant morbidity and mortality in both the fetus and the mother (Jones *et al.*, 2004; Le *et al.*, 2004). There are several studies documenting causal relationship between sex practices and elevated risk of UTIs; sexual activity (e.g. intercourse frequency or the use of spermicide-based contraception) is associated with an increased risk of UTIs and their recurrences (Foxman *et al.*, 2002; Karkkainen *et al.*, 2000; Kontiokari *et al.*, 2003). These activities partly explain the observed relation of UTIs to the age, recurrent UTIs being most common in women of 25 to 29 years of age. The second UTI risk group is comprised of women over 55 years of age (Foxman and Brown, 2003; Hu *et al.*, 2004). In this age group, the frequent and recurrent UTIs are explained with an increasing urinary incontinence, a decreased estrogen level, and an altered ABO-blood group secretor status (Hu *et al.*, 2004; Jackson *et al.*, 2004; Rozenberg *et al.*, 2004). Men are relatively rarely affected by UTI and prostatis; approximately 14% of men have experienced one or more UTIs (Griebling, 2004). UTI, urosepsis, and UTI-associated meningitis are the major causes of bacterial infections in infants younger than 90 days (Bachur and Caputo; 1995). Male infants are above average affected by UTIs and UTI-associated sepsis, and the main causes of infant urosepsis are anatomical or functional abnormalities of the urinary tract (Bachur and Caputo, 1995; Ginsburg an McCracken, 1985; Honkinen *et al.*, 2000).

Finally, domestic animals can serve as reservoirs for ExPEC and may increase the risk to acquire UTI (Chen *et al.*, 2003; Murray *et al.*, 2004). Dietary factors, such as frequent consumption of fresh berry juices and fermented milk products, use of probiotic micro-organisms, and breast feeding of neonates, have been shown to protect against UTI (Hoesl and Altwein, 2005; Kontiokari *et al.*, 2003; Marild *et al.*, 2004; Martin-Sosa *et al.*, 2002). However, the causal relationships between the host's immunological status or life style and UTIs are not clear.

1.3.1 Histology of the urinary tract

Urothelium, the epithelium lining the urinary tract is composed of three cell types: undifferentiated basal cells, intermediate cells, and superficial facet cells (Fig. 2). It is covered with microvilli and mucus. Superficial cell layers are connected by tight junctions, which a physical barrier to the infection. These cells express four types of integral membrane proteins (i.e. uroplakins Ia, Ib, II and III), which are recognized by UPEC. Uroplakins Ia and II are strictly specific for the urothelium (Olsburgh *et al.*, 2003). The turnover rate of the normal bladder epithelium is extremely slow and its regeneration in humans is estimated to take 6 to 12 months (Southgate *et al.*, 1999).

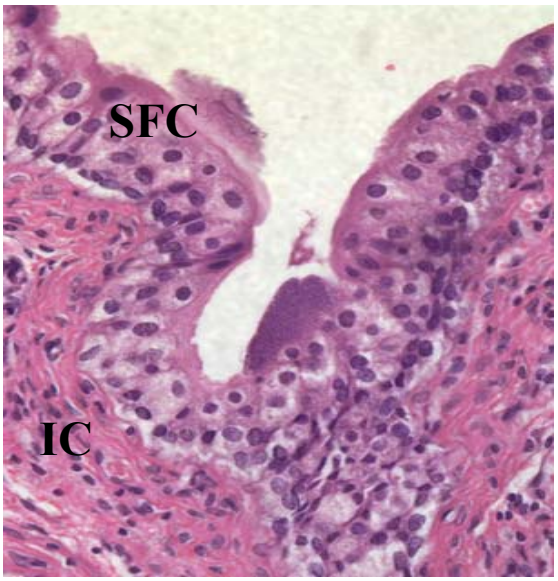

Figure 2. Haematoxylin and eosin stained mouse urothelium
SFC = superficial facet cells,
IC = intermediate cells
Figure kindly provided by K. Hildinger
(Max von Pettenkofer-Insitut).

1.3.2 Host response to UTIs

UPEC interact with the host at two levels: at the mucous surfaces lining the urinary tract and at the tissues of the kidney and the bladder (Mulvey, 2002). In the first phase of UTI, invading pathogens bind to receptors (e.g. uroplakins; Chapter 1.3.1.) on the mucosal surfaces coating the urothelium (Fig. 3). During the acute phase of UTI, widening of the tight junctions and extensive exfoliation of the urothelial cells take place. This active exfoliation of the infected cells is initiated by the host and serves to eliminate the bulk of adherent bacteria, but it also detrimentally exposes the underlying undifferentiated basal cells (Fig. 3; Mysorekar *et al.*, 2002). The loss of epithelial integrity is complemented by a rapid renewal of the urothelium and by the accumulation of epithelial proteins (Mysorekar *et al.*, 2002). In addition, receptors for the leukocyte integrins and circulating leukocytes are highly expressed on the infected urothelium (Mysorekar *et al.*, 2002). These leukocyte adhesion molecules play a key role in the inflammatory processes; they activate and assist in the epithelial transmigration of circulating leukocytes (e.g. polymorphonuclear neutrophils and macrophages), the key effectors against UPEC (Fig. 3; Detmers *et al.*, 1990).

The host's first line of defense against UPEC (i.e. the. innate immunity) involves a complex interplay of unspecific antimicrobial proteins, including the urinary Tamm-Horsfall protein, defensins, and complement proteins, as well as cytokines and inflammatory cells (Fig. 3; Nitschke *et al.*, 2002; Saemann *et al.*, 2005; Schilling *et al.*, 2003). The main cytokines secreted during UTI include transforming growth factor β, tumor necrosis factors α and β, interferon γ, and interleukins 6 and 8 (IL; Fig. 3). These mediators are produced by the bladder and renal epithelial cells and are present in highly elevated levels in the urine of patients during the acute phase of UTI (Farmaki *et al.*, 2005; Funfstuck *et al.*, 2001; Schilling *et al.*, 2003). In the response to cytokines and leukocyte adhesion molecules expressed by the infected urothelium, phagocytic cells migrate through capillary cell walls (Fig. 3; Gbadegesin *et al.*, 2002; Godaly *et al.*, 2001). Once at the site of infection, phagocytes recognize opsonised bacteria through specific receptors (Durand *et al.*, 2001; Nishimura *et al.*, 2001). Cytoplasmic granules within these cells contain a number of antimicrobial products and reactive oxygen intermediates (Fang, 2004; Wiedow and Meyer-Hoffert, 2005); following phagocytosis, these

granules fuse with the phagosomes leading to effective killing of indigested bacteria (Chowdhury *et al.*, 2004).

The innate and adaptive immune systems collaborate to provide protection against invading pathogens. A couple of days after the initial infection phagocytic cells are accompanied by the antibody producing and cytotoxic immune cells. The UPEC specific antibodies elicited in response to UTI are primarily immunoglobulins A (IgA), IgM, and IgG; they are commonly directed against LPS, outer membrane proteins, and fimbriae (e.g. Type I and P fimbriae; Gibb and Edmond, 1992; Kantele *et al.*, 2003; Kawahara *et al.*, 1994; Nicolle *et al.*, 1989; Salit *et al.*, 1988). The secretory-IgA plays a pivot role in the immunological defense of the urinary tract by preventing the initial bacterial adherence to the urothelium (Deo and Vaidya, 2004).

Cells of the human immune system as well as the bladder and renal epithelial cells, express 10 (11 in mice) distinct types of TLRs on their surfaces. These receptors are involved in the recognition of conserved molecular patterns unique for microbes, such as the bacterial LPS, lipoproteins, fimbriae, and flagellin as well as viral and bacterial RNA and DNA (Finberg *et al.*, 2004; Takeda and Akira, 2004). During UTI, the LPS-specific signaling through TLR4 (and TLR11 in mice) is necessary for host's cytokine production, activation of the immune-cells, and the tissue integrity control (Takeda and Akira, 2005; Zhang *et al.*, 2005).

Intact function of the urothelium and innate immune-defenses are essential for the resistence and clearance of UTIs caused by UPEC; deficiencies in the urothelial integrity, phagocyte activation and function, secretion of antimicrobial proteins, cytokines, and antibodies have been described in the patient groups susceptible to UTIs (Chapter 1.3.; Chowdhury *et al.*, 2004; Williams and Schaeffer, 2004). The inability of the host to induce an appropriate immune response may result in recurrent or chronic infections and provoke additional pathological conditions, including autoimmune diseases and renal dysfunction (Stojanovic *et al.*, 2004).

Untreated pyelonephritis is a life threatening condition associated with the scarring of the renal tissue, loss of glomerular function, and an eventual loss of renal function (Jahnukainen *et al.*, 2005). Although the mechanisms responsible for the renal scars are obscure, there is strong evidence that the tissue damage is caused by uncontrolled inflammation resulting from an unbalanced cytokine or TLR control rather than by any direct bacterial influence (Chowdhury *et al.*, 2004; Jahnukainen *et al.*, 2005).

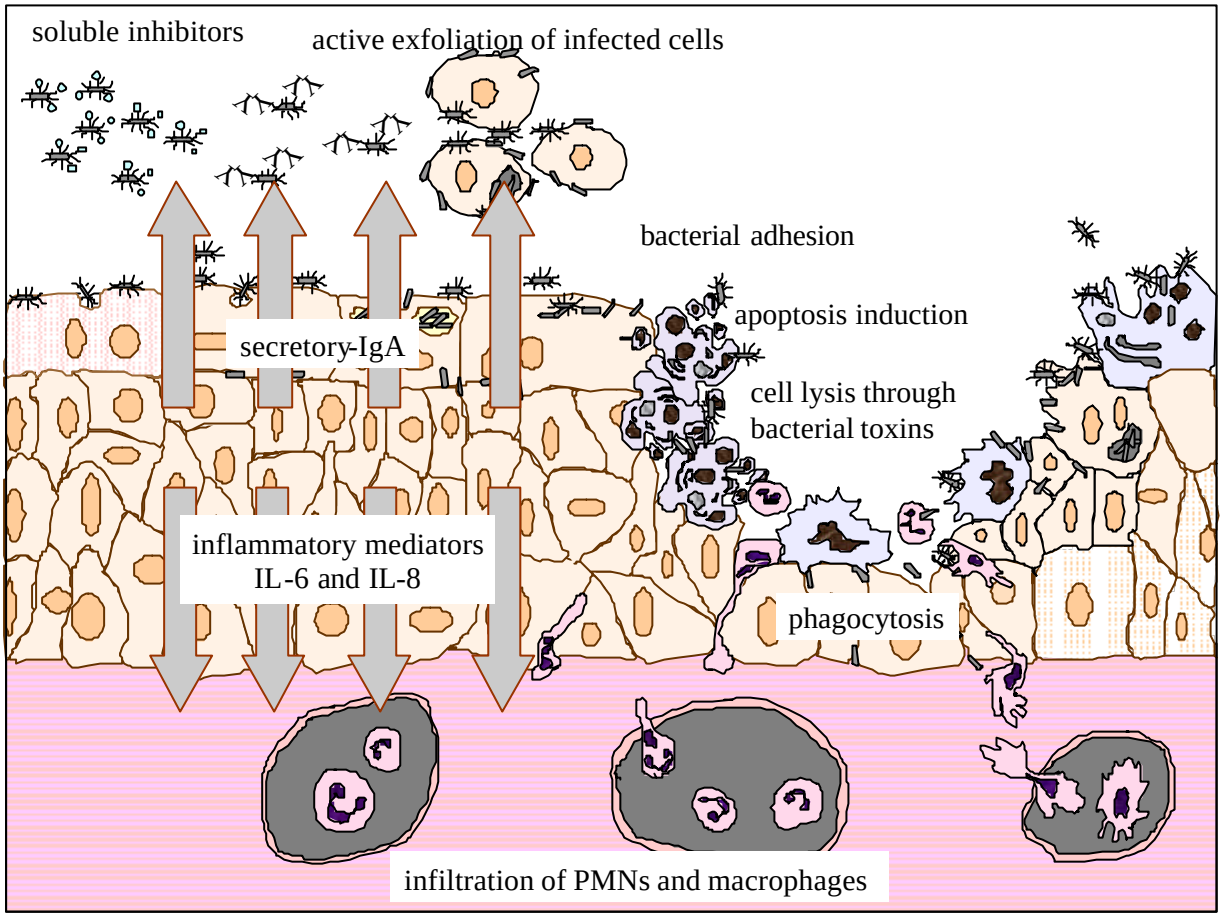

Figure 3. Loss of mucosal integrity during the acute phase of UTI. During the acute phase of UTI, UPEC disrupts the mucosal integrity by secreting cytotoxins, inducing apoptosis, and foremost, by triggering a strong inflammatory reaction in the host's urothelium. Host responses to the bacterial invasion include active exfoliation of the infected cells and secretion of soluble bacterial inhibitors (e.g. Tamm-Horsfall protein and secretory-IgA) and inflammatory cytokines (e.g. IL-6 and IL-8) as well as expression of receptors for the circulating leukocytes. In the response to cytokines and leukocyte adhesion molecules expressed by the infected urothelium, phagocytic cells (e.g. polymorphonuclear neutrophils [PMN] and macrophages) migrate through capillary cell walls into the site of infection. Once at the site of infection, PMNs and macrophages secrete cytokines and reactive oxygen intermediates, which further enhance the inflammation.

1.4 Genomic structure of ExPEC

1.4.1 Clonal origin

Phylogenetic analyses have shown that the natural populations of *E. coli* are basically clonal and that *E. coli* strains fall into four main phylogenetic groups (PGs; A, B1, B2, and D; Johnson and Russo, 2005; Ochman and Selander, 1984; Pupo *et al.*, 1997). ExPEC strains belong mainly to the PG B2 and to a lesser extent to the group D, whereas most commensal *E. coli* isolates belong to the PGs A and B1 (Clermont *et al.*, 2000; Clermont *et al.*, 2001; Picard *et al.*, 1999). Recent studies indicate that most of ExPEC strains have evolved from an ancestral strain by a stepwise accumulation of virulence factors through horizontal gene transfer (Bidet *et al.*, 2005; Dobrindt and Hacker, 2001). The specific genetic background (e.g. the PG B2) may be required for the acquisition and expression of these virulence factors (Bidet *et al.*, 2005; Dobrindt *et al.*, 2002; Escobar-Paramo *et al.*, 2004; Le Gall *et al.*, 2005; Picard *et al.*, 1999). However, the sole association of an *E. coli* strain with a certain PG does not necessarily render the strain pathogenic or non-pathogenic as the adaptation of a strain to a host also contributes significantly in determining its virulence (Duriez *et al.*, 2001; Souza *et al.*, 1999).

1.4.2 Genomic variation

During the last decade, high-throughput techniques have been developed that allow the sequencing of bacterial chromosomes in a short time. To date, about 140 bacterial chromosomes have been completely sequenced and the genome sequences are available at the databases of the National Center for Biotechnological Information (NCBI; http://www.ncbi.nlm.nih.gov) and the European Bioinformatics Institute (EBI; http://ebi.ac.uk). From the numerous sequence data available in these databases, it has become obvious that *E. coli* genomes are highly diverse mosaic structures in permanent flux. Comparison of the genome sequences of three archetypal *E. coli* strains, namely the EHEC strain EDL933 (5.5 mega base pairs [Mb]; NC_002655), the UPEC strain CFT073 (5.2 Mb; NC_004431), and the commensal *E. coli* strain MG1655 (4.6 Mb; NC_000913) revealed a conserved *E. coli* core genome with an estimated size of 3.9 Mb (Table 3; Blattner *et al.*, 1997; Perna *et al.*, 2001; Welch *et al.*, 2002). This core genome has a G+C content of 50.8% and displays a homogeneous codon usage pattern (Blattner *et al.*, 1997; dos Reis *et al.*, 2003; Perna *et al.*, 2001; Sharp and Li, 1987; Welch *et al.*, 2002). Besides the core genome a strain-specific flexible gene pool exists (predictably up to 1.3 Mb) that is formed by horizontally transferred mobile genetic elements, such as PAIs, plasmids, phages, transposons, and insertion sequences (IS; Blattner *et al.*, 1997; Perna *et al.*, 2001; Welch *et al.*, 2002). These mobile "genetic modules" facilitate rapid adaptation of *E. coli* to the changing environmental conditions and hence extend the spectrum of sites that can be infected (Hacker and Carniel, 2001; Lerat *et al.*, 2005; Ochman *et al.*, 2005; Ochman and Jones, 2000). The strain specific extragenous gene pool is largely different from the set of established house keeping genes involved in metabolic and catabolic functions and differs both in the G+C composition and the codon usage pattern (Daubin *et al.*, 2003; Lawrence and Ochman, 1997; Medigue *et al.*, 1991; Ochman *et al.*, 2005).

1.4.3 Pathogenicity islands

The concept of "pathogenicity island" was introduced in the late 1980s by Hacker and colleagues to describe chromosomal gene islands (Hacker *et al.*, 1990). Nowadays, there is a tendency towards describing PAIs as genomic islands (GI) because most of the traditional PAI-associated genes do not directly account for virulence and they are found in non-virulent strains too. In view of that, GIs are classified as PAIs, fitness islands, metabolic islands, or symbiotic islands according to their established or putative functions (Fig. 4; Carroll *et al.*, 2004; Chouikha *et al.*, 2006; Dobrindt *et al.*, 2004).

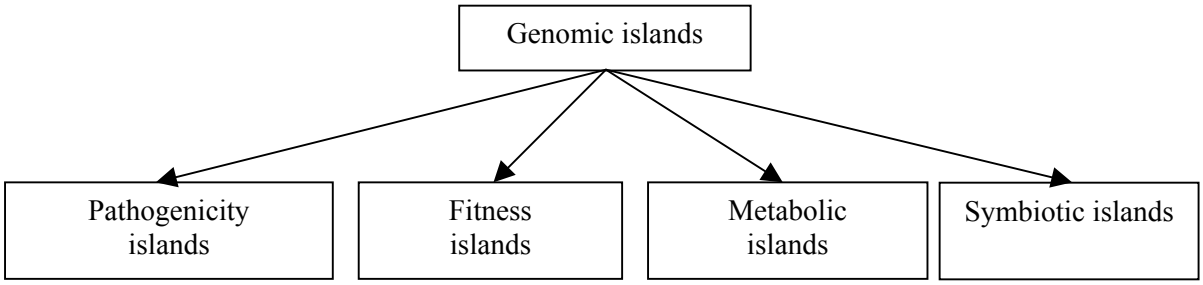

Figure 4. GIs are subclassified according to their established or putative functions as PAIs, fitness islands, metabolic islands, or symbiotic islands.

Several GIs have been identified in ExPEC and most of the known urovirulence factors are traditionally thought to be PAI-associated (e.g. siderophore systems and fimbrial adhesins). The model organisms for the study of GIs of ExPEC are the UPEC isolates 536, J96, and CFT073 (Bingen-Bidois *et al.*, 2002; Dobrindt *et al.*, 2003; Houdouin *et al.*, 2002; Johnson *et al.*, 2002). Genomic islands carry a number of common genetic features, which are outlined below. Points i and ii are the criteria for a GI to be termed a PAI. For a current review, see references Hochhut *et al.*, 2005 and Schmidt and Hensel, 2004.

i. PAI carry one or more virulence genes.
ii. PAI are present in the genomes of pathogenic bacteria but absent from the genomes of the non-pathogenic representatives of the same species or closely related species.
iii. GIs occupy relatively large genomic regions; the majority of GIs are in the range of 10 to 200 kb (kilo base pair).
iv. GIs often differ from the core genome with respect to the base composition and the codon usage; recently acquired GIs still have the base composition and the codon usage pattern of the donor species.
v. GIs are frequently located adjacent to transfer-RNA (tRNA) genes; genes encoding tRNAs are highly conserved among various bacterial species and are assumed to provide an integration point for foreign DNA (by providing substrate for homologous recombination or phage transduction).
vi. GIs are frequently associated with mobile (functional or cryptic) genetic elements, like integrases, ISs, plasmids, transposons, and bacteriophages or remnants of these elements. In addition, GIs are often flanked by direct repeats and recognition sites for the integration and excision of bacteriophages; the mobile genetic elements may contribute to the instability of GIs flanked by direct repeats.

vii. GIs are often unstable and are deleted with distinct frequencies; the same genetic elements that contribute to mobilization are also responsible for intrinsic genetic instability.

viii. GIs represent mosaic-like structures rather than homogeneous segments of horizontally acquired DNA; during the evolution, several genetic elements have been acquired independently from different hosts and at different time points.

1.5 DNA microarrays in the research of extraintestinal infections caused by ExPEC

First DNA microarrays were described in the early 1990s (Schena *et al.*, 1995). Today, the DNA and oligonucleotide microarray technologies play increasingly important roles in life sciences and they have been applied in various formats ranging from low-density arrays with few genes to higher density arrays including hundreds of probes. DNA microarrays have been employed in genotyping (e.g. phylogenetic and pathogenetic classification of bacteria, discrimination of micro-organisms in mixed or non-cultivable samples, or drug surveillance of clinically relevant micro-organisms) and in monitoring the transient gene expression in response to altered environmental conditions (e.g. *in vitro versus in vivo* growth of a pathogen; Bryant *et al.*, 2004; Choudhuri, 2004; Dharmadi and Gonzalez, 2004; Hofman, 2005; Stoughton, 2005).

A DNA microarray is a solid surface, such as a glass slide or silicon substrate, on which extended DNA fragments or short DNA oligonucleotides are attached (Fig. 5; Bras *et al*., 2004). In the DNA array format, DNA probes homologous with the desired gene or DNA region are generated by PCR (polymerase chain reaction) using either complementary-DNA (cDNA) or genomic DNA as template. These DNA probes are typically 300 to 800 nucleotides (nt) in length (Chou *et al.*, 2004; Schena *et al.*, 1995). In the oligonucleotide-based arrays, short target-specific DNA probes of 20 to 80 nt in length are either synthesized prior to spotting or *in situ* directly on the solid substrate (Cheng *et al.*, 2002; Chou *et al.*, 2004; Zarrinkar *et al.*, 2001). It should be noted that in this thesis the immobilized DNA molecules are referred to as "DNA probes" and the labeled DNA molecules (chromosomal DNA) as "DNA targets". Before array hybridization, the target DNA, which is either chromosomal DNA (genotyping) or cDNA (gene expression), is labeled with fluorochrome dyes, such as cyanines Cy5 (red) or Cy3 (green; Fig. 5). Multiple DNA targets labeled with miscellaneous fluorochromes can be employed simultaneously. The DNA-DNA hybridizations and subsequent stringency washes are performed as in the conventional Southern hybridization procedure with certain refinements (Ausubel *et al.*, 1989). Subsequently, the hybridized slides are scanned and the fluorescence emission of Cy5 and Cy3 dyes are measured at wavelengths of 635 nm and 532 nm, respectively (Fig. 5; Bilban *et al.*, 2002). The fluorescence intensity in the close vicinity of each spot is individually measured in order to estimate and eventually eliminate the contaminating background emission (signal-to-noise ratio; Martinez *et al.*, 2003). The different data sets (hybridizations) are normalized using normalization algorithms, such as the variance stabilization normalization algorithm developed by Huber and colleagues (Huber *et al.*, 2002).

To date, there are several studies, which describe the use of DNA microarrays in genotyping of pathogenic *E. coli* (e.g. pathoarrays) and expression profiling of infecting *E. coli* populations (Bekal *et al.*, 2003; Dobrindt *et al.*, 2003; Fukiya *et al.*, 2004; Korczak *et al.*, 2005; Ochman and Santos, 2005; Snyder *et al.*, 2005; van Ijperen *et al.*, 2002; Yu *et al.*, 2004). Pathoarrays are designed for screening of the pathotype-specific genetic markers either in defined clinical isolates or in mixed cultures, such as in stool, blood, or cerebrospinal fluid samples (Cleven *et al.*, 2006; Liu *et al.*, 2005; Wang *et al.*, 2004; Wu *et al.*, 2004). So far, four prototypic ExPEC pathoarrays with different sets of DNA probes have been described (Bekal *et al.*, 2003; Dobrindt *et al.*, 2003; Korczak *et al.*, 2005; van Ijperen *et al.*, 2002). These pilot studies showed that the ExPEC pathoarrays can be applied for differentiation of extraintestinal and intestinal pathogenic *E. coli* and commensal *E. coli* strains (Bekal *et al.*, 2003; Dobrindt *et al.*, 2003; Korczak *et al.*, 2005; van Ijperen *et al.*, 2002).

DNA microarrays comprised of all translatable open reading frames (ORF) of the commensal *E. coli* strain MG1655 (4290 ORFs) have been applied to determine the global genetic diversity among ExPEC, non-pathogenic *E. coli*, and closely related *Shigella* spp. The comparative genomic

hybridizations of ExPECs and *Shigella* spp. with the *E. coli* strain MG1655 showed that the core set of genes involved in the growth maintenance would comprise of 3100 ORFs in *E. coli* and 2586 ORFs in *Shigella* spp. (Dobrindt *et al.*, 2003; Fukiya *et al.*, 2004).

One of the promising applications of DNA microarrays is the monitoring of transient gene expression in response to an environmental stimulus (i.e. functional genomics; Balazsi *et al.*, 2005; Beloin *et al.*, 2004; Gerdes *et al.*, 2003; Hua *et al.*, 2004; Ren *et al.*, 2004). Snyder and colleagues used a sophisticated array to profile gene expression of the UPEC strain CFT073 during an experimental UTI. They identified a number of genes that were down- or upregulated in response to altered growth environment (i.e. Luria-Bertani broth, urine, and the urinary tract; Snyder *et al.*, 2004). This study is discussed more closely in Chapter 1.2. As an accurate inflammatory response induced by the host (e.g. in response to bacterial LPS) influences significantly the outcome of an UTI, monitoring of transient gene expression of the host at the different stages of infection will most likely be of particular importance in the future (Dou *et al.*, 2005; Bliss *et al.*, 2005; Higgins *et al.*, 2003; Saban *et al.*, 2002).

Unfortunately, the routine use of microarrays in diagnostic laboratories is still limited due to high costs, requirement of specialized equipment and well-trained personnel and hence sophisticated microarray technology is so far established only in research laboratories.

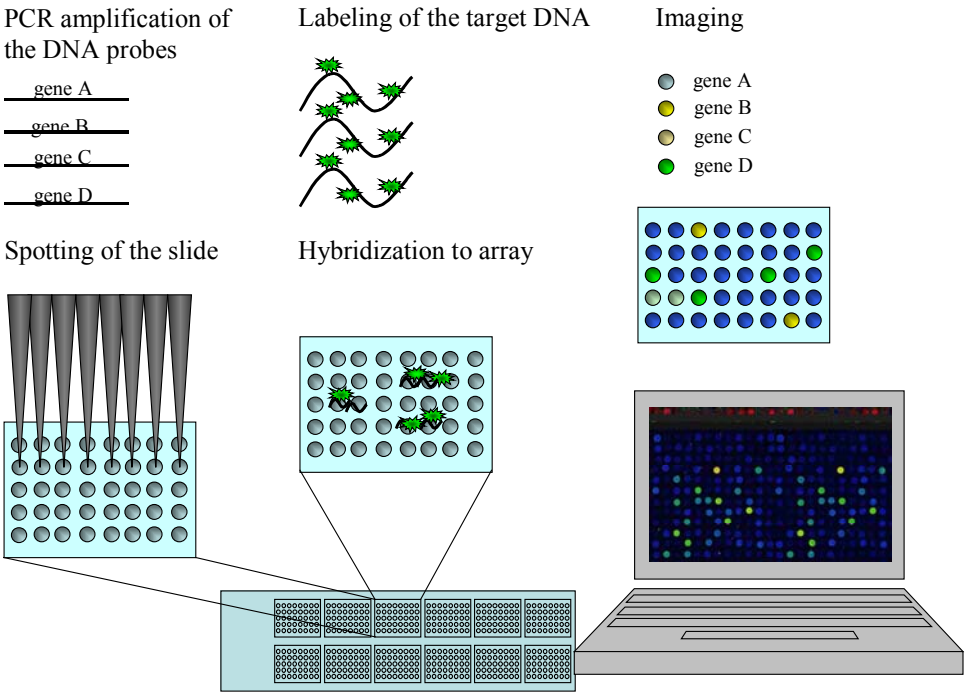

Figure 5. Generation and detection of a DNA microarray. DNA probes homologous with desired DNA regions are generated by PCR and printed on the slide surface. Subsequently, the target DNA is fluorescently labeled and hybridized on the array. The hybridized arrays are scanned for the fluorescence emission and the signal intensities are measured ("imaging").

1.6 Suppressive subtractive hybridization of genomic DNA

Diatchenko and colleagues developed the suppressive subtractive hybridization technique in 1996 (Fig. 6; Diatchenko *et al.*, 1996). This method is based on an antecessor technique of the subtractive genomic hybridization, where sequences present in one gene pool (tester-genome or tester-cDNA library) are enriched at the expense of another gene pool (driver-genome or driver-cDNA library) by the selective concentration (e.g. by affinity chromatography) of the tester-specific sequences after hybridization of the two sequence-pools (Welcher *et al.*, 1986). In the subtraction technique developed by Diatchenko, the requirement for purification of the subtracted DNA fragments is circumvented by using PCR amplification with oligonucleotide primers homologous with the linker-sequences ligated on the tester-DNA pool (Fig. 6). These PCR primers contain long inverted terminal repeats, which selectively suppress the PCR amplification of abundant sequences, thus balancing differences in the abundances of individual mRNAs or genes in the genome (Diatchenko *et al.*, 1996). Today, a few commercial kits are available for subtraction experiments and several authors have successfully used suppressive subtractive hybridization for identification of pathotype specific sequences in *E. coli* (Bahrani-Mougeot *et al.*, 2001; Janke *et al.*, 2001; Miyazaki *et al.*, 2002; Zhang and Foxman, 2003).

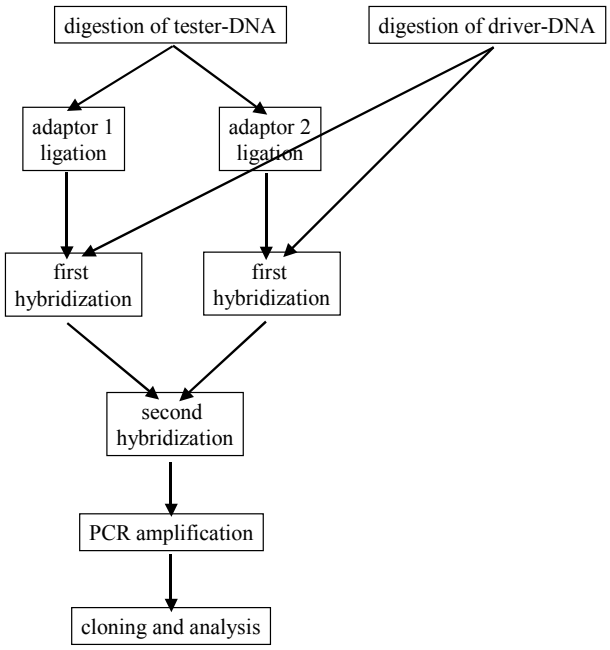

Figure 6. Suppressive subtractive hybridization method developed by Diatchenko and colleagues. Two cycles of hybridization ensure elimination of the sequences shared by the tester- and driver-strains. The DNA fragments unique for the tester-strain are PCR amplified using oligo-nucleotide primers specific for the adaptor sequences ligated to the tester-DNA prior to hybridizations.

2. AIMS OF THE PROJECT

At the beginning of this project in 1999, the vast genomic diversity among species *E. coli* was only anticipated; the initial aim of this project was to characterize the genomic variation in ExPEC strains, which cause urinary tract infections (i.e. UPEC). During this project, it became clear that UPEC strains differ significantly not only from intestinal pathogenic and non-pathogenic *E. coli* strains but also from each other. This extensive heterogeneity posed us a challenge to characterize the fraction of unique DNA in two clinically relevant UPEC strains in order to learn how these flexible genomes are assembled and which mechanisms contribute to the observed genomic mobility. Moreover, we hypothesized that the mobile gene fraction must encode a number of novel virulence determinants, which might contribute to the infection of an extraintestinal site (i.e. the urinary tract). As ExPEC strains seemed to lack a common denominator (i.e. a virulence trait or traits), which could be used to distinguish highly virulent clones from less virulent and non-pathogenic *E. coli*, we decided to develop a novel algorithm, which described the probability of some combination of ExPEC virulence determinants to specify the extraintestinal virulent phenotype.

For these purposes, the following questions were addressed:

- Do genomic contents of ExPEC strains, which belong to the same pathotype (i.e. UPEC) and infect the same anatomical site (i.e. the urinary tract), differ from each other?
- If they do, how extensive is this variation?
- How does extragenous DNA (e.g. genomic islands) differ from the conserved *E. coli* core genome, which encodes catabolic and metabolic functions?
- How are UPEC genomes assembled and how is the observed genomic mobility achieved?
- Do extragenous DNA encode putative virulence or fitness determinants, which might contribute to the ascending urinary tract infections?
- Are there common denominators for ExPEC or UPEC?
- Is it possible to differentiate highly virulent ExPEC clones from less virulent isolates and from commensal *E. coli* strains, which inhabit the gut?

3. MATERIALS AND METHODS

The experimental methods used in this project are described in the original publications and summarized in Table 2. References for the published methods can be found in the original publications (see "List of original publications"). The DNA microarray technology and the suppressive subtractive hybridization method used in this project are summarized in chapters 1.5. and 1.6. Bacterial strains, plasmids, and cosmids are described in detail in the original publications and summarized in table 3. The bacterial strains were stored at -80ºC until use. The oligonucleotide primer sequences used in this project are described in the original publications and summarized in table 4.

3.1 Methods

Table 2. Methods used in this project.

Method	Described and used in study
Culture conditions	
General culture conditions for *E. coli*	I-IV
Culture conditions in bacterial matings and selection of transconjugants	I
Induced iron starvation	I
Methods of molecular biology	
Catecholate siderophore uptake assays (feeding bioassay)	I
Construction of the *iroN* and *iroE* mutant *E. coli* strains by the EZ::Tn-transposon insertion	I
Dot blot hybridization	I, II
Generation of cosmid libraries and selection of the cosmids of interest	I, III
Isolation of chromosomal and plasmid DNA	I-IV
Isolation of mRNA	I
Reverse transcription PCR (RT-PCR)	I
DNA sequencing	I-IV
Suppressive subtractive hybridization of genomic DNA	II
Microarray technology	
Generation and hybridization of DNA microarrays	III, IV
Imaging and analysis of the DNA microarray hybridization signals	III, IV
Determination of the gene-specific threshold values	III, IV
Sequence analysis and statistical tests	
Gaussian approximation algorithm	III, IV
K-mean clustering algorithm	IV
Phylogenetic analysis of nucleotide and protein sequences	I
Sequence alignment using public sequence databases	I-IV
Variance stabilization normalization algorithm	III, IV
Virulence gene algorithm	IV
χ^2 test	I-IV

3.2 Bacterial strains, plasmids, and cosmids

Table 3. Bacterial strains, plasmids, and cosmids used in this project.

Strain, plasmid, or cosmid	Relevant characteristic(s) and description	Reference	Study
E. coli strain			
536	Pyelonephritis, O6:K15:H31, PG[a] B2	Hacker *et al.*, 1992	I-IV
A30	Cystitis, O75	Archambaud *et al.*, 1988	IV
AD110	Cystitis, O6:K2:H1:F7, PG B2	Labigne-Roussel and Falkow, 1988	IV
BL21	Laboratory strain, PG A, F$^-$ *ompT hsdS*(r_B^- m_B) *dcm*+ Tet^r *gal* λ(DE3) *endA* Hte (pLysS Cam^r)	Stratagene, La Jolla, CA, USA	IV
CFT073	Pyelonephritis, O6:K2:H1, PG B2	Mobley *et al.*, 1990	I-IV
CP9	Bacteremia, O4:K54:H5, PG B2	Russo *et al.*, 1998	I-IV
D9	Cystitis, PG A	Schubert[b]	III, IV
DH5α	Laboratory strain, PG A, F$^-$ Φ80d*lacZ*ΔM15Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR*17(r_k-m_k+) *phoA supE*44-*thi*-1 *gyrA*96 *relA*1	Invitrogen, Paisley, UK	I-IV
Do764	Cystitis PG B2	Dobrindt[c]	III, IV
Do768	Cystitis, PG B2	Dobrindt	III, IV
DSM6601	Stool, O6:K5:H1, PG B2	Nissle, 1959	I-IV
EAEC 042	Diarrhea	Henderson *et al.*, 1999	IV

EB1	Intra-abdominal wound infection, O8:K43	Verweij-van Vught *et al.*, 1985	IV
ECOR31	Stool, PG D	Ochman and Selander, 1984; Schubert *et al.*, 2004	I-IV
ED2403	Cystitis, PG A	Dobrindt	III, IV
EDL933	Hemorrhagic diarrhea, O157:H7	O'Brien *et al.*, 1993; Perna *et al.*, 2001	III, IV
H5058	*aroB malT tsx thi cir fiu fepA*	Baumler *et al.*, 1998	I
HB101	*hsdR⁻ hsdM⁻ recA13 supE44 lacz4 leuB6 proA2 thi-l* Sm^r, PG A	Takara, Shiga, Japan	IV
HE300	Pyelonephritis, R:H-, PG A	Sorsa[e]	I-IV
IHE1041	Pyelonephritis, O1:K1:H7, PG B2	Vaisanen-Rhen *et al.*, 1984	IV
IHE1049	Pyelonephritis, O1:K1:H7, PG B2	Vaisanen-Rhen *et al.*, 1984	IV
IHE1086	Pyelonephritis, O4:K12:H1, PG B2	Vaisanen-Rhen *et al.*, 1984	IV
IHE1167	Pyelonephritis, O4:K12:H1, PG B2	Korhonen[d]	IV
IHE1190	Pyelonephritis, O18:K5:H7, PG B2	Vaisanen-Rhen *et al.*, 1984	IV
IHE1210	Pyelonephritis, O1:K1:H7, PG B2	Korhonen	IV
IHE1402	Pyelonephritis, O6:K2:H1, PG B2	Vaisanen-Rhen *et al.*, 1984	IV
IHE1431	Pyelonephritis, O6:K2:H1, PG B2	Vaisanen-Rhen *et al.*, 1984	IV
IHE3040	Newborn meningitis, O18:K1:H7, PG B1	Korhonen *et al.*, 1985	IV
IHE3041	Newborn meningitis, O18:K1:H7, PG B2	Korhonen	IV
IHE3034	Newborn meningitis, O18:K1:H7, PG B2	Korhonen *et al.*, 1985	IV
IHE3047	Newborn meningitis, O18:K1:H7, PG B2	Korhonen *et al.*, 1985	IV
IHE3080	Newborn meningitis, O18:K1:H7, PG B2	Korhonen *et al.*, 1985	IV
IHE11015	Pyelonephritis, O4:K12:H7, PG B2	Korhonen	IV

IHE11020	Pyelonephritis, O6:K2:H1, PG B2	Korhonen	IV
IHE11035	Pyelonephritis, O21:K14:H4, PG A	Korhonen	IV
IHE11048	Pyelonephritis, O75:K100:H5, PG B2	Korhonen	IV
IHE11158	Pyelonephritis, O6:K13:H1, PG B2	Korhonen	IV
IHE11290	Pyelonephritis, O16:K1:H6, PG B2	Korhonen	IV
J96	Pyelonephritis, PG B2	Hull *et al.*, 1981	I-IV
JS293	Chronic cystitis, PG B2, host suffers of *diabetes mellitus*	Schubert	I-IV
JS295	Cystitis, PG B1	Schubert	I-IV
JS297	Cystitis, PG B2, host suffers of *diabetes mellitus*	Schubert	I-IV
JS299	Cystitis, O6:H-, PG B2	Schubert	I-IV
JS304	Pyelonephritis, O1:H7, PG B1	Schubert	I-IV
JS305	Pyelonephritis, PG B2	Schubert	I-IV
JS309	Cystitis, PG B2	Schubert	I-IV
JS322	Pyelonephritis, O77:H18, PG D	Schubert	I-IV
JS323	Pyelonephritis, PG B2	Schubert	I-IV
JS325	Pyelonephritis, PG B2	Schubert	I-IV
JS326	Pyelonephritis, PG B2	Schubert	I-IV
JS355	Pyelonephritis, PG B2	Schubert	I-IV
MG1655	Non-pathogenic K-12 strain, PG A	Blattner *et al.*, 1997	I-IV
Nu14	Cystitis, O18:K1:H7, PG B2	Johnson *et al.*, 2001	I-IV

RS218	Newborn meningitis, O18:K1:H7, PG B2	Achtman *et al.*, 1983	I-IV
S17-1λPIR	Laboratory strain, *tpr* Sm^r *recA thi pro hsdM*$^+$ RP4-2-Tc::Mu::Km *Tn*7 λ*pir*, *hsdR*	Alexeyev and Shokolenko, 1995	I
S5	Bacteriemia, O15:K+:H21	Morris and Drouin, 2004	IV
S2207	Catheter-associated cystitis, PG A	Schubert	I-IV
S2265	Catheter-associated pyelonephritis, PG B2	Schubert	I-IV
S2287	Catheter-associated pyelonephritis, PG D	Schubert	I-IV
S2288	Catheter-associated pyelonephritis, PG B1	Schubert	I-IV
S2413	Urosepsis, PG B2	Schubert	I-IV
S2523	Urosepsis, PG B2	Schubert	I-IV
S2536	Urosepsis, PG D	Schubert	I-IV
S2572	Urosepsis, PG D	Schubert	I-IV
S2608	Cystitis, PG A	Schubert	I-IV
S27710	Pyelonephritis, PG B2	Schubert	I-IV
S4000	Stool, PG D	Schubert	I-IV
S4005	Stool, PG B1	Schubert	I-IV
S4010	Stool, PG B2	Schubert	I-IV
S4011	Stool, PG B2	Schubert	I-IV
S4012	Stool, PG A	Schubert	I-IV
S4013	Stool, PG D	Schubert	I-IV
S4014	Stool, PG D	Schubert	I-IV
S4015	Stool, PG B1	Schubert	I-IV

S4016	Stool, PG B1	Schubert	I-IV
S4017	Stool, PG B1	Schubert	I-IV
S4020	Stool, PG B2	Schubert	I-IV
S4021	Stool, PG B2	Schubert	I-IV
S4022	Stool, PG D	Schubert	I-IV
S4023	Stool, PG A	Schubert	I-IV
S4026	Stool, PG A	Schubert	I-IV
S4039	Stool, PG B1	Schubert	I-IV
S4050	Stool, PG B1	Schubert	I-IV
S4134	Stool, PG A	Schubert	I-IV
S4136	Stool, PG A	Schubert	I-IV
S4138	Stool, PG A	Schubert	I-IV
S4140	Stool, PG A	Schubert	I-IV
S4141	Stool, PG A	Schubert	I-IV
S4147	Stool, PG B1	Schubert	I-IV
S4148	Stool, PG A	Schubert	I-IV
S4149	Stool, PG B2	Schubert	I-IV
S4151	Stool, PG B1	Schubert	I-IV
S4152	Stool, PG A	Schubert	I-IV
S4153	Stool, PG B1	Schubert	I-IV
S4158	Stool, PG B1	Schubert	I-IV

TH2	Laboratory strain, PG A, *supE44 hsdS20* ($r_B^-m_B^-$) *recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1 thi trpR624*	Takara	I
TOP10	Laboratory strain, PG A, F$^-$ *mcrA*(*mrr-hsdRMS-mcrBC*)Φ80*lac*ΔM15Δ*lacX*74 *recA*1 *ara*Δ139Δ(*ara-leu*)7697 *galU galK rpsL* (StrR) *endA1 nupG*	Invitrogen	I-IV
XL-1 Blue MRF´	Laboratory strain, PG A, Δ(*mcrA*)*183*Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F´ *proAB lacI*^{q}*Z*Δ*M15 Tn*10 (TetR)]	Stratagene	I, IV

Plasmids and cosmids:

p300	Plasmid of the UPEC strain HE300 carrying *iroBCDEN* gene cluster	Sorsa	I, IV
pSuperCos1	Cosmid vector	Stratagene	I, III
pJS332	Cosmid vector pSuperCos1 with 32 kb inserted DNA originating from plasmid p300	Sorsa	I
pCR4-TOPO for sequencing	TA cloning vector	Invitrogen	I, II
p[*iroEN*]	Cloning vector pCR4-TOPO with *iroEN* genes originating from plasmid p300	Sorsa	I
p[*iroBCDE*]	Cloning vector pCR4-TOPO with *iroBCDE* genes originating from plasmid p300	Sorsa	I
p[*iroN*:Tp^r-*iroE*]	Plasmid p[*iroEN*] with trimethoprim cassette inserted into *iroN* gene	Sorsa	I
p[*iroN-iroE*:Tp^r]	Plasmid p[*iroEN*] with trimethoprim cassette inserted into *iroE* gene	Sorsa	I
pJS448	Cosmid vector pSuperCos1 carrying chromosomal DNA of the UPEC strain JS299	Sorsa	III, IV
pJS666	Cosmid vector pSuperCos1 carrying chromosomal DNA of the UPEC strain JS299	Sorsa	III, IV

pPJS700	Cosmid vector pSuperCos1 carrying chromosomal DNA of the UPEC strain JS299	Sorsa	III, IV
pJS706	Cosmid vector pSuperCos1 carrying chromosomal DNA of the UPEC strain JS299	Sorsa	III, IV

a Phylogenetic group (PG) defined by the method described by Clermont *et al.* 2000
b Strain kindly provided by Dr. Sören Schubert, Max von Pettenkofer-Institut, University of Munich, Germany
c Strain kindly provided by Dr. Ulrich Dobrindt, Department of Infection Biology, University of Würzburg, Germany
d Strain kindly provided by Dr. Timo Korhonen, Department of Biological and Environmental Sciences, University of Helsinki, Finland
e This project

3.3 Oligonucleotide primer sequences

Table 4. Oligonucleotide primer sequences used in the project.

Target gene	Accession number or reference	Template source	Forward primer 5'→3'	Reverse primer 5'→3'	Study
Homo sapiens					
β-actin	M28424	Human	CTGTGGCATCCACGAAACTA	AGCAATGATCTTGATCCTTCA	IV
E. coli					
adk	NC_004431	CFT073	ATTCTGCTTGGCGCTCCGGG	CCGTCAACTTTCGCGTATTT	IV
afa/draBC	Johnson and Stell, 2000	A30	GGCAGAGGGCCGGCAACAGGC	CCCGTAACGCGCCAGCATCTC	IV
afaD3	X76688	A30	CGGAGATGAACGGGAGTATAAG	GGTATTCACCAGGAGCAATGTC	IV
afaD8	AF072900	IH11165	GAGTCTTAATACCAGTGATGGAAGG	AATATTATGAGCATTCTCCGCTAAC	IV
aufA	AE016768	CFT073	CCATTATTCGTCGTACCTTTGG	TATCCCAGGCTCACTGATATGG	IV
aufC	AE016768	CFT073	GAAAGGTTGCGATACACTAGCC	ATTTCAGGGGGAATTAATTTGG	IV
aufG	AE016768	CFT073	TAGTTTTCCACCAATCTGAACG	TAAATATTCAGGTTCCGCAAGG	IV
$c2397_{serU\ \text{Island}}$	AE016762	CFT073	TGCTCACATTGTTGAAAGTTCC	TCAAGGTAGCTGTTTGACATGG	IV
$c2398_{serU\ \text{Island}}$	AE016762	CFT073	TCATATCCCATGCAAAAGAGG	AGTTTGCAGTGCCATTTTATCC	IV
$c2400_{serU\ \text{Island}}$	AE016762	CFT073	AAGATGGATTCAAGGTCATTGC	ATCCGGCATTGTAAGAAAAGC	IV
$c2416_{serU\ \text{Island}}$	AE016762	CFT073	GGATTTACTGCGTAGGTTCTGG	ATCAAAGGAAATGGATTGTTGG	IV
$c3607_{\text{PAI II}}$	AF447814	CFT073	GCTCGGTTGAAAGTGCC	TCTCAGCGTTGCATGG	IV
c4836	AE016770	CFT073	TGGGTCGTATTCAGCGG	GGGTTCGCAACTCACG	IV
cadA	AE016771	MG1655	GTGGGATACCACCAAGAATACC	TTCTGTACTCCTGGTCACATGG	IV
cdtB	AJ508930	S5	TGCTCTGGTTAGCAATCG	GCTTGTAGCGGAAATGG	IV
chuA	NC_004431	CFT073	CTTCAGCCACTGATTTGC	GCGGTTTCACCATTGC	IV
cnf1	AF483829	J96	GACGTTGATGGCTCAGG	ATTGCCTCTTATCATACGGC	IV

cnf2	U01097	S5	ATTCCTGATGAGGCTCCAGTAG	CTCGAGCAGAATTTGATACACG	IV
csgA	NC_000913	MG1655	AGCAATCGTATTCTCCGGTAGC	GTTACCAAAGCCAACCTGAGTC	IV
cvaC	AJ223631	EB1	AGATTCTGTTTCTGGTGGTGCT	TTCGGGTTTTTGCTTAATTGTT	IV
dsdA	NC_004431	CFT073	TGGAATACGAGCAAGATTATGG	TGGACGCCTAATAACATACACG	IV
dsdC	NC_004431	CFT073	GATATCGGCGTGTAAAGTCACC	CCCTTCGTGAAATAAGAAATCG	IV
east1	AF411067	EAEC 042	CGAACGCTTCCAGTGC	AACAAATCGGTGTTCCTGG	IV
entF	NC_004431	CFT073	TAGCGATATCCCCAATTTAACC	GGCTTCACCACTACAGAAAACC	IV
fecD	AE016759	CFT073	ATTATGCCTGCTATTGCTACCG	CACCCAGTTACCGTCAATTACC	IV
fepA	NC_004431	CFT073	GAACTCCAGACCAATCTCCTTG	GTACAGAGTGGAATCAGCAACG	IV
fimH	AE016771	CFT073	CGGCGTGTTATCTAGTTTTTCC	TAGGTAATACCCCAGGTTTTGG	IV
focA	AF29820	AD110	GGCTGTATTTTCAGCTCTGACC	AGCTTGCAGTTCCATCTAAAGG	IV
fyuA	NC_003143	536	TGCAGCTTTACTCTGGCC	TAGACGGAAATGTTCATCCG	IV
hbp	AJ223631	EB1	TCCCAGTTCGTGCTTACG	ATTTGTCATGCGTACAGACC	IV
hek	AJ494981	536	GGAAGTCATAGTACGCATTCAGC	TCTGTGGTGTCACTTTCAGACC	IV
hlyA	AJ494981	536	ATGCAGATACCGGAACTAAAGC	CTGAAATTATCCCCGTAACAGC	IV
ibeA	AF289032	RS218	AAGCAGGGCAATAATTTACTCG	TTTCCTTTCGCAAGACTTTACC	IV
iha	AF447814	CFT073	TTGATGGTGTTCGTCAGG	TATCATAACGCTGCCGG	IV
ilvD	NC_004431	CFT073	GACCAAACGTTACTACGAACAGG	GTCGTATTGTTCCAGTGTTTGC	IV
*ireA*1	AF320691	CP9	TGGGAAATTTGGTAACTCAACC	TCCGGTTTTCTTGACTCGTAAT	IV
*ireA*2	AF320691	CP9	TGACGAAAGTCTCAATGGC	TACATTAGCCCAATAACGGC	IV
iroBCDEN	AY205565	HE300	TGGTCAGGTTGCGGAGGCTAT'-	GCGAAGGCGAGACTATCAGGA	II
iroD	AY205565	HE300	GGGGCTGAGAAATATCAACATC	GAAGATGAGCGAGAAGTGACCT	IV
iroE	AY205565	HE300	CGATCGATATTCAGTCCCTGTT	CGATAAAGGCTCCGTGTTTTAC	IV
iroN	AY205565	HE300	CAAAACAGTTCTTCCAGTGCAG	GAACCGGCTCAATGTTATCTTC	IV
iss	AY205565	HE300	TTCCAGCGGAGTATAGATGC	ATCACATAGGATTCTGCCG	IV
iucC	NC_004431	CFT073	TTGATAAAATCGCCTTTCAGGT	ACCAGGAAATGCTTGGTGTTAT	IV
iutA	NC_004431	CFT073	CAGACCAGCATTGAACAGAAAG	CAGTACGGCATGAAACTGACTC	IV
mchF	AF302690	536	TATTCGTCTGGTGACATACGG	TTTCATGTAGTGCAATATCTGC	IV
mdh	NC_004431	CFT073	GGCGATATCTTTCTTCAGCG	ATGAAAGTCGCAGTCCTC	IV
modD	NC_004431	CFT073	TTTGTCGAGCTGAAGTACG	TTACCCTGATGGCAATATCG	IV
ompT	NC_004431	CFT073	TTCATCGTTATCAAATGCTTCC	ACACAACTCAATTATGCCAACG	IV
matB	AF325731	IHE3034	TCTATTTGACGTGGCTATCGAG	TAGCGTCGAACTGTACGCTAAC	IV
c1648$_{pmA\text{-}modD\text{-}Island}$	AF081283	CFT073	GTATAGACGCGAAGCCG	AATTGACCTGCAAGGCG	IV

orf4$_{\text{PAI II}}$	AJ494981	536	GCCACTCCCAGAATTACG	TCCGTCACTTGCCATACC	IV
orf40$_{\text{PAI II}}$	AJ494981	536	GGAGGCAATCACCGC	ACAGTGGTCGGGATACG	IV
orf41$_{\text{PAI II}}$	AJ494981	536	TTACGGGCAGGTCTCC	ACAACTGGTTTGTCCGC	IV
orf11$_{\text{PAI III}}$	X16664	536	TTATTGCACACAGAGAAACTACG	TCATTTCAACACCTTCGC	IV
orf12$_{\text{PAI III}}$	X16664	536	CTTCACAAATCTTATAGCGAAGG	CCGCAGACGAAGAGGC	IV
orf36$_{\text{PAI III}}$	X16664	536	CAGTTAAAAGGCGGATATTTCG	AGAGTTACTGACGCCTCAAAGC	IV
orf29$_{\text{p300}}$	AY205565	HE300	GATCACAACAAGCCGGATGT	AGCCAATGAGCGCCAGA	I, IV
orf24 (*iss*)$_{\text{p300}}$	AY205565	HE300	GCTTCCAGCGGAGTATAGAT	CGCTCTGGCAATGCTTATT	I, IV
P761 (*scrA*)	DQ079865	JS299	AGCGGCTTCAGTAATCCCTA	CTTATCTTGGCGCGACAT	III, IV
P762 (*scrY*)	DQ079865	JS299	CCACTCTCCGCCAGTA	GCCAGCAATTCTCGCACAAA	III, IV
P763 (*ukoI*)	DQ079860	JS299	ATGTCATACTCTTTCGCATT	CAAGAAGCCGTGGAC	III, IV
P765 (*ukoC*)	DQ079861	JS299	GGCATTACTTCTGAGTATTG	ATTCATAATCAGGACCAGAT	III, IV
P767 (*ukoE2*)	DQ079861	JS299	GTTTCCCATTCCTTGACTGC	GCGGATGAGATTGCCTATG	III, IV
P769 (*c2163*)	DQ079863	JS299	AATTTCTGTCTGGATTACGG	GACGACGAAGATCAGTGG	III, IV
P700 (*ukoA*)	X16664	536	GTTGCCCGGGGTGATACATCC	TCACTGAGCTGTAATCAAT	III, IV
P774 (*ukoE1*)	DQ079861	JS299	GTTTGAGGTCTGACGCGATA	AGTGCGCTTGCTGATACCAG	III, IV
P776 (*ukoJ*)	DQ079860	JS299	CGCTCTCAATCACTGGAA	CACGCTGACATCGGTGTTGT	III, IV
P780 (*scrB*)	DQ079865	JS299	ACTGGCAGGAAGATAACGA	CAGTACATTGCCATTGGTCA	III, IV
P785 (*ukoK*)	DQ079863	JS299	TTCTCTGACTTCTGGCAACC	TTATGTTCATGGCAGGGAA	III, IV
P792 (*ukoF*)	DQ079860	JS299	GGCCACGCAGACGCAGGGAG	GGGCGACCACACGGTGCT	III, IV
papA	NC_004431	CFT073	AAGTCAGGTTGAAATTCGC	CACCATGTGGTATTGATGC	IV
papF	AE016771	CFT073	GCTGTCACTCTGTAACCATTTCC	TTATATCGTTGCTTCTGACATCG	IV
papGI	AF240678	CP9	AAGGTTATTGACCAACCTCAGC	CTGGCAAAAAGACCCTGAATAC	IV
papGII	AE016771	CFT073	CGTCTGACTGAGAAATTTGACG	TTCTTACCATGGCTGTATGTCG	IV
papGIII	AF237473	CP9	ATGGATTTACCTGGACTCATGG	AACGAAGAAGGGATTTTGTAGC	IV
picU	AF097644	EAEC 042	GTTACCCCGTATCTGCTGAATC	CCAGTAAAAGCTGGGATTTTTG	IV
prfG	AJ494981	536	CCTTCCTAAGGGGGATTATACG	ACCATCTCATCGTTGTCTCTCC	IV
prrA	U85771	CFT073	TGCGGTCATAGTACGAAAGC	TTACAGCGACGGTGCC	IV
sat	NC_004431	CFT073	CTCGTTGTCATGGTGAACG	CGTATGGGTGATCTGCG	IV
sfa/foc	Johnson *et al.*, 2005b	J96	CTCCGGAGAACTGGGTGCATCTTAC	CGGAGGAGTAATTACAAACCTGGCA	IV
sfaA	X16664	536	TTTAAGGGGGAAGTTGTTGATG	TGCCTGGAAAGGAATTTTATTG	IV
sfaS	X16664	536	GGATACGACGATTACTGTGACG	TAACCTGGCCTTTACTCACTGC	IV

SAK450	AY428160	JS322	ATCCCCTCTGGCCGTGGTTC	AGAAGTCATCCGCGAAATAGC	II, IV
SAK448	DQ679231	JS322	ATAGGCGCGTGATACAGG	TAATAACGTAAAATCAATGCTTCC	II, IV
SAK464	DQ409208	JS322	GCCAGGGCAGAGCTTGTTTTA	ATGGTAATAATAGATGTTTTC	II, IV
SPL107	AY428158	HE300	GATACATAGCTTTTGGGTCTG	GCTGCAGGGTAGGGTGAT	I, II, IV
SPL135	DQ662933	HE300	AGTTATTTATCTTTGAGTGG	CCAAGCATGCGGGTGTG	II, IV
SPL220	DQ409207	JS304	GGGCTTCATCAGTAATCAGGGTAA	TGGTCCAGGCACAGATAGAAGTAG	II, IV
SPL247	AL627270	JS304	ATGCGGTTTATCTCAATCTT	GGTACTGGGCTTCCGTTGCTGGTT	II, IV
SPL284	DQ409206	JS304	GCGGCCGAGGTACTGGTTCC	CATGATGCCGTTTTCTGTAT	II, IV
SPL341 (*ukoD/E*)	DQ079861	JS299	CGTATTTTGTAAAACAGTATA	AGCGAGGGAACAGTTCAGGG	II, III, IV
SPL345	AY428148	JS299	GACTATCGCTCGAACCTCCTCCTT	TTGATTATGCCTTACGCCAGATGC	II, IV
SPL346 (*ukoG*)	DQ079860	JS299	GTTGATATGAAAATCAGGGG	CAAACCGCGCTGAATGGTGATT	II, III, IV
SPL359 (*ukoK/c2163*)	DQ079863	JS299	GTACCCTTGTTCTGCCGATTT	GGTGTTTCAAAATGATTTAGG	II, III, IV
SPL373 (*scrB*)	DQ079865	JS299	TACCATTGGCTGTTGTGCTTC	GTACGACCGTAATGGTTGTT	II, III, IV
SPL386	DQ409201	JS299	TCAGGTATATTCAGTGGCGTCAGT	TATTGAAGGCGCGGACGACCAT	II, IV
SPL395 (*ukoH*)	DQ079860	JS299	CCTAACAATGCCAGAAGTGT	ATTATGGTTATGGTAAGCTAT	II, III, IV
SPL764	DQ079860	JS299	GTCCAGTTCGGCAGTCCAC	TCGCAGTGTCATTCCCGTAT	II, IV
T3/T7	Evans *et al.*, 1989	Standard cloning vectors	ATTAACCCTCACTAAAG	AATACGACTCACTATAG	II, II, III
tonB	NC_004431	CFT073	GATTTCTGTCACGATGGTTGC	CTTGGCTGAGAGGATTTGTACG	IV
usp	AB056440	536	TGACATTCACGGCAAGC	CTCCTTCAGGTAAACATAGATGG	IV
$ydfB_{p300}$	AY205565	HE300	ATATCACGGGAAAGGAAGCAC	GTCCGGCAAGGTTTATCA	II
yjaA	NC_000913	MG1655	TGAAGTGTCAGGAGACGC	GAGTACCGCTGATTGCG	IV

3.4 Nucleotide sequence accession numbers of the sequences

Table 5. The nucleotide sequence accession numbers (NCBI) of the sequences described in this project.

Sequence designation	Accession number	Study
Cosmid p332	AY205565	I, IV
Cosmid pJS448	DQ079861	III, IV
Cosmid pJS666	DQ079860	III, IV
Cosmid pJS700	DQ079863	III, IV
Cosmid pJS706	DQ079865	III, IV
Subtraction fragment SAK448	DQ679231	II, IV
Subtraction fragment SAK450	AY428160	II, IV
Subtraction fragment SAK464	DQ409208	II, IV
Subtraction fragment SPL107	AY428158	II, IV
Subtraction fragment SPL135	DQ662933	II, IV
Subtraction fragment SPL220	DQ409207	II, IV
Subtraction fragment SPL247	AL627270	II, IV
Subtraction fragment SPL284	DQ409206	II, IV
Subtraction fragment SPL341	DQ409205	II, IV
Subtraction fragment SPL345	AY428148	II, IV
Subtraction fragment SPL346	DQ409204	II, IV
Subtraction fragment SPL359	DQ409203	II, IV
Subtraction fragment SPL373	DQ409202	II, IV
Subtraction fragment SPL386	DQ409201	II, IV
Subtraction fragment SPL395	DQ409200	II, IV
Subtraction fragment SPL764	DQ079860	II, IV

4. RESULTS AND DISCUSSION

4.1 Analysis of genomic variation of uropathogenic *E. coli* (UPEC; Study II)

ExPEC are a diverse group of *E. coli* strains, which inhabit the gut, and occasionally infect different extraintestinal sites, such as the meninges, the blood stream, and the urinary tract. Since beginning of the century, it is known that the ability of these strains to cause a disease is dependent upon specific gene repertoires (i.e. virulence factors). Although a number of such virulence factors have been identified in these strains, none of the factors is uniquely known to define these pathogens. Even if whole-genome comparisons of *E. coli* strains causing different diseases (i.e. intestinal and extraintestinal infections) have shown great differences in gene contents (Chen *et al.*, 2006; Perna *et al.*, 2001; Welch *et al.*, 2002), there is growing epidemiological, clinical, and experimental evidence, however, to suggest that differences exist even within ExPEC strains that cause the same disease and target the same host tissues (Dobrindt *et al.*, 2003; Johnson *et al.*, 2005b; Johnson *et al.* 2005d; Russo and Johnson, 2003). To identify potential virulence genes or genetic markers of ExPEC, we used genomic subtractive hybridization to determine genetic differences between four clinical UPEC isolates, an archetypal UPEC isolate, and a commensal *E. coli* isolate. Evaluation of the degree of similarity between these five strains, which cause the same disease, revealed a high degree of diversity, with only a few shared sequences. Subsequently, additional *E. coli* strains of clinical and fecal origin were screened by dot blot hybridization, and the results provided further evidence for the existence of a high degree of genome plasticity among the ExPEC pathogroup. In addition, we identified yet unknown sequences associated with the UPEC phenotype, which might be of medical importance in that they could serve as valuable genetic markers to identify and discriminate potentially pathogenic UPEC strains from less pathogenic UPEC isolates and from the commensal *E. coli* strains belonging to the normal intestinal flora.

4.1.1 Genomic subtraction of the UPEC strains HE300, JS299, JS304, and JS322

Genomic DNA of four previously uncharacterized UPEC isolates HE300, JS299, JS304, and JS322 (tester-strains) was each subtracted from the chromosomes of the archetypal UPEC strain CFT073 and the commensal *E. coli* strain MG1655 (driver-strains; Blattner *et al.*, 1997; Welch *et al.*, 2002) using the suppressive subtractive hybridization method (Diatchenko *et al.*, 1996). The driver-strains CFT073 and MG1655 are well-characterized strains representing UPEC and non-pathogenic *E. coli*, respectively. The genomic sequences of these strains are available at the public sequence databases of the NCBI and the EBI, which allowed direct sequence comparisons. The suppressive subtractive hybridization method is discussed in detail in chapter 1.6.

We sequenced and characterized a total number of 384 subtraction fragments originating from four separate subtraction fragment pools (Table 6). Of these 384 fragments, 172 (45%) revealed sequences specific for the four tester-strains and were absent from the genomes of the driver-strains CFT073 and MG1655. In addition, a half of the tester-specific subtraction fragments (n=86) was shown to be highly homologous (>85% nucleotide sequence identity) with sequences deposited in the public nucleotide databases (Table 6). These 86 sequences represented mobile genetic elements, including bacteriophages, transposable elements (n=10), and *E. coli* and *Salmonella* sp. plasmids (n=27), GIs of the UPEC strain 536 (n=22; Dobrindt *et al.*, 2002), *S. flexneri* 2a strain YSH600 (n=3: Turner *et al.*, 2003), and the EHEC strain EDL933 (n=14; Perna *et al.*, 2001) as well as other genomic loci (n=10; Table 6).

Table 6. Summary of the 384 subtraction fragments obtained from the UPEC strains HE300, JS299, JS304, and JS322.

Subtraction library	Clinical characteristics	PG group[a]	Number of sequences (%)			
			Tester-specific	>85% homology[b]	<85% homology[b]	No homology
HE300	Pyelonephritis	A	34 (100%)[c]	17 (50%)	14 (41%)	3 (9%)
JS299	Cystitis	B2	62 (100%)	32 (52%)	29 (47%)	1 (1%)
JS304	Pyelonephritis	B2	42 (100%)	25 (60%)	13 (31%)	4 (9%)
JS322	Pyelonephritis	D	34 (100%)	12 (35%)	21 (62%)	1 (3%)
Total			172 (100%)[d]	86 (50%)	77 (45%)	9 (5%)

a Phylogenetic group (PG) defined by the triplex-PCR method described by Clermont *et al*. 2000
b Percentage of the homology regarding to the total sequence length
c Percentage relating to the total number of sequences in a subtraction library
d Percentage relating to the total number of sequences in the four subtraction libraries

Interestingly, each of the four UPEC strains used as tester-strains (HE300, JS299, JS304, and JS322) gave a distinct pattern of subtraction fragments with regard to the origin of their respective homologue (Table 7). The pyelonephritogenic *E. coli* tester-strain HE300, which belonged to the PG A, revealed almost exclusively subtraction fragments, which were highly homologous (>95% identity) with the *S. enterica* serotype Typhimurium plasmid pR64 (Gyohda *et al.*, 2004) and different *E. coli* plasmids. Subtraction of the cystitis strain JS299 (PG B2) essentially generated sequences associated with the PAIs I, II, and III of the UPEC strain 536. Subtraction library of the pyelonephritogenic strain JS304 (PG B2) revealed DNA fragments similar to the genomic DNA of *Shigella* spp. and the EHEC strain EDL922. Finally, subtraction library of the pyelonephritogenic strain JS322 (PG D) consisted mainly of sequences homologous with the chromosome of strain EDL922. Even though only a few UPEC strains have been reported to possess plasmids (Leflon-Guibout *et al.*, 2004; Ojo *et al.*, 2003; Rodriguez-Siek *et al.*, 2005), half of the strain-specific sequences (n=27) originating from strains HE300, JS304, and JS322 were highly similar to conjugative plasmids described in *E. coli*, *Klebsiella pneumoniae*, *Shigella* spp., and *Salmonella* sp. (Table 7). The second large category of subtraction fragments (n=14) was homologous with the GIs of EHEC. These subtraction fragments were detected in all UPEC strains studied (Table 7).

Table 7. Summary of the 86 subtraction fragments highly homologous (>85%) with known sequences deposited in the public sequence databases (n=86)

Subtraction library	Number of sequences (%)						
	Total	Plasmids	Mobile genetic elements	UPEC GIs	EHEC GIs	*Shigella* GIs	Others
HE300	17 (100%)[a]	10 (58%)	3 (18%)	0 (0%)	2 (12%)	0 (0%)	2 (12%)
JS299	32 (100%)	0 (0%)	0 (0%)	22 (69%)	2 (6%)	0 (0%)	8 (25%)
JS304	25 (100%)	15 (60%)	4 (16%)	0 (0%)	3 (12%)	3 (12%)	0 (0%)
JS322	12 (100%)	2 (17%)	3 (25%)	0 (0%)	7 (58%)	0 (0%)	0 (0%)
Total	86 (100%)[b]	27 (31%)	10 (12%)	22 (26%)	14 (16%)	3 (3%)	10 (12%)

a Percentage relating to the sequence number of highly homologous sequences in a subtraction library
b Percentage relating to the sequence number of highly homologous sequences in the four subtraction libraries

Another half of the tester-specific subtraction fragments (n=86) displayed either low (55% to 82% identity; n=77) or no homology (n=9) with sequences in the public nucleotide databases (Table 6). These DNA fragments represented the "low homology fraction" and exhibited short sequence stretches related to the chromosome of the UPEC strain CFT073 (n=10) as well as to putative genes of bacteriophages, prophages, and GIs of the EHEC strain EDL933 (n=11; Table 8). However, majority of the tester-specific subtraction fragments with low homology (n=56) were similar to sequences of 39 different bacterial species, both Gram-negative and Gram-positive. Interestingly, 51 subtraction fragments (59%) shared short sequences with putative protein coding genes with conserved functional domains. This might indicate the presence of novel groups of genes with conserved function (Table 8).

Table 8. Summary of the 86 subtraction fragments with low (<85%) or no sequence homology with known sequences deposited in the public sequence databases (n=86)

Subtraction library	Number of sequences (%)				
	Total	Conserved genes	EHEC GIs or bacteriophages	Intergenic region	No homology
HE300	17 (100%)[a]	8 (46%)	3 (18%)	3 (18%)	3 (18%)
JS299	30 (100%)	18 (60%)	6 (20%)	5 (17%)	1 (3%)
JS304	17 (100%)	8 (46%)	1 (6%)	4 (24%)	4 (24%)
JS322	22 (100%)	17 (77%)	1 (5%)	3 (13%)	1 (5%)
Total	86 (100%)[b]	51 (59%)	11 (13%)	15 (17%)	9 (11%)

a Percentage relating to the sequence number of low homology sequences in a subtraction library
b Percentage relating to the sequence number of low homology sequences in the four subtraction libraries

4.1.2 Virulence association study of the 86 subtraction fragments showing no or only low homology with known sequences deposited in the public databases

Subtractive hybridization of the UPEC strains HE300, JS299, JS304, and JS322 with strains CFT073 and MG1655 yielded a pool of 86 novel DNA fragments, which had no or only low homology with known sequences in the public databases (Table 8). In order to determine if these sequences were associated with the UPEC phenotype, we performed dot blot hybridizations of chromosomal DNA (Ausubel *et al.*, 1989) to look for the presence of the specific sequences in 90 additional strains of the ExPEC pathotype and fecal *E. coli*. The results indicated that among the 86 sequences examined, 12 sequences were significantly associated ($P<0.05$) with ExPEC strains when compared to commensal *E. coli* strains isolated from stool samples (Table 9). Interestingly, two subtraction fragments SPL135 and SPL345, which were recovered from the UPEC strains HE300 and JS299, respectively, were significantly more prevalent among strains of the commensal phenotype. Reliability of the dot hybridizations in the case of the most prominent virulence gene candidates (i. e. regarding to their prevalence among ExPECs and commensal *E. coli* isolates, and their putative function in bacterial virulence) was further validated by a DNA microarray analysis (Study IV).

Interestingly, only few ExPEC-associated sequences were shared by the five UPEC strains even though they cause clinically similar diseases. In addition, most of the subtraction fragments were nonuniformly distributed among ExPEC strains (e.g. not according to the phylogenetic distribution or the clinical characteristics of strains) or were present in only a relatively small subset of strains; therefore, they may have been acquired separately by rather recent horizontal gene transfers. This

comparison of genomic differences between fecal *E. coli* and highly pathogenic ExPEC isolates predicted potential virulence-associated elements, whose characterization could facilitate identification of factors responsible, in part, for the UPEC pathogenicity. To further describe the DNA regions adjacent to five ExPEC-associated subtraction fragments; we constructed cosmid libraries of the total DNA of the UPEC strains HE300 and JS299. Using the ExPEC-associated subtraction fragments SPL107, SPL341, SPL359, SPL373, and SPL395 (Table 9; highlighted with an asterisk) as DNA probes to screen the cosmid libraries HE300 (SPL107) and JS299 (SPL341, SPL359, SPL373, and SPL395), five novel cosmid clones were identified. These five cosmids were partially sequenced and characterized and the structural and the epidemiological data are presented in studies I and III.

Table 9. BlastN (NCBI) homology search of the 14 subtraction fragments highly associated (P<0.05) with the extraintestinal virulent or the commensal *E. coli* phenotype

Subtraction fragment	Size (nt)	Origin[a]	Sequence homology or designation	Source	Accession number[b]	Association P[c]
SAK448	374	JS322	*ratA,* putative outer membrane protein	*E. coli* CFT073	AE014075	0.05
SAK450	347	JS322	*blr0961*, hypothetical protein	*Bradyrhizobium japonicum*	AP005938	0.01
SPL107*	156	HE300	*lin0414*, hypothetical protein	*Listeria innocua*	AL596164	0.007
SPL135	430	HE300	*sso_0233*, hypothetical Hcp-like protein	*S. sonnei*	CP000038	<0.001
SPL220	735	JS304	*estA*, ABC transporter	*E. coli* A2363 pAPEC-ColV	AY545598	0.01
SPL247	199	JS304	ST41626, tail core protein	*S. enterica*	AL627270	0.01
SPL284	572	JS304	16S ribosomal DNA	*Caulobacter* sp.	AY568471	0.01
SPL341*	606	JS299	No homology			0.06[d]
SPL345	557	JS299	No homology			<0.001
SPL346	230	JS299	Intergenic region	*Nostoc* sp.	BA000019	0.001
SPL359*	374	JS299	16S ribosomal DNA	*Acinetobacter radioresistens*	AY568469	0.1[d]
SPL373*	234	JS299	*ygcF*, hypothetical protein	*E. coli* T19	AJ639630	0.004
SPL386	512	JS299	16S ribosomal DNA	*Acinetobacter radioresistens*	AY568469	<0.001
SPL395*	589	JS299	16S ribosomal DNA	Uncultured bacterium	AB128887	<0.001

a The UPEC origin of a subtraction fragment
b Accession number of the respective homologue
c Significance of association with the extraintestinal virulent or commensal *E. coli* phenotype was determined by a dot blot hybridization assay and by a DNA microarray analysis. Using χ^2 test a *P* value of <0.05 was considered as statistically significant.
d The subtraction fragments SPL341 and SPL359 showed significant association with the ExPEC strain collection used in the dot blot hybridization analysis (data not shown)
* Subtraction fragments used as DNA probes to screen the cosmid libraries of the UPEC strains HE300 and JS299

4.1.3 The key aspect of the UPEC research; the UPEC pathogroup is a highly diverse group of strains with a considerable amount of unknown DNA

The genome sequence determination of UPEC and EHEC, and the genomic comparison with a commensal *E. coli* strain MG1655 revealed that the *E. coli* chromosome is comprised of a 4.1 Mb backbone chromosome common to all *E. coli* isolates and of up to 1.4 Mb strain specific sequences. Strain specific sequences are inserted at various sites on the chromosome backbone, and encode strain specific genes, many of which are virulence related (e.g. the type III secretion system of intestinal pathogenic *E. coli*; Bergthorsson and Ochman, 1998; Blattner *et al.*, 1997; Dobrindt and Hacker, 2001; Perna *et al.*, 2001; Welch *et al.*, 2002).

In this study, we analyzed genomes of four ExPEC strains, which presented similar pathotypes (i.e. UPEC) and infected the same anatomical site (i.e. the urinary tract) by a whole genome genome comparison analysis called suppressive subtractive hybridization. In this analysis, we hybridized genomic sequences of four clinical UPEC isolates (HE300, JS299, JS304, and JS322) with chromosomes of the archetypical UPEC strain CFT073 and the commensal *E. coli* strain MG1655, and searched for DNA fragments unique for the four unknown UPEC strains. This analysis provided a good view on the genomic variety of UPEC, and revealed that there is an unexpectedly high degree of diversity among UPEC strains; the subtraction libraries of the UPEC strains differed not only from the commensal *E. coli* strain MG1655 but also from the archetypal UPEC strain CFT073 and from each other. The subtractions of some strains largely resulted in PAI-associated subtraction fragments, whereas other strains gave DNA fragments homologous with the genomic DNA of EHEC or plasmids of *E. coli* and *Salmonella* sp.

The obtained sequences were divided into three categories, as shown in table 6. In summary, the first category included 170 sequences, which were shared either by the four clinical UPEC isolates and by strain CFT073 or by strain MG1655. These sequences were homologous with the *E. coli* core genome or with putative or verified ExPEC virulence factors, such as components of iron uptake systems, toxins, and fimbriae. The second category included 86 sequences, which were absent from the driver strains, but were highly homologous (>85%) with known sequences in the public databases (Table 6). These sequences represented diverse ExPEC virulence genes as well as mobile genetic elements such as bacteriophages, transposons, plasmids, and genomic islands (Table 7). The third category included 86 sequences with no known function that could be divided into two types: fragments with similarity to sequences of unknown function (e.g. conserved hypothetical proteins) and fragments for which no similarity was found (Table 8).

None of the known ExPEC virulence determinants was uniformly found in all the five UPEC strains tested. This and the fact that only 75% of the subtracted sequences were previously characterized indicated that there is substantial genomic variability in gene content even among "typical" UPEC strains with similar pathogenic properties. Secondly, it suggested that the UPEC pathogenesis is a complex phenomenon resulting from a combination of gene products rather than being determined by a single locus. The presence of novel sequences in highly virulent UPEC strains, which are significantly associated with fecal *E. coli*, implied that adaptation to diverse environmental niches (i.e. the gut and the urinary tract) might be the driving force for the genetic variation of the ExPEC genomes. Moreover, the striking extent of genome variation in this set of isolates suggested that analysis of UPEC and other ExPEC isolates with a broader pathogenic representation might reveal even greater variation than that found in the current study.

A large fraction of DNA sequences was homologous with mobile genetic elements (56%; Tables 7 and 8) and was variably present in different strains. Especially the subtractions of strains

HE300 and JS304 resulted in unexpected large number of plasmid-derived sequences (58% and 60% of the sequences, respectively). This showed that mobile genetic elements, in particular plasmids, are likely to have a significant impact on the spread of virulence traits throughout the UPEC genomes, and on the genomic diversity of these strains in general. However, we are left with many questions about the genomic diversity and its effect on the gene contents or the pathogenic characteristics of each strain.

A large fraction of the subtraction fragments (25%) had no or only very low (<85%) sequence homology with known sequences in the public databases (Tables 6 and 8). It is an interesting speculation that the strain-specific gene fraction of UPEC encodes a number of novel virulence or fitness determinants or diagnostically important genetic markers associated with the UPEC phenotype. The most outstanding result of this study was the identification of 14 DNA sequences significantly more prevalent among ExPEC strains when compared with commensal *E. coli* isolates. The identification of new virulence genes in the UPEC pathotype would contribute to our understanding of how these bacteria cause disease and provide diagnostic tools to distinguish potentially pathogenic ExPEC strains from their less virulent counterparts.

4.2 Identification and functional characterization of a plasmid-encoded *iroBCDEN* gene cluster (Study I)

In the previous study (II), we analyzed genome sequences of four highly virulent UPEC isolates by selectively subtracting the chromosomal fraction shared by ExPEC and commensal *E. coli*. In this analysis, we searched for novel genomic sequences, which would be preferentially present in ExPEC but absent from *E. coli* of fecal origin. This analysis revealed 14 hitherto unknown DNA sequences, which were significantly associated with the ExPEC phenotype. In the present study (I), we characterized DNA regions adjacent to the virulence associated subtraction fragment SPL107 originating from the pyelonephritogenic *E. coli* strain HE300. By construction of a cosmid library, followed by colony hybridization, selection, and DNA sequencing, a novel virulence plasmid (p300) was identified, which carried a chromosomal virulence gene island (*iroBCDEN*) encoding an iron acquisition system (i.e. siderophore salmochelin and its receptor IroN). The plasmid encoded *iroBCDEN* gene cluster was shown to be functional as it promoted uptake of different catecholate siderophores in a siderophore negative *E. coli* strain. Moreover, of capital importance is that plasmid p300, and with it the functional *iroBCDEN* gene cluster, were shown to be transmissible in *E. coli* and to have the potential capacity to spread throughout bacterial populations by conjugation.

4.2.1 Characterization of the cosmid pJS332 sequence

The ExPEC-associated subtraction fragment SPL107 (Table 9) was identified as part of cosmid pJS332 originating from the pyelonephritogenic *E. coli* strain HE300. Cosmid pJS332 carried a 32-kb DNA insert, which appeared to be a composite of genes derived from plasmids of *Salmonella* sp. and *E. coli*, and from GIs of the UPEC strain 536 and the EHEC strain EDL933.

A total of 31 ORFs were identified in the pJS332 sequence (ORFs 1 to 31; Fig. 6). Seven ORFs had no significant similarity to nucleotide or protein sequences in the public databases, while 24 putative polypeptides were very similar to characterized proteins of *S. enterica* and *E. coli* (Table 10). With regard to the sequence homologies and the G+C contents, the pJS332 sequence exhibited a modular structure with three distinct regions (regions I to III; Fig. 6). Region I revealed 99% homology with the *S. enterica* plasmid R64. These genes encoded components of a putative ABC transporter system (YcjA, YdaA, IbfA, YdfA, and YdfB) and proteins involved in plasmid maintenance, transfer, and replication (Mck, Kor, YdiA, and YdjA). The 5'- and 3'- ends of this region were flanked by sequences highly homologous with *E. coli* transposons (Fig. 6; Table 10). A total of 11 putative ORFs (ORFs 2 to 13) were identified in this region. Region II harbored ORFs 15 to 20 and was highly homologous with the chromosomal *iroBCDEN* gene clusters of the UPEC strains CFT073 and 536 (Russo *et al.*, 2002, Welch et al, 2002). The 5'-border of region II was composed of a truncated IS carrying ORFs 21 and 22 (Fig. 6; Table 10). Region III extended from ORFs 23 to 31 and revealed sequences related to ISs and plasmids as well as a number of hypothetical ORFs with no significant similarity to sequences in the public databases (Table 10). The deduced AA sequence of ORF 29 (region III) was highly homologous with the Iss virulence determinant of APEC, which is involved in increased serum resistence and survival of these strains in the chicken respiratory tract infection model (Tivendale *et al.*, 2004). The subtraction fragment SPL107 used as the initial DNA probe was localized on non-coding DNA of region III (Fig. 6; Table 9).

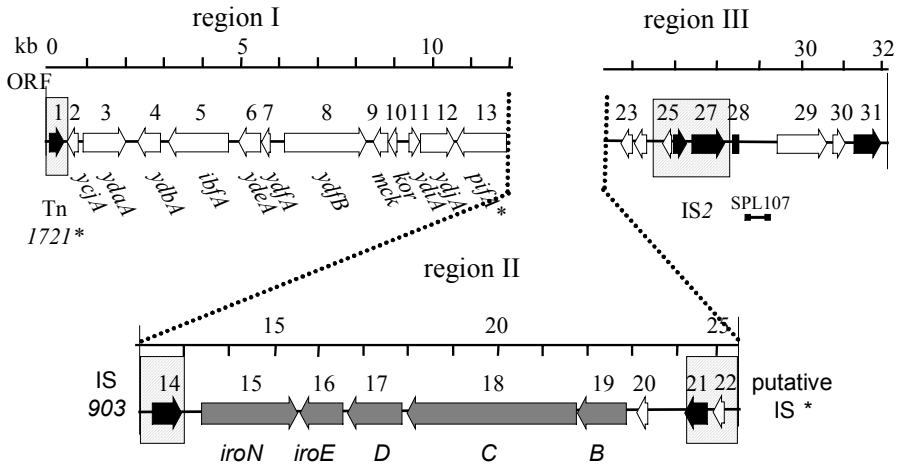

Figure 6. Schematic presentation of the cosmid pJS332 sequence originating from the UPEC strain HE300.

The overall G+C content of the pJS332 DNA insert was 49%, which is about the average for the *E. coli* core genome (51%). However, the G+C content differed significantly within the pJS332 sequence (23% to 72%) displaying peaks and troughs. These variations corresponded to different functional units (e.g. mobile genetic elements and plasmids) indicating a mosaic DNA structure composed of elements of apparently different origins (Fig. 7).

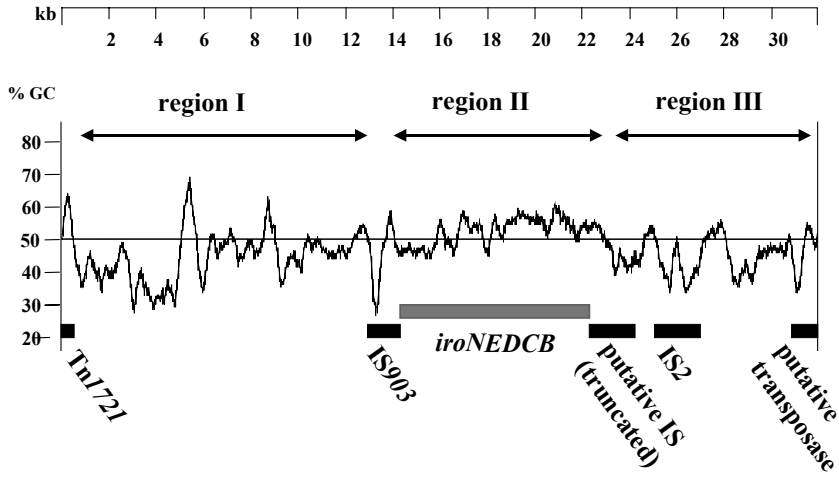

Figure 7. Schematic presentation of the G+C content variation (%) within the cosmid pJS332 sequence.

Table 10. BLASTP (NCBI) homology search of the cosmid pJS332 sequence. A total of 31 ORFs were identified in the cosmid pJS332 sequence, which originated from *S. enterica* and *E. coli* plasmids, GIs of UPEC and from additional diverse and unknown sources.

ORF (gene)	Size (AA)[a]	ORF location (nt)[b]	Source	Protein sequence homology or designation	Identity		Accession number[c]	DNA probe[d]
					%	AA		
1	136	413	*E. coli*	Truncated ORF, TnpA of *Tn*1721	95	135	P51565	
2	107	791-68	*S. enterica* pR64	YcjA, putative bacterial repressor protein	100	107	BAB91595	
3	377	896-2029	*S. enterica* pR64	YdaA, putative permease	90	377	BAB91596	
4	205	2939-2322	*S. enterica* pR64	YdbA, hypothetical protein	96	205	BAB91597	
5	518	4686-3130	*S. enterica* pR64	IbfA, putative ABC transporter	95	518	BAB91598	
6	196	5539-4949	*S. enterica* pR64	YdeA, hypothetical protein	100	196	BAB91599	
7	85	5796-5539	*S. enterica* pR64	YdfA, putative ABC transporter	100	85	BAB91600	
8	712	6150-8288	*S. enterica* pR64	YdfB, putative ABC transporter	100	712	BAB91601	*ydfB*
9	133	8851-8450	*S. enterica* pR64	Mck, involved in coordination of plasmid replication with cell division	99	133	BAB91602	
10	76	9093-8863	*S. enterica* pR64	Kor, involved in coordination of plasmid replication with cell division	100	76	BAB91603	
11	96	9389-9679	*S. enterica* pR64	YdiA, 93% identical to protein YebA on *E. coli* plasmid F	100	96	BAB91604	
12	299	9669-10568	*S. enterica* pR64	YdjA; 100% identical to protein YebB on *E. coli* plasmid F	100	299	BAB91605	
13	443	11949-10618	*S. enterica* pR64	Truncated ORF, PifA, phage T7 exclusion protein	100	443	BAB91606	
14	224	12299-12973	*IS*903	TnpA, transposase of *IS*903	99	223	AAF30382	
15	725	13385-15562	*E. coli* 536 PAI III	IroN, enterochelin and dihydrobenzoic acid receptor	99	725	CAC43424	*iroN*
16	318	16563-15607	*E. coli* 536 PAI III	IroE, putative hydrolase	99	318	CAC43425	*iroE*
17	409	17877-16648	*E. coli* 536 PAI III	IroD, putative ferric enterochelin esterase	99	409	CAC43426	*iroD*
18	1261	21766-17981	*E. coli* 536 PAI III	IroC, putative ABC transporter	99	1261	X16664	
19	371	22895-21780	*E. coli* 536 PAI III	IroB, UDP-glucoronosyl and UDP-glucosyl transferase	100	371	X16664	

20	89	23363-23094	No homology	Hypothetical protein				
21	173	24692-24171	*E. coli* K5	Hypothetical protein, putative integrase	59	100	CAA54707	
22	96	25164-24874	*E. coli* K5	Hypothetical protein	59	48	CAA54710	
23	97	26026-25733	*E. coli*	Hypothetical protein	80	76	AAB40752	
24	102	26347-26039	*E. coli* 102	Iss, increased serum survival and complement resistance	100	102	AAD41540	*orf24*
25	80	26948-26706	*E. coli* EDL933	Hypothetical protein	100	26	AE005326	
26	112	26959-27297	*E. coli*	Putative transposase of IS*2*	98	109	P51026	
27	263	27404-28195	*E. coli*	Putative integrase of IS*2*	97	262	P51026	
28	54	28530-28366	*Y. enterocolitica*	Hypothetical protein, *trp*1328 on IS*1328*	63	50	CAB46575	
29	397	29418-30611	*E. coli* CFT073	Hypothetical protein, putative cobalamin synthesis protein	65	397	NP_753181	P*orf29*
30	103	30750-31061	*S. flexneri*	Hypothetical protein, putative integrase	59	82	NP_709264	
31	207	31248-31869	*S. enterica* pHCM1	Truncated ORF, HCM1.201, putative transposase	95	207	NP_569402	

a Predicted polypeptide size (AA)
b Location of a gene on the cosmid pJS332 sequence (nt)
c Accession number of the respective homologue
d DNA probes used in dot blot and DNA microarray analyses (Studies II and IV). The DNA probe homologous with *ydfB* was excluded from the DNA microarray analysis.

4.2.2 Identification of the conjugative plasmid p300 carrying *iroBCDEN* gene cluster

Region I of pJS332 displayed sequences highly homologous with the *S. enterica* plasmid R64, which suggested a plasmid origin of the DNA inserted into this cosmid (Table 10). We affirmed that pJS332 originated from a novel 79-kb plasmid denoted as p300, which conferred resistance to trimethoprim, streptomycin, and tetracycline. Plasmid p300 isolated from the UPEC strain HE300 was transformed into the *E. coli* laboratory strain S17-1λPIR, which carried a chromosomal transfer region required for the conjugative transfer of plasmids (Alexeyev and Shokolenko, 1995). From this strain, p300 was further mobilized into the *E. coli* laboratory strain Th2 by mating.

In this study, we showed that the conjugative transfer of the *iroBCDEN* gene cluster-carrying plasmid p300 was highly efficient in *E. coli*. This and the fact that plasmid p300 seemed to originate from the *S. enterica* plasmid R64, supported the possibility that plasmid p300 is transmissible by conjugation into commensal and pathogenic bacteria such as *E. coli* and *S. enterica*. It is also likely that this plasmid transfer will occur naturally, for example in the gut environment, and thus can play an important role in the spread of virulence and drug resistance factors among *E. coli* and related bacteria.

4.2.3 Distribution of the p300-derived sequences among ExPECs and commensal *E. coli* isolates

To determine the distribution of the plasmid-carried *iroBCDEN* gene cluster among ExPECs, dot blot hybridizations were performed using DNA probes homologous with *iroN* (region II) and *ydfB* (region I; Table 10). The *iroN* sequence was found in 29 out of 50 ExPEC strains studied (58%). The DNA probe homologous with the *S. enterica* plasmid pR64 (*ydfB*) was identified in six ExPEC strains (12%). Only two strains were positive for both *iroN* and *ydfB*. This data demonstrated that the *iroBCDEN* gene cluster described in this project was not liked to the plasmid p300 in the vast majority of UPEC strains, but rather was chromosomally located or at least linked to a different genetic backbone.

4.2.4 Functional characterization of the *iroBCDEN* gene cluster encoded by the plasmid p300

To investigate the function of the plasmid encoded *iroBCDEN* gene cluster, the ability of this gene cluster to promote uptake of catecholate siderophores enterochelin and 2,3-dihydroxybenzoyl-D-ornithine was tested by a plate assay. We showed that complementation of the siderophore-negative *E. coli* strain H5058 with *iroN*, but not with *iroB*, *iroC*, *iroD,* or *iroE* was sufficient to promote the siderophore uptake by this siderophore-negative *E. coli* strain (Fig. 8). This confirmed that plasmid p300-encoded IroN functions as receptor for siderophores enterochelin and 2,3-dihydroxybenzoyl-D-ornithine.

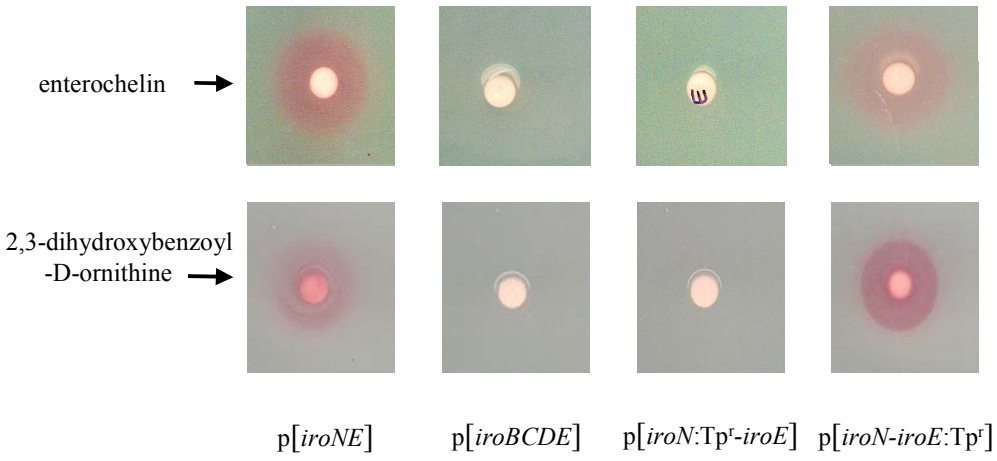

Figure 8. The feeding bioassay (plate assay) to detect the function of plasmid encoded IroN$_{p300}$. The *E. coli* [*fepA cir fiu aroB*] mutant strain H5058 complemented with *iroN*$_{p300}$ (p[*iroNE*], p[*iroN-iroE*:tp^r]), but not with *iroBCDE* (p[*iroBCDE*], p[*iroN*:tp^r-*iroE*]) was able to utilize the catecholate siderophores enterochelin and 2,3-dihydroxybenzoyl-D-ornithine as sole iron sources. Bacteria were grown in an iron minimal medium where the growth was visible as halos surrounding the disk plates soaked into enterochelin or 2,3-dihydroxybenzoyl-D-ornithine.

4.2.5 Virulence determinants are effectively transferred in, between, and from the ExPEC pathogroup through conjugation

In this study, a novel 75-kb virulence plasmid from an UPEC strain was partially sequenced and analyzed, providing new insights into molecular mechanisms underlying the dispersion of virulence factors among UPEC. This plasmid, p300, contained a chromosomal virulence gene cluster *iroBCDEN*, which has been shown to contribute significantly to the virulence of UPEC and NMEC *in vivo* (Hantke *et al.*, 2003; Negre *et al.*, 2004; Russo *et al.*, 2002). It also carried resistance against three antibiotics. Furthermore, plasmid p300 was shown to be transmissible by conjugation.

Sequence differences of individual ORFs located in the plasmid p300 suggested that various regions of this plasmid had different origins; this plasmid encoded genes highly similar to a chromosomal PAI, transmissible plasmids of *S. enterica*, as well as to a number of other mobile genetic elements. This indicated that pathogenic ExPEC isolates might evolve gradually under selective pressure of the host environments by adopting genetic elements, such as plasmids and PAIs, from closely and distantly related species. In addition, mobile genetic elements, such as phages, transposons, and ISs , might be involved in this process either directly by transferring genetic material or indirectly by providing substrate for the homologous recombination (Kuzminov, 1999).

Recently, a number of ExPEC virulence factors (e.g. salmochelin, aerobactin, and yersiniabactin siderophore systems as well as the Hbp protease homologue Tsh) have been identified in APEC, an *E. coli* subgroup very closely resembling ExPEC (Janben *et al.*, 2001; Rodriguez-Siek *et al.*, 2005; Tivendale *et al.*, 2004). Several of these virulence determinants in APEC are localized on large transmissible plasmids similar to plasmids pR64 of *S. enterica* and p300 described in this study (Dozois *et al.*, 2003; Gyohda *et al.*, 2004; Johnson *et al.*, 2004; Johnson *et al.* 2005e; Rodriguez-Siek *et al.*, 2005). Interestingly, plasmids have been shown to be transferable from poultry to human *E. coli* isolates (Johnson *et al.*, 2005e; Levy *et al.*, 1976).

These results support our data of the novel urovirulence plasmid p300, and indicate that conjugative transfer is indeed an important means of transfer of virulence factors in UPEC. This transfer is likely to occur both in natural environments and in the host (e.g. in the gut). Co-localization of virulence factors and antibiotic resistence genes on the same transmissible plasmid may promote selection of *E. coli* containing such plasmids in humans and animals treated with antibiotics (Levy *et al.*, 1976; Laporta *et al.*, 1986; Johnson *et al.*, 2005e). In this scenario, virulence factors are readily transferred through plasmids not only between ExPEC strains but also between ExPECs and their close relatives, such as APEC, enhancing the virulence of infecting pathogens in general. Furthermore, this also suggest that farm animals colonized with *E. coli* carrying p300 -like plasmids may serve as a reservoir of resistance and virulence genes for other bacteria of animal and human health importance. Further studies of UPEC plasmids, such as p300, should facilitate identification of factors responsible, in part, for the ExPEC gene transmission and pathogenicity.

4.3 Characterization of four novel DNA sequences of the UPEC strain JS299 highly associated with the extraintestinal virulent phenotype (Study III)

In this study, we investigated the genetic information encoded by the nonhomologous fraction of the UPEC strain JS299 genome. Furthermore, we examined whether these sequences might be a source for horizontal gene transfer and for generation of genomic diversity in ExPEC. Sequencing and comparison of four cosmid sequences, JS448, JS666, JS700, and JS706, identified both common *E. coli* and *Shigella* sp. sequences as well as for strain JS299 unique ORFs, suggesting that individual cosmids expressed distinct features. From database comparisons, we identified putative genes involved in the regulation of conjugative transfer (*finO*, *sfh*), genes found in genomic islands of UPEC (*orf44* of PAI II_{536}) and in the chromosome of EPEC as well as gene sequences homologous to that of an operon present in the EPEC strain T19 responsible for sucrose utilization. Most unexpected, the *scr* gene cluster of cosmid pJS706 had the same organization as the corresponding chromosomal genes of strain T19, which suggested an exchange of gene clusters between chromosomes of EPEC and UPEC.

The epidemiological distributions of 17 novel cosmid-derived DNA sequences were determined among ExPEC and commensal *E. coli*. From this, 13 sequences were found at significantly higher prevalence among pathogenic isolates and thus could be suitable as new genetic markers for ExPEC isolates. Interestingly, obtained segregation patterns of DNA probes excluded the genetic linkage of different genetic modules, which provided new insights into assembly of the UPEC and the ExPEC genomes in general.

4.3.1 Sequence analysis of the cosmids pJS448, pJS666, pJS700, and pJS706

The ExPEC-associated subtraction fragments SPL341, SPL395, SPL359, and SPL373 (Table 9), which were derived from the highly virulent UPEC strain JS299, were found in four distinct cosmid sequences pJS448, pJS666, pJS700, and pJS706 (Table 11). DNA regions neighboring these subtraction fragments were partially sequenced by primer walking, which revealed DNA sequences of 3200 nt to 6100 nt (Fig. 9 to 12; Table 11). Separate cosmid regions differed markedly in their respective nucleotide compositions; the G+C contents of distinct cosmid inserts varied not only from the average G+C content of the *E. coli* chromosome, but also between the adjacent ORFs (29 to 61%; Table 11.). A total of 19 ORFs were identified, 12 of which were highly homologous with hypothetical genes of diverse *E. coli* and *Shigella* spp. strains. The remaining seven ORFs had no or only very low homology (<85%) with the sequences in the public databases (Table 11). Figure 9 shows the partially analyzed cosmid sequences (pJS448 [Fig. 9a], pJS666 [Fig. 9b], pJS700 [Fig. 9c], and pJS706 [Fig. 9d]). Subfigure **A** illustrates the variation of the G+C content in a cosmid sequence (EMBOSS; http://emboss.sourceforge.net). Additionally, subfigure **B** shows the compositionally different DNA regions, which are denoted by roman numbers and separated by vertical dotted lines. Predicted ORFs are marked as arrows showing the direction of translation. Black arrows mark the conserved *E. coli* core genome, gray arrows represent the DNA regions specific for a certain *E. coli* or *Shigella* strain, and white arrows mark the unique DNA sequences of the UPEC strain JS299. The hatched boxes highlight DNA regions with similar genetic composition. Black bars below the genes show the location of DNA probes used in the DNA microarray analysis (Study IV). The significance of association with the extraintestinal virulent phenotype (Chapter 4.4.2) is given in parenthesis. Significant association with the ExPEC phenotype ($P<0.05$) is indicated with an asterisk.

The subtraction fragment SPL341 was identified on cosmid pJS448 (Fig. 9a; Table 11). The 5'-boundary of the inserted DNA of pJS448 was highly homologous with the *E. coli* core genome (*yicO*; NC_000913; Fig. 9a). In contrast, the 3'- end of pJS448 had sequences similar to the recently characterized Type I site-specific restriction-modification system of *S. boydii* (*hsdR*; NZ_AAKA01000002; Fig. 9a). Analysis of the 5363-nt DNA insert of pJS448 revealed a mosaic structure of three different DNA regions with regard to the genetic configuration, the sequence homology, and the G+C contents (regions I, II, and III; Fig. 9a). Five putative protein coding genes (*ukoA* [unknown ORF; ORF1], *ukoB* [ORF2], *ukoC* [ORF3], *uko*D [ORF4], and *ukoE* [ORF5]) located in this sequence, showed polypeptides similar to proteins involved in regulatory functions as well as to proteins previously identified in mobile genetic elements (Table 11). Interestingly, the 5'-truncated *ukoA* (region I) was identical with the 3'-end of a hypothetical gene located on PAI II of the UPEC strain 536 (*orf44*; AJ494981; Fig. 9a, Table 11).

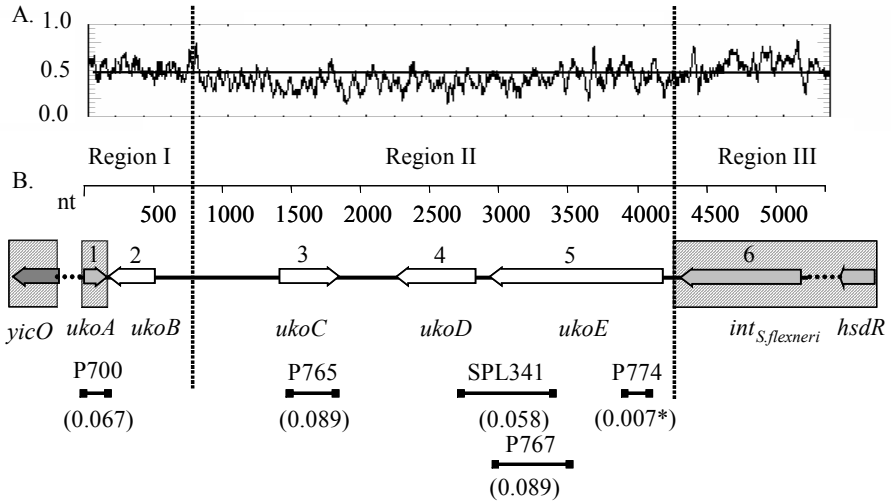

Figure 9a. Schematic presentation of the partial cosmid pJS448 sequence originating from the UPEC strain JS299.

The subtraction fragment SPL395 was localized on cosmid pJS666 (Fig. 9b; Table 11). The 5'-end of pJS666 revealed sequences highly homologous with the *E. coli* core genome (*ppdB*; NC_000913; Fig. 9b). The 3'- sequence of the inserted DNA in pJS666 disclosed the 3'- border of the characterized 4960-nt sequence read (Fig. 9b). Cosmid pJS666 was assembled of three compositionally different DNA regions (regions I, II, and III; Fig. 9b). Regions I (*ukoF* [ORF1], *ukoG* [ORF2]) and III (*ukoJ* [ORF5]) were highly homologous with the genomic DNA of EPEC. However, region II was unique for the UPEC strain JS299. This region was composed of two ORFs (*ukoH* [ORF3] and *ukoI* [ORF4]), which encoded putative proteins without any considerable sequence identities to known sequences in the public databases (Fig. 9b; Table 11). The G+C content of pJS666 varied from 10% to 87% with observed peaks and troughs that corresponded to different regions (Fig. 9b).

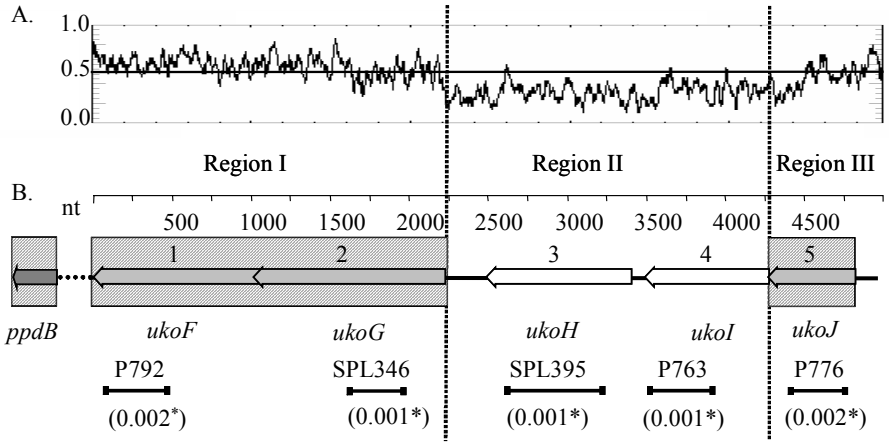

Figure 9b. Schematic presentation of the partial cosmid pJS666 sequence originating from the UPEC strain JS299.

The subtraction fragment SPL359 was located on cosmid pJS700 (Fig. 9c; Table 11). The 5'- and 3'- boundaries of the inserted DNA of pJS700 were highly homologous with the *E. coli* core genome (*katE*, *sppA*; NC_000913; Fig. 9c). The 3229-nt sequence of pJS700 revealed a duplication of two DNA fragments (direct repeats 1 and 2; Fig. 9c). Three putative ORFs were identified in the pJS700 sequence (the 3'- truncated *c2163*$_{JS299}$' [ORF1], *c2163*$_{JS299}$ [ORF3], and *ukoK* [ORF2]; Fig. 9c). The two intact ORFs (*c2163*$_{JS299}$ and *ukoK*) located in this sequence encoded hypothetical proteins predictably involved in protein-protein interactions (pfam06251; Table 9c). Interestingly, the degree of homology with known sequences differed within the c2163$_{JS299}$ sequence; the 5'-end of c2163$_{JS299}$ shared 95% AA sequence identity with its homologue c2163 in the UPEC strain CFT073, whereas AA residues at the 3'- end of c2163$_{JS299}$ were only 36% identical to the respective parts of protein c2163$_{CFT073}$ (Table 11). The composite structure of *c2163*$_{JS299}$ was also reflected by a decrease of the G+C content from 46% (nt 1 to nt 600) to 34% (nt 601 to nt 1092; Fig. 9c).

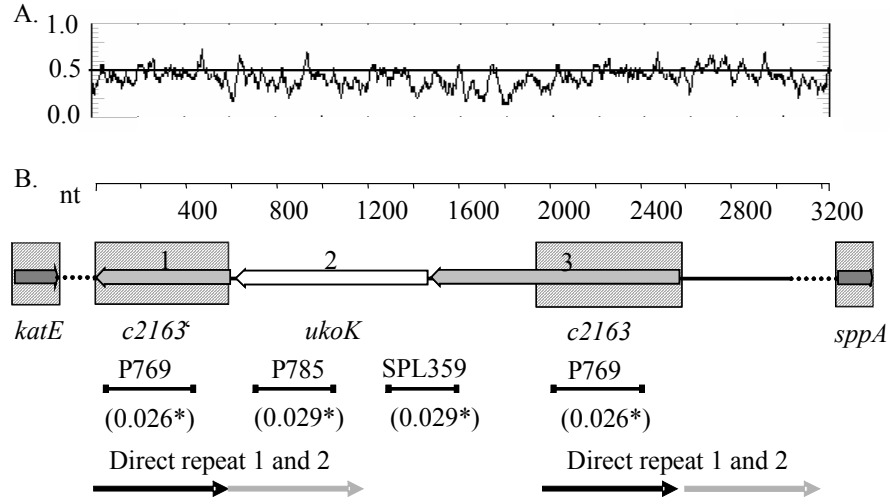

Figure 9c. Schematic presentation of the partial cosmid pJS700 sequence originating from the UPEC strain JS299.

The subtraction fragment SPL373 was identified on cosmid pJS706 (Fig. 9d; Table 11). The 5'- and 3'- boundaries of pJS706 were homologous with the *E. coli* core genome (*pyrG*, *ygbO*; NC_000913; Fig. 9d). The characterized 6183-nt DNA region of pJS706 had G+C content of 47% and could be divided into regions I (*ygcF*) and II (*scrR*, *scrB*, *scrA*, and *scrY*; Fig. 9d). Region I (*ygcF*) was conserved among *E. coli*, *Shigella* spp., and *Salmonella* sp. ORFs *scrR* to *scrY* (region II) were highly homologous (99%) with a recently described *scr* gene cluster encoding a novel sucrose utilization system in the EPEC strain T19 (Fig. 9d; Table 11; Cowan *et al.*, 1991; Gosset, 2005). The border of regions I and II was marked by a 59- nt AT-rich sequence (Fig. 9d). Interestingly, highly similar sucrose-utilizing systems (i.e. *scr* gene clusters) have been identified also in other enteric bacteria, such as *K. pneumoniae* and *Salmonella* sp. In *Salmonella*, the conjugative plasmid pUR400 confers the ability to utilize sucrose, whereas the *scr* regulon of *K. pneumoniae* is located on the chromosome (Sprenger *et al.*, 1988; Wohlhieter *et al.*, 1975).

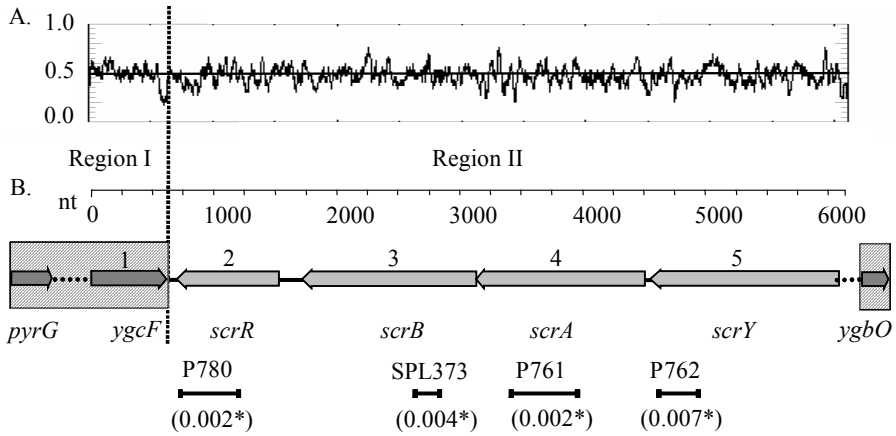

Figure 9d. Schematic presentation of the partial cosmid pJS706 sequence originating from the UPEC strain JS299.

In summary, comparison of the four novel cosmid sequences showed that there is substantial genomic variation among UPEC isolates, particularly in the strain specific gene pool of the genome. In addition, we observed modular genomic structures with insertion of sequence elements of apparently different origins. This indicated the existence of a "mix- and match" -combinatorial system, which might contribute to the high level of genome plasticity among ExPEC.

Table 11. BLASTN (NCBI) homology search of the cosmid pJS448, pJS666, pJS700, and pJS706 sequences. A total of 19 ORFs were identified, which were derived from genomes of *E. coli* and *S. flexneri* and from additional diverse and unknown sources.

ORF (gene)	Size (AA)[a]	ORF location (nt)[b]	G+C %	Source	Protein sequence homology or designation	Identity %	Identity AA	Accession number[c]	DNA probe[d]
Cosmid pJS448									
1 (*ukoA*)	19	1-58	53	*E. coli* 536	Truncated ORF, Orf44, hypothetical protein of PAI II	100	58	CAD42059	P700
2(*ukoB*)	114	179-520	52	*E. coli* ECOR27	FinO, bacterial conjugation repressor	33	82	AAC44278	-
3 (*ukoC*)	145	1414-1848	36	*S. flexneri*	Sfh, H-NS–like protein	46	78	AAN38840	P765
4 (*ukoD*)	190	2265-2834	37	*E. carotovora*	ECA0829, putative membrane protein	51	159	YP_048939	SPL341
5 (*ukoE*)	418	2935-4188	43	*V. cholerae*	Deoxyguanosine-triphosphate triphosphohydrolase –related protein	50	406	AAF96216	P774 P767
6 (*int*)	289	4318-5184	55	*S. flexneri*	S1093, integrase of IS*3*	100	287	NP_836727	-
Cosmid pJS666									
1 (*ukoF*)	354	1062	61	*E. coli* E22	EcolE2_01003258, hypothetical protein	96	307	ZP_00728432	P792
2 (*ukoG*)	405	1014-2228	56	*E. coli* E22	EcolE2_01003259, hypothetical protein	96	391	ZP_00728433	SPL346
3 (*ukoH*)	307	2485-3405	31	*E. coli* CFT073	C1890, hypothetical protein	31	181	NP_753788	SPL395
4 (*ukoI*)	263	3490-4278	34	*E. coli* CFT073	C1889, hypothetical protein	34	64	NP_753787	P763
5 (*ukoJ*)	184	4271-4822	45	*E. coli* E22	EcolE2_01003260, hypothetical protein	80	172	ZP_00728434	P776
Cosmid pJS700									
1 (*c2136*-)	197	593	45	*E. coli* CFT073	Truncated ORF, C2163, hypothetical protein	87	197	NP_754057	-
2 (*ukoK*)	282	621-1466	40	*Magnetococcus* sp.	COG0790, TPR repeat protein	35	237	ZP_00291368	P785 SPL359
3 (*c2136*)	364	1478-2569	40	*E. coli* CFT073	C2163, hypothetical protein	64	360	NP_754057	P769

Cosmid pJS706

1 (*ygcF*)	204	1-611	51	*E. coli* CFT073	Truncated ORF, YgcF, hypothetical protein	92	202	NP_755213	-
2 (*scrR*)	279	696-1531	46	*K. pneumoniae*	ScrR, sucrose utilization repressor protein	79	221	CAA47977	P780
3 (*scrB*)	470	1729-3137	49	*E. coli* T19	ScrB, sucrose 6-phosphate hydrolase	99	338	CAG25843	SPL373
4 (*scrA*)	457	3137-4507	46	*E. coli* T19	ScrA, sucrose and maltose specific permease of the PTS system	92	456	CAG25844	P761
5 (*scrY*)	506	4568-6086	48	*E. coli* T19	ScrY, sucrose and maltose porin	89	359	CAG25845	P762

a Predicted polypeptide size
b Location of an ORF in a cosmid sequence
c Accession number of the respective homologue
d DNA probe used in the DNA microarray analysis

4.3.2 Association of the novel DNA sequences with the ExPEC phenotype

Since sequence data does not provide adequate information on the epidemiology of novel DNA sequences of the UPEC strain JS299, we determined the distribution of 15 previously uncharacterized cosmid-associated genes (17 DNA probes) in a group of clinical ExPECs and commensal *E. coli* isolates by means of DNA microarray analyses (Studies III and IV).

In the ExPEC population, these DNA probes ranged in prevalence from less than 20% (*scrY, ukoC, ukoD/E, ukoE1, ukoK*, and *ukoK/c2163*) to more than 60% (*ukoA* and *ukoF*; Table 12). In total, 13 out of 17 novel DNA sequences were found at significantly higher frequency ($P<0.05$) among ExPEC isolates when compared to commensal *E. coli* strains isolated from stool samples (Table 12). These DNA probes included one DNA fragment targeting the pJS448 sequence (*ukoE*; Fig. 9a), five probes marking the pJS666 sequence (*ukoF, ukoG, ukoH, ukoI*, and *ukoJ*; Fig. 9b), three DNA fragments specific for genes *c2136* and *ukoK* located on the cosmid pJS700 sequence (Fig. 9c) and finally, four DNA fragments homologous with the *scr* gene cluster of cosmid pJS706 (*scrA, scrB, scr* and *scrY*; Fig. 9d). Of note, the *scr* gene cluster was explicitly found in ExPEC strains, as all *E. coli* strains isolated from stool samples failed to hybridize with the pJS706-specific DNA probes (Table 12). Interestingly, the obtained hybridization patterns of distinct genes differed significantly among cosmid sequences. For example, the unique genes *ukoC*, *ukoD*, and *ukoE* of cosmid pJS448, *ukoH,* and *ukoI* of cosmid pJS666 as well as the *scr* gene cluster of pJS706 showed a pattern of segregation different of that in surrounding sequences.

Here we presented 13 novel DNA sequences originating from an UPEC strain that are highly associated with the ExPEC phenotype, and which may serve as new genetic markers for ExPEC isolates. In addition, the data of this study suggested that the analyzed cosmid sequences were genetically flexible and thus they might segregate independently of one another. In the future, expression- and function studies of the novel genes described here should provide evidence of whether these clusters of hitherto unknown genes contribute to the UPEC pathogenesis.

Table 12. Prevalence of 17 cosmid-associated DNA sequences among ExPECs and commensal *E. coli* isolates.

DNA microarray probe	ExPEC n=53 (Prevalence%)	Commensal *E. coli* n=37 (Prevalence%)	P[a]
Cosmid pJS448			
P700 (*ukoA*)	34 (64%)	17 (46%)	0.067
P765 (*ukoC*)	7 (13%)	1 (3%)	0.089
SPL341 (*ukoD/E*)	8 (15%)	1 (3%)	0.058
P767 (*ukoE1*)	7 (13%)	1 (3%)	0.089
P774 (*ukoE2*)	15 (28%)	2 (5%)	0.007*
Cosmid pJS666			
P792 (*ukoF*)	47 (89%)	22 (59%)	0.002*
SPL346 (*ukoG*)	20 (38%)	2 (5%)	0.001*
SPL395 (*ukoH*)	16 (30%)	1 (3%)	0.001*
P763 (*ukoI*)	16 (30%)	1 (3%)	0.001*
P776 (*ukoJ*)	18 (34%)	2 (5%)	0.002*
Cosmid pJS700			
P769 (*c2136*)	18 (34%)	5 (14%)	0.026*
P785 (*ukoK*)	7 (13%)	0 (0%)	0.029*
SPL359 (*ukoK/c2136*)	7 (13%)	0 (0%)	0.029*
Cosmid pJS706			
P780 (*scrR*)	12 (23%)	0 (0%)	0.002*
SPL373 (*scrB*)	11 (21%)	0 (0%)	0.004*
P761 (*scrA*)	12 (23%)	0 (0%)	0.002*
P762 (*scrY*)	10 (19%)	0 (0%)	0.007*

a The significance of association with the ExPEC phenotype was determined by a DNA microarray analysis. Using the χ^2 test, a P value of <0.05 was considered as statistically significant. The significant association is indicated with an asterisk.

4.3.3 The impact of metabolic functions on the ExPEC virulence

In this study, we argue that resolving the molecular mechanisms of sugar metabolism, in particular sucrose and its regulation might provide an important theme for investigating the host-pathogen interactions in intestinal and extraintestinal environments. We identified an *E. coli* sucrose specific phosphotransferase system (the *scr* gene cluster of cosmid pJS706) of UPEC, which was found in 20% of the clinical ExPEC isolates but in none of the commensal *E. coli* strains studied (Fig. 9d). Interestingly, this gene cluster was highly homologous (>90%) with the previously described *scr* operon of EPEC.

The impact of metabolic functions on *E. coli* pathogenesis became evident in the past decades, when effective iron acquisition systems were shown to be required for the bacterial virulence (Bullen *et al.*, 2005). More recently, evidence has been accumulated that besides iron utilization other metabolic functions may also account for the ability of *E. coli* strains to colonize and infect a host. For instance, the ability to use deoxyribose of degraded DNA or mucosal saccharides, such as gluconate, N-acetylglucosamine, and mannose, as carbon sources enables *E. coli* effectively colonize the mucous surfaces of the intestine (Bernier-Febreau *et al.*, 2004; Chang *et al.*, 2004). In addition, metabolic genes may not only be involved in the accumulation of limited nutrients, but may also be implicated in the regulation of virulence factors. The catabolic enzyme D-serine deaminase (DsdA) of UPEC has been shown to modulate growth and gene expression (e.g. flagella) in response to the urinary D-serine levels and thus influence the *E. coli* uropathogenesis (Roesch *et al.*, 2003). The finding of an ExPEC-associated sucrose utilization gene cluster presented in our study was unexpected, as less than 50% of the *E. coli* strains are able to use sucrose as the sole carbon source (Cowan *et al.*, 1991). As the *scr* gene cluster described here was explicitly detected in clinical *E. coli* isolates, we hypothesize that sucrose metabolism may play a role in the ExPEC fitness or infectivity. This role could be mediated either directly by conferring upon the pathogen the ability to use this nutrient or indirectly as a mediator of regulatory functions. However, the results presented here must be considered with certain caution, as we provided no data considering the prevalence of the *scr* gene cluster among intestinal pathogens, such as EPEC, *Shigella* spp., or *Salmonella* sp. In addition, there are yet no studies available regarding the impact of sucrose catabolism on the extraintestinal virulence or colonization. Nevertheless, the detection of metabolic genes highly associated with the ExPEC phenotype may further emphasize a causal relationship between nutritional factors and the pathogenesis of ExPEC.

4.3.4 The "puzzle theory": UPEC genomes are mosaics of individual genetic modules of different origins

The sequences derived from the genome of the UPEC strain JS299 revealed complex mosaic structures (Study III). These mosaics were assembled of compositionally different, short DNA modules integrated into the common *E. coli* backbone (Fig. 9). The novel genes were characterized by atypical G+C contents (Table 11), which indicated a foreign origin of these DNA sequences (Sharp and Li, 1987; Wang *et al.*, 2001). In addition, the distinct cosmid-associated genes exhibited divergent patterns of distribution among ExPECs (Table 12), which excluded the genetic linkage of individual DNA modules.

As ExPEC have shown to inhabit widely diverse niches, such as the gut, the vagina, and the urinary tract (Johnson *et al.*, 1998; Siitonen *et al.*, 1992), adaptation to the changing environments might be the driving force for the genetic variation in these strains. However, the mechanisms leading to the diversification of the ExPEC genomes are largely unknown. The UPEC sequences described in

this study showed genetically flexible modular structures, "gene puzzles", which might enable integration or deletion of genetic modules. This makes such non-conserved gene islands ideal candidates for an efficient adaptation mechanism in ExPEC. The possibility to combine a gene or a domain from one gene with another gene, some regulatory element (e.g. promoter or transcriptional regulator), or different genetic background (e.g. plasmid) might produce chimeric genes and gene islands that demonstrate unexpected functional properties. The advantage of such a modular organization is somewhat obvious; ExPEC could obtain and distribute a high number of diverse genetic modules in and from *E. coli* and other bacterial populations. Different combinations of modules could have been created by recombination (deletion and insertion) and selected for by the needs of ExPECs in their individual hosts. Due to the modular structure, ExPEC might either pick up chromosomal genes of other organisms or integrate sequence modules from foreign plasmids, which are taken up by the bacteria during their natural transformation competence. From conjugative plasmids, such as p300 (Study I), the novel sequences could be rapidly distributed within the UPEC and other ExPEC populations, and exchange novel plasmid-carried sequences with the bacterial chromosomes. Such events might help to explain the vast diversity among ExPEC strains and their ability to effectively colonize and inhabit different anatomical sites and hosts (e.g. domestic animals and humans).

The "puzzle theory" was further supported by a finding of novel DNA sequences derived from UPEC that were highly homologous with the genomic DNA of enteropathogenic *E. coli* (Table 11). Moreover, these new DNA sequences had a significant association with the extraintestinal virulent phenotype (Table 12). Interestingly, although EPEC and UPEC strains infect anatomically and therefore physiologically different body sites (i.e. the intestine and the urinary tract, respectively), some virulence determinants, such as the cytotoxic necrotizing factors and the Afa/Dr family adhesins, are shared by both pathotypes (Bouguenec and Servin, 2006; Nougayrede *et al.*, 2005). Many of the potential virulence factors in both intestinal and extraintestinal pathogenic *E. coli* are encoded by mobile genetic elements, including plasmids, PAIs, and bacteriophages, and the association between DNA transfer and virulence is well documented (Hacker and Carniel, 2001; Ochman *et al.*, 2005). The identification of ExPEC-associated sequences common to two pathotypes is a further indication of the dynamic and extensive DNA exchange, which continuously remodels genome contents of pathogenic *E. coli*.

4.4 Development and validation of a clinical DNA microarray specific for ExPEC causing UTIs (Study IV)

ExPEC are intestinal commensals as well as pathogens of the urinary tract and other extraintestinal sites. Due to the extreme heterogeneity of the UPEC and the ExPEC populations in general, routine surveillance of these pathogens is very difficult. A number of phenotypic and genotypic methods, including serotyping, ribotyping, phylogenetic typing, and virulence gene typing, have been developed to detect and categorize virulent ExPEC clones (Bingen-Bidois *et al.*, 2002; Clermont *et al.*, 2000; Clermont *et al.*, 2001; Morin and Hopkins, 2002). Although these DNA-based methods have been useful for the identification of ExPEC strains, they are limited in that they identify only a few kinds of strains (e.g. according to the serotype, the phylogenetic group, or a limited number of virulence genes) at a time and thus do not often correlate to the virulence of ExPEC strains. Because common clinical ExPECs belong to a wide range of serotypes as well as possess numerous virulence genes and genetic markers, the aforementioned methods cannot meet the requirement of quick and simultaneous identification of various ExPEC clones.

In this study, we describe a sensitive and specific method for rapid and accurate identification of clinical and resident ExPECs within a single DNA microarray ("ExPEC pathoarray"). This pathoarray was composed of 44 PCR probes homologous with known ExPEC virulence genes encoding toxins, proteases, iron acquisition systems, and adhesion factors. Additional 33 PCR probes represented putative markers of the ExPEC phenotype, which in part, were identified during this project. These probes included hypothetical genes located in GIs of UPEC (Study III) and the virulence plasmid p300 (Study I) as well as a number of ExPEC associated subtraction fragments (Study II). Chromosomal DNA of *E. coli* isolates (target DNA) was labeled with the fluorescent dye Cy^5-dCTP.

As the obtained information volume of the ExPEC pathoarray described here would be largely inapprehensible for routine diagnostic laboratories, the array was combined with a novel virulence gene algorithm (VGA) for estimation of the extraintestinal virulence potential (VP, pathogenicity risk) of clinically relevant ExPECs and fecal *E. coli* isolates. Performance of the ExPEC pathoarray was tested by estimating VPs of a number of known and unknown ExPECs (n=53) and commensal *E. coli* strains (n=37; Table 14). Furthermore, relations of the VPs, clonal origins (i.e. phylogenetic groups), clinical data, and virulence gene profiles of distinct strains were determined. We showed that both the ExPEC pathoarray and the VGA described in this study could be employed for detection and classification of unknown ExPEC strains both in epidemiological surveys and in routine diagnostics.

4.4.1 The prevalence of indicator genes among clinical ExPECs and commensal *E. coli*

Applicability of the ExPEC pathoarrays was validated using 90 clinical and commensal *E. coli* strains, which were isolated in hospitals in Finland or Germany (Table 14). In ExPEC strains investigated, DNA probes ranged in prevalence from less than 30% (e.g. the Afa/Dr adhesins, cytotoxic necrotizing factors, and S fimbriae) to more than 70% (e.g. the yersiniabactin siderophore receptor FyuA and Auf fimbriae; Study IV). Of the 77 DNA probes included on the pathoarray, 61 sequences were identified as good indicators of the extraintestinal virulent phenotype (i.e. were associated with ExPEC isolates; $P<0.05$). These probes included 25 PCR fragments homologous with known ExPEC virulence factors (e.g. toxins, adhesins, and siderophore receptors; Fig. 10) and 36 PCR fragments homologous with putative virulence determinants (e.g. hypothetical ORFs associated with GIs of UPEC; Fig. 11). The discriminative power of individual DNA probes between the extraintestinal virulence and the

commensalisms was expressed as the "correct classification rate", and varied from less than 0.3 (e.g. subtraction fragments SPL135 and SPL345 (Study II), which were associated with the commensal *E. coli* phenotype) to more than 0.7 (e.g. the known ExPEC virulence determinants *aufA*, *aufG*, *chuA*, *fyuA, iroD, iroE, iroN, iucC, papF, papGII, prfG, prrA,* and *usp*; Fig. 10). This value characterized the potential of a single gene to represent the extraintestinal virulent phenotype. Classification rates close to one corresponded to nearly perfect indictors, whereas rates closer to zero indicated indifferent genes (i.e. not representative of the ExPEC phenotype). Interestingly, none of the ExPEC-associated indicator genes could be uniformly detected in all ExPEC strains or in an ExPEC subgroup (e.g. according to the clonal origin or the clinical signs caused).

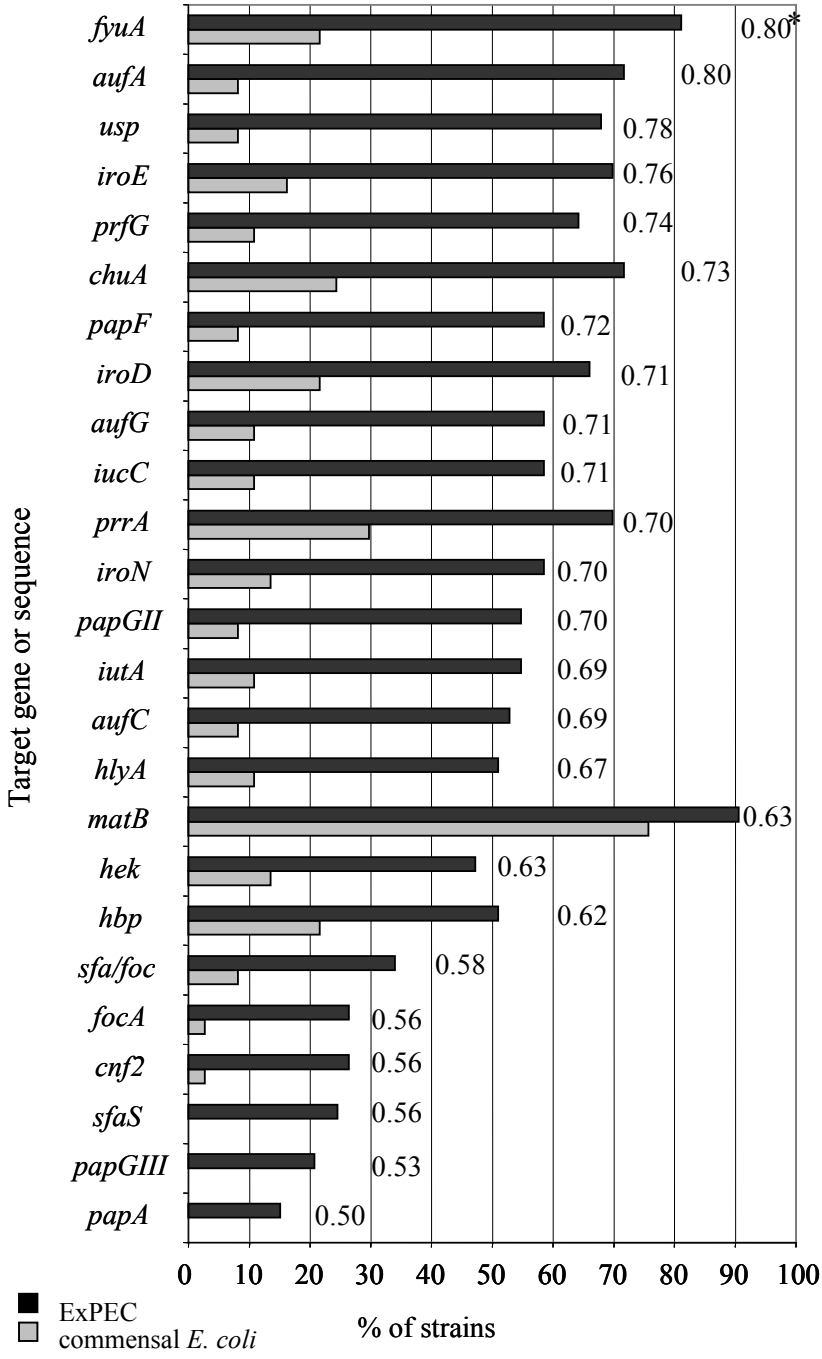

Figure 10.
Prevalence of the indicator genes among ExPEC and commensal *E. coli* isolates. 25 PCR probes homologous with the known ExPEC virulence factors were significantly associated ($P<0.05$) with the extraintestinal virulent phenotype. These sequences had classification rates (*) of 0.50 to 0.80 (1.0 = perfect indicator of the ExPEC phenotype, 0.0 = indifferent gene, not representative of the ExPEC phenotype) and thus were good indicators of the extraintestinal virulence.

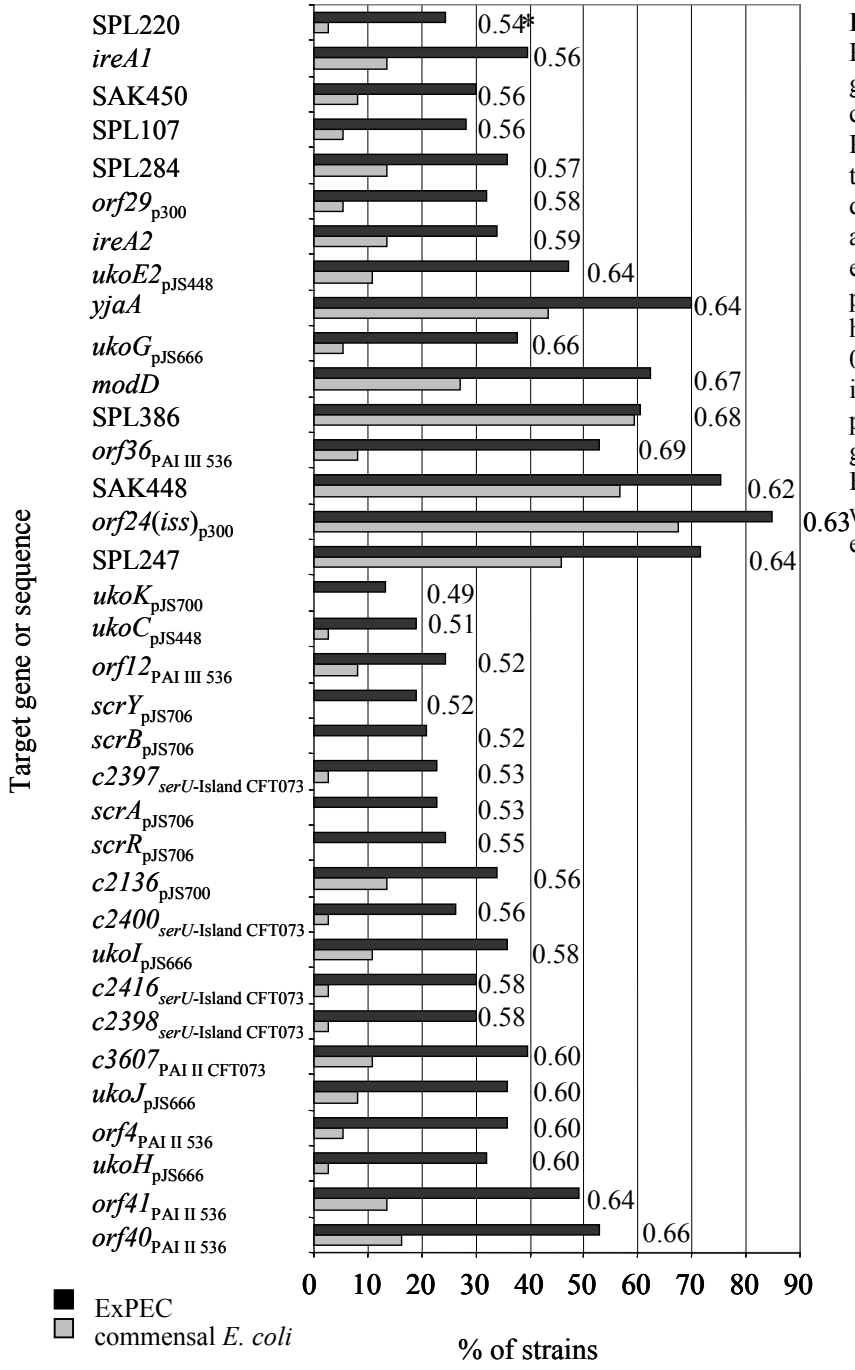

Figure 11.
Prevalence of the indicator genes among ExPEC and commensal *E. coli* isolates. 36 PCR probes homologous with the putative ExPEC virulence determinants were significantly associated ($P<0.05$) with the extraintestinal virulent phenotype. These sequences had classification rates (*) of 0.49 to 0.68 (1.0 = perfect indicator of the ExPEC phenotype, 0.0 = indifferent gene, not representative of the ExPEC phenotype) and thus were good indicators of the extraintestinal virulence.

4.4.2 Development and validation of the virulence gene algorithm for differentiation and classification of clinical ExPECs and fecal *E. coli* isolates according to their extraintestinal virulence potentials

In order to concretely describe the pathogenicity risk (i.e. the capacity to cause an infection) of an ExPEC isolate, the extraintestinal virulence potentials of ExPEC and commensal *E. coli* strains were calculated using a novel "virulence gene algorithm" (VGA). The VGA was based on the assumption that the virulence of ExPEC strains would be strongly dependent upon the available virulence gene

repertory and that the extraintestinal virulence potential of an *E. coli* strain could be estimated according to the assortment of these virulence genes (i.e. indicator genes; Figs. 10 and 11). The VGA was based on logistic regression analysis, which modeled the extraintestinal virulence by virtue of the co-occurrence of individual DNA probes (Hartigan and Wong, 1979).

Applicability of the VGA was validated by determining the correlation between clinical data, phylogenetic data, virulence gene repertoires, and the estimated "virulence potentials" (VPs) in the group of known and unknown clinical ExPECs and commensal *E. coli* isolates. 85% of ExPECs and 81% of commensal *E. coli* strains were accurately classified as pathogenic or non-pathogenic correspondingly to their estimated VPs (Table 14; Fig. 12). In contrast, eight UPEC strains and seven fecal *E. coli* isolates must be reclassified (Table 14; marked with an asterisk). Review of the clinical data and the virulence gene profiles of these strains warranted the estimated VPs. The UPEC strains with atypically low risk-ratios (VP<0.5) were largely isolated from patients with predestined medical history, such as diabetes or urinary-catheter, or from patients with advanced age. The ExPEC have shown to effectively colonize the bowel (Jantunen *et al.*, 2001; Nowrouzian *et al.*, 2005; Siitonen, 1992) and thus the fraction of rectal *E. coli* isolates with the VPs characteristic of ExPEC (VP>0.5) was likely to represent this diagnostically important fraction of resident, potentially pathogenic strains deposited in the gut. Although ExPEC strains are normal inhabitants of the intestinal flora, the fecal and vaginal colonization with ExPEC might predispose towards UTI especially in susceptible patients (Foxman and Brown, 2003). Furthermore, the results obtained from this study correlated with previous research projects and further emphasized the opportunistic nature of UTIs caused by UPEC (Gordon *et al.*, 2005).

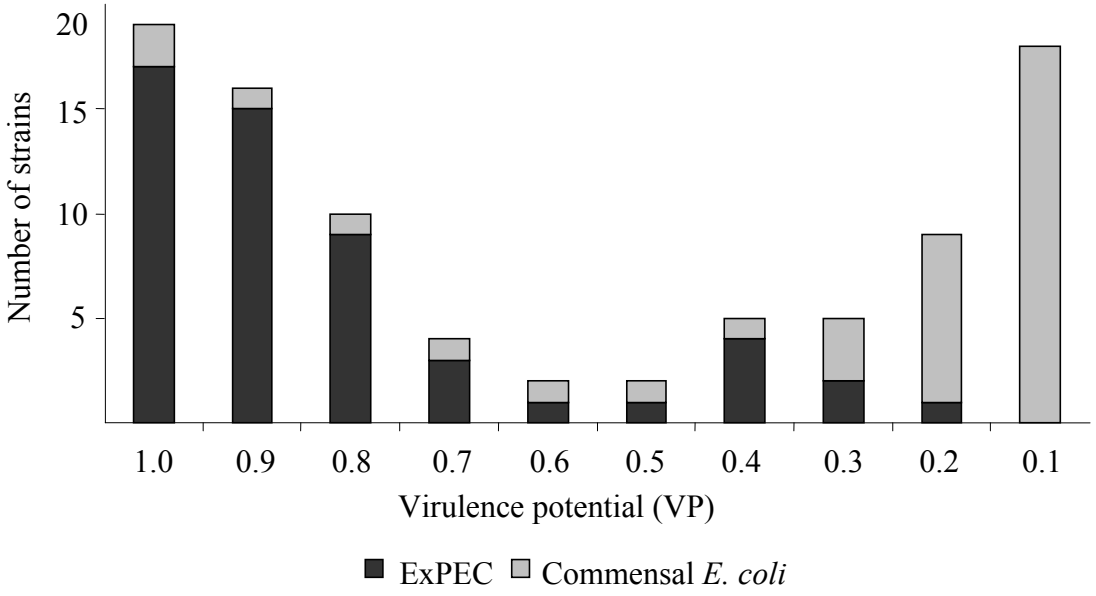

Figure 12. Estimation of the extraintestinal virulence potential of unknown ExPECs (n=53) and commensal *E. coli* (n=37) isolates using the VGA (VP<0.5 = low virulence, VP ≈ 0.5 = opportunistic virulence, VP>0.5 = high virulence). A large fraction of ExPEC strains (79%) displayed VPs highly characteristic of the extraintestinal virulent phenotype (0.7 to 1.0). In contrast, 6% of clinical ExPEC strains showed moderate VPs of 0.6 to 0.7 and 15% of the ExPEC strains were classified according to their VPs (0.2 to 0.5) as non-pathogenic. 81% of commensal *E. coli* isolates showed VPs less than 0.3. Two commensal *E. coli* strains had VPs of 0.5 to 0.6 and five fecal *E. coli* strains showed VPs highly indicative of the ExPEC phenotype (VPs of 0.7 to 1.0).

Table 14. Evaluation of the extraintestinal virulence potential of clinical ExPECs and fecal *E. coli* isolates using the VGA developed in this project.

E. coli strain	Clinical characteristic(s)	PG[a]	VP[b]
ExPEC			
536	Pyelonephritis	B2	0.92
AD110	Cystitis	B2	0.97
CFT073	Pyelonephritis	B2	0.97
CP9	Bacteriemia	B2	0.99
D9	Cystitis	A	0.77
Do764	Cystitis	B2	1.00
Do768	Cystitis	B2	1.00
ED2403	Cystitis	A	0.37*
HE300	Pyelonephritis	A	0.95
IHE1041	Pyelonephritis	B2	0.90
IHE1049	Pyelonephritis	B2	0.93
IHE1086	Pyelonephritis	B2	0.79
IHE11015	Pyelonephritis	B2	0.82
IHE11020	Pyelonephritis	B2	0.98
IHE11035	Pyelonephritis	A	0.48*
IHE11048	Pyelonephritis	B2	0.84
IHE11158	Pyelonephritis	B2	0.95
IHE11290	Pyelonephritis	B2	0.90
IHE1167	Pyelonephritis	B2	0.97
IHE1190	Pyelonephritis	B2	0.99
IHE1210	Pyelonephritis	B2	0.89
IHE1402	Pyelonephritis	B2	0.97
IHE1431	Pyelonephritis	B2	0.77
IHE3034	Newborn meningitis	B2	0.85
IHE3040	Newborn meningitis	B1	0.77
IHE3041	Newborn meningitis	B2	0.98
IHE3047	Newborn meningitis	B2	0.87
IHE3080	Newborn meningitis	B2	0.85
J96	Pyelonephritis	B2	0.95
JS293	Chronic cystitis, host suffers of *diabetes mellitus*	B2	0.44*
JS295	Cystitis	B1	0.35*
JS297	Cystitis, host suffers of *diabetes mellitus*	B2	0.85
JS299	Cystitis	B2	0.98
JS304	Pyelonephritis	B1	0.67
JS305	Pyelonephritis	B2	0.95
JS309	Cystitis	B2	0.98
JS322	Pyelonephritis	D	0.44*
JS323	Pyelonephritis	B2	0.24*
JS325	Pyelonephritis	B2	0.86
JS326	Pyelonephritis	B2	0.66
JS355	Pyelonephritis	B2	0.92
Nu14	Cystitis	B2	0.81
RS218	Newborn meningitis	B2	0.97

S2207	Catheter-associated cystitis	A	0.27*
S2265	Catheter-associated pyelonephritis	B2	0.97
S2287	Catheter-associated pyelonephritis	D	0.88
S2288	Catheter-associated pyelonephritis	B1	0.41*
S2413	Urosepsis	B2	1.00
S2523	Urosepsis	B2	0.83
S2536	Urosepsis	D	0.95
S2572	Urosepsis	D	0.87
S2608	Cystitis	A	0.74
S27710	Pyelonephritis	B2	0.62
Commensal *E. coli*			
BL21	Laboratory strain	A	0.05
DH5α	Laboratory strain	A	0.08
DSM6601	Stool	B2	0.98*
ECOR31	Stool	D	0.19
HB101	Laboratory strain	A	0.09
MG1655	Stool	A	0.08
S4000	Stool	D	0.09
S4005	Stool	B1	0.17
S4010	Stool	B2	0.63*
S4011	Stool	B2	0.21
S4012	Stool	A	0.27
S4013	Stool	D	0.35
S4014	Stool	D	0.34
S4015	Stool	B1	0.11
S4016	Stool	B1	0.06
S4017	Stool	B1	0.09
S4020	Stool	B2	0.93*
S4021	Stool	B2	0.98*
S4022	Stool	D	0.52*
S4023	Stool	A	0.27
S4026	Stool	A	0.13
S4039	Stool	B1	0.79*
S4050	Stool	B1	0.74*
S4134	Stool	A	0.12
S4136	Stool	A	0.18
S4138	Stool	A	0.09
S4140	Stool	A	0.12
S4141	Stool	A	0.07
S4147	Stool	B1	0.11
S4148	Stool	A	0.15
S4149	Stool	B2	0.24
S4151	Stool	B1	0.07
S4152	Stool	A	0.19
S4153	Stool	B1	0.11
S4158	Stool	B1	0.11
TOP10	Laboratory strain	A	0.07
XL-1 Blue MRF´	Laboratory strain	A	0.24

a Phylogenetic group (PG) was defined using the DNA probes described by Clermont *et al.*, 2000.

b The extraintestinal virulence potential was defined using the virulence gene algorithm. The strains with an atypical VP are indicated with an asterisk.

4.5 Future aspects: preventative screening of ExPECs in patient groups susceptible of life-threatening extraintestinal infections (Studies I to IV)

E. coli is the predominant facultative anaerobe of the human colonic flora and most strains are harmless. However, certain *E. coli* strains cause extraintestinal diseases in humans. These infections caused by ExPEC account for serious morbidity and mortality in the elderly, in young children, and in immune-compromised and hospitalized patients (Foxman and Brown, 2003). ExPEC strains are heterogeneous both with respect to symptoms developed by infected hosts and with respect to virulence genes (Foxman *et al.*, 2000; Gaynes and Edwards, 2005). Currently available diagnostic tools for extraintestinal infections have low sensitivity and specifity, and often provide no accurate information about the different serovars of ExPEC (Bingen-Bidois *et al.*, 2002; Clermont *et al.*, 2000; Clermont *et al.*, 2001; Morin and Hopkins, 2002). The present major focus in the research of extraintestinal infections is to discover gene loci that are conserved among ExPEC, and may provide genetic markers for a broad range of ExPEC.

In the current project, we identified 61 potential indicator genes of the extraintestinal virulence using literature search (Study IV), subtractive genomic hybridization (Study II), genetic characterization of plasmid- and chromosome-borne virulence genes (Studies I and III), DNA microarray analyzes (Studies I, III, and IV), and bioinformatics (Study IV). These candidate genes not only included known virulence determinants of ExPEC (n=25), but also a number of novel DNA sequences (n=36) with only a putative function in the ExPEC virulence. As none of these genetic markers was present in all ExPEC isolates studied here, these molecular probes *per se* could not be used in routine clinical laboratory testing for identification of different ExPEC serovars. Consequently, we assessed the strains' extraintestinal virulence potential globally using a novel virulence gene algorithm, which based on the co-occurrence of all analyzed sequence probes (Study IV). To our knowledge, this study described for the first time the application of the DNA microarray technology in combination with a quantification method for the rapid and reliable detection of ExPEC strains.

To validitate the method described here, we carried out molecular typing for a collection of 53 ExPECs and 37 fecal *E. coli* isolates. As the microarray combined with the VGA was able to detect and clearly identify the ExPEC isolates, this system offered an interesting potential for clinical microbiology laboratories. With the development of more aggressive therapeutic regimens, the incidence of multiresistant ExPEC strains has increased (Johansen *et al.*, 2006; Paterson, 2006). The early initiation of accurate antibiotic therapy is critical in reducing the high mortality rates of susceptible patients due to extraintestinal infections. However, this is only possible if the pathogen is identified early and accurately. The previously described molecular methods for ExPEC diagnosis feature only limited multiplexing capability, resulting in time delay and high costs if all relevant variants must be considered (Bingen-Bidois *et al.*, 2002; Clermont *et al.*, 2000; Clermont *et al.*, 2001; Morin and Hopkins, 2002). An economically more efficient approach would be the application of a protocol, which is capable of identifying a panel of relevant characteristics (e.g. virulence and antibiotic resistance genes) in a parallel fashion. For example, the standard methods to detect antibiotic resistances in pathogenic *E. coli* isolates take more than one day. Using the microarray technology in a combination with a highly sensitive and selective PCR system, the assay time could be reduced for a few hours. Simultaneously the sensitivity and the specificity would be increased, which are the most important factors for a clinical application (Bruant *et al.*, 2006 Cassone *et al.*, 2006). Furthermore, identification of pathogenic or multiresistant *E. coli* strains in the mixed culture samples (e.g. in the urine and the stool samples, or in the biofilms on the catheter surfaces) could be possible (Kostic *et al.*, 2007; Maynard *et al.*, 2005). Besides the clinical diagnostics of acute infections, the analysis on the

molecular level could be used for the long term monitoring of patients with indwelling catheters or during an antibiotic treatment. Here, we described such an ExPEC identification array by making use of the high multiplexing capacity of DNA microarrays. The fact that highly virulent ExPEC isolates, which carry a perilous configuration of extraintestinal virulence factors, could be effectively differentiated from less pathogenic clinical ExPECs and from commensal *E. coli* strains opened up new avenues for the acute and the preventive diagnostics. By manifolding the scarce target sequences in a clinical sample using an elaborate PCR amplification (e.g. random extension of genomic DNA sequences followed by the target-directed amplification) prior to microarray analysis, it might be possible to circumvent laborious and time-consuming bacterial cultivation and DNA isolation procedures, and thus provide a system suitable for routine use.

In summary, DNA microarrays, such as the ExPEC pathoarray described here, present a promising platform for the parallel identification virulence factors and the detection of genes that confer antibiotic resistance. However, to our knowledge, there are no such commercial arrays available so far, and thus this topic provides an interesting issue for future applications for optimization of diagnostics of microbial infections.

5. CONCLUSIONS AND FUTURE PERSPECTIVES

This project was initiated to characterize the genomic variation of ExPEC; a diverse group of *E. coli* strains infecting extraintestinal sites, such as the bloodstream, the meninges, and the urinary tract. During this project, genomic sequences of three *E. coli* strains (i.e. the UPEC strain CFT073, the commensal *E. coli* strain MG1655, and the EHEC strain EDL933) and two closely related *S. flexneri* strains became available (Blattner *et al.*, 1997; Jin *et al.*, 2002; Perna *et al.*, 2001; Welch *et al.*, 2002; Wei *et al.*, 2003). The accumulated sequence data as well as data obtained from further research projects have aided to compile a portrait of the ExPEC pathogroup. Accordingly, the ExPEC genomes are extremely dynamic and differ considerably with regard to their sizes, the gene contents, and the number of known virulence factors (Dobrindt and Hacker, 2001; Johnson and Russo, 2005 Oelschlaeger *et al.*, 2002). Therefore, in spite of best efforts, no research team has managed to decipher the genes or factors, which are uniformly required by ExPEC or by some ExPEC subgroup to initiate or establish an infection. These strains have obtained a significant proportion of their genetic diversity through acquisition of DNA from related or non-related organisms. Different mobile genetic elements, such as phages, transposons, plasmids, and PAIs are involved in this horizontal gene transfer (Hacker and Carniel, 2001; Ochman *et al.*, 2005). However, it is known neither how these very dynamic genomes have been assembled or controlled, nor which mechanisms are responsible for the observed genomic mobility.

Our results corroborate the previous genomic studies, which document high genomic variability among the ExPEC pathogroup and strongly emphasize the impact of lateral gene transfer as the driving force in shaping the ExPEC genomes (Dobrindt and Hacker, 2001; Hacker and Carniel, 2001). We showed that ExPEC strains are different, not only from commensal *E. coli* strains belonging to the normal intestinal flora, but from also each other. This extensive variation was not restricted to distinct ExPEC subgroups (e.g. NMEC or UPEC) as also UPEC strains studied differed significantly from each other with respect to their genomic contents and the virulence gene repertoires. A large fraction of the UPEC genomes (up to 25%) seems to be composed of unique DNA without any known sequence counterpart or function. Given that the amount of transferred strain-specific DNA in ExPEC can only be estimated, it is likely that yet unknown virulence or fitness factors are encoded by the unique DNA fraction. The fact that not many virulence factors are common in UPEC strains was unexpected since all UPEC strains cause the same disease. Our findings imply that a number of alternative virulence factors can mediate each step in the infection process and that each strain may have a unique combination of such factors. Detection of novel ExPEC-associated DNA sequences in UPEC, shared by both intestinal pathogenic and extraintestinal pathogenic *E. coli*, further emphasized the importance of understanding the mechanisms and the environmental signals UPEC use for recognition of environmental changes and different ecological niches (e.g. the gut and the urinary tract).

An assortment of virulence genes is apparently made possible by a variety of genetic factors, which contribute to the genomic plasticity. The novel ExPEC-associated DNA sequences characterized in this project revealed composite DNA structures of unknown genes and mobile genetic elements of different origins, such as pathogenicity islands, insertion sequences, transposable elements, phages, and plasmids. The presence of such mobile elements adjacent to virulence gene clusters has previously been described in another contexts and it has been suggested that these sequences mediate the mobilization of the respective chromosomal gene clusters (Dobrindt *et al.*, 2002; Neilands, 1992; Purdy and Payne, 2001). This project provided circumstantial evidence that IS elements indeed facilitate site-specific recombination events and in this manner enable the transfer of single genes or gene clusters from their original locations to new sites. As follows, we assumed that

the integration of IS elements in the sites adjacent to a chromosomal *iroBCDEN* gene cluster led to its mobilization into *S. enterica* plasmid, an event which resulted the novel urovirulence plasmid p300. However, the sole presence of mobile elements did not explain the mosaic DNA structures characteristic of the novel DNA sequences originating from the UPEC strain JS299. These sequences displayed short, compositionally different DNA fragments, which showed individual patterns of distribution among ExPEC strains. It appeared likely that also homologous recombination, even within coding genes, contributed to the observed mosaicism in ExPEC, and that besides *en block* transfer of large DNA regions (e.g. *iroBCDEN* cluster of plasmid p300) rearrangements of small DNA modules might also provide a means of genomic plasticity. Interestingly, sequences highly homologous with the genomic islands of the archetypal UPEC strains showed that even parts of such supposedly inert virulence gene clusters could be mobilized individually. An additional and important conclusion from this data was the indication that the conjugative transfer of plasmids provides indeed means for effective lateral gene transfer in ExPEC strains. The accumulation of virulence factors and antibiotic resistance genes on the same horizontally transferable vehicle may lead to the simultaneous transfer of multiple virulence and antibiotic resistance determinants to further bacterial strains, species, or genera.

Many of the sequences examined in this project were of genes that had an unknown function or no known homologues in the public databases, but which were significantly associated with the extraintestinal virulent phenotype. Although we did not attribute functions to these novel genes in the present project, they do represent novel markers for urovirulence, which are well suited for the subclassification and subtyping ExPECs and other *E. coli* isolates. Subsequent studies will have to define the function of these novel determinants on the ability of ExPEC strains to colonize and infect the urinary tract in order to decipher the puzzle of ExPEC virulence.

In the last part the project, we demonstrated that highly virulent ExPEC isolates, which carry a perilous configuration of extraintestinal virulence factors, could be effectively differentiated from less pathogenic clinical ExPECs and from commensal *E. coli* strains. This was achieved using a novel ExPEC pathoarray combined with a virulence gene algorithm, which described the potential of a certain gene combination to present the extraintestinal virulent phenotype. The preventive diagnostics of extraintestinal infections is of great interest especially in immunocompromised subjects susceptible to severe life-threatening infections (e.g. diabetic or elderly patients, neonates, or pregnant women) or in patients treated at the intensive care units with invasive interventions (e.g. the urinary catheter). The exact knowledge of the genotype (e.g. antibiotic resistence) of a clinical sample, would allow purposive prevention in a sufficient time. We show that diagnostic microarrays provide better and more rational use of available resources of the public health care by focusing on target-directed prevention (i.e. targeted treatment according to the known genotype of an infecting strain), instead of massive therapeutic interventions of these life-threatening infections.

In summary, this project has attempted to make clear that genomics (i.e. DNA sequencing of partial and entire *E. coli* genomes) and rapid screening methods, such as the DNA microarray analyses, have deepened our appreciation of genetic diversity and horizontal gene transfer as central features of the ExPEC pathogroup. In addition, we argue that unraveling the molecular mechanisms of gene transfer and its regulation provides a unifying theme for investigating the host-pathogen interactions in a wide range of extraintestinal infectious diseases.

6. ACKNOWLDEDGEMENTS

This project was accomplished at the Max von Pettenkofer-Institut, Ludwig Maximilians University, Munich, with the educational support from the Division of General Microbiology, Department of Biological and Environmental Sciences, University of Helsinki, during the years 1999 to 2006. The project was financed by a grant from the Bundesministerium für Bildung und Forschung (Kompetenznetzwerk PathoGenoMik), by a grant from the Deutsche Forschungsgemeinschaft (SCHU1494/2-1 to S. Schubert), and by a grant for finishing doctoral dissertations at the University of Helsinki.

I warmly thank Professor Kielo Haahtela and Professor Timo Korhonen at the Division of General Microbiology for the support and guidance during my undergraduate studies. Furthermore, I want to thank Professor Jürgen Heesemann at the Max von Pettenkofer-Institut for the opportunity to complete my project and for providing excellent working and educational facilities.

To my supervisor, Dr. Sören Schubert at the Max von Pettenkofer-Institut, I want to express my gratitude for his altruistic efforts to educate me. Moreover, I want to thank him for the indispensable help in the process of writing the publications referred to and, finally yet importantly, this project. Without his visions, knowledge, enthusiasm, and willingness to explain the unexplainable this project would not have been realized.

I sincerely thank Dr. Emeka Igwe at the Department of Pathobiochemistry, Zentrum der Biologischen Chemie, Frankfurt, for editing this dissertation. Without his constructive and perspicacious criticism these thesis would not have been accomplished in this form.

I want also warmly thank Professor Anja Siitonen at the Department of Bacterial and Inflammatory Diseases of the Finnish National Public Health Institute and Dr. Benita Westerlund-Wikström at the Division of General Microbiology for their competent advice during the revision of this project.

I sincerely thank Dr. Ulrich Dobrindt at the Institut für Molekulare Infektionsbiologie, University of Würzburg, for accepting the invitation for being the opponent.

I sincerely thank Dr. Harald Hoffman, Dr. Rainhard Hoffman, Dr. Emeka Igwe, Dr. Thomas Müller, Dr. Klaus Panthel, Dr. Holger Rüssmann, Christina Nölting, and Dr. Gottfried Wilharm as well as my Russian friends Dr. Alexander Rakin, Vladimir Antonenka, Roman Ansimow, and Andrey Golubov for their analytical and technical expertise and for their valuable assistance with the small and sometimes trickier problems.

I warmly thank Dr. Kaarina Lähteenmäki and Dr. Ritva Virkola at the Division of General Microbiology for excellent educational guidance during my early academic career.

I sincerely thank my biology teachers Leena and Veikko Makkonen in Kitee for patient quidance to the fashinating world of living organisms.

I want to express my great affection to my dear colleagues and trusted friends Christine, Chrissie, Emeka, Friederike, Harald, Jeanette, Kristin, Kirsten H. and Kirsten W., Klaus, Orsi, and Severin. I want to thank them for sharing the good and bad days and for encouragement and inspiration, in both the scientific work and the private life. You made it worth the effort.

I thank my family and friends for providing the right perspective. Moreover, I thank them for invaluable support, patience, and love they gave me at the difficult moments. Without you all I would only be a small fraction of myself and I am deeply sorry not to tell it often enough.

7. REFERENCES

Achtman, M., A.Mercer, B.Kusecek, A.Pohl, M.Heuzenroeder, W.Aaronson, A.Sutton, and R.P.Silver. 1983. Six widespread bacterial clones among *Escherichia coli* K1 isolates. Infect. Immun. 39:315-335.

Alexeyev, M.F. and I.N.Shokolenko. 1995. Mini-Tn10 transposon derivatives for insertion mutagenesis and gene delivery into the chromosome of gram-negative bacteria. Gene 160:59-62.

Anderson, G.G., K.W.Dodson, T.M.Hooton, and S.J.Hultgren. 2004. Intracellular bacterial communities of uropathogenic *Escherichia coli* in urinary tract pathogenesis. Trends Microbiol. 12:424-430.

Anderson, K.L., J.Billington, D.Pettigrew, E.Cota, P.Simpson, P.Roversi, H.A.Chen, P.Urvil, M.L.du, P.N.Barlow, M.E.Medof, R.A.Smith, B.Nowicki, C.Le Bouguenec, S.M.Lea, and S.Matthews. 2004. An atomic resolution model for assembly, architecture, and function of the Dr adhesins. Mol. Cell 15:647-657.

Andrews, S.C., A.K.Robinson, and F.Rodriguez-Quinones. 2003. Bacterial iron homeostasis. FEMS Microbiol. Rev. 27:215-237.

Anisimov, R., D.Brem, J.Heesemann, and A.Rakin. 2005. Transcriptional regulation of high pathogenicity island iron uptake genes by YbtA. Int. J. Med. Microbiol. 295:19-28.

Archambaud, M., P.Courcoux, and A.Labigne-Roussel. 1988. Detection by molecular hybridization of pap, afa, and sfa adherence systems in *Escherichia coli* strains associated with urinary and enteral infections. Ann Inst. Pasteur Microbiol 139:575-588.

Austin, P.F., A.J.Casale, M.P.Cain, R.C.Rink, and S.J.Weintraub. 2003. Lipopolysaccharide and inflammatory cytokines cause an inducible nitric oxide synthase-dependent bladder smooth muscle fibrotic response. J. Urol. 170:645-648.

Ausubel, F.M., R.Brent, R.E.Kingston, D.D.Moore, J.G.Seidman, J.A.Smith, and K.Struhl. 1989. Current protocols in molecular biology. John Wiley&Sons, New York.

Azpiroz, M.F., E.Rodriguez, and M.Lavina. 2001. The structure, function, and origin of the microcin H47 ATP-binding cassette exporter indicate its relatedness to that of colicin V. Antimicrob. Agents Chemother. 45:969-972.

Backhed, F., S.Normark, E.K.Schweda, S.Oscarson, and A.Richter-Dahlfors. 2003. Structural requirements for TLR4-mediated LPS signalling: a biological role for LPS modifications. Microbes. Infect. 5:1057-1063.

Bagshaw, S.M. and K.B.Laupland. 2006. Epidemiology of intensive care unit-acquired urinary tract infections. Curr. Opin. Infect Dis 19:67-71.

Bahrani-Mougeot, F.K., E.L.Buckles, C.V.Lockatell, J.R.Hebel, D.E.Johnson, C.M.Tang, and M.S.Donnenberg. 2002. Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia col*i virulence determinants in the murine urinary tract. Mol. Microbiol. 45:1079-1093.

Bahrani-Mougeot, F.K., S.Pancholi, M.Daoust, and M.S.Donnenberg. 2001. Identification of putative urovirulence genes by subtractive cloning. J. Infect. Dis. 183 Suppl 1:S21-S23.

Bachur R. and G.L. Caputo. 1995. Bacteremia and meningitis among infants with urinary tract infections. Pediatr Emerg Care. 11:280-284.

Balazsi, G., A.L.Barabasi, and Z.N.Oltvai. 2005. Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A 102:7841-7846.

Baumler, A.J., T.L.Norris, T.Lasco, W.Voight, R.Reissbrodt, W.Rabsch, and F.Heffron. 1998b. IroN, a novel outer membrane siderophore receptor characteristic of *Salmonella enterica*. J. Bacteriol. 180:1446-1453.

Bekal, S., R.Brousseau, L.Masson, G.Prefontaine, J.Fairbrother, and J.Harel. 2003c. Rapid identification of *Escherichia coli* pathotypes by virulence gene detection with DNA microarrays. J. Clin. Microbiol. 41:2113-2125.

Bellini, E.M., W.P.Elias, T.A.Gomes, T.L.Tanaka, C.R.Taddei, R.Huerta, F.Navarro-Garcia, and M.B.Martinez. 2005. Antibody response against plasmid-encoded toxin (Pet) and the protein involved in intestinal colonization (Pic) in children with diarrhea produced by enteroaggregative *Escherichia coli*. FEMS Immunol. Med. Microbiol. 43:259-264.

Beloin, C., J.Valle, P.Latour-Lambert, P.Faure, M.Kzreminski, D.Balestrino, J.A.Haagensen, S.Molin, G.Prensier, B.Arbeille, and J.M.Ghigo. 2004. Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. Mol. Microbiol. 51:659-674.

Berger, C.N., O.Billker, T.F.Meyer, A.L.Servin, and I.Kansau. 2004. Differential recognition of members of the carcinoembryonic antigen family by Afa/Dr adhesins of diffusely adhering *Escherichia coli* (Afa/Dr DAEC). Mol. Microbiol. 52:963-983.

Bergsten G., M. Samuelsson, B. Wullt, I. Leijonhufvud, H. Fischer, and C. Svanborg. 2004. PapG-dependent adherence breaks mucosal inertia and triggers the innate host response. J Infect Dis. 189:1734-1742.

Bergthorsson, U., and H. Ochman. 1998. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol Biol Evol. 15:6-16.

Bernier-Febreau, C., M.L.du, E.Turlin, V.Labas, J.Ordonez, A.M.Gilles, and B.C.Le. 2004. Use of deoxyribose by intestinal and extraintestinal pathogenic *Escherichia coli* strains: a metabolic adaptation involved in competitiveness. Infect. Immun. 72:6151-6156.

Best, A., R.M.La Ragione, A.R.Sayers, and M.J.Woodward. 2005. Role for flagella but not intimin in the persistent infection of the gastrointestinal tissues of specific-pathogen-free chicks by shiga toxin-negative *Escherichia coli* O157:H7. Infect. Immun. 73:1836-1846.

Bidet, P., S.Bonacorsi, O.Clermont, C.De Montille, N.Brahimi, and E.Bingen. 2005b. Multiple insertional events, restricted by the genetic background, have led to acquisition of pathogenicity island IIJ96-like domains among *Escherichia coli* strains of different clinical origins. Infect. Immun. 73:4081-4087.

Bielaszewska, M., B.Sinha, T.Kuczius, and H.Karch. 2005. Cytolethal distending toxin from Shiga toxin-producing *Escherichia coli* O157 causes irreversible G2/M arrest, inhibition of proliferation, and death of human endothelial cells. Infect. Immun. 73:552-562.

Bilban, M., L.K.Buehler, S.Head, G.Desoye, and V.Quaranta. 2002. Normalizing DNA microarray data. Curr. Issues Mol. Biol. 4:57-64.

Bingen-Bidois, M., O.Clermont, S.Bonacorsi, M.Terki, N.Brahimi, C.Loukil, D.Barraud, and E.Bingen. 2002b. Phylogenetic analysis and prevalence of urosepsis strains of Escherichia coli bearing pathogenicity island-like domains. Infect. Immun. 70:3216-3226.

Bister, B., D.Bischoff, G.J.Nicholson, M.Valdebenito, K.Schneider, G.Winkelmann, K.Hantke, and R.D.Sussmuth. 2004. The structure of salmochelins: C-glucosylated enterobactins of *Salmonella enterica*. Biometals 17:471-481.

Blattner, F.R., G.Plunkett, III, C.A.Bloch, N.T.Perna, V.Burland, M.Riley, J.Collado-Vides, J.D.Glasner, C.K.Rode, G.F.Mayhew, J.Gregor, N.W.Davis, H.A.Kirkpatrick, M.A.Goeden, D.J.Rose, B.Mau, and Y.Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453-1474.

Bliss, T.W., J.E.Dohms, M.G.Emara, and C.L.Keeler, Jr. 2005. Gene expression profiling of avian macrophage activation. Vet. Immunol. Immunopathol. 105:289-299.

Blomgran, R., L.Zheng, and O.Stendahl. 2004. Uropathogenic *Escherichia coli* triggers oxygen-dependent apoptosis in human neutrophils through the cooperative effect of type 1 fimbriae and lipopolysaccharide. Infect. Immun. 72:4570-4578.

Boddicker, J.D., N.A.Ledeboer, J.Jagnow, B.D.Jones, and S.Clegg. 2002. Differential binding to and biofilm formation on, HEp-2 cells by *Salmonella enterica* serovar Typhimurium is dependent upon allelic variation in the *fimH* gene of the fim gene cluster. Mol. Microbiol. 45:1255-1265.

Bouckaert, J., J.Berglund, M.Schembri, E.De Genst, L.Cools, M.Wuhrer, C.S.Hung, J.Pinkner, R.Slattegard, A.Zavialov, D.Choudhury, S.Langermann, S.J.Hultgren, L.Wyns, P.Klemm, S.Oscarson, S.D.Knight, and H.De Greve. 2005. Receptor binding studies disclose a novel class of high-affinity inhibitors of the *Escherichia coli* FimH adhesin. Mol. Microbiol. 55:441-455.

Bouckaert, J., J.Mackenzie, J.L. de Paz, B. Chipwaza, D. Choudhury, A. Zavialov, K. Mannerstedt, J. Anderson, D. Pierard, L. Wyns, P.H. Seeberger, S. Oscarson, H. De Greve, and S.D. Knight. 2006. The affinity of the FimH fimbrial adhesin is receptor-driven and quasi-independent of *Escherichia coli* pathotypes. Mol Microbiol. 61:1556-1568.

Bouguenec, C., and A.L.Servin. 2006. Diffusely adherent *Escherichia coli* strains expressing Afa/Dr adhesins (Afa/Dr DAEC): hitherto unrecognized pathogens. FEMS Microbiol Lett. 256:185-194.

Bower, J.M., D.S.Eto, and M.A.Mulvey. 2005. Covert operations of uropathogenic *Escherichia coli* within the urinary tract. Traffic. 6:18-31.

Bras, M., V.Dugas, F.Bessueille, J.P.Cloarec, J.R.Martin, M.Cabrera, J.P.Chauvet, E.Souteyrand, and M.Garrigues. 2004. Optimisation of a silicon/silicon dioxide substrate for a fluorescence DNA microarray. Biosens. Bioelectron. 20:797-806.

Braun, V. 2003. Iron uptake by *Escherichia coli*. Front Biosci. 8:s1409-s1421.

Bryant, P.A., D.Venter, R.Robins-Browne, and N.Curtis. 2004. Chips with everything: DNA microarrays in infectious diseases. Lancet Infect. Dis. 4:100-111.

Bruant, G., Maynard C., Bekal S., Gaucher I., Masson L., Brousseau R., and J. Harel. 2006. Development and validation of an oligonucleotide microarray for detection of multiple virulence and antimicrobial resistance genes in *Escherichia coli*. Appl Environ Microbiol. 72:3780-3784.

Buckles, E.L., F.K.Bahrani-Mougeot, A.Molina, C.V.Lockatell, D.E.Johnson, C.B.Drachenberg, V.Burland, F.R.Blattner, and M.S.Donnenberg. 2004. Identification and characterization of a novel uropathogenic *Escherichia coli*-associated fimbrial gene cluster. Infect. Immun. 72:3890-3901.

Bullen, J.J., H.J.Rogers, P.B.Spalding, and C.G.Ward. 2005a. Iron and infection: the heart of the matter. FEMS Immunol. Med. Microbiol. 43:325-330.

Caprioli, A., S.Morabito, H.Brugereb, and E.Oswald. 2005. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. Vet. Res 36:289-311.

Carniel, E., I.Guilvout, and M.Prentice. 1996. Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. J. Bacteriol. 178:6743-6751.

Carroll, I.M., A.A.Khan, and N.Ahmed. 2004. Revisiting the pestilence of *Helicobacter pylori*: insights into geographical genomics and pathogen evolution. Infect. Genet. Evol. 4:81-90.

Cassone M, D'Andrea MM, Iannelli F, Oggioni MR, Rossolini GM, and G.Pozzi. 2006. DNA microarray for detection of macrolide resistance genes. Antimicrob Agents Chemother. 50:2038-2041.

Chang, D.E., D.J.Smalley, D.L.Tucker, M.P.Leatham, W.E.Norris, S.J.Stevenson, A.B.Anderson, J.E.Grissom, D.C.Laux, P.S.Cohen, and T.Conway. 2004. Carbon nutrition of *Escherichia coli* in the mouse intestine. Proc. Natl. Acad. Sci. U. S. A 101:7427-7432.

Chen S.L., C.S. Hung, J. Xu, C.S. Reigstad, V. Magrini, A. Sabo, D. Blasiar, T. Bieri, R.R. Meyer, P. Ozersky, J.R. Armstrong, R.S. Fulton, J.P. Latreille, J. Spieth, T.M. Hooton, E.R. Mardis, S.J. Hultgren, and J.I. Gordon.. 2006. Identification of genes subject to a positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proc Natl Acad Sci U S A. 103:5977-5982.

Chen, Y.M., P.J.Wright, C.S.Lee, and G.F.Browning. 2003. Uropathogenic virulence factors in isolates of *Escherichia coli* from clinical cases of canine pyometra and feces of healthy bitches. Vet. Microbiol. 94:57-69.

Cheng, J.Y., H.H.Chen, Y.S.Kao, W.C.Kao, and K.Peck. 2002. High throughput parallel synthesis of oligonucleotides with 1536 channel synthesizer. Nucleic Acids Res 30:e93.

Chou, C.C., C.H.Chen, T.T.Lee, and K.Peck. 2004. Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. Nucleic Acids Res. 32:e99.

Choudhuri, S. 2004. Microarrays in biology and medicine. J. Biochem. Mol. Toxicol. 18:171-179.

Chouikha, I., P.Germon, A.Bree, P.Gilot, M.Moulin-Schouleur, and C.Schouler. 2006. A selC-associated genomic island of the extraintestinal avian pathogenic *Escherichia coli* strain BEN2908 is involved in carbohydrate uptake and virulence. J Bacteriol. 188:977-987.

Chowdhury, P., S.H.Sacks, and N.S.Sheerin. 2004. Minireview: functions of the renal tract epithelium in coordinating the innate immune response to infection. Kidney Int. 66:1334-1344.

Chung, J.W., S.J.Hong, K.J.Kim, D.Goti, M.F.Stins, S.Shin, V.L.Dawson, T.M.Dawson, and K.S.Kim. 2003. 37-kDa laminin receptor precursor modulates cytotoxic necrotizing factor 1-mediated RhoA activation and bacterial uptake. J. Biol. Chem. 278:16857-16862.

Clarke, S.C., R.D.Haigh, P.P.Freestone, and P.H.Williams. 2003. Virulence of enteropathogenic *Escherichia coli*, a global pathogen. Clin. Microbiol Rev. 16:365-378.

Clermont, O., S.Bonacorsi, and E.Bingen. 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl. Environ. Microbiol. 66:4555-4558.

Clermont, O., C.Cordevant, S.Bonacorsi, A.Marecat, M.Lange, and E.Bingen. 2001. Automated ribotyping provides rapid phylogenetic subgroup affiliation of clinical extraintestinal pathogenic *Escherichia coli* strains. J. Clin. Microbiol. 39:4549-4553.

Cleven, B.E., M.Palka-Santini, J.Gielen, S.Meembor, M.Kronke, and O.Krut. 2006. Identification and characterization of bacterial pathogens causing bloodstream infections by DNA microarray. J Clin Microbiol 44:2389-2397.

Connell, H., L.K.Poulsen, and P.Klemm. 2000. Expression of type 1 and P fimbriae *in situ* and localisation of a uropathogenic *Escherichia coli* strain in the murine bladder and kidney. Int. J. Med. Microbiol. 290:587-597.

Cooke, E.M. and S.P.Ewins. 1975. Properties of strains of *Escherichia coli* isolated from a variety of sources. J. Med. Microbiol. 8:107-111.

Cowan, P.J., H.Nagesha, L.Leonard, J.L.Howard, and A.J.Pittard. 1991. Characterization of the major promoter for the plasmid-encoded sucrose genes *scrY*, *scrA*, and *scrB*. J. Bacteriol. 173:7464-7470.

Cross, A., L.Asher, M.Seguin, L.Yuan, N.Kelly, C.Hammack, J.Sadoff, and P.Gemski, Jr. 1995. The importance of a lipopolysaccharide-initiated, cytokine-mediated host defense mechanism in mice against extraintestinally invasive *Escherichia coli*. J. Clin. Invest 96:676-686.

Cross, M.A., V.Koronakis, P.L.Stanley, and C.Hughes. 1990. HlyB-dependent secretion of hemolysin by uropathogenic *Escherichia coli* requires conserved sequences flanking the chromosomal hly determinant. J. Bacteriol. 172:1217-1224.

Daubin, V., N.A.Moran, and H.Ochman. 2003. Phylogenetics and the cohesion of bacterial genomes. Science 301:829-832.

Demir, M. and I.Kaleli. 2004. Production by *Escherichia coli* isolates of siderophore and other virulence factors and their pathogenic role in a cutaneous infection model. Clin. Microbiol. Infect. 10:1011-1014.

Deo, S.S. and A.K.Vaidya. 2004. Elevated levels of secretory immunoglobulin A (sIgA) in urinary tract infections. Indian J. Pediatr. 71:37-40.

Detmers, P.A., S.K.Lo, E.Olsen-Egbert, A.Walz, M.Baggiolini, and Z.A.Cohn. 1990. Neutrophil-activating protein 1/interleukin 8 stimulates the binding activity of the leukocyte adhesion receptor CD11b/CD18 on human neutrophils. J. Exp. Med. 171:1155-1162.

Dharmadi, Y. and R.Gonzalez. 2004. DNA microarrays: experimental issues, data analysis, and application to bacterial systems. Biotechnol. Prog. 20:1309-1324.

Diatchenko, L., Y.F.Lau, A.P.Campbell, A.Chenchik, F.Moqadam, B.Huang, S.Lukyanov, K.Lukyanov, N.Gurskaya, E.D.Sverdlov, and P.D.Siebert. 1996b. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. U. S. A 93:6025-6030.

Dobrindt, U., F.Agerer, K.Michaelis, A.Janka, C.Buchrieser, M.Samuelson, C.Svanborg, G.Gottschalk, H.Karch, and J.Hacker. 2003. Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. J. Bacteriol. 185:1831-1840.

Dobrindt, U., G.Blum-Oehler, G.Nagy, G.Schneider, A.Johann, G.Gottschalk, and J.Hacker. 2002. Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic *Escherichia coli* strain 536. Infect. Immun. 70:6365-6372.

Dobrindt, U. and J.Hacker. 2001. Whole genome plasticity in pathogenic bacteria. Curr. Opin. Microbiol. 4:550-557.

Dobrindt, U., B.Hochhut, U.Hentschel, and J.Hacker. 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. 2:414-424.

Dodson, K.W., J.S.Pinkner, T.Rose, G.Magnusson, S.J.Hultgren, and G.Waksman. 2001. Structural basis of the interaction of the pyelonephritic *E. coli* adhesin to its human kidney receptor. Cell 105:733-743.

dos Reis, R.M., L.Wernisch, and R.Savva. 2003. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. Nucleic Acids Res. 31:6976-6985.

Dou, W., S.Thompson-Jaeger, S.J.Lauledeerkind, J.W.Becker, J.Montgomery, E.Ruiz-Bustos, D.L.Hasty, L.R.Ballou, P.S.Eastman, B.Srichai, M.D.Breyer, and R.Raghow. 2005. Defective Expression of Tamm-Horsfall Protein/Uromodulin in Cyclooxygenase-2 Deficient Mice Increases Their Susceptibility to Urinary Tract Infections. Am. J. Physiol Renal Physiol. 289:49-60.

Doye, A., A.Mettouchi, G.Bossis, R.Clement, C.Buisson-Touati, G.Flatau, L.Gagnoux, M.Piechaczyk, P.Boquet, and E.Lemichez. 2002. CNF1 exploits the ubiquitin-proteasome machinery to restrict Rho GTPase activation for bacterial host cell invasion. Cell 111:553-564.

Dozois, C.M., F.Daigle, and R.Curtiss, III. 2003. Identification of pathogen-specific and conserved genes expressed *in vivo* by an avian pathogenic *Escherichia coli* strain. Proc. Natl. Acad. Sci. U. S. A 100:247-252.

Dudgeon,L.S.E.W.a.F.B. 1921. On *Bacillus coli* infections of the urinary tract especially in relation to haemolytic organisms. Journal of Hygiene 20:137-164.

Duguid, J.P., I.W.Smith, G.Dempster, and P.N.Edmund. 1955. Non-flagellar filamentous appendages (fimbriae) and haemagglutinating activity in *Bacterium coli*. J. Pathol. Bacteriol. 70:335-348.

Durand, V., J.O.Pers, Y.Renaudineau, A.Saraux, P.Youinou, and C.Jamin. 2001. Differential effects of anti-Fc gamma RIIIb autoantibodies on polymorphonuclear neutrophil apoptosis and function. J. Leukoc. Biol. 69:233-240.

Duriez, P., O.Clermont, S.Bonacorsi, E.Bingen, A.Chaventre, J.Elion, B.Picard, and E.Denamur. 2001. Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations. Microbiology 147:1671-1676.

Eden, C.S., L.A.Hanson, U.Jodal, U.Lindberg, and A.S.Akerlund. 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.

Eden, C.S. and H.A.Hansson. 1978. *Escherichia coli* pili as possible mediators of attachment to human urinary tract epithelial cells. Infect. Immun. 21:229-237.

Escobar-Paramo, P., O.Clermont, A.B.Blanc-Potard, H.Bui, C.Le Bouguenec, and E.Denamur. 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Mol. Biol. Evol. 21:1085-1094.

Evans, G.A., K.Lewis, and B.E.Rothenberg. 1989. High efficiency vectors for cosmid microcloning and genomic analysis. Gene 79:9-20.

Falzano, L., M.G.Quaranta, S.Travaglione, P.Filippini, A.Fabbri, M.Viora, G.Donelli, and C.Fiorentini. 2003. Cytotoxic necrotizing factor 1 enhances reactive oxygen species-dependent transcription and secretion of proinflammatory cytokines in human uroepithelial cells. Infect. Immun. 71:4178-4181.

Fang, F.C. 2004. Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. Nat. Rev. Microbiol. 2:820-832.

Fang, L., B.J.Nowicki, P.Urvil, P.Goluszko, S.Nowicki, S.L.Young, and C.Yallampalli. 2004. Epithelial invasion by *Escherichia coli* bearing Dr fimbriae is controlled by nitric oxide-regulated expression of CD55. Infect. Immun. 72:2907-2914.

Farmaki, E., F.Papachristou, R.M.Winn, N.Karatzas, J.Sotiriou, and E.Roilides. 2005. Transforming growth factor-beta1 in the urine of young children with urinary tract infection. Pediatr. Nephrol. 20:180-183.

Faro, S. and D.E.Fenner. 1998. Urinary tract infections. Clin. Obstet. Gynecol. 41:744-754.

Farthing, M.J. 2004. Bugs and the gut: an unstable marriage. Best. Pract. Res Clin Gastroenterol. 18:233-239.

Fetherston, J.D. and R.D.Perry. 1994. The pigmentation locus of *Yersinia pestis* KIM6+ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. Mol. Microbiol. 13:697-708.

Finberg, R.W., F.Re, L.Popova, D.T.Golenbock, and E.A.Kurt-Jones. 2004. Cell activation by Toll-like receptors: role of LBP and CD14. J. Endotoxin. Res. 10:413-418.

Fiorentini, C., L.Falzano, A.Fabbri, A.Stringaro, M.Logozzi, S.Travaglione, S.Contamin, G.Arancia, W.Malorni, and S.Fais. 2001. Activation of rho GTPases by cytotoxic necrotizing factor 1 induces macropinocytosis and scavenging activity in epithelial cells. Mol. Biol. Cell 12:2061-2073.

Fiorentini, C., L.Falzano, S.Travaglione, and A.Fabbri. 2003. Hijacking Rho GTPases by protein toxins and apoptosis: molecular strategies of pathogenic bacteria. Cell Death. Differ. 10:147-152.

Fischer, D. and D.Eisenberg. 1999. Finding families for genomic ORFans. Bioinformatics. 15:759-762.

Fischer, H., M.Yamamoto, S.Akira, B.Beutler, and C.Svanborg. 2006. Mechanism of pathogen-specific TLR4 activation in the mucosa: fimbriae, recognition receptors and adaptor protein selection. Eur. J Immunol. 36:267-77.

Fleckenstein, J.M., L.E.Lindler, E.A.Elsinghorst, and J.B.Dale. 2000. Identification of a gene within a pathogenicity island of enterotoxigenic *Escherichia coli* H10407 required for maximal secretion of the heat-labile enterotoxin. Infect Immun. 68:2766-2774.

Foxman, B., R.Barlow, H.D'Arcy, B.Gillespie, and J.D.Sobel. 2000. Urinary tract infection: self-reported incidence and associated costs. Ann. Epidemiol. 10:509-515.

Foxman, B. and P.Brown. 2003. Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. Infect Dis Clin. North Am. 17:227-241.

Foxman, B., S.D.Manning, P.Tallman, R.Bauer, L.Zhang, J.S.Koopman, B.Gillespie, J.D.Sobel, and C.F.Marrs. 2002. Uropathogenic *Escherichia coli* are more likely than commensal *E. coli* to be shared between heterosexual sex partners. Am. J. Epidemiol. 156:1133-1140.

Frisan, T., X.Cortes-Bratti, and M.Thelestam. 2002. Cytolethal distending toxins and activation of DNA damage-dependent checkpoint responses. Int. J. Med. Microbiol. 291:495-499.

Fukiya, S., H.Mizoguchi, T.Tobe, and H.Mori. 2004. Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* Strains revealed by comparative genomic hybridization microarray. J. Bacteriol. 186:3911-3921.

Funfstuck, R., S.Franke, M.Hellberg, U.Ott, B.Knofel, E.Straube, M.Sommer, and J.Hacker. 2001. Secretion of cytokines by uroepithelial cells stimulated by *Escherichia coli* and *Citrobacter* spp. Int. J. Antimicrob. Agents 17:253-258.

Garcia, M.I., P.Gounon, P.Courcoux, A.Labigne, and C.Le Bouguenec. 1996. The afimbrial adhesive sheath encoded by the afa-3 gene cluster of pathogenic *Escherichia coli* is composed of two adhesins. Mol. Microbiol. 19:683-693.

Garcia, M.I., M.Jouve, J.P.Nataro, P.Gounon, and C.Le Bouguenec. 2000. Characterization of the AfaD-like family of invasins encoded by pathogenic *Escherichia coli* associated with intestinal and extra-intestinal infections. FEBS Lett. 479:111-117.

Gaynes, R. and J.R.Edwards. 2005. Overview of nosocomial infections caused by gram-negative bacilli. Clin Infect Dis 41:848-854.

Gbadegesin, R.A., S.A.Cotton, B.M.Coupes, A.Awan, P.E.Brenchley, and N.J.Webb. 2002. Plasma and urinary soluble adhesion molecule expression is increased during first documented acute pyelonephritis. Arch. Dis. Child 86:218-221.

Gerdes, S.Y., M.D.Scholle, J.W.Campbell, G.Balazsi, E.Ravasz, M.D.Daugherty, A.L.Somera, N.C.Kyrpides, I.Anderson, M.S.Gelfand, A.Bhattacharya, V.Kapatral, M.D'Souza, M.V.Baev, Y.Grechkin, F.Mseeh, M.Y.Fonstein, R.Overbeek, A.L.Barabasi, Z.N.Oltvai, and A.L.Osterman. 2003. Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. J. Bacteriol. 185:5673-5684.

Gerlach, G.F., S.Clegg, and B.L.Allen. 1989. Identification and characterization of the genes encoding the type 3 and type 1 fimbrial adhesins of *Klebsiella pneumoniae*. J. Bacteriol. 171:1262-1270.

Gibb, A.P. and D.M.Edmond. 1992. Urinary IgG antibody against mixed heat-killed coliform antigen and lipopolysaccharide core antigen. J. Clin. Pathol. 45:161-164.

Gillor, O., B.C.Kirkup, and M.A.Riley. 2004. Colicins and microcins: the next generation antimicrobials. Adv. Appl. Microbiol. 54:129-146.

Ginsburg C.M., and G.H.McCracken, Jr. 1982. Urinary tract infections in young infants. Pediatrics. 69:409-412.

Godaly, G., G.Bergsten, L.Hang, H.Fischer, B.Frendeus, A.C.Lundstedt, M.Samuelsson, P.Samuelsson, and C.Svanborg. 2001. Neutrophil recruitment, chemokine receptors, and resistance to mucosal infection. J. Leukoc. Biol. 69:899-906.

Goldberg, M.B., S.A.Boyko, J.R.Butterton, J.A.Stoebner, S.M.Payne, and S.B.Calderwood. 1992. Characterization of a *Vibrio cholerae* virulence factor homologous to the family of TonB-dependent proteins. Mol. Microbiol. 6:2407-2418.

Goluszko, P., E.Goluszko, B.Nowicki, S.Nowicki, V.Popov, and H.Q.Wang. 2005. Vaccination with purified Dr Fimbriae reduces mortality associated with chronic urinary tract infection due to *Escherichia coli* bearing Dr adhesin. Infect. Immun. 73:627-631.

Goni, F.M. and H.Ostolaza. 1998. *E. coli* alpha-hemolysin: a membrane-active protein toxin. Braz. J. Med. Biol. Res. 31:1019-1034.

Gordon, D.M., S.E.Stern, and P.J.Collignon. 2005. Influence of the age and sex of human hosts on the distribution of *Escherichia coli* ECOR groups and virulence traits. Microbiology 151:15-23.

Gosset, G. 2005. Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. Microb. Cell Fact. 4:14.

Griebling T.L. 2004. Urinary Tract Infection in Men. Urologic Diseases in America. 5512: 187–209.

Grodberg, J., M.D.Lundrigan, D.L.Toledo, W.F.Mangel, and J.J.Dunn. 1988. Complete nucleotide sequence and deduced amino acid sequence of the ompT gene of *Escherichia coli* K-12. Nucleic Acids Res. 16:1209.

Grozdanov, L., C.Raasch, J.Schulze, U.Sonnenborn, G.Gottschalk, J.Hacker, and U.Dobrindt. 2004. Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. J. Bacteriol. 186:5432-5441.

Guignot, J., M.F.Bernet-Camard, C.Pous, L.Plancon, C.Le Bouguenec, and A.L.Servin. 2001. Polarized entry of uropathogenic Afa/Dr diffusely adhering *Escherichia coli* strain IH11128 into human epithelial cells: evidence for alpha5beta1 integrin recognition and subsequent internalization through a pathway involving caveolae and dynamic unstable microtubules. Infect. Immun. 69:1856-1868.

Gunther IV, N.W., J.A.Snyder, V.Lockatell, I.Blomfield, D.E.Johnson, and H.L.Mobley. 2002. Assessment of virulence of uropathogenic *Escherichia coli* type 1 fimbrial mutants in which the invertible element is phase-locked on or off. Infect. Immun. 70:3344-3354.

Guyer, D.M., I.R.Henderson, J.P.Nataro, and H.L.Mobley. 2000. Identification of *sat*, an autotransporter toxin produced by uropathogenic *Escherichia coli*. Mol. Microbiol. 38:53-66.

Guyer, D.M., J.S.Kao, and H.L.Mobley. 1998. Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. Infect. Immun. 66:4411-4417.

Guyer, D.M., S.Radulovic, F.E.Jones, and H.L.Mobley. 2002. Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. Infect. Immun. 70:4539-4546.

Gyohda, A., N.Furuya, A.Ishiwa, S.Zhu, and T.Komano. 2004. Structure and function of the shufflon in plasmid R64. Adv. Biophys. 38:183-213.

Haas, H. 2003. Molecular genetics of fungal siderophore biosynthesis and uptake: the role of siderophores in iron uptake and storage. Appl. Microbiol Biotechnol. 62:316-330.

Hacker, J., L.Bender, M.Ott, J.Wingender, B.Lund, R.Marre, and W.Goebel. 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. Microb. Pathog. 8:213-225.

Hacker, J. and E.Carniel. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep. 2:376-381.

Hacker J., H. Kestler, H. Hoschutzky, K. Jann, F. Lottspeich, and T.K. Korhonen. 1993. Cloning and characterization of the S fimbrial adhesin II complex of an *Escherichia coli* O18:K1 meningitis isolate. Infect Immun. 61:544-550.

Hacker, J., M.Ott, G.Blum, R.Marre, J.Heesemann, H.Tschape, and W.Goebel. 1992. Genetics of *Escherichia coli* uropathogenicity: analysis of the O6:K15:H31 isolate 536. Zentralbl. Bakteriol. 276:165-175.

Hantke, K., G.Nicholson, W.Rabsch, and G.Winkelmann. 2003. Salmochelins, siderophores of *Salmonella enterica* and uropathogenic *Escherichia coli* strains, are recognized by the outer membrane receptor IroN. Proc. Natl. Acad. Sci. U. S. A 100:3677-3682.

Hartigan,J.A. and M.A.Wong. 1979. A K-means clustering algorithm. Applied Statistics 28:100-108.

Hasan, R.J., E.Pawelczyk, P.T.Urvil, M.S.Venkatarajan, P.Goluszko, J.Kur, R.Selvarangan, S.Nowicki, W.A.Braun, and B.J.Nowicki. 2002. Structure-function analysis of decay-accelerating factor: identification of residues important for binding of the *Escherichia coli* Dr adhesin and complement regulation. Infect. Immun. 70:4485-4493.

Hayashi, K., Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H, and T.Horiuchi. 2006. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol Syst Biol. 2:2006.0007. Epub 2006 Feb 21.

Hedlund, M., B.Frendeus, C.Wachtler, L.Hang, H.Fischer, and C.Svanborg. 2001. Type 1 fimbriae deliver an LPS- and TLR4-dependent activation signal to CD14-negative cells. Mol. Microbiol. 39:542-552.

Heimer, S.R., D.A.Rasko, C.V.Lockatell, D.E.Johnson, and H.L.Mobley. 2004. Autotransporter genes *pic* and *tsh* are associated with *Escherichia coli* strains that cause acute pyelonephritis and are expressed during urinary tract infection. Infect. Immun. 72:593-597.

Henderson, I.R., J.Czeczulin, C.Eslava, F.Noriega, and J.P.Nataro. 1999. Characterization of *pic*, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. Infect. Immun. 67:5587-5596.

Hernday, A., B.Braaten, and D.Low. 2004. The intricate workings of a bacterial epigenetic switch. Adv. Exp. Med. Biol. 547:83-89.

Hidalgo-Tenorio, C., J.Jimenez-Alonso, L.J.de Dios, M.Tallada, A.Martinez-Brocal, and S.J.Mario. 2004. Urinary tract infections and *lupus erythematosus*. Ann. Rheum. Dis. 63:431-437.

Higgins, M.A., B.R.Berridge, B.J.Mills, A.E.Schultze, H.Gao, G.H.Searfoss, T.K.Baker, and T.P.Ryan. 2003. Gene expression analysis of the acute phase response using a canine microarray. Toxicol. Sci. 74:470-484.

Hochhut, B., U.Dobrindt, and J.Hacker. 2005. Pathogenicity islands and their role in bacterial virulence and survival. Contrib. Microbiol. 12:234-254.

Hoepelman, A.I., R.Meiland, and S.E.Geerlings. 2003. Pathogenesis and management of bacterial urinary tract infections in adult patients with *diabetes mellitus*. Int. J. Antimicrob. Agents 22 Suppl 2:35-43.

Hoesl, C.E. and J.E.Altwein. 2005. The probiotic approach: an alternative treatment option in urology. Eur. Urol. 47:288-296.

Hoffmann, C. and G.Schmidt. 2004. CNF and DNT. Rev. Physiol Biochem. Pharmacol. 152:49-63.

Hoffmann, H., M.W.Hornef, S.Schubert, and A.Roggenkamp. 2001. Distribution of the outer membrane haem receptor protein ChuA in environmental and human isolates of *Escherichia coli*. Int. J. Med. Microbiol. 291:227-230.

Hofman, P. 2005. DNA microarrays. Nephron Physiol 99:85-89.

Honkinen, O., T. Jahnukainen, J. Mertsola, J. Eskola, and O.Ruuskanen. 2000. Bacteremic urinary tract infection in children. Pediatr Infect Dis J. 19:630-634.

Honkinen, O., O.P.Lehtonen, O.Ruuskanen, P.Huovinen, and J.Mertsola. 1999. Cohort study of bacterial species causing urinary tract infection and urinary tract abnormalities in children. BMJ 318:770-771.

Horwitz, M.A. and S.C.Silverstein. 1980. Influence of the *Escherichia coli* capsule on complement fixation and on phagocytosis and killing by human phagocytes. J. Clin. Invest 65:82-94.

Houdouin, V., S.Bonacorsi, N.Brahimi, O.Clermont, X.Nassif, and E.Bingen. 2002. A uropathogenicity island contributes to the pathogenicity of *Escherichia coli* strains that cause neonatal meningitis. Infect. Immun. 70:5865-5869.

Hu, K.K., E.J.Boyko, D.Scholes, E.Normand, C.L.Chen, J.Grafton, and S.D.Fihn. 2004. Risk factors for urinary tract infections in postmenopausal women. Arch. Intern. Med. 164:989-993.

Hua, Q., C.Yang, T.Oshima, H.Mori, and K.Shimizu. 2004. Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures. Appl. Environ. Microbiol. 70:2354-2366.

Huber, W., A.von Heydebreck, H.Sultmann, A.Poustka, and M.Vingron. 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 18 Suppl 1:S96-104.

Hull, R.A., W.H.Donovan, M.Del Terzo, C.Stewart, M.Rogers, and R.O.Darouiche. 2002. Role of type 1 fimbria- and P fimbria-specific adherence in colonization of the neurogenic human bladder by *Escherichia coli*. Infect. Immun. 70:6481-6484.

Hull, R.A., R.E.Gill, P.Hsu, B.H.Minshew, and S.Falkow. 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. 33:933-938.

Ishikawa, K., B.Sundelin, R.Mollby, S.Normark, and J.Winberg. 2004. Receptors for Escherichia coli adhesins in the genitourinary tract in a non-human primate. Scand. J. Urol. Nephrol. 38:3-14.

Jackson, S.L., E.J.Boyko, D.Scholes, L.Abraham, K.Gupta, and S.D.Fihn. 2004. Predictors of urinary tract infection after menopause: a prospective study. Am. J. Med. 117:903-911.

Jahnukainen, T., M.Chen, and G.Celsi. 2005. Mechanisms of renal damage owing to infection. Pediatr. Nephrol. 20:1043-1053.

Janben, T., C.Schwarz, P.Preikschat, M.Voss, H.C.Philipp, and L.H.Wieler. 2001. Virulence-associated genes in avian pathogenic *Escherichia coli* (APEC) isolated from internal organs of poultry having died from colibacillosis. Int. J. Med. Microbiol. 291:371-378.

Janke, B., U.Dobrindt, J.Hacker, and G.Blum-Oehler. 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.

Jantunen, M.E., H.Saxen, S.Lukinmaa, M.Ala-Houhala, and A.Siitonen. 2001. Genomic identity of pyelonephritogenic *Escherichia coli* isolated from blood, urine and faeces of children with urosepsis. J Med Microbiol 50:650-652.

Jin, Q., Z.Yuan, J.Xu, Y.Wang, Y.Shen, W.Lu, J.Wang, H.Liu, J.Yang, F.Yang, X.Zhang, J.Zhang, G.Yang, H.Wu, D.Qu, J.Dong, L.Sun, Y.Xue, A.Zhao, Y.Gao, J.Zhu, B.Kan, K.Ding, S.Chen, H.Cheng, Z.Yao, B.He, R.Chen, D.Ma, B.Qiang, Y.Wen, Y.Hou, and J.Yu. 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucleic Acids Res. 30:4432-4441.

Johansen T.E., M. Cek, K.G. Naber, L. Stratchounski, M.V. Svendsen, and P. Tenke; PEP and PEAP-study investigators; Board of the European Society of Infections in Urology. 2006. Hospital acquired urinary tract infections in urology departments: pathogens, susceptibility, and use of antibiotics. Data from the PEP and PEAP-studies. Int J Antimicrob Agents. 28:91-107. Review.

Johnson, J.R., J.J.Brown, U.B.Carlino, and T.A.Russo. 1998. Colonization with and acquisition of uropathogenic *Escherichia coli* as revealed by polymerase chain reaction-based detection. J. Infect. Dis. 177:1120-1124.

Johnson J.R., O. Clermont, M. Menard, M.A. Kuskowski, B. Picard, and E. Denamur. 2006. Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits, phylogenetic group, and ecological source. J Infect Dis. 194:1141-1150.

Johnson, J.R., S.Jelacic, L.M.Schoening, C.Clabots, N.Shaikh, H.L.Mobley, and P.I.Tarr. 2005a. The IrgA homologue adhesin Iha is an *Escherichia coli* virulence factor in murine urinary tract infection. Infect. Immun. 73:965-971.

Johnson, J.R., C.E.Johnson, and J.N.Maslow. 1999. Clinical and bacteriologic correlates of the *papG* alleles among *Escherichia coli* strains from children with acute cystitis. Pediatr. Infect. Dis. J. 18:446-451.

Johnson, J.R., M.A.Kuskowski, A.Gajewski, S.Soto, J.P.Horcajada, M.T.Jimenez de Anta, and J.Vila. 2005b. Extended virulence genotypes and phylogenetic background of *Escherichia coli* isolates from patients with cystitis, pyelonephritis, or prostatitis. J. Infect. Dis. 191:46-50.

Johnson, J.R., M.A.Kuskowski, K.Smith, T.T.O'Bryan, and S.Tatini. 2005c. Antimicrobial-resistant and extraintestinal pathogenic *Escherichia coli* in retail foods. J Infect Dis 191:1040-1049.

Johnson, J.R., I.Orskov, F.Orskov, P.Goullet, B.Picard, S.L.Moseley, P.L.Roberts, and W.E.Stamm. 1994. O, K, and H antigens predict virulence factors, carboxylesterase B pattern, antimicrobial resistance, and host compromise among *Escherichia coli* strains causing urosepsis. J. Infect. Dis. 169:119-126.

Johnson, J.R., E.Oswald, T.T.O'Bryan, M.A.Kuskowski, and L.Spanjaard. 2002. Phylogenetic distribution of virulence-associated genes among *Escherichia coli* isolates associated with neonatal bacterial meningitis in the Netherlands. J. Infect. Dis. 185:774-784.

Johnson, J.R., K.Owens, A.Gajewski, and M.A.Kuskowski. 2005d. Bacterial characteristics in relation to clinical source of *Escherichia coli* isolates from women with acute cystitis or pyelonephritis and uninfected women. J. Clin. Microbiol. 43:6064-6072.

Johnson, J.R. and T.A.Russo. 2002. Extraintestinal pathogenic *Escherichia coli*: "the other bad *E. coli*". J Lab Clin. Med 139:155-162.

Johnson, J.R. and T.A.Russo. 2005. Molecular epidemiology of extraintestinal pathogenic (uropathogenic*) Escherichia coli*. Int. J. Med. Microbiol. 295:383-404.

Johnson, J.R. and A.L.Stell. 2000. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. J. Infect. Dis. 181:261-272.

Johnson, J.R., S.J.Weissman, A.L.Stell, E.Trintchina, D.E.Dykhuizen, and E.V.Sokurenko. 2001. Clonal and pathotypic analysis of archetypal *Escherichia coli* cystitis isolate NU14. J. Infect. Dis. 184:1556-1565.

Johnson T.J, K.E. Siek, S.J. Johnson, and L.K. Nolan. 2005e. DNA sequence and comparative genomics of pAPEC-O2-R, an avian pathogenic *Escherichia coli* transmissible R plasmid. Antimicrob Agents Chemother. 49:4681-4688.

Johnson, T.J., J.Skyberg, and L.K.Nolan. 2004. Multiple antimicrobial resistance region of a putative virulence plasmid from an *Escherichia coli* isolate incriminated in avian colibacillosis. Avian Dis 48:351-360.

Jones, B., K.Peake, A.J.Morris, L.M.McCowan, and M.R.Battin. 2004. *Escherichia coli*: a growing problem in early onset neonatal sepsis. Aust. N. Z. J. Obstet. Gynaecol. 44:558-561.

Justice, S.S., C.Hung, J.A.Theriot, D.A.Fletcher, G.G.Anderson, M.J.Footer, and S.J.Hultgren. 2004. Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis. Proc. Natl. Acad. Sci. U. S. A 101:1333-1338.

Kanamaru, S., H.Kurazono, S.Ishitoya, A.Terai, T.Habuchi, M.Nakano, O.Ogawa, and S.Yamamoto. 2003. Distribution and genetic association of putative uropathogenic virulence factors *iroN*, *iha*, *kpsMT*, *ompT* and *usp* in *Escherichia coli* isolated from urinary tract infections in Japan. J. Urol. 170:2490-2493.

Kantele, A., T.Mottonen, K.Ala-Kaila, and H.S.Arvilommi. 2003. P fimbria-specific B cell responses in patients with urinary tract infection. J. Infect. Dis. 188:1885-1891.

Karkkainen, U.M., R.Ikaheimo, M.L.Katila, and A.Siitonen. 2000. Recurrence of urinary tract infections in adult patients with community-acquired pyelonephritis caused by *E. coli*: a 1-year follow-up. Scand. J. Infect. Dis. 32:495-499.

Kawahara, M., L.G. Human, J.S. Winningham, and G.J. Domingue. 1994. Antibodies to *Escherichia coli* 06 porins cross-react with urinary pathogens. Immunobiology 192:65-76.

Keenleyside, W.J., D. Bronner, K. Jann, B. Jann, and C. Whitfield. 1993. Coexpression of colanic acid and serotype-specific capsular polysaccharides in *Escherichia coli* strains with group II K antigens. J. Bacteriol. 175:6725-6730.

Kingsley, R., W.Rabsch, P.Stephens, M.Roberts, R.Reissbrodt, and P.H.Williams. 1995. Iron supplying systems of *Salmonella* in diagnostics, epidemiology and infection. FEMS Immunol. Med Microbiol 11:257-264.

Klumpp D.J., M.T. Rycyk, M.C. Chen, P. Thumbikat, S. Sengupta, and A.J. Schaeffer. 2006. Uropathogenic *Escherichia coli* induces extrinsic and intrinsic cascades to initiate urothelial apoptosis. Infect Immun. 74:5106-5113.

Koczura, R. and A.Kaznowski. 2003a. Occurrence of the *Yersinia* high-pathogenicity island and iron uptake systems in clinical isolates of *Klebsiella pneumoniae*. Microb. Pathog. 35:197-202.

Koczura, R. and A.Kaznowski. 2003b. The *Yersinia* high-pathogenicity island and iron-uptake systems in clinical isolates of *Escherichia coli*. J. Med. Microbiol. 52:637-642.

Konopka, K. and J.B.Neilands. 1984. Effect of serum albumin on siderophore-mediated utilization of transferrin iron. Biochemistry 23:2122-2127.

Kontiokari, T., J.Laitinen, L.Jarvi, T.Pokka, K.Sundqvist, and M.Uhari. 2003. Dietary factors protecting women from urinary tract infection. Am. J. Clin. Nutr. 77:600-604.

Korczak, B., J.Frey, J.Schrenzel, G.Pluschke, R.Pfister, R.Ehricht, and P.Kuhnert. 2005. Use of diagnostic microarrays for determination of virulence gene patterns of *Escherichia coli* K1, a major cause of neonatal meningitis. J. Clin. Microbiol. 43:1024-1031.

Korhonen, T.K., M.V.Valtonen, J.Parkkinen, V.Vaisanen-Rhen, J.Finne, F.Orskov, I.Orskov, S.B.Svenson, and P.H.Makela. 1985. Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. Infect. Immun. 48:486-491.

Kostic, T., A.Weilharter, S.Rubino, G.Delogu, S.Uzzau, K.Rudi, A.Sessitsch, and L. Bodrossy. 2007. A microbial diagnostic microarray technique for the sensitive detection and identification of pathogenic bacteria in a background of nonpathogens. Anal Biochem. 360:244-254.

Kukkonen, M. and T.K.Korhonen. 2004. The omptin family of enterobacterial surface proteases/adhesins: from housekeeping in *Escherichia coli* to systemic spread of *Yersinia pestis*. Int. J. Med. Microbiol. 294:7-14.

Kurazono, H., S.Yamamoto, M.Nakano, G.B.Nair, A.Terai, W.Chaicumpa, and H.Hayashi. 2000. Characterization of a putative virulence island in the chromosome of uropathogenic *Escherichia coli* possessing a gene encoding a uropathogenic-specific protein. Microb. Pathog. 28:183-189.

Kusecek, B., H.Wloch, A.Mercer, V.Vaisanen, G.Pluschke, T.Korhonen, and M.Achtman. 1984. Lipopolysaccharide, capsule, and fimbriae as virulence factors among O1, O7, O16, O18, or O75 and K1, K5, or K100 *Escherichia coli*. Infect. Immun. 43:368-379.

Kuzminov A. 1999. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. Microbiol Mol Biol Rev. 63:751-813. Review.

Labigne-Roussel, A. and S.Falkow. 1988. Distribution and degree of heterogeneity of the afimbrial-adhesin-encoding operon (*afa*) among uropathogenic *Escherichia coli* isolates. Infect. Immun. 56:640-648.

Landraud, L., C.Pulcini, P.Gounon, G.Flatau, P.Boquet, and E.Lemichez. 2004. *E. coli* CNF1 toxin: a two-in-one system for host-cell invasion. Int. J. Med. Microbiol. 293:513-518.

Lane, M.C., V.Lockatell, G.Monterosso, D.Lamphier, J.Weinert, J.R.Hebel, D.E.Johnson, and H.L.Mobley. 2005. Role of motility in the colonization of uropathogenic *Escherichia coli* in the urinary tract. Infect. Immun. 73:7644-7656.

Langermann, S., R.Mollby, J.E.Burlein, S.R.Palaszynski, C.G.Auguste, A.DeFusco, R.Strouse, M.A.Schenerman, S.J.Hultgren, J.S.Pinkner, J.Winberg, L.Guldevall, M.Soderhall, K.Ishikawa, S.Normark, and S.Koenig. 2000. Vaccination with FimH adhesin protects cynomolgus monkeys from colonization and infection by uropathogenic *Escherichia coli*. J. Infect. Dis. 181:774-778.

Lankford,C.E. 2005. Bacterial assimilation of iron. Crit. Rev. Mikrobiol. 273-331.

Laporta, M.Z., M.L.Silva, I.C.Scaletsky, and L.R.Trabulsi. 1986. Plasmids coding for drug resistance and localized adherence to HeLa cells in enteropathogenic *Escherichia coli* O55:H- and O55:H6. Infect Immun. 51:715-717.

Lara-Tejero, M. and J.E.Galan. 2001. CdtA, CdtB, and CdtC form a tripartite complex that is required for cytolethal distending toxin activity. Infect. Immun. 69:4358-4365.

Larsson, A., J.Ohlsson, K.W.Dodson, S.J.Hultgren, U.Nilsson, and J.Kihlberg. 2003. Quantitative studies of the binding of the class II PapG adhesin from uropathogenic *Escherichia coli* to oligosaccharides. Bioorg. Med. Chem. 11:2255-2261.

Lawrence, J.G., and H.Ochman. 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. 44:383-397.

Le Bouguenec, C., L.Lalioui, M.L.du, M.Jouve, P.Courcoux, S.Bouzari, R.Selvarangan, B.J.Nowicki, Y.Germani, A.Andremont, P.Gounon, and M.I.Garcia. 2001. Characterization of AfaE adhesins produced by extraintestinal and intestinal human *Escherichia coli* isolates: PCR assays for detection of Afa adhesins that do or do not recognize Dr blood group antigens. J. Clin. Microbiol. 39:1738-1745.

Le Gall, T., P.Darlu, P.Escobar-Paramo, B.Picard, and E.Denamur. 2005. Selection-driven transcriptome polymorphism in *Escherichia coli*/*Shigella* species. Genome Res. 15:260-268.

Le, J., G.G.Briggs, A.McKeown, and G.Bustillo. 2004. Urinary tract infections during pregnancy. Ann. Pharmacother. 38:1692-1701.

Lee, Y.M., P.A.DiGiuseppe, T.J.Silhavy, and S.J.Hultgren. 2004. P pilus assembly motif necessary for activation of the CpxRA pathway by PapE in *Escherichia coli*. J. Bacteriol. 186:4326-4337.

Leflon-Guibout, V., C.Jurand, S.Bonacorsi, F.Espinasse, M.C.Guelfi, F.Duportail, B.Heym, E.Bingen, and M.H.Nicolas-Chanoine. 2004. Emergence and spread of three clonally related virulent isolates of CTX-M-15-producing *Escherichia coli* with variable resistance to aminoglycosides and tetracycline in a French geriatric hospital1. Antimicrob. Agents Chemother 48:3736-3742.

Lemichez, E., G.Flatau, M.Bruzzone, P.Boquet, and M.Gauthier. 1997. Molecular localization of the Escherichia coli cytotoxic necrotizing factor CNF1 cell-binding and catalytic domains. Mol. Microbiol. 24:1061-1070.

Lerat, E., V.Daubin, H.Ochman, and N.A.Moran. 2005. Evolutionary origins of genomic repertoires in bacteria. PLoS. Biol. 3:e130.

Levy, S.B., G.B.FitzGerald, and A.B.Macone. 1976. Changes in intestinal flora of farm personnel after introduction of a tetracycline-supplemented feed on a farm. N Engl J Med. 295:583-588.

Leying, H., S.Suerbaum, H.P.Kroll, D.Stahl, and W.Opferkuch. 1990. The capsular polysaccharide is a major determinant of serum resistance in K-1-positive blood culture isolates of *Escherichia coli*. Infect. Immun. 58:222-227.

Liu, Y., J.X.Han, H.Y.Huang, and B.Zhu. 2005. Development and evaluation of 16S rDNA microarray for detecting bacterial pathogens in cerebrospinal fluid. Exp. Biol. Med (Maywood.). 230:587-591.

Lutwyche, P., R.Rupps, J.Cavanagh, R.A.Warren, and D.E.Brooks. 1994. Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglutinin 1, an integral membrane hemagglutinin from *Escherichia col*i O9:H10:K99. Infect Immun. 62:5020-5026.

Malorni, W., M.G.Quaranta, E.Straface, L.Falzano, A.Fabbri, M.Viora, and C.Fiorentini. 2003. The Rac-activating toxin cytotoxic necrotizing factor 1 oversees NK cell-mediated activity by regulating the actin/microtubule interplay. J. Immunol. 171:4195-4202.

Marild, S., S.Hansson, U.Jodal, A.Oden, and K.Svedberg. 2004. Protective effect of breastfeeding against urinary tract infection. Acta Paediatr. 93:164-168.

Marrs, C.F., L.Zhang, and B.Foxman. 2005. *Escherichia coli* mediated urinary tract infections: Are there distinct uropathogenic *E. coli* (UPEC) pathotypes? FEMS Microbiol. Lett. 252:183-190.

Martin-Sosa, S., M.J.Martin, and P.Hueso. 2002. The sialylated fraction of milk oligosaccharides is partially responsible for binding to enterotoxigenic and uropathogenic *Escherichia coli* human strains. J. Nutr. 132:3067-3072.

Martinez, M.J., A.D.Aragon, A.L.Rodriguez, J.M.Weber, J.A.Timlin, M.B.Sinclair, D.M.Haaland, and M.Werner-Washburne. 2003. Identification and removal of contaminating fluorescence from commercial and in-house printed DNA microarrays. Nucleic Acids Res. 31:e18.

Maynard, C., F.Berthiaume, K.Lemarchand, I.Harel, P.Payment, P.Bayardelle, L.Masson, and R.Brousseau. 2005. Waterborne pathogen detection by use of oligonucleotide-based microarrays. Appl Environ Microbiol. 71:8548-8557.

Medigue, C., T.Rouxel, P.Vigier, A.Henaut, and A.Danchin. 1991. Evidence for horizontal gene transfer in *Escherichia coli* speciation. J. Mol. Biol. 222:851-856.

Mehnert-Kay, S.A. 2005. Diagnosis and management of uncomplicated urinary tract infections. Am. Fam. Physician 72:451-456.

Mey, A.R., E.E.Wyckoff, A.G.Oglesby, E.Rab, R.K.Taylor, and S.M.Payne. 2002. Identification of the *Vibrio cholerae* enterobactin receptors VctA and IrgA: IrgA is not required for virulence. Infect. Immun. 70:3419-3426.

Mills, M., K.C.Meysick, and A.D.O'Brien. 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.

Miyazaki, J., W.Ba-Thein, T.Kumao, H.Akaza, and H.Hayashi. 2002. Identification of a type III secretion system in uropathogenic *Escherichia coli*. FEMS Microbiol. Lett. 212:221-228.

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

Moon, Y.H., T.Tanabe, T.Funahashi, K.Shiuchi, H.Nakao, and S.Yamamoto. 2004. Identification and characterization of two contiguous operons required for aerobactin transport and biosynthesis in *Vibrio mimicus*. Microbiol. Immunol. 48:389-398.

Moreno, E., I.Planells, G.Prats, A.M.Planes, G.Moreno, and A.Andreu. 2005. Comparative study of *Escherichia coli* virulence determinants in strains causing urinary tract bacteremia versus strains causing pyelonephritis and other sources of bacteremia. Diagn. Microbiol Infect Dis 53:93-99.

Morin, M.D. and W.J.Hopkins. 2002. Identification of virulence genes in uropathogenic *Escherichia coli* by multiplex polymerase chain reaction and their association with infectivity in mice. Urology 60:537-541.

Morris, R.T. and G.Drouin. 2004. Ectopic gene conversions in four *Escherichia coli* genomes: increased recombination in pathogenic strains. J. Mol. Evol. 58:596-605.

Morschhauser, J., V.Vetter, T.Korhonen, B.E.Uhlin, and J.Hacker. 1993. Regulation and binding properties of S fimbriae cloned from *E. coli* strains causing urinary tract infection and meningitis. Zentralbl. Bakteriol. 278:165-176.

Mulvey, M.A. 2002. Adhesion and entry of uropathogenic *Escherichia coli*. Cell Microbiol. 4:257-271.

Mulvey, M.A., J.D.Schilling, J.J.Martinez, and S.J.Hultgren. 2000. Bad bugs and beleaguered bladders: interplay between uropathogenic *Escherichia coli* and innate host defenses. Proc. Natl. Acad. Sci. U. S. A 97:8829-8835.

Munro, P., G.Flatau, A.Doye, L.Boyer, O.Oregioni, J.L.Mege, L.Landraud, and E.Lemichez. 2004. Activation and proteasomal degradation of rho GTPases by cytotoxic necrotizing factor-1 elicit a controlled inflammatory response. J. Biol. Chem. 279:35849-35857.

Murray, A.C., M.A.Kuskowski, and J.R.Johnson. 2004. Virulence factors predict Escherichia coli colonization patterns among human and animal household members. Ann. Intern. Med. 140:848-849.

Mysorekar, I.U., M.A.Mulvey, S.J.Hultgren, and J.I.Gordon. 2002. Molecular regulation of urothelial renewal and host defenses during infection with uropathogenic *Escherichia coli.* J. Biol. Chem. 277:7412-7419.

Nagy, G., U.Dobrindt, G.Schneider, A.S.Khan, J.Hacker, and L.Emody. 2002. Loss of regulatory protein RfaH attenuates virulence of uropathogenic *Escherichia coli*. Infect. Immun. 70:4406-4413.

Nataro, J.P. 2005. Enteroaggregative *Escherichia coli* pathogenesis. Curr. Opin. Gastroenterol. 21:4-8.

Negre, V.L., S.Bonacorsi, S.Schubert, P.Bidet, X.Nassif, and E.Bingen. 2004. The siderophore receptor IroN, but not the high-pathogenicity island or the hemin receptor ChuA, contributes to the bacteremic step of *Escherichia coli* neonatal meningitis. Infect. Immun. 72:1216-1220.

Neilands,J.B. A crystalline organo-iron pigment from the smut fungus *Ustitago sphaerogena*. 1989. J. Am. Chem. Soc. 74:4846-4847.

Neilands, J.B. 1992. Mechanism and regulation of synthesis of aerobactin in *Escherichia coli* K12 (pColV-K30). Can. J Microbiol 38:728-733.

Nicolle, L.E. 2005. Catheter-related urinary tract infection. Drugs Aging 22:627-639.

Nicolle, L.E., J.Brunka, E.Ujack, and L.Bryan. 1989. Antibodies to major outer membrane proteins of *Escherichia coli* in urinary infection in the elderly. J. Infect. Dis. 160:627-633.

Nishimura, H., A.Gogami, Y.Miyagawa, A.Nanbo, Y.Murakami, T.Baba, and S.Nagasawa. 2001. Bactericidal/permeability-increasing protein promotes complement activation for neutrophil-mediated phagocytosis on bacterial surface. Immunology 103:519-525.

Nissle, A. Explanations of the significance of colonic dysbacteria & the mechanism of action of *E. coli* therapy (mutaflor). 1959. Medizinische. 4:1017-1022.

Nitschke, M., S.Wiehl, P.C.Baer, and B.Kreft. 2002. Bactericidal activity of renal tubular cells: the putative role of human beta-defensins. Exp. Nephrol. 10:332-337.

Nougayrede, J.P., P.J.Fernandes, and M.S.Donnenberg. 2003. Adhesion of enteropathogenic *Escherichia coli* to host cells. Cell Microbiol. 5:359-372.

Nougayrede, J.P., F.Taieb, J.De Rycke, and E.Oswald. 2005. Cyclomodulins: bacterial effectors that modulate the eukaryotic cell cycle. Trends Microbiol. 13:103-110.

Nowicki, B., R.Selvarangan, and S.Nowicki. 2001. Family of *Escherichia coli* Dr adhesins: decay-accelerating factor receptor recognition and invasiveness. J Infect Dis 183 Suppl 1:S24-S27.

Nowrouzian, F., B.Hesselmar, R.Saalman, I.L.Strannegard, N.Aberg, A.E.Wold, and I.Adlerberth. 2003. *Escherichia coli* in infants' intestinal microflora: colonization rate, strain turnover, and virulence gene carriage. Pediatr. Res 54:8-14.

Nowrouzian, F.L., A.E.Wold, and I.Adlerberth. 2005. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. J. Infect. Dis. 191:1078-1083.

O'Brien, A.D., A.R.Melton, C.K.Schmitt, M.L.McKee, M.L.Batts, and D.E.Griffin. 1993. Profile of *Escherichia coli* O157:H7 pathogen responsible for hamburger-borne outbreak of hemorrhagic colitis and hemolytic uremic syndrome in Washington. J. Clin. Microbiol. 31:2799-2801.

O'Brien, I.G. and F.Gibson. 1970. The structure of enterochelin and related 2,3-dihydroxy-N-benzoylserine conjugates from *Escherichia coli*. Biochim. Biophys. Acta 215:393-402.

Ochman, H. and I.B.Jones. 2000. Evolutionary dynamics of full genome content in *Escherichia coli*. EMBO J. 19:6637-6643.

Ochman, H., E.Lerat, and V.Daubin. 2005. Examining bacterial species under the specter of gene transfer and exchange. Proc. Natl. Acad. Sci. U. S. A 102 Suppl 1:6595-6599.

Ochman, H. and S.R.Santos. 2005. Exploring microbial microevolution with microarrays. Infect. Genet. Evol. 5:103-108.

Ochman, H. and R.K.Selander. 1984a. Evidence for clonal population structure in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A 81:198-201.

Ochman, H. and R.K.Selander. 1984b. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. 157:690-693.

Oelschlaeger, T.A., U.Dobrindt, and J.Hacker. 2002. Virulence factors of uropathogens. Curr. Opin. Urol. 12:33-38.

Ojo, K.K., C.Kehrenberg, H.A.Odelola, and S.Schwarz. 2003. Structural analysis of the tetracycline resistance gene region of a small multiresistance plasmid from uropathogenic *Escherichia coli* isolated in Nigeria. J. Antimicrob. Chemother. 52:1043-1044.

Olsburgh, J., P.Harnden, R.Weeks, B.Smith, A.Joyce, G.Hall, R.Poulsom, P.Selby, and J.Southgate. 2003. Uroplakin gene expression in normal human tissues and locally advanced bladder cancer. J. Pathol. 199:41-49.

Olsen, A., M.J.Wick, M.Morgelin, and L.Bjorck. 1998. Curli, fibrous surface proteins of *Escherichia coli*, interact with major histocompatibility complex class I molecules. Infect Immun. 66:944-949.

Oswald, E., M.Sugai, A.Labigne, H.C.Wu, C.Fiorentini, P.Boquet, and A.D.O'Brien. 1994. Cytotoxic necrotizing factor type 2 produced by virulent *Escherichia coli* modifies the small GTP-binding proteins Rho involved in assembly of actin stress fibers. Proc. Natl. Acad. Sci. U. S. A 91:3814-3818.

Ott, M., H.Hoschutzky, K.Jann, D.Van, I, and J.Hacker. 1988. Gene clusters for S fimbrial adhesin (*sfa*) and F1C fimbriae (*foc*) of *Escherichia coli*: comparative aspects of structure and function. J Bacteriol. 170:3983-3990.

Otto, B.R., S.J.van Dooren, J.H.Nuijens, J.Luirink, and B.Oudega. 1998. Characterization of a hemoglobin protease secreted by the pathogenic *Escherichia coli* strain EB1. J. Exp. Med. 188:1091-1103.

Oxhamre, C., A.Richter-Dahlfors, V.P.Zhdanov, and B.Kasemo. 2005. A minimal generic model of bacteria-induced intracellular Ca^{2+} oscillations in epithelial cells. Biophys. J. 88:2976-2981.

Pal, C., B.Papp, and M.J.Lercher. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. 37:1372-1375.

Parham, N.J., U.Srinivasan, M.Desvaux, B.Foxman, C.F.Marrs, and I.R.Henderson. 2004. PicU, a second serine protease autotransporter of uropathogenic *Escherichia coli*. FEMS Microbiol. Lett. 230:73-83.

Paterson DL. 2006. Resistance in gram-negative bacteria: *enterobacteriaceae*. Am J Med. 119:20-28; discussion: 62-70. Review.

Patzer, S.I., M.R.Baquero, D.Bravo, F.Moreno, and K.Hantke. 2003. The colicin G, H and X determinants encode microcins M and H47, which might utilize the catecholate siderophore receptors FepA, Cir, Fiu and IroN. Microbiology 149:2557-2570.

Peres, S.Y., O.Marches, F.Daigle, J.P.Nougayrede, F.Herault, C.Tasca, J.De Rycke, and E.Oswald. 1997. A new cytolethal distending toxin (CDT) from *Escherichia coli* producing CNF2 blocks HeLa cell division in G2/M phase. Mol. Microbiol. 24:1095-1107.

Perna, N.T., G.Plunkett, III, V.Burland, B.Mau, J.D.Glasner, D.J.Rose, G.F.Mayhew, P.S.Evans, J.Gregor, H.A.Kirkpatrick, G.Posfai, J.Hackett, S.Klink, A.Boutin, Y.Shao, L.Miller, E.J.Grotbeck, N.W.Davis, A.Lim, E.T.Dimalanta, K.D.Potamousis, J.Apodaca, T.S.Anantharaman, J.Lin, G.Yen, D.C.Schwartz, R.A.Welch, and F.R.Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529-533.

Pettis, G.S., and M.A.McIntosh. 1987. Molecular characterization of the *Escherichia coli* enterobactin cistron *entF* and coupled expression of *entF* and the *fes* gene. J. Bacteriol. 169:4154-4162.

Picard, B., J.S.Garcia, S.Gouriou, P.Duriez, N.Brahimi, E.Bingen, J.Elion, and E.Denamur. 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infect. Immun. 67:546-553.

Plancon, L., M.L.du, S.Le Friec, P.Gounon, M.Jouve, J.Guignot, A.Servin, and C.Le Bouguenec. 2003. Recognition of the cellular beta1-chain integrin by the bacterial AfaD invasin is implicated in the internalization of *afa*-expressing pathogenic *Escherichia coli* strains. Cell Microbiol. 5:681-693.

Pluschke, G., J.Mayden, M.Achtman, and R.P.Levine. 1983. Role of the capsule and the O antigen in resistance of O18:K1 *Escherichia coli* to complement-mediated killing. Infect. Immun. 42:907-913.

Pourbakhsh, S.A., M.Dho-Moulin, A.Bree, C.Desautels, B.Martineau-Doize, and J.M.Fairbrother. 1997. Localization of the *in vivo* expression of P and F1 fimbriae in chickens experimentally inoculated with pathogenic *Escherichia coli*. Microb. Pathog. 22:331-341.

Prager, R., U.Strutz, A.Fruth, and H.Tschape. 2003. Subtyping of pathogenic *Escherichia coli* strains using flagellar (H)-antigens: serotyping versus *fliC* polymorphisms. Int. J. Med. Microbiol. 292:477-486.

Pupo, G.M., D.K.Karaolis, R.Lan, and P.R.Reeves. 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. Infect. Immun. 65:2685-2692.

Purdy, G.E. and S.M.Payne. 2001. The SHI-3 iron transport island of *Shigella boydii* 0-1392 carries the genes for aerobactin synthesis and transport. J. Bacteriol. 183:4176-4182.

Raymond, K.N., E.A.Dertz, and S.S.Kim. 2003. Enterobactin: an archetype for microbial iron transport. Proc. Natl. Acad. Sci. U. S. A 100:3584-3588.

Redford, P., P.L.Roesch, and R.A.Welch. 2003. DegS is necessary for virulence and is among extraintestinal *Escherichia coli* genes induced in murine peritonitis. Infect. Immun. 71:3088-3096.

Reisner, A., J.A.Haagensen, M.A.Schembri, E.L.Zechner, and S.Molin. 2003. Development and maturation of *Escherichia coli* K-12 biofilms. Mol. Microbiol. 48:933-946.

Ren, D., L.A.Bedzyk, S.M.Thomas, R.W.Ye, and T.K.Wood. 2004. Gene expression in *Escherichia coli* biofilms. Appl. Microbiol. Biotechnol. 64:515-524.

Rizvi, M., and S.Kumar. 2003. Serum resistance of *Escherichia coli* strains causing urinary tract infection and diarrhoea in relation to alpha haemolysin production and O type. Indian J. Pathol. Microbiol. 46:504-506.

Roberts, J.A., M.B.Kaack, G.Baskin, M.R.Chapman, D.A.Hunstad, J.S.Pinkner, and S.J.Hultgren. 2004. Antibody responses and protection from pyelonephritis following vaccination with purified *Escherichia coli* PapDG protein. J. Urol. 171:1682-1685.

Rodriguez-Siek, K.E., C.W.Giddings, C.Doetkott, T.J.Johnson, M.K.Fakhr, and L.K.Nolan. 2005. Comparison of *Escherichia coli* isolates implicated in human urinary tract infection and avian colibacillosis. Microbiology 151:2097-2110.

Roesch, P.L., P.Redford, S.Batchelet, R.L.Moritz, S.Pellett, B.J.Haugen, F.R.Blattner, and R.A.Welch. 2003. Uropathogenic *Escherichia coli* use d-serine deaminase to modulate infection of the murine urinary tract. Mol. Microbiol. 49:55-67.

Rozenberg, S., A.Pastijn, R.Gevers, and D.Murillo. 2004. Estrogen therapy in older patients with recurrent urinary tract infections: a review. Int. J. Fertil. Womens Med. 49:71-74.

Ruiz, J., K.Simon, J.P.Horcajada, M.Velasco, M.Barranco, G.Roig, A.Moreno-Martinez, J.A.Martinez, d.A.Jimenez, J.Mensa, and J.Vila. 2002. Differences in virulence factors among clinical isolates of *Escherichia coli* causing cystitis and pyelonephritis in women and prostatitis in men. J. Clin. Microbiol. 40:4445-4449.

Russo, T., J.J.Brown, S.T.Jodush, and J.R.Johnson. 1996. The O4 specific antigen moiety of lipopolysaccharide but not the K54 group 2 capsule is important for urovirulence of an extraintestinal isolate of *Escherichia coli*. Infect. Immun. 64:2343-2348.

Russo, T.A., U.B.Carlino, and J.R.Johnson. 2001. Identification of a new iron-regulated virulence gene, *ireA*, in an extraintestinal pathogenic isolate of *Escherichia coli*. Infect. Immun. 69:6209-6216.

Russo, T.A., U.B.Carlino, A.Mong, and S.T.Jodush. 1999. Identification of genes in an extraintestinal isolate of *Escherichia coli* with increased expression after exposure to human urine. Infect. Immun. 67:5306-5314.

Russo, T.A. and J.R.Johnson. 2003. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes. Infect 5:449-456.

Russo, T.A., C.D.McFadden, U.B.Carlino-MacDonald, J.M.Beanan, T.J.Barnard, and J.R.Johnson. 2002. IroN functions as a siderophore receptor and is an urovirulence factor in an extraintestinal pathogenic isolate of *Escherichia coli*. Infect. Immun. 70:7156-7160.

Russo, T.A., S.Wenderoth, U.B.Carlino, J.M.Merrick, and A.J.Lesse. 1998. Identification, genomic organization, and analysis of the group III capsular polysaccharide genes *kpsD*, *kpsM*, *kps*T, and kpsE from an extraintestinal isolate of *Escherichia coli* (CP9, O4/K54/H5). J. Bacteriol. 180:338-349.

Saban, M.R., N.B.Nguyen, T.G.Hammond, and R.Saban. 2002. Gene expression profiling of mouse bladder inflammatory responses to LPS, substance P, and antigen-stimulation. Am. J. Pathol. 160:2095-2110.

Saemann, M.D., T.Weichhart, W.H.Horl, and G.J.Zlabinger. 2005. Tamm-Horsfall protein: a multilayered defence molecule against urinary tract infection. Eur. J Clin. Invest 35:227-235.

Salit, I.E., J.Hanley, L.Clubb, and S.Fanning. 1988. The human antibody response to uropathogenic *Escherichia coli*: a review. Can. J. Microbiol. 34:312-318.

Schembri, M.A., D.W.Ussery, C.Workman, H.Hasman, and P.Klemm. 2002. DNA microarray analysis of fim mutations in *Escherichia coli*. Mol. Genet. Genomics 267:721-729.

Schena, M., D.Shalon, R.W.Davis, and P.O.Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-470.

Schilling, J.D., S.M.Martin, D.A.Hunstad, K.P.Patel, M.A.Mulvey, S.S.Justice, R.G.Lorenz, and S.J.Hultgren. 2003. CD14- and Toll-like receptor-dependent activation of bladder epithelial cells by lipopolysaccharide and type 1 piliated *Escherichia coli*. Infect. Immun. 71:1470-1480.

Schlager, T.A., T.S.Whittam, J.O.Hendley, J.L.Bhang, C.L.Wobbe, and A.Stapleton. 2003. Variation in frequency of the virulence-factor gene in *Escherichia coli* clones colonizing the stools and urinary tracts of healthy prepubertal girls. J. Infect. Dis. 188:1059-1064.

Schmidt, H., and M.Hensel. 2004. Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol. Rev. 17:14-56.

Schneider, G., U.Dobrindt, H.Bruggemann, G.Nagy, B.Janke, G.Blum-Oehler, C.Buchrieser, G.Gottschalk, L.Emody, and J.Hacker. 2004. The pathogenicity island-associated K15 capsule determinant exhibits a novel genetic structure and correlates with virulence in uropathogenic *Escherichia coli* strain 536. Infect. Immun. 72:5993-6001.

Scholes, D., T.M.Hooton, P.L.Roberts, K.Gupta, A.E.Stapleton, and W.E.Stamm. 2005. Risk factors associated with acute pyelonephritis in healthy women. Ann Intern. Med 142:20-27.

Schubert, S., S.Dufke, J.Sorsa, and J.Heesemann. 2004. A novel integrative and conjugative element (ICE) of *Escherichia coli*: the putative progenitor of the *Yersinia* high-pathogenicity island. Mol. Microbiol. 51:837-848.

Schubert, S., B.Picard, S.Gouriou, J.Heesemann, and E.Denamur. 2002. *Yersinia* high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. Infect Immun. 70:5335-5337.

Schubert, S., A.Rakin, and J.Heesemann. 2004. The *Yersinia* high-pathogenicity island (HPI): evolutionary and functional aspects. Int. J. Med. Microbiol. 294:83-94.

Schubert, S., A.Rakin, H.Karch, E.Carniel, and J.Heesemann. 1998. Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. Infect. Immun. 66:480-485.

Seed, P.C. and S.J.Hultgren. 2005. Blueprinting the regulatory response of *Escherichia coli* to the urinary tract. Trends Microbiol. 13:246-248.

Servin, A.L. 2005. Pathogenesis of Afa/Dr Diffusely Adhering *Escherichia coli*. Clin. Microbiol. Rev. 18:264-292.

Sharp, P.M. and W.H.Li. 1987. The codon Adaptation Index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281-1295.

Siitonen, A. 1992. *Escherichia coli* in fecal flora of healthy adults: serotypes, P and type 1C fimbriae, non-P mannose-resistant adhesins, and hemolytic activity. J Infect Dis 166:1058-1065.

Sircili, M.P., M.Walters, L.R.Trabulsi, and V.Sperandio. 2004. Modulation of enteropathogenic *Escherichia coli* virulence by quorum sensing. Infect. Immun. 72:2329-2337.

Snyder, J.A., B.J.Haugen, E.L.Buckles, C.V.Lockatell, D.E.Johnson, M.S.Donnenberg, R.A.Welch, and H.L.Mobley. 2004. Transcriptome of uropathogenic *Escherichia coli* during urinary tract infection. Infect. Immun. 72:6373-6381.

Snyder, J.A., B.J.Haugen, C.V.Lockatell, N.Maroncle, E.C.Hagan, D.E.Johnson, R.A.Welch, and H.L.Mobley. 2005. Coordinate Expression of Fimbriae in Uropathogenic *Escherichia coli*. Infect Immun. 73:7588-7596.

Soderblom, T., A.Laestadius, C.Oxhamre, A.Aperia, and A.Richter-Dahlfors. 2002. Toxin-induced calcium oscillations: a novel strategy to affect gene regulation in target cells. Int. J. Med. Microbiol. 291:511-515.

Sokurenko, E.V., M.Feldgarden, E.Trintchina, S.J.Weissman, S.Avagyan, S.Chattopadhyay, J.R.Johnson, and D.E.Dykhuizen. 2004. Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in *Escherichia coli.* Mol. Biol. Evol. 21:1373-1383.

Sokurenko, E.V., M.A.Schembri, E.Trintchina, K.Kjaergaard, D.L.Hasty, and P.Klemm. 2001. Valency conversion in the type 1 fimbrial adhesin of *Escherichia coli.* Mol. Microbiol. 41:675-686.

Soloaga, A., M.P.Veiga, L.M.Garcia-Segura, H.Ostolaza, R.Brasseur, and F.M.Goni. 1999. Insertion of *Escherichia co*li alpha-haemolysin in lipid bilayers as a non-transmembrane integral protein: prediction and experiment. Mol. Microbiol. 31:1013-1024.

Southgate, J., P.Harnden, P.J.Selby, D.F.Thomas, and L.K.Trejdosiewicz. 1999. Urothelial tissue regulation. Unraveling the role of the stroma. Adv. Exp. Med. Biol. 462:19-30.

Souza, V., M.Rocha, A.Valera, and L.E.Eguiarte. 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. Appl. Environ. Microbiol. 65:3373-3385.

Sprenger, G. A., and J. W. Lengeler. 1998. Analysis of sucrose catabolism in *Klebsiella pneumoniae* and Scr+ derivates of *Escherichia coli* K-12. J Gen Microbiol. 134:1635–1644.

Srinivasan, U., B.Foxman, and C.F.Marrs. 2003. Identification of a gene encoding heat-resistant agglutinin in *Escherichia coli* as a putative virulence factor in urinary tract infection. J. Clin. Microbiol. 41:285-289.

Staiger, D. 2002. Chemical strategies for iron acquisition in plants. Angew. Chem. Int Ed Engl. 41:2259-2264.

Stamm, W.E. 1982. Recent developments in the diagnosis and treatment of urinary tract infections. West J Med 137:213-220.

Stojanovic, M., A.Inic-Kanada, Z.Popovic, I.Zivkovic, and L.Dimitrijevic. 2004. Changes in pools of autoantibodies and anti-bacterial antibodies in patients suffering from recurrent infections of the urinary tract and undergoing bacterial immunization treatment. Immunol. Lett. 94:123-133.

Stoughton, R.B. 2005. Applications of DNA Microarrays in Biology. Annu Rev Biochem.74:53-82

Strausberg, R.L., E.A.Feingold, L.H.Grouse, J.G.Derge, R.D.Klausner, F.S.Collins, L.Wagner, C.M.Shenmen, G.D.Schuler, S.F.Altschul, B.Zeeberg, K.H.Buetow, C.F.Schaefer, N.K.Bhat, R.F.Hopkins, H.Jordan, T.Moore, S.I.Max, J.Wang, F.Hsieh, L.Diatchenko, K.Marusina, A.A.Farmer, G.M.Rubin, L.Hong, M.Stapleton, M.B.Soares, M.F.Bonaldo, T.L.Casavant, T.E.Scheetz, M.J.Brownstein, T.B.Usdin, S.Toshiyuki, P.Carninci, C.Prange, S.S.Raha, N.A.Loquellano, G.J.Peters, R.D.Abramson, S.J.Mullahy, S.A.Bosak, P.J.McEwan, K.J.McKernan, J.A.Malek, P.H.Gunaratne, S.Richards, K.C.Worley, S.Hale, A.M.Garcia, L.J.Gay, S.W.Hulyk, D.K.Villalon, D.M.Muzny, E.J.Sodergren, X.Lu, R.A.Gibbs, J.Fahey, E.Helton, M.Ketteman, A.Madan, S.Rodrigues, A.Sanchez, M.Whiting, A.Madan, A.C.Young, Y.Shevchenko, G.G.Bouffard, R.W.Blakesley, J.W.Touchman, E.D.Green, M.C.Dickson, A.C.Rodriguez, J.Grimwood, J.Schmutz, R.M.Myers, Y.S.Butterfield, M.I.Krzywinski, U.Skalska, D.E.Smailus, A.Schnerch, J.E.Schein, S.J.Jones, and M.A.Marra. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc. Natl. Acad. Sci. U. S. A 99:16899-16903.

Suerbaum, S., S.Friedrich, H.Leying, and W.Opferkuch. 1994. Expression of capsular polysaccharide determines serum resistance in *Escherichia coli* K92. Zentralbl. Bakteriol. 281:146-157.

Svanborg, C., G.Bergsten, H.Fischer, B.Frendeus, G.Godaly, E.Gustafsson, L.Hang, M.Hedlund, D.Karpman, A.C.Lundstedt, M.Samuelsson, P.Samuelsson, M.Svensson, and B.Wullt. 2001. The 'innate' host response protects and damages the infected urinary tract. Ann. Med. 33:563-570.

Svanborg, C., B.Frendeus, G.Godaly, L.Hang, M.Hedlund, and C.Wachtler. 2001. Toll-like receptor signaling and chemokine receptor expression influence the severity of urinary tract infection. J. Infect. Dis. 183 Suppl 1:S61-S65.

Takeda, K., and S.Akira. 2004. Microbial recognition by Toll-like receptors. J Dermatol. Sci. 34:73-82.

Takeda, K., and S.Akira. 2005. Toll-like receptors in innate immunity. Int Immunol. 17:1-14.

Tambyah, P.A., V.Knasinski, and D.G.Maki. 2002. The direct costs of nosocomial catheter-associated urinary tract infection in the era of managed care. Infect Control Hosp. Epidemiol. 23:27-31.

Tarr, P.I., S.S.Bilge, J.C.Vary, Jr., S.Jelacic, R.L.Habeeb, T.R.Ward, M.R.Baylor, and T.E.Besser. 2000. Iha: a novel *Escherichia coli* O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure. Infect. Immun. 68:1400-1407.

Tenorio, E., T.Saeki, K.Fujita, M.Kitakawa, T.Baba, H.Mori, and K.Isono. 2003. Systematic characterization of *Escherichia coli* genes/ORFs affecting biofilm formation. FEMS Microbiol. Lett. 225:107-114.

Tivendale, K.A., J.L.Allen, C.A.Ginns, B.S.Crabb, and G.F.Browning. 2004. Association of *iss* and *iucA*, but not *tsh*, with plasmid-mediated virulence of avian pathogenic *Escherichia coli*. Infect. Immun. 72:6554-6560.

Torres, A.G., and S.M.Payne. 1997. Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7. Mol. Microbiol. 23:825-833.

Torres, A.G., P.Redford, R.A.Welch, and S.M.Payne. 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.

Turner, S.A., S.N.Luck, H.Sakellaris, K.Rajakumar, and B.Adler. 2003. Molecular epidemiology of the SRL pathogenicity island. Antimicrob. Agents Chemother. 47:727-734.

Uhlen, P., A.Laestadius, T.Jahnukainen, T.Soderblom, F.Backhed, G.Celsi, H.Brismar, S.Normark, A.Aperia, and A.Richter-Dahlfors. 2000. Alpha-haemolysin of uropathogenic *E. coli* induces Ca^{2+} oscillations in renal epithelial cells. Nature 405:694-697.

Vaisanen, V., J.Elo, L.G.Tallgren, A.Siitonen, P.H.Makela, C.Svanborg-Eden, G.Kallenius, S.B.Svenson, H.Hultberg, and T.Korhonen. 1981. Mannose-resistant haemagglutination and P antigen recognition are characteristic of *Escherichia coli* causing primary pyelonephritis. Lancet 2:1366-1369.

Vaisanen-Rhen, V., J.Elo, E.Vaisanen, A.Siitonen, I.Orskov, F.Orskov, S.B.Svenson, P.H.Makela, and T.K.Korhonen. 1984. P-fimbriated clones among uropathogenic *Escherichia coli* strains. Infect. Immun. 43:149-155.

van Ijperen, C., P.Kuhnert, J.Frey, and J.P.Clewley. 2002. Virulence typing of *Escherichia coli* using microarrays. Mol. Cell Probes 16:371-378.

Vandemaele, F.J., S.M.Hensen, and B.M.Goddeeris. 2004. Conservation of deduced amino acid sequence of FimH among *Escherichia coli* of bovine, porcine and avian disease origin. Vet. Microbiol. 101:147-152.

Wang L, D. Rothemund, H. Curd, and P.R. Reeves. 2003. Related A. Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. J Bacteriol. 185:2936-2943.

Verweij-van Vught, A.M., F.Namavar, M.Sparrius, W.A.Vel, and D.M.MacLaren. 1985. Pathogenic synergy between *Escherichia coli* and *Bacteroides fragilis*: studies in an experimental mouse model. J. Med. Microbiol. 19:325-331.

Wang, H.C., J.Badger, P.Kearney, and M.Li. 2001. Analysis of codon usage patterns of bacterial genomes using the self-organizing map. Mol. Biol. Evol. 18:792-800.

Wang, R.F., M.L.Beggs, B.D.Erickson, and C.E.Cerniglia. 2004. DNA microarray analysis of predominant human intestinal bacteria in fecal samples. Mol. Cell Probes 18:223-234.

Watt, S., P.Lanotte, L.Mereghetti, M.Moulin-Schouleur, B.Picard, and R.Quentin. 2003. *Escherichia coli* strains from pregnant women and neonates: intraspecies genetic distribution and prevalence of virulence factors. J. Clin. Microbiol. 41:1929-1935.

Wei, J., M.B.Goldberg, V.Burland, M.M.Venkatesan, W.Deng, G.Fournier, G.F.Mayhew, G.Plunkett, III, D.J.Rose, A.Darling, B.Mau, N.T.Perna, S.M.Payne, L.J.Runyen-Janecky, S.Zhou, D.C.Schwartz, and F.R.Blattner. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. Infect. Immun. 71:2775-2786.

Welch, R.A., V.Burland, G.Plunkett, III, P.Redford, P.Roesch, D.Rasko, E.L.Buckles, S.R.Liou, A.Boutin, J.Hackett, D.Stroud, G.F.Mayhew, D.J.Rose, S.Zhou, D.C.Schwartz, N.T.Perna, H.L.Mobley, M.S.Donnenberg, and F.R.Blattner. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A 99:17020-17024.

Welch, R.A., and S.Pellett. 1988. Transcriptional organization of the *Escherichia coli* hemolysin genes. J. Bacteriol. 170:1622-1630.

Welcher, A.A., A.R.Torres, and D.C.Ward. 1986. Selective enrichment of specific DNA, cDNA and RNA sequences using biotinylated probes, avidin and copper-chelate agarose. Nucleic Acids Res. 14:10027-10044.

Whitfield, C., and I.S.Roberts. 1999. Structure, assembly and regulation of expression of capsules in *Escherichia coli.* Mol. Microbiol. 31:1307-1319.

Wiedow, O., and U.Meyer-Hoffert. 2005. Neutrophil serine proteases: potential key regulators of cell signalling during inflammation. J. Intern. Med. 257:319-328.

Williams, D.H., and A.J.Schaeffer. 2004. Current concepts in urinary tract infections. Minerva Urol. Nefrol. 56:15-31.

Wohlhieter, J.A., J.R.Lazere, N.R.Snellings, E.M.Johnson, R.M.Synenki, and L.S.Baron. 1975. Characterization of transmissible genetic elements from sucrose-fermenting *Salmonella* strains. J Bacteriol. 122:401–406.

Wright, K.J., P.C.Seed, and S.J. Hultgren. 2005. Uropathogenic *Escherichia coli* flagella aid in efficient urinary tract colonization. Infect Immun. 73:7657-7668.

Wu, L., D.K.Thompson, X.Liu, M.W.Fields, C.E.Bagwell, J.M.Tiedje, and J.Zhou. 2004. Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications. Environ. Sci. Technol. 38:6775-6782.

Yamamoto, K., K.Tominaga, M.Sukedai, T.Okinaga, K.Iwanaga, T.Nishihara, and J.Fukuda. 2004. Delivery of cytolethal distending toxin B induces cell cycle arrest and apoptosis in gingival squamous cell carcinoma *in vitro*. Eur. J. Oral Sci. 112:445-451.

Yamamoto, S., M.Nakano, A.Terai, K.Yuri, K.Nakata, G.B.Nair, H.Kurazono, and O.Ogawa. 2001. The presence of the virulence island containing the *usp* gene in uropathogenic *Escherichia coli* is associated with urinary tract infection in an experimental mouse model. J. Urol. 165:1347-1351.

Yu, X., M.Susa, C.Knabbe, R.D.Schmid, and T.T.Bachmann. 2004. Development and validation of a diagnostic DNA microarray to detect quinolone-resistant *Escherichia coli* among clinical isolates. J. Clin. Microbiol. 42:4083-4091.

Zarrinkar, P.P., J.K.Mainquist, M.Zamora, D.Stern, J.B.Welsh, L.M.Sapinoso, G.M.Hampton, and D.J.Lockhart. 2001. Arrays of arrays for high-throughput gene expression profiling. Genome Res. 11:1256-1261.

Zhang, D., G.Zhang, M.S.Hayden, M.B.Greenblatt, C.Bussey, R.A.Flavell, S.Gosh, M.Leist, P.Ghezzi, G.Grasso, R.Bianchi, P.Villa, M.Fratelli, C.Savino, M.Bianchi, J.Nielsen, J.Gerwien, P.Kallunki, A.Larsen, L.Helboe, S.Christensen, L.Pedersen, M.Nielsen, L.Torup, T.H.Sager, A.Sfacteria, S.Erbayaktar, Z.Erbayaktar, N.Gokmen, O.Yilmaz, C.Cerami-Hand, Q.Xie, T.H.Coleman, A.Cerami, M.Brines, V.E.Torres, X.Wang, Q.Quian, S.T.Somlo, P.Harris, and V.Gattone. 2005. A Novel Molecule Designed to Keep Bacteria out of the Urinary Tract: A Toll-Like Receptor that Prevents Infection by Uropathogenic Bacteria. Science 303: 1522-1526, 2004. J. Am. Soc. Nephrol. 16:4-7.

Zhang, L., and B.Foxman. 2003. Molecular epidemiology of *Escherichia coli* mediated urinary tract infections. Front Biosci. 8:e235-e244.

Zimecki, M., G.Chodaczek, M.Kocieba, and M.L.Kruzel. 2004. Lethality in LPS-induced endotoxemia in C3H/HeCr mice is associated with prevalence of proinflammatory cytokines: lack of protective action of lactoferrin. FEMS Immunol. Med. Microbiol. 42:167-172.

8. APPENDICES: Original publications