Genome sequence of *Escherichia coli* 536: insights into uropathogenicity through comparison with genomes of *Escherichia coli* MG1655, CFT073, and EDL933

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von

Elżbieta Barbara Brzuszkiewicz

aus Warschau, Polen

Göttingen, 2005

D 7

Referent:	**Prof. Dr. G. Gottschalk**
Korreferent:	**Prof. Dr. W. Liebl**
Tag der mündlichen Prüfung:	30.06.2005

Data is not information.

Information is not knowledge.

Knowledge is not understanding.

Understanding is not wisdom.

Cliff Stoll and Gary Schubert

TABLE OF CONTEST

LIST OF ABBREVIATIONS

°C	degree Celsius
µ	micro (10^{-6})
µg	microgram
µl	microlitre
µm	micrometer
µM	micromolar
A	adenine
A	ampere
aa	amino acid
ABI	Applied Biosystem Instrument
Ala	alanine
Amp	ampicillin
AP	alkaline phosphatase
APS	ammonium persulphate
Asp	aspartate
Asx	asparagine
ATP	adenosine triphosphate
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
bp	base pair
BrET	ethidium bromide
C	cytosine
ca.	circa
cDNA	complementary desoxyribonucleic acid
cfu	colony forming units
CIP	calf intestine phosphatase
cm	centimetre
co	company
Da	Dalton
dH_2O	distilled Water

DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide phosphate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
et al.	et alei (and others)
Fig.	figure
g	gram
G	guanine
Gln	glutamine
Glu	glutamate
Gly	glycine
h	hour(s)
HPLC	High Performance Liquid Chromatography
IPTG	Isopropyl-s-thiogalactopyranoside
Iva	isovaline
kb	kilo base
kDa	kilo Dalton
l	liter
LB	Luria Broth
Leu	leucine
LPS	lipopolysaccharides
Lys	lysine
M	molar (moles per litre)
Mbp	mega base
MCS	Multicloning
Met	methionine
mg	milligram
min	minute(s)
ml	millilitre
mM	millimolar
mRNA	messenger RNA
MW	molecular weight
n	nano (10^{-9})

NaAc	sodium acetate
NCBI	National Center for Biotechnology Information
ng	nanogram
nl	nanolitre
nt	nucleotide
OD	optical density
ON	over night
ORF	open reading frame
Orn	ornithine
p	pico (10^{-12})
PCR	Polymerase chain reaction
pH	negative logarithm of hydrogen ion concentration
pmol	picomole
Phe	phenylalanine
Pheol	phenylalaninol
Pro	praline
psi	pounds per square inch
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolution per minute
RT	room temperature
SDS	sodium dodecyl sulphate
sec	second(s)
sp	species (singular)
spp	species (plural)
Taq	*Thermus aquaticus*
TE	Tris EDTA
TEMED	N,N,N,N,-Tetramethy –Ethylenediamine
TRIS	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Trp	tryptophan
U	unit
UPEC	uropathogenic *Escherichia coli*
UTI	urinary tract infection

UV	ultraviolet
V	potential in volts
Val	valine
VF	virulence factor(s)
v/v	volume per volume
w/v	weight per volume
w/w	weight per weight
wt	wild type

1. INTRODUCTION

Microbial genome sequencing and analysis is a rapidly expanding and increasingly important strand of microbiology. Until whole genome analysis become available, life sciences have been based on a reductionist principle – dissecting cell and systems into fundamental components for further study. Recent studies on whole genomes and whole genome comparisons in particular give us a possibility to receive more information leading to better understanding of the general feature of pathogenic and nonpathogenic lifestyle.

The urinary tract is among the most common sites of both community-acquired and nosocomial infections, and *Escherichia coli* is the most common etiologic agent of these infections. The ability to inhabit the different niches during the ascending urinary tract infections (UTIs) and cause particular pathologies at each site resides largely in the island genes specific to uropathogenic *E. coli*. The genome sequence has revealed many possible factors that may contribute to colonisation of the urinary tract tissues and the disease. Analysis of whole genome of uropathogenic *E. coli* 536 strain provides insight into pathogenecity of *E. coli*.

E. coli, a venerable "workhorse" for biochemical and genetic studies, and for the large-scale production of recombinant proteins, is one of the most intensively studied organisms.

E. coli is the head of the large bacterial family, *Enterobacteriaceae*, the enteric bacteria, which are Gram-negative rods that live in the intestinal tract of animals. The *Enterobacteriaceae* are among the most important bacteria medically. A number of genera within the family are human intestinal pathogens (e.g. *Salmonella*, *Shigella*, *Yersinia*). Several others are normal colonists of the human gastrointestinal tract (GI) (e.g. *Escherichia*, *Enterobacter*, *Klebsiella*), but these bacteria, as well, may occasionally be associated with disease.

1.1. PATHOGENESIS OF *E. COLI*

Macroorganisms live intimately together with microorganisms throughout their entire live, and to their mutual benefit in the overwhelming number of cases. Infection is just an interruptive aberration of the balance. Nevertheless, the experience can be anything from mild to extremely dramatic or even fatal. Bacteria express various virulence factors such as: toxins, adhesins, factors promoting endocytosis or preventing phagocytosis and factors aiding to overcome mechanisms of host defence.
The natural habitat of *E. coli* is the gastrointestinal tract of warm-blooded animals and the human body, where it is the most common facultative anaerobe in the gut. Although most strains exist as harmless symbionts, there are many pathogenic E. *coli* strains that are capable of causing a variety of diseases in animals and humans. In addition, from an evolutionary perspective, strains of the genus *Shigella* are so closely related phylogenetically that they should also be included in the group of organisms recognized as *E. coli.* [1,2]

Pathogenic *E. coli* strains differ from those that predominate in the feces of healthy individuals in that they are more likely to express virulence factors (molecules directly involved in pathogenic processes but ancillary to normal metabolic functions). Known virulence determinants of pathogenic *E. coli* are summarised in Table 1.
In addition to their role in disease, these virulence factors presumably enable the pathogenic strains to explore niches unavailable to commensal strains, and thus to spread and persist in the bacterial community. Yet, for all the differences among pathogenic and non-pathogenic *E. coli*, they all evolved from a common ancestor with *Salmonella enterica* some 100 million years ago.

E. coli is responsible for three types of infections in humans: intestinal diseases (gastroenteritis), neonatal meningitis and urinary tract infections (UTI). These three diseases depend on a specific array of pathogenic (virulence) determinants.

TABLE 1. SUMMARY OF KNOWN VIRULENCE DETERMINANTS OF PATHOGENIC *E. COLI*	
ADHESINS	CFAI/CFAII, Type 1 fimbriae, P fimbriae, S fimbriae, intimin (non-fimbrial adhesin)
INVASINS, FITNESS	Hemolysisn, siderophores and siderophore uptake systems, Shigella-like "invasins" for intracellular invasion and spread
MOTILITY/CHEMOTAXIS	Flagella
TOXINS	LT toxin, ST toxin, Shiga-like toxin, cytotoxins, endotoxin LPS
ANTIPHAGOCYTIC SURFACE PROPERTIES	Capsules, K antigens, LPS
DEFENCE AGAINST SERUM BACTERICIDAL REACTIONS	LPS, K antigens
DEFENCE AGAINST IMMUNE RESPONSES	Capsules, K antigens, LPS, antigenic variation
GENETIC ATTRIBUTES	Genetic exchange by transduction and conjugation, transmissible plasmids, R factors and drug resistance plasmids, toxin and other virulence plasmids

1.1.1. INTESTINAL DISEASES CAUSED BY *E. COLI*

Five classes of *E. coli* that cause diarrheal diseases are now recognised: enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enterohemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), and enteroaggregative *E. coli* (EAggEC). Each class falls within a serological subgroup and manifests distinct features in pathogenesis. A short characterization of each group is presented in Table 2.

TABLE 2. SUMMARY OF VIRULENCE CHARACTERISTICS OF INTESTINAL PATHOGENS					
	ETEC	**EIEC**	**EPEC**	**EAGGEC**	**EHEC**
ADHESIN	Fimbrial adhesins	Nonfimbrial adhesins, possibly OMP	Nonfimbrial adhesin (intimin)	Adhesins not characterized	Adhesins not characterized, probably fimbriae
INVASIN	Non invasive	Invasive (penetrates and multiplies within epithelial cells)	Moderately invasive	Non invasive	Moderately invasive
TOXINS	Produce LT and/or ST toxin	Does not produce shiga toxin	Does not produce LT or ST; some reports of shiga-like toxin	Produces ST-like toxin and hemolysin	Does not produce LT or ST but does produce shiga toxin
SYMPTOMS	Watery diarrhea in infants and travellers; no inflammation, no fever	Dysentery-like diarrhea (mucous, blood); severe inflammation, fever	Usually infantile diarrhea, watery diarrhea similar to ETEC; some inflammation, no fever	Persistent diarrhea in young chilidren; without inflammation, no fever	Pediatric diarrhea, copious bloody discharge (hemorrhagic colitis), intense inflammatory response, may be complicated by hemolytic uremia

1.1.2. EXTRAINTESTINAL DISEASES CAUSED BY *E. COLI*

1.1.2.1. NEONATAL MENINGITIS

Neonatal meningitis affects 1,000-4,000 infants per year in the United States. *E. coli* strains invade the blood stream of infants from the nasopharynx or GI tract and are carried to the meninges. The K-1 antigen is considered the major determinant of virulence among strains of *E. coli* that cause neonatal meningitis. K-1 is a homopolymer of sialic acid. It inhibits phagocytosis, and responses from the host's immunological mechanisms. K-1 may not be the only determinant of virulence as siderophore production and endotoxin are also likely to be involved.

1.1.2.2. URINARY TRACT INFECTION

1.1.2.2.1. ETIOLOGY

UTI is defined as a significant bacteraemia in accompanied by symptoms. The bacteria most often seen in UTIs are of fecal origin. These organisms are a subset of the organisms found in the feces. Strict anaerobic bacteria rarely cause UTIs. More than 90% of acute UTIs in patients with normal anatomic structure and function are caused by certain strains of *E. coli*, these are currently defined as uropathogenic *Escherichia coli* (UPEC).

This is especially true of spontaneous UTI in females (cystitis and pyelonephritis). Other strains are less common, including *Proteus mirabilis* and more rarely Gram-positive microbes. Among the latter, *Staphylococcus saprophyticus* deserves special mention, as this Gram-positive pathogen is responsible for 5% to 15% of such primary infections, is not detected by the leukocyte esterase dipstick, and is resistant to antimicrobial agents that are active on Gram-negative rods. [3, 17]

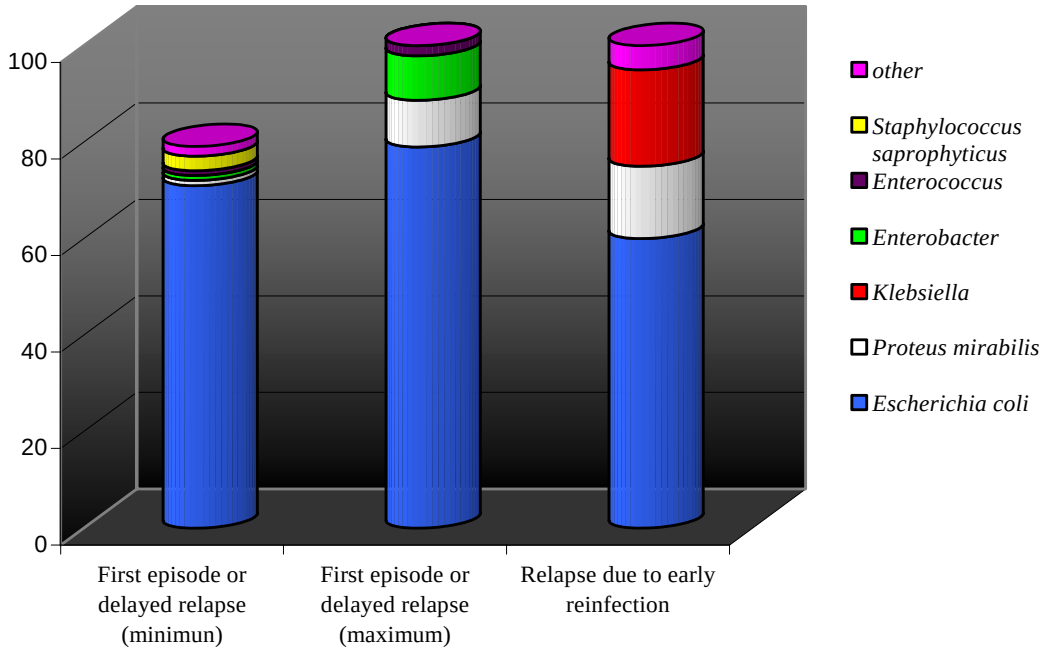


Figure 1. Main pathogens responsible for urinary tract infections (UTIs)

1.1.2.2.2. EPIDEMIOLOGY

UTI is an extremely common medical problem, with an unpredictable natural history. Many infections resolve spontaneously, but others can progress to destroy the kidney, and in case of Gram-negative sepsis, the entire host. Intelligent management of UTI requires a thorough knowledge of the pathophysiology involved and a medical care, which is tailored to the clinical setting.

The incidence of UTI's ranks second after respiratory infections in the United States of America and Germany. Each year, urinary tract infections (UTIs) account for more than 8 million physician visits in the United States (mostly for cystitis) and are estimated to cause more than 100,000 hospital admissions per year in this country (mostly for acute pyelonephritis) [4]. UTI is a common diagnosis among patients evaluated in the emergency department as well. [5] The majority of the cases seen in the doctor's office are women (30:1, female: male ratio). 40% of all women have at least one episode of a UTI at some time in their lives. Up to 20% of young women with acute cystitis develop recurrent UTIs.

Males experience a rapid increase in the incidence UTIs sometime in their 40s. This is about the time that males are experiencing prostate gland hypertrophy. The direct cost of treating UTIs in the United States is estimated at $ 1 billion per year or greater. [6]

1.1.2.2.3. PATHOGENESIS OF INFECTION

The urinary system is structured in a way that prevents bacterial infection. The ureters and bladder normally prevent urine from flowing backwards towards the kidneys, and the flow of urine from the bladder washes bacteria out of the body. Apart from the lowermost part of the urethra (tube that carries urine from the bladder), which is usually colonized with various species of bacteria, the urinary tact is normally a sterile system that is protected from the nearby colonic microflora by non-specific defence including the epithelial barrier, the antibacterial properties of the bladder mucosa and the flow of urine. The low pH and osmolarity of urine is inhibitory to bacterial growth and the salts, urea and organic acids present in urine can reduce bacterial survival within urinary tract. Despite these natural safeguards, symptomatic urinary tract infections (UTIs) are among the most frequent infections in both the general community and in hospital.

The colonizing bacteria derived from the feces or perineal region ascend the urinary tract to the bladder. Depending on the location of infecting bacteria, distinct diseases can be defined, such as cystitis (infection of the bladder) or (pyelo-)nephritis, when the kidneys are involved (Figure 2.). Table 3 outlines individual factors increasing the risk of getting UTIs.

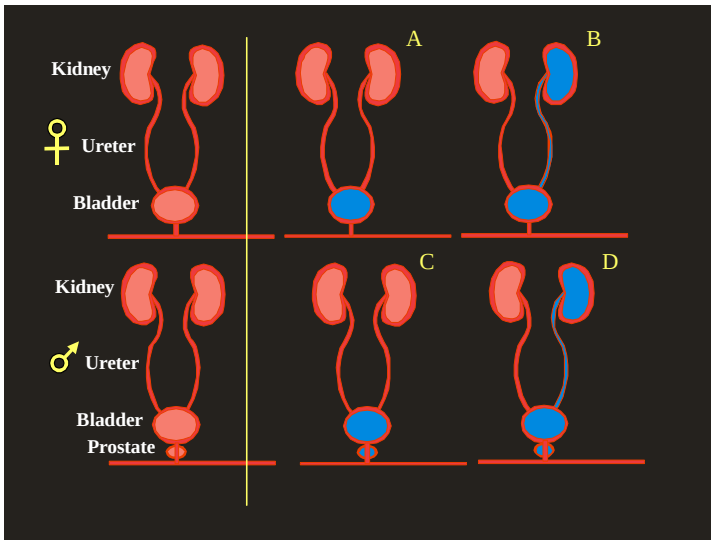


Figure 2. Classification of Urinary Tract Infections. Blue colour in this figure symbolises the presence of bacteria and leukocytes (*i.e.*, infection) in the relevant organ. A. cystitis, B. pyelonephritis in a female patients; C. prostatitis, D. pyelonephritis in males.

TABLE 3. RISK FACTORS

- ANY ABNORMALITY OF THE URINARY TRACT THAT OBSTRUCTS OR SLOWS THE FLOW OF URINE MAKES IT EASIER FOR BACTERIA TO GROW
- SHORT URETHRA
- A STONE IN THE KIDNEY OR ANY PART OF THE URINARY TRACT CAN FORM SUCH A BLOCKAGE, CREATING THE CONDITIONS FOR A UTI
- IN MEN, AN ENLARGED PROSTATE GLAND CAN OBSTRUCT URINE FLOW AND MAKE INFECTION DIFFICULT TO TREAT
- ONE OF THE MOST COMMON SOURCES OF INFECTION IS CATHETERS, OR TUBES, PLACED IN THE BLADDER
- PEOPLE WHO HAVE DIABETES MELLITUS
- IMMUNOSUPPRESSED PATIENTS UTIS OCCUR IN A SMALL PERCENTAGE OF INFANTS DUE TO CONGENITAL ABNORMALITIES THAT SOMETIMES REQUIRE SURGERY
- FOR MANY WOMEN, SEXUAL INTERCOURSE SEEMS TO PRECIPITATE UTIS
- WOMEN WHO USE THE DIAPHRAGM AND/OR SPERMICIDES ARE MORE LIKELY TO DEVELOP A UTI THAN WOMEN WHO USE OTHER FORMS OF CONTRACEPTION
- PATIENTS WITH A NEUROGENIC BLADDER OR BLADDER DIVERTICULUM
- POSTMENOPAUSAL WOMEN WITH BLADDER OR UTERINE PROLAPSE
- PREGNANT WOMEN ARE MORE SUSCEPTIBLE TO UTIS
- INCREASED UROEPITHELIAL RECEPTOR DENSITY THAT CREATES GENETIC PREDISPOSITION
- DELAYED POSTCOITAL URINATION

1.1.2.3. VIRULENCE FACTORS ASOCIATED WITH UTI

The ability of *E. coli* to cause an illness is influenced by the production of virulence factors (VFs). These include tissue-specific adhesins, toxins, siderophores, secretion systems, etc. VFs allow the colonization of specific host surfaces, avoiding or subverting the host defence system and stimulating a noxious host inflammatory response.

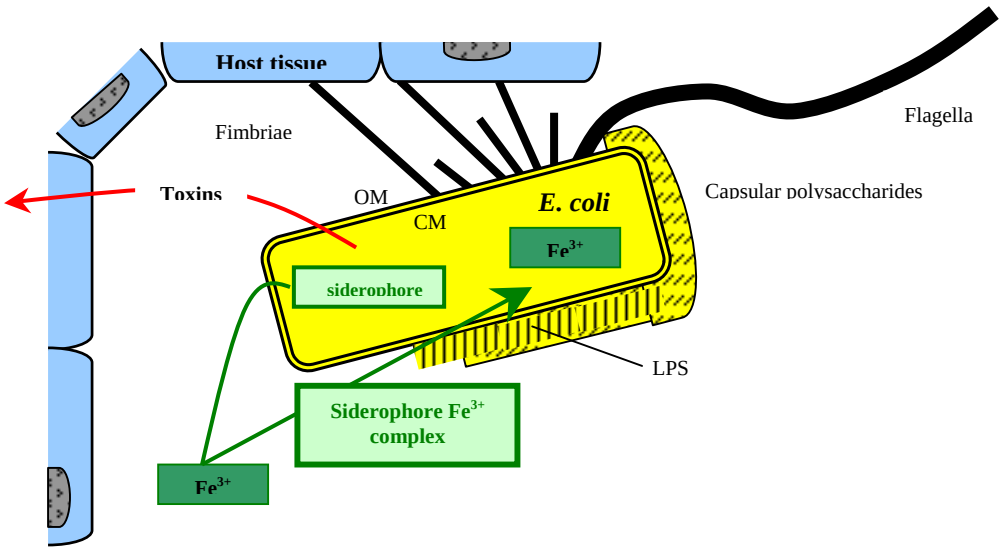


Figure 3. Schematic representation of an *E. coli* cell interacting with host tissue, highlighting features relevant to bacterial pathogenicity OM, Outer membrane; CM, cytoplasmic membrane; LPS, lipopolysaccharide [7, adopted]

1.1.2.3.1. ADHERENCE

UTI-causing *E. coli* have to be able to adhere to the urinary tract epithelium in order to prevent being washed out. They do so by expressing adherence factors. In addition, the physical attachment of bacteria to host cells can also serve as a signal for the activation of genes involved in bacterial virulence via signal transduction. A specific adhesion is associated with the host and tissue tropism (Figure 4.). The adhesive

organelles that are associated with UPEC strains include P and Type 1 fimbriae, S pili and Dr family adhesins.

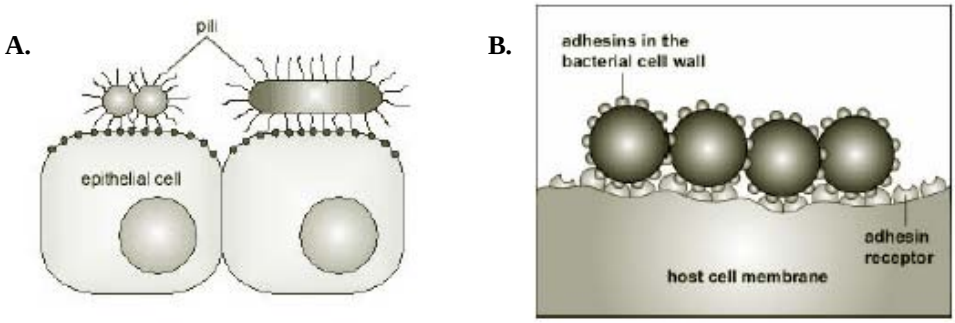


Figure 4. Two types of adhesions: A. pili and B. afimbrial adhesions

The adhesin that has been most closely associated with uropathogenic *E. coli* is the P fimbriae (or pyelonephritis-associated pili [PAP] pili). The designation is derived from the ability of P fimbriae to bind specifically to the P blood group antigen, which contains a D-galactose-D-galactose residue. The fimbriae bind not only to red blood cells but also to a specific galactose dissaccharide that is found on the surface of uroepithelial cells in approximately 99% of the population.

Uropathogenic strains of *E. coli* possess other determinants of virulence in addition to P fimbriae. *E. coli* with P fimbriae possess also the genes for Type 1 fimbriae, and there is strong evidence that P fimbriae are derived from Type 1 fimbriae by insertion of a new fimbrial tip protein to replace the mannose-binding domain of Type 1 fimbriae. In any case, Type 1 fimbriae could provide a supplementary mechanism of adherence or play a role in aggregating the bacteria to a specific mannosyl-glycoprotein that occurs in urine. This type of fimbriae consist of a helical rod made up of repeating FimA subunits and a distal tip containing two adaptor proteins (FimF and FimG) and the adhesin FimH. The interaction between FimH and bladder epithelial cells induces host-signalling cascades that lead to cytoskeletal rearrangements and the envelopment of the attached bacterium. Invasion of bladder epithelial cells triggers the production of proinflammatory cytokines, chemokines and epithelial cells apoptosis, leading to the recruitment of inflammatory cells and exfoliation, respectively. To prevent bacterial clearance by exfoliation of superficial bladder epithetical cells UPEC may invade the underlying bladder epithelial cells, potentially facilitating long-term persistence of bacteria within the bladder mucosa.

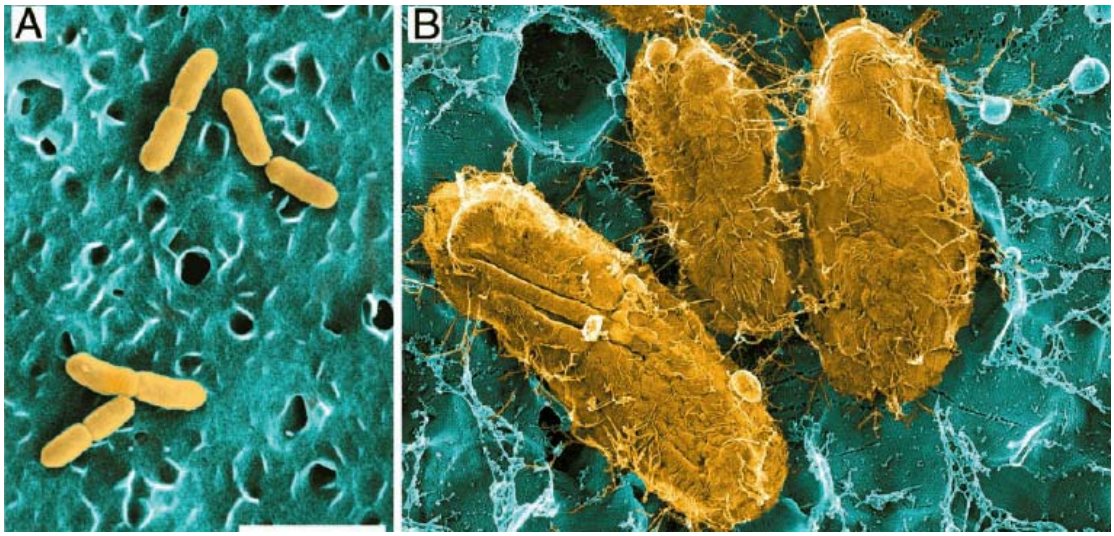


Figure 5. Type 1 pilus-mediated bacterial attachment to the bladder epithelium. After inoculation of C57BLy6 mice with type 1-piliated UPEC, numerous bacteria (yellow) can be found attached to the luminal surface of the bladder (blue) as detected by scanning EM (A) and high-resolution freeze-dryydeep-etch EM (B) [8]

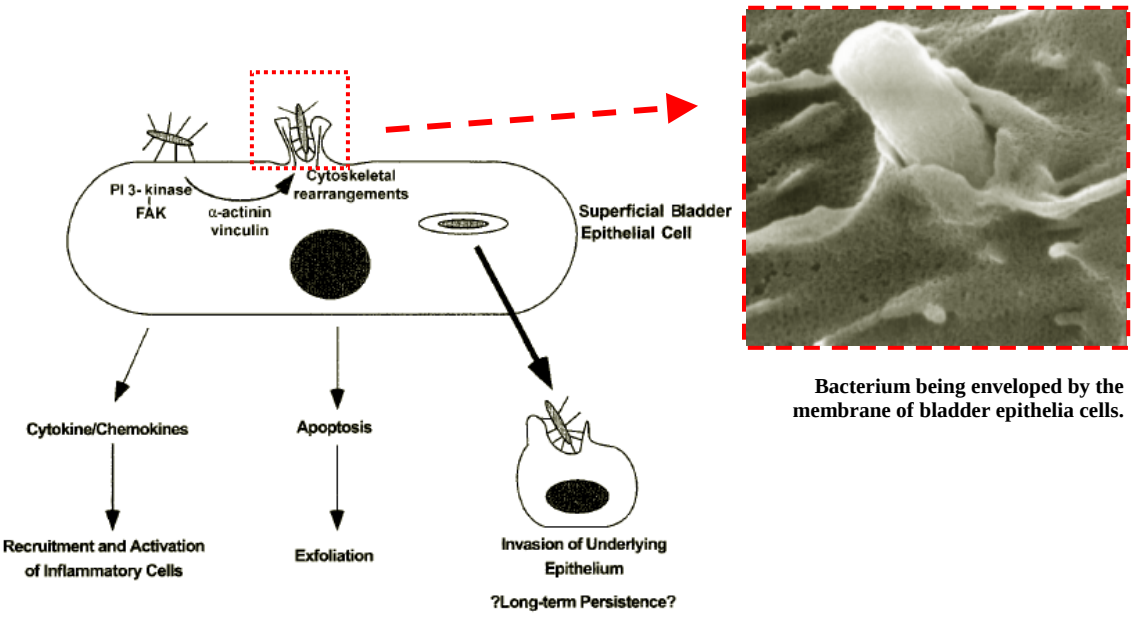


Figure 6. Type 1 fimbriae mediated invasion [9]

S fimbriae are designated according their specific binding to terminal sialyl-galactoside residues. Expression of S fimbriae can be modified by phase variation. Binding sites for S fimbriae are found on epithelial cells of the proximal and distal tubes, collecting ducts, and glomerulus, in the renal interstitium and on renal vascular endothelium. Recognize sialosyloligosaccharide residues on host cells and may help promote colonization of the upper urinary tract by UPEC.

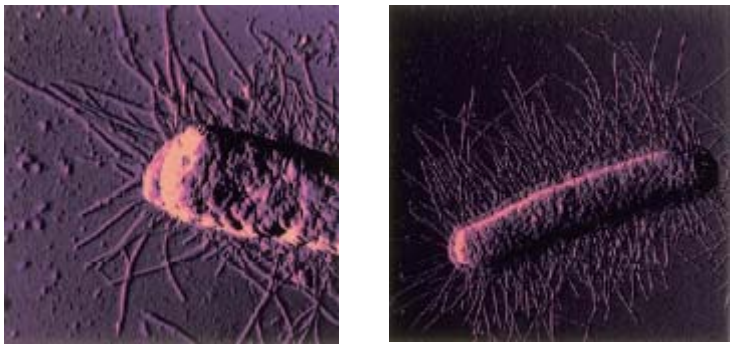


Figure 7. S-fimbriated bacteria, P-fimbriated bacteria [11]

Dr family adhesins, including Dr fimbriae and the afimbrial adhesins AFA-I and AFA-II, bind to the Dra blood group antigen present on decay accelerating factors and may also facilitate ascending colonization of the urinary tract.

1.1.2.3.2. CAPSULES AND SURFACE COMPONENTS

The cell surface of *E. coli* is a complex array of proteins and glycoconjugates. The capsular polysaccharides (CPSs) and the O-polysaccharides of the lipopolysaccharide (LPS) molecules are the major surface polysaccharides expressed.

These polymers are serotype-specific and give rise to the K- and O-antigens, respectively. Different K and O serotypes result from variation in sugar composition

and linkage specificity in the polysaccharides, as well as their substitution with non-carbohydrate residues. There are 167 different O-serogroups and more than 80 polysaccharide K antigens in *E. coli*. Capsular polysaccharides are linear polymers of repeating carbohydrate subunits that sometimes include a prominent amino acid or lipid component. They coat the cell, interfere with O-antigen detection and protect the cell from host defence mechanisms. The capsules are mostly thin, patchy, acidic, thermostable, and highly anionic.

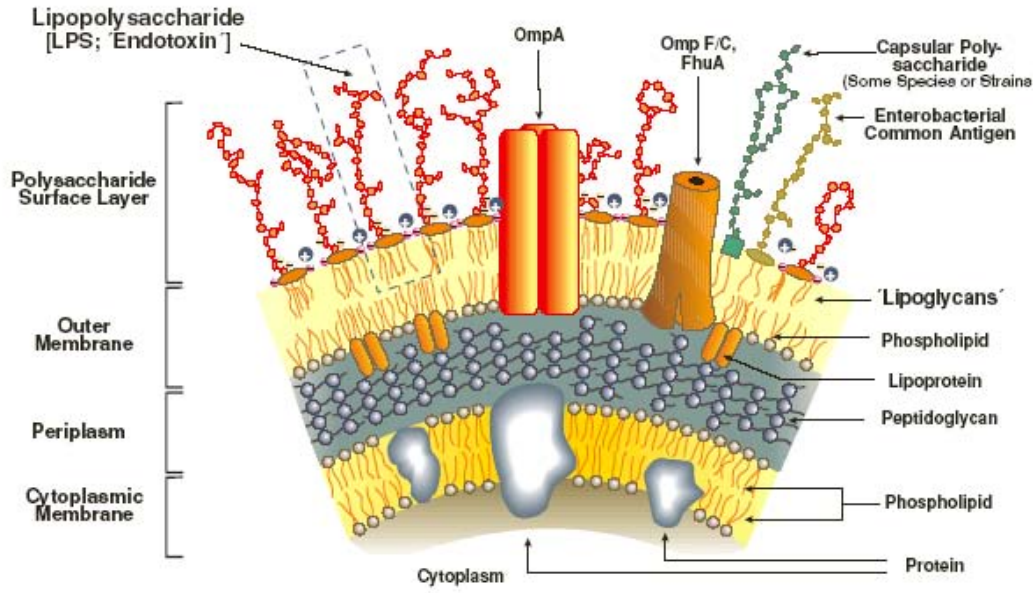


Figure 8. Bacteria membrane - diagram [12]

The K1 polysaccharide is a homopolymer of sialic acid units linked C2-C8 and randomly acetylated at C7 and C9. The presence of K antigens is associated with upper urinary tract infections, and antibody to the K antigen has been shown to afford some degree of protection in experimental infections. Regardless of their chemistry, these capsules may be able to promote bacterial virulence by decreasing the ability of antibodies and/or complement to bind to the bacterial surface, and by preventing phagocytes to recognise and engulf the bacterial cells. The best studied K antigen, K-1, repetition has besides being antiphagocytic, the additional property of being an antigenic disguise.

1.1.2.3.3. BIOFILM

Uropathogenic strains of *E. coli* have got the ability to adhere to epithelial cells and they form a biofilm. [13.14] They secrete a slime matrix, which surrounds and protects them (Table 3.)

TABLE 3. SOME ADVANTAGES OF BIOFILM FORMATION

Protection from the environment: macrophages, antibodies do not penetrate readily
Facilitiates nutrient sharing, metabolic cooperativity, syntrophisms
Horizontal gene transfer favoured: possibility of acquiring new genetic traits
Sessile cells are more resistant to antibiotics than are planctonic cells - very difficult to eradicate biofilmic infections

Host defence mechanisms and antibiotic therapy kill most of the bacteria during infection, though some can remain within the bladder tissue. But a bacterial biofilm persists in the urinary tract and on catheter surfaces because microorganisms, which are organized in a biofilm, are strongly resistant to host defence and antibiotic treatment. [15,16,17] Biofilms facilitate the survival of bacteria because the bacteria are protected from the urine flow and the antimicrobial attacks. However, the mechanisms involved in bacterial persistence within the urinary tract are still unclear. Gregory G. Anderson and colleagues showed that *E. coli* forms intracellular bacterial biofilm-like pods in bladder, favouring the recurrence of UTIs. [18] They showed, using the scanning electron microscopy, that *E. coli* invaded the bladder superficial cells, and these intracellular bacteria matured into biofilms, creating pod-like bulges on the luminer surface of the bladder. The pods contained bacteria encased in polysaccharide-rich matrix surrounded by a protective shell of uroplakin.

1.1.2.3.4. ENDOTOXINS (LPS)

Bacterial lipopolisaccharides (LPS) are the major outer surface membrane components consisting of a predominantly lipophilic region, lipid A and a covalently hydrophilic polyoligosaccharide region.

Bacteria are killed by normal human serum through the lytic activity of the complement system. The alternative pathway is activated by bacteria in the absence of a specific antibody and plays a more important role in serum killing than does the classical pathway. Lipid A can activate the classical pathway in the absence of an antibody, but its location deep within the outer membrane probably makes it inaccessible to complement components. Bacterial resistance to serum complement results from the individual or combined effects of capsular polysaccharides, O-polysaccharide side chain, and surface proteins. Figure 9.

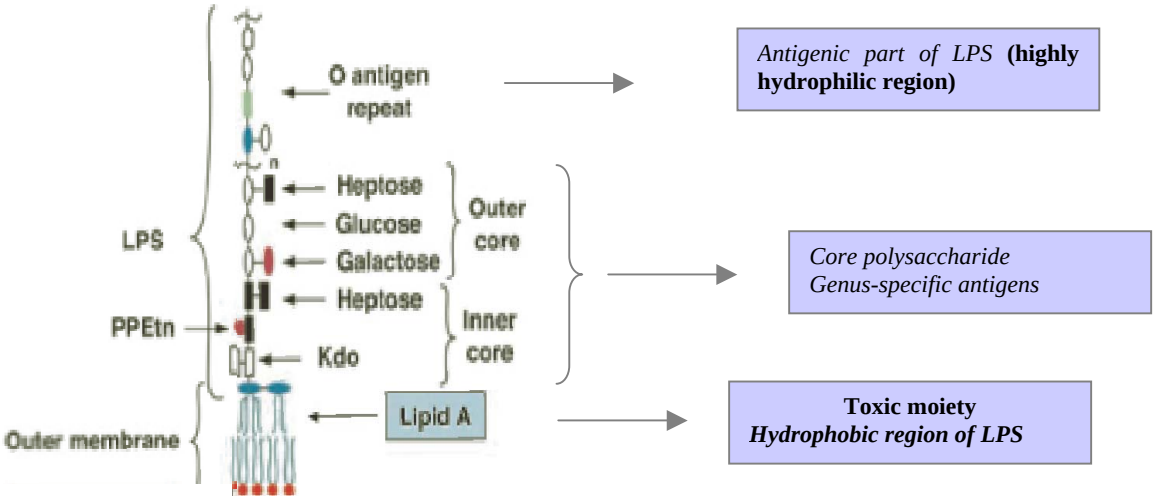


Figure 9. LPS – diagram [19; adopted] Kdo – 2-keto-3-deoxyoctonic acid – unusual sugar, unique to LPS and invariably present in LPS.

1.1.2.3.5. EXOTOXINS

1.1.2.3.5.1. HEMOLYSIN (HLY)

UPEC produces hemolysins that are cytotoxic due to the formation of transmembrane pores in the host cells. They belong to the RTX (repeat-in-toxin) toxin family. A possible strategy for obtaining iron and other nutrients for bacterial growth may involve the lysis of host cells to release these substances. The activity of hemolysins is not limited to red blood cells since the alpha-hemolysins of *E. coli* also lyse lymphocytes, and the beta-hemolysins inhibit phagocytosis and chemotaxis of neutrophils.

1.1.2.3.5.2. CYTOTOXIC NECROTIZING FACTOR (CNF1)

Among the putative virulence factors produced by *E. coli* are the cytotoxic necrotizing factors (CNFs). Cytotoxic necrotizing factor type 1 (CNF1) is chromosomally encoded - 115-kDa toxin that catalyzes the deamidation of the small GTPases RhoA, Rac, and Cdc42. [20] Rippere-Lampe et al. propose that CNF1 production increases the capacity of UPEC strains to resist killing by neutrophils, which in turn permits these bacteria to gain access to a deeper tissue and persist better in the lower urinary tract. [21]

1.1.2.3.5.3. AUTOTRANSPORTER TOXIN (SAT)

UPEC strains secrete specific proteins that may contribute to pathogenesis of UTI. Among these proteins, the most recently described is the secreted autotransporter toxin (Sat). Sat is a large (107kDa) serine protease, which exhibits cytopathic activity on kidney, bladder and other cell lines, as well as eliciting a strong antibody response after experimentally infection. [22] The protein appears to fall within one subgroup of autotransporters recently classified as the SPATE (serine protease autotransporters of Enterobacteriaceae) family. There are eight known SPATE proteins: Sat of uropathogenic *E. coli* [30], Pet of enteroaggregative *E. coli* [23], EspC

of enteropathogenic *E. coli* [24], Pic of enteroaggregative *E. coli* and *Shigella* [25], SigA of *Shigella* [26], SepA of *Shigella* [27], Tsh of avian *E. coli* [28], and EspP of enterohemorrhagic *E. coli* [29]. SPATE autotransporters are identified by the presence of a serine protease active site motif in the passenger domain. It has been shown that the motif is necessary for phenotypic functions (e.g., adhesin, invasin, protease, or cytotoxin) as demonstrated by site-directed mutagenesis with Pet and Tsh [30]. Interestingly, no SPATE proteins have been identified in nonpathogenic organisms.

1.1.2.3.6. IRON ACQUISITION

The host environment is ideal for bacteria in every way except for one: iron is plentiful but tightly bound in haem, ferritin, transferrin, or lactoferin. It can be a limiting factor in growth. Therefore, uropathogenic strains of *E. coli* have evolved virulence factors that mediate the release of the host iron. As an example siderophores play an essential role in iron acquisition for the bacteria during or after colonisation. Siderophores are low-molecular weigh, high-affinity iron-binding compounds. [31]

TABLE 4. IRON ACQUISITION SYSTEMS

SIDEROPHORE	CHARACTERISATION
ENTEROBACTIN	(Also known as a enterochelin) It is a cyclic trimester of 2,3-dihydroxy-N-benzoyl-L-serine that removes iron-binding proteins and transports it into the bacterial cell. The products of seven chromosomal genes are required to synthesise enterobactin, which can be used only once, because the molecule is cleaved within the cell and the remnants discarded for the iron to become available. It is thus very energy expensive process.
AEROBACTIN	Aerobactin is a conjugate of 6-(N-acetyl-N-hydroxylamine)-2-aminohexanoic acid and citric acid. It forms an octahedral complex with ferric iron. This siderophore can be recycled and unlike enterobactin, it is not bound by serum albumin.
TONB AND THE TONB-DEPENDENT IRON UPTAKE SYSTEMS	TonB is known to be required by *E. coli* for high-affinity transport and utilization of several nutrients including chelated iron and vitamin B12.

2. MATERIALS

2.1. BACTERIAL STRAIN AND PLASMIDS

The bacterial strains and plasmids used in this study are described in Table 5.

TABLE 6. BACTERIAL STRAINS, MAMMALIAN CELL LINE AND PLASMIDS

STRAINS	GENOTYPE/FENOTYPE	SOURCE
BACTERIAL STRAINS		
E. coli 536	WT urinary tract isolate of the serogrup O6:K15:H31	U. Dobrindt *
E. coli IHE3034	WT meningitis isolate of the serogrup 18ac:K1:H7	U. Dobrindt *
IHE3034ΔPKS	IHE3034 mutant, deleted PKS gene cluster	
J96-M1ΔPKS	J96-M1 mutant, deleted PKS gene cluster	
E. coli DH5α	supE44, Φ80lacZΔM15, ΔlacU169, hsdR17, recA1, endA1, gyrA96, thi-1, relA1	Gibco [33]
E. coli TOP10	F-, mcrA, -80lacZΦM15, Δ(mrrhsdRMS-mcrBC), ΔlacX74, recA1,araD139, (ara-leu)7697, galU, galK,rpsL, (StrR), endA1, nupG	Invitrogen
MAMMALIAN CELL LINE		
HeLa	Human cervix epithelioid carcinoma	U. Dobrindt *
	PLASMIDS	
pTZ19R	2,8kb, lacZα, MCS, Amp^R	Fermentas
PCR4-TOPO	3.9 kb, Amp^R, Kan^R, lacZα-ccdB	Invitrogen
IHE3034 BAC11	IHE3034, PKS coding region	S. Homburg *
pAKR2	Amp^R, Km^R, p15or, delivered from pTC ETEC 2173	G. Blum-Oehler *

* Institut für Molekulare Infektionsbiologie, Würzburg, Germany

All bacterial strains were stored in 20% glycerol at –80°C.

2.2. CULTURE MEDIA

All culture media were sterilized by autoclavation (121˚C, 40 minutes) just after preparation (Table 6.). pH of the medium was adjusted to 7.2 [32]

TABLE 7. MEDIA

BACTERIAL MEDIA	
LIQUID MEDIA	
LB (Luria-Bertani medium)	10 mg/ml bacto-tryptone 5 mg/ml yeast extract 10 mg/ml NaCl (pH 7.4)
SOLID MEDIA	
To obtain solid media (hard agar-based culture plates for growth of bacterial colonies), 15 g/l agar were added to the liquid medium (LB) before autoclavation. After cooling, the agarose-medium was poured directly into Petra dishes, allowing about 30-35 ml per 85mm dish. Blood agar plate (BAP) contains LB supplemented with 15 g/l agar and 5% sheep blood.	
MAMMALIAN CELL LINE MEDIUM	
MEM – Minimal Essential Medium with Earle's Salts supplemented with: 10% FCS (heat-inactivated fetal calf serum) 2 mM glutamine 1 ml/100 ml MEM Non-Essential Amino Acids DMEM – Dulbecco's Minimal Essential Medium with high Glucose supplemented with: 10% FCS (heat-inactivated fetal calf serum) 2 mM glutamine	
The cell culture media and supplements were obtained from GIBCO BRL.	

Antibiotics were used at the following final concentrations: ampicillin – 100 µg/ml, kanamycin – 30-50 µg/ml, tetracycline 12,5 µg/ml, and chloramphenicol 12,5 µg/ml.

2.3. BUFFERS, SOLUTIONS AND CHEMICAL REAGENTS

All solutions and buffer stocks were prepared using chemicals of analytical grade or better. Aqueous solutions were prepared using bidistilled water only. All solutions were sterilized by autoclavation (121˚C, 30min) or filtration (pore size 0,22 m).

TABLE 8. BUFFERS, SOLUTIONS AND REAGENTS

STOCK SOLUTIONS	
Agarose, 2,5-0,8%	2,5-0,8 g agarose dissolve in 100 ml TAE Buffer. Biozym GmbH, Hessisch Oldendorf, Gibco
$CaCl_2$, 1 M	147 g/l $CaCl_2$. Autoclaved, stored at room temperature. Merck
EDTA, 0.5 M	186 g/l ethylenediamine tetraacetic acid (disodium salt and 10 N NaOH were used to set pH 8.0), autoclaved, stored at room temperature. Serva
Ethidium bromide	10 mg/ml; used to visualize DNA: emits fluorescence when excited at 250-320 nm. Stored at 4 ˚C in dark. Handle with gloves and avoid inhalation! Strong mutagen! Sigma
Glucose, 20%	20 g glucose dissolved in 100 ml water, sterile filtered, stored at 4 ˚C
IPTG, 100 mM	23,8 mg isopropyl-D-thiogalactopyranoside in 1 ml H_2O, stored at –20˚C. Biomol
NaAc, 3 M	408 g CH_2COONa $3H_2O$ in 1 liter H_2O. Acetic acid to pH 5,2; autoclaved
PEG6000/NaCl	20% PEG6000 (Polyethylene Glycol)/2.5 NaCl: 40 g PEG600, 29.22 g NaCl in 200 ml of water, autoclaved
SDS 20%	200 g SDS (Sodium dodecyl sulfate) dissolved in 1-liter water at 60˚C, not autoclaved!
Sucrose gradient solutions: 40%, 32,5%, 25%, 17,5% end 10%	40%, 32,5%, 25%, 17,5% end 10% dissolve in Buffer: 1M NaCl, 20mM Tris-HCl, 5mM EDTA pH 8,0
Xgal, 2%	2 mg Xgal (5-bromo-4-chloro-3-indolyl-D-galactoside, 1 ml of dimethylformamide (DMF). Stored dark at –20˚C. Biomol

BUFFERS	
EBSS (washing buffer)	382,5 ml water, 50 ml EBSS1, 50 ml EBSS2, 10 ml EBSS3, 15 ml EBSS4
EBSS1	1.06 g $CaCl_2$, 0.79 g $MgSO_4$ in 400 ml water, autoclaved
EBSS2	1.58 g KCl_2, 27.35 g NaCl, 0.62 g NaH_2PO_4 in 400 ml water, autoclaved
EBSS3	5.55 g Glucose in 25 ml water
EBSS4	7.5 g $NaHCO_3$ in 100 ml water
Gel loading buffer, 5x	50% glycerol, 0,5% bromophenol blue, 0,5% xylene cyanol in 5x TAE, stored at 4 ˚C
Loading dye	5 vol. Deionized formamide, 1 vol. 25 mM EDTA (pH 8,0) with blue dextran (50 mg/ml)
Shearing buffer	TE, pH 8 containing 10% glycerol
TAE Buffer, 50x	242 g Tris base, 57,1 ml glacial acetic acid, 37,2 g $Na_2EDTA\text{-}2H_2O$ (2mM) dissolved in 1000 ml water, (pH 8,0). Stored at room temperature
TE buffer	10 mM Tris-Cl (pH 7,4) and 1 mM EDTA (pH 8,0), stored at room temperature
TfbI	30 mM potassium acetate, 100 mM rubidium chloride, 10 mM calcium chloride, 50 mM manganese chloride, 15% glycerol, pH 5,8 with dilute acetic acid
TfbII	10 mM MOPS, 75 mM calcium chloride, 10 mM rubidium chloride, 15% glycerol, pH 6,5 with dilute NaOH

CHEMICAL REAGENTS	**SOURCE**
Acrylamide	Serva
Ampicillin	Sigma
APS	Sigma
dATP	MBI, Fermentas
Agar-Agar	Roth
Tryptone	Oxoid
Bisacrylamide	Serva
DMSO	(Dimethylsulfoxid), Sigma
Ethanol	Merc
dNTPmix	Promega
Glycerol	Fuka Biochemica

Glycine	Serva
Isopropanol	Roth
MOPS	Sigma
Phenol	Biomol
Primers	Gibco
TEMED	(N,N,N,N,-Tetramethy –Ethylenediamine), Sigma
Tris HCl	Sigma
Urea	Serva
Yeast extract	Oxoid

2.4. DNA SIZE MARKERS

For estimation of the DNA fragments size GeneRuler™ DNA Ladder Mix (MBI, Fermentas) was used (Figure 10).

Figure 10. DNA Ladder Mix

2.5. ENZYMES AND COMMERCIALLY PREPARED KITS

TABLE 9. ENZYMES AND KITS

ENZYME	SOURCE
Taq DNA Polymerase	Qiagen, Fermentas, Promega, TaKaRa Bio inc., Bioline
T4 Polymerase	MBI, Fermentas
T4-DNA-Ligase	MBI, Fermentas
Calf Intestin Alkaline Phosphatase	MBI, Fermentas
Klenow Fragment	DNA Polymerase I large fragment, MBI Fermentas
Proteinase K	Boehringer Mannheim GmbH
Rnase (ribonuclease)	Boehringer Mannheim GmbH
EcoRI	MBI, Fermentas
AluI	MBI, Fermentas
SmaI	MBI, Fermentas
Lysosym	Biomol

KITS	SOURCE
Qiaprep Spin Miniprep Kit	Qiagen
QIAEX II Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
DirectPrep 96 BioRobot Kit	Qiagen
Qiagen Plasmid Mini Kit	Qiagen
BigDye Terminator Cycle Sequencing Kit	Applied Biosystem
TopoTA Cloning Kit For Sequencing with Top 10 Cells	Invitrogen
AquaPure Genomic DNA Isolation Kit	Bio-Rad
DYEnamic ET Dye Terminator Kit	Amersham Bioscience
Megabase 4000 Long Read Matrix	Amersham Bioscience

3. METHODS

3.1. BACTERIAL GROWTH CONDITIONS

In most cases, cells were grown to early log phase (OD_{600} 0,2 to 0,4); LB medium was used in most experiments, supplemented with glucose and/or antibiotic. Cultures were grown at 37˚C and shaken at 180-350 rpm.

3.2. COMPETENT CELLS PREPARATION

1ml from overnight culture was inoculated into 100 ml LB medium and incubate at 37˚C with aeration to early log phase (OD_{600} 0,2 to 0,4). The sample was incubated on ice for 15 minutes. Cells were pelleted in appropriate centrifuge tube 3-5000 x g 5 min (~5000 rpm in a Sorvall SS-34 rotor). Pellet was resuspended with 40 ml TfbI and incubated on ice for 15 minutes. After this time samples were centrifuge (with the same condition) and the obtained pellet was resuspended in 40 ml TfbII,and incubated on ice for 15 minutes and either use immediately or quick freeze at -70C for storage. Samples were usually saved these in 0,25 to 0,5 ml aliquots.

3.3. RUBIDIUM CHLORIDE METHOD FOR TRANSFORMATION COMPETENT *E. COLI* CELLS

Methods for *E. coli* transformation were adapted from Sambrook et al. (1989). [32] Approximately 40-100 µl of competent cells were transferred to microcentrifuge tubes and 6 µl pCR4 Topo ligation sample to Top 10 cells (3.7.3.) or1-2 µl of plasmid were added (approximately 50 ng plasmid per 10 µl of competent cells). The samples were incubated on ice for 10 minutes, followed by incubation in 42 °C water bath for 45 sec. 400-800 µl of LB medium were added. Next, the samples were incubated in 37 °C for 45-60 minutes. 50-100 µl of the putative transformed cells were transferred to LB plates containing selective antibiotic and Xgal and IPTG. Plates were then incubated at 37 °C overnight.

3.4. ELECTROCOMPETENT CELLS PREPARATION

1ml from overnight culture was inoculated into 100 ml LB medium and incubate at 37˚C with aeration to early log phase (OD_{600} 0,2 to 0,4). The sample was incubated on ice for 15 minutes. Cells were pelleted in appropriate centrifuge tube 3-5000 x g 5 min (~5000 rpm in a Sorvall SS-34 rotor). Pellet was washed with prechilled, sterile water two times and resuspended with 1 ml of 10% glycerol. Samples were use immediately or quick freeze at -70C for storage. Samples were usually saved these in 0,25 to 0,5 ml aliquots.

3.5. ELECTROPORATION AND BLUE-WHITE SELECTION

40 µl electrocompetent cells were thawed on ice and transferred into a prechilled 0.1 cm electrode Gene Pulser Cuvette (Bio-Rad). 1 µl DNA solution (1 ng/µl) or 1.5 µl ligation product was added directly into the competent cells and mixed well by gently flicking. Then, the surface of the cuvette was completely dried and the electroporation was performed using Gene Pulser (Bio-Rad) under the condition of 1.8 kV voltage, 200 Ω resistance, 25 µF capacitance. Afterwards, 700-1000 µl prewarmed LB medium was immediately supplied to the electroporated *E. coli* for recovery. The cells were recovered at 37 °C rotating for 1 hour, followed by plating 100 µl and 900 µl on separate plates containing selective antibiotic, X-gal and IPTG.

3.6. GENERAL DNA MANIPULATION METHODS

3.6.1. ISOLATION OF PLASMID DNA

3.6.1.1. PLASMID ISOLATION USING A KIT

The plasmids were isolated from the fluid *E. coli* cultures using Qiaprep Spin Miniprep Kit (Qiagen, Valencia, CA). The procedures were followed as described by Qiagen instructions.

3.6.1.2. QIAGEN BIO ROBOT 9600

DNA was isolated using a Qiagen Bio Robot 9600 (Qiagen GmbH, Germany).
The BioRobot 9600 is designed to automate routine liquid-handling tasks and nucleic acid purification for general molecular biology research laboratories and core facilities. This workstation integrates proven QIAGEN chemistries, ready-to-run protocols, and dependable instrumentation for consistent, reliable results in molecular biology projects.

Figure 11. Qiagen Bio Robot 9600

3.6.2. PURIFICATION AND PRECIPITATION OF DNA FRAGMENTS

3.6.2.1. PURIFICATION USING A KIT

Before performing ligation reactions and sequencing reactions, all DNA fragments were purified using PCR Purification Kit (Qiagen, Valencia, CA).

3.6.2.2. DNA PRECIPITATION

Precipitation was performing by adding to the DNA sample 1/10 volume NaAC and 3 volumes of EtOH absolute. Mixture was chilled for at least 10 min. Frequently; this precipitation was performed by incubation at –20°C overnight. To recover the precipitated DNA, sample was centrifuged for at 4°C for 30 min at 12000 rpm, supernatant discarded and the DNA palled washed with: 70 % EtOH (cold), and centrifuged at 4°C, 12000 rpm, 3 min. The pellet was dried and resuspended in water.

3.6.3. CHROMOSOMAL DNA ISOLATION

Genomic DNA was extracted from *E. coli* 536 urinary tract isolate using the AquaPure Genomic DNA Isolation Kit (Bio-Rad) according to manufacture protocol.

3.6.4. DNA RESTRICTION ENDONUCLEASE DIGESTION

Restriction digests were carried generally using the specific buffer provided by the manufacturer. Multiple restrictions were combined in a single digest provided the buffers are compatible. Digestion volumes varied depending on the amount of DNA being digested, normally 200-300 ng DNA were digested in a 20 ml volume using 1 unit of enzyme for 1 hour at 37°C. Digestions were stopped by the addition of DNA sample loading solution.

3.6.5. DNA: FILLING IN 3'-RECESSED ENDS OF DNA, KLENOW

DNA polymerase I large fragment (Klenow fragment) (2U/µl, Fermentas) was used to fill-in the ends of 5'-overhang DNA fragment. After clean up of the digested DNA pellet (0,1 – 4 µg) were resuspended in 32 µl of water, 4 µl of 10X Klenow buffer and 4 µl dNTP solution (0,05 mM – final concentration) and 1 unit of Klenow Fragment per microgram of DNA. The reaction was incubated at 37°C for 10 minutes and stopped by heating the mixture at 75°C for 10 minutes.

3.6.6. T4 DNA POLYMERASE

3'-overhang DNA fragment was sharpened blunt using T4 DNA Polymerase (5U/µl, Fermentas). T4 DNA polymerase catalyzes 5'→ 3' synthesis of DNA and has got a 3' → 5' exonuclease activity.

Reaction mixture was prepared on an ice by adding the following:

5x reaction buffer	10 µl
digested/sheared DNA	20-25 µl (5 µg)
2 mM dNTPmix	1,5 µl
T4 DNA Polymerase	1 µl (5 U/µl)
water	to 50 µl,

then incubated the mixture at RT for 1,5 h.

3.6.7. MODIFICATION OF 3' ENDS WITH A TAQ DNA POLYMERASE

The adenine overhangs at the 3'-blunt ends were created using the template-independent terminal transferase activity of Taq DNA polymerase.

The reaction was performed on an ice by adding the following:

DNA	50 µl
2 mM dATP	6 µl
25 mM $MgCl_2$	6 µl
10x PCR Buffer without $MgCl_2$	7 µl
Taq DNA Polymerase 5U/µl	0.5 µl,

then incubated the mixture at 72°C for 25 min.

3.6.8. DEPHOSPHORYLATION

The enzyme catalyzes the release of 5'-phosphate groups from DNA.
5'-ends of DNA fragment dephosphorylation was performed by directly adding:

DNA	50 µl
10x CIAP Buffer	10 µl
CIAP 1U/µl	2 µl
water	40 µl,

mixture was incubated at 37 °C for 30 min. (max. 35 min.!)

3.6.9. LIGACJA Z VECTOREM

25-100 ng purified vector fragment was mixed with 3-10 folds (molecular ratio) of purified insert fragment, 1 µl 10×T4 DNA ligase buffer (MBI Fermentas), and 1 µl T4 DNA ligase (3 U/µl, MBI Fermentas) in a total volume of 10 µl. This ligation mixture was incubated at RT for at least 2 hours or at 15°C overnight.

3.6.10. DNA AMPLIFICATION (PCR, POLYMERASE CHAIN REACTION)

The Polymerase Chain Reaction makes possible the synthesis of specific DNA sequences using two oligonucleotides complement to the two DNA strands. The reaction requires a Thermocycler, which controls the exact temperature for each of the different steps that characterize the PCR: melting, annealing and amplification. The amplification reaction is carried out by the special heat resistant Taq-polymerase. The reaction provides several factors, beside of the temperature, to adjust the reaction to optimal conditions, as the concentration of magnesium chloride. PCR reactions were performed according to instructions for each Polymerase used.

Standardized PCR protocol:

Reaction mix:	
DNA	10-200ng
PCR-buffer*	1x
dNTPmix	200µM
$MgCl_2$	1mM - 5mM
Primer1	100pmol
Primer2	100pmol
Taq-polymerase	2,5 U/50µl reaction volume
H_2O	to a final volume of 50µl or 20µl

* PCR buffer 10x: 500mM KCL, 200mM Tris-HCl pH 8,6, 1% Triton x-100

Cycle parameters:	
Initial denaturation:	3 min 94°C
3-step cycling	
Denaturation:	0.5–1 min 94°C
Annealing:	0.5–1 min 50–68°C (approximately 5°C below Tm of primers)
Extension:	1 min 72°C For PCR products longer than 1 Kb, increase the extension time by approximately 1 min per kb DNA
Number of cycles:	25–35
Final extension:	10 min 72°C

3.6.11. DNA GEL ELECTROPHORESIS

Molecules in a mixture can be separated according to size by electrophoresis, a technique dependent on the fact that dissolved molecules in an electric field move at a speed determined by their charge-mass ratio. Nucleic acids in solution generally have a negative charge because their phosphate groups are ionized; thus they migrate toward a positive electrode. However, nucleic acid molecules consisting of long

chains have almost identical charge-mass ratios, whatever their length, because each residue contributes about the same charge and mass. Therefore, if the electrophoresis of nucleic acids were simply carried out in solution, little or no separation of molecules of varying lengths would occur.

Agarose gels were prepared in 1x TAE Buffer containing 250 ng/µl ethidium bromide and the appropriate percentage of agarose according to the size of fragments being separated: 2.5 % agarose gels were used for electrophoresis of fragments below 1 kb; 1.0% agarose gels were used for analysis of larger fragments. Electrophoresis was performed at 50 - 90 V for 15 - 45 minutes depending on the separation required.

3.7. GENE BANK PREPARATION

Prior to beginning the process of gene bank preparation, agar plates (hard agar-based culture plates supplemented with Amp, X-gal and IPTG) must be prepared.

3.7.1. SHEARING THE DNA

The DNA was shred using a nebulizer. It is a small plastic device which allow to atomize liquids with a compressed air to shear DNA to 0.5-6 kb in just a few seconds. The size of the sheared DNA is inversely proportional to the amount of pressure.

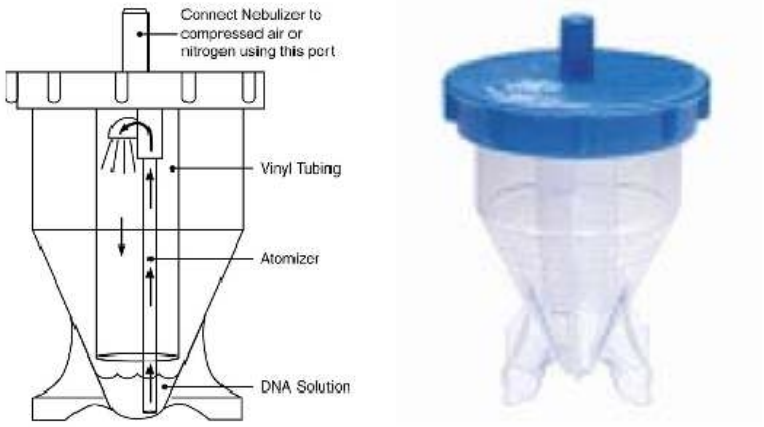


Figure 12. Nebulizer

After connection the nebulizer to the compressed air source, the following steps were performed:

- 3-25 µg DNA was added to 750 µl of shearing buffer into the bottom of the nebulizer
- to keep the DNA cold, nebulizer was placed in an ice
- the DNA was sheared for 45-90 s at 9-10 psi (0.62 bar)
- the sheared DNA was transferred to a sterile microcentrifuge tube
- sodium acetate precipitation was done and the DNA was resuspended to a concentration of ~50 ng/µl in a sterile water

3.7.2. SUCROSE GRADIENT – SIZE FRACTIONATION AND PURIFICATION OF SHEARED DNA

One of the most important parameters determining the success of gene bank construction is the quality and the fragments size of the DNA. For this purpose sucrose density gradient ultracentrifugation was used.

Procedure:

1. Prepare sucrose gradient solutions: 40%, 32,5%, 25%, 17,5% end 10% (Table. 7)
2. Add 1.9 ml the following sucrose solution sequentially into the 10 ml ultracentrifugation tube (use a glasspipette or cut off approximately 2 mm from the top of a 1 ml pipette tip; after each sucrose solution incubate a tube in –20°C for ca. 45 min.)
    a. 1.9 ml of 40% sucrose solution
    b. 1.9 ml of 32.5% sucrose solution
    c. 1.9 ml of 25% sucrose solution
    d. 1.9 ml of 17.5% sucrose solution
    e. 1.9 ml of 10% sucrose solution
3. Keep the tube at –20°C for at least 12-16 h (i.e., overnight). A linear gradient will form.
4. Layer ca. 850 µl sheared DNA on a top of the sucrose gradient solution

5. Gently insert the tube into a pre-chilled Swing Out Rotor TH641, balance the rotor and centrifuge the sucrose density gradient using ultracentrifuge Sorvall Ultra Pro 80, Sorvall ODT50B or ODT55B at 20°C for 20 h at 27000U.
6. After centrifugation prepare 26 microcentrifuge tubes
7. Remove the centrifuge tube from the rotor and carefully transfer 400 µl of fraction from the top of the gradient to the microcentrifuge tubes with 1 ml pipette tip.
8. Add 200 µl PEG6000/NaCl (Table 7.) and mix the samples.
9. Incubate samples at RT, overnight.
10. DNA purification:
    a. Centrifuge the samples for 30 min, RT and descent the supernatant
    b. Mix the samples with 700 µl of 70% EtHO
    c. Repeat the centrifugation.
    d. Dry the DNA pellet in the Speed Vac at 30°C for 20 min.
    e. Resuspend the DNA in a 25 µl of sterile water.

3.7.3. TOPO CLONING

Topo cloning it is a few step procedure:

1. Sharpening blunt 3'-overhang DNA fragment using T4 DNA Polymerase (3.6.6.)
2. Modification of 3' ends with a Taq DNA Polymerase (3.6.7.)
3. Dephosphorylation (3.6.8)
4. Ligation with a Topo vector

   Reaction mix:

DNA	4 µl
Salt solution	1 µl
Vector pCR4Topo (includes topoisomerase)	1 µl

   Incubation: 2h, RT
5. Transformation using competent Top 10 cells and clone selection (3.3.)

Figure shows the principles of Topo cloning strategy using a pCR4Topo cloning vector.

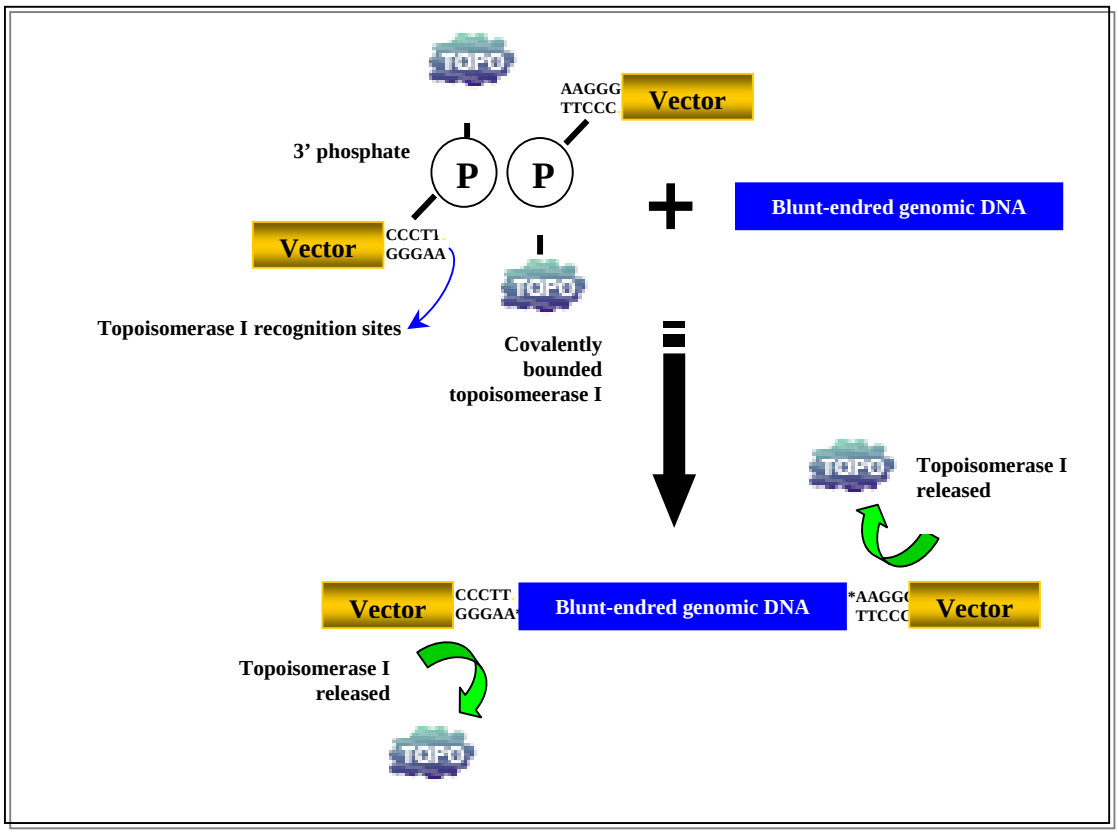


Figure 13. Topo cloning strategy

3.8. DNA SEQUENCING

3.8.1. SAMPLE PREPARATION

The sequencing reactions are carried out with AmpliTaqTM DNA Polymerase FS dye terminator cycle sequencing chemistry using the "ABI PRISM. BigDye. Terminator Cycle Sequencing Ready Reaction Kit". This Kit contains the four ddNTPs with different fluorescence labels. Reactions components and parameters were shown in a Table 9.

After sequencing reaction samples were purify with NaOAc (3M, pH4.6) and 99% ethanol. Finally the DNA pellet was resuspended in 2 µl loading dye.

TABLE 10. THE SEQUENCING REACTION COMPONENTS AND PATAMETERS

REACTION COMPONENTS		REACTION PARAMETERS**		
Premix*	4µl	01 LIDHT	110°C	ON
DMSO	1µl	02 PAUSE	98°C	FOREVER
Betain (5M)	1µl	03 TEMP	98°C	2min
Primer (5pmol/ µl)	1µl	04 LOOP [	---25x---	---------
H2O	xµl	05 TEMP	96°C	20 sec (degradation)
DNA (400ng)	xµl	06 TEMP	50°C	15 sec (annealing)
	---------	07 TEMP	60°C	4 min (extension)
	20µl	08 LOOP]	---------	---------
		09 LIDHT	OFF	---------
		10 TEMP	4°C	FOREVER
		11 END		

* ABI Big Dye Terminator Kit, reactions mix.
** Reaction cycle with 96er PCR-Thermo block (MWG-Primus 96 plus)

3.8.2. DNA SEQUENCING WITH THE ABI PRISM® 377 AUTOMATED SEQUENCER (APPLIED BIOSYSTEMS)

The ABI PRISM® 377 DNA Sequencer automatically analyzes DNA molecules labeled with multiple fluorescent dyes (Figure 13.). After samples are loaded onto the system's vertical gel, they undergo electrophoresis, laser detection, and computer analysis. Electrophoretic separation can be viewed on-screen in real-time, and final data can be output in a variety of formats. The 377 systems accept gel plates in different lengths for flexibility in sample analysis. The systems ship with the Apple Data Collection and Analysis software modules. Software versatility enables this system to support sequencing and fragment analysis applications Reads up to and beyond 900 bases per sample generating long reads Electrophoresis optimizes speed, resolution, and read length (Figure 14.).

Figure 14. The ABI PRISM® 377 DNA automated DNA sequencer

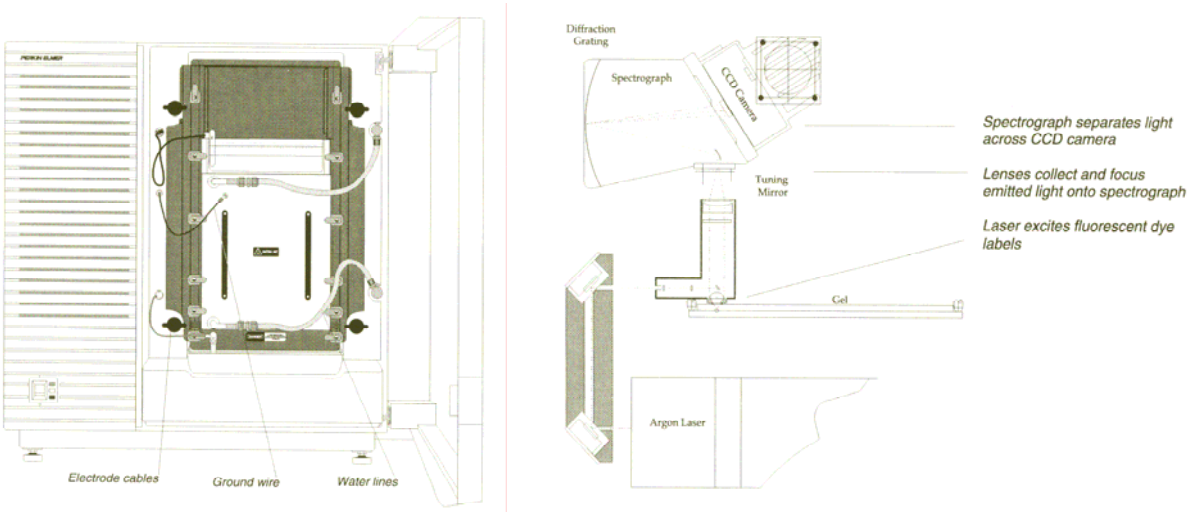


Figure 15. The ABI PRISM® 377 DNA automated DNA sequencer - diagram

3.8.3. DNA SEQUENCING USING MEGABACE DNA SEQUENCING SYSTEMS 1000 AND 4000 (AMERSHAM BIOSYSTEMS)

DNA samples were all sequenced using BigDye Terminator chemistry and resolved on the MegaBACE 1000 and 4000 Sequence Analyzer. The instrument is a 96-capillary array fluorescence-based DNA sequencer with interchangeable filter sets, confocal detection, and NT workstation. It automates gel matrix replacement, sample injection, DNA separation, and data analysis; uses a linear polyacrylamide separation matrix; and has a turnaround time of 2 hr per run. Minimal sample loading volume is 5 µL and the samples are introduced into the system in a 96-well format. Read lengths was excess of 800 bp.

Figure 16. MegaBace4000 DNA sequencer.

3.8.4. DNA SEQUENCING WITH THE ABI 3730XL AUTOMATED SEQUENCER (APPLIED BIOSYSTEMS)

For pARK2 and IHE3034 BAC11 the ABI 3730xl DNA Analyzer was used. All inserts from shotgun library were sequenced using BigDye Terminator chemistry and run on ABI 3730xl DNA Analyzer. The ABI 3730xl uses a 96-capillary electrophoresis system that creates a sensitive detection system, long sequence reads (up to a 1000 bases for high quality DNA). The instrument is an automated system (sample loading, separation matrix preparation, and sequence analysis), which coupled with the facility's liquid handling robot, dramatically reduces the introduction of human error.

3.9. RAW DATA PROCESSING

Sequences were processed first with Phred and Pregap4, assembled into contigs by using Phrap assembling tool and edited with Gap4, witch is a part of the Staden package software (http://staden.sourceforge.net/). [34, 35]

3.9.1. PHRED – BASE-CALLING SOFTWARE WITH QUALITY INFORMATION

Phred reads DNA sequencer trace data, calls bases, assigns quality values to the bases, and writes the base calls and quality values to output files. Phred can read trace data from SCF files and ABI model 373 and 377 DNA sequencer chromate files, automatically detecting the file format. After calling bases, Phred writes the sequences to files in either FASTA format, the format suitable for XBAP, PHD format, or the SCF format. Quality values for the bases are written to FASTA format files or PHD files, which can be used by the Phrap sequence assembly program in order to increase the accuracy of the assembled sequence.

3.9.2. PREGAP4 - PREPARING SEQUENCE TRACE DATA FOR ANALYSIS OR ASSEMBLY

Before entry into a Gap4 database the raw data from sequencing instruments needs to be passed through several processes, such as screening for vectors, quality evaluation, and conversion of data formats. Pregap4 is used to pass a batch of readings through these steps in an automatic way. It provides an interface for setting up and configuring the processing and for controlling the passage of the readings through each stage. The separate tasks are termed "modules" and a dedicated program typically manages each module. Pregap4 wraps all of these modules into a single easy to use environment, whilst maintaining the flexibility to select and extend the processing modules.

3.9.3. PHRAP - SEQUENCE-ASSEMBLY PROGRAM

Phrap ("phragment assembly program", or "phil's revised assembly program") is a program for assembling shotgun DNA sequence data. Some of its key features are: (1) allows use of entire read, not just highest quality part; (2) uses a combination of user-supplied and internally computed data quality information to improve accuracy of assembly in the presence of repeats; (3) constructs contig sequence as a mosaic of the highest quality parts of reads, rather than a consensus. Phrap does not provide editing or viewing capabilities; these are available with consed or phrapview (http://www.phrap.org).

3.9.4. GAP4 - A GENOME ASSEMBLY AND EDITING PROGRAM

The program contains all the tools that would be expected from an assembly program plus many unique features and a very easily used interface. [36]
Gap4 is very big and powerful tool: performs assembly, contig joining, and assembly checking, repeat searching, experiment suggestion, and read pair analysis and contig editing. Has graphical views of contigs, templates, readings and traces which all scroll in register. Contig editor searches and experiment suggestion routines use Phred confidence values to calculate the confidence of the consensus sequence and hence only identify places requiring visual trace inspection or extra data.

3.10. SEQUENCE ANALYZE

3.10.1. ORF FINDING

Initial gene prediction was accomplished using a YACOP program. The predictions generated by this tool are based on the output of existing gene finding programs. In present YACOP supports the boolean combination of predictions from Critica105b (with wublast2), Glimmer2.02 or Glimmer2.10 (with RBSfinder), Orpheus (with dps) and ZCurve1.0. The current version is realized in Perl (v.5.6.1). [37] The output was verified and edited manually using criteria such as the presence of a ribosome-binding site (consensus Shine-Delgarno sequence for prokaryotic organisms: $^{5'}$AGGAGG$^{3'}$) and codon usage analysis. This was done with an Artemis - free genome viewer and annotation tool that allows visualization of sequence features and the results of analyses within the context of the sequence, and its six-frame translation. www.sanger.ac.uk/Software/Artemis)

3.10.2. ANNOTATION

Automatic annotation was done using the ERGO tool (Integrated Genomics, Inc., http://www.integratedgenomics.com). The predictions were verified and modified manually by searching derived protein sequences against public nucleotide or protein domain databases such as GenBank or PFAM, respectively. All software and databases are listed in Table 10.

TABLE 11. SOFTWARE AND DATABASES USED FOR AN ANNOTATION

PROGRAM	DESCRIPTION
BLAST 2.0	Basic Local Alignment Search Tool (BLAST) is a program for rapid searching of nucleotide and a protein database developed at NCBI, and is instrumental in identifying genes and genetic features. Sequence alignments provide a powerful way to compare novel sequences with previously characterized genes. Both functional and evolutionary information can be inferred from well-designed queries and alignments. http://www.ncbi.nlm.nih.gov/BLAST
COGs	Clusters of Orthologous Groups of proteins (COGs) were delineated by comparing protein sequences encoded in complete genomes, representing major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogs from at least 3 lineages and thus corresponds to an ancient conserved domain. [38] http://www.ncbi.nlm.nih.gov/COG
Pfam	Protein Families database of alignments and HMMs (Pfam) is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. For each family in Pfam you can: ook at multiple alignments, view protein domain architectures, examine species distribution, follow links to other databases, view known protein structures. [39, 40] http://pfam.wustl.edu/hmmsearch.shtml , http://www.sanger.ac.uk/Software/Pfam
PROSITE	It is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family (if any) a new sequence belongs. [41] http://ca.expasy.org/prosite/
LipoP 1.0	The LipoP 1.0 server produces predictions of lipoproteins and discriminates between lipoprotein signal peptides, other signal peptides and n-terminal membrane helices in Gram-negative bacteria. [42] http://www.cbs.dtu.dk/services/LipoP/

SignalP 3.0	SignalP 3.0 server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes. The method incorporates a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks and hidden Markov models. [43] http://www.cbs.dtu.dk/services/SignalP
DAS	Dense Alignment Surface (DAS) - prediction of transmembrane alpha helices in prokaryotic membrane proteins. [44] http://www.sbc.su.se/~miklos/DAS
TMpred	Prediction of Transmembrane (Tmpred) regions and orientation - program makes a prediction of membrane-spanning regions and their orientation. The algorithm is based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins. The prediction is made using a combination of several weight-matrices for scoring. [45] http://www.ch.embnet.org/software/TMPRED_form.html

3.11. DETECTION OF PKS TOXIN-PRODUCING *ESCHERICHIA COLI* ISOLATES BY PCR SCREENING

3.11.1. PCR SCREENING FOR THE POLYKETIDE-PAI (*ASNW*) IN DIFFERENT *E. COLI* ISOLATES

In order to investigate the occurrence of the identified PKS cluster in an *E. coli* strain collection PCR screening was applied.

A total of 185 isolates (Table 12) were subjected to PCR assay. Based on a conserved region within the PKS cluster of the 536 sequence eight pairs of primers were designed (Table 13). Bacterial strains were grown on agar plates at 37ºC overnight. Samples were collected from plates and resuspended in 500 µl of sterile water and boiled for 5 min.1 µl of the lysate was used as template DNA in a 20-µl PCR. PCR was performed, using eight pairs of primers to amplify 1.4 to 2.4 kb fragments. The

amplification programme consisted of 30 cycles of 30 seconds at 94ºC, annealing for 30 seconds at 55ºC for and a 2-minute extension at 72ºC with a final extension at 72ºC. The PCR products were separated in a 1.7% agarose gel. Electrophoresis was performed in 0.5 X Tris-borate-EDTA buffer for 1 h at 100 V. Amplicons were visualized with UV light after the agarose gel had been soaked in ethidium bromide solution (0.5 µg/ml) for 15 min.

TABLE 12. BACTERIAL STRAINS USED FOR PCR SCREENING

STRAIN	SEROTYPE	DESCRIPTION
MG1655	K12	Laboratory strain
EcoR50	O2:HN	UPEC (Gr.D)*
EcoR40	O7:NM	UPEC (Gr.D)*
EcoR48	ON:HM	UPEC (Gr.D)*
EcoR62	O2:NM	UPEC (Gr.B2)*
EcoR55	O25:H1	UPEC (Gr.B2)*
EcoR60	O4:HN	UPEC (Gr.B2)*
EcoR14	OM:HN	UPEC (Gr.A)*
EcoR71	O78:NM	UPEC (asymptomatic bacteuria) (Gr.B1)*
2980	018ac:K5:H-	UPEC
AE5	NS	UPEC
2A1	NS	UPEC
2A2	NS	UPEC
3B5	NS	UPEC
3N1	NS	UPEC
3N2	NS	UPEC
3N5	NS	UPEC
5A1	NS	UPEC
5A1U	NS	UPEC
8B2	NS	UPEC
8F4	NS	UPEC
16A3	NS	UPEC
16B1	NS	UPEC
19B1	NS	UPEC
A284	NS	UPEC
764	O18:K5:H5/11	UPEC
764-2	O18:K5:H5/11	UPEC
J96	O4:K-:H5	UPEC
J96-M1	O4:K:H5	UPEC
E-B35	O4:K12:H5	UPEC
AD110	O4:K6	UPEC
CFT073	O6:H1:K2	UPEC
RZ411	O6:K-:H1	UPEC
RZ532	O6:K-:H31	UPEC
RZ479	O6:K+:H-	UPEC
RZ451	O6:K+:H31	UPEC
RZ-436	O6:K13:H1	UPEC
RZ422	O6:K14:H-	UPEC

RZ505	O6:K14:H-	UPEC
536	O6:K15:H31	UPEC
RZ454	O6:K2:H-	UPEC
RZ-458	O6:K2:H1	UPEC
RZ-443	O6:K5:H-	UPEC
RZ-495	O6:K5:H-	UPEC
RZ-439	O6:K5:H1	UPEC
RZ-441	O6:K5:H1	UPEC
RZ-468	O6:K5:H1	UPEC
RZ-475	O6:K5:H1	UPEC
RZ-500	O6:K5:H1	UPEC
RZ-525	O6:K5:H1	UPEC
RZ-526	O6:K5:H1	UPEC
RZ-446	O6:K53:H1	UPEC
20A1	NS	UPEC
20A1U	NS	UPEC
20A2	NS	UPEC
22B2U	NS	UPEC
7521/94-1	NS	UPEC
2E2U	NS	UPEC
3D5	NS	UPEC
13A1	NS	UPEC
1G1U	NS	UPEC
1H1	NS	UPEC
1H1U	NS	UPEC
19A1	NS	UPEC
1G1	NS	UPEC
2E1U	NS	UPEC
16A-2U	NS	UPEC
4405/1	NS	UPEC
S5	NS	UPEC
8B1	NS	UPEC
100.	NS	UPEC
RB9	NS	(Sheep)
HK8 269/93	NS	Sepsis (blood culture)
HK24 10413/93	NS	Sepsis (blood culture)
HK1	NS	Sepsis (Blood culture)
HK8	NS	Sepsis (Blood culture)
AC/I	O78	Sepsis (Avian)
HK4	NS	Sepsis
HK17	NS	Sepsis
HK19	NS	Sepsis
HK25	NS	Sepsis
E642	NS	Sepsis
HK54	NS	Sepsis
B616	NS	Sepsis
HK2	NS	Sepsis
E457	NS	Sepsis
IHE3036	O18:K1:H7	NBM
IHE3080	O18:K1:H7	NBM
IHE3034	O18:K1:H7/9	NBM

RS218	O18ac:H7:K1	NBM
RS226	O18ac:H7:K1	NBM
RS176	O7:K1: H /3	NBM
B13155	NS	NBM
Ve1140	NS	NBM
Ve239	NS	NBM
BK658	O75:K1:H7	NBM
B10363	NS	NBM
DSM6601 (Nissle 1917)	O6:K5	Faecal isolate
EcoR52	O25:H1	Faecal isolate (Orangutan) (Gr.B2)*
EcoR07	O85:HN	Faecal isolate (Orangutan) (Gr.A)*
EcoR58	O112:H8	Faecal isolate (Lion) (Gr.B1)*
EcoR31	O79:H43	Faecal isolate (Leopard) (Gr.E)*
EcoR16	ON:H10	Faecal isolate (Leopard) (Gr.A)*
EcoR42	ON:H26	Faecal isolate (Gr.E)*
EcoR43	ON:HN	Faecal isolate (Gr.E)*
EcoR35	O1:NM	Faecal isolate (Gr.D)*
EcoR51	O25:HN	Faecal isolate (Gr.B2)*
EcoR53	O4:HN	Faecal isolate (Gr.B2)*
EcoR54	O25:H1	Faecal isolate (Gr.B2)*
EcoR56	O6:H1	Faecal isolate (Gr.B2)*
EcoR59	O4:H40	Faecal isolate (Gr.B2)*
EcoR61	O2:NM	Faecal isolate (Gr.B2)*
EcoR63	ON:NM	Faecal isolate (Gr.B2)*
EcoR01	ON:HN	Faecal isolate (Gr. A)*
EcoR57	ON:NM	Faecal isolate (gorilla) (Gr.B2)*
EcoR32	O7:H21	Faecal isolate (giraffe) (Gr.B1)*
EcoR23	O86:H43	Faecal isolate (Elephant) (Gr.A)*
EcoR25	ON:HN	Faecal isolate (dog) (Gr.A)*
EcoR65	ON:H10	Faecal isolate (Celebese ape) (Gr.B2)*
EcoR66	O4:H40	Faecal isolate (Celebese ape) (Gr.B1)*
EcoR18	O5:NM	Faecal isolate (Celebese ape) (Gr.A)*
EcoR28	O104:NM	Faecal isolate (Gr.B1)*
EcoR24	O15:HN	Faecal isolate (Gr.A)*
F18	NS	Faecal isolate
F18Col -	NS	Faecal isolate
E351	NS	Faecal isolate
147/1	O:128:H-	ETEC
C9221a	O6:K15:H16	ETEC
E 1392/75	06:H16	ETEC
E57 IMI775	O138:K81	ETEC
H10407	O78:H11	ETEC
	O149:K88	ETEC
156A	055:H7	EPEC
179/2	055:H6	EPEC
182A	NS	EPEC
37-4	055:H-	EPEC
E2348/69	O127:H6	EPEC
12860	NS	EIEC
76-5	O143	EIEC
EDL-7284	NS	EIEC

78-5	NS	EIEC
4608-58	NS	EIEC
107-11	NS	EIEC
W7062	NS	EIEC
EDL-1284	O124:H-	EIEC
95004730	O111:H-	EHEC
2-45	NS	EHEC
2455/99	O128:H2	EHEC
2576/97	O103:H2	EHEC
2851/96	O103:H2	EHEC
2905/96	O103:H2	EHEC
2907/97	NS	EHEC
3115/97	O128:H2	EHEC
3117/98	O128:H-	EHEC
3172/97	O128:H2	EHEC
3574/92	O157:H7	EHEC
3697/97	NS	EHEC
3978/91	O157:H-	EHEC
4356/96	O26:H-	EHEC
4417/96	O111:H-	EHEC
4797/97	O91:H-	EHEC
5291/92	O157:H-	EHEC
5714/96	O103:H2	EHEC
5720/96	026:H11	EHEC
6037/96	O111:H-	EHEC
6105/96	O26:H-	EHEC
6578/93	O157:H7	EHEC
78/92	O111:H-	EHEC
86-24	O157:H7	EHEC
9167/91	O157:H7	EHEC
933W	O157:H7	EHEC
ED 147	O26:H11	EHEC
ED142	O111	EHEC
EDL-880	NS	EHEC
PIG E57	NS	EHEC
SF493/89	O157:H-	EHEC
189	NS	EAggEC
17-2	O3:H2	EAggEC
7484/94	NS	EAggEC
DDC 4441	NS	EAggEC
5464/95	NS	EAggEC
DPA065	O119	EAggEC
5477/94	O86:H7	EAggEC
042	NS	EAggEC
TA109	NS	(Rattus fuscipers)
TA349	NS	(Nyctophilus geoffroyi)
TA024	NS	(Sarcophilus harrisii)
TA083	NS	(Dasyurus geoffroii)

ON, HN = non-typeable with standard antisera NM = non-motile strain NS = not serotyped * = EcoR phylogenetical group

All bacterial isolates were provided by prof. J. Hacker (Institut für Molekulare Infektionsbiologie, Würzburg, Germany) and were stored in 20% glycerol at –80°C.

TABLE 13. PRIMERS USED FOR SCREENING

PRIMER'S NAME (POSITION)	SEQUENCE 5'→ 3'	PRODUCT SIZE
asnW-PAIleftend1 (784-803)	aatcaacccagctgcaaatc	1824 bp
asnW-PAIleftend2 (2588-2607)	caccccatcattaaaaacg	
asnW-PAIrightend1 (51940-51959)	agccgtatcctgctcaaaac	1413 bp
asnW-PAIrightend2 (53333-53352)	tcggtatgtccggttaaagc	
ORF1907-1908.1 (6271-6290)	tctgtcttggtcgcgttagtg	2304 bp
ORF1907-1908.2 (8555-8574)	tcagttcgggtatgtgtgga	
ORF1911-1912.1 (19912-19931)	attcgatagcgtcacccaac	2119 bp
ORF1911-1912.2 (22011-22030)	taagcgtctggaatgcagtg	
ORF1913-1914.1 (29674-29693)	cgcttcatcaacacgcttta	2118 bp
ORF1913-1914.2 (31772-37191)	cgcatcaggatgttctgcta	
ORF1915-1918.1 (36626-36645)	tcatcgcaatttggatttca	2255 bp
ORF1915-1918.2 (38861-38880)	tgatgaacgtggcggtaata	
ORF1919-1920.1 (41353-41372)	cctcgctaaagaaggtgacg	2421 bp
ORF1919-1920.2 (43754-43773)	accgttgactgtgatggaca	
ORF1920-1922.1 (50211-50230)	attcgccctgatattgtcg	2460 bp
ORF1920-1922.2 (52651-52670)	ccttcgttggcagattgatt	

3.11.2. IDENTIFICATION OF HEMOLYSIN-PRODUCING *E. COLI* ISOLATES

Culture of bacteria on blood agar (BAP) for the purpose of hemolysis classification was performed for overnight at 37°C in the presence of 5% CO_2. Hemolysis indicates a zone of clearing in the blood agar in the area surrounding a bacterial colony.

3.11.3. MAMMALIAN CELL LINE GROTH CONDITIONS

HeLa cells were cultured in MEM (Minimal Essential Medium) supplemented with 10% of FCS, glutamine and non-essential amino acids.

Adherent cell lines will grow in vitro until they have covered the surface area available or the medium is depleted of nutrients. At this point the cell line was sub-cultured in order to prevent the culture dying. Before splitting, cultures needed to be

viewed using an inverted microscope to assess the degree of confluency (cells should be passaged after reaching 70-80% confluency) and confirm the absence of bacterial and fungal contaminations.

For splitting of HeLa cells, medium was removed, cells were treated with pre-warmed to 37°C 0.05% trypsin /0.02% EDTA in PBS (1 ml per 25 cm^2 of surface area) until they no longer adhered to the plate (approximately 5 min of incubation at 37°C). The side of the flask could be gently tapped to release any remaining attached cells. Trypsin/EDTA was removed; cells were resuspend in 5 ml of fresh and pre-warmed to 37°C MEM medium to inactivate the trypsin. Required number of cells (dilution 1:10 or 1:20) was transferred to a new-labeled flask containing pre-warmed medium. Cells were grown at 37°C, 5% CO_2, 95% humidity and splited every 3-4 days.

3.11.4. TOXIN ACTIVITY ASSAYS

For the toxin activity assay eukaryotic cells were seeded into 96-well plates (Falcon) and incubated in MEM medium overnight at 37°C. Prior to the start of the toxin test the MEM medium was exchanged to 150 µl DMEM medium. 10 µl of overnight culture (OD_{600} = 0.6) were added to confluent monolayer cells per well and incubated for 3.5 h at 37°C in a 5% CO_2 and 95% air atmosphere. After this time cells were washed two times with the EBSS buffer and once with MEM medium containing of gentamicin (100 µg/ml). Fresh MEM medium with gentamicin was added. Cells were grown at 37°C, 5% CO_2, 95% humidity and controlled after 24-48 hours.
Assay for each isolate was conducted in quadruplicate and was independently repeated at least three times. Toxin effect was viewed using an inverted microscope.

4. RESULTS

4.1. GENOME PROJECT *E. COLI* 536

The uropathogenic *Escherichia coli* strain 536 (O6:K15:H31), isolated from a patient with acute pyelonephritis, is one of the model organisms of extraintestinal *E. coli*.

In an attempt to analyze the genetic basis of urovirulence the entire genome of strain 536 were sequenced and compared to the genome sequence of strain CFT073 (O6:K2:H1), another well-studied highly virulent uropathogenic *E. coli* (UPEC) strain, as well as to the genomes of non-pathogenic MG1655 (K12) and enterohemorrhagic strain EDL933 (O157:H7).

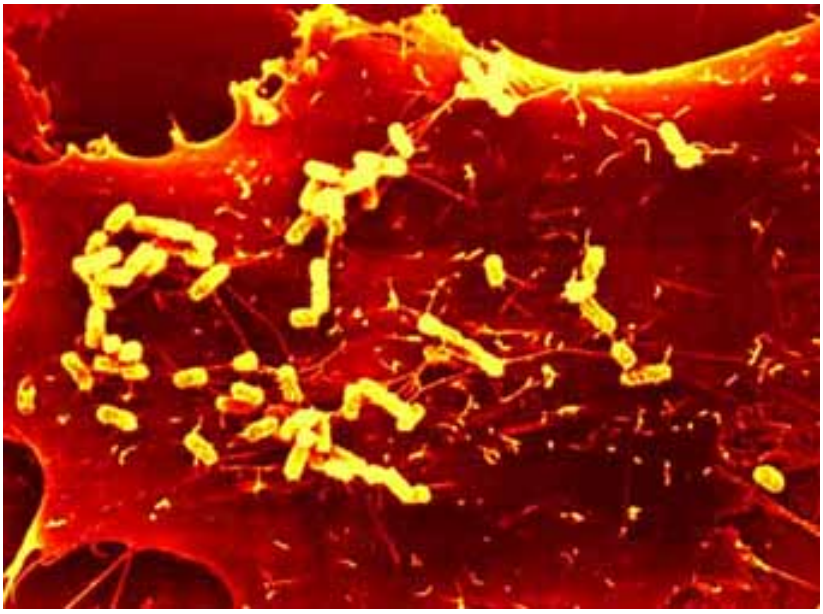


Figure 17. Bladder epithelial cell with adherent uropathogenic *E. coli* strain 536 cells (www.uni-wuerzburg.de/)

4.1.1. SEQUENCING

4.1.1.1. SHOTGUN STRATEGY

According to the whole genome shotgun (WGS) sequencing approach, developed by Sanger, the entire DNA was first randomly broken into fragments (usually about 1-3 kb) and subcloned into the cloning vector pCR4Topo (Figure 18) or pTZ19R.

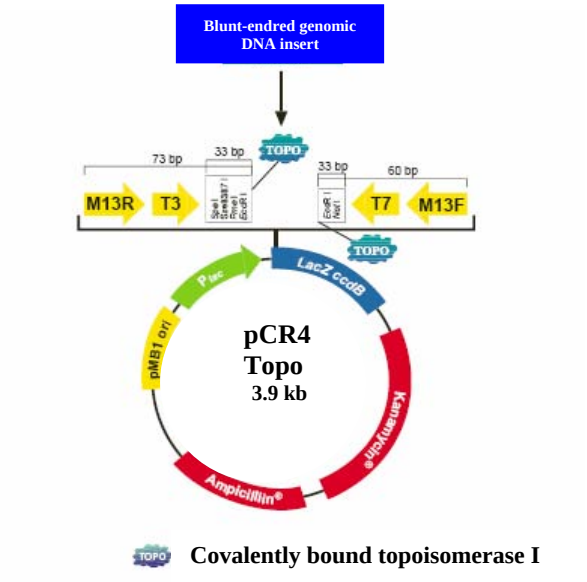


Figure 18. pCR4Topo vector used for shotgun library preparation.

Three shotgun libraries were used: two small insert libraries (DNA sheared using a nebulizer and partially digested DNA with *Alu*I enzyme) and a large insert library (cosmide library provided by U. Dobrindt). The small insert library was used for random shotgun sequencing and the cosmide library for gap closure and finishing procedures. In the first shotgun sequencing phase a number of 45800 sequence reads were generated using universal forward and reverse primers, which directed sequencing from within the cloning vector into the inserts. Subsequently, these sequence reads have been assembled into contigs by exploiting sequence overlaps (Figure 22). A contig is a set of overlapping sequence reads. The sequence reads in a contig can be summed to form a contiguous consensus sequence, which corresponds to the length of the contig. Sequencing strategy is illustrated in Figure 19.

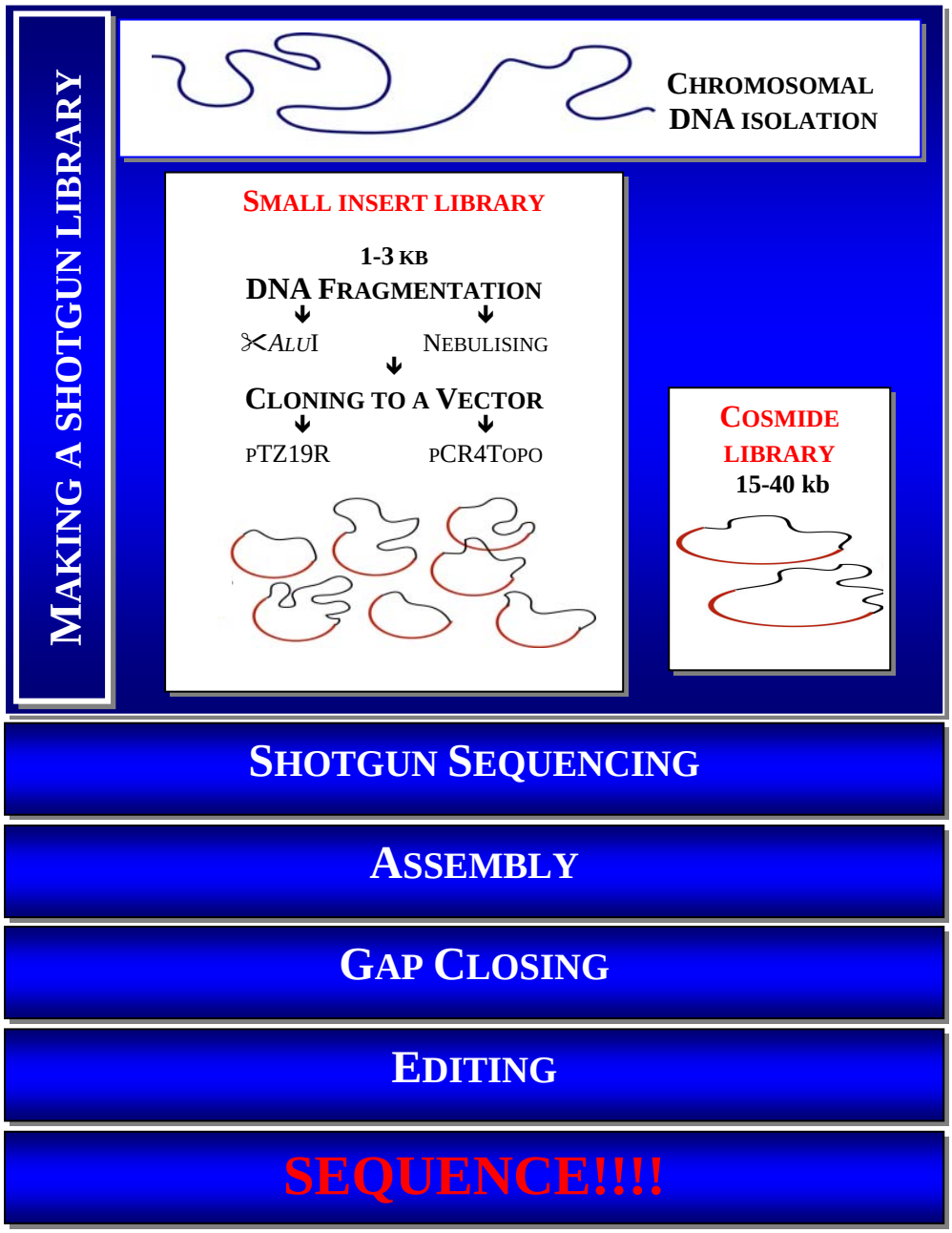


Figure 19. Shotgun strategy

4.1.1.2. RAW SEQUENCE PROCESSING AND ASSEMBLY

Lander and Waterman suggested that the number of times a base is sequenced follows a Poisson distribution. [46] There are two key assumptions they made. First, reads are randomly distributed in the genome, and second the ability to detect an overlap between two truly overlapping reads does not vary from clone to clone.

$$P = e^{-Nw/L}$$

Figure 20. Lander and Waterman formula; P – probability of any base being unsequenced, e – constant, e=2.718, N – number of clones, w – average read length in bp, and L – genome length

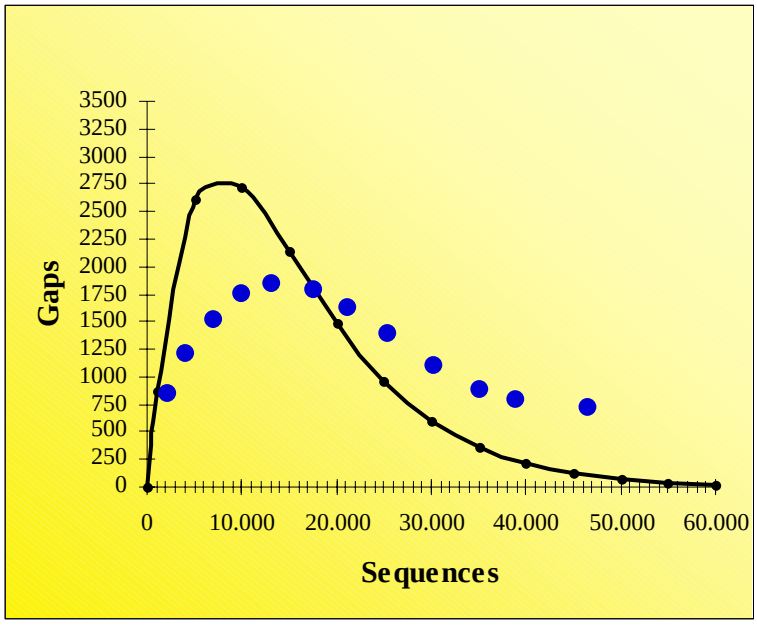


Figure 21.The end of raw sequencing phase; In black: optimal progression (genome size 5 Mb, average reading lenth 650 bp), In bleu: observed progression

The formula was used to plan and predict the progression of a shotgun-sequencing project. It helps to estimate the number of clones to be sequenced in order to get a certain average coverage of each base. In the *E. coli* 536 project 42,400 readings were generated, which led to an average coverage of 5.9.

In the assembly phase (Figure 22), all sequence reads were first compared to each other in order to identify overlapping parts. Identities between the sequences of different reads were noted, and these identities are used to align the sequences into sequence contigs. 739 contigs were obtained out of the first raw data processing event. 389 of them were bigger then 3 kb, and the predicted genome size was 4,880,500 bp.

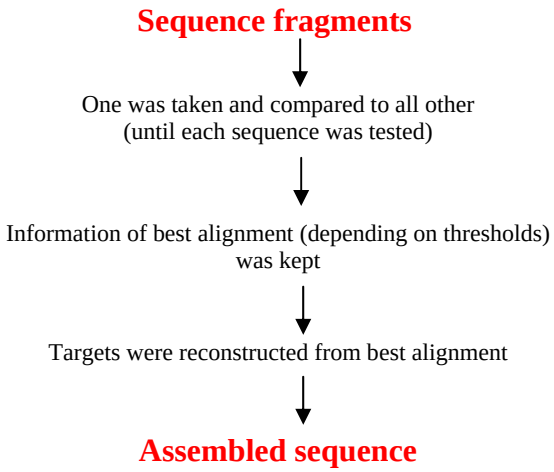


Figure 22. Assembly steps

4.1.1.3. EDITING AND GAP CLOSING

The first stage of the editing and gap-closing phase is resolving of misassembly problem. Misassemblies are caused by several factors, detailed below.

Assembly problems:

- repetitive elements
- clone gaps (cloning bias):
  - missing large insert clones
  - missing small insert clones

  - unclonable sequences
  - low coverage (the coverage can be defined as the average number of times any given base in the genome is sequenced)
- sequencing gaps (sequencing bias):
  - homopolymeric runs
  - high GC regions
  - inverted repeats
- chimeric clones

First all misassemblies were localized in all contigs bigger than 3 kb, disassembled and reassembled manually. An example of a misassembly is shown below (Figure 23). The gap4 Contig Editor is designed to allow rapid checking and editing of characters in assembled readings. The figure 23 shows a screendump from the Contig Editor that contains segments of improperly aligned readings; this misassembled region was identified manually during the finishing phase. Disagreements between the consensus and individual base calls are shown in green. Notice that these disagreements are not in poor quality base calls (reads a. and f. on a picture have got a based confidence more then 80%).

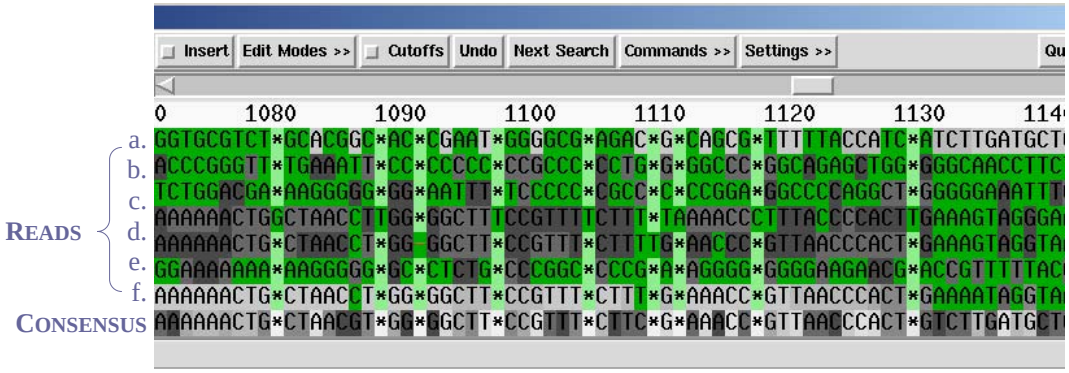


Figure 23. An example of misassembly. The gap4 Contig Editor window: the bottom line gives the contig sequence (consensus), and above it are the read sequences (a.-f.). Along the top of the editor window is a row of command buttons and menus. The background grey scale indicates base quality, with the highest quality being white and the lowest quality black. (*) A pad that is inserted by Phrap to align reads that have insertions and deletions. Disagreements between the consensus and individual base calls are shown in green.

4.1.1.3.1. STRATEGIES FOR CLOSING SEQUENCE GAPS

Gap closing was a challenging phase in the sequencing project, in particular since genome assemblies were complicated by the presence of repeat elements (like rRNA operons), insertion sequences and other similar factors that contribute to sequence misassemblies. In order to close the gaps a contig order was established. Several, described below, strategies were used to achieve that.

Generally speaking, there are two main types of gaps. One type of gaps is spanned (captured) by one or more plasmid read pairs (at least one clone spanning the gap).

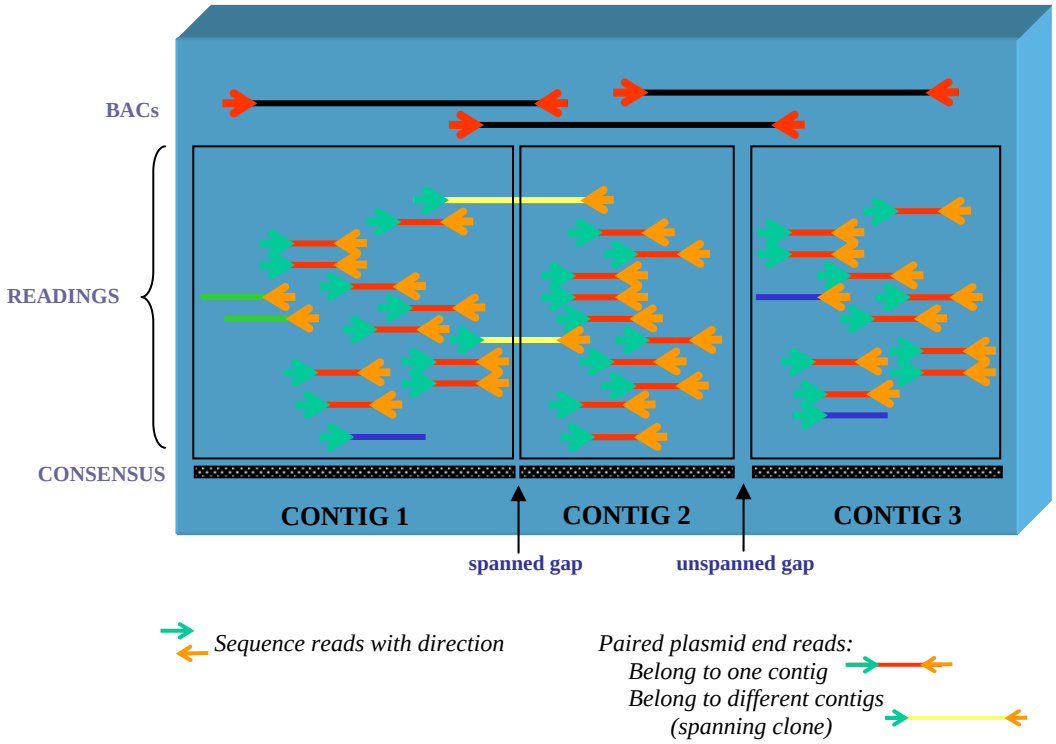


Figure 24. Anatomy of sequence assembly (described in details in a text above). Arrows represent reads, lines connecting arrows represent plasmid insert, line colour code: red – forward, reverse reads in a proper distance in one contig, yellow – forward, reverse reads belong to separate contigs, blue – missing complementary read, light green – complementary read belongs to not depicted contig.

The second type is unspanned (uncaptured) gaps (Figure 24). In case of spanned gaps (contig 1 and contig 2 on Figure 24) the order and the orientation of the contigs were easily established. Unspanned gaps (contig 2 and contig 3, Figure 24.) gave no information about the adjacent contigs, or about the DNA spanning the gap. Contigs

separated with unspanned (physical) gaps were usually spanned by PCR on genomic DNA using primers from each end of the contigs and the PCR products could then be sequenced to close the gaps.

4.1.1.3.1.1. STRATEGY FOR CLOSING THE GAPS IN A SEQUENCE – GAP4

One of the most effective ways of contig ordering and gap filling was the usage of linking clones. The terminal sequences (from either end of the insert) of linking clones belong to different contigs. Mate pair analysis permitted to identify clones that bridge two contigs. Mate pairs are the two end sequences from one single clone; in the assembly they must point towards each other. Contigs were linked, if the orientation of the sequences and the distance to the end of the contig were compatible with the size of the insert.

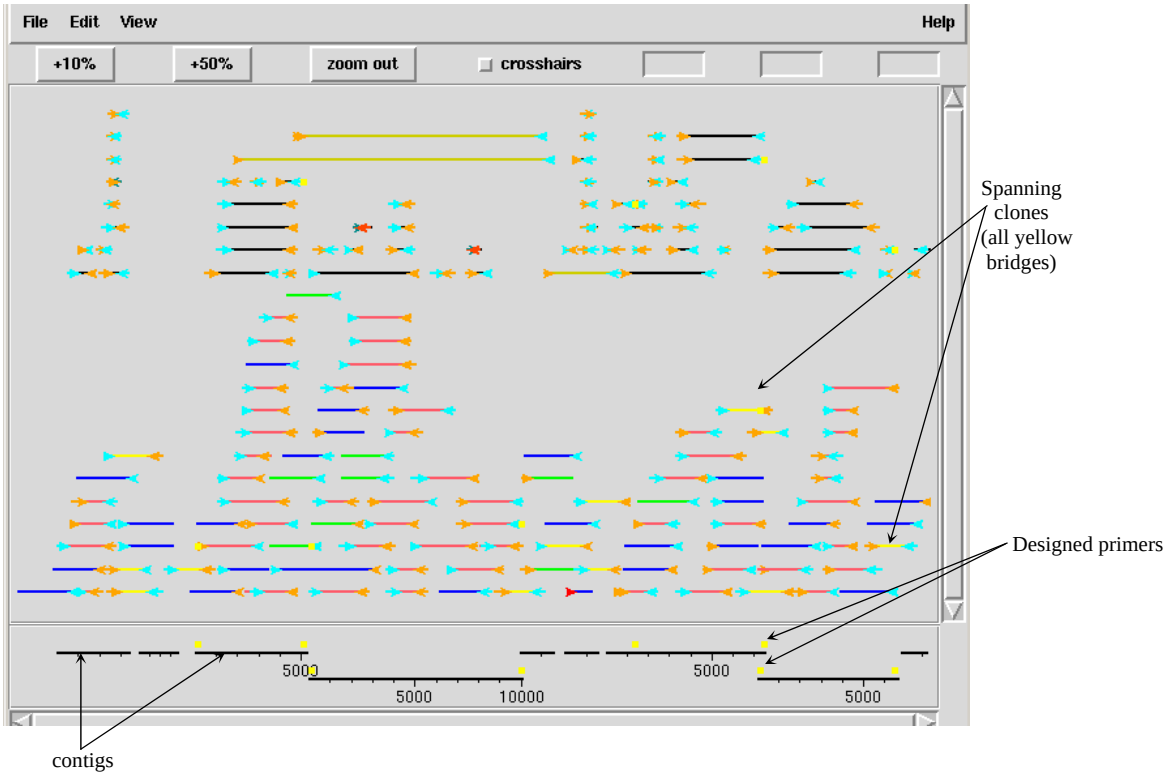


Figure 25. Contig order according to Gap4 program. Arrows represent reads, lines connecting arrows represent plasmid insert, line colour code: red – forward, reverse reads in a proper distance in one contig, yellow – forward, reverse reads belong to separate contigs, blue – missing complementary read, light green – complementary read belongs to not depicted contig.

The Gap4 program was used for the detection of bridging clones. When they were available, gaps were closed by primer walking. Figure 25 shows an example of linking 9 contigs.
The larger the insert the more likely a clone will be a linking one. Therefore a large insert library is essential for gap closure.

4.1.1.3.1.2. PRIMER WALKING

Primer walking is the method of designing internal primers in order to "walk" along the DNA template. The DNA of interest was either a plasmid insert or a PCR product. Sequencing by primer walking is an effective strategy for the complete sequence determination. When the sequencing DNA fragment was between 1,5 to 5 kb series of sequencing reaction were performed. In this case, initial sequence data, obtained using the first pair of primers, was a base for designing the next primers. This process was repeated to complete the sequence of the entire DNA fragment.
After the first round of gap closing and primer walking in the genome project 243 contigs were obtain, 128 of them bigger then 3 kb.

4.1.1.3.1.3. STRATEGY FOR CLOSING THE GAPS IN A SEQUENCE – LARGE FRAGMENT LIBRARY (COSMIDS)

The assembly of the complete genome was largely facilitated due to the availability of a large clone library. It helped in the gap closure phase, allowed to cope with repeat sequences, and permitted to validate the genome assembly. 290 BAC (Bacterial Artificial Chromosome) readings were assembled. The BAC library was provided by Melanie Emmerth (Institut für Molekulare Infektionsbiologie der Universität Würzburg). Contigs number 2 and 3 (Figure 24.) show an example where it is possible to link two contigs without having spanned clones. The BAC library -

derived sequence pairs allowed the mapping of the corresponding clone onto the existing contigs. The contig order was confirmed by PCR.

4.1.1.3.1.4. STRATEGY FOR CLOSING THE GAPS IN A SEQUENCE – MUMMER SOFTWARE

It is well known that the organization of genetic information is conserved between closely related species. This phenomenon was exploited to increase the efficiency of gap closing.

The MUMmer system is useful for aligning whole genome sequences, whether in complete or draft form. [47] The system finds all maximal, unique, matching sequences (MUMs), extracts the longest possible set of matches that occur in the same order in both genomes and can be used for contigs ordering.

Previously described strategies led to 18 contigs bigger then 10 kb. Finally, 18 contigs from the shotgun sequencing project were aligned to the related genome of *E. coli* CFT073 using MUMmer. This alignment allowed to order contigs (Figure 26a); it also revealed major differences between these two genomes, especially with regard to the location and content of four major *E. coli* 536 pathogenic islands as well as CFT073 specific clusters (phage regions, pathogenic islands and other) (Figure 26). Primer walking strategy was applied to confirm the contig order which were shown by MUMmer comparison plots

At that stage of the project it was possible to estimate the extent of genome syntheny- by whole genome DNA alignment. Few genome sequences were used for genome alignments: *E. coli* K12 (NC_000913), *E. coli* O157:H7 EDL933 (NC_002655), *E. coli* O157:H7 (Sakai) (NC_002695), *Shigella flexneri* 2a (NC_004337), *Salmonella typhimurium* LT2 (NC_003197), and *Salmonella enterica subsp. enterica serovar Typhi* Ty2 (NC_004631). Figure 27.

The comparison confirmed the conservation of organization of genetic information, detected numerous large-scale inversions in bacterial genomes, and also indicated “hot-spot” rearrangement loci, where gene transfer could have occurred (pathogenic islands loci).

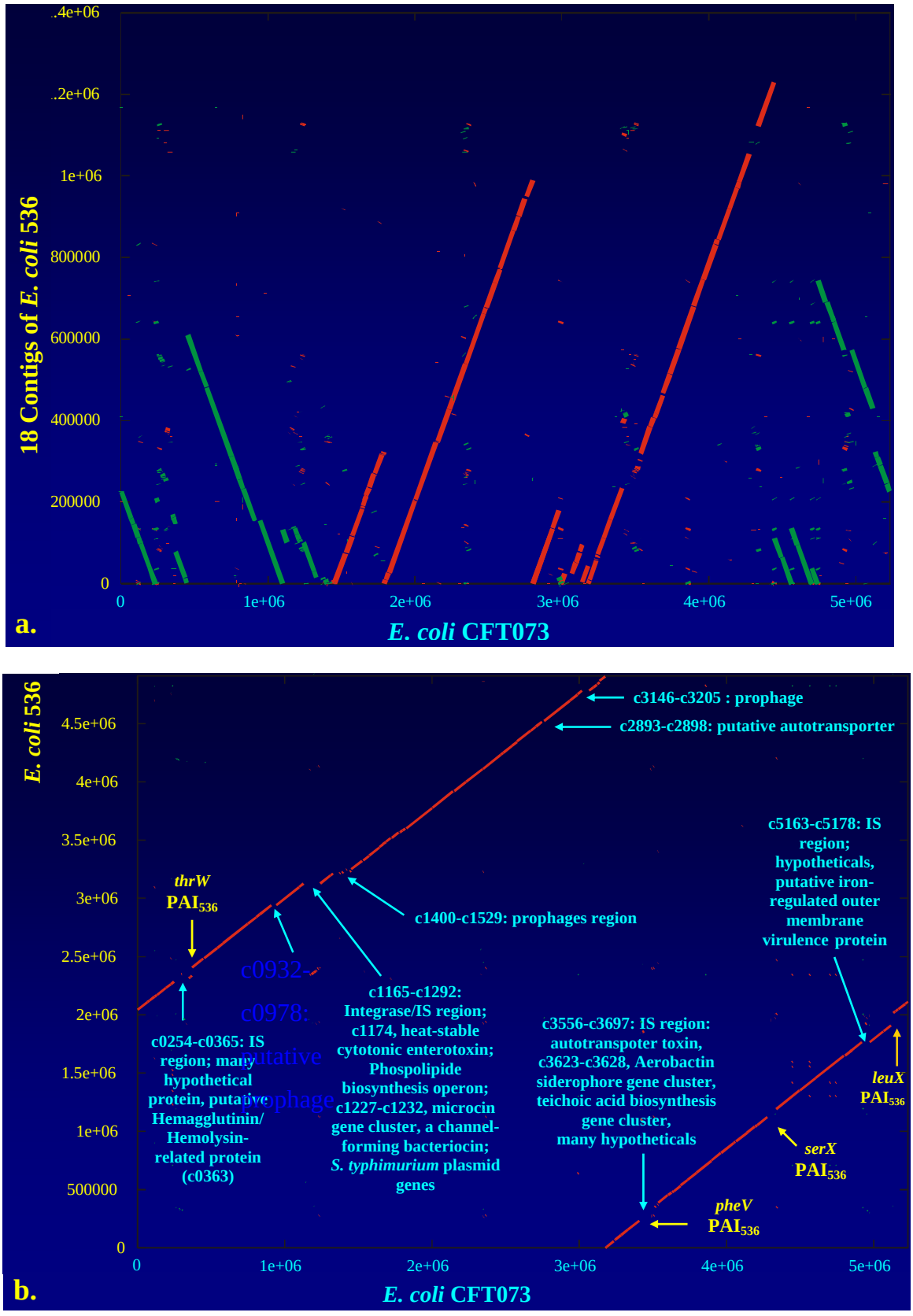


Figure 26. Dot-plot alignment of 18 contigs of *E. coli* 536 to a chromosome of *E. coli* CFT073: a. random order; b. order according to *E. coli* CFT073 genome with identified major differences. Aligned segments are represented as dots. This alignment was generated by the MUMmer an open source package.

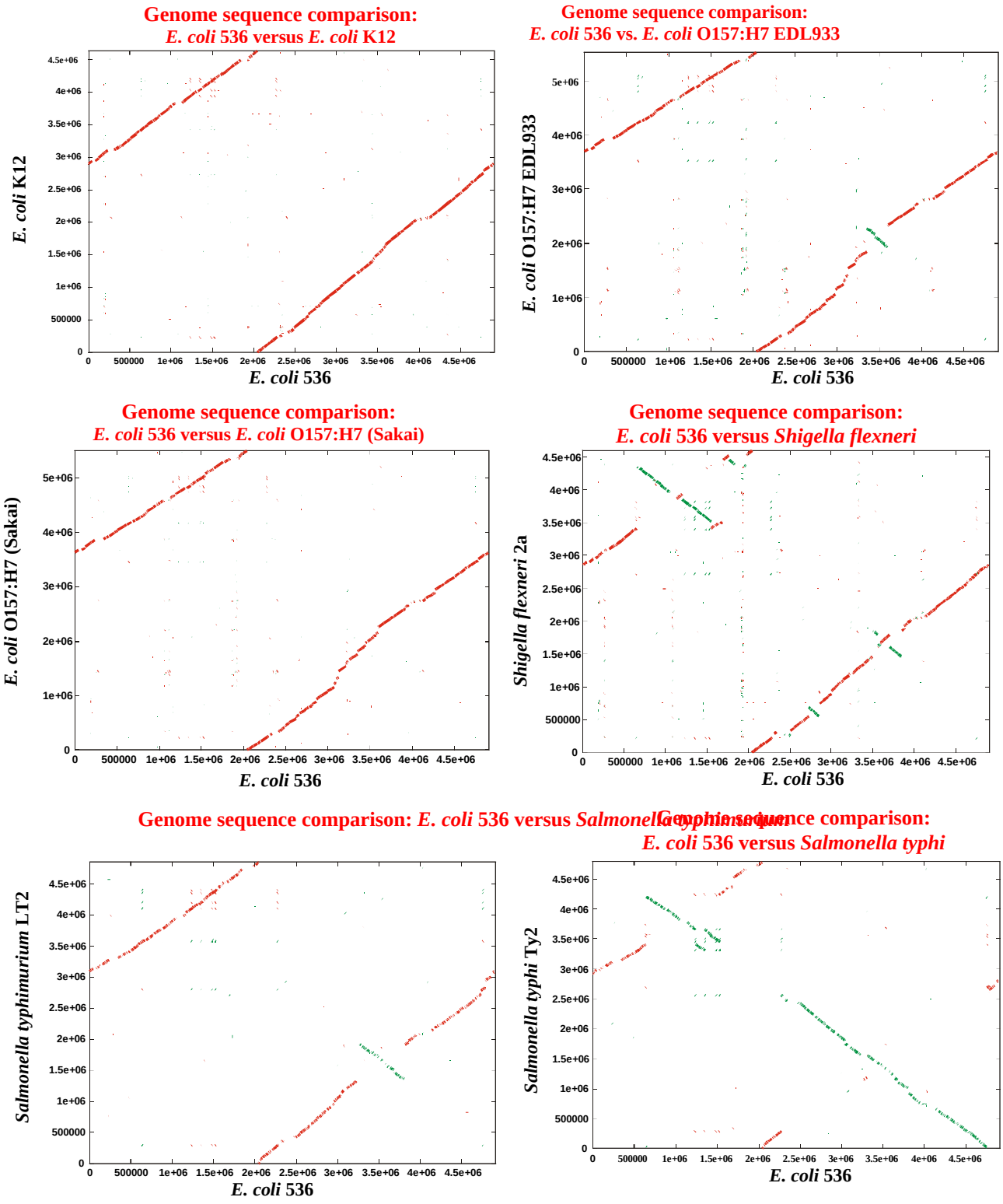


Figure 27. Between-species whole genome DNA alignment. Plot of maximally unique matching subsequences (MUMs) between genomes as identified by the MUMmer program. A dot indicates a DNA sequence that occurs once within each genome, at location x in one genome and y in the other genome. The matching sequences may occur on either the forward (red dots) or the reverse (green dots) strand.

4.1.1.3.2. REPEATS – ASSEMBLY PROBLEMS

Repetitive sequences frequently caused troubles in genome assembly. Defining the unique flanks and mate pairs permitted to unravel the assembly problem, which can be either a misassembly or a gap. Figure 28 illustrates the anatomy of repeat misassembly and the way of problem solution.

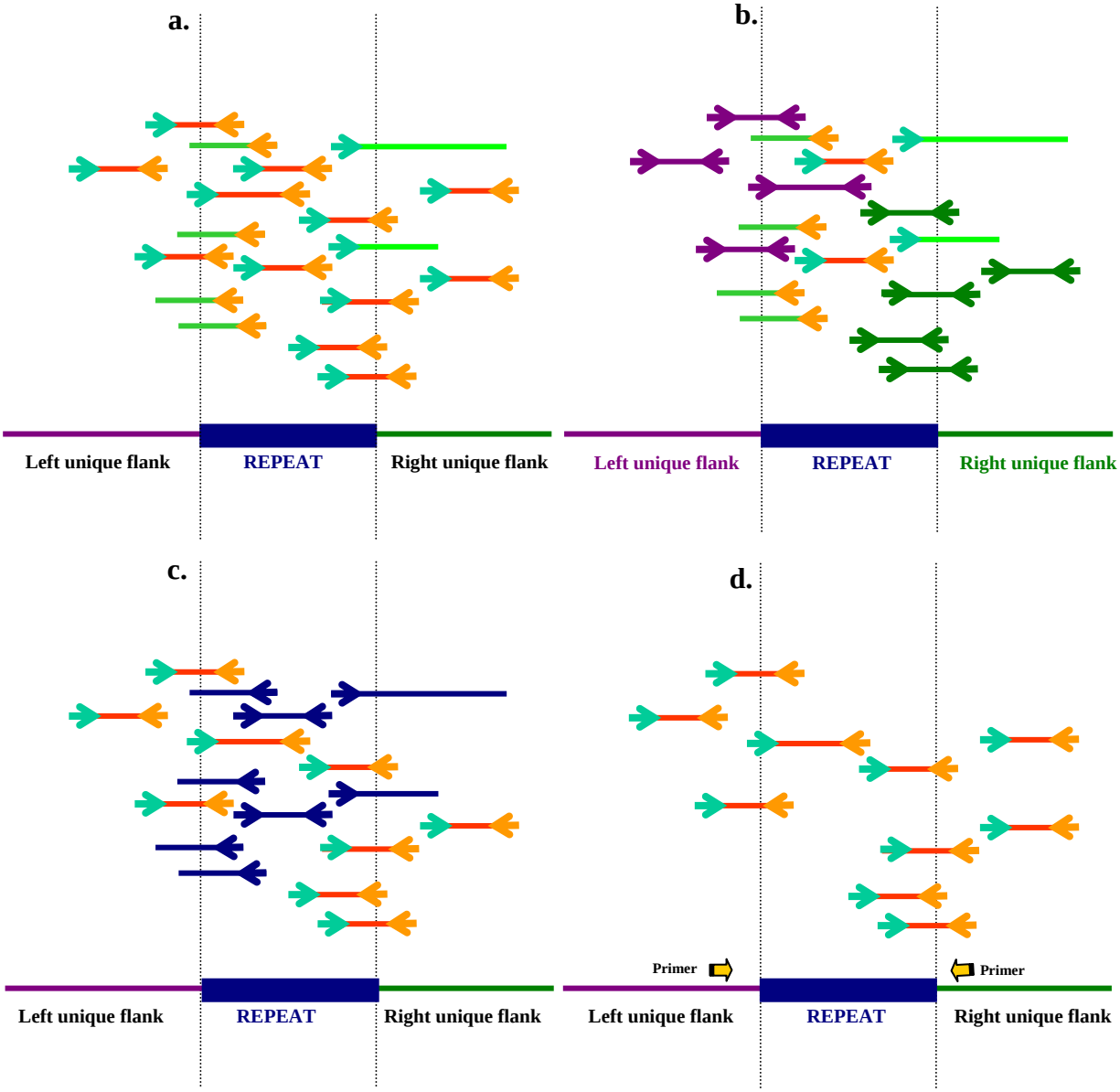


Figure 28. Misassembly caused by repeat. a. identification of unique and repeated sequences; b. mate pair analysis; c. identification of misassembled clones (dark blue); d. PCR amplification in order to prove the sequence. Arrows represent reads and lines bridges between reads: red – forward, reverse reads in a proper distance in one contig, light green – complementary read belongs to not depicted contig.

When the specific readings were identified (readings which definitely belongs to this genome region; Figure 28b; violet and dark green marked clones), it was possible to disassembly all incorrectly joined sequences (Figure 28c; dark blue readings). Depending o the gap size, primers walking sequencing reactions were performed. In case of a gap within a repeat, the procedure was the same (identification of the unique sequence and mate pair analysis and the and of contigs).

The best examples of those kind problematical regions are rRNA operons. The *E. coli* genome carries seven rRNA (rrn) operons. A single rRNA operon contains three rRNA genes: 16S, 23S and 5S. Genes encoding transfer RNA (tRNA) are located in the internally transcribed spacer (ITS) region. The ITS is 355 bp in length. Figure 29 shows a schema of rRNA operon. Assembler program joins all readings within rRNA operon producing misassembly and preventing a proper consensus sequence obtaining.

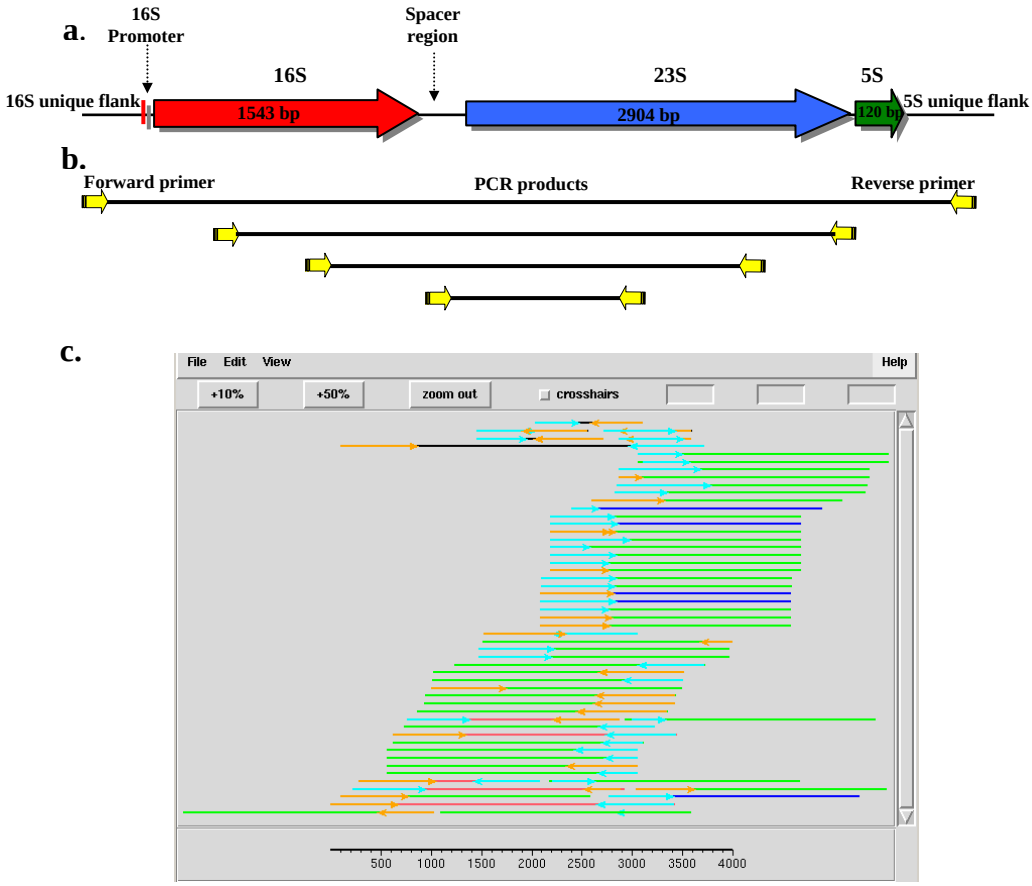


Figure 29. Typical rRNA operon (a.), sequencing strategy (primer walking steps) (b.), and template display from Gap4 program (c.)

4.1.1.3.3. EDITING OF CONSENSUS SEQUENCE

Gap4, the contig building software, takes into account the "quality" of each base in a read (where quality represents the confidence that the base has been called correctly) and can display the trace files. Regions of insufficient sequence quality (base confidence lower then 65%) were identified and solved using primer walking techniques and additional sequencing reactions. A better way of displaying the accuracy of bases is to shade their surroundings so that the lighter the background the better the data. In the figure above, this grey scale encoding of the base accuracy or confidence has been activated for bases in the readings and the consensus. The Figure 30 shows a typical display from the contig editor.

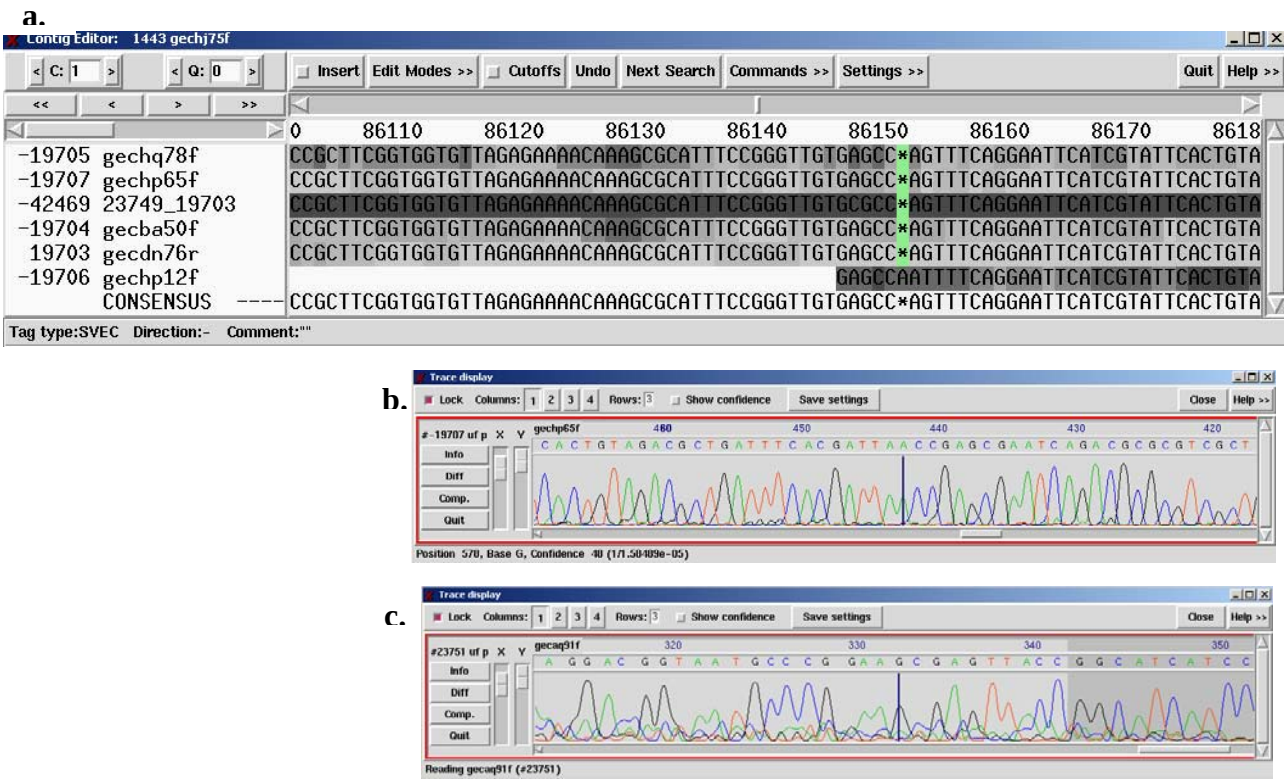


Figure 30. Contig Editor (greyed section represents the currently displayed segment; intensity of the color indicates each base confidence) with traces (a.); high and (b.) c. low quality. The bottom line of the Contig Editor gives the contig sequence (consensus), and above it are the read sequences.

4.1.2. SEQUENCE ANALYZING

4.1.2.1. ORF FINDING

All possible open reading frames ORFs have to be identified. Automatic ORF prediction was accomplished using the YACOP program. The correctness of the predicted protein inventory might be insufficient, especially for regions with a very high GC content (like pathogenic islands). Furthermore, the gene finder has generated many false positives ORFs. The output was verified and edited manually using criteria such as the presence of a ribosome-binding site, codon usage analysis and sequence homology analysis. 4,687 ORFs were identified.

4.1.2.2. ANNOTATION

A microbial genome sequence alone is only raw data – it needs to be interpreted in order to be of any scientific significance. The process of predicting the function of all possible coding sequences (genes) in a genome sequence is known as annotation. Although an annotated genome sequence provides a large amount of important information it is still merely a starting point for characterizing an organism.

Automatic function annotation was done using the ERGO tool (Integrated Genomics, Inc., http://www.integratedgenomics.com).

Manual annotation was done to evaluate the validity of automatic function assignments. All predicted ORF's were verified and modified manually by searching derived protein sequences against public (translated) nucleotide or protein domain databases such as GenBank or PFAM, respectively. All software and databases are listed in Table 10. Manual annotation confirmed full-length sequence homology to avoid problems resulting from partial, distant, or mistaken sequence homology. In case of partial sequence homology, it ensured that only functions related to the homologous region are transferred. It ensured that the function assigned to the homologous protein was experimentally validated and not the result of an incorrect function prediction. In addition, manual annotation identified and annotated complex

genetic, biochemical, and physiological data, e.g. signal transduction cascades, metabolic pathways, transcription units.

For each ORF there were three possible results:

- clear sequence homology indicating function
- blocks of homology to defined functional motifs (these should be confirmed experimentally)
- no significant homology or homology to proteins of unknown function

Many of the ORFs identified could not be assigned to a specific function based on homology. These ORFs were divided into two categories:

- conserved hypothetical proteins – ORFs with no homology to proteins of known function but with significant homology to unidentified ORFs of other species (these ORFs are therefore functionally conserved across numerous species and may represent components of central metabolism and others that have not yet been identified. The more universal the distribution of these ORFs the more likely they have a fundamental role in cellular processes)
- hypothetical ORFs – ORFs without homologues – these are ORFs that have no homology to any known sequences – these may represent genes related to more specific traits of the organism

On one hand, function prediction is an advantage as many of the predicted functions turn out to be correct. On the other hand, function prediction can be a disadvantage. Although many of the predicted functions are correct, a significant fraction of the predicted functions is incorrect. The most prominent problem is overprediction of function based on partial, distant, or mistaken sequence homology. Overprediction or, more precisely, overly specific prediction, occurs during automathed transfer functional information from unreliable best hit. Far-reaching similarity is necessary to make similarity-based annotation sufficiently reliable. For enzymes it is often impossible to predict the specific substrates. The problem is enhanced, as most databases do not permit to discriminate between experimentally validated and merely predicted functions. Incorrectly predicted functions, once added to the database, will be further propagated to homologous sequences.

4.2. THE 536 GENOME SEQUENCE

4.2.1. GENERAL FEATURES

Eventually, the genome sequence of *Escherichia coli* human pyelonephritis isolate 536 was completed and would be compared to other completely sequenced *E. coli* genomes. The genome was sequenced with a six-fold coverage. The 45,800 sequence runs were assembled. The general properties of the sequenced *E. coli* genomes are shown in Table 14.

TABLE 14: GENERAL FEATURES OF *E. COLI* 536 COMPARED TO OTHER SEQUENCED *E. COLI* STRAINS

	536	**CFT073**	**MG1655**	**EDL933**	**EDL933 Sakai [48]**
	O6:K15:H31	O6:K?:H1	K12	0157:H7	0157:H7
CHROMOSOME LENGTH, BP	4,938,875	5,231,428	4,639,221	5,528,445	5,498,450
PLASMIDS	-	-	-	-	pO157 (92,721) pOSAK1 (3,308)
DNA CODING SEQUENCE (%)	88	89	89	88	88
ORFS	4,747 [a]	5,533 (5,379 [a])	4,411	5,361	5,981
G+C CONTENT	50.5%	50.5%	50.8%	50.5%	50.5%
tRNA GENES	82	88	87	100	103
rRNA GENES	22	22	22	22	22
PREDICTED MISC. RNAS [B]	45	51	47	52	13
BACKBONE (%) [A]	77	71	81	67	n.d.
STRAIN SPECIFIC ORFS (%)	374 (8)	867 (16)	406 (9)	1270 (24)	n.d.
PROPHAGES	1	5	10	13	24

[a] using the ORF prediction program YACOP [b] using the Rfam database (http://www.sanger.ac.uk/rfam)

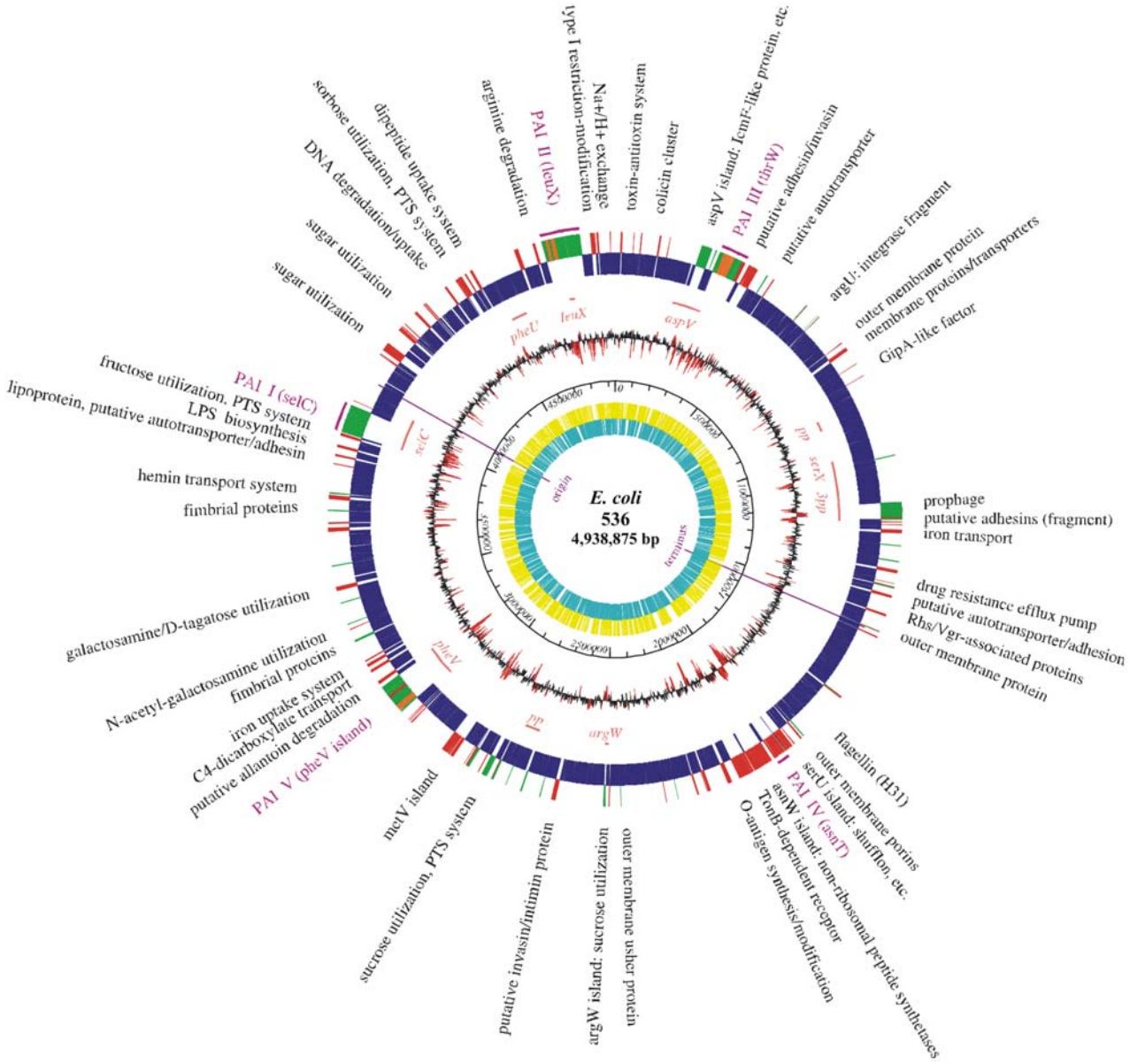


Figure 31: Genetic map of the chromosome of *E. coli* 536. The inner circle represents all putative genes, depending on ORF orientation. The next circle gives the scale. The fourth circle shows the G+C distribution. Thereby, regions with a highly aberrant G+C content (above or below two-fold standard deviation) are highlighted in red. The most outer circles show the results of a three-way genome comparison with the genomes of CFT073 and MG1655: in blue, backbone genes found in all three organisms; in red, genes present in 536 and CFT073 but absent from MG1655; in green, genes found in 536 only; in orange, genes of 536, which are present in CFT073 but located within a different genomic region. Pathogenic islands (PAIs) are highlighted; flanking tRNAs are given in brackets

The single circular chromosome of *E. coli* 536 is composed of 4,938,875 bp. The whole chromosome encodes 4,747 putative genes, no extrachromosomal plasmid was found. The genome of strain 536 contains only one cryptic prophage region, which harbours among structural phage proteins several DNA modification enzymes, such as recombinases and nucleases, as well as an outer membrane porin.
From analyses of the G+C variation throughout the genome it became apparent that divergences from the mean G+C content (50.5%) were found often in genomic regions, which were not present in strain MG1655. These regions include already known PAIs (pathogenic islands) as well as smaller clusters, which are often flanked by mobile elements. (Figure 31)

In order to visualize the differences in *E. coli* genomes sequenced so far a four-way comparison was done (Figure 32), revealing a highly mosaic genome structure. It is apparent that genomic differences are not exclusively linked to the presence or absence of large PAIs. Instead, also the distribution of smaller gene clusters seems to confer strain-specific traits.

4.2.1.1. INSERTION ELEMENTS

Insertion sequences (IS) are mobile genetic elements, usually less than 2.5 kb in size, that are widely distributed in the genomes of most bacteria. IS elements are commonly defined as carrying only the genetic information related to their transposition and its regulation, in contrast to transposons that also carry genes involved in other functions (e.g., antibiotic resistance). IS activities cause mutations, promoting genetic diversity and sometimes adaptation. They are suggested to be involved in chromosomal rearrangements including the incorporation of foreign DNA.
Seven types of IS elements were found in the 536 genome (Figure 32) (IS Finder; http://www-is.biotoul.fr/. [49] Interestingly, the 536 genome possess less IS in comparison to other pathogenic strains as well as laboratory strain MG1655 (Figure 33). Table 15 shows a list of found IS for five complete *E. coli* genomes. Most of 536 IS belong to the IS3 and IS21 families.

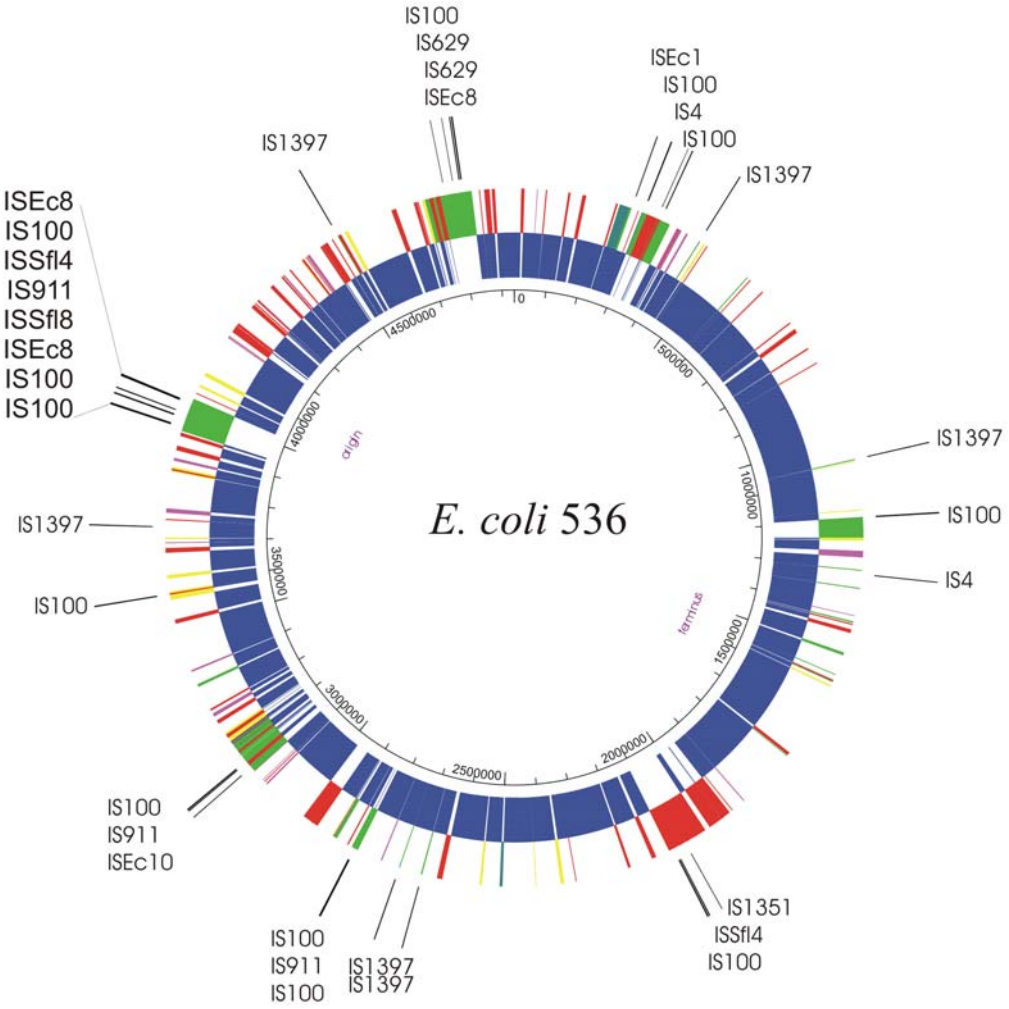


Figure 32. Distribution of IS found in the 536 *E. coli* genome. The inner circle gives a scale, The most outer circles show the results of a three-way genome comparison with the genomes of CFT073 and MG1655: in blue, backbone genes found in all three organisms; in red, genes present in 536 and CFT073 but absent from MG1655; in green, genes found in 536 only; in orange, genes of 536, which are present in CFT073 but located within a different genomic region.

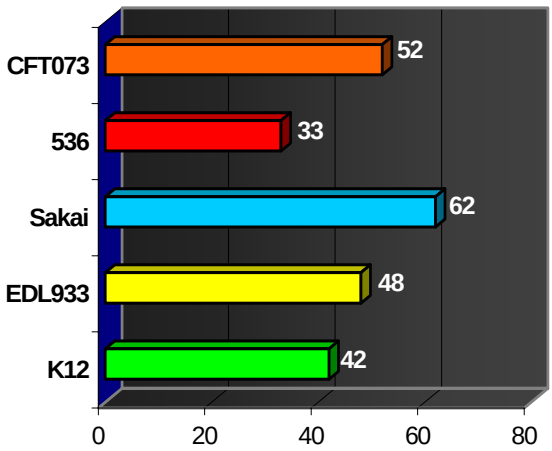


Figure 33. Total number of IS elements present in *E. coli* genomes – comparison.

One of the largest families of ISs is the IS3 family, which is widely distributed in both Gram-negative and Gram-positive bacteria. All *E. coli* genomes encode several types and a significant copy number of IS elements belongs to this family.

This is in contrast to IS21 distribution. Although members of the IS21 family are widespread, they have been found in RP4, in *Shigella sonnei*, various clinical isolates of *E. coli*, in the EHEC plasmid p0157, flanking a high pathogenicity island in *Y. pestis Y. pseudotuberculosis*, and in the *Y. pestis* virulence plasmid pCD1, but they are absent from the genome of both 0157:H7 and MH1655 strains. [49] UPEC strains differ significantly in number of IS21 copies. There are 13 copies in the 536 genome and 7 in case of CFT073. ISSfl18 and ISEc1 were found only in the 536 genome

TABLE 15. IS FAMILY MEMBERS [COPY NUMBER] IN *E. COLI* GENOMES.

Total number of copies is indicated between square brackets together with those that are only partial copies, indicated by a lower case "p".

IS FAMILY	MG1655	O157:H7 EDL933	O157:H7 SAKAI	536	CFT073
IS1	1[7]	1 [2]	1[4/2p]	0	1[1/1p]
IS110	0	0	1[1/1p]	1[2]	1[1]
IS1111	0	0	0	1[1]	0
IS200*	0	0	0	0	19
IS21	0	0	0	2[13/3p]	2[7]
IS256	0	0	1[4/3p]	0	0
IS3	5[16]	3[39/13p]	5[44/6p]	4[12/2p]	6[15/3p]
IS30	1[4]	1[1]	1[4/3p]	0	1[1]
IS4	2[4]	1[1/1p]	1[1/1p]	1[2/2p]	1[3/2p]
IS5	1[11]	0	0	0	0
IS605	0	1[4]	1[3]	0	0
IS630	0	1[2/2p]	0	0	1[1]
IS66	0	0	0	1[2/1p]	1[4/1p]
ISAs1	0	0	0	1[1]	0
ISNCY	0	0	1[1/1p]	0	0

* In the case of IS200 only orfA is present.

4.2.2. COMPARATIVE GENOMICS

The genome sequence of 536 was searched for putative genes: 4,747 coding sequences were predicted. The four-way comparison revealed 3650 (77%) "backbone" genes, which have highly similar orthologues in MG1655 (Figure 33 and Figure 35). These "backbone" genes are present in all *E. coli* strains and are essentially co-linear. However, the backbone sequence is interrupted by numerous segments of strain-specific DNA, bearing features of genomic islands, mobile elements or phages (Figure 33, 35).

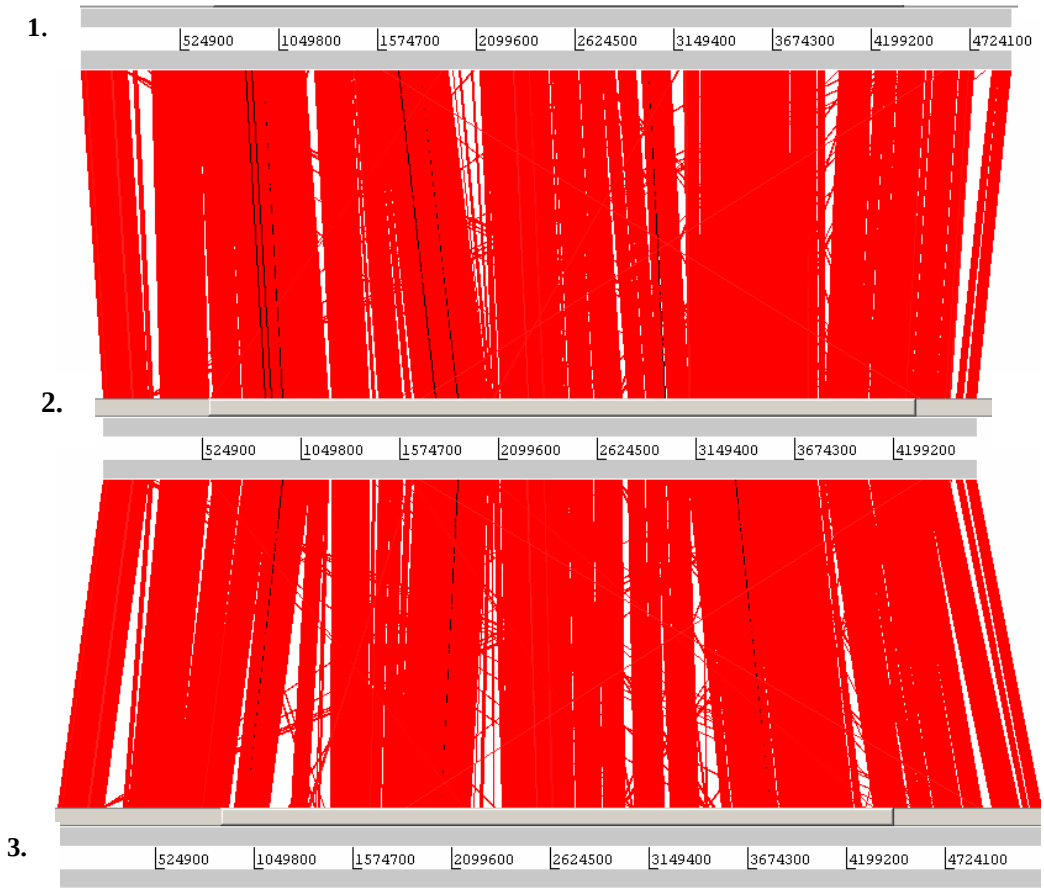


Figure 34. Linear DNA sequence comparison of whole genomes of *E. coli* 536 (1.) MG1655 (2.), and CFT073 (3.). Red blocks represent corresponding regions with a high similarity (98% or more). White spaces indicate strain differences

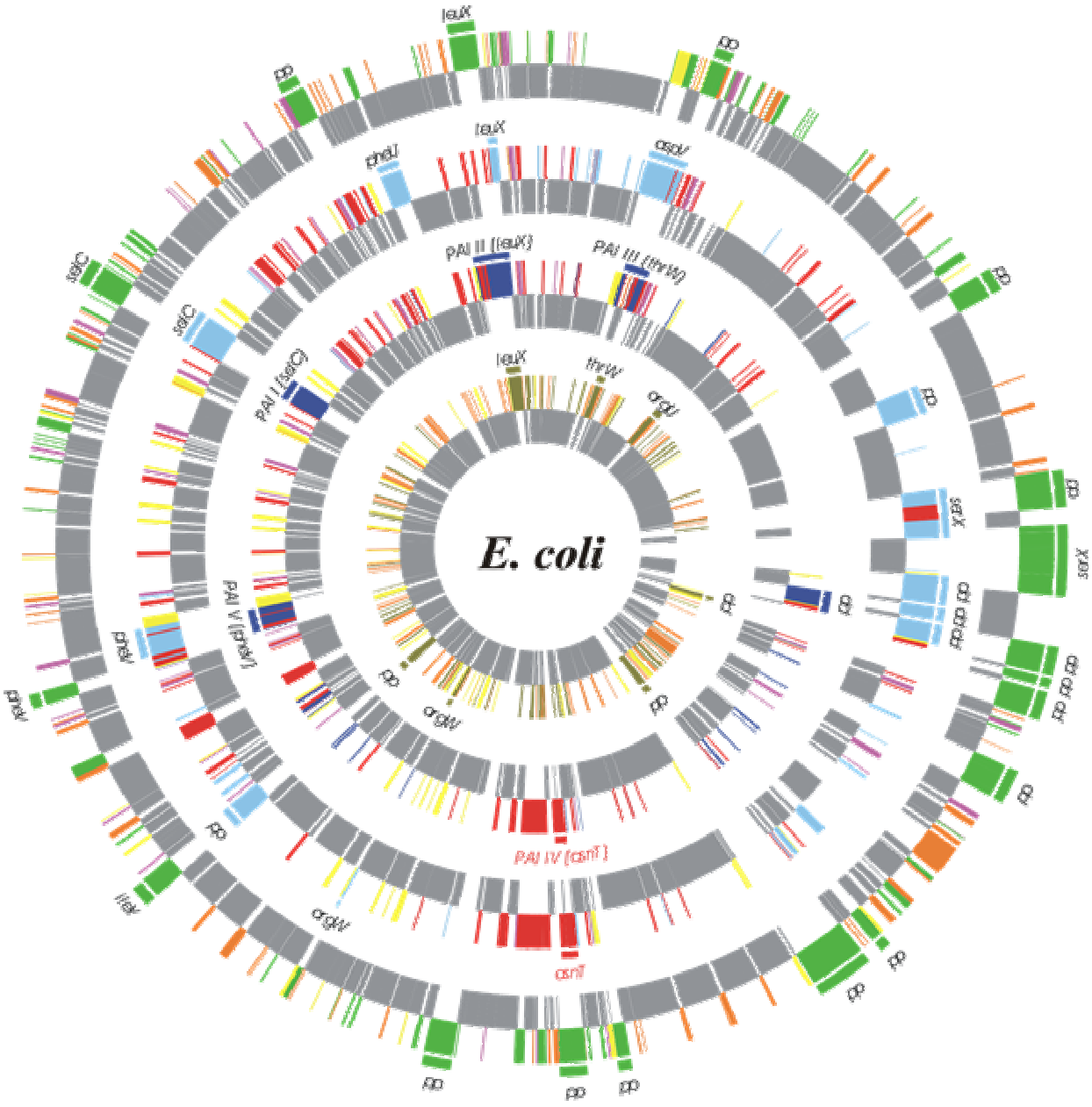


Figure 35. Four-way comparison of *E. coli* genomes. The most inner ring represents all putative genes of the genome of strain MG1655, the second and third rings show all ORFs of the uropathogenic strains 536 and CFT073, respectively. The most outer ring represents the ORFs of the genome of EDL933 (For simplicity reasons the genome of the Sakai strain has been left out). Larger islands are named according to their flanking tRNA genes. In grey, genes present in all four genomes (*E. coli* backbone); in green, light blue, dark blue, dark yellow, genes specific to the respective genome; in red, genes specific to both uropathogenic strains; in pink, genes present only in the three pathogenic strains; in orange, genes present only in MG1655 and EDL933; in yellow, others (e.g. genes present in two genomes only). Abbreviations: pp, prophage (smaller prophages in MG1655 are not named), PAI, pathogenic islands

From the remaining ORFs, 563 are present also in CFT073, which means that 90% of all ORFs of *E. coli* 536 have got highly similar orthologues in the genome of CFT073 (Figure 35.). The remaining 472 ORFs are located mainly within a region of a cryptic prophage and within the major PAIs. 374 of these ORFs are strain 536-specific, whereas the remaining 100 ORFs have orthologues in the genome of EDL933.

The comparative analysis shown, that genomic differences are not exclusively linked to the presence or absence of large PAIs. Apart from the large pathogenic islands, other regions within the genome of *E. coli* 536 were identified, which are absent from MG1655. Table 19 summarizes the location and functional category of some genes or gene clusters found in such regions. Generally speaking, the majority of the regions encode proteins, which have either an implication on the provision or modification of cell membrane/surface components; putative virulence factors like adhesins or are involved in specialized metabolic activities (sugar utilization systems, mechanisms involve in pH homeostasis). Discussed below.

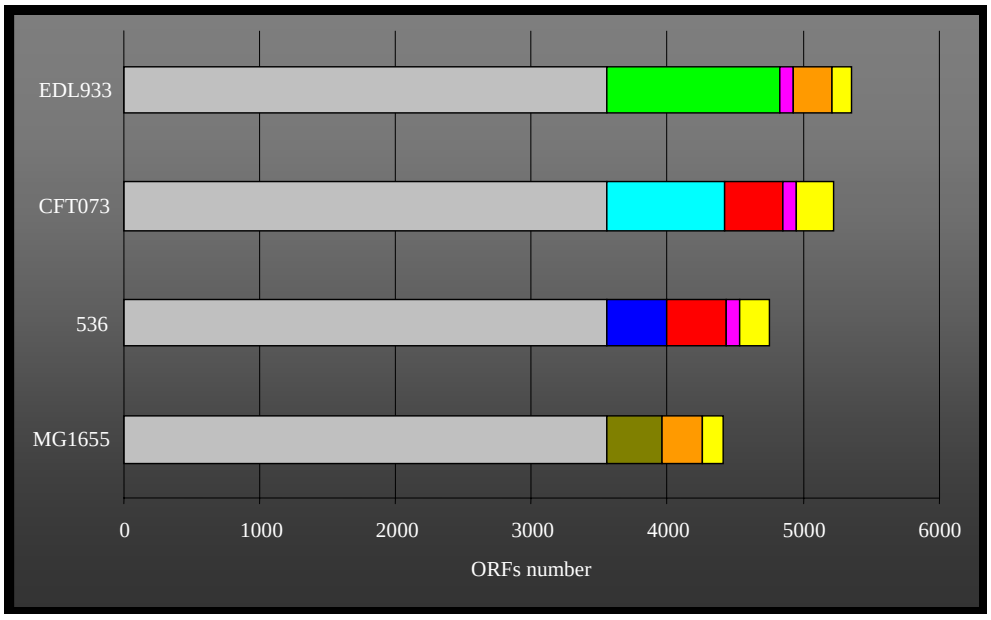


Figure 35. Occurrence of homologous ORFs in *E. coli* MG1655, 536, CFT073 and EDL933.

Colour code: in grey, genes present in all four genomes (*E. coli* backbone); in green, light blue, dark blue, dark yellow, genes specific to the respective genome; in red, genes specific to both UPEC strains; in pink, genes present only in the three pathogenic strains; in orange, genes present only in MG1655 and EDL933; in yellow, others (e.g. genes present in two genomes only).

Comparative genomics with the known *E. coli* genome sequences revealed a lower number of 536-specific genes (374 OFRs) than in the cases of other sequenced *E. coli* genomes. Moreover, we could identify a number of genes present in all pathogenic strains; a large fraction of genes is present only in the UPEC strains. (Figure 33, 35, 36) The genome comparison of three pathogenic and one non-pathogenic strain of *E. coli* revealed only 92 genes, which are present exclusively in all three pathogenic *E. coli*. These genes, arranged in small gene clusters that are scattered around the genome, can be classified in several heterogeneous functional categories such as transport (35 genes), iron acquisition (17), drug resistance (5), DNA modification (4), and others. No main virulence determinants are among these proteins. Rather, they seem to have an implication in improving the pathogens' fitness with respect to e.g. utilization of additional nutrients. Apparently, no broad 'pathogenic backbone' has evolved in disease-causing *E. coli* isolates.

4.3. SEQUENCING AND ANALYSIS OF PLASMID pARK2 ENTEROTOXIGENIC *ESCHERICHIA COLI* STRAIN 2173

4.3.1. INTRODUCTION

Enterotoxigenic *Escherichia coli* strains (ETEC) are the major cause of diarrhoeal diseases in humans and newborns and weaned piglets. ETEC produce enterotoxins. Several types of *E. coli* enterotoxins have been characterized [50]:

- Heat-stable enterotoxins (ST):
    - Heat-stable enterotoxin I (ST-I). There are two slightly different toxins under this designation: ST-Ia and ST-Ib. Both have a molecular weight of around 2000, are resistant to 100°C for short times, resistant to proteolytic enzymes, as well as resistant to acids. ST-Ia and ST-Ib share a sequence motif of 11 amino acids but differ in their N-terminal sequences. Six cysteine residues are involved in disulfide bridges. They are antigenically related. ST-I production is under control of the catabolite repression, being optimally produced in the absence of glucose. ST-I binds to a heterogeneous group of glycoprotein receptors on the brush borders of intestinal epithelial cells. The genes for ST-I are generally plasmid-encoded, often associated with genes for fimbriae, drug resistance and colicinogeny. While porcine, bovine and human ETEC produce ST-Ia only human isolates produce ST-Ib in addition.
    - Heat-stable enterotoxin II (ST-II). The mechanism of action of this 48amino acid long peptide with two disulfide bonds is at present not known. Although ST-II causes observable fluid responses in an experimental mouse model, the production of ST-II appears restricted predominantly to porcine ETEC, with occasional reports of ST-II producing strains isolated from humans with diarrhoea. The ST-II genes are generally plasmid-encoded and often found to be associated with genes for ST-I and the heat-labile enterotoxin LT.

- Heat -abile enterotoxins (LT)

    - Heat-labile enterotoxin I (LT-I) is the best studied heat-labile enterotoxin and often referred to as just LT. LT is an oligomeric protein of about 88kDa comprising one A subunit of 30kDa above a central aqueous channel formed by the five B subunits each 11.5 kDa in size. The doughnut-shaped B pentamer is a very stable structure. Two disulfide bonds, which are part of the B subunit structure, are an integral part of the molecule. The A and B subunits are synthesized separately. The A subunits are required for the formation of the B pentamer, which also requires the disulphide bonds of the *B subunits*. The last four residues of the C-terminal sequence of the A subunit are essential for the stability of the final toxin. Protein assembly occurs in the periplasmic space from pre-A and pre-B subunits, separately transported across the cytoplasmic membrane. Iron starvation and the presence of bile salts and trypsin in the concentrations normally present in the intestine stimulate the release of LT from *E. coli*. While the structural genes of LT-1 are on plasmids, chromosomally encoded proteins may affect levels of expression. These plasmids additionally harbour resistance factors to antimicrobial agents as well as genes coding for the adherence factors associated with ETEC. There is a strong homology between LT and cholera toxin (CT), suggesting that the LT genes were derived from *Vibrio cholerae*. LT-I and CT are also closely related immunologically.

    - Heat-labile enterotoxin II (LT-II): There are two serologically distinguishable variants LT-IIa and LT-IIb. Both are not neutralised by anti-CT or anti-LT-I. Structurally, LT-II is similar to LT-I but its receptors are slightly different. The enzymatic activity of LT-II is similar to LT-I. Structural genes for LT-II are located on the chromosome and have been detected in *E. coli* isolates from human and bovine sources.

Many ETEC strains have fimbrial structures on their surface, whose genes are, like the ones for LT, ST enterotoxins, plasmid-mediated. These fimbriae are termed colonization antigens (K88, K99, F41 and P987), and enable the bacteria to colonize the epithelial surface of the small intestine.

The complete DNA sequence of plasmid pAKR2 of ETEC strain 2173 is presented here. Plasmid pARK2 is a result of a transposition experiment with the wild-type plasmid pTC used as a "transposition-donor" plasmid and vector pACYC177 (Amp^R, Km^R, *p15ori*) – recipient. [51] Plasmid pTC (*sta*$^+$, *stb*$^+$, Tet^R) was isolated from strain *E. coli* 2173 (O147, STa^+, STb^+, $F18ac^+$). *E. coli* 2173 is the representative strain of porcine post weaning enterotoxigenicity.

4.3.2. SEQUENCING OF THE PARK2 PLASMID

To sequence the pARK2 plasmid, the same shotgun approach was used as in the case of the *E. coli* 536-genome project. The DNA was first randomly broken into fragments using the nebulizer technique and size fractioned by sucrose density gradients in the range of 1.0 to 4.0 kb. Purified, end repaired fragments were subcloned into the cloning vector pCR4Topo (Figure 18) forming a random library. Sequencing data were collected on ABI 3730xl sequencer. The 1914 sequence runs were assembled and the resulting contigs were ordered. Remaining gaps were closed using PCR-based methods. The plasmid was sequenced with a 9-fold coverage.

4.3.3. SEQUENCE ANALYZING

4.3.3.1. ORGANIZATION AND PHYSICAL PROPERTIES

The complete sequence of the 88,480 bp *E. coli* 2173 pARK2 plasmid was deciphered. Figure 36 represents a circular representation of the pARK2 genome. The sequence was searched for potential open reading frames: 104 ORFs were identified, most of them are encoded on the same DNA strand. The coding regions account for

83.6% of the plasmid. Analysis of base composition showed the G+C content of the plasmid is 51%, which is a little higher than in *E. coli* 536 (50.5%). pARK2 comprises a mosaic structure of potential pathogenesis-associated genes, IS elements, maintenance genes, transfer genes and unknown ORFs.

Plasmid pARK2 consists of two large distinct regions. The first is 71.7 kb large and is essentially homologous to the non-R-determinant region of plasmid R100, except of several point mutations and a few insertion elements (hypothetical elements RARK0008, RARK0009, RARK0010) and the plasmid stability inheritance proteins StbA and StbB (RARK0087, RARK0086). This region includes transfer genes, the origin of replication and genes involved in plasmid maintenance and stability.

The second region is 16.7 kb in size, which is an extended version of the previously described 10 kb fragment designated toxin-specific locus (TSL) [2]. TSL carries the *sta* and *stb* genes encoding heat-stable enterotoxins. Interestingly, in contradiction to previous data the distance between *sta* and *stb* genes was established as 9,570 bp and not 4,047 bp. The previously determined TSL locus is disrupted by a 5,523 bp IS coding region (TSL1 and TSL2 on Figure 36). The whole TSL locus corresponds to a 16,765 bp fragment consisting of 25 ORFs, five of which are hypothetical proteins with no significant homology to any known protein, two encode toxins and eighteen ORFs are known and newly identified insertion elements (IS Finder; http://www-is.biotoul.fr/). [49] Directly downstream of the *sta* gene two IS elements were identified. Both belong to the well-known class IS21. One corresponds to IS640 from *Shigella sonnei*. The second is 98 % identical to IS21, which was first observed in the plasmid R68.45 as a direct duplication of a 2 kb fragment of the original R68 plasmid. It is involved in conduction of the bacterial chromosome and the pTi plasmid, via "co-integrate formation", but also promotes adjacent deletions. Upstream of the *sta* gene, IS203, IS1 and three copies of IS91 were identified. IS1 is present in about 14% of the *Salmonella* strains tested but in more of 90% of *E. coli*. [52] IS91 produces both cointegrates (10-6) and simple insertions (10-4). IS91 always inserts with the same orientation into its specific targets. IS91 ORF is 36 % identical to that of IS801, and they are both related to the rolling-circle-type replication proteins of the pUB110 family of plasmids, with which they share 4 conserved motifs.

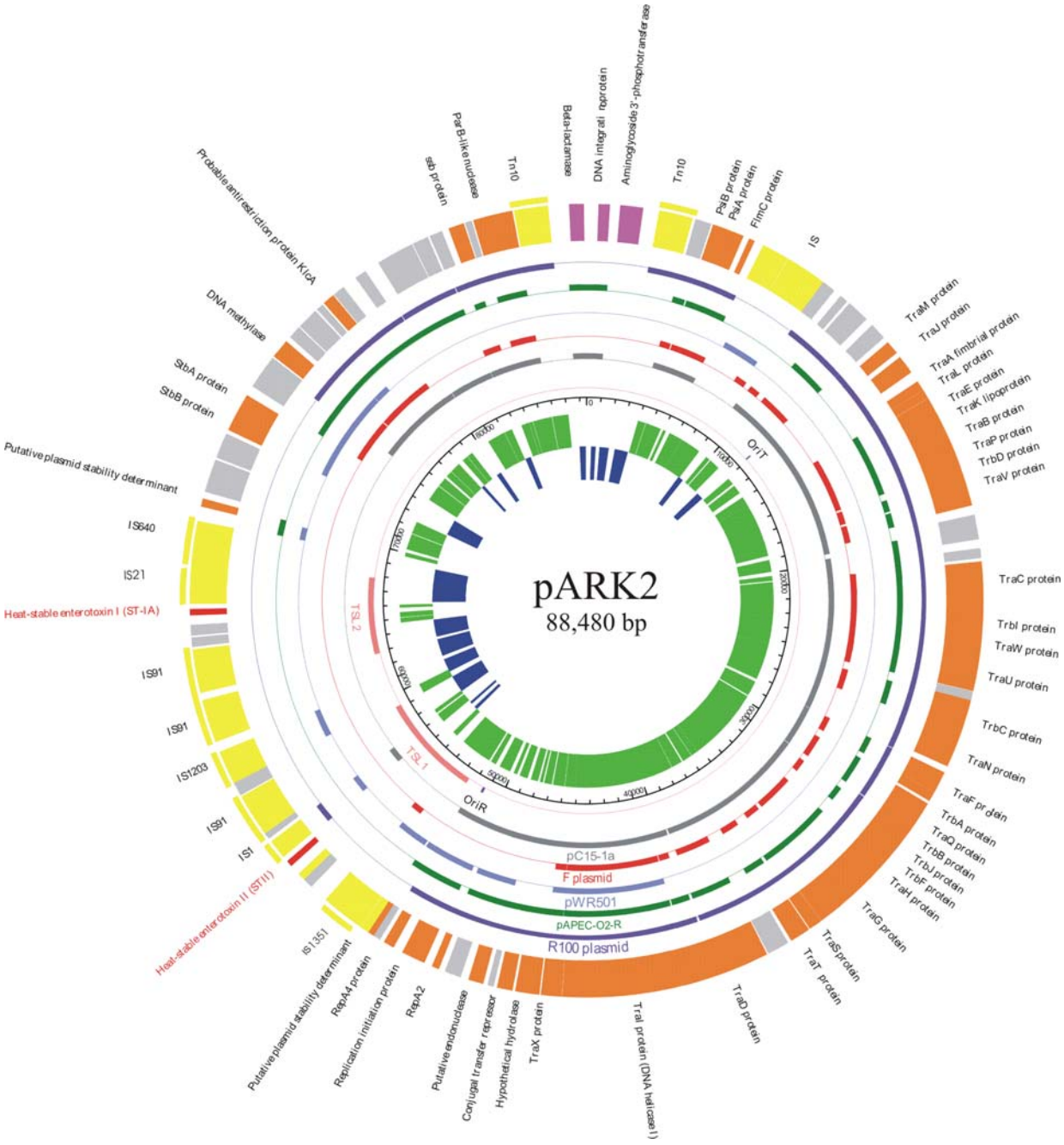


Figure 36. Circular representation of the pARK2 genome. Outer ring depicts all identified ORFs, color-coded according to functional categories: yellow – IS elements (known IS indicated by additional yellow block above); orange – genes identical or with essentially similar to known proteins; red – toxins encoding genes; grey – ORFs with unknown function. Next five rings indicate regions with high (at least 95%) identity to selected plasmids: violet – R100 plasmid, green - pAPEC-O2-R, pale blue - *Shigella flexneri* large virulence plasmid pWR501, red – F plasmid, grey – pC15-1a. Next ring shows positions of origin of replication oriR, origin of transfer oriT and localisation of toxin-specific locus (pale pink) TSL1 and TSL2 (described in a text). The eighth ring is a scale bar (in base pairs). The most inner ring represents all ORFs with their orientations.

Directly upstream of the *stb* gene, a 92 aa polypeptide was predicted. This ORF (RARK0062) shows 96% (83/86) identity to IS1351, which belongs to IS3 family originally found in *Salmonella enteritidis*).

4.3.3.2. ANNOTATION

Of the 104 potential ORFs, 35 (34%) encode genes required for conjugation transfer (a 33 kb *tra* region). Figure 37 visualises functional classification of pARK2 ORFs. Twenty-two ORFs (21%) are IS elements. Together, twenty-four putative proteins had insufficient homology to known protein to be assigned a putative function. These constitute the unknown ORFs, of which 16 (15%) had significant similarity to conserved hypothetical ORFs, and 8 had no significant similarity to any proteins in the databases. They form a group of strain specific proteins of unassigned function (Table 16).

In addition to well-characterized ORFs from the *tra* region, twenty ORFs (19% of pARK2 OFRs) showed significant homology to proteins of known function.

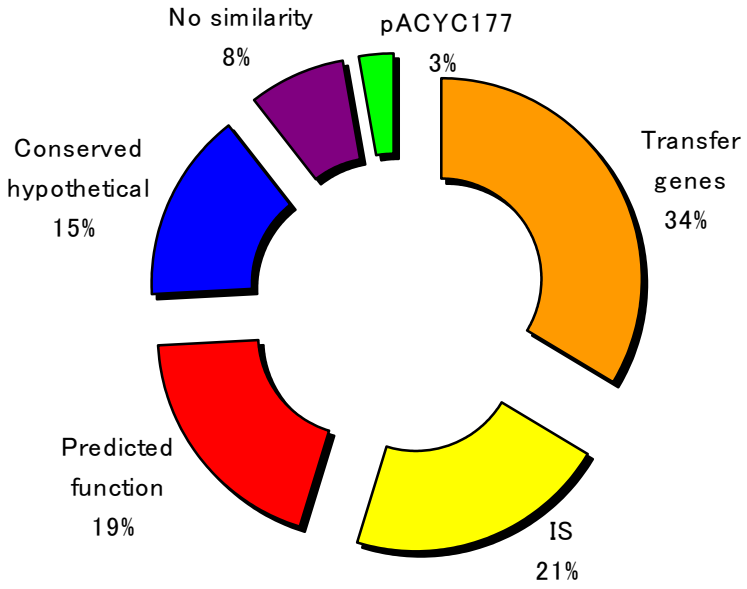


Figure 37. Functional classification of pARK2 ORFs.

TABLE 16. OFR OF UNKNOWN FUNCTION

ORF	POSITION	FIRST CODON	NO. OF AA	ORGANISM OF PROTEIN TO WHICH MOST SIMILAR	HOMOLOG PROTEIN	EXTENT OF SIMILARITY
RARK0004	3,927 - 4,499	gtg	191	*E. coli*, plasmid pC15-1a	YefA	178/191 (93%)
RARK0011	9,486 - 9,004	atg	161	*Shigella flexneri*, plasmid R100	YehA	108/125 (86%)
RARK0012	9,787 - 10,080	atg	98	*Shigella flexneri*, plasmid R100	YeiA	98/98 (100%)
RARK0013	10,194 - 11,012	atg	273	*Shigella flexneri*, plasmid R100	32 protein	273/273 (100%)
RARK0025	19,045 - 19,455	atg	137	*Shigella flexneri*, plasmid R100	YfhA	137/137 (100%),
RARK0026	19,451 - 19,921	atg	157	*Shigella flexneri*, plasmid R100	YfiA	129/157 (82%),
RARK0027	20,250 - 20,594	atg	115	*Shigella flexneri*, plasmid R100	YfiC	115/115 (100%),
RARK0032	25,378 - 25,686	atg	103	*Shigella flexneri*,, plasmid R100	YgeA	103/103 (100%),
RARK0045	36,743 - 37,543	gtg	267	*Shigella flexneri*, plasmid R100	YhfA	224/260 (86%),
RARK0051	47,578 - 47,787	atg	70	*Shigella flexneri*, plasmid R100	YigA	70/70 (100%),
RARK0053	48,789 - 49,376	atg	196	*Shigella flexneri*, plasmid R100	YihA	173/173 (100%)
RARK0057	51,949 - 52,215	atg	89	*E. coli*, plasmid pAPEC-O2-R	YacA	85/89 (95%),
RARK0061	55,031 - 55,396	atg	122	No homology		
RARK0066	57,310 - 57,597	atg	96	No homology		
RARK0076	64,715 - 65,041	gtg	109	No homology		
RARK0077	65,118 - 65,489	gtg	124	No homology		
RARK0084	70,285 - 71,469	atg	395	No homology		
RARK0088	74,397 - 75,320	atg	308	*Salmonella typhimurium*, plasmid pU302L		289/308 (93%),
RARK0089	75,287 - 75,628	gtg	114	*E. coli*, plasmid pAPEC-O2-R		110/114 (96%),
RARK0091	76,593 - 77,057	atg	155	*Salmonella typhimurium*, plasmid pU302L		143/144 (99%),
RARK0092	77,111 - 77,884	atg	258	*Shigella flexneri*, plasmid R100	YchA	221/257 (85%),
RARK0093	78,219 - 77,968	atg	252	*E. coli*, plasmid pAPEC-O2-R		28/46 (60%),
RARK0095	78,772 - 79,191	atg	140	*E. coli*, plasmid pAPEC-O2-R		140/140 (100%),
RARK0096	80,156 - 79,737	ttg	140	*E. coli*, plasmid pAPEC-O2-R		138/140 (98%),
RARK0097	80,702 - 82,060	atg	453	Plasmid R100, pC15-1a, pAPEC-O2-R, pU302L	YdbA	436/453 (96%),
RARK0098	82,110 - 82,670	atg	187	*Shigella flexneri*, plasmid R100	YdcA	187/187 (100%),
RARK0099	83,220 - 82,762	atg	153	*E. coli* ,plasmid pC15-1a	YddA	122/153 (79%),
RARK0101	84,101 - 84,337	atg	79	*Shigella flexneri*, plasmid R100	YdeA	79/79 (100%),

Among proteins with an assigned function are those, which are involved in plasmid replication, stabilisation and maintenance:

- Three ORFs (RARK0054, RARK0055, RARK0056) code for the replication regulatory protein RepA2, the replication initiation protein, and the RepA4 protein, respectively. RepA2 protein is essential for plasmid replication; it is involved in controlling the plasmid copy number. Initiation of DNA replication requires the presence of a replication initiator protein and of a *cis*-acting DNA sequence, the origin of replication (*oriR*). The binding of the initiator protein to *oriR* is the first step in DNA replication. The *oriR* of the pARK2 plasmid was proposed to lie within a 150-bp sequence located 339 bp downstream of the coding sequence of the replication initiation protein (RARK00055) and 27 bp upstream of RepA4 protein (RARK00056). Analysis of the origin of replication revealed that pARK2 belongs to the IncFII group of self-transmissible plasmids. Members of this incompatibility group include R1, R6-5, pC15-1a, pTUC100, pO157, p1658/97, pWR100, pAPEC-02-R and R100 plasmids. These plasmids have a low copy number of 2-5 plasmids per cell.

- Additionally, the conserved genes *psiB* (RARK0005), which acts to inhibit induction of bacterial SOS stress response in conjugation, and *ssb* (RARK0100), which encodes a single stranded DNA binding protein required for bacterial DNA replication, were found. [53-56] Plasmid *ssb* and *psiB* genes are thought to function in bacterial conjugation, as their expression is controlled indirectly by fertility inhibition and their products accumulate in a transient burst in the conjugatively infected cell.

- RARK0094 probably endodes the antirestriction protein KlcA, which could be involved in overcoming restriction barriers after conjugative transfer.

- RARK0007 encodes the FlmC protein (F-plasmid maintenance protein C). [57]

- A locus, *parB*, which is necessary for the stable maintenance of the IncFII plasmid R1 was also mapped (RARK0102). Functional studies have shown that *parB* exerts plasmid maintenance by host killing of plasmidless cells. [58, 59] The product, a 52–aa polypeptide, is a lethal membrane associated protein, which causes a loss of the membrane potential.

- RARK0050 encodes the fertility inhibition protein (conjugal transfer repressor). This is one of the components on the finOP fertility inhibition complex, that inhibits the expression of traJ gene, which in turn regulates the expression of some 20 transfer genes. RNA-binding that interacts with the traJ mRNA and its antisense RNA, *finP*, stabilizing finP against endonucleolytic degradation and facilitating sense-antisense RNA recognition.

- Several ORFs involved in plasmid stability were detected: ORFs RARK0058 and RARK0083 are homologs to YacB. Plasmid stable inheritance proteins StbA and StbB (RARK0086, RARK0087) were also identified on plasmid pARK2.

- RARK0090, a DNA methylase, is probably involved in protecting host DNA by its methylation against degradation by restriction enzymes.

- Two toxin-encoded genes are present on pARK2: RARK0063 – Heat-stable enterotoxin II precursor (STII, enterotoxin B, ST-B) and RARK0078 – Heat-stable enterotoxin ST-IA/ST-P – published previously [51].

- Among other ORFs of predicted function are: RARK0052 – endonuclease, and RARK0014 and RARK0049: protein with transglycosylase SLT domain and protein with alpha/beta-hydrolase domain, respectively.

4.1.1.1. TRANSFER REGION

A block of genes located between sequences coordinates 12,135 bp and 52,215 bp is almost identical to known transfer genes (up to 98% identity). In pARK2, this block has got the same genetic organization as the corresponding genes on plasmid R100, pC15-1a, F plasmid, pAPEC-O2-R, p1658/97, raising the possibility that the entire region was acquired by horizontal transfer from R100 (Figure 38). Conjugative plasmids are important mediators for genetic transfer between prokaryotes. During bacterial conjugation, a conjugative plasmid directs transfer of a copy of itself, in single-stranded form, to a recipient cell.

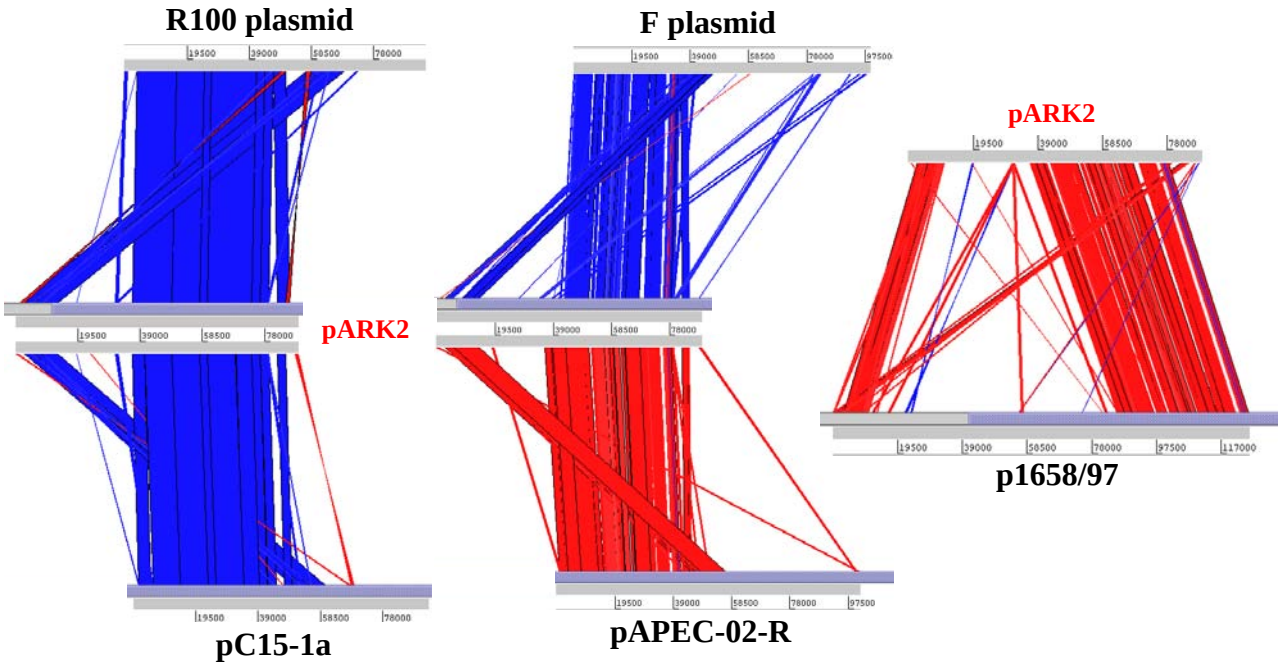


Figure 38. Linear DNA sequence comparison of plasmids pARK2 and F plasmid, R100, pC15-1a, p1658/97, pAPEC-02-R. Red and blue bars represent corresponding regions with a high similarity, 98% or more (forward and reverse strand, respectively).

4.2. SEQUENCING AND ANALYSIS OF A POLYKETIDE SYNTHASE GENE CLUSTER (PKS-PAI (*asnW*)) OF THE NEWBORN MENINGITIS ISOLATE *E. COLI* IHE3034

4.2.1. INTRODUCTION

Escherichia coli is a normal inhabitant of the gut. However, certain variants of *E. coli* are able to cause either intestinal or extraintestinal infectious diseases. Extraintestinal *E. coli* pathogens include those, which are able to evoke urinary tract infections (UTIs), sepsis, and newborn meningitis (NBM).

Bacterial meningitis is a major cause of neonatal mortality worldwide and is associated with a high incidence of neurological sequels. *E. coli* is the second common cause of bacterial meningitis in newborns after group B streptococci (GBS). [60]

The pathogenesis of *E. coli* neonatal meningitis is characterized by high-level bacteremia following the penetration of the blood-brain barrier (BBB). [61-64] The capsular polysaccharide K1 is a virulence factor with a key role in the bacteremic phase. [65, 66] Other bacterial traits contribute to *E. coli* survival in serum and may thus be involved in the pathogenicity of MNEC isolates. These factors include iron-uptake systems encoded by the loci *iro* [67,68] and *chu* [69], the siderophores aerobactin [70] and yersiniabactin [71-73], and other factors, such as hemolysin that may facilitate access to iron and or resistance to phagocytic cells. [74-76] Several specific virulence factors facilitate BBB penetration, such as K1 [77], fimbrial adhesin S (*sfaS*) [78-80], the invasin IbeA (*ibeA*) [81, 82], and cytotoxic necrotizing factor (*cnf1*). [83, 84] *E. coli* strains invade the blood stream of infants from the nasopharynx or GI tract and are carried to the meninges, layers of tissue covering the brain and spinal cord.

Most cases of *E. coli* meningitis occur in newborn babies and. *E. coli* causes between one-quarter and one-third of cases of meningitis in newborns, but less than 2% of cases of meningitis at all other ages. Almost all infections of newborns are caused by one type of *E. coli*, the K-1 strain. Like many other forms of bacterial meningitis, it affects in particular those with a suppressed immune system (removal of the spleen, cancer, organ transplants, AIDS, etc.). *E. coli* meningitis is treated with antibiotics, but, like all forms of bacterial meningitis, it is a life-threatening disease.

4.2.2. PKS/NRPS

Microorganisms make a wealth of unusual metabolites that have a secondary role in the organism's ontogeny, such as self-defence, competitive aggression, or communication, as a response to environmental challenges. Such secondary metabolites are known to possess a wealth of pharmacologically important activities, including antimicrobial, antifungal, antiparasitic, antitumor, immunosuppressive and agrochemical properties. These metabolites are ubiquitous in distribution and have been reported from organisms as diverse as bacteria, fungi, amoeba, plants, insects, dinoflagellates, molluscs and sponges. The wide spectrum of acticvity of polyketides makes them economically, clinically and industrially the most sought-after molecules. The preliminary investigation of the meningitis-causing E. *coli* strain IHE3034 led to the hypothesis that it might secrete a PKS-derived compound, which has got a cytotoxic effect on eukaryotic cells (Dobrindt U., not published).

A better understanding of the molecular pathogenicity of MNEC, especially which regard to the impact of secondary metabolites in disease formation, is necessary in orderto develop new preventive strategies.

4.2.3. SEQUENCING

To sequence the IHE3034 BAC11 insert, the same shotgun approach was used as in the case of genome sequencing project of 536 *E. coli*. The DNA was first randomly broken into fragments using nebulizer technology and size fractionation was performed by applying a sucrose density gradient in order to obtain fragments in the range of 1.0 to 4.0 kb. Purified, end repaired fragments were subcloned into the cloning vector pCR4Topo (Figure 18) forming a random DNA library. Data were collected on ABI 3730xl sequencer. The 1,300 sequence runs were assembled and the resulting contigs were ordered. Remaining gaps were closed using PCR-based methods. The plasmid was sequenced with a 9-fold coverage.

4.2.3.1. SEQUENCE ANALYZING

Essentially, the whole sequence of IHE3034 BAC11 insert was obtained and analysed (Figure 39). It is composed of 67,932 bp. The sequence was search for potential open reading frames. 41 ORFs were identified most of them are oriented in the same direction. Obtained protein-coding regions include the PKS-PAI associated with the *asnW* tRNA gene and 15 flanking ORFs. The PKS-PAI shows 99% identity on nucleotide level to the homologous region of the sequenced uropathogenic strains *E. coli* 536 and CFT073. Analysis of base composition showed the G+C content of the insert is 53,2%, which is higher than in the *E. coli* 536 whole genome (50,52%).

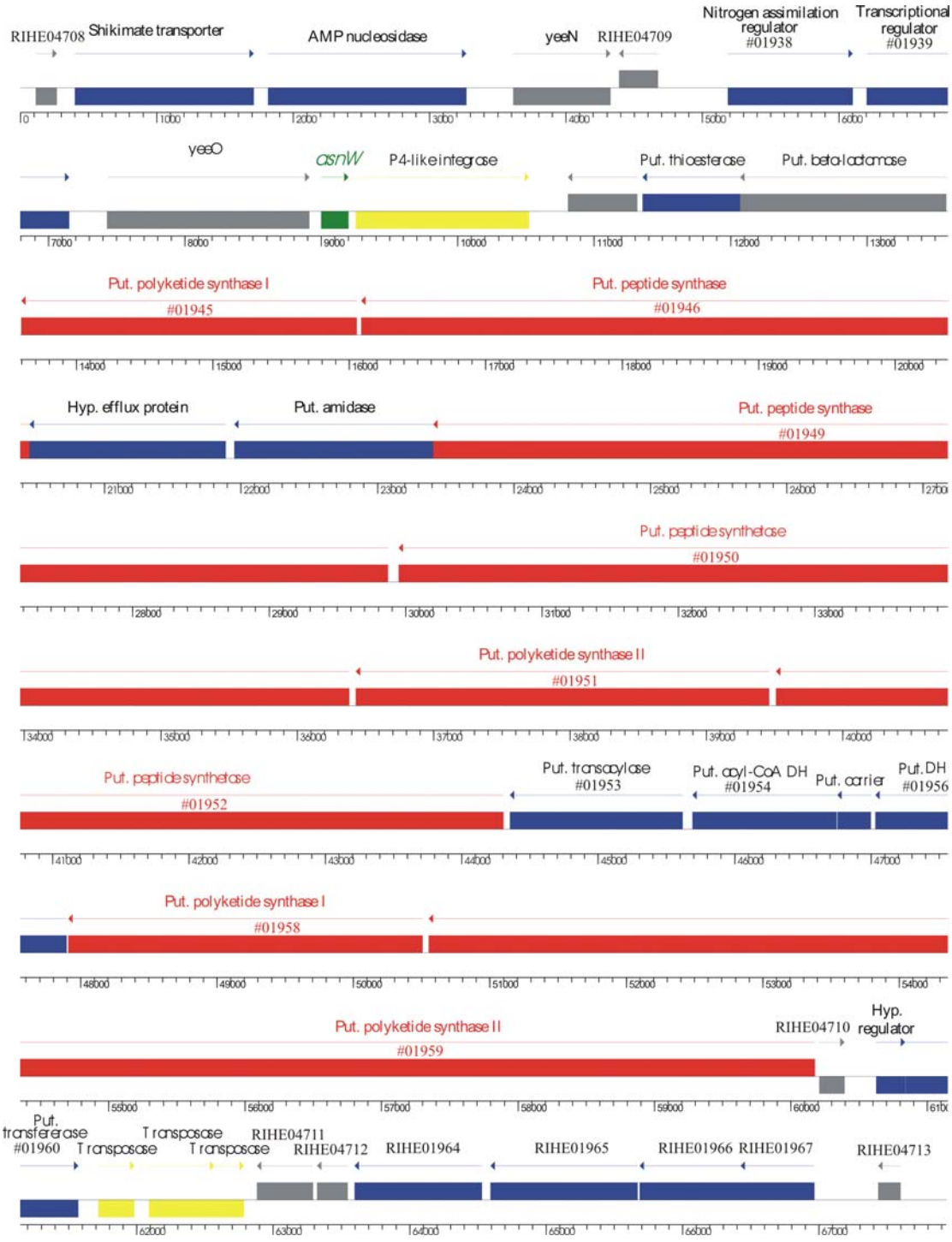


Figure 39. Linear representations of the IHE3034 BAC11 insert sequence. Red blocks represent all putative PKS genes, yellow – IS elements, grey – hypothetical proteins, green – tRNA gene, and dark blue – others. Put. – putative, Hyp. – hypothetical, DH - dehydrogenase

4.2.4. DETECTION OF THE POLYKETIDE-PAI (*ASNW*) TOXIN -PRODUCING *ESCHERICHIA COLI* ISOLATES BY PCR SCREENING

In order to investigate the occurrence of the identified PKS cluster in an *E. coli* strain collection PCR screening was applied. A total of 185 isolates (Table 12) were subjected to PCR assay. Based on a conserved region within the PKS cluster eight pairs of primers were designed (Table 13).

The PCR results demonstrated that 61 (33%) of 185 strains were PKS-positive (Figure 40). Table 17 shows all tested strains carrying a PKS gene cluster.

TABLE 17. *E. COLI* ISOLATES CARRYING A PKS-PAI ACCORDING TO PCR RESULTS.

STRAIN	SEROTYPE	PATHOTYPE	COMMENTS	HLY TEST
DSM6601	O6:K5	Faecal isolate	probiotic Nissle 1917 strain	Hly-
EcoR51	O25:HN	Faecal isolate	EcoR gr. B2; kps1,2; sfa1,2; pap1,2,3; hly2	Hly+
EcoR52	O25:H1	Faecal isolate	EcoR gr. B2; kps1,2; sfa1,2; pap1,2,3; hly2, (orangutan)	Hly+
EcoR53	O4:HN	Faecal isolate	EcoR gr. B2; kps1,2; sfa1,2; pap1,2,3; hly2	Hly+
EcoR54	O25:H1	Faecal isolate	EcoR gr. B2; kps1,2; sfa1,2; hly2	Hly+
EcoR56	O6:H1	Faecal isolate	EcoR gr. B2; kps1,2; pap1,2,3; hly2	Hly+
EcoR57	ON:NM	Faecal isolate	EcoR gr. B2; kps1,2; sfa1,2; pap1,2,3; hly2, (gorilla)	Hly+
EcoR63	ON:NM	Faecal isolate	EcoR gr. B2; kps1,2; sfa1,2; pap1,2,3; hly2	Hly+
EcoR65	ON:H10	Faecal isolate	EcoR gr. B2; sfa1,2; hly2, (*Celebese ape*)	Hly+
F18	NS	Faecal isolate	fimA	Hly-
F18Col -	NS	Faecal isolate	fimA	Hly-
536	O6:K15:H31	UPEC		Hly+
764	O18:K5:H5/11	UPEC		Hly+
2980	018ac:K5:H-	UPEC		Hly+
16A-2U	NS	UPEC		Hly+
19A1	NS	UPEC		Hly+
1G1U	NS	UPEC		Hly+
1H1U	NS	UPEC		Hly+
1H1	NS	UPEC		Hly+
20A1	NS	UPEC		Hly+
20A1U	NS	UPEC		Hly+
20A2	NS	UPEC		Hly+
22B2U	NS	UPEC		Hly+
2E1U	NS	UPEC		Hly-
2E2U	NS	UPEC		Hly-
7521/94-1	NS	UPEC		Hly+
764-2	O18:K5:H5/11	UPEC		Hly-

CFT073	O6:H1:K2	UPEC		Hly+
EcoR55	O25:H1	UPEC	EcoR gr. B2; kps1,2; pap1	Hly-
EcoR60	O4:HN	UPEC	EcoR gr. B2; sfa1,2; pap1,2,3; hly2	Hly+
J96	O4:K-:H5	UPEC		Hly+
J96-M1	O4:K:H5	UPEC		Hly-
RZ-411	O6:K-:H1	UPEC		Hly-
RZ-422	O6:K14:H-	UPEC		Hly+
RZ-439	O6:K5:H1	UPEC		Hly-
RZ-441	O6:K5:H1	UPEC		Hly+
RZ-451	O6:K+:H31	UPEC		Hly+
RZ-454	O6:K2:H-	UPEC		Hly+
RZ-468	O6:K5:H1	UPEC		Hly+
RZ-475	O6:K5:H1	UPEC		Hly+
RZ-479	O6:K+:H-	UPEC		Hly+
RZ-495	O6:K5:H-	UPEC		Hly-
RZ-500	O6:K5:H1	UPEC		Hly-
RZ-505	O6:K14:H-	UPEC		Hly+
RZ-525	O6:K5:H1	UPEC		Hly-
RZ-526	O6:K5:H1	UPEC		Hly+
RZ-532	O6:K-:H31	UPEC		Hly+
BK658	O75:K1:H7	MENEC	expresses F1A and F1B fimbriae and S/F1C-related fimbria	Hly+
IHE3034	O18:K1:H7/9	MENEC		Hly-
B13155	NS	MENEC		Hly-
IHE3036	O18:K1:H7	MENEC		Hly+
IHE3080	O18:K1:H7	MENEC		Hly-
RS218	O18ac:H7:K1	MENEC		Hly+
RS226	O18ac:H7:K1	MENEC		Hly+
B616	NS	Sepsis		Hly-
E642	NS	Sepsis		Hly+
HK2	NS	Sepsis		Hly-
HK4	NS	Sepsis		Hly+
HK54	NS	Sepsis		Hly+
HK8 269/93	NS	Sepsis		Hly+
HK24 10413/93	NS	Sepsis		Hly+

hly (a-hemolysin), kps (type II capsule), sfa (S-type. fimbrial adhesin), pap (P-type fimbrial adhesin), fim – fimbriae, NS = not serotyped, Last column: Hly- non-hemolytic and Hly+ hemolylit isolate

Of 70 UPEC strains, 36 (51%) were positive. The PCR was also successful in screening for the PKS cluster in faecal isolates (38%), and in meningitis and sepsis causing *E. coli* strains (46% in both cases). So there is no conncetion between the occurence of the PKS-cluster to a special *E. coli*-caused disease.

Tested *E. coli* isolates included 33 well-characterised isolates belong to EcoR collection (The *Escherichia coli* Reference Collection). [85] Ten of them were positive and all of these belong to theB2 phylogenetic group. [86]

Results show that polyketide-PAI (*asnW*) occurred only among extraintestinal strains. None of enteropathogenic, enteroinvasive, enterotoxigenic, enteroaggregative or enterohemorrhagic strains amplified the PKS-PAI (*asnW*) sequences.

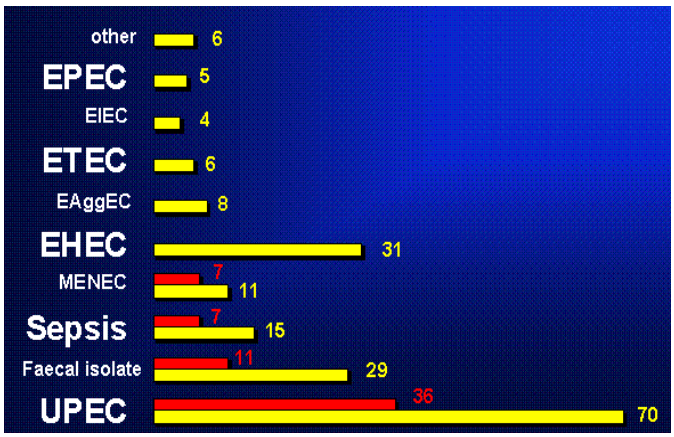


Figure 40. *E. coli* isolates possess a PKS-PAI according to PCR results (in red).

4.2.5. IDENTIFICATION OF HEMOLYSIN-PRODUCING *E. COLI* ISOLATES

Production of hemolysin can be plasmid or chromosomally determined. Haemolytic *E. coli* are frequently isolated from the faeces of normal pigs, and those with oedema disease or diarrhoea. Alpha-haemolysin is also frequently produced by strains isolated from cases of human urinary tract infections and other extraintestinal infections. It is produced early in the growth phase. It lyses erythrocytes from cattle, pigs, horses, rabbits, guinea pigs, chickens and humans, and acts as a pore-forming cytolysin, causing dissipation of transmembrane ion gradients, while retaining cytoplasmic proteins. Thus water moves into the cell as a result of the increased intracellular osmotic pressure, causing the cell to swell and finally burst. Ca^{2+} ions are required for proper haemolysin function. [87-89] Haemolysin affects not only erythrocytes but has cytocidal effects on a variety of cell types including HeLa cells.

A total of 61 PKS-PAI (*asnW*)-positive isolates were tested for hemolysin production: 17 were established as non-hemolysin-producing isolates (Table 17). These strains were used for toxin activity assay.
A total of 61 PKS-PAI (*asnW*)-positive isolates were tested for hemolysin production: 17 were established as non-hemolysin-producing isolates (Table 17). These strains were used for toxin activity assay.

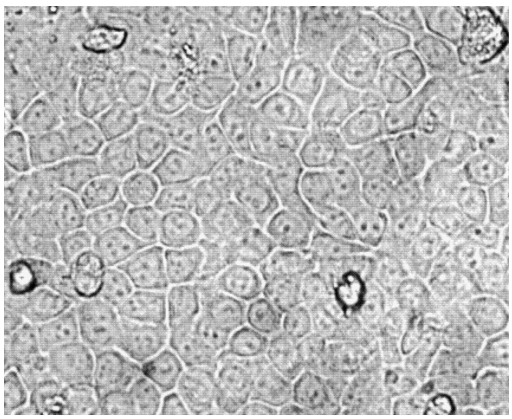
4.2.6. TOXIN ACTIVITY ASSAYS

The goal of this study was to gain first insight into the function of the PKS cluster. For this reason, the effect of potentially secreted compounds was examined on human cervix epithelioid carcinoma cells (HeLa cells). Initially, some control experiments with prepared genetic constructs were performed. All constructs were obtained from Stefan Homburg (Institut für Molekulare Infektionsbiologie der Universität Würzburg).
HeLa cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum and glutamine. For cytotoxicity assays, cells were seeded into 96-well plates for 24 hours. Then monolayer cells were infected with overnight cultures:

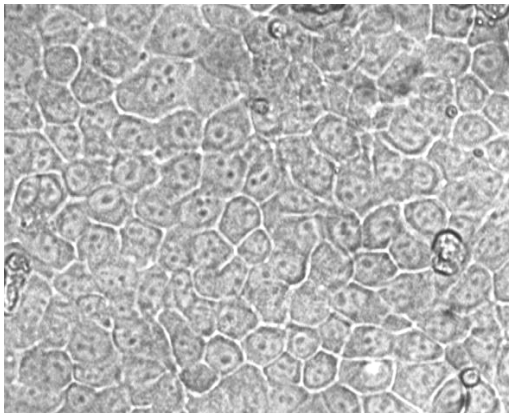
- MG1655 – laboratory strain – negative control – no effect was suspected;
- IHE3034 – MNEC strain;
- IHE3034ΔPKS – IHE3034 mutant with deleted PKS genes cluster – again, no effect was suspected;
- MG1655/IHE3034 BAC11 – effect would confirm, that cytotoxity activity is combine with a PKS genes cluster
- J96-M1 – strain possessing a PKS-PAI and suspected to produce a PKS-toxin
- J96-M1ΔPKS – J96-M1 mutant with deleted PKS genes cluster – no effect was suspected;

After 3.5 hour of co-incubation, the growth media were removed and cells were washed three times with washing buffer. Next, cells were grown at 37°C, 5% CO_2, 95% humidity and controlled after 24-48 hours.

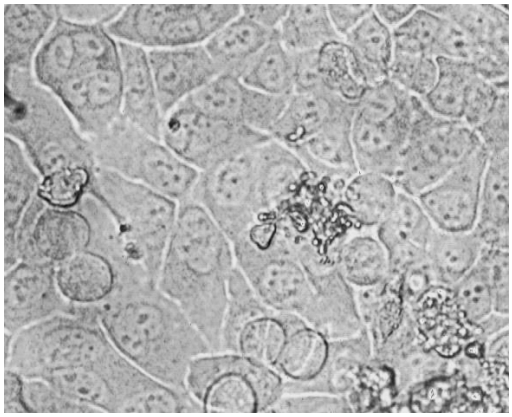
The results of that experiment are shown on Figure 41. The remarkable effects on HeLa cells were observed in both suspected strains (IHE3034 and J96-M1) as well as in case of MG1655/IHE3034 BAC11 strain. The HeLa cells were significantly bigger. No toxic activity was detected, when PKS-PAI was deleted from the chromosome. PKS^+ strains are those, which possess PKS-PAI and demonstrate morphologic changes in the HeLa cells. This results confirmed suspicions that PKS cluster encode a compound with modulatory activity on eukaryotic cells.



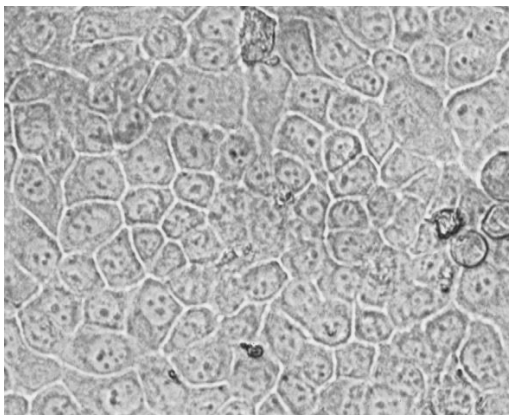
Control: non infected HeLa cells



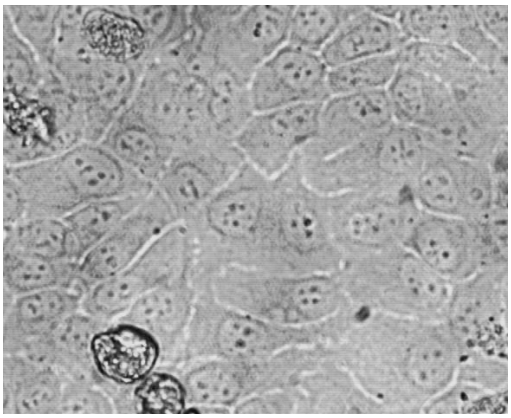
MG1655



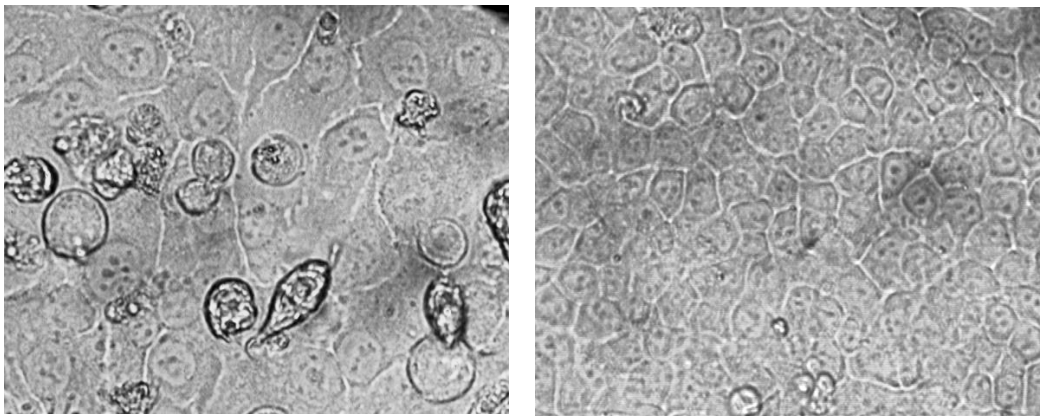
IHE3034



IHE3034ΔPKS



K12/BAC_PKS3034



J96-M1 **J9696 M1ΔPKS**

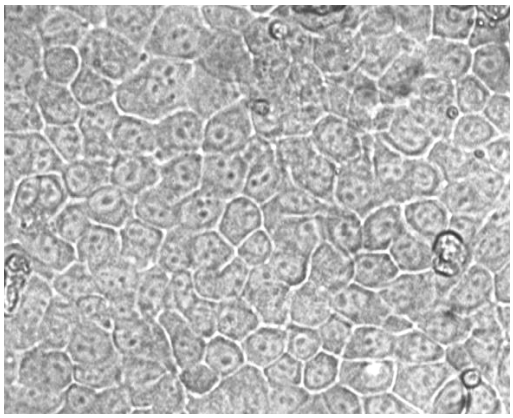
Figure 41. Assay of the toxicity of the *E. coli* isolates possess PKS-PAI to HeLa cells – preliminary test. Cells were photographed with the 40× objectives on a Nikon microscope that was fitted with a Nikon camera, after 48 h from co-incubation. Non-infected HeLa cells and cells infected with laboratory strain MG1566, as a negative control, showed no changes. Coincubation with strain IHE3034 and J96-M1 induced morphologic changes on HeLa cells. No change was observed in case of both PKS deletion mutants, when toxic effect was observed for K12/BAC_PKS3034 (laboratory strain with BAC clone possessing whole PKS cluster)

TABLE 18. E. COLI STRAINS WHICH WERE TESTED FOR TOXIC ACTIVITY

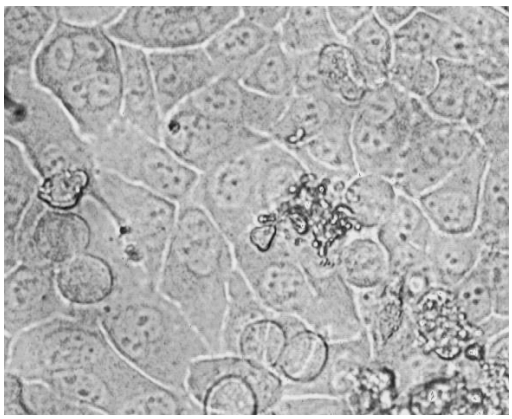
STRAIN	PATHOTYP	PHENOTYPE
-	non infected HeLa cells	PKS^-
MG1655	laboratory strain	PKS^-
536hly^- *	UPEC	PKS^-
EcoR55	UPEC	PKS^-
RZ-411	UPEC	PKS^-
F18	Faecal isolate	PKS^-
F18Col^-	Faecal isolate	PKS^-
IHE3080	MENEC	PKS^-
IHE3034ΔPKS **	constructed	PKS^-
J96-M1 ΔPKS **	constructed	PKS^-
MG1655/IHE3034 BAC11**	constructed	PKS^+
IHE3034	MENEC	PKS^+
B13155	MENEC	PKS^+
DSM6601	Faecal isolate	PKS^+
CFT073hly^- ***	UPEC	PKS^+
J96-M1	UPEC	PKS^+
2E2U	UPEC	PKS^+
764-2	UPEC	PKS^+
RZ-439	UPEC	PKS^+
RZ-495	UPEC	PKS^+
RZ-500	UPEC	PKS^+
RZ-525	UPEC	PKS^+
B616	Sepsis	PKS^+
HK2	Sepsis	PKS^+

* Gabor Nagy
** Stefan Homburg
*** Ulrich Dobrindt

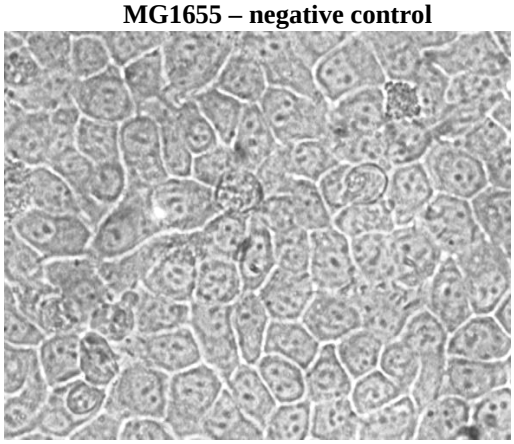
After this series of preliminary tests, all seventeen previously selected strains (Table 18) were tested for polyketide-derived toxic activity. Within 48 hours after co-incubation 11 (58%) PCR-positive strains showed the cytopathic effects, while control cells (infected with MG1655, as well as non infected HeLa monolayer calls) and six remaining strains showed a normal morphology (Figure 42). A correlation between pathotype and PKS-phenotype was not observed. Morphologic changes were found in 8 (42%) uropathogenic *E. coli*. Among the bacteria tested, both sepsis cause strains had an obvious effect, as well as two out of three MNEC isolates were PKS^{+}. The citotoxity effect could be detected in one faecal- derived isolate, when two remaining had no obvious activity.



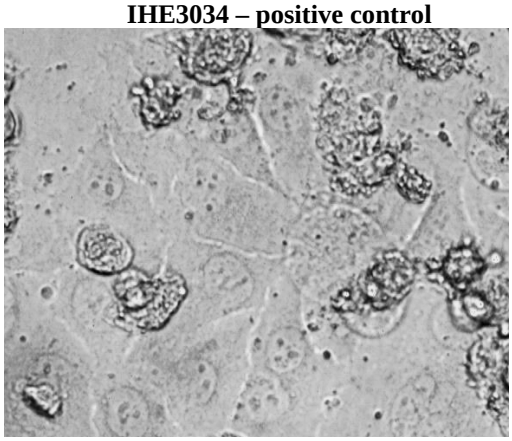
MG1655 – negative control



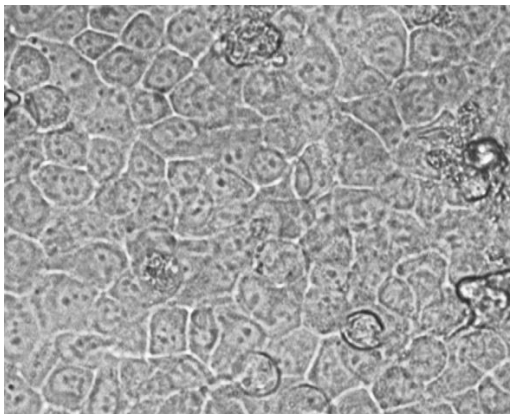
IHE3034 – positive control



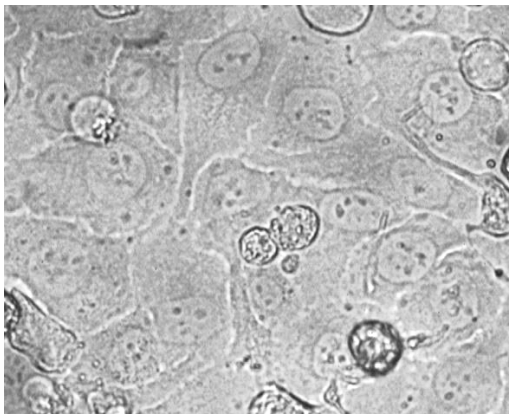
536hly^{-}



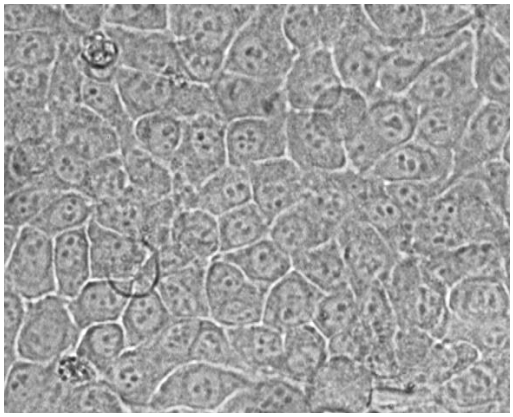
CFT073hly $^{-}$



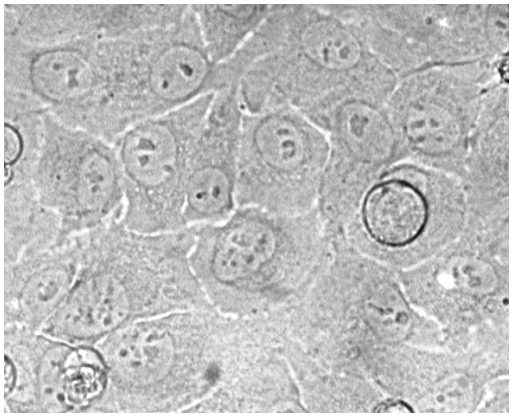
F18



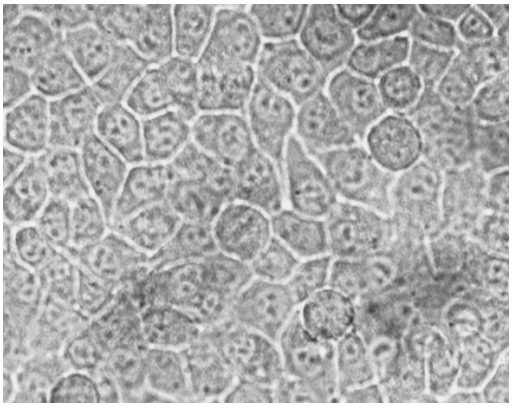
DSM6601



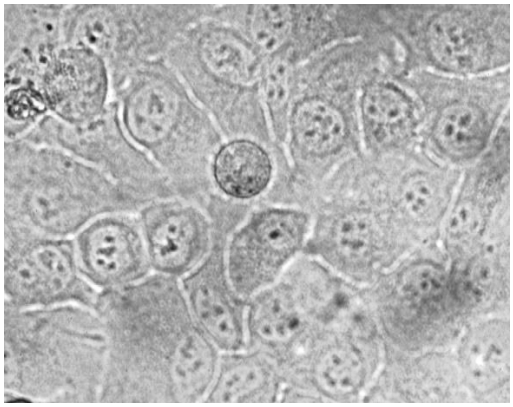
F18Col^{-}



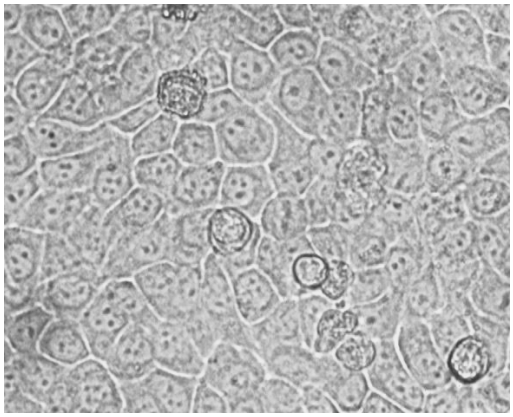
HK2



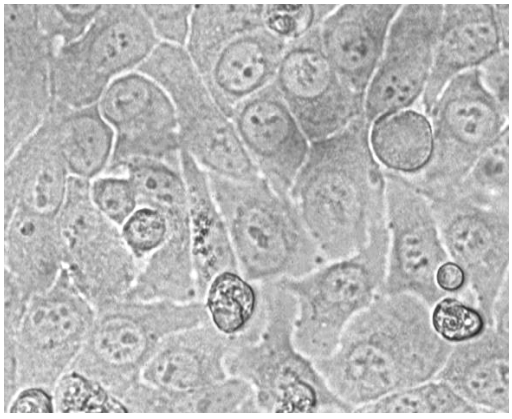
IHE3038



B13155



EcoR55



B616

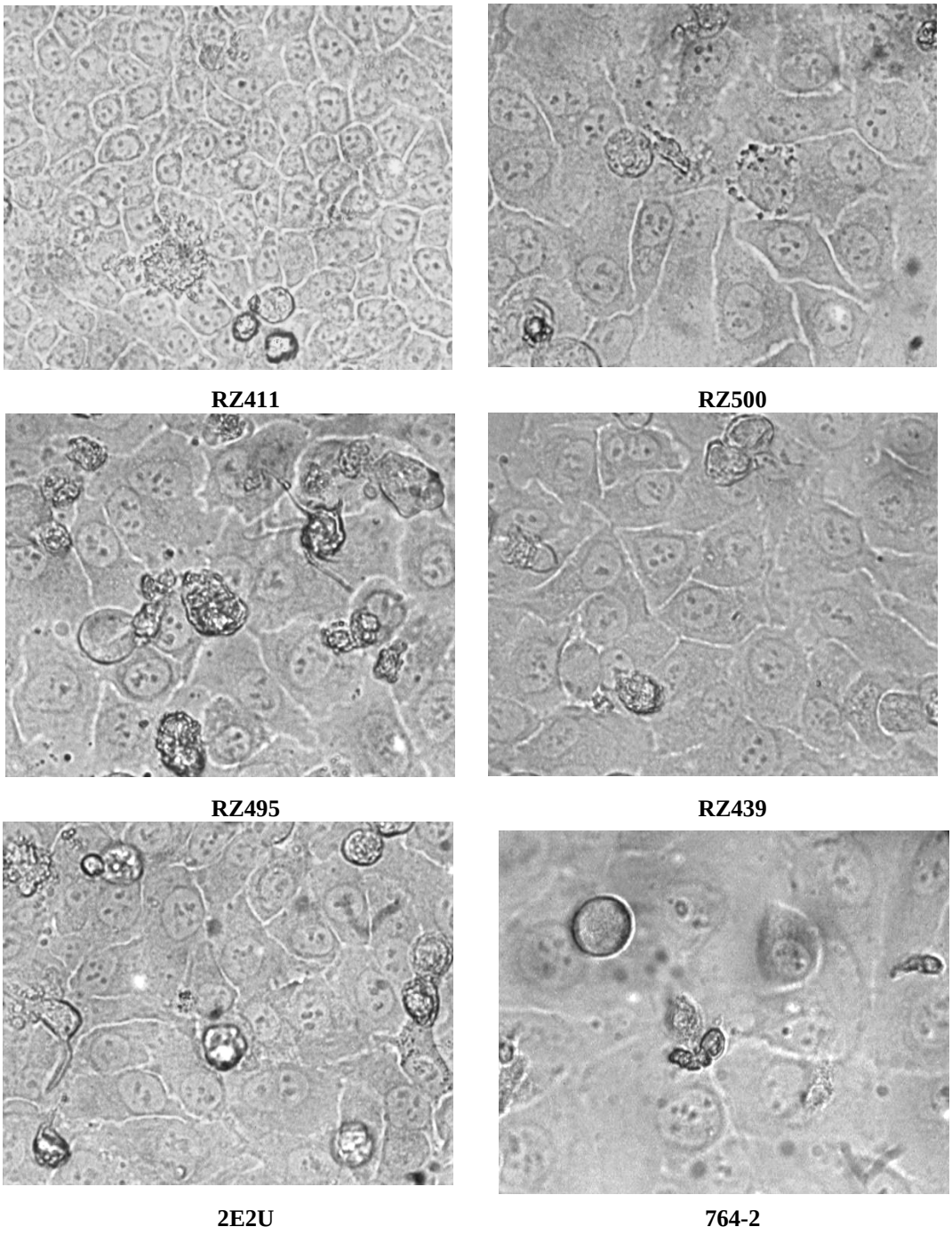


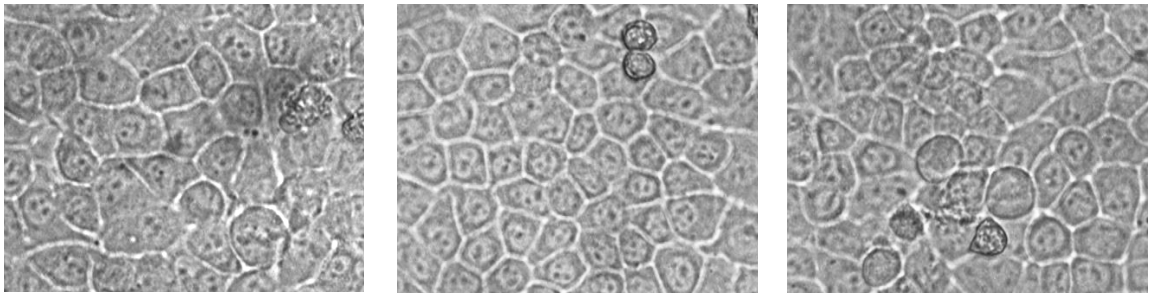
Figure 42. Assay of the toxicity of the *E. coli* isolates possess PKS-PAI to HeLa cells. Non infected HeLa cells (A.) and cells infected with laboratory strain MG1566 (B.), as a control, showed no changes, when toxic delivered effect were observed in case of strains: CFT073hly-, DSM6601, KK2, RZ500, RZ495, rz439, 3E2U and 764-2. No effects were observed after coinbubation with strains: 536hly-, F18, F18Col-, and RZ411. Cells were photographed with the 40× objectives on a Nikon microscope that was fitted with a Nikon camera, after 48 h from co-incubation.

Interestingly, induced changes, which were detected, in some cases seemed to be clearer, more explicit. To determine the basis for the unexpected differences in morphology, 96-hours experiment was performed. After appropriate time of incubation (48, 72 or 96 h), the cells condition was controlled (Figure 43). Within 96 hours after co-incubation, no changes were observed in case of non-infected cell and cells infected with laboratory strain MG1655 as well as all strains previously showing PKS^- phenotype. Cells were able to grow and proliferate. However, differences in cytotoxic effect induced by PKS-derived toxin were still observable. Only IHE3034-infected cells were able to survive 96 hours after co-incubation. In the remaining cases apoptosis was observed within 72 hours after bacterial infection.

In conclusion, of 19 *E. coli* strains, experiment successfully identified 13 positives isolates. The cytotoxicity assay with selected *E. coli* strains showed that the presence of the *asnW*-associated PKS island is not a guarantee for an effective secretion of PKS toxin. At the same time it was possible to prove that PKS gene cluster encodes a compound inducing morphologic changes on human cervix epithelioid carcinoma cells. The experiment showed that a polyketide-derived toxin causes cythopathicity and cell cycle arrest in HeLa cell culture.

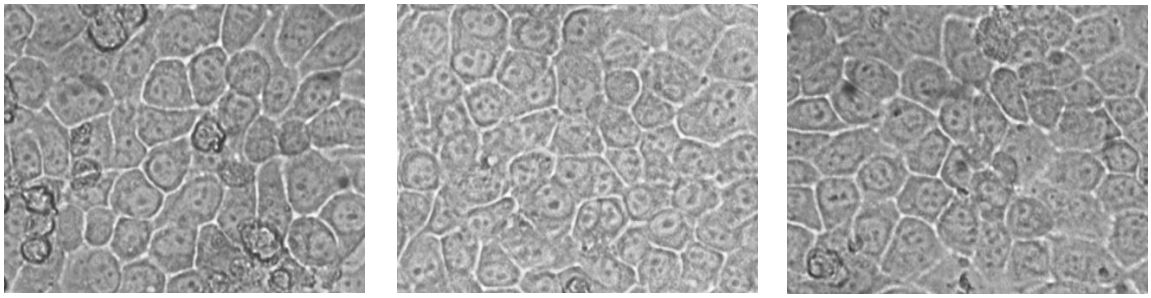
To further understand the function, these *in vitro* results obtained with cell culture must be expanded in conjunction with supporting or associated *in vivo* studies, biochemical analysis and assay with the isolated and characterized compound.

A.



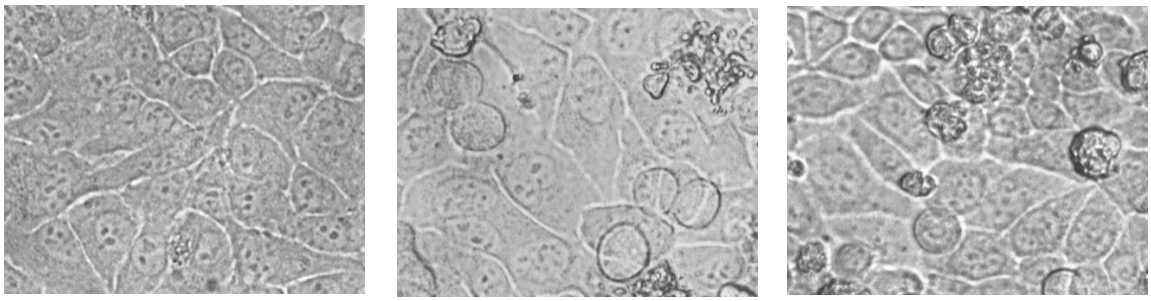
Non infected

B.



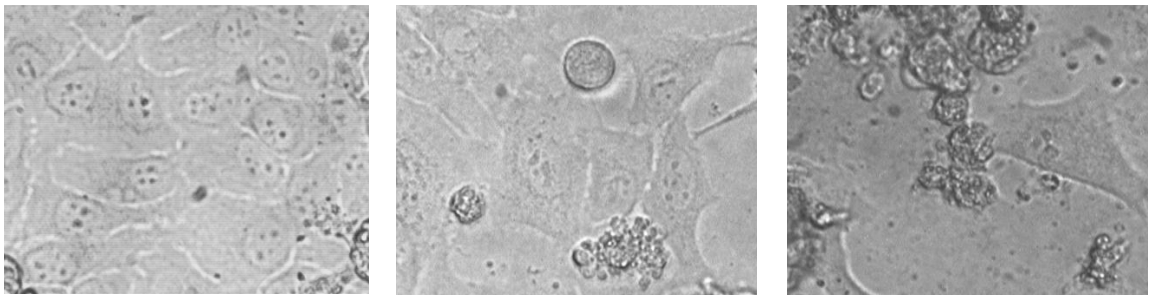
MG1655

C.



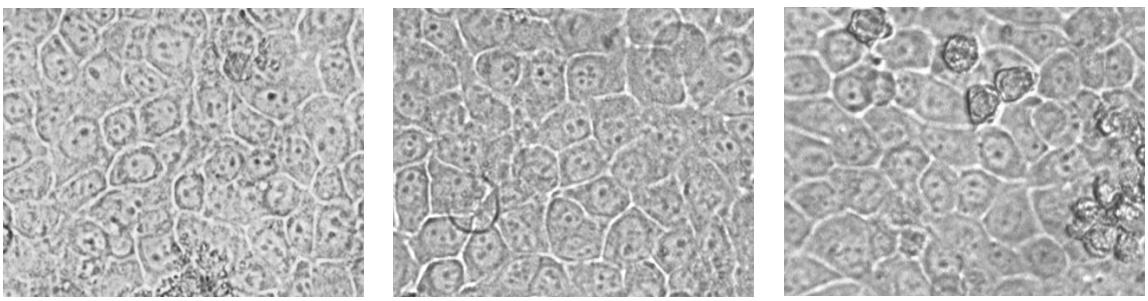
IHE3034

D.



K12/BAC_PKS

E.



536hly-

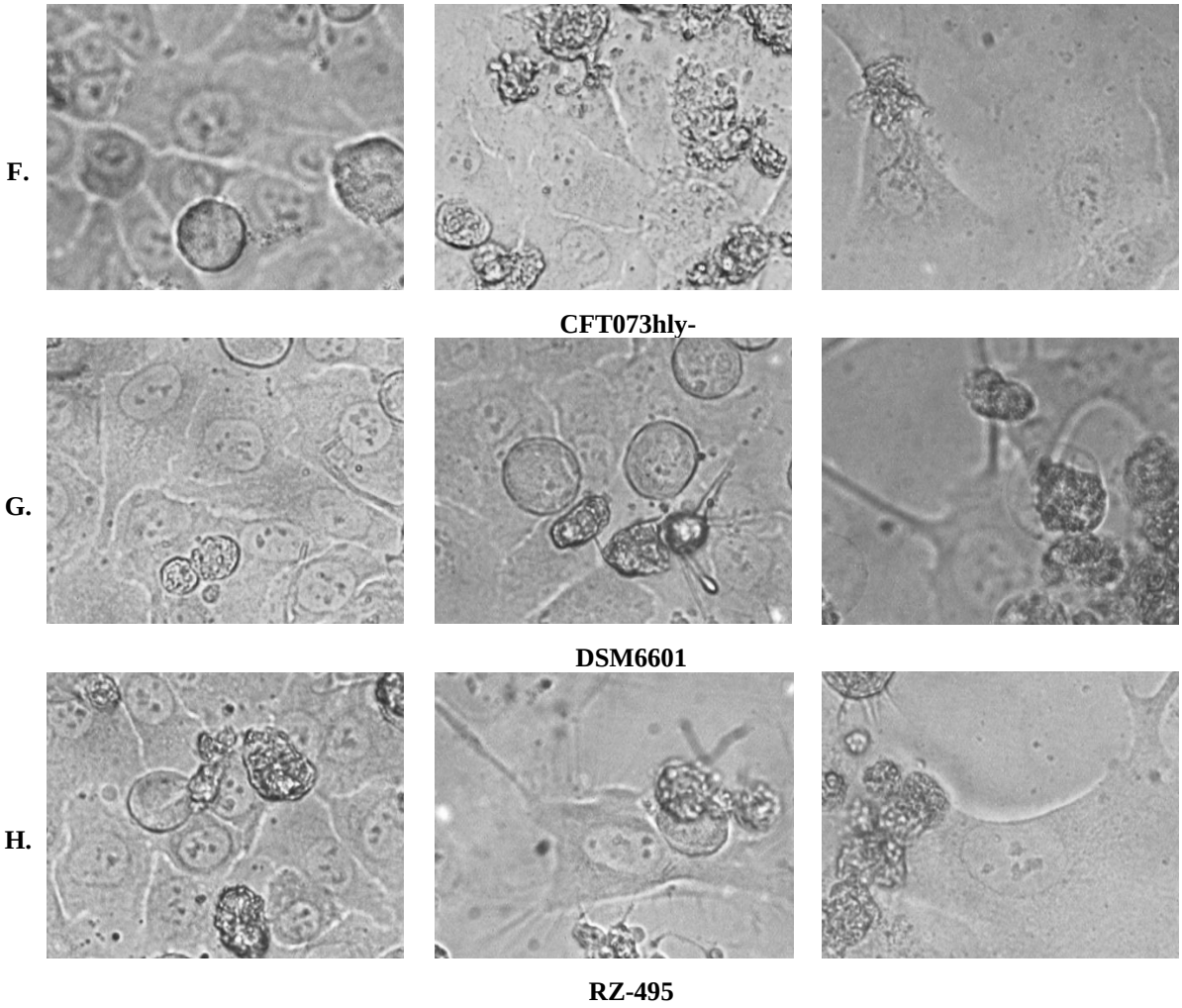


Figure 43. Time-depending assay of the toxicity of the selected *E. coli* isolates to HeLa cells. Each row shows the same sample after 24, 48 and 96 hours, respectively. Non infected HeLa cells (A.) and cells infected with laboratory strain MG1566 (B.), as a control, showed no changes, when apoptosis were observed in case of strains: K12/BAC_PKS (D.), CFT073hly- (F.), DSM6601 (G.) and RZ-495 (H.) after 96 h. Coincubation with the 536hly- strain (E.) indicated no morphologic changes on HeLa cells. Coincubation with strain IHE3034 (C.) induced cell arrest but no apoptosis 96 hour after infection. Cells were photographed with the 40× objectives on a Nikon microscope that was fitted with a Nikon camera.

5. DISCUSSION

Uropathogenic *E. coli* are intestinal commensals as well as pathogens of the urinary tract. Identification of genes specific to UPEC that permit replication, survival and disease in the urinary tract is critical to understanding their pathogenesis.
The approach was to investigate loci common in UPEC strains, but missing in enteric pathogenic *E. coli* strains as well as laboratory MG1655 strain.

5.1. INSERTION ELEMENTS

The general characteristics and genetics of insertion sequence (IS) elements are well established. For *E. coli* IS elements, mechanism of transposition and mutation are known and their recombinogenic role in bacterial genome has been investigated. [90,91] In *E. coli* the IS elements are dispersed around the chromosome (also found on the pARK2 plasmid). Despite their widespread occurrence and potential for horizontal gene transfer, the observed distribution remains somehow enigmatic. Natural populations of *E. coli* harbour at least 20 classes of IS and are highly polymorphic with respect to their distribution and copy numbers. [92, 93]
Most of the 536 IS elements are found within PAIs. However some are present outside of large pathogenic islands.
Six copies of IS1397 were found in the 536 genome (two in CFT073). IS1397 elements are IS3 family members, which are specifically inserted into the loop of palindromic units (PUs). IS1397 is shown to transpose into PUs with sequences close or identical to the *E. coli* consensus, even in other enterobacteria (*Salmonella enterica serovar Typhimurium, Klebsiella pneumoniae*, and *Klebsiella oxytoca*). [94] Interesting is that there is no IS1397 element within PAIs and the localization of all IS1397 is always specific. In two cases significant differences in genetic organization were observed. In corresponding MG1655 regions additional metabolic pathways were found: the cyanate uptake system and 3 -phenylpropionate degradation operon.
Cyanase is an inducible enzyme in *E. coli* that catalyzes bicarbonate-dependent decomposition of cyanate. It is encoded as part of an operon we have named the cyn

operon, which includes three genes in the following order: cynT (cyanate permease), cynS (cyanase), and cynX (protein of unknown function).Cyanate at concentrations of about 1 mM is toxic to strains of *E. coli* lacking the cyanase gene, but strains in which the inducible gene for cyanase is present can grow on cyanate as the sole source of nitrogen at concentrations as high as 20 mM. Uptake of cyanate involves the product of the permease gene in an energy-dependent process. It appears that the cyn operon has evolved to function in detoxification/decomposition of cyanate arising from both intra- and extracellular sources. [95]
3-phenylpropionate degradation is one of only two aromatic-ring-cleavage pathways found in *E. coli*. That *E. coli* can degrade aromatic acids suggests a wider natural distribution for *E. coli* than the anaerobic environment of the animal gut. [96] However both systems are absent from *E. coli* 536 as well as CFT073.

Although there is a significant number of IS100 elements (12 in whole genome) only three are encoded outside PAIs. The most obscure seems to be a presence of two IS100 elements (RECP02670, RECP02671) together with IS911 (RECP04737, RECP02673) within a 4-kb 536 specific sequence. This region encodes fourteen short ORFs showing no similarity to any known proteins. Interestingly one of the IS100 element (RECP04681, RECP04682) disrupts phage integrase. This phenomenon occurs only in case of *E. coli* 536. Corresponding phage integrases from *E. coli* MG1655, CFT073 and 0157:H7 are not disrupted.

There are 19 copies of IS200 in the CFT073 genome. IS200, the smallest known element, is confined to the salmonellae and several lineages of *E. coli*. However it is absent from the rest fully sequenced *E. coli* genomes.
IS 200 is an insertion element abundant in the genus *Salmonella* [97] and sporadically found in strains of *Shigella* [98], *Escherichia coli* [99] and *Yersinia* [100] The chromosome of *Salmonella typhimurium* LT2 contains six copies of IS 200 [97, 101, 102]; in other *S. typhimurium* strains the number of IS 200 copies ranges from 1 to >12 [98].

The classes of IS detected within *E. coli* are generally not present in the species typed to *Salmonella*. In contrast to the situation in *E. coli*, a single class of element-IS200- has been recovered from *S. typhimurium*. A comprehensive analysis of all known IS200 (Beuzon et al., 1997) suggests a very low transposition frequency and supports the hypothesis that IS200 is an ancestral element of the *Enterobacteriaceae*. [103]

5.2. VIRULENCE FACTORS

5.2.1. PATHOGENIC ISLANDS

About 450 genes of strain 536 are found in CFT073 but are absent from MG1655 as well as from EDL933. These genes represent candidates, which could be involved in generating and maintaining an urovirulent phenotype. Furthermore, about 370 genes of *E. coli* 536 (Figure 31) are absent from all *E. coli* genomes sequenced so far, indicating their implication in an *E. coli* 536-specific phenotype. Many of the genes of these two groups are organized in so-called pathogenic islands (PAIs), five of which have been previously identified and described in detail *in E. coli* 536. [104, 105] PAI I, II, III and V contain genes, which either are not found in strain CFT073, or which are located in similar islands located elsewhere on the genome of CFT073, whereas PAI IV is very similar in both uropathogenic strains. In the following, we will mainly focus on the description of island-like genomic regions (20 kb or bigger) outside of the large PAIs. Interestingly, four of such additional island-like regions can be detected in the vicinity of PAIs III and IV, designated PAIs IIIa, IIIb, IVa and IVb (Figure 31). They exhibit a different G+C content as compared to the core genome and are flanked by tRNAs and/or mobile elements – features, which have been found to be characteristic to pathogenic islands in a variety of *E. coli* strains and some other species. [106-109]

An island-like region is located upstream of PAI III, associated with a tRNA (*aspV*) as well as with a Rhs-element-like region (rearrangement hot-spots). This *aspV*-associated island, designated PAI IIIa, is absent from CFT073 but present in the EDL933 genome. It encodes 28 proteins (RECP00217-245), most of which of unknown function. However, some proteins show similarity to virulence factors. For instance, a gene with strong similarity to *icmF* of *Vibrio cholerae* is present (RECP00220), *an in vivo* induced gene involved in the regulation of motility and adherence to epithelial cells. [112] This protein shows also significant similarity to IcmF of *Legionella pneumophila*, a protein belonging to the icm cassette responsible for macrophage killing and intracellular survival of the organism. [110] Another protein (RECP00234) is highly similar to Hcp, a 28-kDa secreted protein of *Vibrio cholerae*, which is regulated co-ordinately with the hemolysin HlyA. [111] Recently, a comparative *in silico* study was undertaken with 27 genomes containing icmF homologs. [114] A partially conserved gene cluster surrounding the icmF genes was identified in several genomes including *Yersinia*, *Vibrio* and *Salmonella* species, designated IAHP (IcmF associated homologous proteins) cluster. Gene composition and organization varies greatly among different species, indicating a strong genomic rearrangement. PAI IIIa of *E. coli* 536 can be regarded as a member of the IAHP cluster. Interestingly, a second IAHP-like cluster can be found within the genome of strain 536. This cluster (RECP02765-2793), designated PAI VI, is also associated with a tRNA (*metV*) and exhibits an aberrant G+C content. It is highly homologous to a *metV*-associated cluster in strain CFT073. [113] Like PAI IIIa, it encodes homologs of IcmF and Hcp. However, its gene composition and organization differs greatly from PAI IIIa. Furthermore, weak, if detectable, similarities of proteins of PAI IIIa to proteins of PAI IV indicate that these islands must have been acquired independently from different sources. It has to be shown if these two IAHP clusters in strain 536 complement each other or represent independent systems, which may be involved in surface architecture or secretion. [114]

A few kb downstream of PAI III another island-like region (PAI IIIb) flanked by a phage integrase gene is found, which is present also in CFT073 but absent from other *E. coli* strains. Some putative virulence-associated genes are encoded here, such as an adhesion and a protein harbouring an autotransporter domain.

The island PAI IV of *E. coli* 536, highly similar to a part of the so-called ‘high pathogenicity island’ of *Yersinia* species, encodes among others a system homologous

to the Yersiniabactin (Ybt) synthetase, an iron-chelating siderophore. [115, 116] Two of its proteins, HMWP1 (RECP01925) and HMWP2 (RECP01924), belong to the class of non-ribosomal peptide synthetases/polyketide synthases (NRPS/PKS), multi-domain enzymes with 9 and 7 subunits, respectively ([117], Figure 32). Unlike in strain 536, both highly homologous ORFs in CFT073 are disrupted. This corresponds to findings, which have proven the absence of a Ybt-like siderophore in strain CFT073, in contrast to strain 536, which was tested Ybt-positive (Dobrindt, U., unpublished). Since iron acquisition is of crucial importance for bacterial growth in a mammalian host, this difference may also explain to a certain extent that virulence of strain 536 is higher compared to strain CFT073 when tested in the mouse model (Dobrindt, U., unpublished).

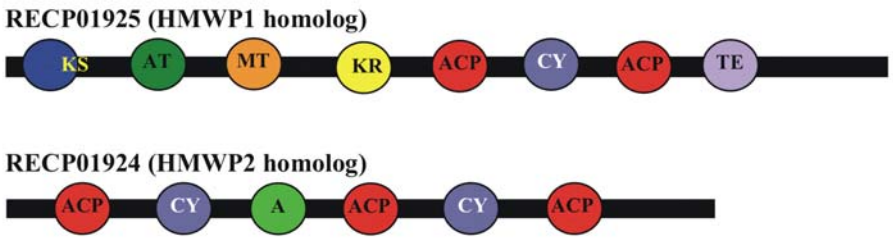


Figure 44. Domain structure analysis of RECP01924 and RECP01925 *E. coli* 536 genes. HMWP1/P2 are the main subunits of the Yersiniabactin synthetase of *Y. pestis*. The domain prediction programs PFAM and SEARCHPKS were used in combination. A, Adenylation domain (AMP-binding domain); ACP, Acyl/Aryl/Peptidyl carrier protein; AT, Acyltransferase; Cy, Cyclization/Condensation domain; ER, Enoyl reductase; KR, Ketoacyl reductase; KS, Ketoacyl synthase; MT, Methyltransferase (MT domains are weakly predicted by PFAM and SEARCHPKS); TE, Thioesterase

Downstream of PAI_{536} IV a region could be identified, which is flanked by a phage integrase gene and associated with a tRNA gene (*asnW*). The region, designated PAI_{536} IVb, encodes several large proteins with a predicted molecular weight of up to 350kDa, eight of which fall into the class of the above-mentioned NRPS/PKS. This class of enzymes has attracted special attention recently, due to their identification in the course of various genome-sequencing projects. [118] It has been shown that they are responsible for the synthesis of a variety of lipopeptides involved in, for instance,

intercellular communication, iron acquisition, competition or self-defence. [118-120] More details of this gene cluster will be discussed later.

Further downstream of the NRPS/PKS genes several IS elements with truncated transposase genes are present as well as a siderophore system cluster composed of four genes (RECP01996-99).

Examining the upstream region of PAI IV, a small island again flanked by a tRNA (serU) and a phage integrase gene is present, designated PAI IVa. Except of an ORF with similarity to a shufflon encoded on the plasmid R64 of *E. coli*, which is involved in the biological swhich for generating alternative type IV pili, PAI IVa contains mainly genes of unknown function. [121]

5.2.2. SMALL GENE CLUSTER/SINGLE GENES SPECIFIC TO UROPATHOGENIC STRAINS

Apart from the large pathogenic islands, other regions within the genome of E. coli 536 were identified, which are absent from MG1655 and present/absent from other pathogenic strains. Table 19 summarizes the location and functional category of some genes or gene clusters found in such regions, they will be discussed here (for instance C1-C15 (C-gene cluster) marked on the Figure 33).

Generally speaking, the majority of the regions encode proteins, which have either an implication on the provision or modification of cell membrane/surface components and putative virulence factors like adhesins or are involved in special metabolic activities (substrate uptake, sugar utilization systems, mechanisms involved in pH homeostasis, etc.).

TABLE 19: SELECTION OF GENES AND CLUSTERS OF *E. COLI* 536 NOT FOUND IN MG1655, WHICH ARE LOCATED OUTSIDE OF THE LARGE PAIS.

CATEGORY	ORF NUMBER	FUNCTION	CFT 073	EDL 933
virulence-associated clusters (proteins involved in adhesion, surface modification, cell-cell interaction, invasion, resistance, iron acquisition, etc.)	RECP00109-113	colicin cluster (HNH endonuclease, immunity protein)	+/-	-
	RECP00361	putative adhesin/invasin	+	+
	RECP00374	putative autotransporter/adhesin	+	-
	RECP00423	putative autotransporter/pertactin	+	+
	RECP00675	outer membrane protein	+	-
	RECP01198-1199	putative adhesins (fragment)	+	-
	RECP01227-1232	iron transport	+	+
	RECP01331-1335	drug resistance, transmembrane proteins	+	+
	RECP01379-1380	drug resistance efflux pump	-	-
	RECP01398	putative autotransporter/adhesin	+	-
	RECP01444-1445	Rhs/Vgr-family proteins	+	+
	RECP01846	flagellin (H31)	-	-
	RECP01855	outer membrane porin	+	-
	RECP01887	outer membrane porin	+	-
	RECP01969	TonB-dependent outer membrane receptor	+	+
	RECP02038-2046, C2	O-antigen synthesis/modification	+	-
	RECP02343	outer membrane usher protein	+	+
	RECP02476	outer membrane protein, RatA homolog (Salmonella)	+	-
	RECP02477	outer membrane protein, SinI homolog (Salmonella)	+	-
	RECP02478	putative invasin/intimin, SinH homolog (Salmonella)	+	-
	RECP02897	outer membrane lipoprotein	+	+
	RECP03074-3079	iron uptake system	+	+
	RECP03142-3146, C3	fimbrial proteins	-	-
	RECP03472-3481, C4	fimbrial proteins	+	-
	RECP03558-3567	hemin transport system	+	+
	RECP03664-3665	lipoprotein, putative autotransporter/adhesin	+	+
	RECP03682-3694, C1	LPS biosynthesis	+	-
	RECP03963	hemolysin-coregulated protein (Hcp)	+	-
	RECP04046	putative RafY-like glycoporin	+	+
metabolism-associated clusters (proteins involved in uptake, substrate utilization, etc.)	RECP00690-697	membrane proteins/transporter; Fe-dehydrogenase	+	-
	RECP02352-2355, C5	sucrose utilization – PTS-independent system	-	+
	RECP02716-2720, C6	sucrose utilization – PTS-dependent system	-	-
	RECP02882-2887	putative ABC-type cobalt transport system	+	+
	RECP03056-3059	dehydrogenases	+	-
	RECP03060-3062, C15	C4-dicarboxylate transport system	+	-
	RECP03180-3183, C8	(N-acetyl)galactosamine utilization, PTS-dependent	+	+
	RECP03301-3308	putative galacticol/D-tagatose utilization	+	-
	RECP03719-3726, C9	putative fructose utilization, PTS-dependent	+	-
	RECP03964-3983, C12	putative sugar utilization, carbamate kinase	+	-
	RECP04021-4028, C10	putative sugar utilization	+	-
	RECP04033	lipase	+	-
	RECP04116-4117	hippuricase system	+	+
	RECP04166-4172, C7	sorbose utilization – PTS-dependent system	+	+
	RECP04208-4216, C13	2-oxoglutarate utilization system	+	-
	RECP04248-4252	dipeptide uptake system	+	-
	RECP04377-4388	several oxidoreductases	+	-
	RECP04427-4431, C11	arginine degradation	+	-
	RECP04583-4585, C14	C4-dicarboxylate transport system	+	-
	RECP04586	Na^+/H^+ antiporter	+	-
others	RECP00047-48	putative toxin-antitoxin system	+	+
	RECP00780	putative GipA-like factor	+	-
	RECP02893	membrane protease	+	-
	RECP03489-3490	DNA processing enzymes	+	+
	RECP04088-4092	DNA uptake/degradation	+	-
	RECP04112	starvation sensing protein	+	-
	RECP04572-4578	type I restriction-modification system	+	+

TABLE 20. *E. COLI* 536 LPS BIOSYNTHESIS CLUSTER (C1) – COMPARISON

ORF NUMBER	FUNCTION	PERCENTAGE OF IDENTITY		
		CFT073	EDL933	K12
RECP03682	2-amino-3-ketobutyrate coenzyme A ligase	96	95	95
RECP03683	ADP-L-glycero-D-manno-heptose-6-epimerase	100	99	99
RECP03684	ADP-heptose-LPS heptosyltransferase II	95	96	95
RECP03685	Lipopolysaccharide heptosyltransferase-1	100	97	94
RECP03686	Lipid A-core surface polymer ligase WaaL	92	<20	28
RECP03687	Putative beta1,3-glucosyltransferase WaaV	98	24	33
RECP03688	UDP-galactose:(galactosyl) LPS alpha1,2-galactosyltransferase WaaW	100	36	38
RECP03689	Lipopolysaccharide core biosynthesis protein WaaY	100	53	52
RECP03690	UDP-galactose:(glucosyl) LPS alpha1,2-galactosyltransferase WaaT	99	43	42
RECP03691	UDP-glucose:(glucosyl) LPS alpha1,3-glucosyltransferase WaaO	97	55	51
RECP03692	Lipopolysaccharide core biosynthesis protein WaaP	93	92	78
RECP03693	UDP-glucose:(heptosyl) LPS alpha1,3-glucosyltransferase WaaG	100	99	90
RECP03694	Lipopolysaccharide core biosynthesis glycosyl transferase WaaQ	95	95	70

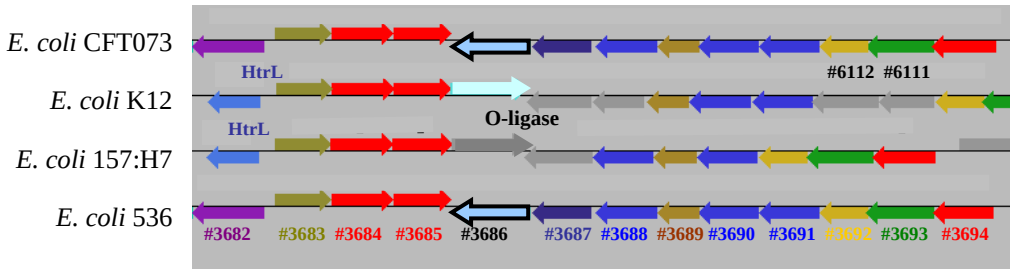


Figure 45. Contig regions for orthologs of RECP03686-94 *E. coli* 536.

Belonging to the former category, a LPS biosynthesis cluster is present (RECP03682-94, C1), which differs significantly from the one of strain MG1655, in terms of sequence similarity as well as the G+C content of this region. A separate cluster was found (RECP02038-46, C2), which encodes several glycosyltransferases unique to strains 536 and CFT073.

These enzymes are probably involved in O-antigen synthesis and modification. Again, the cluster exhibits an aberrant G+C content. MG1655 cluster possess three genes in addition: lipopolysaccharide 1,6-galactosyltransferase (REC06111), lipopolysaccharide core biosynthesis protein RfaS (REC06112) and HtrL involved in lipopolysaccharide biosynthesis phenotype (Figure 45). Table 20 shows a comparison of genes belonging to the hypervariable LPS biosynthesis cluster.

Two additional fimbrial gene clusters were found, one of which (RECP03142-46, C3) is unique to strain 536. These genes could be involved in the assembly of cable-like pili similar to the ones found in *Burkholderia cepacia*. [122] The second one (RECP03472-81, C4) is present only in the uropathogenic strains.

Two ORFs are disrupted in MG1655 but intact in strain 536, such as ORF RECP01969, a TonB-dependent receptor, showing highest similarity to outer membrane receptors for ferrienterochelin and colicins, and ORF RECP02343, an outer membrane usher protein, located within a region encoding fimbrial genes.

Some clusters encoding specialized metabolic activities will be discussed in the following. A mechanism to modulate infection of the murine urinary tract has been identified and characterized recently: the D-serine catabolism, carried out by a D-serine deaminase system, is thought to provide a growth advantage in the urinary tract. [123] The system is composed of the deaminase, a transcriptional activator and a putative D-serine permease. Examining the CFT073-homologous genomic region in strain 536 only a truncated cluster could be found, replaced partly by a sucrose utilization gene cluster. This is also the case in strain EDL933. Instead and unlike in CFT073, a second full length D-serine deaminase cluster can be found within PAI II of strain 536, which may indicate that the ability to catabolize D-serine is indeed of essential benefit in the course of an urinary tract infection.

Worth mentioning is the presence of two sucrose utilization systems, both of which are absent from CFT073 and MG1655.

The ability to utilize sucrose (Scr) as a sole carbon source is a highly variable phenotype among enteric bacteria. More than 90% of wild-type strains of *Klebsiella* spp. but less than 50% of *E. coli* and less than 10% of *Salmonella strains* are Scr^+. [124] The corresponding genes belong to the set of optional genes, which may be present in an individual genome of a species or may be lacking. [125, 126] The

characterization of such optional genes supplies valuable information about the mechanisms of horizontal gene transfer and gene adaptation in new hosts. [127]

The genes of a phosphotransferase system (PTS)-independent cluster (RECP02352-55, C5) replace the permease and regulator genes of the above-mentioned D-serine deaminase cluster. The PTS-independent system is composed of a sucrose permease (putative sucrose:H+ symporter), a fructokinase, a sucrase and a transcriptional regulator. The genes are flanked by a tRNA gene (*argW*). It has been suggested, that this tRNA gene is used by a mobile genetic element for site-specific integration into the chromosome in *E. coli* EC3132. [128]

The second sucrose utilization operon encodes a PTS-dependent system (RECP02716-20, C6), composed of a fructokinase, a sucrose porin, a sucrase, the sucrose-specific IIBC component of PTS and a sucrose operon repressor. Homologous systems are absent from *E. coli* strains CFT073, EDL933 and MG1655, but are encoded on plasmids of *Salmonella* species, *Klebsiella pneumoniae* and some *E. coli* isolates. [129-131]

A similar PTS-dependent utilization system specific for L-sorbose (RECP04166-72, C7) is present in strain 536 and other pathogenic *E. coli*, but absent from MG1655. It is composed of seven genes, which catalyze the uptake (four genes) and conversion of L-sorbose to D-fructose-6-phosphate, catalyzed by L-sorbose 1-phosphate reductase and sorbitol-6-phosphate 2-dehydrogenase. [132]

Five additional regions encoding sugar utilizing systems have been detected, RECP03180-83 (C8), RECP03719-26 (C9), RECP04021-28 (C10), RECP03964-83 (C12), and RECP03301-08 all of which are absent from MG1655. Three systems with unclear specificities occur only in the UPEC strains. One system (C8) is probably involved in the uptake and degradation of N-acetyl-galactosamine (PTS-dependent) and one (RECP03301-08) is probably responsible for galacticol/D-tagatose utilization. [133, 134]

Several genes present only in uropathogenic strains may be involved in pH homeostasis. The pH of urine usually tends to be rather acidic (varies from 4.5 to 8.0, averaging 5.5 to 6.5.). In order to respond to such acidic conditions there are several systems available in *E. coli* 536. For instance, two additional Na^+/H^+ antiporters, RECP04586 and RECP04522, could be involved in regulating the intracellular H^+ concentration. The gene of the latter antiporter is located within PAI II and absent from other *E. coli* strains.

Lysine decarboxylase has been shown to be involved in conferring acid tolerance by consuming protons and neutralizing the acidic by-products of carbohydrate fermentation. [135] Three copies, each composed of the decarboxylase gene and a lysine/cadaverine antiporter, are present in the genome of *E. coli* 536, one of which is located within the PAI III (RECP00319-20). An arginine catabolism system (RECP04427-31, C11), which is present in CFT073 but absent from other *E. coli*, could have a similar role in pH homeostasis, as it has been shown for the oral pathogen *Streptococcus gordonii*. [136] The cluster contains five genes, whose gene products catalyze the conversion of arginine to ornithine using arginine deiminase and ornithine carbamoyltransferase. An arginine/ornithine antiporter guarantees the interchangeability of substrate and product. In addition to its role in pH homeostasis, the system can also be of importance in substrate-level ATP synthesis in the course of anaerobic arginine fermentation. Carbamoyl phosphate, the by-product of the above-mentioned ornithine carbamoyltransferase reaction, is converted to NH_3 and CO_2 catalyzed by carbamate kinase, the gene of which is also encoded within the cluster. A second carbamate kinase specific to the two UPEC strains is located within a cluster (RECP03967-83, C12) encoding several metabolic enzymes with unknown specificities. A further system with implications in substrate-level ATP synthesis preferably under anaerobic conditions could be the 2-oxoglutarate degradation cluster (RECP04208-16, C13). The ATP is generated in the succinyl-CoA synthetase reaction. The cluster also contains the genes of a two-component system putatively responding to C4-dicarboxylates as well as a dicarboxylate transporter. Two other C4-dicarboxylate transport systems, RECP04583-85 (C14) and RECP03060-62 (C15), could be detected in the genome of *E. coli* 536, both of which are not present in MG1655 and EDL933. In general, C4-dicarboxylate transport systems allow C4-dicarboxylates like succinate, fumarate, and malate to be taken up. [137]

In both prokaryotic and eukaryotic organisms, peptides exhibit an important nutritional role. [138] Systems to transport peptides occur in many species and generally function to accumulate intact peptides intracellularly, where they are hydrolysed. This mechanism confers many advantages, being energetically more favourable than one of extracellular peptide cleavage followed by uptake of constituent amino acids; and, by avoiding competition between individual residues, it also allows a more balanced uptake of amino acids. Uropathogenic specific dipeptide uptake system has been localized (RECP04248-52) in the 536 genome.

It is worth to mention a hippuricase system. The ability to hydrolyze N-benzoylglycine (hippurate) to benzoic acid and glycine seems to be a strict pathogenic feature among *E. coli* strains. It is present in genomes of both uropathogenic strains, EDL933 and in genomes of such bacteria like *Pseudomonas aeruginosa* and *Burkholderia pseudomallei*.

Our bodies convert benzoic acid, which is found in most foods, into hippuric acid, a simple amide, in order to detoxify it. Furthermore hippurate keeps the urine accidic that is a natural way against bacterial infection.

Methenamine hippurate (previously known as hexamine hippurate in the UK) is an antibiotic that is used to treat and prevent urinary tract infections. It is absorbed from the gut and passes into the urine where it releases the chemical formaldehyde (however this can only occur if the urine is acidic). This chemical causes the breakdown of proteins essential to the bacteria, which ultimately results in the death of the bacteria (http://www.netdoctor.co.uk/medicines/100004085.html). According to Lee, et al. (2002), methanamine hippurate or mandelamine has not yet been conclusively shown to be useful for preventing UTI's or bacteriuria although it may reduce bacteriuria in some populations. [139]

E. coli 536 possesses two genes RECP04116 and RECP04117 (hippurate hydrolase and transmembrane transporter protein, respectively) which could create a effective hippuricase system.

5.2.3. AUTOTRANSPORTERS

All *Escherichia* strains carry a wide range of determinants involved in host interaction and virulence, many of which have been extensively studied. Numerous epidemiologic studies have suggested a role for P, type I and other fimbriae, afimbrial adhesins, hemolysin, aerobactin, capslule, and cytotoxic necrotyzing factor in the pathogenesis of urinary track infections. Comparisons of many of these systems in the four species identify differences that might bear virulence profiles of these organisms.

In contrast to Gram-positive bacteria, which possess only one biological membrane, called the cytoplasmic membrane, the cell envelope of Gram-negative bacteria is composed of two biological membranes, the inner membrane, i.e. the cytoplasmic membrane, and the outer membrane. This dual membrane envelope presents a real challenge for the transport of molecules into and out of these microorganisms. The passenger domain of an autotransporter protein corresponds to the secreted effector molecule. To date, all of the functionally characterized autotransporters have been implicated in bacterial virulence either by displaying enzymatic activity (protease, peptidase, lipase, esterase), mediating actin-promoted bacterial motility, or acting as adhesins, toxins/cytotoxins, immunomodulatory proteins, or permitting maturation of another virulent protein. Table 21 shows autotransporters found in *E. coli* genomes.

Antigen 43 (Ag43) is a self-recognizing surface adhesin. Expression of Ag43 confers aggregation and fluffing of cells, promotes biofilm formation and is associated with enhanced resistance to antimicrobial agents. Although ortholog of *flu* is present in each of the four species, we found additional genes encoding Ag43-like proteins in case of pathogenic strains (both uropathogenic (RECP00323, RCFT01230) and also the enterohaemorrhagic strain EDL9333 (RECS06274)). [140]

Several of putative autotransporters show a high homology to pertactin, or haemagglutinin protein virulence factors. Pertactin and filamentous haemagglutinin have been identified as *Bordetella* adhesins. Both proteins contain an arg-gly-asp (RGD) motif that promotes binding to integrins, known to be important in cell mobility and development. [141, 142] There is present only one such an ortholog in UPEC strains (RECP03665 and RCFT04329). These autotransporter families are characteristic for enterohaemorrhagic strain (RECS04480, RECS06178, RECS01707, RECS03081, RECS06796) but and are also present in non-pathogenic K12 srtain (REC04813, REC05325, REC05586).
Several types of autotransporters are present only in case of uropathogenic species.

We identified *Shigella flexneri* orthologs SepA and Pic in the genomes of uropathogenic srtains. SepA, extracellular serine protease, which induces mucosal atrophy and tissue inflammation, indicating that it is involved in tissue invasion by *Shigella flexneri* Construction and phenotypic characterization of a *sepA* mutant

suggested that SepA plays a role in epithelium invasion or destruction. [143] This ortholog is present in both UPEC strains (RECP00340, RCFT00374), in contrast to protein Pic, which we could find only in CFT073 strain (RCFT00333). Pic (protein involved in intestinal colonization) is an extracellular serine protease that displays in vitro mucinolytic activity, serum resistance, and hemagglutination. Both SepA and Pic belong to the Tsh subfamily, otherwise known as the SPATE (Serine Protease Autotransporters of *Enterobacteriaceae*) subfamily, and both possess a highly conserved serine protease motif shared by other members of the group. [144]

Sat is a large (107kDa) autotransporter toxin; which is secreted only by *E. coli* CFT073, a highly virulent strain isolated from blood and urine of a woman with acute pyelonephritis. Sat exhibited cytopathic activity on kidney, bladder and other cell line, as well as elicited a strong antibody response after experimentally infection. [145]
Two putative autotransporters are present in the 536 genome (RECP00423, RECP01398), which are disrupted by transposase genes in MG1655. ORF RECP00423 is present in all pathogenic strains and contains also a pertactin domain, found in cell adhesion proteins, for instance in the immunogenic outer membrane protein P69 of *Bordetella pertussis*. [146] ORF RECP01398 is 536-specific gene and encodes a putative 250kDa protein, which contains several cell-cell adhesion domains.

Hemolysins (cytolysins) were described to be important virulence factors of bacteria causing extraintestinal diseases and are active on different cells, such as lymphocytes, granulocytes, erythrocytes, and renal tubular cells. They are bacterial RTX-toxins (Repeat In Toxin) that function by assembling identical subunits into a membrane-spanning pore. Cell lysis (and death) is caused by the leakage of small molecules and ions through the large water-filled central channel.
The best characterized is α-hemolysin. [147, 148] These protein orthologs we found only in genomes of uropathogenic strains. In contrast, only in case of enterohaemorrhagic strains we located hemolysin E (HlyE) that is a novel pore-forming toxin. HlyE is unrelated to the α-hemolysins (HlyA), and the enterohaemolysin encoded by the plasmid borne ehxA gene of *E. coli* 0157:H7 Sakai. However, it is evident that expression of HlyE gives a hemolytic phenotype in *E. coli* and may contribute to the colonization of the host. [149]

TABLE 21. AUTOTRANSPORTERS ENCODED IN THE GENOMES OF *E. COLI* 536, *E. COLI* CFT073, *E. COLI* 0157:H7 SAKAI, *E. COLI* 0157:H7 AND *E. COLI* K12

	536	CFT073	O157:H7 Sakai	O157:H7	K12
SepA, serine protease	RECP00340	RCFT00374	n	n	n
Pic serine protease	n	RCFT00333	n	n	n
Autotransporter Sat	n	RCFT03537	n	n	n
Antigen 43	RECP02965	RCFT03573	RECS06067	RECO05638	REC01948
Antigen 43	RECP00323	RCFT01230	RECS06274	n	n
Hemolysin A	RECP04474	RCFT03488	n	n	n
Hemolysin A	RECP03768	n	n	n	n
Hemolysin E	RECP01213	RCFT01578	RECS01677	RECO01732	REC04802
Putative hemolysin	RECP04493	RCFT00330	n	n	n
Hemolysin in 3 fragments	n	RCFT01167-1169	n	n	n
Putative RTX family exoprotein A	n	RCFT00344	n	n	n
Putative haemagglutinin/invasin	RECP03665	RCFT04329	RECS04480	RECO04448	n
Hemagglutinin/hemolysin-related protein	n	n	RECS06178	RECO05537	n
Putatiive pertactin	n	n	RECS01707	RECO01759	REC04813
Putatiive pertactin	RECP02194/2195	RCFT02657/2658	RECS03081	RECO03026	REC05325
Putatiive pertactin	n	n	RECS06796	RECO06255	REC05586
Putative autotransporter/adhesin	RECP01398	n	RECSO06488/6487	RECO05869/5870	REC01361/1365
Putative adhesion/invasin	RECP00361	RCFT00396	RECS00336	RECO00332	n
Putative adhesion/invasin	n	n	RECS06668	RECO06100/6099	REC01929
Putative invasin	RECP01932	RCFT02374	n	n	n
Putative invasin	n	n	RECS07004	RECO06511	n
Puative autotransporter	RECP00423	RCFT00456	RECSO00424	RECO00420	REC00355/358
Putative autotransporter	n	n	n	n	REC01332
Putative autotransporter	n	RCFT02820	n	n	n
Putative autotransporter	n	n	RECS02117	REO01955	REC04960
Putative autotransporter	n	n	RECS00542	RECO00538	n
Putative autotransporter	n	n	RECS00541	RECO00537	n
Putative autotransporter	RECP00374	RCFT00407	RECSO05907	RECO05332	n
Putative autotransporter	RECP02240	RCFT02704	RECSO03116	RECO03061	REC05356
Putative autotransporter	RECP01203	RCFT01567	RECS06372	RECO05754	REC01135

Pseudogenes are underlined.

5.3. ANALYSIS OF THE POLYKETIDE SYNTHASE GENE CLUSTER (PKS-PAI (*ASNW*)) OF NEWBORN MENINGITIS *E. COLI* STRAIN IHE3034 AND UROPATHOGENIC STRAIN 536

The *asnW*-associated island is interesting as polyketide biosynthesis is uncommon for *E. coli*. The previously sequenced CFT073 genome as well as the genome of *E. coli* 536, as presented in this work, contain a cluster of genes encoding polyketide synthases (PKSs), and polyketide-modifying enzymes. Database searches of ORFs: #01945, #01946 and #01949 show highest similarities, albeit only partially, to characterized systems involved in the synthesis of melithiazol or myxothiazol by myxobacteria, both of which act as electron transport inhibitors of the respiratory chain [150, 151] and of antibiotics such as surfactin and iturin A. [152] The other proteins (#01951, #01952, #01958, #01959) show highest partial similarities to systems of the cyanobacteria *Nostoc* sp. and *Lyngbya majuscula* involved in the biosynthesis of nostopeptolide and barbamide, a chlorinated metabolite with molluscicidal activity, respectively. [153, 154] Strikingly, and in analogy to the Ybt-like synthetase genes, the corresponding genes #01946 and the 10kb gene #01959 are disrupted in CFT073.

In order to get a first glance of their function in *E. coli* 536 and *E. coli* IHE3034, the domain structures were analyzed, indicating a complex composition (Figure 46).

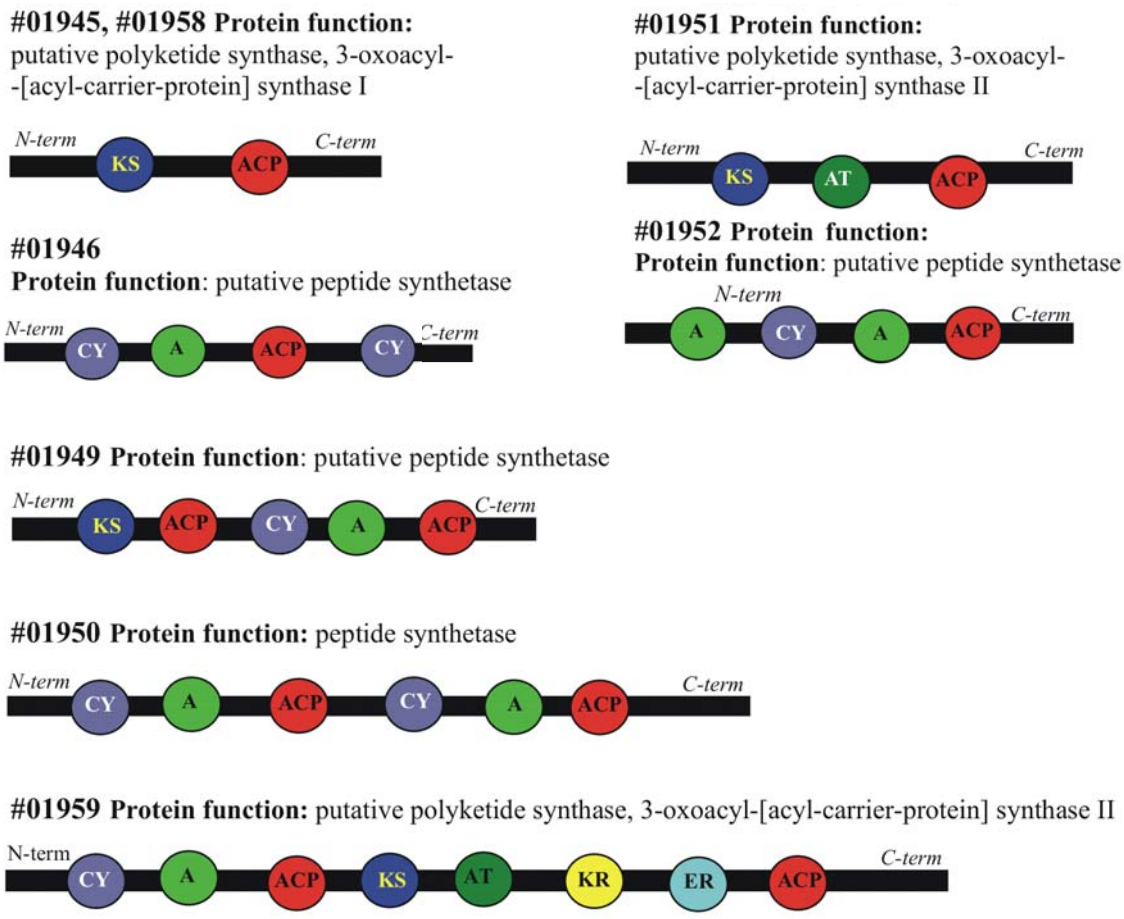


Figure 46. Functional domain structure prediction of non-ribosomal peptide synthetases/ polyketide synthases encoded within PKS- PAI (*asnW*) in *E. coli* 536 (PAI_{536} IVb) and *E. coli* IHE3034 (PKS-$PAI_{IHE3034}$). In both cases the same ORF numbers are used. The domain prediction programs PFAM and SEARCHPKS were used in combination. A, Adenylation domain (AMP-binding domain); ACP, Acyl/Aryl/Peptidyl carrier protein; AT, Acyltransferase; Cy, Cyclization/Condensation domain; ER, Enoyl reductase; KR, Ketoacyl reductase; KS, Ketoacyl synthase; MT, Methyltransferase (MT domains are weakly predicted by PFAM and SEARCHPKS); TE, Thioesterase.

Biosynthesis of polyketides shares striking similarities with the fatty acid biosynthesis. Polyketides are synthesized by sequential reactions catalysed by a collection of enzyme activities called polyketide synthases (PKSs). These are large multienzyme protein complexes that contain a coordinated group of active sites. The PKS genes for a certain polyketide are usually organized in one operon. At least two

architecturally different types of PKSs have been discovered in the microbial world. Type I systems consist of very large multifuctional proteins which can be either processive or iterative. These PKSs are large multidomain proteins carrying all the active sites required for polyketide biosynthesis. The iterative Type II systems consist of complexes of mono-functional proteins. In these synthases, active sites are distributed among several smaller, typically monofunctional polypeptides. Type II synthases catalyse the formation of compounds that require aromatization and cyclization, but not extensive reduction or reduction/dehydration cycles. Generally speaking, the order of modules and domains of a complete polyketide synthase is as follows (in the order N-terminus to C-terminus): starting or loading module: AT-ACP-; elongation or extending modules: -KS-AT-[DH-ER-KR]-ACP-; termination or releasing module: -TE. The polyketide chain and the starter groups are bound with their carboxy group to the SH groups of the ACP and the KS domain through a thioester linkage: R-C(=O)OH + HS-protein <=> R-C(=O)S-protein + H_2O. The growing chain is handed over from one SH group to the next by trans-acylations and is releases at the end by hydrolysis or by cyclization. There are three biosynthesis stages. During the starting stage the starter group, usually acetyl-CoA or malonyl-CoA, is loaded onto the ACP domain of the starter module catalyzed by the starter module AT domain. The elongation stage itself is a complex path where reactions go as following:

- The polyketide chain is handed over from the ACP domain of the previous module to the KS domain of the current module, catalyzed by the KS domain.
- The elongation group, usually malonyl-CoA or methyl-malonyl-CoA, is loaded onto the current ACP domain catalyzed by the current AT domain.
- The ACP-bound elongation group reacts in a Claisen condensation with the KS-bound polyketide chain under CO2 evolution, leaving a free KS domain and an ACP-bound elongated polyketide chain. The reaction takes place at the KS-bound end of the chain, so that the chain moves out one position and the elongation group becomes the new bound group.
- Optionally, additional domains can alter the new fragment of the polyketide chain stepwise. The KR (keto-reductase) domain reduces the β-keto group to a β-hydroxy group, the DH (dehydratase) domain splits off H_2O, resulting in the

α-β-unsaturated alkene, and the ER (enoyl-reductase) domain reduces the α-β-double-bond to a single-bond.

This cycle is repeated for each elongation module. And finally in the termination stage the completed polyketide chain is hydrolyzed be the TE (thio-esterase) domain from the ACP-domain of the previous module. Given that a structure of the *asnW*-associated polyketide might be predicted. Figure 47 shows the proposed structure of the putative polyketide compound (Dobrindt, U., not published.). The order of the different building blocks that are separated by the red line is so far unknown. The order of the genes coding for PKS- or NRPS cluster usually correlates well with the order of building blocks within the final molecule. However, maybe not in case of the *asnW*-associated island: in the KS module of ORF1951, a serine residue replaces a conserved cysteine residue, which is typical for the PKS module that builds the "start unit". However, #01951 is not located at the beginning of the gene cluster, but in between other PKS genes.

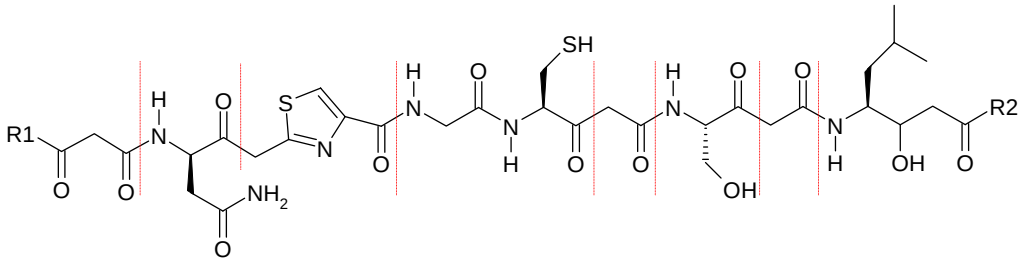


Figure 47. Proposed structure of the putative polyketide produced by the *asnW* island-encoded gene products.

To gain first insights into the function of the PKS clusters and its occurrence in an *E. coli* strains effects on human cervix epithelioid carcinoma cells are examined (HeLa cells). A total of 185 isolates (Table 11) were subjected to PCR assay. The results of PCR demonstrated that 61 (33%) did generate a PKS pattern (Table 17) and polyketide-PAI (*asnW*) occurred only among extraintestinal strains. None of enteropathogenic, enteroinvasive, enterotoxigenic, enteroaggregative or enterohemorrhagic strains amplified the PKS-PAI (*asnW*) sequence. Subsequently, non-hemolysin-producing isolates were tested for cytocidal activity. Of 19 *E. coli* strains, the test successfully identified 13 positives isolates.

Interestingly, the results showed that a polyketide-derived toxin causes cythopathicity and the arrest of the cell cycle in HeLa cell culture, although the presence of the PKS-PAI is not a guarantee for an effective secretion of PKS-derived substances.

To explain this phenomenon, since nucleotide sequence for three pathogenic strains is available, the genetic background was investigated.

To keep in mind, infection with uropathogenic strain 536 induces no morphologic changes after co-incubation with eukaryotic cell line, while the second UPEC CFT073 showed a significant cytotoxic effect. The meningitis causing isolate IHE3034 gives also PKS$^+$ phenotype, although it seems that the IHE3034 PKS toxin has got a weaker activity. The IHE3034 PKS-PAI region shows 99% identity to homologue regions from uropathogenic strains *E. coli* 536 and CFT073. (Figure 48.)

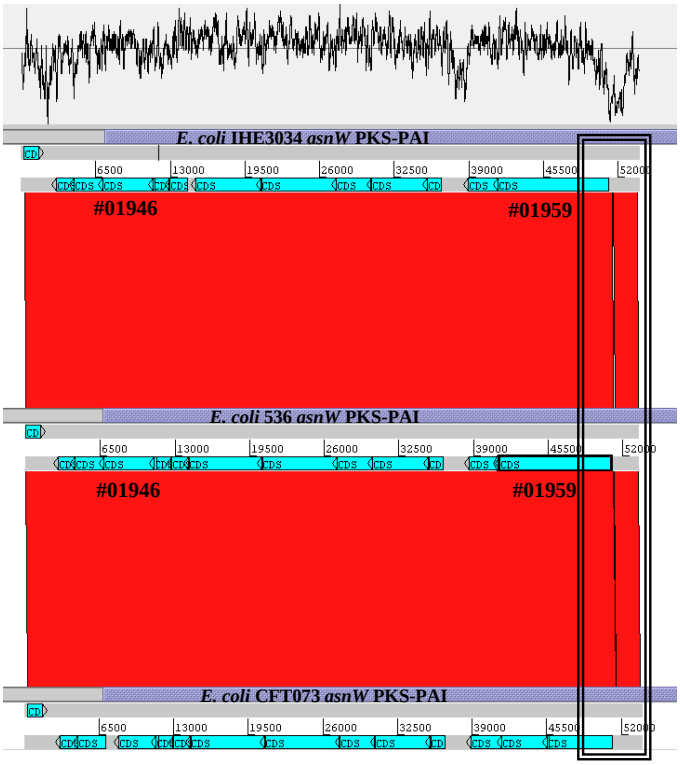


Figure 48. Linear comparison of *asnW* PKS-PAI *E. coli* IHE3034, 536 and CFT073. Block indicates a putative regulatory region, which differs in all strains.

As mention above, two genes from the region are disrupted in CFT073 (#01946 and #01959 homologs). Apparently, sequencing mistakes exist in the genome of CFT073.

The analysis of the short region upstream of #01956 the first gene of the PKS operon and followed by a putative transcriptional regulator gene consists of a series of repeat blocks $[ACAGATAC]_n$ (Figure 49). The amount of blocks is different in all investigated strains. It is 13 for IHE3034, 10 for *E. coli* 536 and finally 7 for CFT073. This observation leads to the hypothesis that the effect of PKS toxin is connected with the number of repeats in this regulatory region. Probably it indicates PKS^+/PKS^- phenotype (536: PKS^- in comparison to IHE3034: PKS^+), as well as it might regulate the expression level (differences between IHE3034 and CFT073).

```
IHE3034     TGCCAAAACAACAGATACACAGATACACAGATACACAGATACACAGATACACAGATACAC
    536     TGCCAAAACAACAGATACACAGATACACAGATACACAGATACACAGATACACAGATACAC
 CFT073     TGCCAAAACAACAGATACACAGATACACAGATACACAGATACACAGATACACAGATACAC
            ************************************************************

IHE3034     AGATACACAGATACACAGATACACAGATACACAGATACACAGATACACAGATACCATTCA
    536     AGATACACAGATACACAGATACACAGATAC------------------------CATTCA
 CFT073     AGATAC----------------------------------------------CATTCA
            ******                                              ******

IHE3034     GGATGGGGGGAAACATGGATAAGTTCAAAGAAAAAAACCCGTTATCTCTGCGTGAAAGAC
    536     GGATGGGGGGAAACATGGATAAGTTCAAAGAAAAAAACCCGTTATCTCTGCGTGAAAGAC
 CFT073     GGATGGGGGGAAACATGGATAAGTTCAAAGAAAAAAACCCGTTATCTCTGCGTGAAAGAC
            ************************************************************
```

Figure 49. Sequence alignment of a regulatory region upstream of #01959.

The discovery that extraintestinal *E. coli* strains produce a polyketide-derived toxin is highly significant. Polyketides are lipid-like molecules that, although relatively small compared with proteineous toxins, have potent biological activities. Well-known polyketides include antibiotics (erythromycin), immunosuppressants, (rapamycin, FK506), antifungal agents (amphotericin B), antihelmetic agents (avermectin), and cytostatins (bafilomycin). Most complex polyketides are made as secondary metabolites by soil bacteria in the order Actinomycetales. [155-157]

However no polyketide causes cythopathicity and cell cycle arrest in cell culture. This has now been shown in *E. coli*. Although, most of the pathogenic *E. coli* strains

remain extracellular, UPEC can invade and replicate within uroepithelial cells, it is possible that they play an important role in intracellular survival or/and protects *E. coli* from predatory eukaryotes in its natural habitat. This complex polyketide may represent the first of a newly discovered class of virulence factors associated with pathology in diseases caused by extraintestinal *E. coli* strains. Next, the compound should be purified and characterized in details.

6. SUMMARY

6.1. GENOME PROJECT *ESCHERICHIA COLI* 536

6.1.1. The genome sequence of *Escherichia coli* human pyelonephritis isolate 536 was obtained and compared to the to the genome sequence of strain CFT073 (O6:K2:H1), another well-studied highly virulent uropathogenic *E. coli* (UPEC) strain, as well as to the genomes of non-pathogenic MG1655 (K12) and enterohemorrhagic strain EDL933 (O157:H7).

6.1.2. The single circular chromosome of *E. coli* 536 is composed of 4,938,875 bp. The genome was sequenced with a 6-fold coverage. The whole chromosome encodes 4,747 putative genes, no extrachromosomal plasmid was found. The genome of strain 536 contains only one cryptic prophage region. Analysis of base composition showed the G+C content of the 536 genome is 50.5%.

6.1.3. The four-way comparison revealed a highly mosaic genome structure. It is apparent that genomic differences are not exclusively linked to the presence or absence of large pathogenic islands PAIs. Instead, also the distribution of smaller gene clusters seems to confer strain-specific traits. The comparison revealed a set of 3650 “backbone” genes (77%), which are present in all *E. coli* strains and are essentially co-linear.

6.1.4. The genome comparison of three pathogenic and one non-pathogenic strain of *E. coli* revealed only 78 genes, which are present exclusively in all three pathogenic strains. These genes, arranged in small gene clusters that are scattered around the genome, can be classified in several heterogeneous functional categories such as transport (35 genes), iron acquisition (17), drug resistance (5), DNA modification (4), and others. No

main virulence determinants are among these proteins. Rather, they seem to have an implication in improving the pathogens' fitness with respect to e.g. utilization of additional nutrients. Apparently, no broad 'pathogenic backbone' has evolved in disease-causing *E. coli* isolates.

6.1.5. Although the genome of *E. coli* CFT073 is about 292 kb larger than the one of strain 536, much of the additional information seems to be of no functional implication on virulence, as over 50% of these additional genes are located within cryptic prophages. In addition, the sequencing project reveals about 470 genes of *E. coli* 536, which are absent from CFT073 and MG1655. These findings indicate that the differences between strains 536 and CFT073 in their potential to cause disease as well as in the graveness of the urinary tract infection are not due to a simple gene loss or gain, respectively. Rather, they seem to be the consequence of using a different set of genes, which results in an individual virulent phenotype.

6.1.6. A major surprise from the uropathogenic *E. coli* genome projects was the discovery that the genomes contain a number of polyketide synthesis (PKS) genes. The region, designated PAI_{536} IVb, encodes several large proteins with a predicted molecular weight of up to 350kDa, eight of which fall into the class of the above-mentioned NRPS/PKS (#01945, #01946 and #01949, #01950, #01951, #01952, #01958, #01959).

6.2. SEQUENCING AND ANALYSIS OF PLASMID pARK2 ENTEROTOXIGENIC *ESCHERICHIA COLI* STRAIN 2173

6.2.1. The complete sequence of the 88,480 bp *E. coli* 2173 pARK2 plasmid was obtained. The plasmid was sequenced with a 9-fold coverage. The coding regions account for 83.6% of the plasmid. Analysis of base composition revealed a G+C content of the plasmid of 51%. pARK2 comprises

a mosaic structure of potential pathogenesis-associated genes, IS elements, maintenance genes, transfer genes and unknown ORFs.

6.2.2. 104 ORFs were identified, most of them are encoded on the same DNA strand. Plasmid pARK2 consists of two large distinct regions: a 71.7 kb large region that is essentially homologous to the non-R-determinant region of plasmid R100, and a 16.7 kb region which carries the *sta* and *stb* genes encoding heat-stable enterotoxins.

6.3. SEQUENCING AND ANALYSIS OF A POLYKETIDE SYNTHASE GENE CLUSTER (PKS-PAI (*ASNW*)) OF THE NEWBORN MENINGITIS ISOLATE *E. COLI* IHE3034

6.3.1. The whole sequence of the IHE3034 BAC11 insert was obtained and analysed. It is composed of 67,932 bp and was sequenced with a 9-fold coverage. 41 ORFs were identified, and most of them are oriented in the same direction.

6.3.2. Obtained protein-coding regions include the PKS-PAI associated with the *asnW* tRNA gene and 15 flanking ORFs. The PKS-PAI shows 99% identity on nucleotide level to the homologous region of the sequenced uropathogenic strains *E. coli* 536 and CFT073. Analysis of base composition showed a G+C content of the insert of 53,2%, which is higher than in the *E. coli* 536 whole genome (50,52%).

6.3.3. In order to investigate the occurrence of the identified PKS cluster in an *E. coli* strain collection PCR screening was applied. The PCR analysis demonstrated that 61 (33%) of 185 strains were PKS-positive. Results show that polyketide-PAI (*asnW*) occurred only among extraintestinal strains. None of enteropathogenic, enteroinvasive, enterotoxigenic,

enteroaggregative or enterohemorrhagic strains amplified the PKS-PAI (*asnW*) sequences.

6.3.4. The goal of this study was to gain first insight into the function of the PKS cluster. For this reason, the effect of potentially secreted compounds was examined on human cervix epithelioid carcinoma cells (HeLa cells). In conclusion, of 19 *E. coli* strains, the test successfully identified 13 positive isolates.

6.3.5. PKS gene cluster encodes a compound inducing morphologic changes on human cervix epithelioid carcinoma cells. It was demonstrated that a polyketide-derived toxin causes cythopathicity and cell cycle arrest in HeLa cell culture.

7. REFERENCES

1. **Whittam, T.S.,** 1996: Genetic variation and evolutionary processes in natural populations of *Escherichia coli*. In *Escherichia coli* and *Salmonella*: cellular and molecular biology. F.C. Neidhardt, editor. ASM Press. Washington, DC, USA. 2708–2720.
2. **Pupo, G.M., Karaolis, D.K.R., Lan, R.T., and Reeves, P.R.** 1997: Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. Infect. Immun. 65:2685–2692.
3. **Stamm, W.E., Hooton, T.M.,** 1993: Management of urinary tract infections in adults. N Engl J Med 329(18):1228-1334.
4. **Warren, J.W., Abrutyn, E., Hebel, J.R., et al.,** 1999: Guidelines for antimicrobial treatment of uncomplicated acute bacterial cystitis and acute pyelonephritis in women. Clin Infect Dis 29:745-58.
5. **Werman, H.A., Brown, C.G.,** 1986: Utility of urine cultures in the emergency department. Ann Emerg Med 15(3):302-7.
6. **Foxman, B., Barlow, R., D'arcy H., et al.,** 2000: Urinary tract infection: self-reported incidence and associated costs. Ann Epidemiol 10:509-15.
7. **Johnson, J.R.,** 1991: Virulence factors in *Escherichia coli* urinary tract infection. Clin Microbiol Rev. 4 (1):80–128.
8. **Mulvey, M.A., Schilling, J.D., Martinez, J.J., and Hultgren, S.J.,** 2000: Bad bugs and beleaguered bladders: Interplay between uropathogenic *Escherichia coli* and innate host defences. Proc Natl Acad Sci USA. 1;97(16):8829-35.
9. **Schilling, J.D., Mulvey, M.A, and Hultgren, S.J.,** 2001: Structure and Function of *Escherichia coli* Type 1 Pili: New Insight into the pathogenesis of Urinary Tract Infections J Infect. Dis. 1;183 Suppl 1:S36-40.
10. **Soto, G.E., Hultgren, S.J.,** 1999: Bacterial Adhesins: Common themes and variations in architecture and assembly. J. Bacteriol. 181: 1059-1071.
11. **Balsalobre, C., Morschhauser, J., Jass, J., Hacker, J., Uhlin, B.F.,** 2003: Transcriptional Analysis of the *sfa* Determinant Revealing Multiple mRNA

Processing Events in the Biogenesis of S Fimbriae in Pathogenic *Escherichia coli* J. Bacteriol. 185: 620-629.

12. **Alexander, C., Rietschel, E.T.,** 2001: Bacterial lipopolysaccharides and innate immunity. Journal of Endotoxin Research, Vol. 7, No. 3,167-202.
13. **Ohkawa, M., Sugata, T., Sawaki, M., Nakashima, T., Fuse, H., and Hisazumi, H.,** 1990: Bacterial and crystal adherence to the surfaces of indwelling urethral catheters. J. Urol. 143:717-721.
14. **Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R, and Lappin-Scott, H.M.,** 1995: Microbial biofilms. Annu. Rev. Microbiol. 49:711-745.
15. **Ladd, T.I., Schmiel, D., Nickel, J.C., and. Costerton, J.W.,** 1987: The use of a radiorespirometric assay for testing the antibiotic sensitivity of catheter-associated bacteria. J. Urol. 138:1451-1456.
16. **Costerton, J.W., Stewart, P.S., and Greenberg, E.P.,** 1999: Bacterial biofilms: a common cause of persistent infections. Science 284:1318-1322.
17. **Hooton, T.M., Stamm, W.E.,** 1997: Diagnosis and treatment of uncomplicated urinary tract infection. Infect Dis Clin North Am 11:551-81.
18. **Anderson, G.G. et al.,** 2003: Intracellular Bacterial Biofilm-Like Pods in Urinary Tract Infections. Science 301:105-107.
19. **Raetz C.R.H, Whitfield, C.,** 2002: Lipopolysaccharide endotoxins. Annual Review of Biochemistry 71, 635-700.
20. **Schmidt, G.R.G., Hofmann, F., and Aktories, K.,** 1998: Activation of Rho GTPases by *Escherichia coli* cytotoxic necrotizing factor 1 increases intestinal permeability in Caco-2 cells. Infect. Immun. 66:5125-5131.
21. **Rippere-Lampe, K.E., O'Brien, A.D., Conran, R., Lockman, H.A.,** 2001: Mutation of the Gene Encoding Cytotoxic Necrotizing Factor Type 1 (*cnf1*) Attenuates the Virulence of Uropathogenic *Escherichia coli* Infect. Immun. 69: 3954-3964.
22. **Guyer, D.M., Henderson, I.R., Nataro, J.P., and Mobley, H.L.T.,** 2000: Identification of Sat, an autotransporter toxin produced by uropathogenic *Escherichia coli*. Molecular Microbiology 38 (1), 53-66.
23. **Eslava, C., Navarro-Garcia, F., Czeczulin, J.R., Henderson, I.R., Cravioto, A., and Nataro, J.P.,** 1998: Pet, an autotransporter enterotoxin from enteroaggregative *Escherichia coli*. Infect. Immun. 66:3155-3163.

24. **Stein, M., Kenny, B., Stein, M.A., and Finlay, B.B.,** 1996: Characterization of EspC, a 110-kilodalton protein secreted by enteropathogenic *Escherichia coli*, which is homologous to members of the immunoglobulin, A protease-like family of secreted proteins. J. Bacteriol. 178:6546-6554.

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

26. **Al Hasani, K., Henderson, I.R., Sakellaris, H., Rajakumar, K., Grant, T., Nataro, J.P., Robins-Browne, R., and Adler, B.,** 2000: The *sigA* gene which is borne on the pathogenicity island of *Shigella flexneri* 2a encodes an exported cytopathic protease involved in intestinal fluid accumulation. Infect. Immun. 68:2457-2463.

27. **Benjelloun-Touimi, Z., Sansonetti, P.J., and Parsot, C.,** 1995: SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion. Mol. Microbiol. 17:123-135.

28. **Provence, D.L., and Curtiss III, R.,** 1994: Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. Infect. Immun. 62:1369-1380.

29. **Brunder, W., Schmidt, H., and Karch, H.,** 1997: EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. Mol. Microbiol. 24:767-778.

30. **Navarro-Garcia, F., Sears, C., Eslava, C., Cravioto, A., and Nataro, J.P.,** 1999: Cytoskeletal effects induced by Pet, the serine protease enterotoxin of enteroaggregative *Escherichia coli*. Infect. Immun. 67:2184-2192.

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

32. **Sambrook, J., Fritsch, E.F., Maniatis, T.,** 1989: Molecular cloning: a laboratory manual (2nd ed.). Cold Spring Habour Laboratory Press, Cold Spring harbour, New York.

33. **Hanahan, D.,** 1983: Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166(4):557–580.

34. **Staden, R., Beal, K.F., Bonfield, J.K.,** 2000: The Staden package, 1998. Methods Mol Biol. 132:115-30.

35. **Bonfield, J.K. and Staden, R.,** 1996: Experiment files and their application during large-scale sequencing projects. DNA Sequence 6, 109-117.

36. **Bonfield, J.K., Smith, K.F. and Staden, R.,** 1995: A new DNA sequence assembly program. Nucleic Acids Res. 24, 4992-4999.

37. **Tech M. and Merkl, R.,** 2003: YACOP: Enhanced gene prediction obtained by a combination of existing methods, In Silico Biology 3, 0037; ©2003, Bioinformation Systems e.V.

38. **Tatusov, R.L., Koonin, E.V., Lipman, D.J.A.,** 1997: Genomic perspective on protein families. Science 24;278(5338):631-7.

39. **Sonnhammer, E.L., Eddy, S.R., Durbin, R.,** 1997: Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins. 28(3):405-20.

40. **Bateman, A, Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones. S., Howe, K.L., Marshall, M., Sonnhammer, E.L.,** 2002: The Pfam Protein Families Database Nucleic Acids Research 30(1):276-280.

41. **Hulo, N., Sigrist, C.J.A., Le Saux, V., Langendijk-Genevaux, P.S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., Bairoch, A.,** 2004: Recent improvements to the PROSITE databaseNucl. Acids. Res. 32:D134-D137.

42. **Juncker, A.S., Willenbrock, H., von Heijne, G., Nielsen, H., Brunak, S. and Krogh, A.,** 2003: Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. 12(8):1652-62.

43. **Nielsen H., Engelbrecht, J., Brunak, S. and von Heijne, G.,** 1997: Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering, 10:1-6.

44. **Cserzo, M., Wallin, E., Simon, I., von Heijne, G. and Elofsson, A.,** 1997: Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the Dense Alignment Surface method; Prot. Eng. vol. 10, no. 6, 673-676.

45. **Hofmann, K., and Stoffel, W.,** 1993: TMbase - A database of membrane spanning proteins segments Biol. Chem. Hoppe-Seyler 374,166.

46. **Lander, E. and Waterman, M. S.,** 1988: Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics 2, 231-239.

47. **Delcher, A.L., Kasif, S., Fleischmann, R.D., Peterson, J., White, O., and Salzberg, S.L.,** 1999: Alignment of whole genomes. Nucleic Acids Research, 27:11, 2369-2376.

48. **Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., Shinagawa, H.,** 2001: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 8, 11-22.

49. **Betley, M.J., Miller, V.L., and Mekalanos, J.J.,** 1986: Genetics of bacterial enterotoxins. Annu. Rev. Microbiol. 40:577-05.

50. **Fekete, P.Z, Schneider, G., Olasz, F., Blum-Oehler, G., Hacker, J.H., Nagy, B.,** 2003: Detection of a plasmid-encoded pathogenicity island in F18+ enterotoxigenic and verotoxigenic *Escherichia coli* from weaned pigs. Int. J. Med. Microbiol. 293 (4) 287-298.

51. **Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.,** 1997: "Gapped BLAST and PSI-BLAST: a new generation of protein database searchprograms", Nucleic Acids Res. 25:3389-3402.

52. **Bisercic, M., Ochman, H.,** 1993: The ancestry of insertion sequences common to *Escherichia coli* and *Salmonella typhimurium*. J Bacteriol. 175(24):7863-8.

53. **Dutreix, M., Backman, A., Celerier, J., Bagdasarian, M.M., Sommer, S., Bailone, A., Devoret, R., Bagdasarian, M.,** 1988: Identification of psiB genes of plasmids F and R6-5. Molecular basis for *psiB* enhanced expression in plasmid R6-5 Nucl. Acids Res. 16: 10669-10679.

54. **Bailone, A., Bäckman, A., Sommer, S., Célérier, J., Bagdasarian, M.M., Bagdasarian, M., Devoret, R.,** 1988: PsiB polypeptide prevents activation of RecA protein in *Escherichia coli*. Mol Gen Genet 214: 389–395.

55. **Howland, C.J., Rees, C.E.D., Barth, P.T., Wilkins, B.M.,** 1989: The *ssb* gene of plasmid ColIb-P9. J Bacteriol 171: 2466–2473.

56. **Golub, E.I., Bailone, A., Devoret, R.**, 1988: A gene encoding an SOS inhibitor is present in different conjugative plasmids. J Bacteriol 170: 4392–4394.

57. **Loh, S.M., Cram, D.S., Skurray, R.A.,** 1988: Nucleotide sequence and transcriptional analysis of a third function (Flm) involved in F-plasmid maintenance. Gene 66:259-268.

58. **Rasmussen, P.B., Gerdes, K., Molin, S.,** 1987: Genetic analysis of the parB+ locus of plasmid R1. Mol Gen Genet. 209(1):122-8.

59. **Gerdes, K, Molin, S.,** 1986: Partitioning of plasmid R1. Structural and functional analysis of the *parA* locus. J Mol Biol. 5;190(3):269-79.

60. **Stoll, B.J.,** 1997: The global impact of neonatal infection. Clin Perinatol 24:121.

61. **Dietzman, D.E., Fischer, G.W., Schoenknecht, F.D.,** 1974: Neonatal *Escherichia coli* septicemia: bacterial counts in blood. J Pediatr; 85:12830.

62. **Kim, K.S.,** 2001: *Escherichia coli* translocation at the blood-brain barrier. Infect. Immun. 69:521722.

63. **Glode, M.P., Sutton, A., Moxon, E.R., Robbins, J.B.,** 1977: Pathogenesis of neonatal *Escherichia coli* meningitis: induction of bacteremia and meningitis in infant rats fed *Escherichia coli* K1. Infect. Immun. 16:7580.

64. **Bortolussi, R., Ferrieri, P., Wannamaker, L.W.,** 1978: Dynamics of *Escherichia coli* infection and meningitis in infant rats. Infect. Immun. 22:4805.

65. **Kim, K.S., Itabashi, H., Gemski, P., Sadoff, J., Warren, R.L., Cross, A.S.,** 1992: The K1 capsule is the critical determinant in the development of *Escherichia coli* meningitis in the rat. J Clin Invest 90:897905.

66. **Pluschke, G., Mayden, J., Achtman, M., Levine, R.P.,** 1983: Role of the capsule and the O antigen in resistance of O18:K1 *Escherichia coli* to complement-mediated killing. Infect. Immun. 42:90713.

67. **Johnson, J.R., Russo, T.A., Tar,r P.I., et al.,** 2000: Molecular epidemiological and phylogenetic associations of two novel putative virulence genes, *iha* and iron *E. coli*, among *Escherichia coli* isolates from patients with urosepsis. Infect. Immun. 68:30407.

68. **Russo, T.A., Carlino, U.B., Mong, A., Jodush, S.T.,** 1999: Identification of genes in an extraintestinal isolate of *Escherichia coli* with increased expression after exposure to human urine. Infect. Immun. 67:530614.

69. **Torres, A.G., Payne, S.M.,** 1997: Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7. Mol Microbiol 23:82533.

70. **Johnson, J.R.,** 1991: Virulence factors in *Escherichia coli* urinary tract infection. Clin Microbiol Rev 4:80128.

71. **Carniel, E.,** 2001: The *Yersinia* high-pathogenicity island: an iron-uptake island. Microbes Infect 3:5619.

72. **Schubert, S., Rakin, A., Karch, H., Carniel, E,, Heesemann, J.,** 1998: Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. Infect. Immun. 66:4805.

73. **Clermont, O., Bonacorsi, S., Bingen, E.,** 2001: The *Yersinia* high-pathogenicity island is highly predominant in virulence-associated phylogenetic groups of *Escherichia coli*. FEMS Microbiol Lett 196:1537.

74. **Bhakdi, S., Muhly, M., Korom, S., Schmidt, G.,** 1990: Effects of *Escherichia coli* hemolysin on human monocytes: cytocidal action and stimulation of interleukin 1 release. J Clin Invest 85:174653.

75. **Welch, R.A., Dellinger, E.P., Minshew, B., Falkow, S.,** 1981: Haemolysin contributes to virulence of extra-intestinal *Escherichia coli* infections. Nature 294:6657.

76. **May, A.K., Gleason, T.G., Sawyer, R.G., Pruett, T.L.,** 2000: Contribution of *Escherichia coli* -hemolysin to bacterial virulence and to intraperitoneal alterations in peritonitis. Infect. Immun. 68:17683.

77. **Hoffman, J.A., Wass, C., Stins, M.F., Kim, K.S.,** 1999: The capsule supports survival but not traversal of *Escherichia coli* K1 across the blood-brain barrier. Infect. Immun. 67:356670.

78. **Parkkinen, J., Korhonen, T.K., Pere, A., Hacker, J., Soinila, S.,** 1988: Binding sites in the rat brain for *Escherichia coli* S fimbriae associated with neonatal meningitis. J Clin Invest 81:8605.

79. **Korhonen, T.K., Valtonen, M.V., Parkkinen, J., et al.,** 1985: Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. Infect. Immun. 48:48691.

80. **Stins, M.F., Prasadarao, N.V., Ibric, L, Wass, C.A., Luckett, P., Kim, K.S.,** 1994: Binding characteristics of S fimbriated *Escherichia coli* to isolated brain microvascular endothelial cells. Am J Pathol 145:122836.

81. **Huang, S.H., Wan, Z.S., Chen, Y.H., Jong, A.Y., Kim, K.S.,** 2001: Further characterization of *Escherichia coli* brain microvascular endothelial cell invasion gene ibeA by deletion, complementation, and protein expression. J Infect Dis 183:10718.

82. **Huang, S.H., Wass, C., Fu, Q., Prasadarao, N.V., Stins, M., Kim, K.S.,** 1995: *Escherichia coli* invasion of brain microvascular endothelial cells in vitro and in vivo: molecular cloning and characterization of invasion gene *ibe10*. Infect. Immun. 63:44705.

83. **Khan, N.A., Wang, Y., Kim, K.J., Chung, J.W., Wass, C.A., Kim, K.S.,** 2002: Cytotoxic necrotizing factor1 contributes to *Escherichia coli* K1 invasion of the central nervous system. J Biol Chem 277:1560712.

84. **Badger, J.L., Wass, C.A., Weissman, S.J., Kim, K.S.,** 2000: Application of signature-tagged mutagenesis for identification of *Escherichia coli* K1 genes that contribute to invasion of human brain microvascular endothelial cells. Infect. Immun. 68:505661.

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

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

87. **Bauer, M.E., and Welch, R.A.,** 1996: Association of RTX toxins with erythrocytes. Infect. Immun. 64:4665-4672.

88. **May, A.K., Sawer, R.G., Gleason, T., Whitworth, A. and Pruett, T.L.,** 1996: *In vivo* cytokine response to *Escherichia coli* alpha-hemolysin determined with genetically engineered hemolytic and nonhemolytic *E. coli* variants. Infect. Immun. 64:2167-2171.

89. **Menestrina, G., Pederzolli, C., Dalla Serra, M., Bregante, M. and Gambale, F.,** 1996: Permeability increase induced by *Escherichia coli*

Hemolysin A in human macrophages is due to the formation of ionic pores: A patch clamp characterization. Journal of Membrane Biology 149:113-121.

90. **Naas, T., Blot, M., Fitch, W.M., and Arber, W.,** 1994: Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. Genetics 136:721-730.
91. **Naas, T., Blot, M., Fitch, W.M., and Arber, W.,** 1995: Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. Mol. Biol. Evol. 12:198-207.
92. **Galas, D.J. and Chandler, M.,** 1989: Bacterial insertion sequences. In D.E. Berg and M.M. Howe (eds.), Mobile DNA. American Society for Microbiology, Washington, D.C., pp. 109–162.
93. **Sawyer, S., Dykhuizen, D., DuBose, R., Green, L., Mutanga- Dura-Mhlanga T., Wolczyk, D., and Hartl, D.,** 1987: Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. Genetics 115: 51–63.
94. **Wilde, C., Bachellier, S., Hofnung, M., Clement, J.M.,** 2001: Transposition of IS1397 in the family Enterobacteriaceae and first characterization of ISKpn1, a new insertion sequence associated with *Klebsiella pneumoniae* palindromic units J. Bacteriol. 183: 4395-4404.
95. **Anderson, P.M., Sung, Y.C., Fuchs, J.A.,** 1990: The cyanase operon and cyanate metabolism. FEMS Microbiol Rev. 7(3-4):247-52.
96. **Burlingame, R., Chapman, P.J.,** 1983: Catabolism of phenylpropionic acid and its 3-hydroxy derivative by *Escherichia coli*. J Bacteriol. 155(1):113-21.
97. **Lam, S.T. and Roth, J.R.,** 1983: IS200: a *Salmonella*-specific insertion sequence. Cell 34:951-960.
98. **Gibert, I., Barbé, J., Casadesús, J.,** 1990: Distribution of insertion sequence IS200 in *Salmonella* and *Shigella*. J. Gen. Microbiol. 136, 2555-2560.
99. **Bisercic, M., and Ochman, H.,** 1993: Natural populations of *Escherichia coli* and Salmonella typhimurium harbor the same classes of insertion sequences. Genetics 133:449-454.
100. **Simonet, M, Riot, B., Fortineau, N., Berche, P.,** 1996: Invasin production by *Yersinia pestis* is abolished by insertion of an IS200-like element within the inv gene. Infect. Immun. 64(1):375-9.

101. **Lam, S., and Roth J.R.,** 1983: Genetic mapping of IS200 copies in *Salmonella typhimurim* strain lT2. Genetics 105: 801-811.

102. **Sanderson, K.E., Sciore, P., Liu, S.L., and Hessel, A.,** 1993: Location of IS200 on the genomic cleavage map of *Salmonella typhimurium* LT2. J. Bacteriol. 175:7624-7628.

103. **Beuzon, C.R., Schiaffino, A., Leori, G., Cappuccinelli, P., Rubino, S., Casadesus, J.,** 1997: Identification of *Salmonella abortusovis* by PCR amplification of a serovar-specific IS200 element. Appl. Environ. Microbiol. 63: 2082-2085

104. **Dobrindt, U., Blum-Oehler, G., Nagy, G., Schneider, G., Johann, A., Gottschalk, G., Hacker, J.,** 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.

105. **Dobrindt, U., et al.,** 2004: Analysis of PAI V from *Escherichia coli* 536, in preparation.

106. **Hacker, J., Kaper, J.B.,** 2000: Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54, 641-679.

107. **Guyer, D.M., Kao, J.S., Mobley, H.L.,** 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.

108. **Rasko, D.A., Phillips, J.A., Li, X., Mobley, H.L.,** 2001: Identification of DNA sequences from a second pathogenicity island of uropathogenic *Escherichia coli* CFT073: probes specific for uropathogenic populations. J. Infect. Dis. 184, 1041-1049.

109. **Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., Shinagawa, H.,** 2001: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 8, 11-22.

110. **Purcell, M., Shuman, H.A.,** 1998: The *Legionella pneumophila* icmGCDJBF genes are required for killing of human macrophages. Infect. Immun. 66, 2245-2255.

111. **Williams, S.G., Varcoe, L.T., Attridge, S.R., Manning, P.A.,** 1996: *Vibrio cholerae* Hcp, a secreted protein coregulated with HlyA. Infect. Immun. 64, 283-289.

112. **Das, S., Chakrabortty, A., Banerjee, R. and Chaudhuri, K.,** 2002: Involvement of *in vivo* induced *icmF* gene of *Vibrio cholerae* in motility, adherence to intestinal epithelial cells and conjugation frequency. Biochem. Biophys. Res. Com. 295: 922-928.

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

114. **Das, S., Chaudhuri, K.,** 2003: Identification of a unique IAHP (IcmF associated homologous proteins) cluster in *Vibrio cholerae* and other proteobacteria through in silico analysis. In Silico Biol. 3, 287-300.

115. **Carniel, E.,** 2001: The *Yersinia* high-pathogenicity island: an iron-uptake island. Microbes Infect. 3, 561-569.

116. **Buchrieser, C., Rusniok, C., Frangeul, L., Couve, E., Billault, A., Kunst, F., Carniel, E., Glaser, P.,** 1999: The 102-kilobase pgm locus of Y*ersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. Infect. Immun. 67, 4851-4861.

117. **Miller, D.A., Luo, L., Hillson, N., Keating, T.A., Walsh, C.T.,** 2002: Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. Chem. Biol. 9, 333-344.

118. **Hutchinson, C.R.,** 2003: Polyketide and non-ribosomal peptide synthases: falling together by coming apart. Proc. Natl. Acad. Sci. USA. 100, 3010-3012.

119. **Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K., Kinashi, H.,** 2003: The large linear plasmid pSLA2-L of *Streptomyces rochei*

has an unusually condensed gene organization for secondary metabolism. Mol. Microbiol. 48, 1501-1510.

120. **Varga, J., Rigo, K., Kocsube, S., Farkas, B., Pal, K.,** 2003: Diversity of polyketide synthase gene sequences in *Aspergillus* species. Res. Microbiol. 154, 593-600.

121. **Yoshida, T., Furuya, N., Ishikura, M., Isobe, T., Haino-Fukushima, K., Ogawa, T., Komano, T.,** 1998: Purification and characterization of thin pili of IncI1 plasmids ColIb-P9 and R64: formation of PilV-specific cell aggregates by type IV pili. J. Bacteriol. 180, 2842-2848.

122. **Sajjan, U.S., Xie, H., Lefebre, M.D., Valvano, M.A., Forstner, J.F.,** 2003: Identification and molecular analysis of cable pilus biosynthesis genes in *Burkholderia cepacia*. Microbiology 149, 961-971.

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

124. **Holt, J.G., Krieg, N.R., Sneath, P.H.A., Staley J.T., and Williams, S.T., ed.** 1994: Bergey's manual of determinative bacteriology, 9th ed. Williams and Wilkins, Philadelphia, Pa.

125. **Sprenger, G.A., and Lengeler, J.W.,** 1987: Mapping of the sor genes for L-sorbose degradation in the chromosome of *Klebsiella pneumoniae*. Mol. Gen. Genet. 209:352-359.

126. **Woodward, M.J., and Charles, H.P.,** 1983: Polymorphism in *Escherichia coli*: *rtl atl* and *gat* regions behave as chromosomal alternatives. J. Gen. Microbiol. 129:75-84.

127. **Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba T., Hattori, M., and Shinagawa, H.,** 2001: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 8:11-22.

128. **Jahreis, K., Bentler, L., Bockmann, J., Hans, S., Meyer, A., Siepelmeyer, J., Lengeler, J.W.,** 2002: Adaptation of sucrose metabolism in the *Escherichia coli* wild-type strain EC3132. J. Bacteriol. 184, 5307-5316.

129. Bockmann J., Heuel, H. and Lengeler, J.W., 1991: A sugar-specific porin, ScrY, is involved in sucrose uptake in enteric bacteria. Mol. Microbiol. 5:941-950.

130. **Hardesty, C., Ferran, C., Di Rienzo, J.M.,** 1991: Plasmid-mediated sucrose metabolism in *Escherichia coli*: characterization of *scrY*, the structural gene for a phosphoenolpyruvate-dependent sucrose phosphotransferase system outer membrane porin. J. Bacteriol. 173, 449-456.

131. **Titgemeyer, F., Jahreis, K., Ebner, R., Lengeler, J.W.,** 1996: Molecular analysis of the *scrA* and *scrB* genes from *Klebsiella pneumoniae* and plasmid pUR400, which encode the sucrose transport protein Enzyme II Scr of the phosphotransferase system and a sucrose-6-phosphate invertase. Mol. Gen. Genet. 250, 197-206.

132. **Wehmeier, U.F. and Lengeler, J.W.,** 1994: Sequence of the sor-operon for L-sorbose utilization from *Klebsiella pneumoniae* KAY2026. Biochim. Biophys. Acta. 1208, 348-351.

133. **Brinkkotter, A., Kloss, H., Alpert, C., Lengeler, J.W.,** 2000: Pathways for the utilization of N-acetyl-galactosamine and galactosamine in *Escherichia coli.* Mol. Microbiol. 37, 125-135.

134. **Blom, N.S., Tetreault, S., Coulombe, R., Sygusch, J.,** 1996: Novel active site in *Escherichia coli* fructose 1,6-bisphosphate aldolase. Nat Struct Biol 3:856-862.

135. **Park, Y.K., Bearson, B., Bang, S.H., Bang, I.S., Foster, J.W.,** 1996: Internal pH crisis, lysine decarboxylase and the acid tolerance response of *Salmonella typhimurium*. Mol. Microbiol. 20, 605-611.

136. **Dong, Y., Chen, Y.Y., Snyder, J.A., Burne, R.A.,** 2002: Isolation and molecular analysis of the gene cluster for the arginine deiminase system from *Streptococcus gordonii* DL1. Appl. Environ. Microbiol. 68, 5549-5553.

137. **Janausch, I.G., Zientz, E., Tran, Q.H., Kroger, A., Unden, G.,** 2002: C4-dicarboxylate carriers and sensors in bacteria. Biochim Biophys Acta 1553:39-56.

138. **Payne, J., and Smith, M.W.,** 1994: Peptide transport by microorganisms. Adv Microb Physiol 36.

139. **Lee, B., Bhuta, T., Craig, J. and Simpson J.,** 2002: Methenamine hippurate for preventing urinary tract infections (Cochrane Review:. Cochrane Database Syst Rev. (1): CD003265.

140. **Klemm, P., Hjerrild, L., Gjermansen, M., and Schembri, M.A.,** 2004: Structure-function analysis of the self-recognizing antigen 43 autotransporter protein from *Escherichia coli*. Mol. Microbiol. 51, 283−296.

141. **Passerini de Rossi, B.N., Friedman, L.E., Gonzalez, Flecha, F.L., Castello, P.R., Franco, M.A., Rossi, J.P.,** 1999: Identification of *Bordetella pertussis* virulence-associated outer membrane proteins. FEMS Microbiol Lett. 1;172(1):9-13.

142. **Leininger, E., Ewanowich, C.A., Bhargava, A., Peppler, M.S., Kenimer, J.G., Brennan, M.J.,** 1992: Comparative roles of the Arg-Gly-Asp sequence present in the *Bordetella pertussis* adhesins pertactin and filamentous hemagglutinin. Infect. Immun. 60(6):2380-2385.

143. **Benjelloun-Touimi, Z., Sansonetti, P.J. and Parsot, C.**, 1995: SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion Mol. Microbiol. 17 (1), 123-135.

144. **Henderson, I.R., J. Czeczulin, Eslava, C., Noriega, F., and Nataro, J.P.,** 1999: Characterization of *pic*, a secreted protease of shigella flexneri and enteroaggregative *Escherichia coli*. Infect. Immun. 67(11): 5587–5596

145. **Guyer, D.M., Henderson, I.R., Nataro, J.P. and Mobley, H.L.T.,** 2000: Identification of Sat, an autotransporter toxin produced by uropathogenic *Escherichia coli*. Molecular Microbiology 38 (1), 53-66.

146. **Romanos, M.A., Clare, J.J., Beesley, K.M., Rayment, F.B., Ballantine, S.P., Makoff, A.J., Dougan, G., Fairweather, N.F., Charles, I.G.,** 1991: Recombinant *Bordetella pertussis* pertactin (P69) from the yeast *Pichia pastoris*: high-level production and immunological properties. Vaccine 9, 901-906

147. **Bhakdi, S., Greulich, S., Muhly, M., Eberspacher, B., Becker, H., Thiele, A., and Hugo, F.,** 1989: Potent leukocidal action of *Escherichia coli* hemolysin mediated by permeabilization of target cell membranes. J. Exp. Med. 169, 737-754,

148. **Ludwig, A. and Goebel, W.,** In The Comprehensive Sourcebook of Bacterial Protein Toxins 2nd ed. (eds. Alouf, J.E. and Freer, J.H.) 330-348 (Academic Press, London, 1999.

149. **Wallace, A.J., Stillman, T.J., Atkins, A., Jamieson, S.J., Bullough, P.A., Green, J., Artymiuk, P.J.,** 2000: *E. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. Cell. 21;100(2):265-76.

150. **Weinig, S., Hecht, H.J., Mahmud, T., Muller, R.,** 2003: Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases. Chem. Biol. 10, 939-952.

151. **Silakowski, B., Kunze, B., Muller, R.,** 2001: Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca*. Gene 275, 233-240.

152. **Yao, S., Gao, X., Fuchsbauer, N., Hillen, W., Vater, J., Wang, J.,** 2003: Cloning, sequencing, and characterization of the genetic region relevant to biosynthesis of the lipopeptides iturin A and surfactin in *Bacillus subtilis*. Curr. Microbiol. 47, 272-277.

153. **Hoffmann, D., Hevel, J.M., Moore, R.E., Moore, B.S.,** 2003: Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224. Gene 311, 171-180.

154. **Chang, Z., Flatt, P., Gerwick, W.H., Nguyen, V.A., Willis, C.L., Sherman, D.H.,** 2002: The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. Gene 296, 235-247.

155. **Labro, M., Int. J.,** 1998: Antibacterial agents - phagocytes: new concepts for old in immunomodulation. Antimicrob. Agents 10, 11-21.

156. **Hopwood, D.A., and Sherman, D.H.,** 1990: Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu. Rev. Genet. 24, 37.

157. **Katz, L. and Donadio, S.,** 1993: Polyketide synthesis: prospects for hybrid antibiotics. Annu. Rev. Microbiol. 47, 875-912.

Acknowledgments

This dissertation was written while I was working at University of Göttingen.
First and foremost, I would like to express my gratitude to my supervisor ***Prof. Dr. G. Gottschalk*** *for giving me an opportunity to work in Göttingen Genomics Laboratory, as well as for his intensive scientific guidance over this period. His expertise, understanding, and patience, added considerably to my graduate experience.*

And I have a few more people to thank.

I am most obviously indebted to ***Dr. Holger Brüggemann****. My gratitude to him is as boundless as the Pacific Ocean. I was very often dependent on his wide knowledge, which he never hesitated to share with me. I appreciate his vast knowledge and skills. I shall remember his excellent help during my first and last steps in "genome's world" and also for the sheer joy of his company. We had the most stimulating and fruitful discussions on many topics not only scientific. Thank you for being a guide and friend both inside and outside the lab.*

My special thanks go to ***Dr. Heiko Liesegang*** *for his encouragement, advice, research support and discussions throughout my doctoral studies. I also truly appreciate his patience and tolerance during my numerous mishaps…*

I am grateful to ***Jarek Sobkowiak*** *for the time he spent with my computer.*

My very special thanks and gratitude go to ***Iwona Decker****,* ***Monika Grzywa****,* ***Silke Dencker****,* ***Silke Steckel****,* ***Frauke Meyer*** *and* ***Mechted Bömeke****. They are not only good in what they do, but they are also nice, and always helpful. Thanks a lot for your cooperation and creating a friendly environment in the lab, which was very important for surviving all these years.*

I am also grateful to ***Dr. Axel Strittmatter****,* ***Dr. Frank Hoster****,* ***Florian Fricke****,* ***Dr. Arnim Wiezer****, Dr.* ***Anke Henne, Kirsten Hahne****,* ***Yvonne Glaeser*** *and* ***Brita Kotrasch*** *for their valuable advices and suggestions during my Ph.D. study.*

Appreciation is also extended to ***Rosalie Bauer***, ***Katrin Groebel***, ***Sonja Vollan***, *and* ***Sönke Andres****. I always enjoy theirs company.*

I take this opportunity to express my sincere thanks to the great members of Institut für Molekulare Infektionsbiologie der Universität Würzburg I was fortunate to have the opportunity to work with a group of energetic people in this lab. I had enjoyed every moment that we had worked together and I sincerely thank them. My special thanks goes to: Dr. Ulrich Dobrindt, Dr. Gabor Nagy, Dr. Gabriele Blum-Oehler, Dr. Melanie Emmerth, Dr. Caroline Wilde, Barbara Plaschke, Sebastian Reidl, Stefan Homburg, Kai Michaelis, and Philippe Bauchart.

Monika, Iwona, Jarek – wielkie dzieki za to, ze bylo do kogo gebe otworzyc.

Finally,
I would like to thank those who mean a lot to me personally.

My wonderful parents and sister whose sacrifice and unconditional love through all these years have helped me achieve one of my dreams. I thank them for giving me free choice in selecting my future profession, for their constant support and tolerance.

At last, but not at least, thanks to all my friends for their indispensable support and encouragement during all this period. They have given me comfort, help and advice whenever I needed them. Without them, there is no way I could possibly have accomplished this work.

My friends Ola and Ania who were always there by my side and always had time to listen to me and my problems. I thank you from the bottom of my heart for being such wonderful friends and giving a shape to my life.

Ela

CURRICULUM VITAE

Elżbieta Barbara Brzuszkiewicz

Date of Birth:	02/06/1977
Place of Birth:	Warsaw, Poland
Sex:	Female
Marital status:	Single

1984-1992:	Primary school, Warsaw, Poland
1992-1996:	Higher secondary school: XXI LO H. Kollataja High School, Warsaw, Poland; biology and chemistry oriented studies
1996-2002:	Faculty of Biology in the Warsaw University, Poland (Department of Bacterial Genetics)
Nov. 2002-present:	Ph.D. studies under the supervision of Prof. Dr. G. Gottschalk on the research project:" Genome sequence of *Escherichia coli* 536: insights into uropathogenicity through comparison with genomes of *Escherichia coli* MG1655, CFT073, and EDL933" at the Institute of Microbiology and Genetics, Georg-August University