Untersuchungen zur Genomorganisation und zur Fitness des apathogener *Escherichia coli* Stammes Nissle 1917 (O6:K5:H1)

Analysis of the genome organization and fitness traits of non-pathogenic *Escherichia coli* strain Nissle 1917 (O6:K5:H1)

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Bayerischen Julius-Maximilians-Universität Würzburg

vorgelegt von

Lubomir Assenov Grozdanov

aus Sofia (Bulgarien)

Würzburg 2003

Eingereicht am:

Mitglieder der Promotionskomission:

Vorsitzende:

1. Gutachter: Prof. Dr. Dr. h. c. J. Hacker
2. Gutachter: PD Dr. M. Kuhn

Tag des Promotionskolloquiums:
Doktorurkunde ausgehändigt am:

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.
Diese Dissertation wurde weder in gleicher noch in ähnlicher Form in einem anderen Prüfungsverfahren vorgelegt.
Des weiteren erkläre ich, dass ich früher weder akademische Grade erworben noch zu erwerben versucht habe.

Würzburg, den 11. Dezember 2003

Lubomir Assenov Grozdanov

Die Experimente zur vorliegenden Arbeit wurden in der Zeit von September 1999 bis August 2003 am Institut für Molekulare Infektionsbiologie der Universität Würzburg durchgeführt.

An dieser Stelle möchte ich mich herzlich bei Herrn Prof. Dr. Dr. h. c. J. Hacker für die Anregung zur vorliegenden Arbeit, für seine Unterstützung, sein Interesse an dieser Arbeit sowie die Möglichkeit zum selbständigen Experimentieren bedanken.
Die Ermittlung von DNA-Sequenzen an Cosmiden wurde im Laboratorium für Genomanalysen der Universität Göttingen durchgeführt. Dafür bedanke ich mich besonders bei Herr Prof. Dr. G. Gottschalk und seinen Mitarbeitern.
Bei Herr. Prof. Dr. E. Th. Rietschel vom Zentrum für Medizin und Biowissenschaften in Borstel und Herr Prof. Dr. P. Nenkov vom Nationalen Zentrum für Parasitologie und Infektionskrankheiten in Sofia bedanke ich mich für Ihre wissenschaftliche Unterstützung.
Der Firma “Ardeypharm“ GmbH und sowie der “Bayerischen Forschungsstiftung“ bin ich für die finanzielle Unterstützung sehr dankbar.

Mein Dank gilt auch besonders Uli Dobrindt und Gabi Blum-Oehler – ohne ihre Geduld, Hilfe und Verständnis hätte ich es nicht geschafft.

Für die gute Atmosphäre in Labor A109, die ständige Hilfsbereitschaft und Diskussionsbereitschaft, sowie für die humorvolle Zeit, möchte ich bei Barbara Plaschke, Gyoergy Schneider, Gabor Nagy, Birgit Schellberg, Simone Werner, Anja Sauer und Anja Waldschmidt bedanken.
Ich bedanke mich auch bei allen “Colis“, “Staphis“, “Schwämme“, “Vibrios“ und “Candidas“, die mich fachlich, moralisch und menschlich in meiner Doktorandenzeiten unterstützt und somit zum Gelingen dieser Arbeit beigetragen haben.

Table of contents

I. Zusammenfassung

Bislang wurden die kompletten Genomsequenzen von mehr als 100 Bakterien ermittelt. Mehr als ein Drittel dieser Organismen wird als pathogen eingestuft. Die Verfügbarkeit dieser Sequenzinformation vergrößert unser Wissen über die bakterielle Genomstruktur und –plastizität sowie über mikrobielle Diversität und Evolution. Diese Daten bilden die Grundlage für viele Fortschritte auf dem Gebiet der Biotechnologie, der industriellen-, Umwelt- und medizinischen Mikrobiologie: neuartige Typisierung-, Diagnostik- und Therapieansätze sowie die Entwicklung neuer Medikamente und Impfstoffe basieren auf und profitieren von der ständig zunehmenden Menge an DNA-Sequenzen. Genomanalysen sind zusammen mit anderen molekularbiologischen Methoden wie PCR, DNA-Chiptechnologie, subtraktive Hybridisierung und Proteomanalysen von zunehmender Bedeutung für die Erforschung von Infektionskrankheiten und das öffentliche Gesundheitswesen.

Ziel dieser Arbeit war die Analyse der Genomstruktur und des Genominhaltes des apathogenen *Escherichia coli* Stammes Nissle 1917 (O6:K5:H1) und der Vergleich mit verfügbaren Daten verschiedener pathogener und apathogener *E. coli* Stämme sowie anderer eng-verwandter Spezies. Eine Cosmid-Genbank des Stammes Nissle 1917 wurde nach Klonen durchsucht, die für Fitnessfaktoren kodieren, welche zur erfolgreichen Kolonisierung des menschlichen Verdauungstraktes und zum probiotischen Charakter dieses Stammes beitragen. Vier genomische Inseln (GEI I-IV$_{\text{Nissle 1917}}$) wurden nachgewiesen und charakterisiert. Auf diesen GEIs befinden sich verschiedene bekannte Fitness-Determinanten (*mch/mcm, foc, iuc, kps, ybt*), bislang nicht charakterisierte ORFs, mobile genetische Elemente und bislang für den Stamm Nissle 1917 nicht beschriebene Gene, die möglicherweise zur Fitness und Adaptabilität dieses Stammes beitragen. Die GEIs I-IV$_{\text{Nissle 1917}}$ sind jeweils mit einem tRNA-Gen assoziiert und ähneln hinsichtlich ihrer Struktur und chromosomalen Lokalisation entsprechenden Inseln im Genom des uropathogenen *E. coli* Stammes CFT073 (O6:K2(?):H1). Interessanterweise fehlen auf diesen wichtige Virulenzgene uropathogener *E. coli* (*hly, cnf, prf/pap*). Eine etwa 30 kb große Region der GEI II$_{\text{Nissle 1917}}$, die von IS*2* Elementen flankiert wird, kann spontan deletieren, was zum Verlust verschiedener Fitnessdeterminanten (*iuc, sat, iha*) führt. Darüber hinaus wurde der chromosomale Sequenzkontext von tRNA-Genen mittels PCR auf die Integration von „Fremd-DNA“ hin untersucht, die durch horizontalen Gentransfer erworben wurde (tRNA Screening), und mit denen anderer apathogener und pathogener *E. coli* Stämme verglichen. Der genomweite Anteil an tRNA-Gen-assoziierter, möglicherweise horizontal erworbener DNA, die im

apathogenen *E. coli* K-12 Stamm MG1655 fehlt, unterschied sich dabei nicht bedeutend im Stamm Nissle 1917 und den uropathogenen *E. coli* O6 Stämmen CFT073 und 536. Die Verbreitung von DNA-Regionen der GEIs des Stammes Nissle 1917 wurde mittels PCR bei apathogenen *E. coli*-Isolaten sowie bei uropathogenen *E. coli* O6:K5-Isolaten untersucht. Nur zwei UPEC O6:K5-Isolate enthielten alle GEI-Bereiche, die in diese Untersuchung einbezogen waren, unterschieden sich jedoch vom Stamm Nissle 1917 durch ihre Phänotypen. Die Makrorestriktionsanalyse der Genomstruktur des *E. coli* Stammes Nissle 1917 zeigte, daß letztere der des uropathogenen *E. coli* Stammes CFT073 sehr ähnelt.

Um die Ursache für den semi-rauhen Phänotyp des Stammes Nissle 1917 zu untersuchen, wurden die *wa** und *wb** Determinanten dieses Stammes, die für die LPS-Biosynthese verantwortlich sind, kloniert und sequenziert. Das bislang unbekannte Serotyp O6-spezifische O-Antigenpolymerase-kodierende Gen *wzy* des Stammes Nissle 1917 wurde charakterisiert und mit dem des rauhen O6 Stammes 536 verglichen. Eine Punktmutation, die zu einem vorzeitigen Translationsstop der *wzy*-Transkripte des Stammes Nissle 1917 führt, wurde als Ursache für den semi-rauhen Phänotyp und damit auch die Serumsensitivität dieses Stammes verantwortlich gemacht.

Zur Untersuchung der Kolonisierungsfähigkeit des *E. coli* Stammes Nissle 1917 wurden verschiedene Faktoren, die an der Biofilmbildung bzw. am multizellulären Verhalten beteiligt sind, sequenziert und näher analysiert. Die Seqenzierung der *fim* Determinante zeigte, daß das *fimB* Gen, das für die Expression der Typ 1-Fimbrien benötigt wird, durch die Insertion eines IS-Elementes inaktiviert wurde. Untersuchungen zum multizellulären Verhalten zeigten, daß der Stamm Nissle 1917 den sogenannten „rdar“ Morphotyp, hervorgerufen durch Expression von Curli-Fimbrien und Cellulose, bei 30 °C und bei 37 °C exprimiert, nicht jedoch die uropathogenen *E. coli* Stämme 536 und CFT073. Das Cellulosebiosynthese-Operon (*bcs*) sowie das Gen *rfaH*, das für einen Transkriptionsantiterminator kodiert, wurden im Stamm Nissle 1917 inaktiviert, um deren Bedeutung für den „rdar“ Morphotyp zu untersuchen. Während Cellulose für die Expression des „rdar“ Morphotyps benötigt wird, hatte die *rfaH*-Inaktivierung keinen Einfluß auf dieses mulizelluläre Verhalten des *E. coli* Stammes Nissle 1917.

Die Ergebnisse dieser Arbeit zeigen, daß der apathogene *E. coli* Stamm Nissle 1917 durch eine spezifische Kombination phänotypischen Eigenschaften gekennzeichnet ist, die ihn von anderen bislang untersuchten *E. coli* Stämmen unterscheidet. An der Evolution dieses Stammes, möglicherweise aus einem pathogenen „Vorfahren“, waren vielfältige Gentransfer- und Deletionsprozesse sowie Punktmutationen beteiligt.

I. Summary

In the last years more than one hundred microbial genomes have been sequenced, many of them from pathogenic bacteria. The availability of this huge amount of sequence data enormously increases our knowledge on the genome structure and plasticity, as well as on the microbial diversity and evolution. In parallel, these data are the basis for the scientific "revolution" in the field of industrial and environmental biotechnology and medical microbiology – diagnostics and therapy, development of new drugs and vaccines against infectious agents. Together with the genomic approach, other molecular biological methods such as PCR, DNA-chip technology, subtractive hybridization, transcriptomics and proteomics are of increasing importance for research on infectious diseases and public health.

The aim of this work was to characterize the genome structure and -content of the probiotic *Escherichia coli* strain Nissle 1917 (O6:K5:H31) and to compare these data with publicly available data on the genomes of different pathogenic and non-pathogenic *E. coli* strains and other closely related species. A cosmid genomic library of strain Nissle 1917 was screened for clones containing the genetic determinants contributing to the successful survival in and colonization of the human body, as well as to mediate this strain's probiotic effect as part of the intestinal microflora. Four genomic islands (GEI I-IV$_{\text{Nissle 1917}}$) were identifed and characterized. They contain many known fitness determinants (*mch/mcm, foc, iuc, kps, ybt*), as well as novel genes of unknown function, mobile genetic elements or newly identified putative fitness-contributing factors (Sat, Iha, ShiA-homologue, Ag43-homologues). All islands were found to be integrated next to tRNA genes (*serX, pheV*, *argW* and *asnT,* respectively). Their structure and chromosomal localization closely resembles those of analogous islands in the genome of uropathogenic *E. coli* strain CFT073 (O6:K2(?):H1), but they lack important virulence genes of uropathogenic *E. coli* (*hly, cnf, prf/pap*). Evidence for instability of GEI II$_{\text{Nissle 1917}}$ was given, since a deletion event in which IS*2* elements play a role was detected. This event results in loss of a 30 kb DNA region, containing important fitness determinants (*iuc, sat, iha*), and therefore probably might influence the colonization capacity of Nissle 1917 strain.

In addition, a screening of the sequence context of tRNA-encoding genes in the genome of Nissle 1917 was performed to identify genome wide potential integration sites of "foreign" DNA. As a result, similar "tRNA screening patterns" have been observed for strain Nissle 1917 and for the uropathogenic *E. coli* O6 strains (UPEC) 536 and CFT073.

The molecular reason for the semi-rough phenotype and serum sensitivity of strain Nissle 1917 was analyzed. The O6-antigen polymerase-encoding gene *wzy* was identified, and it was shown that the reason for the semi-rough phenotype is a frame shift mutation in *wzy*, due to the presence of a premature stop codon. It was shown that the restoration of the O side-chain LPS polymerization by complementation with a functional *wzy* gene increased serum-resistance of strain Nissle 1917.

The results of this study show that despite the genome similarity of the *E. coli* strain Nissle 1917 with the UPEC strain CFT073, the strain Nissle 1917 exhibits a specific set of geno- and phenotypic features which contribute to its probiotic action.

By comparison with the available data on the genomics of different species of *Enterobacteriaceae*, this study contributes to our understanding of the important processes such as horizontal gene transfer, deletions and rearrangements which contribute to genome diversity and -plasticity, and which are driving forces for the evolution of bacterial variants.

At last, the *fim, bcs* and *rfaH* determinats whose expression contributes to the mutlicellular behaviour and biofilm formation of *E. coli* strain Nissle 1917 have been characterized.

II. Introduction

1. Microbial genome research

Since 1995, when the genome of the *Haemophilus influenzae* was completely sequenced (Fleischmann, et al., 1995), hundreds of microbial genomes from prokaryotic, as well as eukaryotic microorganisms, have been sequenced. These data are of crucial importance for diagnostics and therapy of infectious diseases, for the development of vaccines and the identification of novel target genes. Moreover, the increasing knowledge on bacterial genome structure and "comparative genomics" improves our understanding on the driving forces of microbial evolution like horizontal gene transfer, deletion formation, recombination, which result in development of new bacterial species or variants.

In some cases, the genome of several strains of one bacterial species have been sequenced. The complete genome sequence of the *E. coli* K-12 laboratory strain MG1655 gives an idea on the genome structure of a non-pathogenic, commensal microorganism, which has lost many phenotypic features due to the decades of laboratory cultivation. Later on, several genomes of pathogenic *E. coli* strains have been sequenced: two from enterohaemorrhagic *E. coli* (EHEC) strains (EDL933 and "Sakai"; Perna et al., 2001; Hayashi et al., 2001) and one uropathogenic *E. coli* (UPEC) isolate (CFT073; Welch et al., 2002). The comparison of these three genomes revealed an unexpectedly extensive mosaic genome structure within the *E. coli* species, where only 39.2 % of the predicted proteins are shared by all of the investigated *E. coli* types (Fig. 1; Welch et al., 2002), which suggests that horizontal gene transfer and acquisition of foreign DNA elements played and important role during the evolution and adaptation of these variants.

Similarly, the genomes of two *Helicobacter pylori* strains from different sources have been sequenced, namely of the isolate 26695 (Tomb et al., 1997) from duodenum ulcer and of strain J99 (Alm et al., 1999) isolated from a gastritis patient. Interestingly, 5.9 % of the identified ORFs in J99 and 7.3 % in 26695 were strain specific.

In addition to the whole-genome sequencing approach, increasing data from the sequence analysis of partial genomes from different bacterial species is accumulating. The use of e. g. genomic libraries, specific PCR amplification products and of

subtractive hybridization techniques made it possible to characterize several "pathogenicity islands" of the UPEC strain 536, which contain various virulence-associated factors and which are located in different positions of this strain's chromosome (Dobrindt et al., 2002). In other studies, the "long-distance" PCR approach has been used to characterize extremely divergent gene clusters, encoding the enzymes involved in the LPS side-chain biosynthesis, which are always localized at a conserved chromosomal region (Paton et al., 1999; Marolda et al., 1998; Wang et al., 1998; Wang et al., 1998).

In order to create an integrated picture of the genome organization, the sequence data can be combined with a physical map of the bacterial chromosome, using rare-cutting restriction endonucleases and pulsed-field gel electrophoresis, thus avoiding the expensive and time consuming whole genome sequencing approach. In such a way and with the help of mini transposon insertions, an integrated genomic map of the UPEC strain J96 has been created (Lyla et al., 2000).

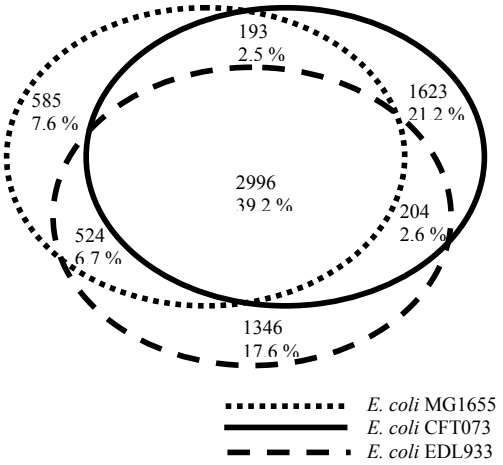

Fig. 1: Comparison of the number of predicted proteins shared by three *E. coli* strains (MG1655, CFT073, EDL933). The number of orthologs in each shared category and the numbers of strain-specific proteins are given. Total number of proteins counted: MG1655, 4288; CFT073, 5016; EDL933, 5063 (modified from Welch et al., 2002). These data are based on the complete genome sequence of the corresponding *E. coli* strains.

One result of the availability of genome sequence data, the DNA-chip technology was developed, which allows genome comparison of multiple isolates. Thousands of specific PCR products or synthetic oligonucleotides are immobilized in an ordered manner on special matrices and hybridized with labeled genomic DNA from the investigated strains, resulting in a whole genome overview of the presence of genes of a particular reference genome. This approach gives the opportunity to study the distribution of specific genetic features associated with metabolic functions or infectious diseases.

Furthermore, the DNA array technology enables genome-wide studies on transcription ("transcriptomics"). After hybridization with isolated RNA or cDNA, the global transcription profile of a distinct strain can be studied, in response to

different environmental conditions. The effect of mutations on the "transcriptome" can also be investigated (Marshall and Hodgson, 1998; Harrington et al., 2000). In this way, the regulation of metabolic pathways, as well as the cellular stress response, the adaptation to different environmental niches, or bacteria-host cell interaction can be examined on the transcriptional level. A commercially available DNA array containing probes specific for all translated ORFs identified in the genome of *E. coli* strain MG1655 was used to examine the differential gene expression of *E. coli* in minimal medium versus LB (Tao et al., 1999). DNA macroarray, containing probes specific for most of the genes encoding for virulence factors of different extraintestinal and intestinal pathotypes of *E. coli* and *Shigella*, enabled large scale epidemiological studies, as well as global examination of virulence gene regulation in UPEC strain 536 (Dobrindt et al., 2003; Michaelis, K., Dobrindt, U., unpublished data).

On the other hand, the proteome approach enables the investigation of the protein modification, as well as and post-transcriptional and post-translation regulation, using two-dimensional gel electrophoresis. In this way, proteins whose expression is affected or which undergo specific modification under certain circumstances can be detected and isolated. Isolated proteins can then be identified by mass spectrometry or by N-terminal sequencing (Van Bogelen et al., 1999).

2. Horizontal gene transfer and genome plasticity

Horizontal gene transfer is an important mechanism contributing to the variability of bacterial genomes, therefore playing a crucial role for the evolution of microorganisms. This process entails the incorporation of genetic information transfered from another organism directly into the genome where they may form "genomic islands", e.g., large blocks of DNA with signatures of mobile genetic elements (Hacker and Carniel, 2001; Fig. 3). Based on these findings, a new concept of "dynamic" bacterial genomes has been developed in the last years. According to this view, the bacterial genome consists of two gene pools, the "flexible gene pool" which includes the majority of the horizontally transferred DNA, and the conserved "core gene pool" which mainly represents the essential genetic information (Fig. 2).

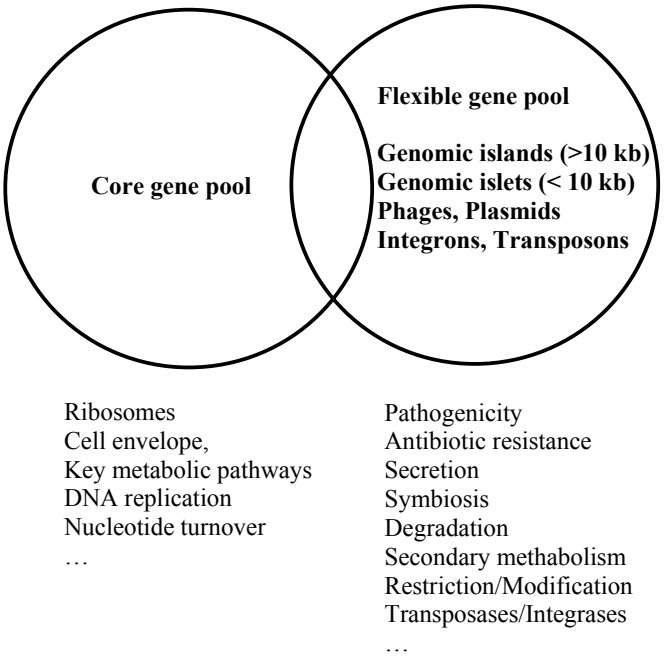

Fig. 2: Model of the DNA pools in the genomes of the prokaryotes (modified from Hacker and Carniel, 2001).

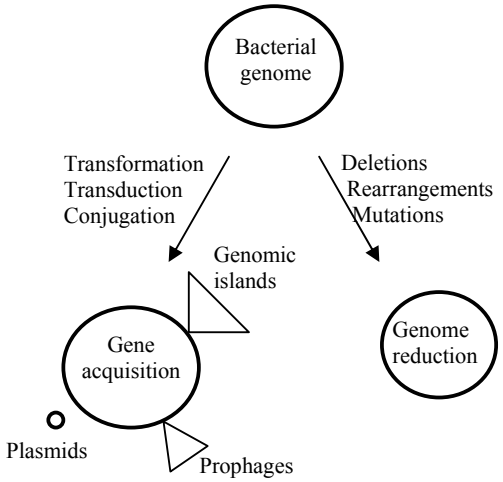

Fig. 3: Genome evolution by means of acquisition and loss of genetic information (modified from Dobrindt and Hacker, 2001).

Whereas the "core gene pool" encodes proteins that play a role in basic cellular processes, the "flexible gene pool" mainly comprises genes whose function reflects the bacterial lifestyle (Dobrindt and Hacker, 2001), and provide advantages for the

bacteria under particular conditions. Therefore the “flexible gene pool” is important for adaptation and the survival in a new ecological environment, e. g. for symbiosis, host-cell interaction and pathogenicity (Hentschel et al., 2000). The number of horizontally acquired genes may vary, from 18 % (*E. coli* K-12) to <1 % (*Mycoplasma*) of the total genome (Lawrence and Ochman, 1998; Ochman et al., 2000). Typically, the laterally transfered DNA differs in the overall G+C content and the codon usage from that of the core genome. Since the “flexible gene pool” consists of DNA fragments of different origin, the G+C content varies dramatically, thus representing the “mosaic” structure of the newly acquired DNA elements. For example, within the *Salmonella* Pathogenicity Island 3 (SPI-3) there are a regions of very low G+C content, despite the fact that the overall codon usage does not differ significantly from those of highly expressed *Salmonella* genes (Blanc-Potard et al., 1999). The *lac* operon in *E. coli* differs in the codon usage compared to the genes from the core genome (Ochmann et al., 1996). This demonstrates that genes, which are part of the species-specific “core genome” and which encode certain metabolic traits, might also be acquired via horizontal gene transfer. In many cases the foreign DNA elements are preferentially integrated at specific positions in the “core chromosome” (“hot-spots”). Such integration sites for lysogenic phages and other mobile genetic elements include highly conserved tRNA-encoding genes (Hou, 1999; Cheetham and Katz, 1995). For example, the high-pathogenicity island (HPI) is typically associated with the *asnT* tRNA-encoding gene in enterobacteria (Carniel et al., 1996), but may also insert at other *asn* tRNA-encoding genes of *Yersinia pseudotuberculosis*. In other cases, one specific tRNA gene represents the chromosomal insertion site of different types of PAIs. Thus the *selC* tRNA gene serves as an insertion site for several different pathogenicity islands in different enterobacterial pathotypes (Blum et al., 1994; Fleckenstein et al., 1998; Moss et al., 1999, McDaniel et al., 1995). However, there are also examples for other chromosomal insertion sites than tRNA-genes such as *galF* of *E. coli* (Tarr et al., 2000).

Many important phenotypic characteristics of bacteria, like antibiotic resistance, virulence traits and particular metabolic pathways are encoded on genes located on horizontally transferable, mobile genetic elements like plasmids, transposons, and phages. For example, the tetracycline resistance determinant and its regulatory genes are part of the Tn*10* transposon, which is able to integrate into the chromosomes of

different bacteria (Lawley et al., 2000). The genes encoding the enzymes responsible for the lipopolysaccharide (LPS) O-side chain biosynthesis of *Shigella sonnei* are located on a plasmid. This gene cluster is very similar to those of the majority of other *E. coli* O-serotypes, which are located in the "*galF* island" on the chromosome. The LEE pathogenicity island of rabbit-pathogenic enterohaemorrhagic *E. coli* strains (REPEC), is integrated at the *pheU* tRNA gene. This PAI was shown to be spontaneously excised from the chromosome. The REPEC LEE-encoded integrase is able to mediate the site-specific integration at the *pheU* tRNA in the *E. coli* strain DH1 (Tauschek et al., 2002). This demonstrates that the entire LEE island of this REPEC isolate represents a distinct mobile genetic element, which can be deleted from the chromosome, transfered and integrated into the recipient`s chromosome..

Together with the acquisition of foreign DNA, the loss of genetic regions can play a significant role for the evolution of bacteria. Big deletions in the chromosome of some species, termed "black holes", are enhancing the virulence of particular strains. The loss of a big fragment of the chromosome of *Shigella spp.* and enteroinvasive *E. coli,* where a gene encoding for lysine decarboxylase (LDC) is located, contributes to their pathogenicity. It was shown that cadaverine, a product of the reaction catalyzed by the LDC, inhibits the *Shigella* enterotoxin activity (Maurelli et al., 1998). Since many of the studied genomic islands are able to delete from the chromosome with frequency which vary dramatically under different environmental conditions (Middendorf, et al., 2001), it seems to be likely that the loss of DNA fragments represents an important adaptive mechanism for multiplying in new ecological niches or host organism. For instance it is tempting to speculate that such reduction of the "flexible gene pool" contributes to the development of a chronic urinary tract infection (UTI) caused by UPEC strains, since such isolates lack many of the virulence factors typical for UPEC strains causing acute UTIs (Dobrindt, U., Hacker, J., personal communication).

3. Pathogenicity islands and genomic islands

Specific genes that encode virulence factors are present in the genome of pathogenic members of a species but absent in non-pathogenic variants. Initially, virulence associated genes were thought to be solely located on horizontally transferable extrachromosomal elements. In the 1980s blocks of chromosomal regions were identified, which carry virulence genes which are exclusively associated with virulent strains of the species. It has been demonstrated that these regions can delete from the

chromosome spontaneously, indicating their mobile character and the possibility of horizontal gene transfer (Hacker et al., 1983; Low et al., 1984). These chromosomal DNA regions have been designated pathogenicity islands (PAIs). The following characteristics are typical of the PAIs (Hacker and Kaper, 1999):

- PAIs carry genes encoding at least one virulence factor, including adhesins, invasins, iron-uptake systems, toxins, type III and type IV secretion systems,
- PAIs are present in virulent but absent from non-virulent members of the same or closely related species,
- PAIs often consist of DNA whose G+C content is different from that of the core genome, and exhibit a different codon usage relative to the chromosomal backbone,
- PAIs are often flanked by direct repeats (DR),
- PAIs are commonly associated with tRNA genes, which frequently represent chromosomal attachment sites of bacteriophages. PAIs often carry bacteriophage integrases-encoding genes, indicating that these tRNA genes may serve as target sites for bacteriophage-derived elements located on PAIs,
- PAIs carry often genes coding for mobility factors like transposases and insertion elements,
- PAIs are frequently unstable genetic regions and deletions may occur upon recombination between the DR located on both ends or via site-specific recombination.

Although PAIs share common features as described above, they are considered a heterogeneous group of genetic elements, according to their size, structure and encoded functions. Often unrelated to pathogenicity genes, cryptic ORFs, pseudogenes and “junk DNA” are also present on PAIs. PAIs seem to undergo constant evolutionary changes due to acquisition and deletion of DNA regions resulting in very complex structure.

In the last years, the definition of PAIs has been extended to the more general concept of “genomic island” (GEI). GEIs increase the bacterial fitness, either directly or indirectly. Thus, being positively selected they may be called “fitness islands”. They can be divided into several subtypes, according to their encoded traits: “ecological islands” in environmental bacteria, and “saprophytic islands”, “symbiosis islands”, or PAIs in bacteria that interact with living hosts (Hacker and Carniel, 2001).

In some cases PAIs of different strains are grouped into families according to similar gene content and structure. This may result from a similar evolution and transfer of a common PAI ancestor, indicative of a similar function in different isolates (Dobrindt et al., 2001). Nevertheless, the localization and sequence context of similar virulence genes may vary considerably. There are chromosomally localized genetic determinants which are clearly acquired through horizontal gene transfer, but share only some of the features of PAIs. The O-antigen-encoding gene clusters in *E. coli* and *Salmonella* have very low G+C content and it is evident that they have originated by means of interspecies lateral transfer and by homologous recombination of flanking DNA regions, but only few of them are associated with pathogenicity, and they are present in all members of these species.

The exact mechanism of PAI acquisition is still not clear. It is speculated that cointegrate structures of bacteriophages and plasmids have been acquired and integrated at specific sites into the chromosome through gene transfer and recombination. The conserved 3'- ends of tRNA genes are the most common target for chromosomal integration (Hou, 1999). Once integrated, those "pre-PAIs" may have undergone deletion events and point mutations, in mobility genes, which lead to stabilization of the island in the chromosome. Some PAIs are able to delete completely from the chromosome, such as the HPI of *Yersinia pseudotuberculosis*. The deletion occurs through homologous recombination between two flanking 17 bp DR at both ends of the island. The HPI encodes a homologue of the bacteriophage P4 integrase-encoding gene *int*, located in the vicinity of a putative bacteriophage P4 attachment (*attP*) site (Buchrieser et al., 1998). The HPI in *Y. enterocolitica* and *E. coli* is not mobilizable as the 17 bp DR are not present, or as the P4-like integrase gene is inactivated by an internal stop codon (Schubert et al., 1999; Bach et al., 1999). The HPI is widely distributed in the genomes of *Enterobacteriaceae,* such as *E. coli, Klebsiella, Salmonella, Enterobacter* and *Citrobacter* (Schubert et al., 1998) including pathogenic as well as in non-pathogenic isolates. The "core" region of this island contains genes required for the expression of the yersiniabactin iron uptake system, thus contributing to the bacterial fitness during colonization and infection.

PAIs are also present in genomes of Gram-positive bacteria. In *Staphylococcus aureus,* a 15.2 kb island flanked by two 17 bp DR has been identified and designated as *Staphylococcus aureus* Pathogenicity Island 1 (SaPI1). The gene coding for the "toxic shock syndrome"-toxin 1 (*tst*) as well as a homologue for superantigen-

encoding determinant are located on this PAI (Lindsay et al., 1998). This island can be mobilized and transfered to other genomes by the bacteriophage 80α.

A 500 kb "symbiosis island" was identified in the genome of *Mesorhizobium loti.* This island is integrated into the *phe* tRNA gene and carries a P4-like integrase gene, IS elements, as well as the genes coding for the nitrogen fixation and tuber development during the symbiosis (Sullivan and Ronson, 1998). Thus this genetic element is considered as a "symbiosis island", contributing to the survival of bacteria in a new ecological niche.

PAIs have been initially identified in the genome of UPECs (Blum et al., 1994). In the genome of UPEC strain J96, at least two PAIs carrying genes coding for toxins, adhesins, and other virulence determinants the have been identified (Swenson et al., 1996). They are integrated into *pheR* and *pheV* tRNA genes, and flanked by imperfect DR.

4. *Escherichia coli*

The *Bacterium coli commune* was isolated and described by Theodor Escherich from the feces of a normal infant in 1885 and later renamed *Escherichia coli.* This is a Gram-negative bacteria species which constitutes about 0.06 % of the normal human intestinal flora, as well as distributed among animals. *E. coli* colonizes the infant bowel within the first hours of life and is usually a harmless commensal member of the intestinal flora. In addition to non-pathogenic variants, several *E. coli* pathotypes exist which cause intestinal and extraintestinal infections of humans and animals. The spectrum of diseases caused by *E. coli* is also due to the diversity of horizontally acquired virulence genes, harbored on plasmids, bacteriophages or PAIs.

The *E. coli* strains are serotyped on the basis of their O (somatic), H (flagellar), and K (capsular) surface antigen profiles (Kauffmann, 1965), where the specific combination of these factors defines the serotype of an isolate. *E. coli* strains of a specific serotype can be associated with certain clinical manifestations. However, the surface antigens alone are not considered to confer pathogenicity themselves. Rather there are specific clonal lineages which have served as "hosts" for horizontally transfered virulence genes resulting in pathogenic clones (Zingler et al., 1992).

4.1 Pathogroups

Pathogenic *E. coli* strains can be subdivided in several pathotypes according to the disease they cause. *E. coli* strains causing intestinal infections (i.e., diarrheagic *E. coli*) can be subdivided in at least six pathogroups: enterohaemorrhagic (EHEC), enterotoxigenic (ETEC), enteropathogenic (EPEC), enteroinvasive (EIEC), enteroaggregative (EAEC), and diffusely adherent *E. coli* (DAEC). Other strains are responsible for extraintestinal infections: uropathogenic *E. coli* (UPEC), newborn meningitis-causing *E. coli* (MENEC) and septicemia-causing *E. coli* (SEPEC). *E. coli* is responsible for one third of all cases of neonatal meningitis with an incidence of 0.1 per 1000 live births (de Louvois et al., 1994), where the fatality rate varies between 25 to 40 %. The virulence properties of MENEC strains are partially identical to those of SEPEC, as the development of meningitis requires preliminary bacteremia. Important virulence factors are K1 capsular polysaccharide, IbeA, OmpA and SfaII fimbriae (Korhonen et al., 1985).

UPEC strains cause 75-90 % of all community acquired urinary tract infections (UTIs) and about 50 % of the nosocomial UTIs (Rubin, 1990; Svanborg and Godaly, 1997). UTI is the most common bacterial infection in the industrialized world: in the USA 7 million patient visits per year are counted with total costs exceeding one billion dollars (Bacheller et al., 1997). As many as 50 % of the women report to have had at least one UTI in their lifetime (Barnett and Stephens, 1997). UTI affects either the bladder (cystitis) or the kidneys and their collecting systems (pyelonephritis), or both. The bacterial colonization of the urinary tract may be completely free of clinical symptoms (“asymptomatic bacteriuria”, ABU). Moreover, a pyelonephritis can be acute or chronic. The last case is a more complex disorder where the bacterial infection plays a dominant role. However, other factors like vesicourethral reflux and obstruction or immunodeficiency are also critically involved in pathogenesis. UTI is normally an ascending infection (in a less common way UTI can be an ascending infection through the bloodstream) where the bacteria are derived from the patient’s own faecal flora. The initial step of the pathogenesis is colonization of the distal urethra and vagina in women by enterobacteria (Sobel, 1997). From the urethra, the pathogens may gain entrance into the bladder. Here, when the natural defense mechanisms (flushing of urine, IgA, uromucoid) are overwhelmed by the virulent bacteria, bacterial adhesion and colonization may occur evolving into UTI. The

colonization of the urinary tract provokes cellular responses e. g., activation of the epithelial cells, secretion of cytokines and neutrophile migration into the urothelium.

4.2 Virulence and fitness factors of *E. coli*

Some of the traits of the *E. coli* isolates contribute to their virulence and others are required for their survival within the host. In this chapter, the features of UPECs will be mainly discussed.

-Adhesins

Adherence factors enable the bacteria to adhere to the epithelial cells and include fimbrial (fimbriae, pili) and afimbrial adhesins. The fimbriae of UPEC are thin, rod-shaped fiber structures, which are heteropolymeric structures. The P- , type 1, S- , and F1C fimbriae exhibit a composite structure, consisting of a rod-shaped shaft of 6-7 nm in diameter comprising over a thousand major subunits and minor subunits. The adhesin is located at the very tip of the fimbriae, often connected with the shaft via the so-called adapter pilus. The role of the major subunits is yet unclear, although they have been proposed to be important for adherence to mammalian extracellular matrix proteins (Korhonen, 2000). The adhesin and some other minor subunits are responsible for the specific binding to carbohydrate moieties on the surface of eukaryotic cells, therefore contributing to specific adherence. The synthesis, export, correct folding and ordered assembly during the fimbrial biogenesis occurs in a co-ordinated manner (Smyth et al., 1996). In extraintestinal pathogenic *E. coli,* the determinants coding for the P- and S- family adhesins are often located on genomic islands. In diarrheagenic *E. coli* strains, some fimbrial or fimbrial-like gene clusters are also located on chromosomal islands or on plasmids, but their role for the virulence remains unclear (Kaper and Hacker, 1999). The P-, S- and F1C-fimbriae are more exclusively associated with extraintestinal *E. coli* isolates and the tip of these adhesins recognize carbohydrate moieties: Galα(1-4)Gal, α-sialyl-2,3-β-galactose, and GalNAcβ(1-4)Galβ, respectively. These fimbriae are factors contributing to the virulence potential of such strains, but they are not necessarily sufficient to cause disease (Mobley et al., 1994).

Most of the UPEC strains also express curli fimbriae, which are composed of only one polymerized subunit. It is suggested that these fimbriae play a role only in the early phase of infection (e.g., adherence to periurethral skin surface), since they are frequently expressed only at 30 °C (Olsen et al., 1993). In the last years, isolates have

been detected in which co-expression of curli fimbriae and cellulose production occurs at 30 °C as well as at 37 °C (rdar morphotype), but the importance of this trait for the survival and colonization in the host organism remains unclear (Zogaj et al., 2001). The genes coding for curli fimbriae and cellulose biosynthesis are located on the core chromosome of *E. coli* (*csg* and *bcs* operons, respectively). It was demonstrated that curli fimbriae are able to mediate internalization by eukaryotic cells (Gophna et al., 2001).

Although type 1 fimbriae are also expressed by commensal *E. coli* strains (encoded by the *fim* cluster, located on the core chromosome), it is one of the most important adhesins of UPEC. It mediates the adhesion to mannose-containing oligosaccharides on bladder epithelial cells and is shown to be involved in bacterial invasion (Martinez et al., 2000). Recently, it has been reported that type 1 fimbriae are involved in uroepithelial apoptosis, a characteristic phenomenon of UTI (Klumpp et al., 2001).

Dr family adhesins, including Dr fimbriae and AFA-I and AFA-II afimbrial adhesins bind the Dr^a blood group antigens and may facilitate ascending colonization of the urinary tract (Nowicki et al., 1990).

The enterohaemorrhagic and enteropathogenic (EHEC and EPEC) express the non-fimbrial adhesion factor intimin, which is necessary for the attaching and effacing (A/E) phenotype during the infection. The intimin-encoding gene *eae* is located on the LEE pathogenicity island (McDaniel et al., 1995). Recently, a new adherence-conferring factor Iha (IrgA Homologue Adhesin) was characterized in EHEC O157:H7 strains (Tarr, et al., 2000). The *iha* gene is also distributed among UPEC strains, but it is not always expressed e. g. in *E. coli* strain CFT073.

-Toxins and bacteriocins

Toxins are prominent virulence factors of bacterial pathogens. Three toxins play a major role during UTI: the cytotoxic necrotizing factor 1 (CNF1), the cytolethal distending toxin (CDT) and α-haemolysin. CNF1 is widely distributed in extraintestinal pathogens (Andreu et al., 1997) and belongs to a toxin family which modifies Rho, a subfamily of small GTP-binding proteins that are regulators of the actin cytoskeleton (Aktories, 1997). The gene for CNF1 is chromosomally located on different pathogenicity islands of UPEC (Blum et al., 1995; Toth et al., 2003). Eukaryotic cells intoxicated with CNF1 exhibit membrane ruffling, formation of focal adhesions and actin stress fibers and DNA replication in absence of cell division. CDT is a secreted protein which has the capacity to inhibit cellular proliferation by

inducing an irreversible cell cycle block at the G_2/M position (Comayras et al., 1997). CDT is composed of three polypeptides (CdtA, B and C) which are all required for CDT activity (Elwell et al., 2001). The direct role of the toxin in uroinfection, however, remains to be proven. The α-haemolysin is a member of the RTX toxin family, which is widely disseminated among pathogenic bacteria and widely distributed in UPEC as well as in EHEC isolates. The *hly* gene cluster encoding the toxin and the enzymes for its biosynthesis is located on PAIs or on plasmids. The type I secretion pathway, a posttranslational maturation and the presence of C-terminal calcium binding domain are characteristics of this pore-forming toxin (Wagner et al., 1983; Holland et al., 1990).

In addition, the so-called autotransporter proteins are widely distributed in *E. coli* (Henderson et al., 1998). This group includes a number of serine proteases in *Enterobacteriaceae* (SPATE) which are considered to be toxins: the plasmid-encoded toxin (Pet) of EAEC, the protease Pic of EAEC and *Shigella flexneri,* EspC of EPEC, EspP of EHEC, Tsh of avian pathogenic *E. coli,* SepA of *Shigella flexneri,* and Sat of UPEC (Dutta et al., 2002). The Sat (Secreted Autotransporter Toxin) is widely distributed in UPEC and was shown to have cytopathic activity (elongation and vacuolation of eukaryotic cells). Sat-specific antibodies were found in the serum of *E. coli*-infected mice. Nevertheless the inactivation of the *sat* gene did not attenuate the the *E. coli* strain CFT073 (Guyer et al., 2000). All SPATEs possess a characteristic GDSGS serine protease motif and it is tempting to speculate that their protease activity may serve as peptide-providing source for the bacteria.

The Stx toxin family (Shiga-like Toxins e. g., Stx1 and Stx2), encoded on lambdoid-like bacteriophages in the chromosome, are the most important virulence factors of EHEC (Kaper and Hacker, 1999).

Other secreted compounds, such as colicins and microcins, are also widespread among *E. coli* strains and are believed to mediate antagonistic relationships, thus contributing to competitiveness and the effective colonization the host. Microcins are peptides of a relatively small size (1.18 to 9.00 kDa). They are considered as modified peptide antibiotics since they are synthesized as peptide precursors which are subsequently modified by other proteins. They exhibit wide range of cellular targets: colicin B17 has been shown to be an inhibitor of the DNA gyrase (Vizàn et al., 1991), colicin C7 inhibits protein synthesis (Guijarro et al., 1995), and colicin V disrupts the membrane potential (Yang and Konisky, 1984). Microcin H47, encoded by the

chromosomally-located *mch* gene cluster, was shown to be ribosomally synthesized as a peptide precursor (Rodriguez et al., 1999).

-O-, K-antigens and serum resistance

Falkenhagen and coworkers reported that the most frequent K antigens determined in 253 UPEC isolates are K1 and K5 (31 % and 35 % of the cases respectively); nevertheless more than 26 different K-antigens were identified. High prevalence of these two capsular serogroups is not astonishing, since both capsular oligosaccharides mimic human antigens, thus preventing effective immune response against bacteria expressing them. The K1 capsule is present in all MENEC isolates and contributes to the the ability to cross of the blood-brain barrier (Kim, 2002). The capsule-encoding *kps* gene clusters of *E. coli* are located on the chromosome as part of different genomic islands (Whitfield and Roberts, 1999; Welch et al., 2002).

Lipopolysaccharide (LPS) is a key component of the outer membrane of Gram-negative bacteria. It comprises three distinct regions: Lipid A, the oligosaccharide core, and commonly a long-chain polysaccharide O antigen that causes a smooth phenotype (Fig. 4).

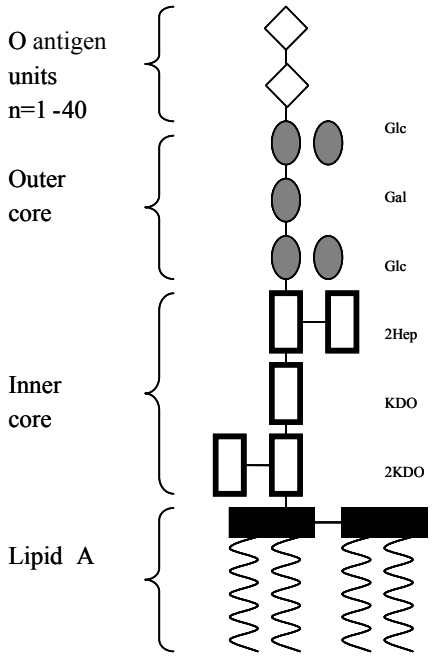

Fig. 4: Schematic representation of LPS structure in Gram-negative bacteria. The O side-chain consists of up to 40 polymerized units.

Lipid A is the most conserved part of LPS. It is connected to the core part, which

links it to the O repeating units (Fig. 4). In *Escherichia coli*, five different core structures (K-12 and R1-R4) have been described (Amor et al., 2000; Holst, 1999; Rietschel et al., 1992). The O repeating units are highly polymorphic, and more than 190 serologically distinguished forms in *E. coli* are known today (Ørskov et al., 1992). The genes coding for LPS core synthesis are located at a conserved position on the *E. coli* K-12 chromosomal map (81-82 min) (Berlyn, 1998). The *wa** (formerly *rfa*) gene clusters contain the genes which code for the enzymes required for the core biosynthesis and assembly and consist of three operons, defined by their first genes *gmhD, waaQ* and *waaA*. Although the O repeating unit-encoding gene cluster (*wb**, former *rfb*) is extremely polymorphic within the species *E. coli*, it is localized at a conserved position on the *E. coli* K-12 chromosome between the genes *galF* and *gnd* (45.4 min.) (Berlyn, 1998). These determinants consist of several sugar transferase-, epimerase- and isomerase-encoding genes, the O antigen flippase (*wzx*), the O antigen polymerase (*wzy*, formerly *rfc*) as well as the genes coding for enzymes involved in carbohydrate biosynthesis pathways. Until now, several *E. coli* O antigen-encoding gene clusters have been studied, e. g. those of serotypes O7, O111, O113, and O157 (Marolda et al., 1999; Paton et al.,1998; Wang et al., 1998; Wang and Reeves, 1998). They show no significant nucleotide homology between each other, with the exception of some common genes such as *manC* and *manB*. However, they contain a conserved range of predicted enzyme activities. The O6 antigen is widely distributed among pathogenic and non-pathogenic faecal *E. coli* isolates and is often found in uropathogenic *E. coli* strains. Since LPS is located on the outer surface of bacterial cells, its expression is known to be responsible for many features of the cell surface of the Gram–negative bacteria, such as resistance to detergents, hydrophobic antibiotics, organic acids, serum complement factors, adherence to eukaryotic cells etc. (Goldmann, et al., 1984; Grossmann et al., 1991; Lukowski et al., 1996; Svanborg-Edén et al., 1987; Barua, et al., 2002 ; Jacques, 1996). It has been suggested that some of these characteristics, especially resistance to the bactericidal effect of the complement system, are dependent on the length of the O side chain (Porat et al., 1992). LPS is believed to significantly contribute to virulence by protecting bacteria from the bactericidal effect of serum complement (Hull et al., 1997; Reeves, 1995; Valvano, 1992). Moreover, it has recently been reported that the K5 capsule does not contribute as much to serum resistance of *E. coli* strains as the O antigen (Burns and Hull, 1998).

-Iron acquisition systems

The availability of Fe(II) ions, an essential factor for bacterial growth, is limited in mammalian hosts. Most of the iron is located intracellularly, primarily bound to ferritin and chelated in haeme. Therefore bacteria have developed different mechanisms to release iron from host proteins. Low molecular weight chelators (siderophores) are secreted from *E. coli*. These molecules "liberate" Fe^{3+} ions from host carriers and – by recognizing specific receptors located in the bacterial outer membrane – transport it into the bacterial cell (Guerinot, 1994). The genes coding for the biosynthesis of such iron-uptake systems in *E. coli* may be located on plasmids or on the chromosome. The gene clusters encoding the enzymes for enterobactin (*ent*) and the ferric dicitrate transport system (*fec*) have a commonly conserved localization in the *E. coli* core genome. However the *fec* gene cluster has been identified to be PAI-encoded in *Shigella flexneri* (Luck et al., 2001). The *iuc* operon coding for aerobactin is either located on plasmids (pColV) or on different genomic islands, whereas the yersiniabactin-encoding HPI (*fyu/irp)* is widely distributed in *Enterobacteriaceae* and shows a rather conserved chromosomal localization at the *asnT* gene. This genetic element is able to "jump" between the *asn* tRNA genes in the chromosome of *Yersinia pseudotuberculosis* (Buchrieser et al., 1998). Recently, the *chu* haeme transport locus in the chromosome of EHEC O157:H7 strain has been characterized (Torres and Payne, 1997). This system enables the bacteria to utilize iron directly from the haeme and is widely distributed among UPEC isolates (Wyckoff et al., 1998). The *iro* gene cluster (coding for the enzymes required for salmochelin biosynthesis) first described for *Salmonella enterica* (Bäumler et al., 1996) is involved in the uptake of catecholate-type siderophore compounds. The *iro* genes are widely distributed among *E. coli* isolates and can be chromosomally or plasmid-encoded (Dobrindt et al., 2003; Sorsa et al., 2003). The ability for iron acquisition of bacteria might be advantageous for their survival in the host organism, therefore it can be considered an important fitness factor.

5. The *Escherichia coli* strain Nissle 1917

The *E. coli* strain Nissle 1917 has been isolated by Prof. Dr. med. Alfred Nissle (1874 – 1965), in 1917, from faeces of a non-commissioned pioneer officer, who, unlike his comrades had not suffered from any of the intestinal disorders then rampant in the region of Dobruja in south-eastern Europe (Nissle, 1918). Investigating the

antagonism of various *E. coli* strains against pathogenic intestinal bacteria, Nissle was the first who administered "antagonistically strong *E. coli* strains" to diarrhea patients. The Mutaflor® probiotic preparation consists of the Nissle 1917 strain (also designated strain DSM6601) that belongs to this group of *E. coli* isolates.

The use of *E. coli* as a probiotic has the advantage that this microorganism is able to colonize successfully the human gut (Malchow et al., 1995). In early studies the Mutaflor® preparation exhibited antagonistic characteristics against *Salmonella typhi* (Nissle, 1925). This probiotic agent is widely used for treatment of chronic constipation, dyspepsia, colitis, enteritis, gastroenteritis, some cases of stomach and duodenum ulcers, as well as for preventing of *Candida albicans* intestinal infections after antibiotic treatment (Kruis et al., 2001; Malchow, 1997; Möllenbrink et al., 1994; Rembacken et al., 1999; Lodnikova-Zadnikova, and Sonnenborn, 1997) . Roerig and Ulrich documented several cases of partially recurring UTIs that were treated orally with Mutaflor®. A certain beneficial effect on cancer patients is also observed, probably due to the restoration of normal microecological conditions in the gut. The *E. coli* strain Nissle 1917 also exhibits an inhibitory effect on *Salmonella typhimurium* invasion into epithelial cells (Ölschläger, A. T., and Altenhöfer, A., unpublished data). Nevertheless, the exact mechanisms of the probiotic activity of this strain are still unclear (Fig. 5). *E. coli* strain Nissle 1917 has been in part pheno- and genotypically characterized (Blum, et al., 1995; Tab. 1).

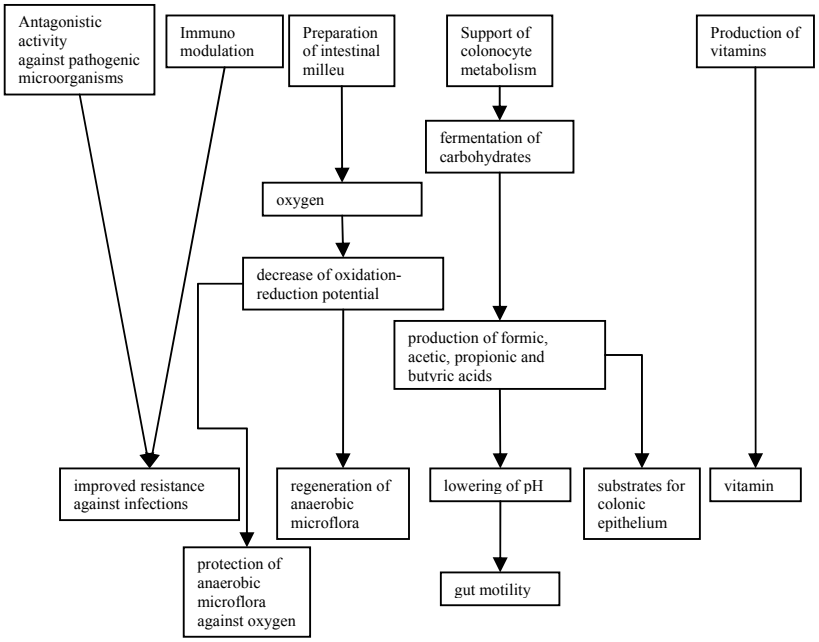

Fig. 5: Modes of action of MUTAFLOR®: application in inflammatory and functional bowel disease (modified from "Mutaflor®-Brief summary of therapeutic principles", Ardeypharm GmbH).

This strain does not express any of the so-called mannose-resistant haemagglutinating factors (MRHA) like P- or S-fimbriae, but expresses type 1 and F1C-fimbriae, which do not agglutinate erythrocytes. It exhibits a semi-rough O6 LPS phenotype, serum sensitivity and does not produce any known virulence-associated toxin. In addition, it produces microcins, exhibits the rdar morphotype, i. e. strongly co-expression of curli fimbriae and cellulose in a temperature-independent manner (30 °C and 37 °C) and possesses a surprisingly high number of iron-uptake systems (Tab. 1).

Table 1: Properties of different *E. coli* O6 strains and the *E. coli* K-12 strain MG1655 (modified from Blum et al., 1995; Welch et al., 2002)

Strains and *properties*	**MG1655 (non-pathogenic, laboratory)**	**Nissle 1917 (non-pathogenic, faecal)**	**536 (uropathogenic, pyelonephritis)**	**CFT073 (uropathogenic, urosepsis)**
Serotype	K-12 rough	O6:K5:H1 semi-rough	O6:K15:H31 smooth	O6:K2(?):H1 smooth
Serum resistance	-	-	+	+
α-haemolysin production	-	-	+	+
CNF1 production	-	-	-	+
P-fimbriae expression	-	-	+	+
S-fimbriae expression	-	-	+	-
Type1 fimbriae expression	+	+	+	+
F1C fimbriae expression	-	+	-	+
Aerobactin expression	-	+	-	+
Enterobactin expression	+	+	+	+
fec	+	+	-	-
fyu/ irp	-	+	+	+
chu	-	+	+	+
Microcin production	-	+	-	+
rdar morphotype				
30 °C	-	+	+	+
37 °C	-	+	-	-
Plasmids	-	+(2)[1]	-	-

[1]Blum-Oehler et al., 2003

Aim of this work

The aim of this work was to characterize the genome structure and genetic organization of determinants contributing to fitness of the *E. coli* strain Nissle 1917, and to compare them with those of the two well-studied uropathogenic *E. coli* O6 strains 536 and CFTO73.

In the first part of the work, the genetic basis for the semi-rough phenotype of strain Nissle 1917 was investigated. The importance of the O6 side-chain synthesis for serum resistance was studied since serum sensitivity of strain Nissle 1917 is important in terms of biological safety and distinguishes this strain from pathogenic isolates.

In the second part, sequence analysis and a detailed characterization of the fitness determinants was performed. In this context the presence of several genomic islands was described. Their genetic organization and localization in the chromosome as well as expression of the encoded genes of interest was studied and compared with available data of pathogenic *E. coli* O6 strains. These studies resulted in a deeper insight into the genetic background and molecular reasons for the probiotic action of strain Nissle 1917.

Furthermore, the multicellular behaviour and the underlying mechanisms as a trait possibly associated with the probiotic character of strain Nissle 1917 were investigated.

III. Materials

1. Bacterial strains

The *E. coli* strains used in this work are listed in Tab. 2a and 2b.

Table 2a: Laboratory *E. coli* strains used in this study

E. coli K-12 strain	Characteristics	Reference
MG1655	F^-, Lam^-, *fnr*-267(del), *rph*-1	Blattner et al., 1997
DH5α	F^-, *end*A 1, *hsdR*17(r_k^-, m_k^-), *supE*44, *thi*-1, *rec*A1, *gyrA*96, λ^-, Δ(*argF-lac*)U169 Φ80*dlacZ*ΔM15	Bethesda Research Laboratories, 1986
AAEC 189	Δ*fim*, Δ*lac, recA*$^-$*, endA*$^-$*, hsdR*$^-$*, hsdM*$^+$	Gally, et al., 1994
Sm10 λ*pir*	*Thi*1, *thr*1, *leu*B6, *sup*E44*, ton*A21, *lac*Y1, *rec*A::RP4-2-Tc::Mu, Km^R, λ*pir*	Miller and Mekalanos, 1988
EN99	*araD, Δlac, aroB, rpsL, thi,* Tc^R, pColV-EN77 but no aerobactin synthesis	Ott, M., et al., 1991
XL1-blue	*SupE*44, *hsdR*17, *recA*1, *endA*1, *gyrA*46, *thi, relA*1, *lac*$^-$, F^-(*proAB*$^+$, *lacI*^{q}, *lacZ* ΔM15, *Tn10* (tet^r))	Bullock et al., 1987

Table 2b: Laboratory and wild-type *E. coli* strains used in this study

E. coli strain	Characteristics	Reference
Nissle 1917 (DSM6601, SK22)	O6:K5:H1 *ent*$^+$*, iuc*$^+$*, fec*$^+$*, chu*$^+$*, foc*$^+$*, fim*$^+$*,* flagella$^+$*,* microcin$^+$	Nissle, 1918
Nissle 1917 (DSM6601, SK22) Sm^r	O6:K5:H31 *ent*$^+$*, iuc*$^+$*, fec*$^+$*, chu*$^+$*, foc*$^+$*, fim*$^+$*,* flagella$^+$, microcin$^+$, Sm^r	This study

Table 2b continued

Nissle 1917 (DSM6601, SK22) CR-	O6:K5:H31 *ent*$^{+}$, *iuc*$^{+}$, *fec*$^{+}$, *chu*$^{+}$, *foc*$^{+}$, *fim*$^{+}$, flagella^{+}, microcin^{+}, congo red binding-negative spontaneous mutant	Gabor, N., unpublished
Nissle 1917 *rfaH::cat*	*rfaH* inactivated by a *cat* cassette, Sm^{R}, Cm^{R}	This study
Nissle 1917λP$_{bla}$*::wzy*$_{536}$	*wzy*$_{536}$ fused with P$_{bla}$ integrated into the bacterial attachment site of the λ bacteriophage of Nissle 1917, Ap^{R}	Grozdanov et al., 2003
Nissle 1917λP*wb**$_{\text{Nissle 1917}}$*::wzy*$_{536}$	*wzy*$_{536}$ fused with the upstream region of the *wb**$_{\text{Nissle 1917}}$ integrated into the attachment site of the λ bacteriophage in the chromosome of Nissle 1917, Ap^{R}	Grozdanov et al., 2003
Nissle 1917-*wzy* $_{536}$	*wzy*$_{\text{Nissle 1917}}$ replaced by *wzy*$_{536}$ in the *wb**$_{\text{Nissle 1917}}$ of the Nissle 1917 chromosome, Sm^{R}	This study
Nissle 1917Δ*bcs*	*bcs* gene cluster deleted from chromosome of Nissle 1917, Sm^{R}	This study
536	O6:K15:H31 *ent*$^{+}$, *chu*$^{+}$, *fec*$^{-}$, *hly*$^{+}$, *fim*$^{+}$, *sfa*$^{+}$, *prf*$^{+}$, flagella^{+}, microcin^{-}, Sm^{R}	Berger et al., 1982
536*rfaH::cat*	Gene *rfaH* inactivated by a insertion of *cat* cassette, Sm^{R}, Cm^{R}	Nagy et al., 2000
536-*wzy* $_{\text{Nissle 1917}}$	*wzy*$_{536}$ replaced by *wzy*$_{\text{Nissle 1917}}$ in *wb**$_{536}$ of the 536 chromosome, Sm^{R}	This study
CFT073	O6:K2(?):H1 *ent*$^{+}$, *iuc*$^{+}$, *fec*$^{-}$, *chu*$^{+}$, *hly*$^{+}$, *foc*$^{+}$, *prf/pap*$^{+}$, *fim*$^{+}$, flagella^{+}, microcin+, Sm^{R}	Mobley et al., 1990
RZ439	O6:K5 *aer*$^{+}$, *hly*$^{-}$, *pap/prf*$^{+}$, *sfa/foc*$^{+}$, *fim*$^{+}$	Zingler et al., 1993

Table 2b continued

RZ442	O6:K5 *aer*$^{+}$, *hly*$^{-}$, *pap/prf*$^{+}$, *sfa/foc*$^{+}$, *fim*$^{+}$	Zingler et al., 1993
RZ526	O6:K5 *aer*$^{+}$, *hly*$^{-}$, *pap/prf*$^{+}$, *sfa/foc*$^{+}$, *fim*$^{+}$	Zingler et al., 1993
RZ533	O6:K5 *aer*$^{+}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{+}$, *fim*$^{+}$	Zingler et al., 1993
RZ446	O6:K53:H1 *aer*$^{-}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{-}$, *fim*+	Zingler et al., 1992
RZ537	O6:K53:H1 *aer*$^{-}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{-}$, *fim*+	Zingler et al., 1992
RZ424	O6:K14:H^{-} *aer*$^{-}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{-}$, *fim*+	Zingler et al., 1992
RZ436	O6:K13:H1 *aer*$^{-}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{+}$, *fim*+	Zingler et al., 1992
RZ447	O6:K13:H1 *aer*$^{-}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{+}$, *fim*+	Zingler et al., 1992
RZ502	O6:K2:H^{-} *aer*$^{-}$, *hly*$^{+}$, *pap/prf*$^{-}$, *sfa/foc*$^{+}$, *fim*+	Zingler et al., 1992
RZ496	O6:K^{+}:H^{-} *aer*$^{-}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{+}$, *fim*+	Zingler et al., 1992
RZ532	O6:K^{+}:H31 *aer*$^{-}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{+}$, *fim*+	Zingler et al., 1992
RZ411	O6:K^{-}:H1 *aer*$^{-}$, *hly*$^{-}$, *pap/prf*$^{+}$, *sfa/foc*$^{-}$, *fim*+	Zingler et al., 1992
RZ412	O6:K^{-}:H1 *aer*$^{+}$, *hly*$^{+}$, *pap/prf*$^{+}$, *sfa/foc*$^{-}$, *fim*+	Zingler et al., 1992

Table 2b continued

RZ418	O6:K$^-$:H1 *aer*$^+$, *hly*$^+$, *pap/prf*$^+$, *sfa/foc*$^+$, *fim*+	Zingler et al., 1992
RZ501	O6:K$^-$:H1 *aer*$^+$, *hly*$^-$, *pap/prf*$^+$, *sfa/foc*$^+$, *fim*+	Zingler et al., 1992

2. Plasmids

The plasmids used in this study are listed in Tab. 3.

Table 3: Plasmids used in this study.

Designation	Description	Reference
pUC19	Ap^R, *ori*f1, *lacZ*	Yanisch-Perron et al., 1998
SuperCos 1	Cosmid vector, Ap^R,	Stratagene
pGEM®-T Easy	Cloning vector, Ap^R	Promega
pBluescript-II KS	Cloning vector, Ap^R	Stratagene
pLDR8	*int* gene expression vector, Km^R	Diederich et al., 1992
pLDR9	Cloning vector for integration of wzy_{536} into the λ-*attB*, Km^R	Diederich et al., 1992
pLDR11	Cloning vector for integration of wzy_{536} into the λ-*attB*, Tc^R	Diederich et al., 1992
pCVD442	*ori*R6K, *mob*RP4, *sac*B, Ap^R	Donnenberg and Kaper, 1991
pCos3YC6	Cosmid clone; insert contains the $fim_{Nissle1917}$ gene cluster, Ap^R	Grozdanov et al., 2004
pCos8YA1	Cosmid clone; insert contains the *selC* tRNA-encoding gene and the $wa*_{Nissle1917}$ gene cluster, Ap^R	Grozdanov et al., 2004

Table 3 continued

pCos9YB4	Cosmid clone; insert containing the *iuc*$_{\text{Nissle1917}}$ gene cluster and *sat*$_{\text{Nissle1917}}$, Ap^{R}	Grozdanov et al., 2004
pCos2YE4	Cosmid clone; insert overlaps with insert of pCos9YB4 and contains *iha*$_{\text{Nissle1917}}$, Ap^{R}	Grozdanov et al., 2004
pCos2RF2	Cosmid clone; insert overlaps with insert of pCos2YE4 and contains the K5-encoding *kps*$_{\text{Nissle1917}}$ gene cluster, Ap^{R}	Grozdanov et al., 2004
pCos3YE4	Cosmid clone; insert contains the *mch/mcm*$_{\text{Nissle1917}}$, *foc*$_{\text{Nissle1917}}$ and *iro*$_{\text{Nissle1917}}$ gene clusters, Ap^{R}	Grozdanov et al., 2004
pCos2RA4	Cosmid clone; insert overlaps with insert of pCos3YE4, Ap^{R}	Grozdanov et al., 2004
pCos1YA7	Cosmid clone; insert contains the *argW*$_{\text{Nissle1917}}$ gene, Ap^{R}	Grozdanov et al., 2004
pCos219	Cosmid clone; insert contains the *glyU*$_{\text{Nissle1917}}$ gene, Ap^{R}	This study
pGBC1917	PCR-amplified *bcs*$_{\text{Nissle1917}}$ gene cluster cloned in pGEM T®-Easy, Ap^{R}	This study
pSMK5	*rfaH*$_{536}$ with Cm^{R} in pCVD442	Nagy et al., 2000
pGDC1917	*Kpn*I/*Msc*I deleted *bcs*$_{\text{Nissle1917}}$ (Δ *bcs*$_{\text{Nissle1917}}$) gene cluster cloned in pGEM® T-Easy, Ap^{R}	This study
pCDB1917	*Sac*I/*Sph*I digested Δ*bcs*$_{\text{Nissle1917}}$ gene cluster subcloned in *Sac*I/*SphI* digested pCVD442, Ap^{R}	This study
pPSD1917	7 kb-*EcoR*I/*Hind*III fragment of pCos9YB4 containing *sat*$_{\text{Nissle1917}}$ subcloned in *EcoR*I/*Hind*III digested pUC19, Ap^{R}	Grozdanov et al., 2004

Table 3 continued

pGLW536	*wzy*$_{536}$ cloned into pGEM®-T Easy, Ap^R	Grozdanov et al., 2003
pGLW1917	*wzy*$_{Nissle1917}$ cloned into pGEM®-T Easy, Ap^R	Grozdanov et al., 2003
pCVW536	*wzy*$_{536}$ subcloned from pGLW536 into *Sac*I/*SphI* digested pCVD442, Ap^R	This study
pCVW1917	*wzy*$_{Nissle1917}$ subcloned from pGLW1917 into *Sac*I/*SphI* digested pCVD442, Ap^R	This study
pGWB1917	PCR-amplified *wb**$_{Nissle\ 1917}$ cloned into pGEM®-T Easy, Ap^R	Grozdanov et al., 2003
pGWB536	PCR-amplified *wb**$_{536}$ cloned into pGEM®-T Easy, Ap^R	Grozdanov et al., 2003
pBWB536	*wb**$_{536}$ subcloned from pGWB536 into pBluescript-II KS, Ap^R	Grozdanov et al., 2003
pGLG2504	PCR-amplified 2504 bp of the 5'-proximal region of *wb**$_{536}$ (including JUMPstart region, *wzx*$_{536}$ and the 986 bp-fragment of *wzy*$_{536}$) cloned into pGEM®-T Easy, Ap^R	Grozdanov et al., 2003
pBLG2504	2504 bp of the 5'-proximal region of *wb**$_{536}$ (including JUMPstart region, *wzx*$_{536}$ and the 986 bp-fragment of *wzy*$_{536}$) subcloned from pGLG2504 into *EcoR*-digested pBluescript-II KS, Ap^R	Grozdanov et al., 2003
pGLG2849	PCR-amplified 2849 bp of the 5'-proximal region of *wb**$_{536}$ (including JUMPstart region, *wzx*$_{536}$ and the 1343 bp intact *wzy*$_{536}$) cloned into pGEM®-T Easy, Ap^R	Grozdanov et al., 2003

Table 3 continued

pBLG2849	2849 bp of the 5'-proximal region of *wb**$_{536}$ (including JUMPstart region, *wzx*$_{536}$ and the 1343 bp intact *wzy*$_{536}$) subcloned from pGLG2849 into *EcoR*I-digested pBluescript-II KS, Ap^R	Grozdanov et al., 2003
pLBW1	P$_{bla}$::*wzy*$_{536}$ cloned into *EcoR*I-digested pLDR11, Tc^R	Grozdanov et al., 2003
pGPW1	PCR-amplified P*wb*$_{\text{Nissle 1917}}$::*wzy*$_{536}$ cloned into pGEM®-T Easy, Ap^R	Grozdanov et al., 2003
pLPW1	P*wb*$_{\text{Nissle 1917}}$::*wzy*$_{536}$ subcloned from pGPW1 cloned into *EcoR*I-digested pLDR9, Km^R	Grozdanov et al., 2003

3. Oligonucleotides

The oligonucleotides (primers) used in this study (Tab. 4) were obtained from commercial sources (Sigma-ARK Oligosys or MWG Biotech GmbH). For primers not designed in this study, a reference is given. Primers used for amplification of PAI III$_{536}$-specific sequences are listed in section VIII, Tab. 1.

Table 4: Primers used in this work.

Designation	Sequence (5' - 3')	Binding site of the primer (reference)
argQ1	GCAAGGCGAGTAATCCTCC	5' of *argQ* (Dobrindt, U., unpublished)
argQ2	TGCCTGAGCAATACGCCAG	in *yqaB* (Dobrindt, U., unpublished)
glnX1	TCGCCATGCGTTGCAGTAC	in *yleB* (Dobrindt, U., unpublished)
glnX2	GTAAGTGCACCCAGTTGGG	5' of *glnX* (Dobrindt, U., unpublished)
lysQ1	GGCAGCATAATCCCGCAAG	5' of *lysQ* (Dobrindt, U., unpublished)
lysQ2	AAGCGCGCCATTTCCAGAG	in *nadA* (Dobrindt, U., unpublished)
thrT1	GCCCGAGACGATAAGTTCG	5' *glyT* (Dobrindt, U., unpublished)
thrT2	CGGGGTGTCGTATTCAACG	in *tufB* (Dobrindt, U., unpublished)
glyU1	GCACTTGCTAAGGAGAGCG	5' of *glyU* (Dobrindt, U., unpublished)
glyU2	CCCTCAACCAGACAGCATC	in *b2863* (Dobrindt, U., unpublished)
serU1	TCCTGGCATCATGGCAACC	5' of *serU* (Dobrindt, U., unpublished)
serU2	GAATGATGCCTCGCCGCAA	in *yodB* (Dobrindt, U., unpublished)
asnT1	TATTCGCCCCGTTCACACG	5' of *asnT* (Dobrindt, U., unpublished)
asnT2	GATACCCGCAGTTAAGCGG	in *b1978* (Dobrindt, U., unpublished)

Table 4 continued

asnW1	ATTGTGGGTGGGCATCGCT	5' of *asnW* (Dobrindt, U., unpublished)
asnW2	GCCCTTGAAGATGACGACG	in *b1983* (Dobrindt, U., unpublished)
asnU1	CGCCGTTAAGATGTGCCTC	5' of *asnU*//in *yeeO* (Dobrindt, U., unpublished)
asnU2	TTCCCGTGGTTTCGTTGGC	in *cbl* (Dobrindt, U., unpublished)
argW1	AGCGGGTATACTCATGCCG	5' of *argW* (Dobrindt, U., unpublished)
argW2	CTTCCAGCAGTTGCATCGG	in *intC* (Dobrindt, U., unpublished)
argU1	CGCGATTACACCGCATTGC	5' of *argU* (Dobrindt, U., unpublished)
argU2	ACATGCAGCGACGAGTTGC	in *intD* (Dobrindt, U., unpublished)
wrbA1	TGGTAGCGAATCGCTACGG	5' of *wrbA* (Dobrindt, U., unpublished)
wrbA2	GAGATGCCCATTCACCGAC	in *yccJ* (Dobrindt, U., unpublished)
ssrA1	GCGGCGCGAATGAACATCT	5' of *ssrA* (Dobrindt, U., unpublished)
ssrA2	AGAGTAAGCCCGTCACCGT	in *intA* (Dobrindt, U., unpublished)
serW1	TATAGTAGCGCCCCGTTGC	5' of *serW* (Dobrindt, U., unpublished)
serW2	CTGCTGTTTGGTTCGCTGG	in *clpA* (Dobrindt, U., unpublished)
serT1	ATAGTGCGCGTCATTCCGG	5' of *serT* (Dobrindt, U., unpublished)
serT2	TTACGTCACCCATTCCGGC	in *yccA* (Dobrindt, U., unpublished)
proL1	ATTAGCAGCCACGAGTCGG	5' of *proL* (Dobrindt, U., unpublished)
proL2	GGGTTGACGGATTGTGGAG	in *yejO* (Dobrindt, U., unpublished)
proK1	TCTGCGCAGTAAGATGCGC	5' of *proK* (Dobrindt, U., unpublished)
proK2	CCAAGCTTCAGCATCCCTG	in *dppA* (Dobrindt, U., unpublished)
pheV1	ACGAGACGAGGCGAATCAG	5' of *pheV* (Dobrindt, U., unpublished)
pheV2	CCGTTGAGCGAACGGATTG	in *yghD* (Dobrindt, U., unpublished)
pheU1	TACGGTTTAATGCGCCCCG	5' of *pheU* (Dobrindt, U., unpublished)
pheU2	ACAAGCCATTCGCCAACGC	in *cadC* (Dobrindt, U., unpublished)
metY1	CGGTACACCAAATCCCAGC	5' of *metY* (Dobrindt, U., unpublished)
metY2	AACGGGCGTAGTGTTCAGC	in *yhbC* (Dobrindt, U., unpublished)
leuU1	GACCAGCGATATCCCGAAC	5' of *leuU* (Dobrindt, U., unpublished)
leuU2	ACCATATGCGCTCCGGAAG	in *yhbX* (Dobrindt, U., unpublished)
asnV1	ATATGCGCCCCGTTCACAC	5' of *asnV* (Dobrindt, U., unpublished)
asnV2	ACACTCGGCGCGAATATGC	in *erfK* (Dobrindt, U., unpublished)
ileY1	CCAAGGTGAATGGGAACGC	5' of *ileY* (Dobrindt, U., unpublished)
ileY2	ATGAGCATAGCCACGCTCC	in *b2651* (Dobrindt, U., unpublished)
proM1	GTATTTCGGCGAGTAGCGC	5' of *prom* (Dobrindt, U., unpublished)
proM2	GCCGTTAGTCTGGAAGCTG	in *aslB* (Dobrindt, U., unpublished)
metV1	GACGCGGTGACGAATTACG	in *mltA* (Dobrindt, U., unpublished)
metV2	GCCGTACTAAAGGCACCAG	in *b2817* (Dobrindt, U., unpublished)
leuZ1	TTCGTTCAGACGCTCTGCG	in *yecA* (Dobrindt, U., unpublished)
leuZ2	CAGGCTGAAAGGCCTGAAG	5' *cysT, leuZ* (Dobrindt, U., unpublished)
leuV1	CCAGGCGGATAAACTCCAG	in *fhuF* (Dobrindt, U., unpublished)
leuV2	CGCACCAAACGAGGCGATA	5' *leuV* (Dobrindt, U., unpublished)
glyY1	GTGCTGTAAGGCACAGACC	5' *glyY* (Dobrindt, U., unpublished)
glyY2	TTGGTCACCTGCGTTGGCT	in *yjeS* (Dobrindt, U., unpublished)

Table 4 continued

valW1	TTGCCACCTTCCCACTCTG	5' of *valV, yalW* (Dobrindt, U., unpublished)
valW2	CTGTCAGGCGAGTTTTGCC	in *b1667* (Dobrindt, U., unpublished)
tyrV1	TTCACATAGACCCTGCTTCG	in *narI* (Dobrindt, U., unpublished)
tyrV2	CCACCATAATTCACCACAGC	in *tyrT* (Dobrindt, U., unpublished)
lysV1	CACCACTTTCTCGCCAGCT	5'*valY* (Dobrindt, U., unpublished)
lysV2	TGATAGCGGACAGCTACGC	in *xapR* (Dobrindt, U., unpublished)
alaX1	ACGTAAACGCATCGTGGGG	5' of *alaX* (Dobrindt, U., unpublished)
alaX2	GAGCATGACCGCAGTCTGT	in *yfeA* (Dobrindt, U., unpublished)
serX1	CTCTCTTGCGCATTTTGATTG	5' *ycdV* (Dobrindt, U., unpublished)
serX2	AGACAGGAACGACAATTTGG	5' *serX* (Dobrindt, U., unpublished)
thrW1	AAGGCCATTGACGCATCGC	5' of *thrW* (Dobrindt, U., unpublished)
thrW2	CCATTTACCCCACTCAGCG	in *yafW* (Dobrindt, U., unpublished)
selC1	GCGTGTATTAGGCGGAAAAAAC	5' of *selC* (Dobrindt, U., unpublished)
selC2	CCCTGAACTTCCCCACAAC	in *yicL* (Dobrindt, U., unpublished)
leuX1	GTGGCGTGCGACAGGTATA	5' of *leuX* (Dobrindt, U., unpublished)
leuX2	GTTTCTCCGGCCCTAAGAC	in yi21_6 (Dobrindt, U., unpublished)
ileX1	GAACACCGACTACACGCTG	3' of *rpoD* (Dobrindt, U., unpublished)
ileX2	CAGATGCAAATCCCTGCCG	in *yqjH* (Dobrindt, U., unpublished)
yicJ-up	ATTGTGCGTCGTGAGTGAGT	in *yicJ* (Blum-Oehler, G., personal communication)
yicJ-down	TTGGTTATGGCATGGGAGAC	in *yicJ,* reverse (Blum-Oehler, G., personal communication)
yicK-up	AACGGCTACCACTCCATCAA	in *yicK,* (Blum-Oehler, G., personal communication)
yicK-down	TAATTCCGTCAACAGAGCCG	in *yicK,* reverse (Blum-Oehler, G., personal communication)
yicL-up	GTTCTGATTGCCGCCGTGTT	in *yicL* (Blum-Oehler, G., personal communication)
yicL-down	ATCAGCAGCGTTCCCAGCCA	in *yicL,* reverse (Blum-Oehler, G., personal communication)
csrA-1	TTCAGCCTGGATACGCTGG	in *csrA*
csrA-2	GATCGTGTGAAAGCAGGGG	in *csrA,* reverse
csrAIPCR-1	CCAGCGTATCCAGGCTGAA	in *csrA,* complementary to csrA-1
csrAIPCR-2	CCCCTGCTTTCACACGATC	in *csrA,* complementary to csrA-2
b1976-1	CCTGGCAGGAAGCACTATC	in *csrA*
b1976-2	AGCGTGATGCAGTCTCTGC	in *csrA*, reverse
yfdC-1	CATGAGCATATCCGCCAGG	in *yfdC*
yfdC-2	AGTAGGTAGTGCGAAGGGC	in *yfdC*, reverse
fimZ-1	TATCGAGCGTTAGCCAGCG	in *fimZ*
fimZ-2	TTGCCGGAACAGACGGTT	in *fimZ*, reverse
b2865-1	TTTGGCTTGCTGCCCTGAG	in *b2865*
b2865-2	CAGGAACGTATTCCGGCTC	in *b2865*, reverse

Table 4 continued

yqgA-1	CTGCTCAGCCAACGCTTAC	in *yqgA*
yqgA-2	CCAGCAGCAATAAACCGCC	in *yqgA,* reverse
yejM-1	CATTCTGTCGACCCGTACG	in *yejM*
yejM-2	CCCAGTAATGACGGCGTTG	in *yejM,* reverse
yfeC-1	GCTATAGCCGCCAGACCAT	in *yfeC*
yfeC-2	TAGCCACACCGAGGCATTC	in *yfeC,* reverse
purU-1	GCGCTGTGCCAGTACTTTG	in *purU*
purU-2	CTCGATAGCGCATTGCCAG	in *purU,* reverse
ybgF-1	ACTGGTTGGTATAGCGGCC	in *ybgF*
ybgF-2	CACCAGCGCAATAGCTGCA	in *ybgF,* reverse
ybgFIPCR-1	GGCCGCTATACCAACCAGT	in *ybgF,* complementary to ybgF-1
ybgFIPCR-2	TGCAGCTATTGCGCTGGTG	in *ybgF,* complementary to ybgF-2
yeeO-1	TCCCATCCAGACACCAACC	in *yeeO*
yeeO-2	ATCCAGCAGCTGCCATCAC	in *yeeO,* reverse
yhjW-1	GCGGGATGTGACACCAGTT	in *yhjW*
yhjW-2	TCGTGAGCAGATTGGTCGG	in *yhjW,* reverse
ycdW-1	CCACCGACCACGGATTTGT	in *ycdW*
ycdW-2	GCCAGGTTTGCAGACTCTG	in *ycdW,* reverse
gltX-1	GACCGTCATAACGCGGCTT	in *gltX*
gltX-2	AAACTCGCTTCGCGCCAAG	in *gltX,* reverse
gltXIPCR-1	AAGCCGCGTTATGACGGTC	in *gltX,* complementary to gltX-1
gltXIPCR-2	CTTGGCGCGAAGCGAGTTT	in *gltX,* complementary to gltX-1
nac-1	CTGCGCCACATAACGAACG	in *nac*
nac-2	TATCGCACAACCAGCGCTC	in *nac,* reverse
proA-1	CCGATGAACTGGAAGCACAAAG	in *proA*
proA-2	TTTGTGTGCTTACCGCCACTTC	in *proA,* reverse
proAIPCR-1	CTTTGTGCTTCCAGTTCATCGG	in *proA,* complementary to proA-1
proAIPCR-2	AAGTGGCGGTAAGCACACAAA	in *proA,* complementary to proA-2
yjgB-1	ACTTTATCGGCACCCATCG	in *yjgB*
yjgB-2	GCATGAGGTGATTGGGCG	in *yjgB,* reverse
yjgBIPCR-1	CGATGGGTGCCGATAAAGT	in *yjgB,* complementary to yjgB-1
yjgBIPCR-2	CGCCCAATCACCTCATGC	in *yjgB,* complementary to yjgB-2
argQCFT073	ACACTATCCAGCAGCATACT	in *c3247*
glyUCFT073	GTCGCTTATCTCATCATCTG	in *c3442*
metVCFT073	CATTACCGTTTCCGTTTAGC	in *c3385,* reverse
serUCFT073	CAGAGAGTTCAGGCGTGG	in *c2416*
asnVCFT073	TTGGTCCTAATGGCTCAATG	in *c2445*
leuXCFT073	GCTGTTGAAGGCGGGAAG	in *c5371,* reverse
thrWCFT073	ACCTCAGGATACACACCTAG	in *c0391,* reverse
selCCFT073	TCTTCCTTTTTTGCCTCCCC	in *intC,* reverse
alaXCFT073	TCAATCCATAACGCCAACGC	in *yfeA*
proLCFT073	AAAGCAACAGGTGATTTCA	in *c2727,* reverse
aspVCFT073a	CATTGTGCGTCAGGCAAGTA	in *dnaQ*

Table 4 continued

aspVCFT073b	GCACAGCAACCACTATGAGG	in *c0253,* reverse
upxCFT073-up	ATGTGCGTCGGAGTGAGTCAGAAGCC	in *c0360*
upxCFT073-up1	AAGAAACTCGCCAACAGAAC	in *c0361*, reverse
upxCFT073-down	AACAACAACCTGCCACTAAACATACATTCA	in *c0364*, reverse
upxCFT073-down1	GGATGGATTTATGGTGGCAGT	in *c0363*
iha-up	ATGCGAATAACCACTCTGGCTTCCG	in *iha*
iha-down	ACTGATAGTTCAGCGACATCCCG	in *iha,* reverse
K5-1-up	AGATGGAAGCGGACCTGCGGAACTTG	in *kpsE*
K5-1-down	GGGGTATTGTCGAAGCCGCTCATGTTT	in *kpsD,* reverse
kfi-up	TTTCAAAATAACATTAGTGATAAAGT	in *kfiD*
kfi-down	TATCTTCTCTCCAGAATTTTTACG	in *kfiD*, reverse
K5reg3-up	CTTATCGTCTGGATGGCAGGT	in *kpsM*
K5reg3-down	CCATTACGACCAATAAAGGCAA	in *kpsT*, reverse
iuc-up	GTATGAGCAGGTTTTCCACGC	in *iucC*
iuc-down	TTTTCCTGCCAGACCTGACTG	in *iucC*, reverse
manC-up	TATCATTATGGCTGGTGGTTCAGGCAGTCGGTTG	in *manC*
manB-down	GTAGCCCTCAATAAACTGCCCTTTTTCGTCAA	in *manB*, reverse
Z2005-down	TGATGGTGGTGGGGGAAGGAT	3' in Z2005
Z4201-down	CAGACAGCATCATAGTGGGTTTC	3' in Z4201
Z1664-down	GACAACAGGCTGACCGACTC	3' in Z1664
Z0696-down	GTAACATCATCAACGATTATCTG	3' in Z0696
Z3672-down	GGATTTAGAAGGGATACATTCAGA	3' in Z3672
Z4314-down	CGAAAGAGAGTTGTAAATCAGGGA	3' in Z4314
Z0307-down	GTATTGAAAAAACAAATCGGTAATCAT	3' in Z0307
Z5878-down	GGGATTGAAGGGGCGGAGTTC	3' in Z5878
shiA-forw	CGCCTGCAGGATGAACGATAGATTATGC	*shiA* (Moss et al., 1999)
shiA-rev	GCGGATCCCTGATACTGGATGCTTGAATG	*shiA*, reverse (Moss et al., 1999)
chu1	GGTATTTATGGTTCAGTGATG	upstream of *shuV* (Nagy, G. personal communication)
chu4	TTTTCTCACTCAAATTGAACG	downstream of *shuS,* reverse (Nagy, G. personal communication)
bcs-up	CTATCTGAAAACTTACCAGTCGGCGTA	downstream of *yhjL* (This study)
bcs-down	AACGGAAAGTCAAAAAGTGAGCAAATTC	upstream of *yhjR* (This study)
curli 1	TGCCGCCACAATCCAGCGTAA	upstream of *csgD* promoter (Dobrindt, U., unpublished)
curli 2	CGCACCCAGTATTGTTAACATCATAAA	downstream of *csgD,* reverse (Dobrindt, U., unpublished)
curli 3	GCTATCAAAAAGCACCAGACAGTC	in *csgD* (Dobrindt, U., unpublished)
ygaG-up	GGCTTTTTTCAATTAATTGTGAAG	in *ygaG* (Dobrindt, U., personal communication)
ygaG-down	CTGGGAAGAAAGAGTTCAGAAAA	in *ygaG*, reverse (Dobrindt, U., personal communication)
ygaGIPCR-1	GCATCAACAGCAACGAAGAACTG	in $ygaG_{Nissle1917}$-specific IPCR product

Table 4 continued

ygaGIPCR-2	CTTTGTTCGGCACGCAGAAGC	in *ygaG*$_{Nissle1917}$-specific IPCR product
ygaGIPCR-uni2	ACGCCACAAGACCCTTGAGCA	in *ygaG*$_{Nissle1917}$-specific IPCR product
ygaGIPCR-rev2	TGCCTCTTCTTTTGCCTGTTG	in *ygaG*$_{Nissle1917}$-specific IPCR product
3YE4-1	CAGAGTGATGTTGTTACTTTGTTGCCT	right-hand end of pCos3YE4 insert
3YE4-2	GGACCAACAATCACAAGCAAGCGAGGA	right-hand end of pCos3YE4 insert, reverse
cva13	GGCATTGTTACTATGCGGTATGCAT	between *mcmA* and *mcmM*, reverse (Patzer et al., 2003)
cva15	TGTTCTTTTATATTCCGGTGTCATT	in *mcmI* (Patzer et al., 2003)
cva21	TATCCCGACGATGATTATCAGTGAC	in *mchE*, reverse (Patzer et al., 2003)
cva28	CTTGAACCGATAAGAAACACAGTGT	between *mchC* and *mchD* (Patzer et al., 2003)
cva31	CTCATCTGGCAGTATTCTCCGTTTC	in *mchC*, reverse (Patzer et al., 2003)
cva34	TGTCTTTGTCTGGGTGAGGTCAGGT	in *mchX* (Patzer et al., 2003)
IS2-1	ACAGAACTTGATGGTATGCCTGCG	in left-hand IS*2* of GEI II$_{Nissle\ 1917}$
ORF-2	CTCAATGCTGTTTCTGCCATCGTC	in *shiF* (GEI II$_{Nissle\ 1917}$), reverse
iha-1	CTGTGACTCCGTGGGATAAGGAAT	in *iha* (GEI II$_{Nissle\ 1917}$)
IS2-2	CTATTTCTCTGGCGTAAGCAATACC	in right-hand IS*2* of GEI II$_{Nissle\ 1917}$, reverse
mutaistyp-1	TTAGCGTTGGCGTAAGGCGAA	in IS-element located in *fim*$_{Nissle\ 1917}$ (*fimB*)
mutaistyp-2	TTCCCTTACGAGATAAAAATACCC	in *fimA*$_{Nissle\ 1917}$, reverse
GEI I-L1	CGCAAAGCCGACGATTTCACC	in ORF5 of GEI I$_{Nissle\ 1917}$
GEI I-L2	TTACTGGCGTATGAAAATGCGGT	in ORF5 of GEI I$_{Nissle\ 1917}$, reverse
GEI I-R1	GGAAACTGGCTACTGGACAATGC	in ORF81 of GEI I$_{Nissle\ 1917}$
GEI I-R2	TCACCTGCTCCTGAATCATTGCC	in ORF81 of GEI I$_{Nissle\ 1917}$, reverse
GEI II-L1	CGACAGGAGAATGACTTTATTGATG	in ORF2 of GEI II$_{Nissle\ 1917}$
GEI II-L2	CAGATTTCCGCACAGTATTTCTG	in ORF2 of GEI II$_{Nissle\ 1917}$, reverse
GEI II-R1	ACAAACCATAGCGAGATAATAAACCA	in ORF75 of GEI II$_{Nissle\ 1917}$
GEI II-R2	CAAGAACTTTTTATTTGTAATGCTGTA	in ORF75 of GEI II$_{Nissle\ 1917}$, reverse
GEI III-rev	GGAAGCGGGCTCTTATATTGCGA	in ORF25 of GEI III$_{Nissle\ 1917}$, reverse
asnT1	CACGATTCCTCTGTAGTTCA	*asnT* (Karch et al., 1997)
int3	TCCTTTTTCGTGTCGTAACCC	*int,* reverse(Karch et al., 1997)
ybtQup	CGGGCGGCCTCTTCTACCT	*ybtQ* (Karch et al., 1997)
ybtQ1lp	GCGATGCGGCGACAAATGC	*ybtQ,* reverse (Karch et al., 1997)
ybtEup	GCAAGATAGACAAAAAAACGCC	*ybtE* (Karch et al., 1997)
fyuybtE	GCTGACAACGGTAGACGAGA	*fyuA,* reverse (Karch et al., 1997)
50A	ATTGATCCACCGTTTTACTC	IS*100* (Karch et al., 1997)
50B	CGAACGAAAGCATGAAACAA	IS*100,* reverse (Karch et al., 1997)
int5	ATGGAATCGGGTTTATGGG	*int* (Karch et al., 1997)
ybtSlp	GCTATTGCTGAACTGGAGG	*ybtS,* reverse (Karch et al., 1997)
ybtQup	GCCGCCAGTCTATCCACA	*ybtQ* (Karch et al., 1997)
ybtA1lp	GAATCGGCCACAATAGGA	*ybtA,* reverse (Karch et al., 1997)

Table 4 continued

ybtAup	GGTATGGATATTTTGCTCTGG	*ybtA* (Karch et al., 1997)
irp2 RP	TCGTCGGGCAGCGTTTCTTCT	*irp2,* reverse (Karch et al., 1997)
irp-1up	TTCCGGTCCCCTGTCTCA	*irp1* (Karch et al., 1997)
ybtTlp	ATCCGCCAATGTCTATCA	*ybtT,* reverse (Karch et al., 1997)
ybtAup	GGTATGGATATTTTGCTCTGG	*ybtA* (Karch et al., 1997)
ybtAlp	GGTAATGCTCTCGAATCGG	*ybtA,* reverse (Karch et al., 1997)
irp1up	TGAATCGCGGGTGTCTTATGC	*irp1* (Karch et al., 1997)
irp1lp	TCCCTCAATAAAGCCCACGCT	*irp1,* reverse (Karch et al., 1997)
ybtTup	TGCAAAAACAGCCCAGTA	*ybtT* (Karch et al., 1997)
fyuA1lp	CATTCCATCCCACATAGG	*fyuA,* reverse (Karch et al., 1997)
irp2 FP	AAGGATTCGCTGTTACCGGAC	*irp2* (Karch et al., 1997)
irp2 RP	TCGTCGGGCAGCGTTTCTTCT	*irp2*, reverse (Karch et al., 1997)
irp1-1up	TTCCGGTCCCCTGTCTCA	in *ybtU* (Karch et al., 1997)
ybtTlp	ATCCGCCAATGTCTATCA	in *ybtU*, reverse (Karch et al., 1997)
III.33	CCACAAATCGTCTTCGGCC	in *iroB* (Dobrindt et al., 2001)
III.34	GGCCGAAGACGATTTGTGG	in *iroB*, reverse (Dobrindt et al., 2001)
IS2-up	CATACATCACTACGCCCACGAAA	upstream of the left-hand copy of IS*2* in GEI $II_{Nissle1917}$
IS2-down	ACAGCACTCTATGAGGACTTAGTA	downstream of the right-hand copy of IS*2* in GEI $II_{Nissle1917}$, reverse
1A	GCCAGCAGCAGGACACGACTT	in 30 kb *serX – iroC* region of *E. coli* CFT073
1B	CCAGTCTGTCGGTGAGACTGAAATCG	in 30 kb *serX – iroC* region of *E. coli* CFT073, reverse
2A	GGCAGAATGGAGAAACAACAGAACCT	in 30 kb *serX – iroC* region of *E. coli* CFT073
2B	CCCGCTTTTCCCCATAATGCCAATCAC	in 30 kb *serX – iroC* region of *E. coli* CFT073, reverse
3A	ATCATCAGGGGCTATTCTACAGCAAAC	in 30 kb *serX – iroC* region of *E. coli* CFT073
3B	TTCGTCACCATAACGGTAAGTCAGTGC	in 30 kb *serX – iroC* region of *E. coli* CFT073, reverse
4A	GCGATTAGCCAGTCGTTTTGGTTATAC	in 30 kb *serX – iroC* region of *E. coli* CFT073
4B	CATCTTCATAGCATTCTTTTCCCTGAAG	in 30 kb *serX – iroC* region of *E. coli* CFT073, reverse
focG-up	GTGAATTAATACTTCCCGCACCAGCAT	in *focG*
focG-down	CTGTTACAGGGAGGGTATTGCCAC	in *focG,* reverse
pheVshiA-1	ACTATGGCACTGACTGACGCAAAAAT	in *int* gene immediately downstream of *pheV*
pheVshiA-2	CCAGTATGTTTCATGCCTGCCGCTGA	in *shiA*, reverse
sat-up	CGGAGATAAATTCTCTTTCCATAATC	in *sat*
sat-down	TAAATGTAAATCCCTGACCGTTTATG	in *sat,* reverse

Table 4 continued

MUTA-up	TCCTGAAAAACAAT	in *wzy*$_{Nissle1917}$
MUTA-down	TAACATTAAATATTT	in *wzy*$_{Nissle1917}$, reverse
Muta-1	ATACTACGACGGTAAATGGT	in the variable central part of *fimA*$_{Nissle1917}$ (Blum-Oehler et al., 2003)
Muta-2	TACATCAGTATCGGTAGCAT	in the variable central part of *fimA*$_{Nissle1917}$, reverse (Blum-Oehler et al., 2003)
Muta-5	AACTGTGAAGCGATGAACCC	non-coding region in pMUT1 (Blum-Oehler et al., 2003)
Muta-6	GGACTGTTCAGAGAGCTATC	non-coding region in pMUT1 (Blum-Oehler et al., 2003)
Muta-7	GACCAAGCGATAACCGGATG	non-coding region in pMUT2 (Blum-Oehler et al., 2003)
Muta-8	GTGAGATGATGGCCACGATT	non-coding region in pMUT2 (Blum-Oehler et al., 2003)
Muta-9	CGGAGGTAACCTCGAACATG	*mobBD* region in pMUT2 (Blum-Oehler et al., 2003)
Muta-10	CGCCGTATCGATAATTCACG	*mobBD* region in pMUT2 (Blum-Oehler et al., 2003)
R1C3l	GGGATGCGAACAGAATTAGT	5' in *wa*$_{O6}$ (Amor et al.)
R1K15	TTCCTGGCAAGAGAGATAAG	3' in *wa*$_{O6}$, reverse (Amor et al.)
waalücke 1	CGCACTCACTGATGCCCAGCA	in *waaC,* reverse
waalücke 2	AGTCCAATCCATGCTTTACGCCAT	in *waaC*
482	CACTGCCATACCGACGCCGATCTGTTGCTTGG	5' in JUMPStart (Coimbra et al.)
412	ATTGGTAGCTGTAAGCCAAGGGCGGTAGCGT	3' in *gnd*, reverse (Coimbra et al.)
M13–uni	TGTAAAACGACGGCCAGT	Promega, Mannheim (Germany)
M13-rev	CAGGAAACAGCTATGAC	Promega, Mannheim (Germany)
SuperCos-F	CGGCCGCAATTAACCCTCAC	SuperCos 1 sequence primer
SuperCos-R	GCGGCCGCATAATACGACTCACT	SuperCos 1 sequence primer
LG1	GTTTCTTGTATTCAGTATGCT	in *wb*$_{O6}$, reverse
LG2	TGG GTT TGC TGT GTA TGA GGC	in *wb*$_{O6}$, reverse
LG3	TAT GAG CCC TGT TAT AAC TTG GGA	in *wb*$_{O6}$, reverse
LG4	CACCTTGCCCTCCTGAACCATTAT	in *wb*$_{O6}$, reverse
LG5	GAATAGTTTACCTGAGGATTTTTTATC	in *wb*$_{O6}$, reverse
LG6	GTCTTCCTACACCCAGCATCTCCA	in *wb*$_{O6}$, reverse
LG7	CCAGCCATAATGATAGGTGTAA	in *wb*$_{O6}$, reverse
LG8	AACCTGAAAGAAGGGGCGAAG	5' in *galF* (position 841–861), amplification of *wb*$_{O6}$ upstream region
LG9	GCTCTAGAGCTTAGGTGTAATTATATTATT	in *wb*$_{O6}$, *Xba*I restriction site, reverse

Table 4 continued

LG15	CCATCGATGGGTGCCTGACTGCGTTAGCAATTTAA CTGTGATAAACTACCGCATTAAAGCTTATCGATGA TAAGAGAGGTTTCTTGTATTCAGTATTGCT	in wb_{O6}, P_{bla} fused to primer LG1
LG16	GCTCTAGAGCGTTTCTTGTATTCAGTATGCT	in wb_{O6}, *Xba*I restriction site fused to primer LG1

4. Chemicals and enzymes used in this study

The chemicals and enzymes used in this work were provided by the following companies:

Amersham Biosciences, Buckinghamshire; Clontech, Heidelberg; Difco, Augsburg; Fluka, Deisenhofen; Invitrogen, Karlsruhe; MBI Fermentas, St. Leon-Rot; Merck, Darmstadt; Molecular Probes, Poort Gebouw; New England Biolabs, Frankfurt am Main; Oxoid, Wesel; Pharmacia Biotech, Promega, Heidelberg; Quiagen, Hilden; Roche Diagnostics, Mannheim; Roth, Karlsruhe; Sigma, Deisenhofen; Schleicher & Schuell, Dassel; Stratagene, Heidelberg.

The following kits are used:

- "ECL™ Direct Nucleic Acid Labeling And detection System", Amersham Biosciences, Buckinghamshire
- "PRISM™ BigDyeTerminator-Cycle Sequenzierkit", PE Applied Biosystems, Weiterstadt
- "Expand Long Template PCR System", Roche, Mannheim
- "pGEM® T-Easy Vector System I", Promega, Heidelberg
- "QIAquick Gel Extraction Kit", Qiagen, Hilden
- "QIAGEN Midi And Maxi Kit", Qiuagen, Hilden
- "Gigapack III XL packaging mixture", Stratagene, Heidelberg
- "EnzCheck® Protease Assay Kit", Molecular Probes, Leiden

5. Equipments

- Autoclave	Webeco
- Camera	Nikon F301
- Centrifuge	Heraeus Biofuge 13R

	Heraeus Megafuge 1.0R
	Eppendorf 5415 C
	Beckman J 2-21
- Cleanbench	Nunc Inter Med
- Computer program	Microsoft Office 2000
- Electronic balance	Chyo Balance Corp.
- Electrophoresis chamber	Bio Rad
- Electroporator	BioRad, Gene Pulser transfection
- X-ray films, autoradiography	Hyperfilm, ECL
- Fluorescence imager	Typhoon 8300
- Freezer	Revco
- Gene linker	BioRad, Gel Doc 2000
- Hot Plate	Eppendorf Thermostat 5320
- Hybridization oven	HybAid
- Ice machine	Scotman AF-20
- Incubator	Mammert Tv40b
	Heraeus B5050E
- Magnetic strirrer	GLW
- Micro pipettes	Gilson, Eppendorf
- Microwave oven	Moulinex
- Nylon membrane	Byodine, Nytran
	Optitran BA-S 85
- PCR-Thermocycler	Biometra, T3 Thermoblock
- pH-Meter	WTW pH523
- Photometer	Pharmacia Ultrospec 3000
- Platform Shaker	Bühler E55 swip
	Innova TM 4300
- Power Supply	BioRad 200mA, 500V
- Printer	Laserjet 2100C
- Rotation mixer	Eppendorf mixer 5432
- Scanner	HP ScanJet IIcx

- Sequence analyser	ABI Prism 310 Genetic Analyzer
- Shaker	GLW
	GFL Wasserbad
	Innova TM 4300
- Speedvac-concentrator	UNIVAPO 150H Uniequipe
- Sterile filter	Schleicher & Schuell 0.22µm
- Fluorescence imager	Typhoon 8600
- Vacuum-blotter	Pharmacia
- Vacuum oven	Heraeus instruments VTR 5036
- Video printer	Mitsubishi, Cybertech Cb1
- Vortexer	GLW
- Water bath	GLW 1083

6. Media and supplements

- Distilled water was used for all media.
- All media were autoclaved for 20 minutes.
- For agar plates, 15 g bacteriological agar (Difco Laboratories, Detroit, MI, USA) per one liter of medium (15 % (w/v)) was added before autoclaving.

6.1 LB- (Luria-Bertani-) medium (Sambrook et al., 1989)

Tryptone	10 g
Yeast Extract	5 g
NaCl	5 g
dH_2O	add 1000 ml; pH 7.5

6.2 X-Gal medium

LB-medium supplemented with the following additives:

IPTG (0.1 M)	0.5 ml/l
X-Gal (2 %, (w/v) in N,N'-dimethylformamide)	3 ml/l

6.3 Yeast-tryptone medium (YT)

Tryptone	16 g
Yeast extract	10 g
NaCl	10 g
dH_2O	add 1000 ml

6.4 Medium for detection of aerobactin expression (Braun, et al., 1983; Ott, et al., 1991)

Aerobactin agar plates:	
Nutrient broth	4 g
NaCl	2.5 g
Agar	6 g
Dipyridyl (200 mM)	5 ml
Titriplex (10 mM)	5 ml
dH_2O	add 500 ml
Soft agar for aerobactin agar plates:	
Nutrient broth	0.8 g
NaCl	0.5 g
Agar	0.75 g
Dipyridyl 200 mM	1 ml
Titriplex 10 mM	1 ml
E. coli strain EN99 overnight culture (YT, Tc)	4 ml
dH_2O	add 100 ml

6.5 M9 medium

The M9 salts, 1M $MgCl_2$, 1M $CaCl_2$ were autoclaved separately and added to the autoclaved and pre-cooled (45 °C) dH_2O, supplemented with 15 % (w/v) agar. The

casamino acids, glucose and thiamine stock solutions were prepared in dH_2O, filter sterilized (0.22 µm) and added to the autoclaved pre-cooled dH_2O supplemented with 15 % (w/v) agar and M9 salts.

5 × M9 salts:	
$Na_2HPO_4 \times 2\ H_2O$	60 g
KH_2PO_4	15 g
NaCl	2.3 g
NH_4Cl	5 g
H_2O	add 1000 ml
casamino acids (10 %)	30 ml
1M $MgCl_2$	2 ml
1M $CaCl_2$	0.1 ml
Glucose (20 %)	10 ml
Thiamine (0.2 %)	1 ml
dH_2O	add 750 ml

6.6 Congo red medium

The congo red and Coomassie brilliant blue stock solutions were sterilized by filtration through a 0.22 µm sterile filter and added to the autoclaved medium.

Trypton	10 g
Congo-red (0.4 mg/ml)	1 ml
Coomassie brilliant blue (0.2 mg/ml)	1 ml
dH_2O	add 1000ml

6.7 Antibiotics

Stock solutions of water soluble antibiotics were prepared in distilled H_2O and sterilized by filtration through a 0.22 µm sterile filter. The ethanol soluble antibiotics were

dissolved in 96 % ethanol. All antibiotics were added after cooling the medium to 45 °C in ratio of 1:1000 (v/v) from the stock solution.

Antibiotics	Stock solution (µg/ml)
Ampicillin	100
Chloramphenicol	30
Kanamycin	50
Streptomycin	50
Tetracycline	15

IV. Methods

-If not explicitly mentioned, an Eppendorf (Centrifuge 5415 C) table centrifuge has been used (13 000 rpm) for centrifugation.

1. Construction of a cosmid genomic library of the *E. coli* strain Nissle 1917 and selection of clones

1.1 Construction of a cosmid genomic library of the *E. coli* strain Nissle 1917

The cosmid library has been constructed using the Gigapack III XL packaging mixture (Stratagene), following the instructions of the manufacturer (Blum-Oehler, G., and Waldtschmidt, A., personal communication), using *E. coli* XL1-blue MR as a host strain.

1.2 Screening of the genomic cosmid library by colony blot hybridisation

The cosmid clones were grown overnight on LB agar plates, supplemented with ampicillin. After cooling to 4 °C, the colonies were transfered from the agar plate to nylon membrane (Biodyne B, 0.45 μm). The bacterial cells were lysed and the DNA was denaturated by laying the membranes (with the colonies on the upper side) for 5 min on Whatman filter paper saturated with solution I. After short drying, the membranes were laid on Whatman filter paper saturated with neutralization solution 2 for 5 min. The DNA was fixed by “baking” for 30 min at 80 °C and afterwards the membranes were incubated in solution III for 6-12 h at 37 °C. After removal of the cell debris for 1 min in 2 × SSC the colony blots were ready for hybridization.

Solution I	0.5 N NaOH; 1.5 M NaCl
Solution II	0.5M Tris-HCl, pH 7.5; 1.5M NaCl
Solution III	50 mM Tris-HCl, pH 7.5; 50 mM NaCl 0.1 % SDS; 100 μg/ml proteinase K

1.3 Screening the genomic cosmid library by PCR

The cosmid clones from one agar plate were collected after overnight growth and suspended in 1 ml sterile dH_2O. The suspension was heated for 10 min at 90 °C and stored at -20°C. For performing the screening PCR (50 µl reaction volume), 1µl of the cell suspension of every plate was used as template.

1.4 Screening the genomic cosmid library by dot blots

Cosmid DNA was isolated as described in IV.2.1.1 and precipitated by addition of 0.7 vol isopropanol followed by incubation on ice for 15 min and 15 min centrifugation. The DNA pellet was washed once with 70 % ethanol (v/v) and dried at 37° C. After dissolving the DNA in 20 µl sterile dH_2O, 5 µl were transfered to a nylon membrane (Biodyne B, 0.45 µm) by pipetting. The DNA was then denaturated and neutralized as described above (see section IV.1.2). The cosmid DNA was fixed to the membrane by exposure to UV for 90 sec. After pre-incubation for 5 min in 2 × SSC, the membranes were ready for hybridization.

Solution I	50mM Tris-HCl pH 8.0; 10 mM EDTA; 100 µg/ml RNaseA
Solution II	0.2 N NaOH; 1 % SDS
Solution III	3 M Na-acetate

2. Isolation of DNA

2.1 Isolation of plasmid (cosmid) DNA

2.1.1 Small-scale preparation of plasmid DNA (Sambrook, et al., 1989) using phenol-chloroform extraction and ethanol precipitation

Bacteria were grown overnight in 5 ml of LB medium supplemented with ampicillin. 1.5 ml of the overnight culture was centrifuged for 2 min. The pellet was resuspended in 100 µl of ice-cold solution I by vigorous vortexing. 200 µl of freshly prepared solution II were added, mixed with the cell suspension by inverting the Eppendorf tube and

incubated on ice for 5 min. After addition of 150 µl ice-cold solution III and carefully mixing by inverting, the tube was stored for 5 min on ice. Cell debris and chromosomal DNA were pelleted by centrifugation (15 min) and the supernatant was transfered to a new Eppendorf tube. After transferring the plasmid DNA into a new Eppendorf tube, 1 vol phenol:chloroform:isoamylalcohol (25:24:1, v/v/v) was added, carefully mixed and centrifuged for 5 min. The supernatant was transfered into a new tube and 1 vol chloroform:isoamylalcohol (24:1, v/v) was added. After careful mixing by inverting the tube and centrifugation for 5 min the supernatant was transfered into a new tube. The DNA was precipitated by addition of 2.5 vol 96 % ethanol followed by 15 min incubation on ice. The DNA was then pelleted by 15 min centrifugation and washed two times with 70 % (v/v) ethanol. The pellet was dried at 37 °C, dissolved in 20 µl sterile dH_2O. and stored at -20 °C.

2.1.2 Large-scale preparation of plasmid (cosmid) DNA (Sambrook et al., 1989, modified)

Bacteria were grown overnight in 50 ml LB medium supplemented with appropriate antibiotics. The cells were harvested by centrifugation for 15 min at 4 °C and suspended in 4 ml solution I (see section IV.1.4) supplemented with 5 mg/ml lysozyme. After 5 min incubation at room temperature (RT), 8 ml solution II (see section IV.1.4) was added and carefully mixed by inverting. The mixture was incubated for 10 min on ice (or until the solution became clear), and then 6 ml solution III (see section IV.1.4) were added. The cell debris and the genomic DNA were precipitated by centrifugation for 30 min. at 4°C. The supernatant was transfered to a new tube. The plasmid DNA in this solution was precipitated by addition of 0.7 vol isopropanol and collected after 30 min incubation at RT by centrifugation for 15 min. The pellet was washed with 2 ml 70 % ethanol and transfered to new Eppendorf tube. After centrifugation and drying at 37 °C, the pellet was dissolved in 0.5 ml sterile dH_2O and subsequently 50 µl 3 M Na-acetate were added. The solution was subjected to phenol- and phenol/chloroform extractions and ethanol precipitation as described above (section IV.2.1.1). Finally, the plasmid (cosmid) DNA was dissolved in 100 µl sterile dH_2O.

2.2 Isolation of genomic DNA (Grimberg, et al., 1989)

Bacteria were cultivated overnight and cells from 1 ml culture were harvested by centrifugation for 2 min in an Eppendorf tube. After washing with 1 ml TNE buffer the cells were centrifuged for 5 min and then suspended in 270 µl TNEX solution. 30 µl lysozyme (5 mg/ml) were added and the cells were incubated for 30 min at 37 °C. Afterwards, 15 µl proteinase K (20 mg/ml) were added and further incubated for 1-2 h at 65 °C (or until the solution became clear). The genomic DNA was precipitated by addition of 0.05 vol 5 M NaCl (15 µl) and 500 µl ice-cold ethanol and then collected by centrifugation for 15 min. After washing two times with 1 ml 70 % (v/v) ethanol, the DNA pellet was dried on air and dissolved in sterile dH_2O.

TNE	10 mM Tris-HCl, pH 8.0; 10 mM NaCl ; EDTA, pH 8.0
TNEX	TNE; 1 % (v/v) Triton X-100

2.3 Isolation of high molecular weight genomic DNA for pulsed-field gel electrophoresis (PFGE)

The bacteria were grown overnight in 10 ml LB medium at 37 °C. 4 ml of the culture were harvested by centrifugation and the pellet was washed two times with SE buffer. After resuspending the cells in 1 ml SE, the OD_{600} was measured (100 µl cells + 900 µl SE buffer). When necessary, the OD_{600} had to be adjusted in the range between 0.6 and 0.7. A 2 % LGT agarose solution was prepared and cooled to 45-50 °C. 900 µl of the bacterial suspension were mixed with 900 µl agarose and poured into appropriate moulds by pipetting. After solidifying at 4 °C, the agarose blocks were cut in pieces with an approximately identical size (0.5 × 0.3 × 0.1 cm), and then incubated in 5 ml NDS solution (freshly supplemented with 2 mg/ml proteinase K) overnight at 50 °C on a shaker. For the complete removal of the proteinase K, the blocks were afterwards washed at least four times for 2 h with TE buffer. The blocks were then stored at 4 °C in TE buffer for at least one week before they could be used for restriction.

SE buffer	75 mM NaCl; 25 mM EDTA, pH 7.5
TE buffer	10 mM Tris-HCl, pH 7.5; EDTA, pH 7.5
NDS buffer	N-laurylsarkosine; EDTA, pH 9.5; proteinase K (2 mg/ml)
LGT buffer	Tris-HCl, pH 7.5; $MgCl_2$; EDTA, pH 7.5; LGT agarose 2 %

3. DNA

3.1 DNA agarose gel electrophoresis

3.1.1 Horizontal gel electrophoresis

For routine analytical and preparative separation of DNA fragments, horizontal gel electrophoresis was performed under non-denaturing conditions. Depending on the size of the DNA fragments to be separated, the agarose concentration varied between 0.8-2 % (w/v) in running buffer (1 × TAE or 1 × TBE). In order to have a visible running front and to prevent diffusion of the DNA, 0.2 vol Stop-Mix was added to the samples before loading. The electrophoresis was carried out using a voltage in the range between 16-120 V. The gels were stained in an ethidium bromide solution (10 mg/ml), then washed with water and photographed under a UV-transilluminator.

50 × TAE buffer:	
Tris	242 g
Acetic acid	57.1 ml
0.5 M EDTA, pH 8.0	100 ml
dH_2O	add 1000 ml
10 × TBE buffer:	
Tris	108 g
Boric acid	55 g
0.5 M EDTA, pH 8.8	40 ml
dH_2O	add 1000 ml
10 × Stop-Mix:	
Bromphenol blue	0.25 %

Xylene cyanol	0.25 %
Ficoll type 400	25 %

3.1.2 Isolation of DNA fragments from agarose gel ("QIAquick® Gel Extraction Kit")

The agarose pieces containing the DNA fragment of interest were cut out of the gel, and subsequently melted for 10 min at 50 °C in QG buffer (supplied by the manifacturer). The DNA was separated from the rest of the solution by applying the mixture to QIAquick® spin columns followed by centrifugation for 1 min. The columns were then washed with 750 µl PE buffer (supplemented with ethanol). Then the PE buffer was completely removed by centrifugation (2 × 1 min). Finally, the DNA was by eluted from the column with 20-50 µl sterile dH_2O.

3.1.3 Determination of the DNA concentration

The DNA concentration was measured at 260 nm in quarz cuvettes with a diameter of 1 cm. An absorption A_{260}=1 corresponds to 50 µg/ml double-stranded DNA or 33 µg/ml single-stranded oligonucleotides. The purity of the DNA was determined by measuring the absorption of the sample at 280 nm. DNA is sufficiently pure if the ratio A_{260}/A_{280} is higher than 1.8.

3.1.4 Pulsed field gel electrophoresis (PFGE)

High molecular weight DNA (see section IV.2.3) was separated on an 0.8 % (w/v) agarose gel (1 × TBE buffer; see section IV.3.1.1) by horizontal electrophoresis. The gels were run for 21-24 h with pulse periods of 0.5-50 s. After staining in an ethidium bromide solution (10 mg/ml), the gels were photographed on UV-transilluminator.

3.1.5 DNA gel size markers

In order to determine the approximate size of the DNA fragments after gel electrophoresis, the following commercially available DNA size markers were used:

Mass Ruller™ DNA Ladder Mix (MBI Fermentas)		Gene Ruler™ DNA Ladder Mix (MBI Fermentas)		Lambda Ladder PFGE Marker (New England Biolabs)		Low Range PFGE Marker (New England Biolabs)	
Fragment	Size bp	Fragment	Size bp	Fragment	Size kb	Fragment	Size kb
1	10000	1	10000	1	1018.5	1	194.0
2	8000	2	8000	2	970.0	2	145.5
3	6000	3	6000	3	921.5	3	97.0
4	5000	4	5000	4	873.0	4	48.5
5	4000	5	4000	5	824.5	5	23.1
6	3000	6	3500	6	776.0	6	9.42
7	2500	7	3000	7	727.5	7	6.55
8	2000	8	2500	8	679.0	8	4.36
9	1500	9	2000	9	630.5	9	2.32
10	1031	10	1500	10	582.0	10	2.03
11	900	11	1200	11	533.5	11	0.56
12	800	12	1031	12	485.0	12	0.13
13	700	13	900	13	436.5	13	-
14	600	14	800	14	388.0	14	-
15	500	15	700	15	339.5	15	-
16	400	16	600	16	291.0	16	-
17	300	17	500	17	242.5	17	-
18	200	18	400	18	194.0	18	-
19	100	19	300	19	145.5	19	-
20	80	20	200	20	97.0	20	-
-	-	21	100	-	48.5	-	-

3.2 DNA restriction

3.2.1 Restriction of plasmid or genomic DNA with restriction endonucleases

The DNA was dissolved in dH_2O and mixed with 0.2 vol 10 × reaction buffer and ca. 5 U of restriction enzyme(s), so that the final volume of the sample was 20 µl. The mixture was incubated at 37 °C or 30 °C, depending of the specific requirements of the enzyme(s). Whereas plasmid and cosmid DNA was digested for one or two hours, respectively, digestion of genomic DNA was carried out overnight. The reaction was stopped with 0.2 vol Stop-Mix (see section IV.3.1.1). When required, inactivation of the restriction enzyme(s) was carried out by heating for 20 min at 65 °C or by phenol-chloroform extraction.

3.2.2 Restriction of high molecular weight DNA

LGT agarose blocks containing the high molecular weight DNA (see section IV.2.3) were transfered into a new Eppendorf tube. The blocks were preincubated 1 h at 50 °C in 1 ml 1 × restriction buffer subsequently. Restriction was carried out overnight at 37 °C in an 150 µl overall reaction mixture containing 1 × restriction buffer and 30 U restriction enzyme.

10 × reaction buffer:	
Tris-acetate	100 mM
K-acetate	500 mM
Mg-acetate	100 mM

3.3 Southern blot analysis

3.3.1 Vacuum southern blotting

A nylon membrane (Biodyne B) of appropriate size was shortly preincubated in dH_2O and then soaked for 10 min in 20 × SSC. Afterwards, DNA was transfered from an agarose gel to the membrane using a vacuum blotter (Amersham-Pharmacia) by applying a 50 mbar vacuum. DNA transfer from a pulsed-field gel required UV irradiation (for fragmentation of the high molecular weight DNA). The following solutions were applied on the surface of the agarose gel during the blotting procedure:

Depurinization solution	0.25 N HCl	8 min
Denaturation solution	0.5 N NaOH; 1.5 M NaCl	8 min
Neutralization solution	0.5 M Tris-HCl, pH 7.5; 1.5M NaCl	8 min
20 × SSC	0.3 M Na-citrate, pH 7.0; 3 M NaCl	45 min – 1 h

After DNA transfer, the nylon membrane was incubated for 1 min in 0.4 N NaOH and 1 min in 0.25 M Tris-HCl, pH 7.5 (for neutralization). The membrane was then shortly dried and the DNA was fixed on the membrane by exposure to UV light.

3.3.2 DNA labelling ("ECL^{TM} Kit", Amersham)

For labelling of DNA probes the "ECL^{TM}-Kit" (enhanced chemo_luminescence) was used. The binding of a DNA probe to the complementary sequence on the nylon membrane was detected by chemo luminescence. Positively charged horseradish peroxidase molecules were mixed with the negatively charged DNA probe. Addition of glutaraldehyd covalently links the horseradish peroxidase molecules with the DNA probes. Reduction of H_2O_2 by the peroxidase requires the oxidation of luminol which results in light emission, which can be detected by suitable light-sensitive films, e.g. the Hyperfilm ECL.

For labelling of the probe 0.1 µg DNA resuspended in 10 µl dH_2O was denatured for 10 min at 90 °C and cooled for 5 min on ice. Then 10 µl labelling reagent and 10 µl glutaraldehyd were added. The mixture was incubated for 10 min at 37 °C and then added to the hybridization reaction.

3.3.3 Hybridization and detection of the membrane

Hybridization of the membrane was carried out overnight at 42 °C in hybridization solution (10-15 ml), after the nylon membrane has been preincubated at 42°C in the hybridization solution for 1 h. On the next day, the membrane was washed twice for 20 min at 42 °C with wash solution I and two times for 10 min at RT with wash solution II. The membrane was placed on Whatman paper to remove the rest of the wash solution,

and then incubated for 1 min in 5-10 ml detection solution I and detection solution II provided with the kit and mixed immediately (1:1) before applycation to the membrane. The membrane was superficially dried on Whatman paper and packed in saran wrap avoiding air bubbles on the top surface of the membrane. Chemoluminescence was detected by exposure of the membrane to Hyperfilm ECL. The exposure time depended on the signal intensity.

Wash solution 1	0.5 × SSC; 0.4 % (w/v) SDS
Wash solution 2	2 × SSC

3.4 Polymerase chain reaction (PCR)

3.4.1 Standart PCR

For routine PCR-amplification *Taq* DNA polymerase kits of different suppliers (QIAGEN, Invitrogen) were used. The reaction was performed in a final reaction volume of 50 µl.

10 × DNA Taq polymerase buffer	5 µl
20 mM dNTP mix	1 µl
0.5 µg/µl primer 1	1 µl
0.5 µg/µl primer 2	1 µl
100 ng/µl template DNA or boiled cells	1 µl
Taq DNA polymerase	0.15 µl
25 mM $MgCl_2$	2 µl
dH_2O	38.85 µl

The PCR cycles were designed according to the annealing temperature of the individual primers and the length of the expected amplification product:

Initial denaturation	2 min, 95 °C

1. Denaturation	30 sec, 95 °C	25-30 cycles
2. Annealing	30 sec (42-72 °C)	
3. Elongation	30 sec-5 min, 72 °C	
4. Final elongation	2 min, 72 °C	

3.4.2 Long distance PCR (Expand Long Template PCR system, Roche)

When DNA fragments longer than 4 kb were amplified, or proof reading was required, the Expand Long Template PCR system was used, following the instructions of the manifacturer. For fragments longer than 10 kb high molecular weight genomic DNA was used as a template. For this purpose, a LGT agarose block prepared for PFGE was melt for 10 min at 65 °C in 100 µl dH_2O. 10 µl of this mixture was used as a template, and the reaction was performed as follows:

Initial denaturation	1 min, 90 °C	
1. Denaturation	10 sec, 90 °C	25-30 cycles
2. Annealing	30 sec, 52-68 °C	
3. Elongation	1-15 min, 68 °C	
Final elongation	1-15 min, 68 °C	

3.4.2. Inverse PCR (IPCR)

In order to amplify unknown DNA with the help of a small region of known DNA sequence inverse PCR was carried out. Amplification of unknown DNA flanking regions of a particular gene by IPCR requires previous digestion of genomic DNA and identification of the corresponding fragment which contains a part of the known sequence as well as unknown flanking sequences by Southern hybridization. The size of the required fragment should not exceed 1 – 3 kb to facilitate amplification by PCR. Genomic DNA was then digested with the chosen restriction enzymes. The sample was diluted and the restriction fragments were ligated in order to obtain circular DNA

fragments. PCR amplification was performed with primers complementary to the ones used for probe, using the circularized genomic DNA fragments as a template.

3.5 Cloning procedures

3.5.1 A/T cloning of PCR products using the pGEM-T® Easy vector system

This kit enables rapid cloning of PCR fragments without digestion by overhanging adenine nucleotides at their 3' ends into a linearized vector that contains overhanging 5' terminal thymidine residues. pGEM-T® Easy vector allows cloning of DNA fragments in a multiple cloning site which is flanked by T7 and SP6 RNA polymerase promoters, respectively. This vector expresses the α-peptide of the β-galactosidase, thus enabling "blue-white" screening of successful DNA integration. The ligation reaction was performed overnight at 4 °C or for 2 h at RT and was prepared as follows:

2 × T4 DNA ligation buffer	10 µl
pGEM-T® Easy vector	1 µl
PCR product	1-8 µl
T4 DNA ligase	1 µl
dH_2O to final volume of	20 µl

3.5.2 Cloning of DNA fragments digested with restriction enzymes

DNA fragments digested with restriction enzymes were cloned in vectors that were cut with appropriate enzymes ("sticky ends ligation"). For ligation, digested vector and insert were mixed in a ratio of 1:3. The ligation reaction was performed overnight at 4 °C or 16 °C, or for 2 h at RT and was prepared as follows:

Linearized vector	0.5-1 µl
Restriction enzyme-digested DNA fragment	1-17.5 µl
5 × T4 ligase buffer	2 µl
T4 ligase (New England BioLabs), 2 U/ µl	1 µl
dH_2O to final volume of	20 µl

When the vector and the insert were cut with only one restriction enzyme or for blunt end ligation, the ends of the linearized vector molecule were dephosphorylated before ligation using calf intestinal alkaline phosphatase in order to prevent religation of the vector. Removal of the 5'-phosphate residue was carried out by addition of 0.1 vol alkaline phosphatase buffer ($ZnCl_2$ 10 mM; $MgCl_2$ 10 mM; Tris-HCl (pH 8.3), 8 mM; $H_2O_{bidest.}$), 1U alkaline phosphatase and incubation for 1 h at 37 °C. The reaction was stopped either by heating for 10 min at 65 °C or by phenol-chloroform extraction followed by ethanol precipitation of the DNA.

3.5.3 Generation of blunt end DNA fragments

For blunt end generation, the following reaction mixture was incubated for 30 min at 25 °C. The reaction was stopped either by heating for 15 min at 65 °C or by phenol-chloroform exctraction.

dNTP mixture (0.5 mM of each nucleotide, dATP, dCTP, dGTP, dTTP)	4 µl
10 × Klenow buffer (Tris-HCl pH 7.2 50 mM; $MgSO_4$, 10 mM; DDT 0.1 mM; BSA, 50 mg/ml; dH_2O)	2 µl
DNA dissolved in dH_2O	12 µl
Klenow DNA polymerase (5 U/ µl)	2 µl

3.6 Transformation and conjugation of bacterial cells

3.6.1 Preparation of competent bacterial cells ($CaCl_2$ procedure) and heat-shock transformation

The bacteria were grown overnight and 1 ml of this culture were used for inoculation of a fresh 50 ml LB culture. The cells were then grown to an OD_{600} of 0.5 and incubated for 30 min on ice. The culture was centrifuged for 5 min at 2000 × g at 4 °C and the pellet was resuspended in 17.5 ml ice-cold 0.1 M $CaCl_2$ and incubated on ice for 30 min. After a second centrifugation, the cells were resuspended in 2.5 ml ice-cold 0.1 M $CaCl_2$ and either used immediately for transformation, or prepared for freezing at – 80 °C by addition of 50% (v/v) glycerol.
50 µl of competent cells were thawed on ice and ca. 100 ng of plasmid DNA was added for transformation. After incubation on ice for 30 min, the cells were incubated at 42 °C for 90 sec and then again placed immediately on ice for 5 min. 1 ml LB was added and the cells were incubated at 37 °C for 1 h, allowing the expression of the selection marker before the mixture was plated on selective LB agar.

3.6.2 Preparation of electrocompetent cells and electroporation

LB medium was inoculated with 500 µl of an overnight culture of the strain of interest and grown OD_{600} of 0.5. The cells were collected by centrifugation at for 10 min at 2000 × g at 4 °C. The pellet was left on ice for 30 min and then washed with 50 ml ice-cold dH_2O. After a second centrifugation step, at the same conditions the pellet was resuspended in 25 ml 10 % glycerol, centrifuged again and finally resuspended in 600 µl 10 % (v/v) glycerol. The cells were stored as 50 µl aliquots at -80 °C. For electroporation, one aliquot was thawed on ice and mixed with ca. 0.5 µg DNA. The mixture was applied into a “Gene pulser” cuvette with an electrode gap of 0.1 cm and incubated for 10 min on ice. The cells were electroporated using a Gene pulser transfection apparatus (BioRad) at the following conditions: 2.0 kV, 600 Ω, 25 µF. 1 ml LB medium was added to the culture, and the mixture was transfered into a new tube and incubated at 37 °C for 1 h before the cells were plated on selective agar plates.

3.6.3 Conjugation

For conjugation the *E. coli* strain Sm10λ*pir* was used as donor strain in order to transfer the plasmid of interest to the recipient strain. Overnight cultures of the donor and recipient strain were used for inoculation of fresh cultures (1 % v/v), which were grown to an OD_{600}=0.7. 100 µl of both cultures were then mixed on LB agar plate and grown overnight without selective pressure. Then, the cells were collected, properly diluted in sterile 0.9 % NaCl and plated on selective LB agar plates.

3.7 DNA sequence analysis (AmpliTaq® FS-BigDye Terminator, PE Applied Biosystems) and annotation

This method allows nucleotide sequence determination by the use of didesoxynucleotides (Sanger et al., 1977) which are each labelled with a different fluorescence dye and unmodified PCR primers are used. For a final reaction volume of 10 µl 4 µl "Premix" and 10 pmol primers were used. The concentration of the DNA template varied between 0.1-1 µg depending on the quality and the source (cosmid, plasmid, or PCR product). The sequencing reaction includes the following steps:

Initial denaturation	1 min, 96 °C	
1.Denaturation	30 sec, 96 °C	26-30 cycles
2. Annealing	15 sec, 50-60 °C	
3. Elongation	4 min, 60 °C	
Final elongation	4 min, 60 °C	

90 µl $H_2O_{bidest.}$ (Merck), 250 µl 100 % ethanol and 10 µl 3 M Na-acetate (pH 4.6) were added after the reaction and the mixture was transfered to a new tube and centrifuged for 15 min. The DNA pellet was washed with 70 % (v/v) ethanol, dried in a Speedvac concentrator, finally dissolved in 30 µl TSR ("Template Suppression" Reagent) and denatured for 2 min at 90 °C.

The large-scale sequence determination of plasmid and cosmid clones was performed at

Göttingen Genomics Laboratory, Georg-August-Universität Göttingen. For this purpose, small insert libraries (2-2.5 kb) were generated by mechanical shearing of DNA (Oefner et al., 1996). After end repair with T4 polymerase, the fragments were ligated into the prepared pTZ19R vector. Isolated plasmids were sequenced from both ends using the dye terminator chemistry and analyzed on ABI337 sequencers (Applied Biosystems). The Phrap software implemented in the Staden software package was used for assembly and editing of the sequence data (Staden et al., 2000).

3.8. Annotation of DNA sequences

Homology searches as well as searches for conserved protein domains were performed with the BLASTN, BLASTX and PSI- and PHI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) programs of the National Center for Biotechnology Information (NCBI) (Altschul et al., 1997). Putative ORFs were identified using Vector NTI® (InforMax) and the NCBI ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) software. Prediction of membrane-spanning regions of proteins was carried out with the TMpred software (http://www.ch.embnet.org/software/TMPRED_form.html). The sequences were submitted to the EMBL nucleotide sequence database.

4. Isolation of lipopolysaccharides (LPS) and proteins

4.1 Isolation of LPS

The cells from 1 ml overnight culture were collected by centrifugation and treated as described for the isolation of genomic DNA (see section IV.2.2; Grimberg et al., 1989). After incubation with proteinase K, 600 µl phenol were added to the clear solution, carefully mixed and centrifuged for 10 min. The supernatant was transfered to a new tube and 20-50 µl were used for electrophoresis.

4.2 Preparation of secreted proteins form culture supernatant

10 ml overnight culture were centrifuged for 5 min at 13000 rpm (Eppendorf C5415 centrifuge) and the supernatant was transfered to a new 15 ml tube. 1 g trichloroacetic acid (TCA) was added and mixed by vortexing until the TCA crystals were completely dissolved. After overnight incubation at 4 °C, the precipitated proteins were collected by centrifugation and dissolved in 100 µl 0.1 N NaOH. After heating for 10 min at 90 °C in loading buffer (see section IV.5.2) the samples could be analysed by PAGE as described above (see section IV.5.2).

5. Analysis of LPS and proteins by polyacrilamyde gel electrophoresis

5.1 Polyacrylamid gel electrophoresis (PAGE, Laemmli et al., 1970)

Protein samples or isolated LPS were separated and analysed by polyacrylamide gels. The size of the gels was 15 × 17.5 × 0.5 cm, and the electrophoresis was performed at RT, overnight at 30 V in electrophoresis buffer. The samples were loaded on the concentration gel after heating for 10 min at 90°C in 4 × loading buffer.

Separation gel (13 % (w/v))	
30 % Acrylamyde, 0.8 % Bisacrylamyde	13 ml
1.5 M Tris-HCl, pH 8.8	7.3 ml
dH_2O	9 ml
10 % SDS	300 µl
APS (100 mg/ml)	150 µl
TEMED	10 µl
Collecting gel (5 % (w/v))	
30 % Acrylamyde, 0.8 % Bisacrylamyde	2.7 ml
1.5 M Tris-HCl, pH 8.8	4 ml
dH_2O	9 ml
10 % SDS	160 µl

APS (100 mg/ml)	160 µl
TEMED	8 µl
10 × Electrophoresis buffer	
Tris	30 g
Glycin	144.4 g
10 % SDS	100 ml
dH_2O to a final volume of	1000 ml
4 × Loading buffer	
0.25 M Tris-HCl, pH 6.4	2.5 ml
SDS	0.25 g
EDTA	4 mg
86 % Glycerin (w/v)	4 ml
β-Mercaptoethanol	0.5 ml
Bromphenol blue	0.4 mg
dH_2O to a final volume of	10 ml

5.2 Staining of proteins in polyacrylamide gels with Comassie Brilliant Blue (Laemmli et al., 1970)

After electrophoresis the polyacrylamide gels were incubated for 2 h in 1 % (w/v) staining solution. Protein bands were visualized by destaining the gel in destaining solution for 2-6 h.

Staining solution	
Coomassie Brilliant Blue	1 g
Destaining solution	100 ml
Destaining solution	
Acetic acid	100 ml
dH_2O	500 ml
Methanol	400 ml

5.3 Staining of proteins or LPS in polyacrylamide gels with silver nitrate (Tsai et al., 1982)

After electrophoresis, the polyacrylamide gels were stained with $AgNO_3$. All used devices were carefully washed with 70 % ethanol. The gels were fixed overnight in 100 ml 1 × fixation solution. The next day, the solution was replaced by 100 ml 1 × periodate solution and the gels were incubated for 5 min. After that the gels were washed three times for 30 min with dH_2O. Then the gels were incubated for 10 min in silver nitrate solution and subsequently washed three times for 10 min with dH_2O. The gels were developed in developing solution preheated to 60 °C, and when the intensity was satisfying, the reaction was stopped by washing with 50 mM EDTA solution for 10 min.

2 × fixation solution	
Isopropanol	250 ml
Acetic Acid	70 ml
dH_2O to final volume of	500 ml
Periodate solution	
Na m-periodate ($NaIO_4$)	0.87 g
1 × Fixation solution	100 ml
Silver nitrate solution	
1 M NaOH	1.4 ml
NH_3 (33 %)	1 ml
dH_2O	70 ml
20 % $AgNO_3$	1.25 ml
Developing solution	
2.5 % Na_2CO_3	100 ml
40 % Formaldehyde	27 µl

6. Western blot analysis of proteins

For the preparation of crude cell extracts, 1 ml overnight culture was adjusted to an OD_{600}=1. The cells were harvested by centrifugation for 5 min at 13000 rpm (Eppendorf C5415 centrifuge) and the pellet was resuspended in 100 µl TE buffer and 25 µl 5 × Laemmli buffer. After heating the mixture for 10 min at 90 °C the cell debris was collected by centrifugation for 5 min. The supernatant was transferred to a new tube and used for PAGE. After PAGE, separated proteins were transfered to a nitrocellulose membrane (Schleicher & Schuell BA S85 reinforced, 0.45 µm). The transfer of the proteins was carried out between two graphite plates in western blotting apparatus, using 12 Whatman paper slices, soaked with Anode buffer I, II or Cathode buffer. The membrane was incubated for 15 min in Anode buffer II. The lower graphite plate (anode) was moistened with water and covered with 6 slices of Whatman paper soaked with Anode buffer I, followed by 12 slices of Whatman papers soaked with Anode buffer II. At the upper side of the whatman papers the nitrocellulose membrane was laid, followed by the polyacrylamide gel and 3 slices of Whatman paper soaked in Cathode buffer. The air bubbles were carefully removed before laying the second graphite plate at the top (cathode). The transfer was carried out applying an electric current of 0.8 mA/cm^2 for 1 h at RT.

Anode buffer I	0.3 M Tris, 20 % methanol
Anode buffer II	25 mM Tris, 20 % methanol
Cathode buffer	25 mM Tris, 40 mM ε-amino-n-caprone acid, 20 % methanol

7. Hybridization and detection of proteins by antibody reaction

After the transfer the proteins the membrane was incubated for 1 h at RT in TBST solution (0.05 M Tris-HCl, pH 7.5; 0.15 M NaCl; 0.05 % Tween 20) supplemented with 10 % fat-free dry milk. Then, the blot was incubated with antibody 1 for 1 h at RT. The concentration of antibody 1 depended on the titre of the serum which was diluted in TBST supplemented with 1 - 3 % dry milk. After washing the membrane with TBST three times for 5 min, the solution (diluted 1:100) containing the secondary antibody

(peroxidase-conjugated) was added and incubated for 1 h at RT. Finally, the membrane was washed three times for 10 min with TBST at RT. Signal detection was performed as described above (see section IV.3.3.3).

8. Phenotypic assays

8.1 Detection of microcin expression

M9 medium agar plates were prepared. 1 ml of an overnight culture of the *E. coli* strain DH5α was mixed with 100 ml M9 soft agar (0.75 %(w/v); see section III.6.5) media, pre-cooled to 42 °C. A thin layer of (0.5 mm) of the mixture was poured on the surface of the M9 agar plates. The bacterial strains to be tested for microcin expression were grown overnight on the plates. Microcin production was assessed by the occurrence of clear zones of growth inhibition of the indicator strain DH5α around the colonies of the tested strains after overnight incubation at 37 °C.

8.2 Detection of aerobactin expression

Aerobactin media plates were prepared. 1 ml of an overnight culture (YT, Tc) of the indicator *E. coli* strain EN99 was mixed with 100 ml aerobactin soft agar (0.75 %(w/v), see section III.6.4) medium, pre-cooled to 42 °C. A thin layer (0.5 mm) of the mixture was poured on the surface of the aerobactin plates. The bacterial strains to be tested were grown overnight in 1 ml LB medium. Sterile susceptibility discs (Oxoid) were soaked with cells of the overnight culture and were placed on the aerobactin plates and incubated overnight at 37 °C. Aerobactin production was assessed by the presence of growth zones of the iron-deficient indicator strain EN99 around the colonies of the tested strains.

8.3 Detection of type 1 fimbrial expression

Overnight cultures of the strains to be tested and of a positive (*E. coli* strain Nissle 1917) and of a negative (*E. coli* strain AAEC189) control were grown. The mannose-dependent yeast agglutination assay was carried out by mixing 10 µl of the different bacterial overnight cultures with 10 µl yeast cells-suspension (1 mg/ml *Saccharomyces cerevisiae*

cells diluted in 0.9 % (w/v) NaCl, with or without 2 % (w/v) mannose) on microscope slides (75:25:1 mm). The slides were kept on ice until the aggregation of bacterial and yeast cells was observed in absence of mannose.

8.4 Detection of F1C fimbrial expression

Overnight cultures of the strains to be tested and of a positive (*E. coli* strain Nissle 1917) and of a negative (*E. coli* strain AAEC189) control were grown. For the immunoagglutination assay a polyclonal α-F1C fimbriae rabbit antibody was used (provided by S. Kahn, Wuerzburg), that was diluted 1:1000 in 1 × PBS. The immunoagglutination assay was carried out by mixing 10 µl of the bacterial overnight culture with 10 µl of the α-F1C fimbriae antibody solution on microscope slides (75:25:1 mm) and incubation on ice until the aggregation of the bacterial cells was clearly observed.

1 × PBS	
NaCl	8g
KCl	0.2 g
Na_2HPO_4	1.4 g
K_2HPO_4	0.24 g
NaOH	add to pH 7.4
dH_2O to a final volume of	1000 ml

8.5 Serum resistance assay

Serum resistance of *E. coli* strains was usually analyzed by incubating the bacteria in 90 % human serum. An overnight culture of the bacteria was diluted 1:100 in LB and grown to 90 Klett units. The bacteria were diluted 1:10 in human serum and incubated at 37 °C. After 0 h, 1 h, 3 h, and 24 h, survival of the strains was tested by plating an aliquot on LB agar plates containing the appropriate antibiotic (Hughes, et al., 1982). In addition, to determine the serum resistance of strain Nissle 1917 after complementation with a single chromosomal copy of the wzy_{536} gene or with a plasmid-encoded entire $wb*_{536}$ gene

cluster, growth of bacterial strains in 50 % human serum was also measured. Reactivity of the serum was routinely checked by incubation of the strains in heat-inactivated serum (56 °C, 30 min). After heat-inactivation, serum-sensitive strains multiplied in the heat-inactivated serum but not in untreated serum.

8.6 Detection of protease activity

For the detection of serine-protease activity, the EnzCheck® Protease Assay Kit E-6638 (Molecular Probes) was used following the instructions of the manifacturer. Briefly, casein derivatives, heavily labelled with pH-sensitive green-fluorescent BODI®PY FL dye, were used as a substrate. Protease-catalyzed hydrolysis releases highly fluorescent BODI®PY FL peptides, and the corresponding fluorescence proportional to the protease activity was measured using the fluorescence imager "Typhoon 8600". 5 ml LB culture were inoculated with 50 µl of an overnight culture of the strain to be tested and then grown to OD_{600}=0.6, the cells were collected by centrifugation for 5 min and diluted 1:100 in LB. 95 µl from the diluted cells were used for performing the reaction in a microtiter plate. 5 µl of the substrate solution (10 µg/ml BODI®PY FL in the supplied 1 × digestion buffer) were added and the mixture was incubated for 1 h at RT protected from light. The fluorescence was measured at wavelength of 500 nm.

V. Results

1. Analysis of the genome structure of *E. coli* strain Nissle 1917

1.1 Characterization of the *E. coli* strain Nissle 1917 genome by PFGE

In order to perform a genome-wide comparison between the three investigated *E. coli* O6 strains, a genomic restriction analysis by PFGE was used. As shown in Fig. 6a the *E. coli* strains Nissle 1917 and CFT073 exhibited similar genomic restriction patterns when digested with one of the restriction enzymes *Not*I, *Xba*I and *Sfi*I. This indicates a common clonal origin of these strains. Nevertheless, some differences in the restriction pattern were evident, which is not surprising, since both strains differ in the presence of at least one genomic island (*pheU*-associated GEI_{CFT073}) and having in mind that also minor genetic events like single nucleotide differences or gene rearrangements and deletions may significantly alter the genomic restriction pattern. The restriction pattern of strain 536 differed more significantly from that of strain Nissle 1917 than the one of strain CFT073.

In addition, the enzymes *Ceu*I and *Avr*II were used to compare the genomic restriction pattern of the three *E. coli* O6 isolates. As seen form Fig. 6b the *Avr*II-specific restriction patterns of the investigated strains differed significantly, but no considerable differences were observed between the strains when digested *Ceu*I and *Avr*II together or with *Avr*II alone. The restriction enzyme *Ceu*I cuts the *E. coli* chromosome specifically in the ribosomal RNA operons (7 restriction sites), whereas *Avr*II cuts 16 times in the *E. coli* strain MG1655 genome and 30 times in the genome of *E. coli* strain CFT073. At least two Nissle 1917-specific aditional *Avr*II restriction sites were identified in the sequence of GEI $II_{Nissle\ 1917}$ (see section V.2.2).

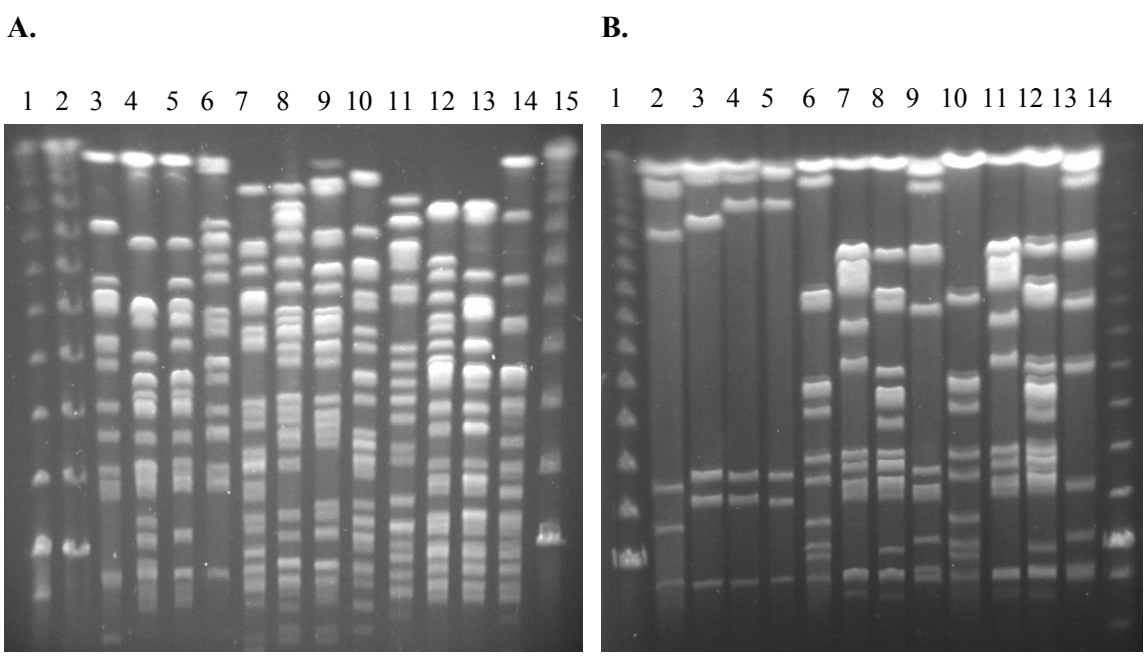

Fig. 6. Restriction analysis of genomic DNA of different *E. coli* strains by PFGE. A. lane 1, Low Range PFGE Marker; lane 2, Lambda Ladder PFGE Marker; lane 3, *Not*I-digested DNA of *E. coli* strain MG1655; lane 4, *Not*I-digested DNA of *E. coli* strain Nissle 1917; lane 5, *Not*I-digested DNA of *E. coli* strain CFT073; lane 6, *Not*I-digested DNA of *E. coli* strain 536; lane 7, *Xba*I-digested DNA of *E. coli* strain MG1655; lane 8, *Xba*I-digested DNA of *E. coli* strain Nissle 1917; lane 9, *Xba*I-digested DNA of *E. coli* strain CFT073; lane 10, *Xba*I-digested DNA of *E. coli* strain 536; lane 11, *Sfi*I-digested DNA of *E. coli* strain MG1655; lane 12, *Sfi*I-digested DNA of *E. coli* strain Nissle 1917; lane 13, *Sfi*I-digested DNA of *E. coli* strain CFT073; lane 14, *Sfi*I-digested DNA of *E. coli* strain 536; lane 15, Lambda Ladder PFGE Marker. B. lane 1, Lambda Ladder PFGE Marker; lane 2, *Ceu*I-digested DNA of *E. coli* strain MG1655; lane 3, *Ceu*I-digested DNA of *E. coli* strain Nissle 1917; lane 4, *Ceu*I-digested DNA of *E. coli* strain CFT073; lane 5, *Ceu*I-digested DNA of *E. coli* strain 536; lane 6, *Ceu*I/*Avr*II-digested DNA of *E. coli* strain MG1655; lane 7, *Ceu*I/*Avr*II-digested DNA of *E. coli* strain Nissle 1917; lane 8, *Ceu*I/*Avr*II-digested DNA of *E. coli* strain CFT073; lane 9, *Ceu*I/*Avr*II-digested DNA of *E. coli* strain 536; lane 10, *Avr*II-digested DNA of *E. coli* strain MG1655; lane 11, *Avr*II-digested DNA of *E. coli* strain Nissle 1917; lane 12, *Avr*II-digested DNA of *E. coli* strain CFT073; lane 13, *Avr*II-digested DNA of *E. coli* strain 536; lane 14, Low Range PFGE Marker.

1.2 Genome-wide tRNA screening

Since tRNA genes are common sites for chromosomal integration of foreign DNA elements, a PCR-based screening for sequence context alterations of tRNA genes of the *E. coli* strains Nissle 1917 and 536 was carried out, using primers which are binding within the ORFs located up- and downstream of tRNA genes in the genome of the *E. coli* MG1655. Usually monocistronic tRNA genes as well as the most promoter distal tRNA gene of polycistronic operons have been included into this study. tRNA-genes located within tRNA-encoding operons have not been included into the tRNA-screening (Dobrindt, U., unpublished). In several cases in which no PCR product was obtained with *E. coli* K-12-specific primers, primer pairs were used which allow amplification of the *E. coli* strain CFT073- or EDL933-specific sequence

context of the specific tRNA gene of interest. For some of the tRNA genes of interest the corresponding upstream genes were amplified and used as probes for dot blot screening of the genomic cosmid library of strain Nissle 1917. In several cases in order to obtain sequence information of unknown DNA regions flanking tRNA gene of interest, inverse PCR (IPCR) was carried out. The results from the tRNA screening approach are summarized in Tab. 5. Generally, an unexpectedly high degree of similarity of the genome organization of *E. coli* strain Nissle 1917 with that of the uropathogenic strain CFT073 and only to a lesser extent to that of the uropathogenic strain 536 was observed.

Table 5: Identification of chromosomally inserted putative horizontally acquired DNA regions by a PCR-based tRNA screening of the *E. coli* Nissle 1917 and 536 genome.

Strain *tRNA gene*	**MG1655** (K-12 primers)	**536** (K-12 primers)	**Nissle 1917** (K-12 primers)	**Upstream located gene**	**Downstream located gene (Nissle 1917)**	**Cloning strategy (Nissle 1917)**
argQ	+	-	-	*csrA*	*c3247*	IPCR
glnX	+	+	+			
lysQ	+	(+)	(+)	*ybgF*		IPCR
thrT	+	+	+			
glyU	+	-	-	*b2865*	*c3442*	PCR
serU	+	-	-	*b1976*	*c2416*	PCR
asnT	+	-	-	*b1976*	HPI	pCos220
asnW	+	(+)	(+)			
asnU	+	+	+			
argW	+	-	-	*yfdC*	*c2893*	pCos1YA7
argU	+	-	-	*fimZ*		
wrbA	+	+	+			
ssrA	+	+	+			
serW	+	(+)	+			
serT	+	+	+			
lysV	+	(+)smaller	(+)smaller	*gltX*		IPCR
proK	+	(+)smaller	(+)smaller			
proL	+	-	-	*yejM*		
pheV	+	-	-	*yqgA*	*Z4313*	PCR
pheU	+	+	+			
metY	+	+	+			
leuU	+	+	+			
asnV	+	-	-	*nac*	*c2445*	PCR
ileY	+	+	+			
proM	+	+	+			
metV	+	-	-	*purU*	*c3385*	PCR
leuZ	+	+	+			
leuV	+	+	+			
glyY	+	+	+			
valW	+	+	+			
tyrV	+	+	+			
serW	+	(+)	(+)			
alaX	+	+	+			
serX	+	-	-	*ycdW, csg*	*c1291*	PCR
leuX	+	-	-	*yjgB*	*c5371*	IPCR
thrW	+	-	-	*proA*	*c0391*	IPCR
aspV	+	-	+	*dnaQ*	*c0253*	PCR
selC	+	-	-	*yicK*	ΔP4 integrase	pCos8YA1

Results from the tRNA screening indicate that 15 of 37 studied tRNA genes serve as integration sites of foreign DNA in the genome of *E. coli* strain Nissle 1917.
Interestingly, the strains Nissle 1917 and 536 exhibited almost an identical "tRNA-screening pattern", indicating that the same tRNA genes in both strains are target sites for chromosomal insertion of horizontally acquired DNA. However, the genetic structure and the size of the integrated fragments differ dramatically, e. g. in the case of the *selC, leuX, thrW, serX,* and *pheV* tRNA genes (Dobrindt et al., 2002). Most of the regions located immediately downstream of the tRNA genes in the genome of Nissle 1917 which are not identical to those of strain MG1655 seem to be identical to those in the *E. coli* strain CFT073, which indicates a possibly similar DNA content and organization of the horizontally acquired DNA in the latter two strains.
A 100 kb *aspV*-associated genomic island, containing a determinant coding for a putative member of the RTX toxin family (*upxBDA*) is present in the genome of CFT073 (Welch et al., 2002). Using primers derived from the E. coli strain CFT073 sequence, a CFT073-specific *aspV* PCR product of about 10 kb was detected only in the strains CFT073 and Nissle 1917, but not in strain 536, for which an *E. coli* K-12-specific PCR product was detected (Tab. 5). In addition, the *upxBDA* gene cluster was successfully amplified (using the primers upx-up, upx-up1, upx-down and upx-down1) from genomic DNA of the strains CFT073 and Nissle 1917, but not from genomic DNA of strain 536. This is an indication that the *aspV*-associated genomic island of *E. coli* strain CFT073 is at least partially present in the genome of strain Nissle 1917, but absent from the genome of strain 536. Interestingly, the strains Nissle 1917 and 536 do not posses *pheU*-associated island (Tab. 5), whereas a *pheU*-associated island is present in the genome of strain CFT073 which carries a *pap* fimbrial determinant. No PCR product for the 536-specific *argW* tRNA-encoding gene was obtained when either *E. coli* K-12- or strain CFT073-specific primers were used.

2. Characterization of genomic islands of *E. coli* strain Nissle 1917 (GEI I – $IV_{Nissle\ 1917}$)

In this work four genomic islands have been detected and characterized in the genome of *E. coli* strain Nissle 1917. These islands are located at tRNA-encoding

genes (*serX, pheV, argW , asnT*; Fig. 7), and contain determinants coding for known fitness-conferring factors, as well as putative fitness factors, mobility genes, ORFs of unknown or putative functions. In addition several genomic islets, most of which localized in the core chromosome, coding for (putative) fitness-conferring factors have been partially characterized.

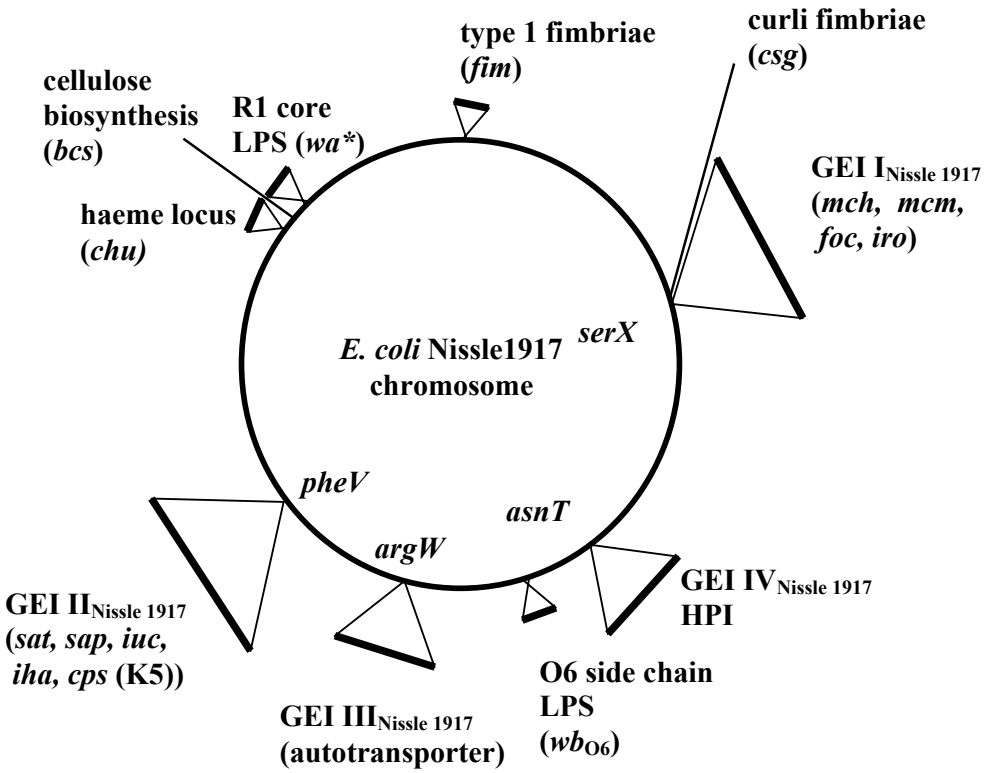

Fig. 7: Chromosomal localization and gene content of *E. coli* Nissle 1917-specific genomic islands and islets.

2.1 Characterization of the Genomic Island I of *E. coli* strain Nissle 1917 (GEI $I_{Nissle\ 1917}$)

In order to identify a cosmid clone containing the entire microcin-encoding genetic determinants of Nissle 1917, the cosmid genomic library was screened by printing the clones on microcin detection plates. The microcin expressing cosmid clone pCos3YE4 was identified (Fig. 8) and further characterized.

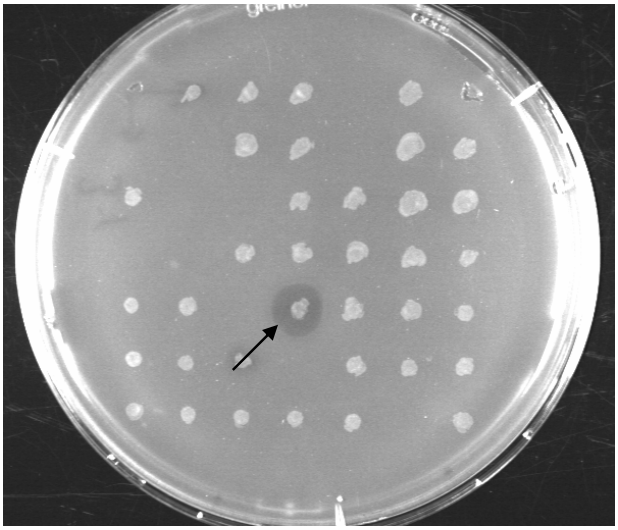

Fig. 8: Identification of the microcin-producing cosmid clone pCos3YE4 (indicated by an arrow).

Using PCR and Southern hybridization analysis, it was demonstrated that the pCos3YE4 cosmid insert contains both *mch* and *mcm* microcin-encoding gene clusters (Patzer et al., 2003), the *foc* operon coding for F1C fimbriae synthesis, and the *iro* cluster encoding the salmochelin siderophore system (Hantke et al., 2003). The expression of F1C fimbriae by *E. coli* XL1-blue harbouring pCos3YE4 as well as by the fimbriae-negative *E. coli* strain AAEC189, when transformed with pCos3YE4, was demonstrated by immunoagglutination using an α-F1C polyclonal rabbit serum (Khan, S., personal communication).

The structure and DNA content of the pCos3YE4 insert resembled those of PAI III_{536} (Dobrindt et al., 2001; Fig. 9; Fig. 10), thus representing a possible genomic island. An overlapping clone (pCos2RA4) was identified, using primers specific for the 3'-end of the pCos3YE4 insert (3YE4-1 and 3YE4-2; Fig. 9; see section VIII, Tab.2). Another cosmid overlapping with the 5' end of pCos3YE4 was also identified and shown to contain the *serX* tRNA gene. Since the structure of this genomic island

(GEI $I_{Nissle\ 1917}$) was very similar to that of the *serX*-associated island in the genome of *E. coli* strain CFT073 (s*erX*-associated $GEI_{CFT073;}$ Welch et al., 2002), overlapping PCR reactions based on the *E. coli* strain CFT073 genome sequence were established, and used for amplification of the CFT073- and Nissle 1917-specific DNA regions between *serX* and *iroC* (ca. 30 kb; primers: 1A + 1B, 2A + 2B, 3A + 3B, 4A + 4B; Fig. 9). In the chromosome of *E. coli* strain CFT073 this region contains mostly putative ORFs coding for hypothetical proteins. The PCR products were subsequently sequenced and assembled with the available sequence data from the cosmid clones. The structure and DNA content of the entire 100 kb region located downstream of the *serX* tRNA-encoding gene is highly similar in both *E. coli* O6 strains. This genomic island contains several determinants which code for fitness-conferring traits (microcins, adhesion, iron uptake).

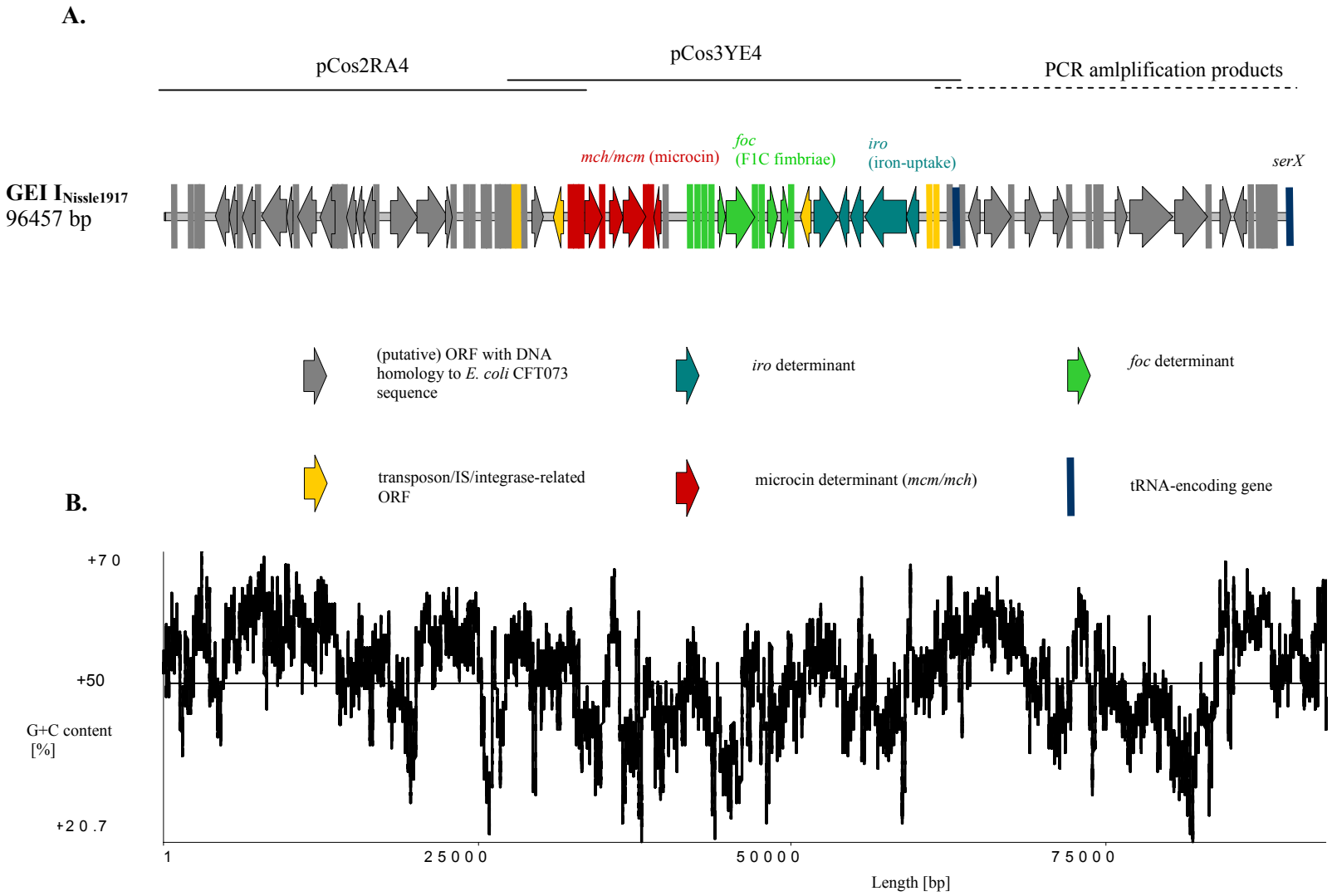

Fig. 9: Characterization of GEI $I_{Nissle\ 1917}$ A. Sequence analysis and structural organization of the ORFs located in GEI $I_{Nissle\ 1917}$. B. G+C content of the DNA in GEI $I_{Nissle\ 1917}$.

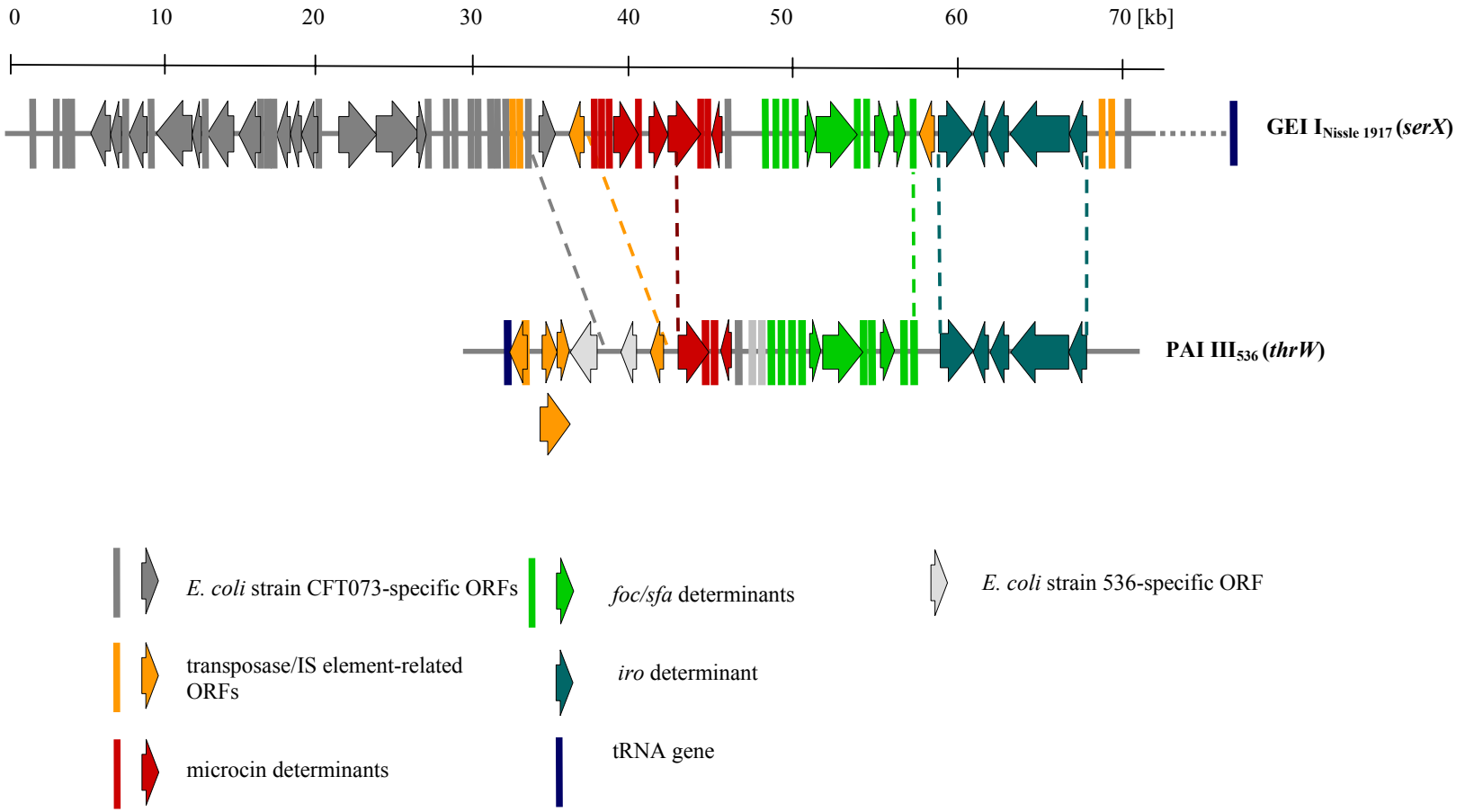

Fig. 10. Sequence alignment of GEI $I_{Nissle\ 1917}$ and PAI III_{536}. The homologous regions between the two islands are depicted by interrupted lines.

The GEI $I_{Nissle\ 1917}$ exhibits typical characteristics of genomic islands. It is associated with the *serX* tRNA gene, carries several mobility-related genetic elements as well as ORFs with unknown or putative function. The genetic organization of GEI $I_{Nissle\ 1917}$ is partially similar to that of PAI III_{536}, where also microcin-related ORFs (but not the entire *mch/mcm* gene cluster) are located together with the *sfa*I cluster (required for the biosynthesis of S-fimbriae) and the *iro* genes (Dobrindt et al., 2001; Schubert et al., 2003). Since *foc* and *sfa* code for members of one and the same fimbrial family and the structural similarity of both islands is high, it is tempting to consider them members of the "S-fimbrial PAI family " with possibly similar evolutional origin. Nevertheless, there are significant differences in the DNA content of GEI $I_{Nissle\ 1917}$ and PAI III_{536} as well as in their chromosomal integration site which are *serX* and *thrW,* respectively. The tRNA screening approach indicated that *thrW* is another putative integration site for foreign DNA in the genome of Nissle 1917 (see section V.1.2). In addition, the PCR-based screening of the 40 kb *E. coli* strain MG1655-specific region located downstream of *thrW* (*thrW – yagU*) was performed, since it is known that this DNA region represents the *E. coli* K-12 strains-specifc CP4-6 prophage and is absent in non-K-12 wild type commensal or pathogenic *E. coli* strains (Dobrindt et al., 2001). The results indicated that this region is absent in the chromosome of *E. coli* Nissle 1917 just like in the chromosome of 536.

2.2 Characterization of the Genomic island II of *E. coli* strain Nissle 1917 (GEI $II_{Nissle\ 1917}$)

The aerobactin gene cluster (*iuc*) is known to be located within different genomic islands on the chromosome of *E. coli, Salmonella* and *Shigella* strains as well as on plasmids (Purdy et al., 2001; McDougall et al., 1984; Moss et al., 1999; Vokes et al., 1999). In order to characterize the aerobactin-encoding gene cluster of *E. coli* strain Nissle 1917 and its chromosomal context, an aerobactin-expressing cosmid clone pCos9YB4 was identified (Fig. 11) by colony blot with a probe generated by PCR using primers iuc-up and iuc-down. Two overlapping cosmid clones pCos2YE4 and pCos2RF2 were identified by PCR, using the primers iha-up, iha-down, and K5-1-up, K5-1-down, respectively. The *iuc, sat, iha* and *kps* genes are located in the *pheV*-associated genomic island of the chromosome of *E. coli* strain CFT073. Since the genome structure of the *E. coli* strain Nissle 1917 is similar to that of the strain CFT073, the presence of the *iuc*, *iha, sat* genes and the genes comprising the three

regions of the K5-capsule-encoding *kps* cluster in the chromosome of the *E. coli* strain Nissle 1917 as well as in the identified cosmid clones was proven by Southern hybridization.

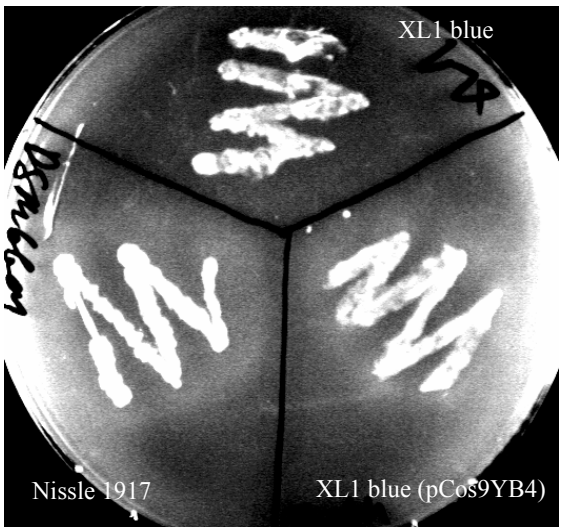

Fig. 11: Detection of aerobactin -xpression of cosmid clone pCos9YB4. The growth zone around the aerobactin-expressing strains is seen as white halos..

The three cosmid clones were sequenced and assembled as depicted in Fig. 12a, thus representing a genomic island of about 100 kb (designated as Genomic Island II of *E. coli* strain Nissle 1917, GEI $II_{Nissle\ 1917}$, see section VIII, Tab. 3). This genomic island contains the determinants coding for several fitness factors, as well as many putative ORFs with so far unknown functions and a considerable number of transposon-related features, IS elements and integrases (Fig. 12a). The G+C content varies considerably from that of the *E. coli* core chromosome (Fig. 12b), indicating a possible heterogeneous origin of the DNA regions assembled within the island. The organization and DNA content of GEI $II_{Nissle\ 1917}$ resembles those of the *pheV*-located pathogenicity island of *E. coli* strain CFT073 (*pheV*-associated GEI $_{CFT073}$; Fig. 13). Moreover, the tRNA screening approach revealed that in strain Nissle 1917 downstream of *pheV* the same putative P4 integrase-encoding gene is located (*c3556, Z4313*) as in the *E. coli* strains CFT073 and EDL933 strains. Since the *pheV* tRNA-encoding gene was not located on the identified cosmids, the *pheV* - *shiA* region of strain Nissle 1917 was amplified by PCR and sequenced using primers derived from the *E. coli* strain CFT073 genome sequence (pheVshiA-1 and pheVshiA-2). Interestingly, the ShiA protein encoded by the *Shigella flexneri* SHI-2 pathogenicity island has been shown to attenuate the host inflammatory response (Ingersoll et al.,

2003), but the importance of the *E. coli* strain Nissle 1917-specific *shiA*-homologue determinant remains so far unknown.

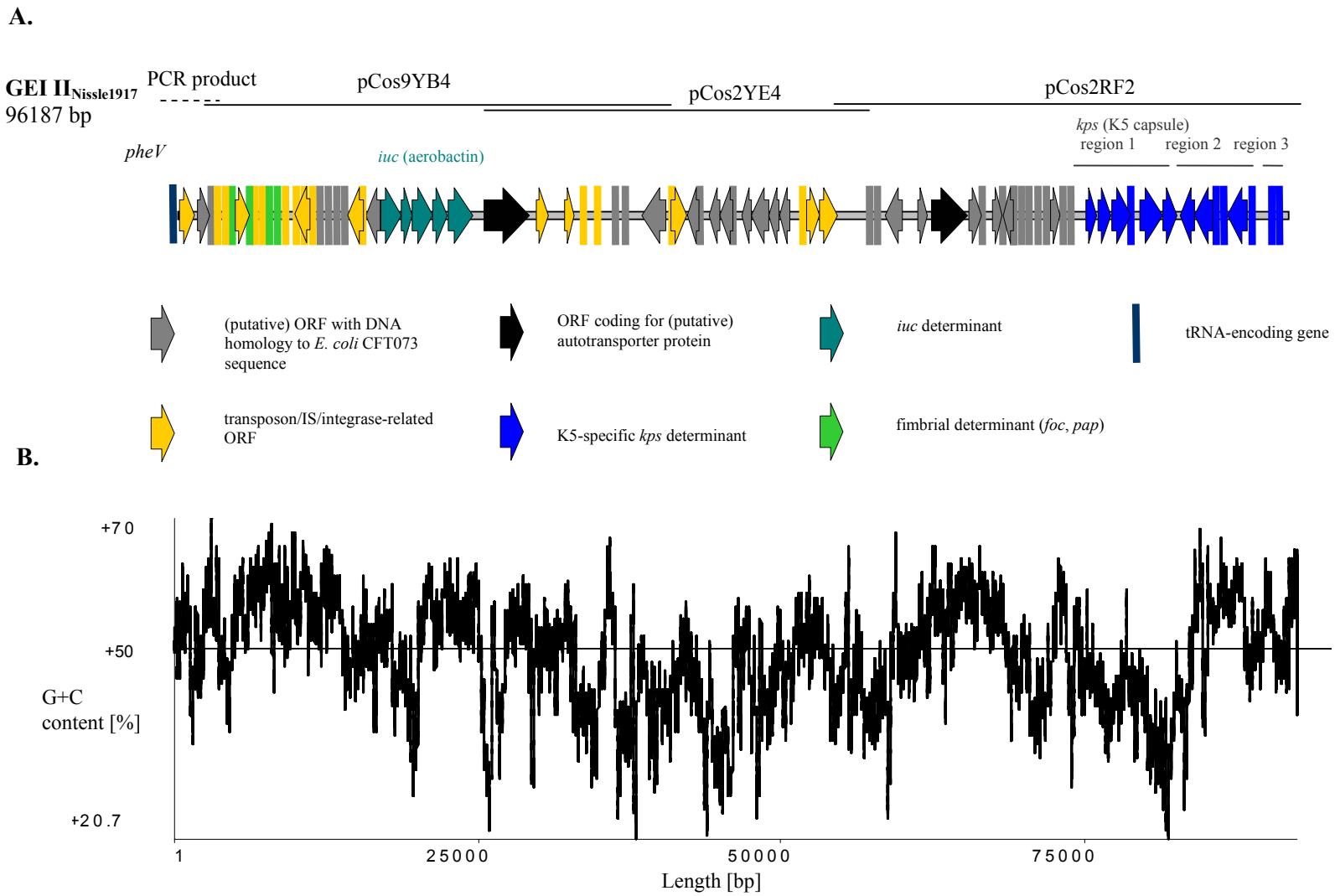

Fig. 12: Characterization of GEI $II_{Nissle\ 1917}$. A. Sequence analysis and structural organization of the ORFs located in GEI $II_{Nissle\ 1917}$. B. G+C content of the DNA in GEI $II_{Nissle\ 1917}$.

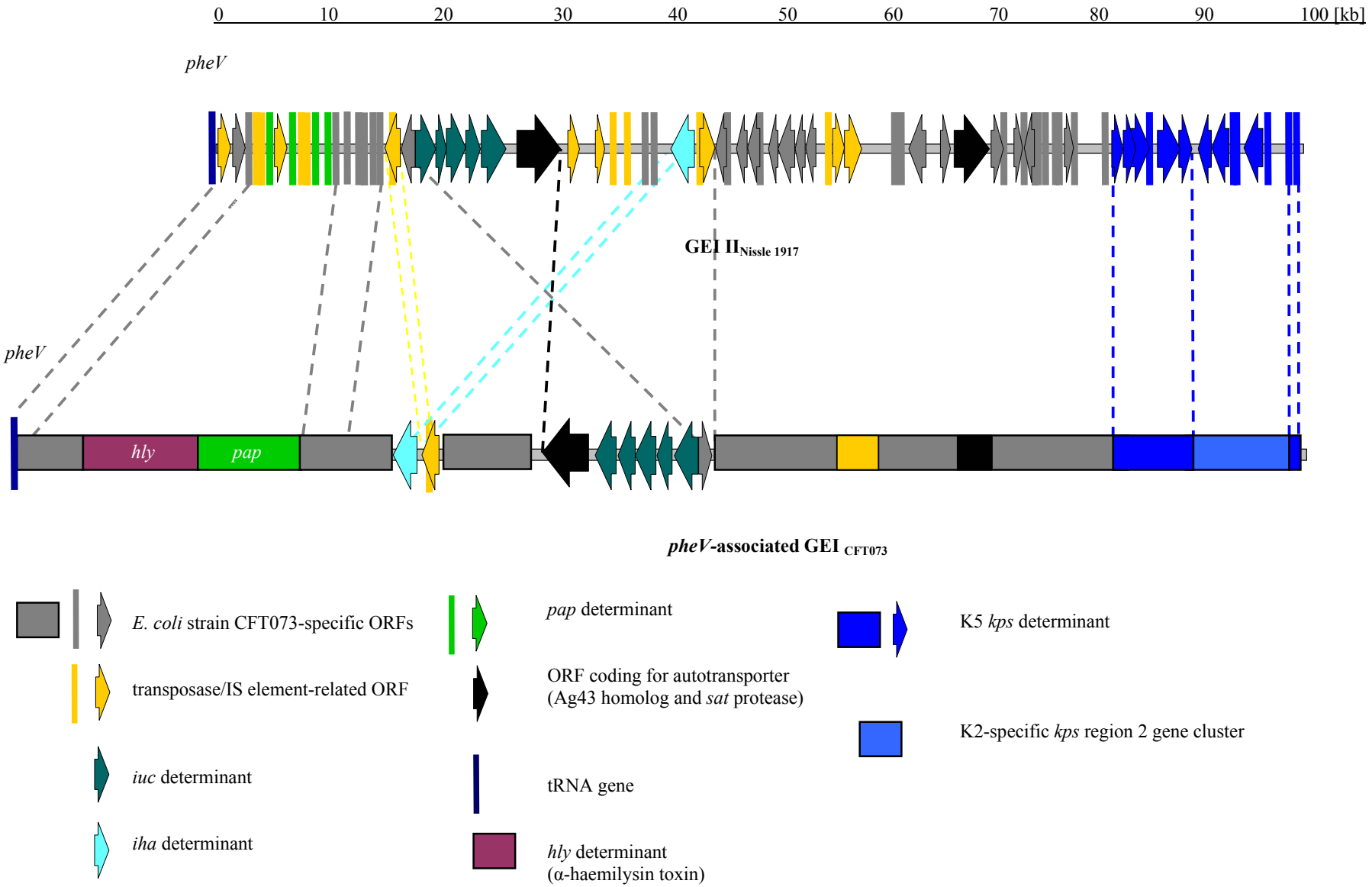

Fig. 13: Sequence alingment of GEI $II_{Nissle\ 1917}$ and *pheV*-associated GEI_{CFT073}. The homologous regions between the two islands are depicted by interrupted lines with the corresponding colour.

Two autotransporter protein-encoding genes are located on GEI $II_{Nissle\ 1917}$: *sat* and *sap*. The *sap* gene is an *agn43* homologue, whose gene product plays a possible role for cell aggregation and biofilm formation, whereas the *sat* gene is coding for a serine protease with cytopathic effect. In order to find out whether the Sat protein is secreted and active, the *E. coli* strain Nissle 1917-specific *sat* gene was cloned and overexpressed (Fig. 14) . For this purpose a 7 kb *Hind*III/*Eco*RI fragment from pCos9YB4 which carries the intact *sat* gene was cloned into *Hind*III/*Eco*RI digested pUC19 thus resulting in pPSD1917. The presence of the overexpressed Sat protein in the culture supernatant of host cells (Fig. 14a) was demonstrated as well as its protease activity (Fig. 14b). Protease activity of the *E. coli* strains Nissle 1917 and CFT073 was not detected probably due to weaker expression of Sat protein compare to the strain DH5α (pPSD1917) and low sensitivity of the assay.

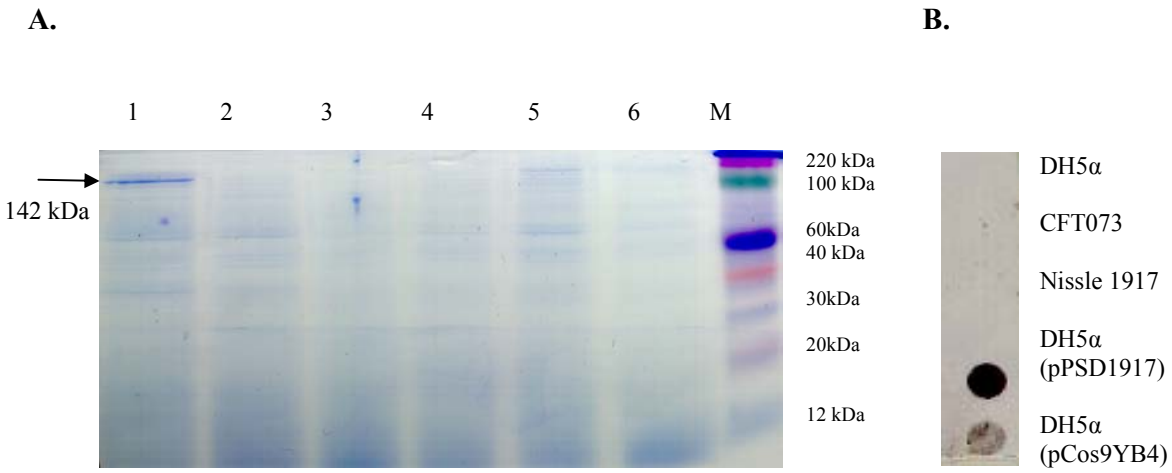

Fig. 14: Analysis of protease activity of the Sat protein encoded on GEI $II_{Nissle\ 1917}$. A. Detection of Sat protease in cluture supernatants. After separation of proteins by PAGE, the polyacrylamid gel was stained with Coomassie brilliant blue. Lane 1, DH5α (pPSD1917); lane 2, DH5α (pCos9YB4); lane 3, DH5α; lane 4, *E. coli* Nissle 1917; lane 5, *E. coli* 536; *E. coli* CFT073. The 142 kDa Sat proteine band is indicated by an arrow. B. Detection of protease activity by fluorescence.

Two regions with a transposon-like structure were identified in GEI $II_{Nissle\ 1917}$ (Fig. 15). One of them is ca. 30 kb in size, contains the *iuc* gene cluster as well as *sat* and *iha* genes, and is flanked by two IS*2* elements in opposite orientation. The second one (ca. 4 kb) exhibits contains a transposase-encoding gene and remnants of the *papX* gene flanked by IS*10* elements.

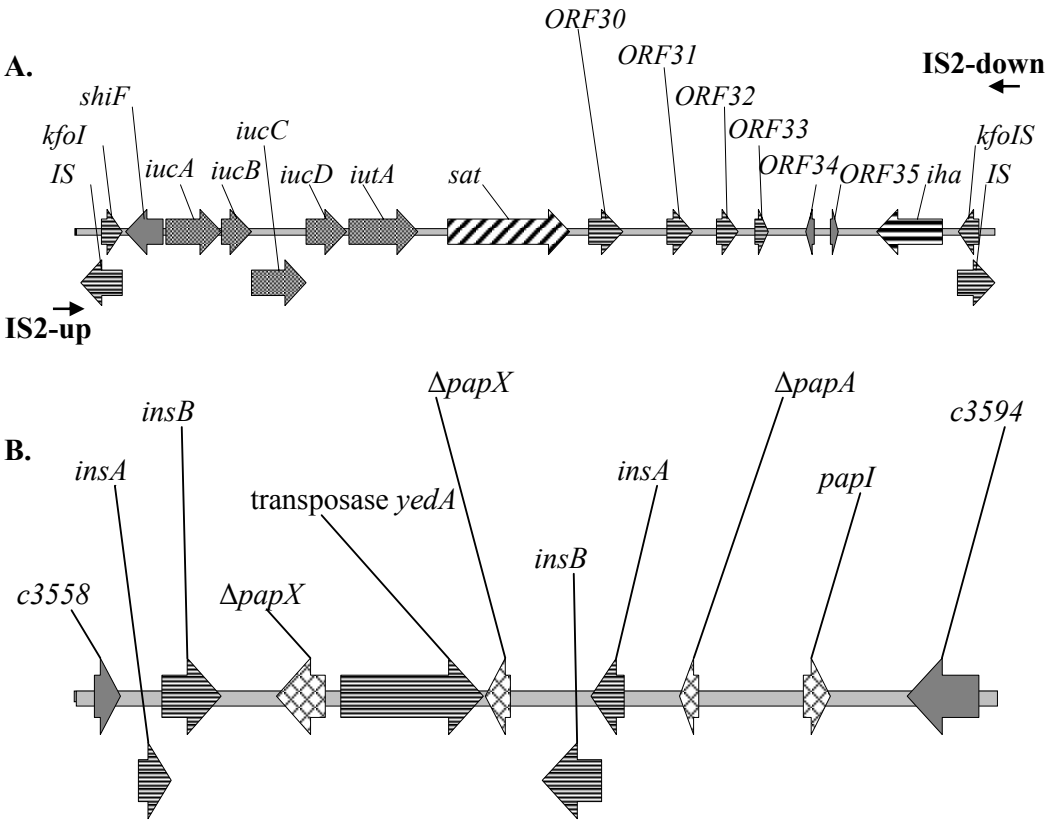

Fig. 15: Structure of the identified putative mobile genetic elements in GEI $II_{Nissle\ 1917}$. The primers used for detection of deletion events are indicated by arrows and designated with bold letters. A. 30 kb *iuc, sat, iha* region flanked by IS*2* elements. B. 4 kb transposon-like element containing remnants of the *pap* determinant.

For the 30 kb DNA region primers were designed which are located up- or downstream of the left-hand and right-hand IS*2* elements, respectively. This primer pair allows the detection of the deletion of the chromosomal region flanked by the IS*2* elements by PCR amplification of a 1.2 kb PCR product (exclusion PCR; Fig. 16).

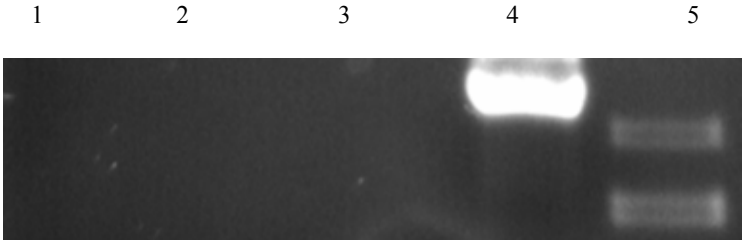

Fig. 16: Detection of the deletion event of 30 kb genetic element within GEI $II_{Nissle\ 1917}$ by exclusion PCR. Lane 1, *E. coli* straim MG1655; lane 2, *E. coli* strain 536; lane 3, *E.coli* strani CFT073; lane 4, *E. coli* strani Nissle 1917; lane 5, 1 kb DNA ladder.

The 1.2 kb PCR product could only be amplified from an overnight culture of strain Nissle 1917, whereas no PCR product was obtained from the strain CFT073 and 536 as well as from the negative control strain MG1655. This indicates that this DNA region can delete by recombination between the flanking IS*2* elements.

The obtained PCR product was directly sequenced. The sequence analysis revealed the presence of two inverted repeats (IRL; Lewis et al., 2001). Interestingly, in the *pheV*-located island of *E. coli* strain CFT073 only one copy of the IS*2* element is present, the order of the included genes differs from that of the 30 kb region of GEI $II_{Nissle\ 1917,}$ and the entire 4 kb transposon-like structure is absent but replaced by a 30 kb region containing the functional *hly* and *pap* determinants. Only remnants of the *pap* gene cluster are present in GEI $II_{Nissle\ 1917}$.

The loss of the 30 kb genetic element due to deletion event may influence the fitness of *E. coli* strain Nissle 1917, since it results in the simultaneous loss of the aerobactin iron-uptake system, the serine protease activity of the Sat protein and the putative adhesion/siderophore receptor activity of Iha.

2.3 Analysis of the sequence context of *argW* tRNA-encoding gene in the genome of *E. coli* strain Nissle 1917 (GEI $III_{Nissle\ 1917}$)

One of the tRNA genes which was found to be a putative integration site of foreign DNA was *argW*. A cosmid clone containing the *argW* tRNA gene was identified by dot blot hybridization with a probe representing the *yfdC* gene located upstream of *argW* . The presence of *argW* in this cosmid was proven by Southern blot hybridization and the cosmid insert was subsequently sequenced (Fig. 17; see section VIII, Tab. 4).

The identified putative ORFs exhibit high overall homology to *E. coli* CFT073 sequences, although the order of the ORFs differs in the *argW* upstream region significantly between the two strains. Interestingly, a large ORF (ORF26) whose gene product exhibits similarity on the protein level to the putative autotransporter YapH of *Y. pestis* seemed to be truncated due to the presence of an internal stop codon in strain CFT073 but not in strain Nissle 1917. ORF 12 – 17 represent a putative fimbriae encoding gene cluster. The *argW* downstream region represents a part of a genomic island (GEI $III_{Nissle\ 1917}$).

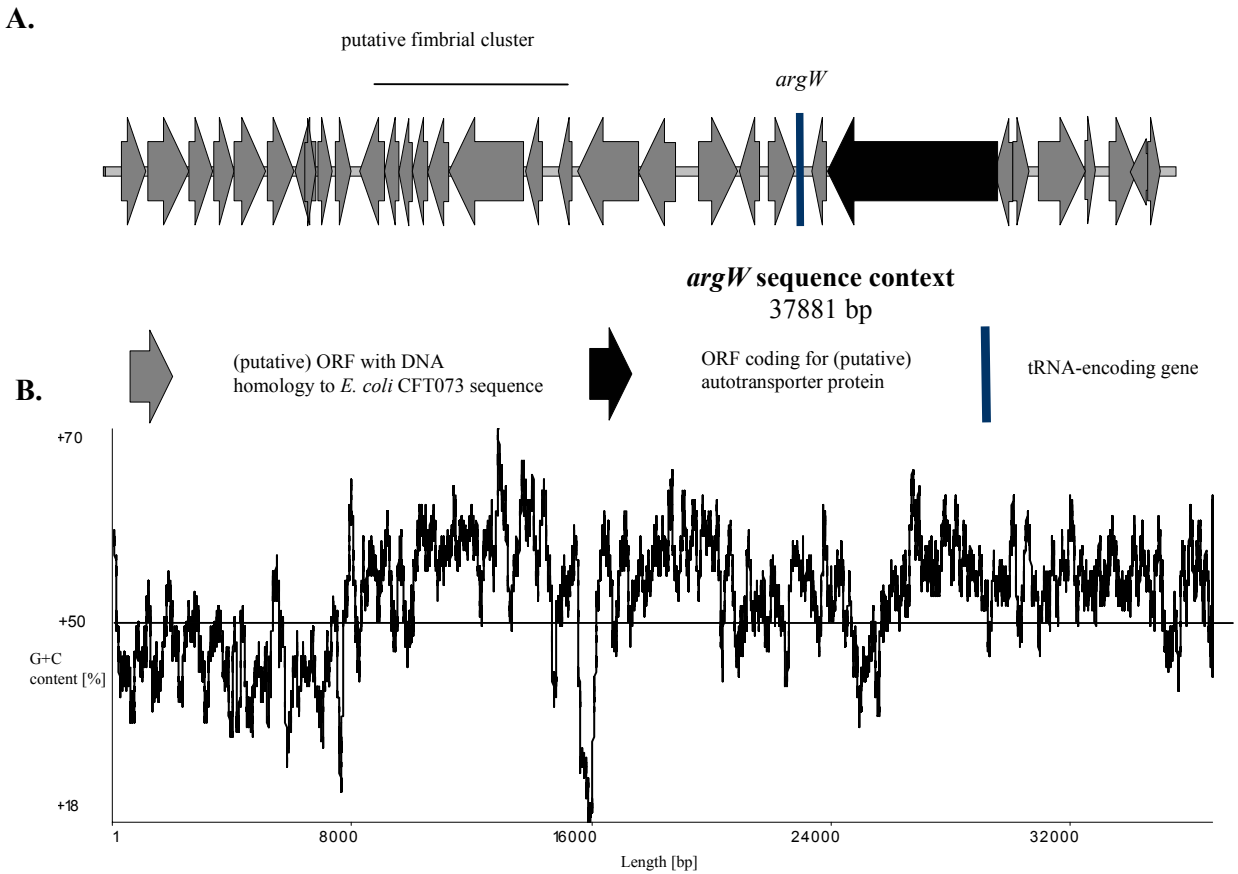

Fig. 17: Characterization of the *argW* sequence context in strain Nissle 1917. A. Genetic structure of the *argW* sequence context. B. G+C content of DNA in the *argW* sequence context in strain Nissle 1917.

2.4 Detection and sequence analysis of GEI $IV_{Nissle\ 1917}$

The presence and genetic organization of the DNA region which encodes the siderophore system yersiniabactin (Carniel et al., 1996) was verified in strain Nissle 1917 by sample sequencing of PCR products obtained with primers described before (Buchrieser et al., 1999; Karch et al., 1999). According to the results, we conclude that by analogy to the HPI of *Y. pseudotuberculosis*, the *asnT*-associated GEI $IV_{Nissle\ 1917}$ is about 30.2 kb in size and has a G + C content of 57 %. The left and right junction of GEI $IV_{Nissle\ 1917}$ have already been sequenced and flanking repeat structures have not been reported (Schubert et al., 1999). This island is also present in the same chromosomal insertion site in uropathogenic *E. coli* O6 strains CFT073 and 536 (Dobrindt et al., 2002; Welch et al., 2002) as well as in many non-human pathogenic *Salmonella* enterica subspecies III and VI and in commensal *E. coli* isolates (Dobrindt et al., 2003; Ölschäger et al., 2003).

3. Molecular analysis of genomic islets in *E. coli* strain Nissle 1917 chromosome

Several genomic islets which possibly contribute to the fitness of *E. coli* strain Nissle 1917 have been detected and characterized. These include: *wa** gene cluster coding for the enzymes required for the R1 type core LPS biosynthesis, *wb** gene cluster coding for the enzymes required for the O6 side chain LPS; determinants which are known to contribute to the rdar morphotype (*csg* and *bcs*); type 1-fimbriae-encoding gene cluster *fim* and the *chu* determinant, which is responsible for the biosynthesis of the haeme iron-uptake system. All of them (except *chu*) are part of the *E. coli* core genome.

3.1 Molecular analysis of the lipopolysaccharide biosynthesis determinants of *E. coli* strain Nissle 1917.

3.1.1 Identification and sequence analysis of the *wa** gene cluster of *E. coli* strain Nissle 1917.

In order to identify a cosmid clone containing the *wa** gene cluster encoding the enzymes for the R1 core of the LPS, the genomic cosmid library was screened by PCR using primers R1C3 and R1K15 (Amor et al., 2000). One cosmid was identified and the presence of the *wa** gene cluster was confirmed by Southern blot hybridization and sequence analysis of the ends of the cloned insert using the primers

SuperCos-F and SuperCos-R, as well as R1K15 and R1C3. The entire cosmid was sequenced and analyzed. The *E. coli* strain Nissle 1917-specific *wa** gene cluster exhibited 97 % homology on the DNA level to the already published R1 core type-encoding *wa** gene cluster of *E. coli* (Heinrichs et al., 1998) and a conserved localization on the *E. coli* chromosomal map upstream of *kdtB* (at 81 min on the *E. coli* K-12 chromosome; Fig. 18; see section VIII, Tab. 5). Since the 5' end of *waaC* was not included into the cosmid insert a PCR was carried out to amlplify the missing part using primers waaClücke-1 and waaClücke-2, that are based on the nucleotide sequence of the already published *wa** gene cluster. The obtained PCR product was subsequently sequenced.

Interestingly, at the 3' end of the insert a tRNA-encoding gene *selC* was identified. Downstream of the *selC* gene an ORF with homology to P4-like integrase was identified, indicating that a genomic island is located at this position in the chromosome of Nissle 1917 (see sections V.1.2 and VIII, Tab. 5). The highest overall DNA sequence homology of the cosmid insert is to the genome of the *E. coli* CFT073 (98 %). ORF32 - 39 are present in the genome of CFT073 but absent from the *E. coli* MG1655 chromosome.

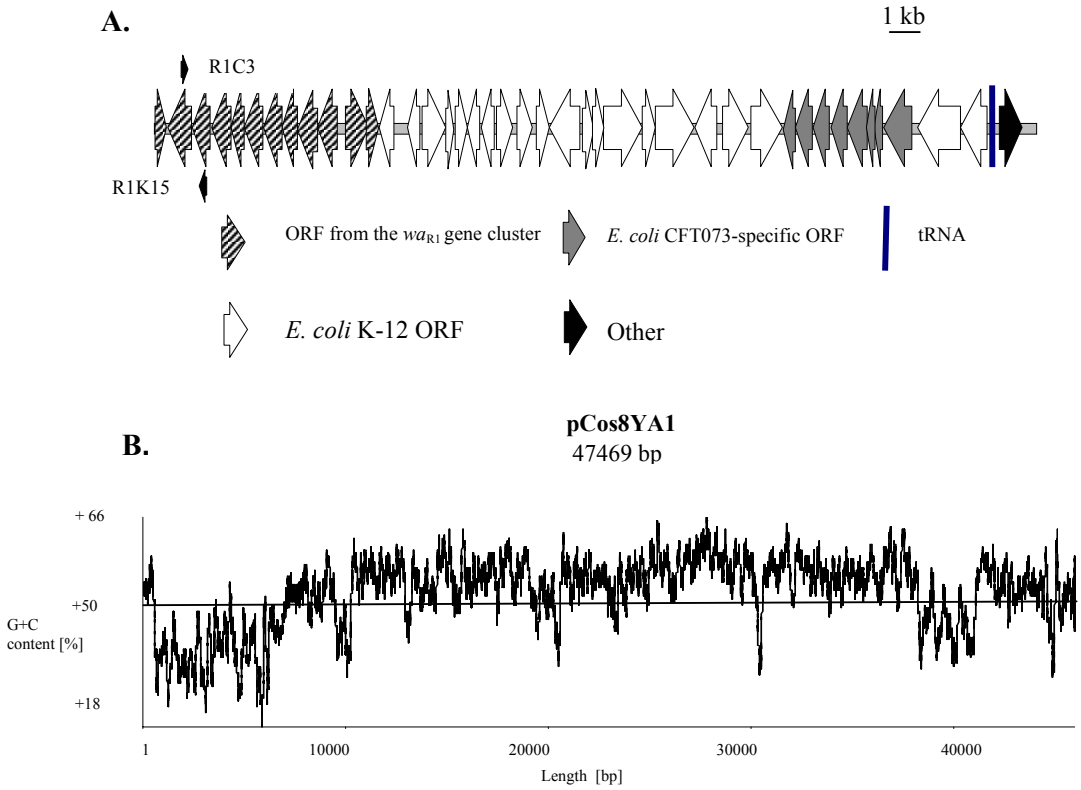

Fig. 18: Analysis of the *wa** determinant of the *E. coli* strain Nissle 1917. A. Genetic organization of the *wa** gene cluster and flanking regions of *E. coli* strain Nissle 1917. B. G+C content of the DNA in the insert.

3.1.2 Identification and sequence analysis of the wb_{O6} gene cluster of *E. coli* strain Nissle 1917.

Since *wb** gene clusters have a conserved localization in the *E. coli* chromosome between *galF* and *gnd* genes (45.4 min on the *E. coli* K-12 chromosome), a PCR amplification using proofreading DNA polymerase was carried out in order to identify wb_{O6} gene cluster. Amplification of the entire wb_{O6} gene cluster of *E. coli* strains Nissle 1917 and 536, using the already published primers 412 and 482 binding in the *wb** flanking regions (Coimbra et al., 1998; Fig. 19), resulted in both *E. coli* strains in a DNA fragment of ~11 kb (Fig. 19a).

The PCR products of both strains were cloned into the plasmid pGEM®-T Easy resulting in the plasmids pGWB1917 and pGWB536, respectively, which exhibited an identical restriction pattern after digestion with *Eco*RI (Fig. 19b).

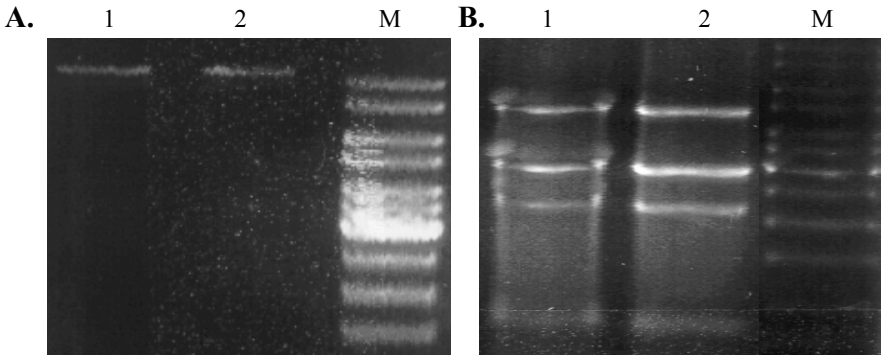

Fig. 19: A. Amplification of the wb_{O6} determinant of *E. coli* strains Nissle 1917 and 536: lane 1, *E. coli* Nissle 1917; lane 2, *E. coli* 536; M, 1 kb DNA ladder. B. Restriction pattern after *EcoR*I digestion of the cloned wb_{O6} gene clusters: lane 1, pGWB1917; lane 2, pGWB536; M, 1 kb DNA ladder.

The *E. coli* strain Nissle 1917-specific DNA fragment was sequenced and the genetic structure was analyzed in detail (Fig. 20). The wb_{O6} gene cluster of strain Nissle 1917 is 11037 bp in size and exhibits an overall G+C content of 36.4 % suggesting an acquisition of the wb_{O6} determinant by horizontal gene transfer. Nine tightly linked, sometimes overlapping putative open reading frames (ORFs) were identified (Fig. 20 and Tab. 3). The G+C content did not vary markedly between the different ORFs with the exception of the genes *manC* and *manB* which exhibited an even higher G+C content than the overall *E. coli* chromosome (50.8 %). The identical genetic organization of the wb_{O6} gene clusters of *E. coli* O6 strains Nissle 1917 and 536 was demonstrated by PCR, using the primers whose position within the wb_{O6} determinant is depicted in Fig. 20 (primers LG1 – LG8). The deduced amino acid sequences of these ORFs were analyzed with regard to the presence of conserved domains and similarity to other protein sequences. Based on the obtained results, the identified putative ORFs of the wb_{O6} determinant were predicted to encode putative glycosyl- or mannosyl transferases (ORFs 3, 4, 5 and 7), a putative O6 antigen flippase Wzx (ORF 1), a putative O antigen polymerase Wzy (ORF 2), a putative UDP-N-acetylglucosamine-4-epimerase or UDP-glucose-4-epimerase (ORF 6), a mannose-1-phosphate guanosyl transferase (ORF 8) and a phosphomannomutase (ORF 9) (Tab. 6). The nucleotide sequences of ORFs 1-7 showed no homology on the DNA level to available sequences from public databases.

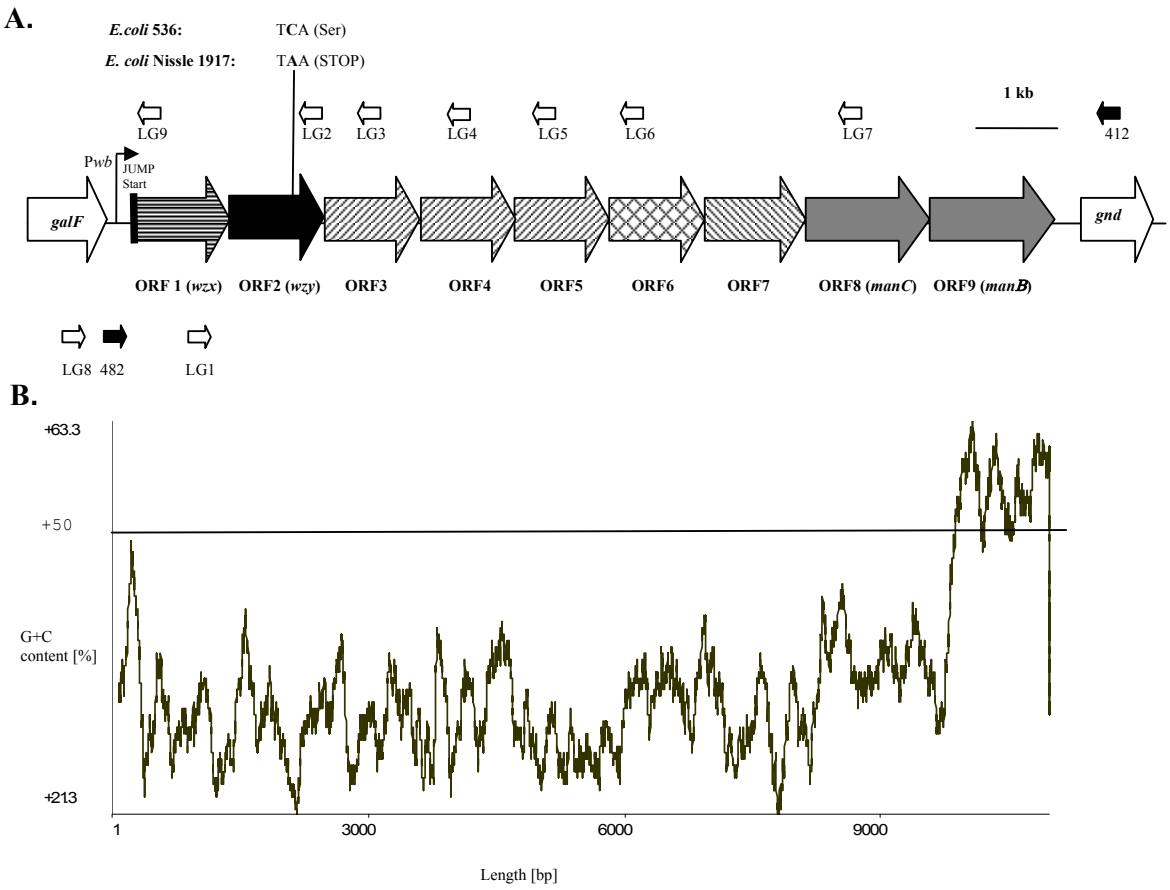

Fig. 20: Analysis of the wb_{O6} determinant of *E. coli* strain Nissle 1917 and 536. A. Genetic structure of the O6 side chain-encoding determinant of *E. coli* strains Nissle 1917 and 536. The identified ORFs and the binding sites of the used primers are marked by arrows. The position of the point mutation in *E. coli* strain Nissle 1917-specific O6 antigen polymerase-encoding gene $wzy_{Nissle\ 1917}$ is indicated by bold letters. B. G + C content of the DNA in *E. coli* Nissle 1917-specific wb_{O6} gene cluster.

Table 6: Characteristics of the ORFs located in the O6-specific *wb* gene cluster of *E. coli* strain Nissle 1917

Putative ORF	Length [bp]	G+C content [%]	No. of aa of encoded protein	Conserved domain (s) of the encoded protein	Similarity to already known proteins (accession no.)	% identity / % similarity (no. of aa)	Putative function of wb_{O6}-encoded protein
ORF1	1256	31.3	418	Polysaccharide biosynthesis proteins	Putative O antigen transporter RfbX protein, *Shigella dysenteriae* (S34963) O antigen transporter *E. coli* K-12 (I69652)	24 / 45 (396) 23 / 44 (415)	O antigen flippase Wzx
ORF2	986 [a] 1343 [b]	30.5	328 [a] 447 [b]	-	Hyp. protein, *Streptococcus pneumoniae* R6 (AAL00026) beta-1,3-glucan synthase GSC-1, *Pneumocystis carinii* (AAG02216)	25 / 48 (447) 24 / 40 (1944)	O antigen polymerase Wzy
ORF3	860	31.2	286	Glycosyl transferase family 2	Glycosyl transferase, *Bacillus halodurans* (BAB07432) putative beta 1,3-glucosyl transferase WaaV, *E. coli* F470 (AAC69672)	38 / 58 (303) 31 / 50 (327)	Glycosyl transferase
ORF4	1032	33.6	343	Glycosyl transferases group 1	Predicted glycosyl transferases, *Thermoanaerobacter tengcongensis* (AAM23571) glycosyl transferase, *Pyrococcus furiosus* DSM 3638 (AAL80431)	28 / 45 (404) 27 / 45 (336)	Glycosyl transferase
ORF5	1145	28	381	-	Glycosyltransferase, *Clostridium acetobutylicum* (AAK80991) putative mannosyltransferase, *Yersinia pestis* (CAC92344)	28 / 43 (393) 25 /39 (380)	Glycosyl transferase
ORF6	1007	35.6	335	NAD dependent epimerase/ dehydratase family	UDP-glucose 4-epimerase, GalE, *Haemophilus influenzae* Rd (AAC22012) UDP-galactose 4-epimerase, *H. influenzae* (CAA40568) UDP-N-acetyl-glucosamine 4-epimerase, *E. coli* O55 (AF461121)	54 / 71 (338) 54 /71 (338) 23 / 39 (331)	UDP-N-acetyl-glucosamine 4-epimerase or UDP-glucose 4-epimerase

Table 6 continued

ORF7	401	30.4	401	Glycosyl transferases group 1	WbaD (Orf7.17) function unknown, *Salmonella enterica* (AAB49389) Glycosyltransferase, *Clostridium acetobutylicum* (AAK80997)	51 / 67 (399) 26 / 47 (420)	Glycosyl transferase
ORF8	483	38.1	483	Mannose-6-phosphate isomerase family 2	Mannose-1-phosphate guanyltransferase, *E. coli* O157:H7, (BAB36277) mannose-1-phosphate guanyltransferase, *E. coli* K-12 (AAC75110)	71 / 84 (478) 71 /84 (478)	Mannose 1-P guanosyl transferase
ORF9	456	55.1	456	Phospho-glucomutase / phosphomannomutase, alpha/ beta/alpha domain I and alpha/beta/ alpha domain II	Phosphomannomutase E. coli O6 (AAG41759) Phosphomannomutase E. coli O41 (AAG41754)	99 / 99 (456) 98 / 98 (456)	Phosphomanno-mutase

[a)] Length of *wzy* or Wzy in semi-rough *E. coli* strain Nissle 1917
[b)] Length of *wzy* or Wzy in smooth *E. coli* 5

Comparison of of the O6-specific O antigen polymerase-encoding gene *wzy* of *E. coli* strains Nissle 1917 and 536

Due to its localization downstream of the putative *wzx* gene of the wb_{O6} determinant, ORF 2 was predicted to be the putative O6 specific O antigen polymerase-encoding gene *wzy*. This was corroborated by the fact that 12 transmembrane helices were predicted from the deduced amino acid sequence of ORF 2 using the TMpred software (data not shown). To find out whether the putative *wzy* genes of the serum-sensitive strain Nissle 1917 and its serum-resistant counterpart strain 536 differ from each other, the corresponding DNA sequences were compared. Interestingly, sequence comparison of the putative *E. coli* Nissle 1917- and 536-specific *wzy* genes demonstrated that in $wzy_{\text{Nissle 1917}}$ a point mutation (C to A transition at position +986 with respect to of the translational start of $wzy_{\text{Nissle 1917}}$) resulted in an internal stop codon (TCA to TAA) and consequently in truncation of the ORF in comparison to wzy_{536} (Fig. 20). This internal stop codon causes premature translation termination of $wzy_{\text{Nissle 1917}}$ transcripts, thus leading to a non-functional O6-specific O antigen polymerase. Therefore, the point mutation within $wzy_{\text{Nissle 1917}}$ is proposed to be the reason for the semi-rough phenotype of *E. coli* strain Nissle 1917. This supports the results of the biochemical analysis of this strain's LPS that consists of only one O repeating unit linked to the R1-type core (Grozdanov et al., 2002).

Determination of O6 side chain length and serum resistance of the wild type strain *E. coli* Nissle 1917 and of different *wzy*-complemented derivatives

To verify the proposed function of the putative *wzy* gene, and to prove that the identified point mutation within $wzy_{\text{Nissle1917}}$ is the reason for the semi-rough phenotype of strain Nissle 1917, complementation experiments were performed. Strain Nissle 1917 was transformed with the entire wb_{O6} gene cluster of *E. coli* strain 536 harbored on a plasmid (pBWB536). In addition, two fragments of the *E. coli* 536-specific wb_{O6} gene cluster were subcloned into pBluescript-II KS, using the *Not*I restriction enzyme, and transfered into strain Nissle 1917. One of these plasmids contained the O antigen flippase gene (*wzx*) and a fragment of wzy_{536} with the size of the truncated ORF $wzy_{\text{Nissle 1917}}$ (pBLG2504). The other one consisted of the complete *wzx* and *wzy* genes of *E. coli* strain 536 (pBLG2849; Fig. 21).

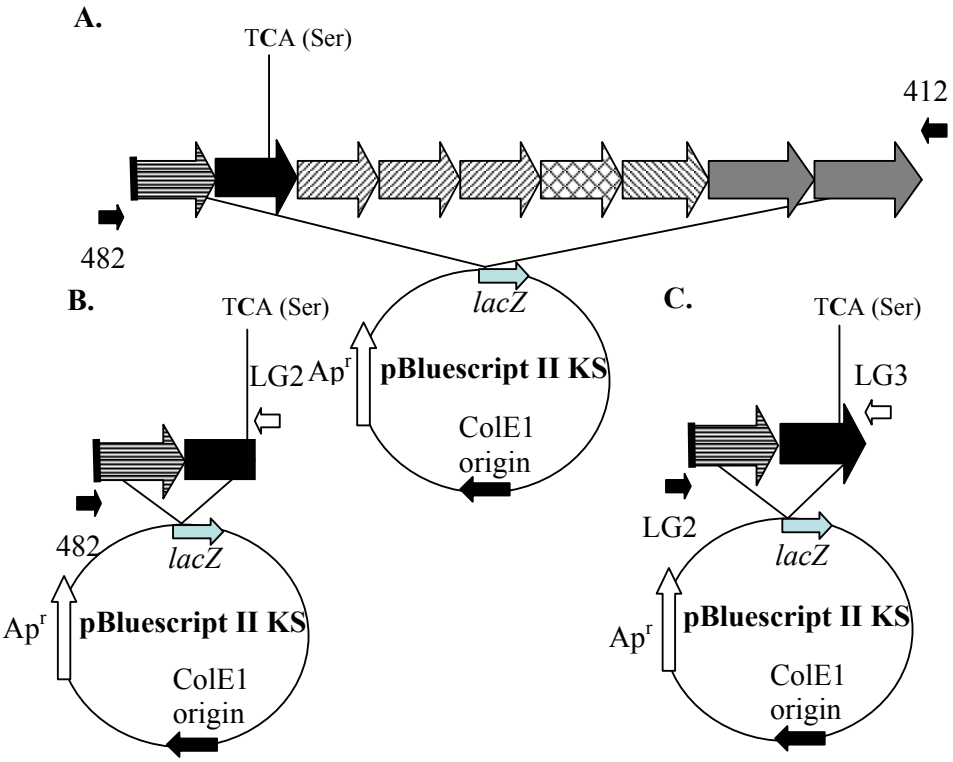

Fig. 21: Plasmids used for complementation analysis of the O6 LPS side chain expression in *E. coli* Nissle 1917 (see Fig. 20). The primers used for amplification of the cloned fragments are indicated with small arrows. A. pBWB536 B. pBLG2504 C. pBLG2894

To avoid gene dosage effects, *E. coli* strain Nissle 1917 was also chromosomally complemented by integration of a single copy of wzy_{536} fused with promoters of the β-lactamase-encoding gene of pBR322 (P_{bla}::wzy_{536}), or with the upstream region of the $wb_{\text{Nissle 1917}}$ gene cluster ($Pwb*_{\text{Nissle1917}}$::wzy_{536}) into the chromosomal attachment site of the bacteriophage λ. The wb_{O6}-specific upstream region of strain Nissle 1917 is 97 % identical to the previously studied wb_{O7} promoter (accession no. U23775). The resulting strains, in which a functional wzy_{536} copy has been stably integrated, were designated *E. coli* Nissle 1917λP_{bla}::wzy_{536} and *E. coli* Nissle 1917λ$Pwb*_{\text{Nissle 1017}}$::$wzy_{536}$, respectively (Fig. 22). The promoters P1 and P2 of the β-lactamase-encoding gene (*bla*) of pBR322 together with a Shine-Dalgarno sequence were fused with the wzy_{536} gene by PCR using primers LG7 and LG10, cloned into pGEM®-T Easy (resulting in pGPW1) and subcloned in plasmid pLDR11. The resulting plasmid pLBW1 (Tab. 5; Fig. 22a) was subsequently used for the integration of the P_{bla}::wzy_{536} fusion into the bacteriophage λ attachment site of the *E. coli* Nissle 1917 chromosome. Using the primers LG15 and LG16, the upstream region of the wb_{O6} gene cluster which is expected to contain the promoter(s) of the wb_{O6} determinant of strain Nissle 1917 was amplified (~450 bp). In parallel, the wzy_{536} gene was amplified

using primers LG8 and LG10. After digestion with *Xba*I, the two fragments were ligated and a PCR reaction was performed with primers LG15 and LG10 using the ligation mixture as a template. The obtained PCR product represents the wzy_{536} gene under transcriptional control of the $wb_{\text{Nissle 1917}}$-specific promoter (Fig. 22b). The fragment was then cloned into pLDR9 vector, thus resulting in pLPW1 which was used for integration of the strain Nissle 1917-specific P$wb^{*}_{\text{Nissle 1917}}$::$wzy_{536}$ fusion into the bacteriophage λ attachment site of the *E. coli* Nissle 1917 chromosome (Fig. 22). For this purpose, pLPW1 and pLBW1 were digested with *Not*I restriction endonuclease and the fragment of interest was re-ligated, leading to closed circular DNA molecules lacking a replication origin and carrying ampicillin resistance, which were then integrated into the *attP* site of the *E. coli* strain Nissle 1917 (Fig. 23).

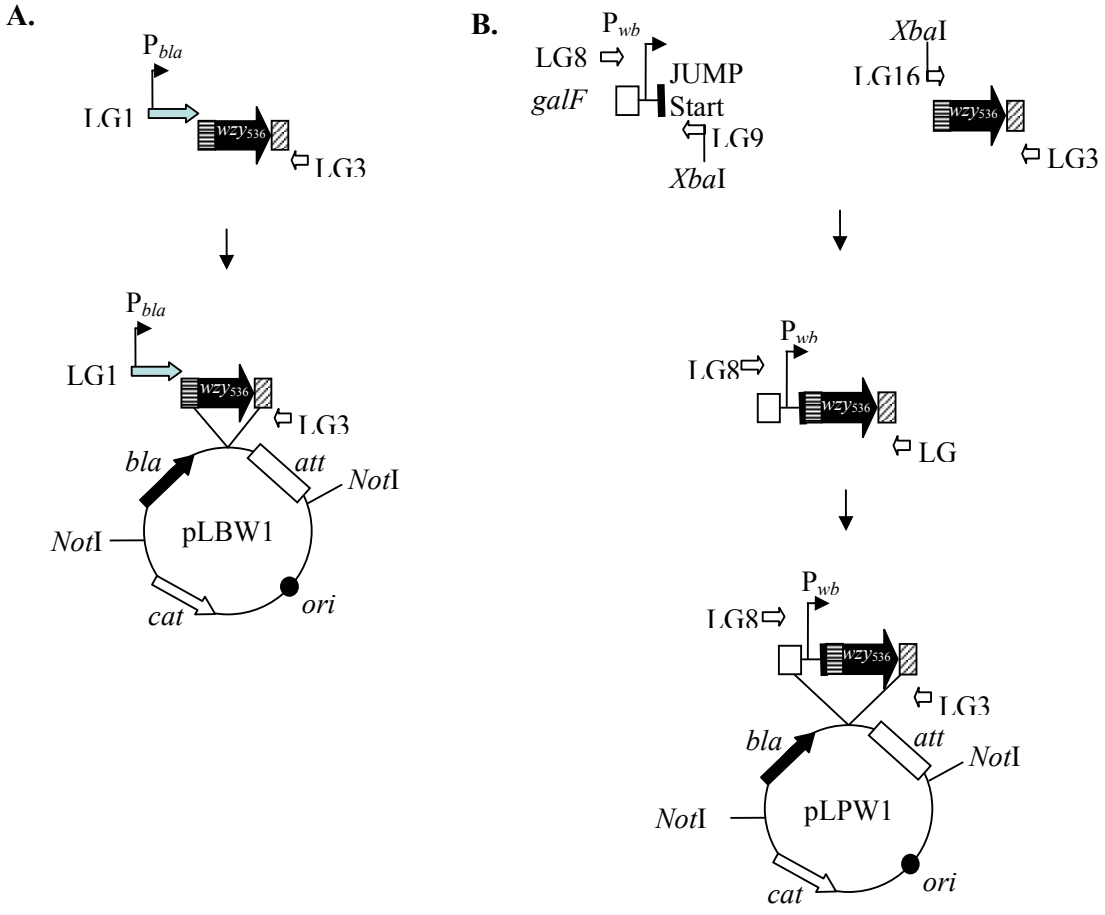

Fig. 22: Construction of different promoter::wzy_{536} fusions used for complementation of *E. coli* strain Nissle 1917. A. Construction of *E. coli* Nissle 1917λP_{bla}::wzy_{536} B. Construction of *E. coli* Nissle 1917λP$wb^{*}_{\text{Nissle 1917}}$::$wzy_{536}$ strains. The primers used for amplification (small arrows). *Not*I digestion of pLBW1 and pLPW1, circulization by ligation and transformation in *E. coli* Nissle 1917 (pLDR8). For details see the text.

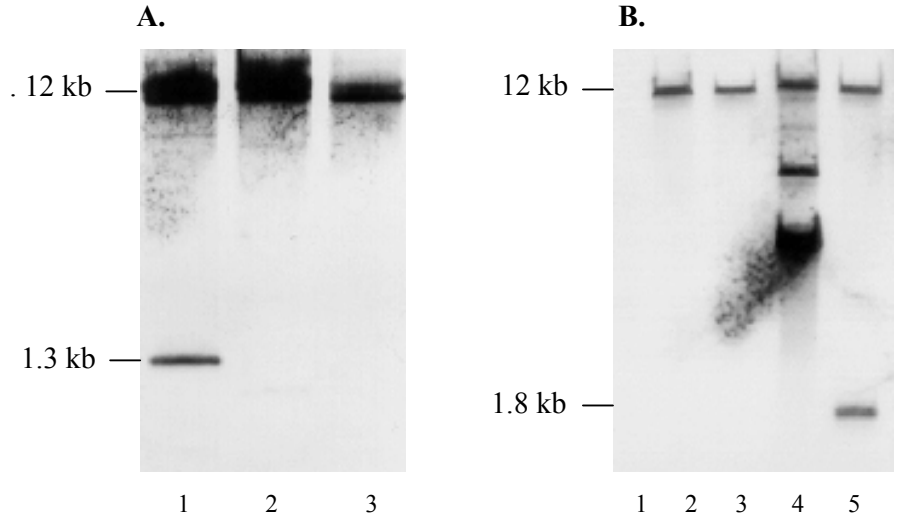

Fig. 23: Verification of integration of wzy_{O6} in the bacteriophage λ attachment site of the *E. coli* Nissle 1917 chromosome by Southern blot hybridization of *EcoR*I-digested genomic DNA. PCR-amplified $wzy_{Nissle\ 1917}$ (primers LG1 and LG3) was used as a probe. A. Lane 1, *E. coli* Nissle 1917λP$_{bla}$::wzy_{536}; lane 2, *E. coli* Nissle 1917; lane 3, *E. coli* 536. B. Lane 1, *E. coli* MG1655; lane 2, *E. coli* Nissle 1917; lane 3, *E. coli* 536; lane 4, *E. coli* Nissle 1917 (pBWB536); lane 5, *E. coli* Nissle 1917λP$wb^*_{Nissle\ 1917}$::wzy_{536}.

As seen from Fig. 23, an additional hybridization signal for the wzy_{O6} gene was detected in *E. coli* strain Nissle 1917λP$_{bla}$::wzy_{536} and *E. coli* strain Nissle 1917λP$wb^*_{Nissle\ 1917}$::wzy_{536} (1.3 kb and 1.2 kb, respectively) compare to the control strains Nissle 1917 and 536.

Expression of O6 side chains was studied by SDS-polyacrylamide gel electrophoresis. With regard to LPS side chain expression, strain Nissle 1917 and its derivatives were grouped into three classes: semi-rough, smooth, and smooth with reduced amount of O antigen (Fig. 24a).

Only transformation of *E. coli* strain Nissle 1917 with a construct containing the entire wb_{O6} determinant (pBWB536) of strain 536 resulted in a smooth phenotype. Introduction of the shortened wzy_{536} fragment (pBLG2504) representing the size of the truncated gene $wzy_{Nissle\ 1917,}$ had no complementing effect (Fig. 24a). Derivatives of strain Nissle 1917 complementation with plasmid-encoded wzy_{536} (pBLG2849) alone as well as the chromosomally complemented strains *E. coli* Nissle 1917λP$_{bla}$::wzy_{536} and *E. coli* Nissle 1917λP$wb^*_{Nissle\ 1917}$::wzy_{536} showed a smooth phenotype with reduced amounts of O antigen. The level of the O6 LPS side chain synthesis in strain Nissle 1917 (pBLG2849) was markedly lower than in the smooth strains Nissle 1917(pBWB536) and 536, but higher than that of the chromosomally complemented strains (Fig. 24a). Therefore, we conclude that the *E. coli* strain 536-specific *wzy* gene encodes the functional O6 antigen polymerase.

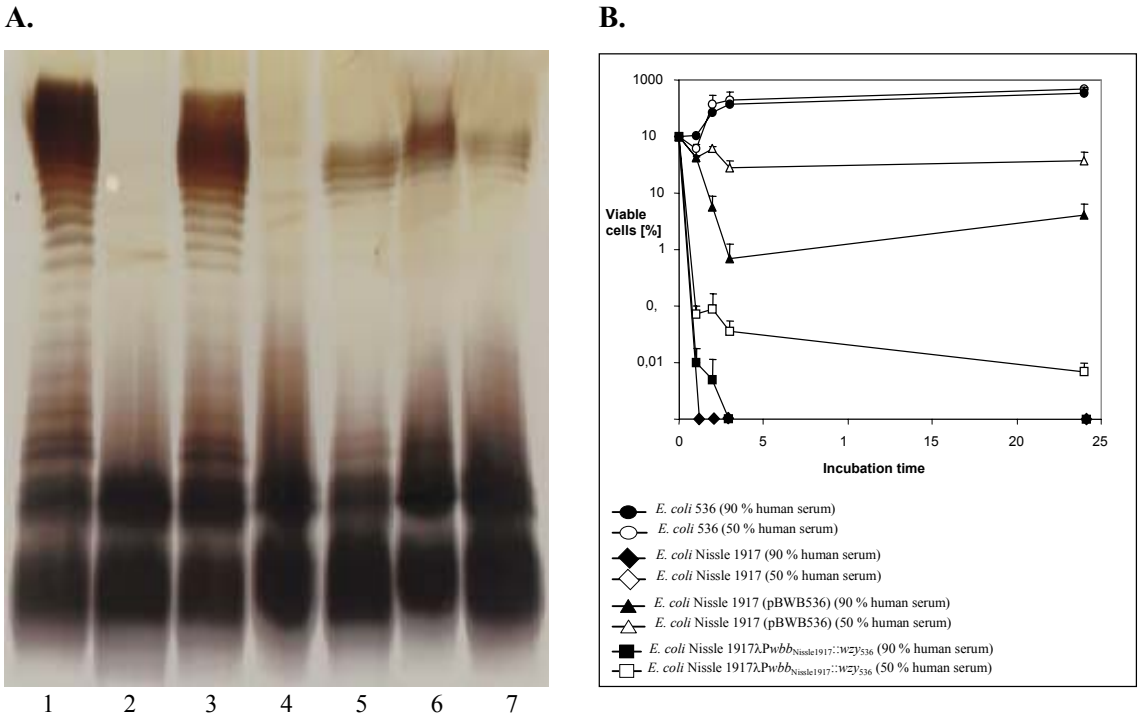

Fig. 24: Influence of O6 LPS side chain expression on serum resistance. A. SDS-PAGE analysis of the O6-specific LPS side chain length of *E. coli* strains Nissle 1917, 536 and derivatives. Lane 1, *E. coli* 536; lane 2, *E. coli* Nissle 1917; lane3 *E. coli* Nissle 1917 (pBWB536); lane 4, *E. coli* Nissle (pBLG2504); lane 5, *E. coli* Nissle (pBLG2849); lane 6, *E. coli* Nissle1917λP*wb**$_{\text{Nissle1917}}$::*wzy*$_{536}$; lane 7, *E. coli* Nissle 1917λP$_{bla}$::*wzy*$_{536}$. B. Serum resistance of *E. coli* strains Nissle 1917, 536 and derivatives. Serum resistance assays were performed in 90 % (indicated by black symbols) and 50 % human serum (indicated by white symbols). The percentage of surviving cells were plotted against incubation time in human serum. Cell numbers within the different inoculi (t = 0) were set 100 %.

With one representative of these three groups of strains (semi-rough, smooth and smooth but reduced amounts of O antigen), serum resistance assays were performed to analyze whether the presence and amount of longer LPS side chains may contribute to serum resistance in *E. coli* strain Nissle 1917 (Fig. 24b). The smooth strain *E. coli* Nissle 1917 (pBWB536) showed a markedly increased resistance to 50 % and 90 % human serum compared to that of the semi-rough wild type strain *E. coli* Nissle 1917, which was not detectable after 1 hour incubation in 50 % and 90 % human serum. In comparison to the wild type strain, the smooth strain with reduced amount of O antigen (strain Nissle 1917λP*wb**$_{\text{Nissle1917}}$::*wzy*$_{536}$) survived better and was still detectable after 24 hours of incubation (0.007 ± 0.001 % survival of the inoculum in 50 % serum) (Fig. 24b). Generally, serum resistance was in accordance with the amount of O6 repeating units produced in the different strains, e.g., serum resistance was higher in strains with higher amounts of O6-specific repeating units. This underlines that *wzy*$_{536}$ encodes the O6–specific antigen polymerase and that the C to A

transition at position +986 in $wzy_{Nissle1917}$, which results in a non-functional O antigen polymerase, is responsible for the semi-rough phenotype of *E. coli* strain Nissle 1917. We therefore identified the O6 antigen polymerase-encoding gene *wzy*, proved its function, and demonstrated that the reason for the semi-rough phenotype of *E. coli* strain Nissle 1917 is a point mutation probably leading to premature translation termination of *wzy*.

Allelic exchange of *wzy* in *E. coli* Nissle 1917 and 536

Since wzy_{536} alone was not able to restore the O6 side chain polymerization in strain Nissle 1917 to the same level as in the *E. coli* strain 536, allelic exchange of $wzy_{Nissle\ 1917}$ by wzy_{536} in strain Nissle 1917 and of wzy_{536} with $wzy_{Nissle\ 1917}$ in strain 536 was carried out. For this purpose wzy_{536} and $wzy_{Nissle\ 1917}$ were amplified using primers LG1 and LG3 and cloned into the suicide vector pCVD442 resulting in the plasmids pCVW536 and pCVW1917. These plasmids were used for allelic exchange of *wzy* in *E. coli* strains Nissle 1917 and 536, respectively (Fig. 25).

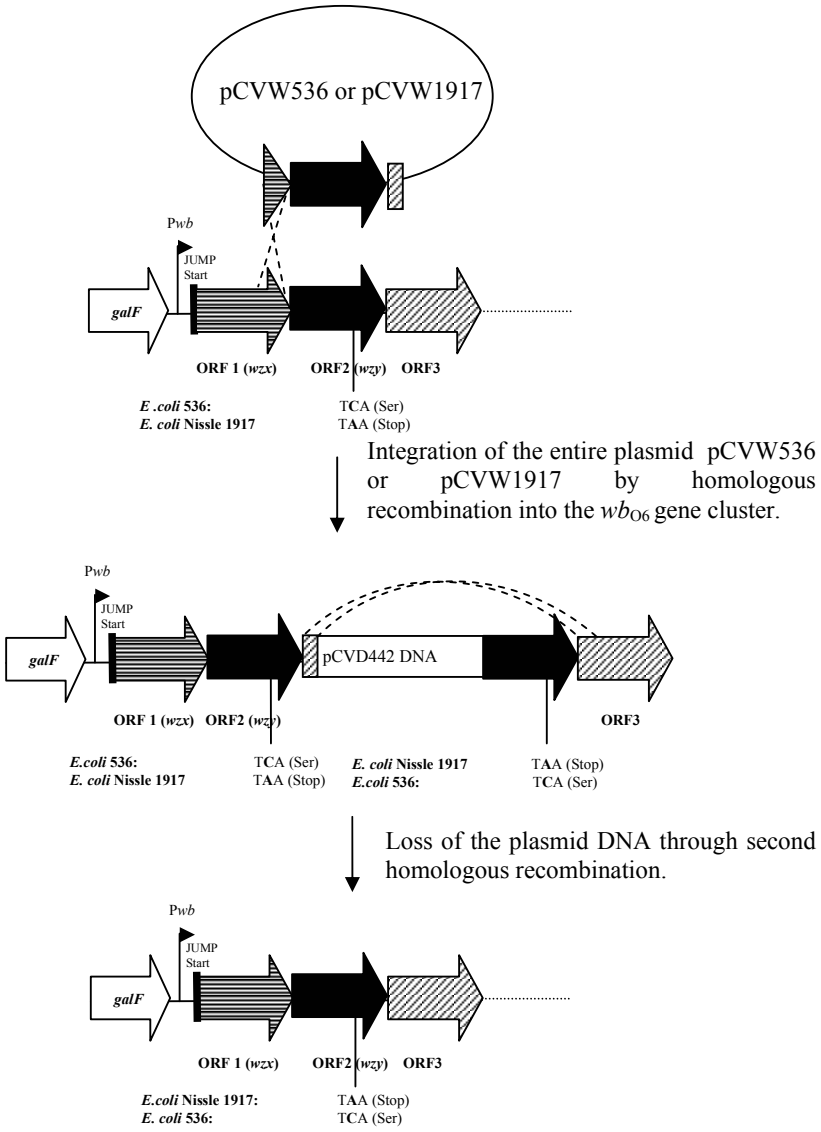

Fig. 25: Allelic exchange of *wzy* in *E. coli* Nissle 1917 and 536. Plasmids pCVW1917 and pCVW536 were transformed in *E. coli* Sm10 λ*pir* and used for conjugation *E. coli* strains 536 and Nissle 1917, respectively.

The presence or absence of the $wzy_{\text{Nissle 1917}}$-specific point mutation was demonstrated by PCR. For this purpose primers which enable detection of the presence of the $wzy_{\text{Nissle 1917}}$-specific mutation were designed (MUTA-up and MUTA-down (Fig. 26a)), and furthermore proven by sequence analysis of the *wzy* gene. The correct integration in the wb_{O6} determinant was demonstrated by PCR (using primers LG8, 482, LG1, LG2, LG3) and Southern hybridization (Fig. 26b).

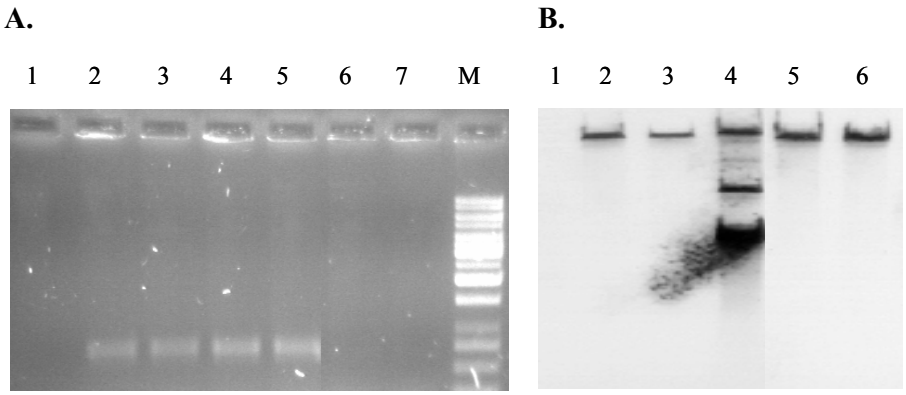

Fig. 26: Verification of proper allelic exchange of *wzy* gene by PCR and Southern blot hybridization. A. PCR detection of $wzy_{\text{Nissle 1917}}$-specific mutation using primers MUTA-up and MUTA-down: lane 1, *E. coli* 536; lane 2 - 4, *E. coli* 536-$wzy_{\text{Nissle 1917}}$; lane 5, *E. coli* Nissle 1917; lane 6 – 7, *E. coli* Nissle 1917-wzy_{536}; lane 8, 1 kb DNA ladder. B. SDS-PAGE analysis of the O6-specific LPS side chain length: lane 1, *E. coli* 536; lane 2, *E. coli* Nissle 1917-wzy_{536}; lane 3, *E. coli* 536-$wzy_{\text{Nissle 1917.}}$ B. Lane 1, *E. coli* MG1655; lane 2, *E. coli* Nissle 1917; lane 3, *E. coli* 536; lane 4, *E. coli* Nissle 1917 (pBWB536); lane 5, *E. coli* Nissle 1917-wzy_{536}; lane 6, *E. coli* 536-$wzy_{\text{Nissle 1971.}}$

Positive PCR signal was detected only in the strains carrying the *E. coli* strain Nissle 1917-specific mutation in the wzy_{O6} gene. Presence of only one hybridization signal of expected size confirmed the proper integration of the wzy_{O6} gene in wb_{O6} in the *E. coli* strains Nissle 1917-wzy_{536} and 536-$wzy_{\text{Nissle 1917.}}$ The clones were tested for O6 side chain polymerization by PAGE (Fig. 27).

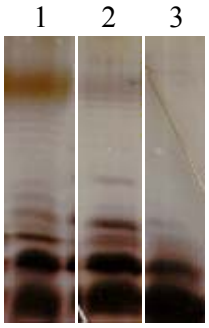

Fig. 27: Detection of O6 side chain expression by SDS-PAGE: lane 1, *E. coli* 536; lane 2, *E. coli* Nissle 1917-wzy_{536}; lane 3, *E. coli* 536-$wzy_{\text{Nissle 1917.}}$

The resulting strain *E. coli* Nissle 1917-wzy_{536} exhibited smooth phenotype with reduced amount of polymerized O antigen, whereas the *E. coli* 536-$wzy_{\text{ Nissle 1917}}$ was semi-rough. Since wzy_{536} alone was not able to restore the O6 side-chain polymerization at the same extent as in *E. coli* 536, it is tempting to speculate that some of the enzymes encoded by $wb_{\text{ Nissle 1917}}$ is less active than in *E. coli* 536, resulting in weaker O6 side-chain units synthesis. This presumption is in accordance

with the observation that the *wb* $_{536}$ (pBWB536) cluster is able to fully restore the O6 side-chain polymerization when transformed in Nissle 1917.

Screening of O6-serotype *E. coli* isolates for the presence of Nissle 1917-specific mutation in the *wzy* gene

In order to find out if the mutation in *wzy*$_{\text{Nissle 1917}}$ is present in other O6 isolates, the MUTA PCR was performed and the results were verified by sample sequencing of the amplified *wzy* genes. The conditions of the PCR amplification were carefully optimized (annealing temperature = 34.9 °C) Sixteen serum-resistant, as well as serum-sensitive O6-serotype *E. coli* strains, expressing different capsule types, were screened (Tab. 7). In no any of them the Nissle 1917-specific mutation was identified. This result indicates that probably this mutation was a unique event in the evolution of Nissle 1917 strain, and is not widely distributed between the O6 strains. The established MUTA PCR can be used for further screening of more O6-serotype *E. coli* isolates.

Table 7: Screeening of different *E. coli* O6 isolates for the presence of the *E. coli* strain Nissle 1917-specific mutation in *wzy*.

Strain	Serotype	Presence of the *E. coli* strain Nissle 1917 - specific mutation in *wzy*	O6 smooth phenotype	Serum resistance
Nissle 1917	O6:K5:H1	+	-	-
536	O6:K15:H31	-	+	+
RZ439	O6:K5	-	+	-
RZ442	O6:K5	-	+	+
RZ526	O6:K5	-	-	-
RZ533	O6:K5	-	+	+
RZ446	O6:K53:H1	-	+	+
RZ537	O6:K53:H1	-	+	-
RZ424	O6:K14:H-	-	+	-
RZ436	O6:K13:H1	-	-	+
RZ447	O6:K13:H1	-	+	-
RZ502	O6:K2:H-	-	+	+
RZ496	O6:K+:H-	-	-	-
RZ532	O6:K+:H31	-	+	-
RZ411	O6:K-:H1	-	+	-
RZ412	O6:K-:H1	-	+	+
RZ418	O6:K-:H1	-	+	-
RZ501	O6:K-:H1	-	+	-

3.2 Molecular characterization of determinants of *E. coli* strain Nissle 1917 required for biofilm formation

The ability of bacterial cells to bind to abiotic surfaces, the surface of eukaryotic cells, as well as to express specific extracellular matrix substances, can result in the formation of multicellular bacterial communities (biofilm). This requires expression of definite factors – fimbrial and non-fimbrial adhesins, surface proteins, extracellular matrix polymers (Stoodley et al., 2002). Biofilm formation of *Enterobacteriaceae* is believed to play a significant role for the colonization and establishment of infections on mucosal surfaces and may also enhance microbial survival in the environment (Boddicker et al., 2002; Cookson et al., 2002). The expression of type 1- and curli fimbriae is known to contribute to the ability of *E. coli* to form biofilm, as well as the cellulose biosynthesis, flagella, colanic acid and *agn43* expression (Zogaj et al., 2001; Römling et al., 1999; Hanna et al., 2003; Ferrieres et al., 2003; Danese et al., 2000).
In addition, deletion mutation of the *rfaH*-encoded transcriptional antiterminator was recently found to attenuate virulence of UPEC strain 536 and to enhance its ability to form biofilm under continuous culture conditions (Michaelis, K., and Dobrindt, U., unpublished data).

3.2.1 Sequence analysis of the *fim*$_{\text{Nissle 1917}}$ –containing pCos3YC6

In order to identify a cosmid clone containing the type 1 fimbriae-encoding gene cluster (*fim*), the genomic library of *E. coli* strain Nissle 1917 was screened by PCR using primers which enable specific amplification of *fimA*$_{\text{Nissle 1917}}$ (Blum-Oehler et al., 2003). The clone pCos3YC6 was identified and transformed in the fimbriae-deficient *E. coli* K-12 strain AAEC189. The expression of type 1 fimbriae was proven by mannose-dependent agglutination of yeast cells of strain AAEC189 (pCos3YC6). The presence of the *fim* genes was also demonstrated by Southern hybridization and sample sequencing. The entire cosmid insert was subsequently sequenced (Fig. 28; see section VIII, Tab. 6). The sequence analysis confirmed that the *fimB* gene, coding for tyrosine recombinase mediating the fimbrial phase switch, was found to be truncated by IS element as reported before (Stentebjerg-Olesen, et al., 1999). The *fimH* gene, coding for the FimH adhesin, is known to be highly variable among the *E. coli* strains and is responsible for the quantitative differences in binding to mannan and for the variability in binding specifities (Sokurenko et al., 1994). The *fimH*$_{\text{Nissle}}$

$_{1917}$ gene shows 99 % amino acid homology to that of the *E. coli* K-12 strain MG1655 and the EHEC strain EDL933 (see section VIII, Tab. 6). The chromosomal DNA context and localization of $fim_{\text{Nissle 1917}}$ does not differ from that of the already sequenced *E. coli* strains.

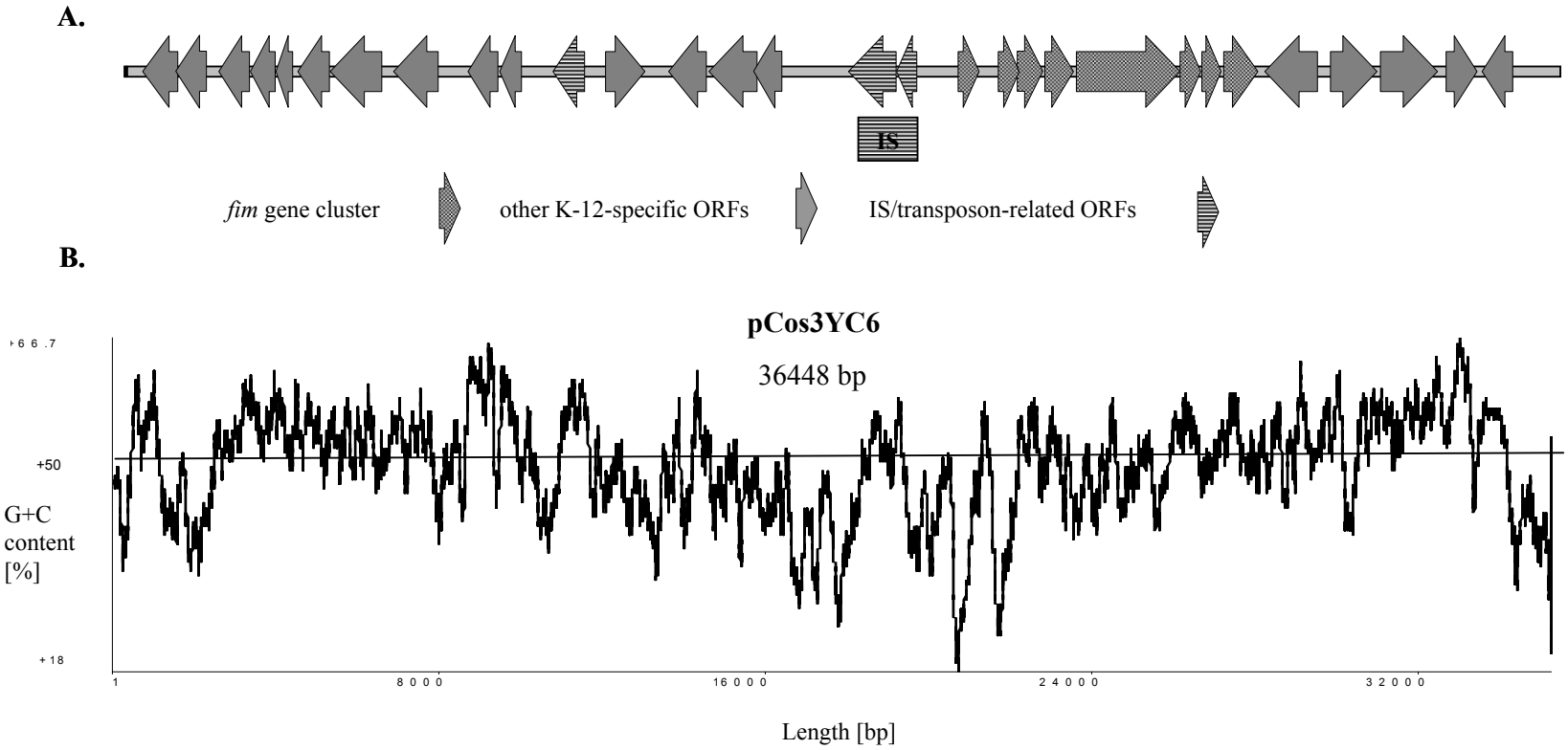

Fig. 28: Characterization of the *fim* determinant of the *E. coli* strain Nissle 1917. Genetic organization of the *fim* gene cluster and flanking regions of *E. coli* strain Nissle 1917. B. G+C content of the DNA in the cosmid insert.

3.2.2 Cloning, sequence-, and mutational analysis of cellulose biosynthesis gene cluster (*bcs*) of the *E. coli* strain Nissle 1917

The *E. coli* strain Nissle 1917 expresses at 30 °C and 37 °C the so-called rdar morphotype on congo red agar plates (Zogaj et al., 2001) which is characterized by intensive red colored colonies with a dry colony surface and strong cell-cell interactions. This morphotype is considered to be due to co-expression of curli fimbriae and cellulose. In comparison, the uropathogenic *E. coli* strain 536 expresses the rdar morphotype only at 30 °C and strain CFT073 does not express rdar morphotype neither at 30 °C or 37 °C.

The cellulose biosynthesis is known to play a significant role for the biofilm formation, but does not contribute to the virulence of *Salmonella enteritidis* (Solano et al., 2002). In order to find out if strain Nissle 1917 expresses cellulose, the entire $bcs_{\text{Nissle 1917}}$ gene cluster was amplified (using primers bcs-up and bcs-down), subsequently sequenced (Fig. 29; see section VIII, Tab. 7), cloned into pGEM-T Easy® thus resulting in pGBC1917. The strain DH5α (pGBC1917) exhibited red-pink-colored colonies when grown on congo red agar plates, thus proving that the $bcs_{\text{Nissle 1917}}$ gene cluster alone is able to mediate cellulose biosynthesis at 30 °C and 37 °C (called pdar morphotype: pink dry and rough; Zogaj et al., 2003). In addition, a $bcs_{\text{Nissle 1917}}$ deletion mutant (strain Nissle 1917Δ*bcs*) was created, as depicted in Fig. 29. Deletion of all coding ORFs in the cloned $bcs_{\text{Nissle 1917}}$ (pGBC1917) operon was created using the enzymes *Kpn*I and *Mcs*I. Consequently, blunt ends were generated and the linearized plasmid was religated resulting in pGDB1917. The insert was cloned into pCVD442 using *Sph*I and *Sac*I thus resulting in pCDB1917. After allelic exchange of the $bcs_{\text{Nissle 1917}}$ in the chromosome of the *E. coli* strain Nissle 1917 using pCDB1917, clones carrying deletion in $bcs_{\text{Nissle 1917}}$ were identified (Fig. 29). Chromosomal deletion of $bcs_{\text{Nissle 1917}}$ was confirmed by PCR, Southern hybridization and phenotypically by growth on congo red agar plates.

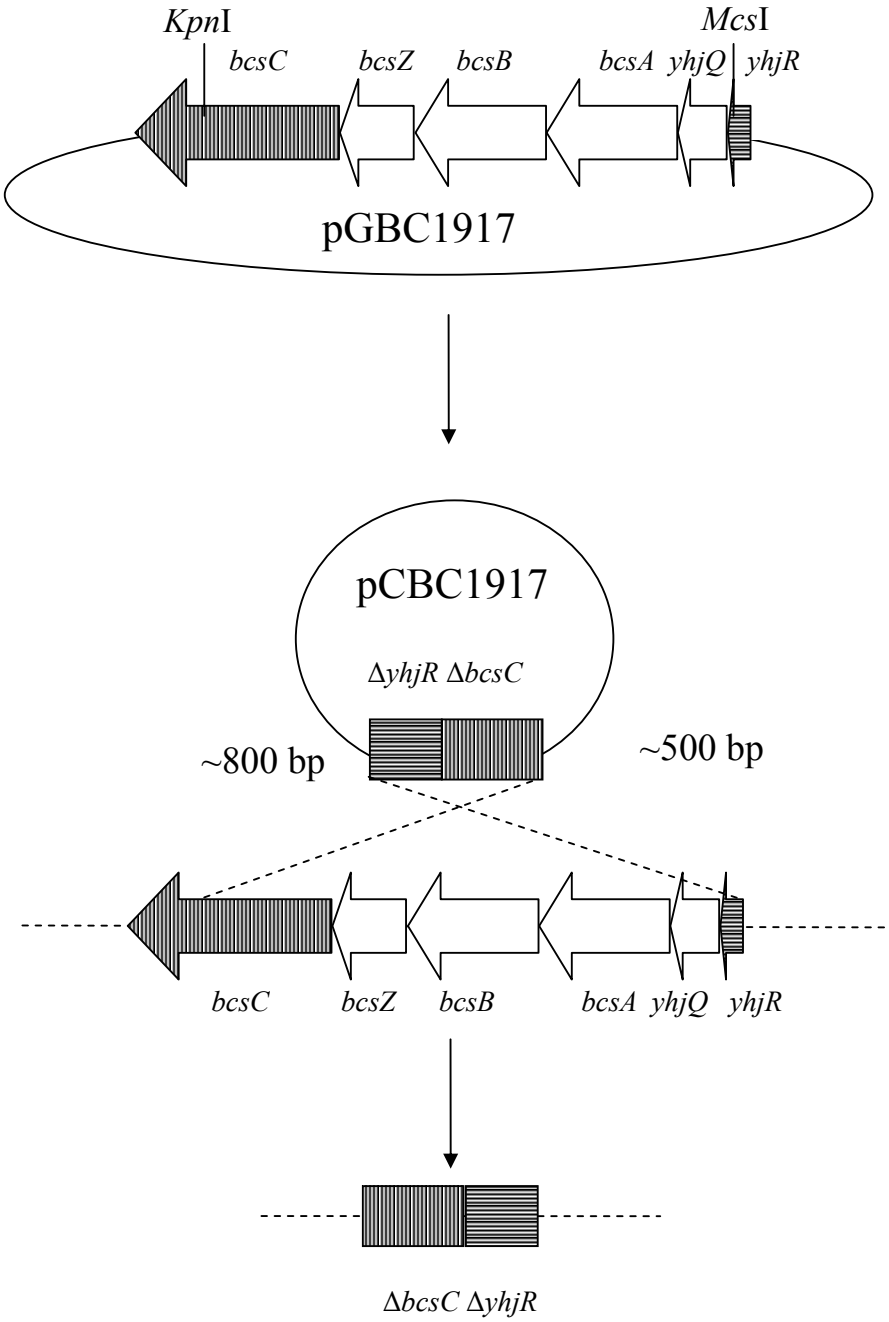

Fig. 29: Construction of *E. coli* strain Nissle 1917Δ*bcs* strain.

The *E. coli* strain Nissle 1917Δ*bcs* lost the ability to express rdar morphotype and exhibited brown-colored colonies due to the expression of curli fimbriae alone, when grown on congo red plates at 30 °C and 37 °C, and thus expressed the bdar morphotype (brown dry and rough; Fig. 30).

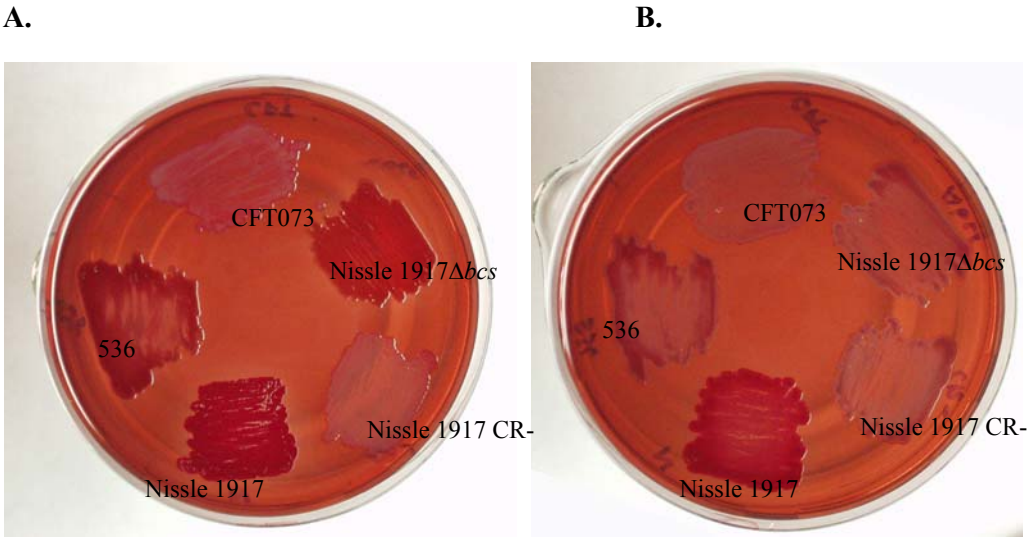

Fig. 30: Manifestation of the rdar morphotype in *E. coli* strains – congo red binding assay. Investigated *E. coli* strains are designated for every colony; *E. coli* Nissle 1917 congo red-negative spontaneous mutant strain is designated as Nissle 1917 CR-. A. 30°C incubation temperature. B. 37°C incubation temperature.

3.2.3 Sequence determination of *csgD* and the promoter region of *csg* gene cluster of *E. coli* strains Nissle 1917 and 536

The *csgD* gene encodes transcriptional regulator protein CsgD which has been shown to positively control the expression of curli fimbriae and cellulose biosynthesis (rdar morphotype; Römling et al., 1998). Mutations in its promoter region may influence the mutlicellular behaviour of *Salmonella typhimurium.* Therefore, using primers curli 1 and curli 2, this region was amplified from chromosomes of *E. coli* strains 536, Nissle 1917 and Nissle 1917 CR- (congo red-negative spontaneous mutant strain) and PCR products were subsequently sequenced (see section VIII.1). Any sequence differences have been detected between the wild type *E. coli* strain Nissle 1917 and the congo red binding-negative mutant strain Nissle 1917 CR-. In addition, no sequence differences in genes which are known to play a role in regulation of the rdar morphotype have been observed between the two strains (*crl, ompR, mlrA*; data not shown). According to these results it is tempting to speculate that there are other factors contributing to the rdar morphotype of strain Nissle 1917.

3.3 Amplification of the *chu* determinant

In order to prove the presence of the haeme iron uptake system in the chromosome of strain Nissle 1917, the entire *chu* gene cluster was amplified (data not shown) using primers chu1 and chu4. The resulting DNA fragment was of expected size (11 kb) and sample sequencing of the PCR product confirmed the presence of the *chu* determinant in the genome of *E. coli* strain Nissle 1917.

3.4 Characterization of the role of *rfaH* for gene expression in *E.coli* strain Nissle 1917

In order to construct a *rfaH* mutant of strain Nissle 1917, the pCVD442-based plasmid pMSK5 was used (Kupfer, M., 2000; Nagy, G., 2000). This plasmid contains the *rfaH* gene inactivated by insertion of the *cat* casette. The plasmid was conjugated into the *E. coli* strain Nissle 1917, and allelic exchange was carried out as described above.

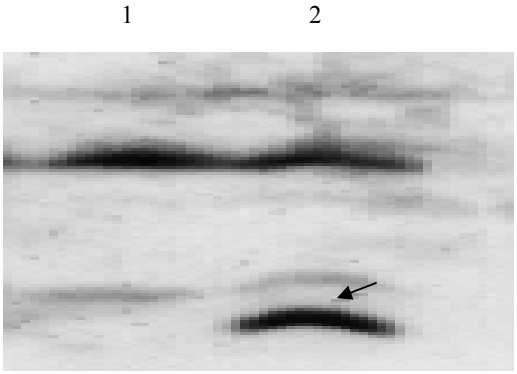

Fig. 31: Western hybridization of whole-cell protein extract. The RfaH-specific band is showed with arrow. lane 1, *E. coli* Nissle 1917*rfaH-*; lane 2, *E. coli* Nissle 1917.

Candidate *rfaH* mutants were screened for proper integration of the *cat* gene into the chromosome by PCR and Southern hybridization, using *cat*- and *rfaH*-specific probes (data not shown). Finally, the lack of RfaH in the *E. coli* strain Nissle 1917*rfaH-* was proven by Western blot hybridization, by using an anti-RfaH rabbit monoclonal antibody (Dobrindt, U., personal communication; Fig. 31).

No effect of the *rfaH* inactivation on the microcin and aerobactin expression of Nissle 1917 was observed when phenotypic tests were performed, as well as no clear influence on the rdar morphotype at 30 °C or 37 °C (data not shown).

4. Screening of commensal and uropathogenic *E. coli* strains for the presence of *E. coli* Nissle 1917-specific sequences and analysis of the genome content of strain Nissle 1917 by DNA-DNA hybridization using DNA arrays

4.1 Screening of commensal and uropathogenic *E. coli* strains for the presence of *E. coli* Nissle 1917-specific sequences by PCR

In order to detect *E. coli* Nissle 1917–specific sequences in the genome of different *E. coli* strains, PCR reactions were designed for the amplification of DNA regions containing genes encoding for (putative) fitness-factors, *E. coli* Nissle 1917-specific insertion sequences, or the regions located at the borders of GEI $I_{Nissle\ 1917}$ and GEI $II_{Nissle\ 1917}$ (see section VIII, Tab. 8). 68 uropathogenic *E. coli* isolates of serotype O6:K5 were investigated (Zingler er al., 1992; Zingler et al., 1993), as well as 40 non-pathogenic faecal isolates (Boyd et al., 1997; Mühldhofer et al., 1996) were screened for the presence of the above mentioned DNA regions. Generally, the results show a wide distribution of the *E. coli* Nissle 1917-specific sequences among the O6:K5 isolates, but also the presence of these DNA regions in non-pathogenic faecal isolates which are phylogenetically unrelated to *E. coli* strain Nissle 1917 (Tab. 8; see section VIII, Tab. 8).

Table 8: Percentage distribution of the investigated *E. coli* Nissle 1917-specific sequences between uropathogenic and commensal *E. coli* strains.

Origin of the strain / PCR (%)	*mcmD*	*focG*	*iroB*	*iuc* (*iucB* – *iucC*)	*sat*	*iha*	GEI III (ORF23 - ORF24)	HPI 3 R (*irp1* – *ybtT*)	HPI 4 L (*irp2*)	GEI I left (Z4866)	GEI I right (c1274)	GEI II left (*shiA*)	GEI II right (*kfiC*)	IS*2*-left (IS*2* – *shiF*)	IS*2*-right (IS*2* – *iha*)	put. IS (pCos3 YC6)-*fimA*
faecal, non-pathogenic strains	12.5	7.5	35	47.5	20	32.5	0	17.5	17.5	35	50	52.5	5	5	10	5
uropathogenic strains	73.5	61.5	95	67.6	58.8	66.1	42.6	97	100	92.6	95.5	100	52.9	10.3	35.3	25

Interestingly, two of the investigated O6:K5 isolates (RZ442 and RZ525) were positive for all of the *E. coli* Nissle 1917 PCR reactions, suggesting a common clonal origin of these strains and strain Nissle 1917. Both strains contain one (pMUT1) of the two plasmids present in strain Nissle 1917, as indicated by a pMUT1-specific PCR reaction (primers Muta5 and Muta6; Blum-Oehler et al., 2003). They were negative for pMUT2 when tested with pMUT2-specific primers (primers MUTA7 and Muta8; Muta9 and Muta10). Despite the high similarity on the genomic level between the *E. coli* strains RZ442, RZ525 and *E. coli* Nissle 1917, they do not express the rdar morphotype, neither on 30 °C or 37 °C. These findings support the view that the *E.*

coli strain Nissle 1917 exhibits a specific pattern of fitness factors which might contribute to its colonization efficiency and survival in the host body, but lacks prominent virulence factors.

4.2 Analysis of the genome content of strain Nissle 1917 by DNA-DNA hybridization using DNA arrays

In data provided from U. Dobrindt the overall genome content was analyzed by DNA-DNA hybridization experiments with *E. coli* K-12-specific arrays and the "*E. coli* Pathoarray" (Dobrindt et al., 2003). The results of the DNA-DNA hybridization of genomic DNA isolated from strain Nissle 1917 with *E. coli* K-12 gene arrays demonstrated that 90.7 % of all translatable ORFs of *E. coli* K-12 strain MG1655 were detectable in strain Nissle 1917. The majority of these missing ORFs can be functionally grouped as coding for hypothetical, not experimentally classified or unknown gene products. A great diversity of ORFs which represent mobile genetic elements or which code for structural components of the cell in the K12 genome were not detectable in strain Nissle 1917 as well. Ten prophages described in *E. coli* strain MG1655 were not detected in strain Nissle 1917 (data not shown). Several variable chromosomal regions among the studied isolates contain ORFs with homology to ORFs of other accessory genetic elements, e.g. IS elements. The chromosomal context of several tRNA-encoding genes (e. g., *serX*, *argW*, *ileY*, *pheV*, *leuX*) was also found to contain alterations in strain Nissle 1917 in comparison to the corresponding sequences in *E. coli* MG1655 implying the presence of horizontally acquired genetic information downstream of these tRNA loci.

The genome of strain Nissle 1917 was also screened for the presence of DNA sequences which belong to the flexible gene pool of UPEC strain 536, as well as for typical virulence- and fitness-associated genes of extraintestinal pathogenic *E. coli* (ExPEC) and intestinal pathogenic *E. coli* (IPEC) including *Shigella* using the previously described "*E. coli* Pathoarray" (Dobrindt et al., 2003). 55 % of the probes specific for PAI I-V_{536}-ORFs and 37 % of the probes specific for other ExPEC virulence-associated genes spottet on the Pathoarray have been detected in the genome of strain Nissle 1917. Generally, the hybridization signals confirmed the results obtained from the characterization of GEIs of strain Nissle 1917. About 18 % of the probes specific for virulence-associated genes of intestinal pathogenic *E. coli*

showed a clear hybridization signal. Most of these probes are complementary to fimbrial determinants of enterotoxigenic *E. coli* (indicating possible cross-reaction of probes designed for the detetction of ExPEC fimbriae-encoding determinants) with or to putative ORFs present on PAIs of the enterohemorrhagic *E. coli* O157:H7 strain EDL933 (Perna et al., 2001). Importantly, known protein toxin-encoding determinants of pathogenic *E. coli* have not been detected by this approach in strain Nissle 1917.

VI. Discussion

1. Genome plasticity and evolution of *E. coli* species

The species *Escherichia coli* comprises a variety of commensal, as well as of pathogenic isolates with striking differences regarding the diseases they cause, i. e., sepsis, local infections of the intestine, kidney, bladder and brain in different hosts (humans, monkeys, horses, birds). The broad spectrum of pathogenic features and different clinical symptoms cause by *E. coli* pathotypes mirrors the presence of different subsets of virulence-associated genes in certain pathotypes which are absent in commensal isolates (Dobrindt et al., 2003). The genomes of different *E. coli* strains exhibit remarkable size variations (4 – 5.5 Mb; Bergthorsson and Ochman, 1998). This indicates the high variability in the gene content among different *E. coli* strains, that is mainly due to the acquisition of foreign DNA elements and the deletion of genetic information. Thus, horizontal gene transfer and genome reduction contribute to the evolution of prokaryotes, leading to the appearance of new species, subspecies and pathotypes. The acquisition of plasmids, phages and large DNA regions termed "genomic islands" is important for the adaptation and microbial fitness in a specific environment or host. The deletion of certain DNA regions or inactivation of specific genes can also contribute to a definite way of successful survival or disease development for some strains (Jin et al., 2002).

Since *E. coli* is part of the normal intestinal microflora of humans (commensal strains) it is of great importance to investigate the genetic mechanisms by which pathogenic variants evolve. It is well known that the normal intestinal flora is the main source of *E. coli* strains causing UTI, new-born meningitis, and sepsis. Therefore, the microbial composition of the intestine plays crucial role for the development of these extraintestinal diseases, as well as for diarrheagenic disorders. It was shown that commensal *E. coli* isolates share some common genetic features with ExPEC (extraintestinal pathogenic *E. coli*) and IPEC (intestinal pathogenic *E.coli* strains) (Dobrindt et al., 2003). Most of these features can be considered fitness-contributing factors (e.g., iron-uptake systems, bacteriocins, fimbriae and other adhesins, some specific O- and K-serotypes). Nevertheless, they are commonly associated with pathogenic isolates. Thus, the question whether certain *E. coli* strains will develop into pathogenic or successful commensal variants can be considered as the

evolutionary balance between the acquisition of sufficient foreign genetic information increasing bacterial fitness to ensure the successful multiplication and colonization of the host on one hand, and the parallel process of acquisition of genetic determinants contributing to pathogenesis on the other. To make this speculative picture more clear, one should also have in mind the accompanying processes of gene silencing and inactivation, deletions and genome reduction.

2. *E. coli* strain Nissle 1917 and UPEC O6 strains: common phenotypic features and differences

The *E. coli* strain Nissle 1917 is a non-pathogenic, faecal isolate which has been used for decades as a probiotic in medicine (Mutaflor®), e. g. for the treatment of wide range of intestinal disorders and reconstruction of the intestinal flora of immuno-supressed individuals. It has been shown that strain Nissle 1917 is serum-sensitive, and non-invasive. This strain does not exhibit a uropathogenic virulence potential and does not provoke intestinal disorders. At the same time the strain is able to multiply successfully in the human intestine, which indicates that it possesses sufficient fitness-conferring determinants ensuring its survival and colonization capacity. These include fimbrial adhesins, whose expression is important for the adhesion to the epithelial cells: the type 1, F1C and curli fimbriae. The type 1 fmbriae (encoded by the *fim* gene cluster) and the curli adhesin (encoded by the *csg* gene cluster) are widely distributed among pathogenic and non–pathogenic isolates exhibit a conserved chromosomal localization. The F1C fimbriae are mostly associated with uropathogenic strains and do not belong to the *E. coli* core genome. Thus, their genes are believed to be acquired by horizontal gene transfer. They are a member of the S family of fimbrial adhesins (Dobrindt et al., 2001) and the corresponding determinant is located on genomic islands. *E. coli* strain Nissle 1917 is also known to express microcins, which probably contribute to its fitness as a result of the antagonistic competition with other inhabitants of the intestine.

This strain also possesses a striking number of iron uptake systems, which facilitate its survival in the human body. The presence of the entire *chu* determinant required for haemine uptake in the genome of Nissle 1917 was demonstrated in this study by long-distance PCR. Some of the iron acquisition systems are frequently detected in pathogenic strains (salmochelin, *iro*; aerobactin, *iuc*; haemine-uptake, *chu*). Others are also widely distributed among non-pathogenic isolates (enterobactin, *ent*;

yersiniabactin, *ybt*; ferric dicitrate uptake system, *fec*). The majority of them have a conserved chromosomal localization (*ent, chu, ybt*). Nevertheless the *ybt* gene cluster represents the "high-pathogenicity island" (Schubert et al., 1999) which is known to be horizontally-transfered mobile genetic element. Although the *fec* genes are present in the core genome of *E. coli* K-12, they might also be part of a pathogenicity island (Luck et al., 2001). The aerobactin-encoding *iuc* gene cluster is either present on plasmid (Braun et al., 1983) or on chromosomal island (Purdy et al., 2001; Vokes et al., 1999).

The *E. coli* strain Nissle 1917 expresses a K5 capsule, which is widely distributed among UPEC strains and presumably the K5-specific *kps* gene cluster is also part of a genomic island. The K5 capsule expression is important for adhesion and colonization, but does not contribute to serum resistance (Herias et al., 1997; Burns et al., 1998). The K5 capsule belongs to the Group 2 capsules, which are encoded by a gene cluster which is organized in three regions (regions 1, 2 and 3). Region 1 and 3 are conserved among all group 2 capsule-encoding operons and code for eight Kps proteins responsible for the transport of the capsule polysaccharides. Region 2 is serotype-specific and encodes enzymes for the polymerization of the polysaccharide molecule and, if necessary, for the biosynthesis of specific monosaccharide components (Whitfield and Roberts, 1999).

UPEC strains express a number of fitness factors, as well as different toxins. The P-related, F1C-, and S-fimbriae (encoded by the *pap/prf, foc, sfa* gene clusters) are widely distributed among UPEC and are usually localized on genomic islands. They play role for cell adhesion, but their inactivation in UPEC strain CFT073 did not alter the virulence of this strain (Mobley et al., 1994; Smyth et al., 1996). Thus, these fimbriae are factors contributing to the infectious potential, but they are not necessarily sufficient by themselves to cause infection and disease.

Toxins expressed by the majority of the UPEC isolates, like α-haemolysin (*hly*), the cytotoxic-necrotizing factor 1 (CNF-1), cytolethal distending (CDT) or some autotransporter toxins (*sat*) are considered prominent virulence factors. Nevertheless, it has been shown that although the secreted autotransporter toxin Sat of the UPEC *E. coli* strain CFT073 exhibits a vacuolating cytotoxic effect on bladder and kidney epithelial cells, its inactivation did not impair the virulence potential of strain CFT073 (Guyer et al., 2002). Thus, this serine protease might be considered more as a fitness factor providing peptides for the bacterium. It is noteworthy to emphasize that the *E.*

coli strain Nissle 1917 does not express any of these toxins connected with virulence of UPEC isolates.

In this study, the genome structure and genetic organization of several fitness- and virulence factor-encoding determinants has been compared between the well studied UPEC strains 536, CFT073 and the non-pathogenic strain Nissle 1917. All three strains exhibit an O6-serotype, but differ in the capsule type they express (K15, K2 (?), and K5, respectively). Although all strains express S-fimbrial adhesins, strain Nissle 1917 does not express P-related fimbriae. Whereas both UPEC strains are serum-resistant, smooth, and express α-haemolysin, strain Nissle 1917 is serum-sensitive, semi-rough and does not express α-haemolysin. The sequence analysis of the genes coding for fitness-contributing factors of *E. coli* strain Nissle 1917 revealed a surprisingly higher structural similarity with the pathogenic isolate CFT073 than with strain 536. The chromosomal localization of "black holes" in the genome of strain Nissle 1917 were identified, where virulence determinants which are present in the genome of the UPEC strains are deleted or replaced by a transposon-like element in GEI $II_{Nissle\ 1917}$, e. g. the *pap*-operon which is partially deleted in GEI $II_{Nissle1917}$. Typical virulence determinants have so far not been detected in the genome of *E. coli* strain Nissle 1917.

In spite of the similarity of the genome organization of strain Nissle 1917 and the UPEC strains 536 and CFT073, this non-pathogenic strain exhibits some intriguing phenotypic features which are not typical for the UPEC isolates, like presence of two plasmids (Blum et al., 2003) and the so-called rdar morphotype (Römling et al., 1998). The latter is due to the strong, temperature-independent curli fimbriae and cellulose expression and is not characteristic for UPEC strains including CFT073 and 536 (Dobrindt et al., personal communication).

3. The lipopolysaccharide of *E. coli* Nissle 1917

Together with *E. coli* O1, O4 and O18 strains, those expressing the O6 antigen belong to the most frequent extraintestinal pathogenic *E. coli* isolates (Ørskov et al., 1977). However, *E. coli* strains of serotype O6 are also commonly detected among intestinal isolates (Bettelheim et al., 1997; Hartley et al., 1979; Lidin-Janson et al., 1978; Sonnenborn and Greinwald, 1991). In the case of uropathogenic O6 strains, the gut may thus serve as a reservoir of infectious microorganisms for recurrent urinary tract infections (Appelmelk et al., 1994). The R1-type core is the most frequently occurring

core type in clinical *E. coli* isolates (Amor et al., 2000).

The entire *wa** and *wb** gene clusters of *E. coli* strain Nissle 1917, which are required for biosynthesis of the *E. coli* LPS R1 core type and of the O6 antigen, respectively, have been cloned, sequenced, and analyzed. The nucleotide sequence of the wa_{O6} determinant is 97 % identical to already known sequences of other R1 core type-specific *wa** gene clusters (e.g. *E. coli* strain F470; Heinrichs et al., 1998). The wb_{O6} gene cluster had not been sequenced so far. In *E. coli* strain Nissle 1917, this determinant is chromosomally located between the genes *galF* and *gnd* as reported for other *wb** clusters (Coimbra et al., 1998). As also described for other O antigen-encoding gene clusters, all putative ORFs of this strain‘s *wb** gene cluster, with the exception of *manC* and *manB*, have a relatively low G+C content suggesting that these genes may have been acquired by horizontal transfer from other species. According to the corresponding deduced amino acid sequences, five putative ORFs specific for the O6 LPS serotype have been identified: one putative ORF coding for the O6 antigen flippase *wzx*, the O6 antigen polymerase-encoding gene *wzy*, four putative glycosyl transferase-encoding genes and a putative epimerase-encoding gene (Fig. 20, Tab. 6). Although ORF 2 shows no marked similarity to other *wzy* genes, it was considered the putative O antigen polymerase-encoding gene as this ORF is located downstream of the putative *wzx* gene. In addition, 12 transmembrane helices have been predicted from the deduced amino acid sequence. This is also the case for the putative Wzy proteins of an *E. coli* O113 and O8:K40 strain (AF172324 and AF013583), respectively. Generally, the number of transmembrane helices of Wzy proteins of other *E. coli* serogroups is variable ranging from 8 (*E. coli* O157:H7 strains EDL933, Sakai and an *E. coli* O7 strain (AAG57099, BAB36267, AF125322), to 10 in isolates of serogroup O55, O104 and O111 (AAL67557, AAK64372, AAD46730) or 11 in serogroup O4 (U39042) and in K-12 (AAB88404). The function of ORF6 could not clearly be defined according to sequence similarity. As N-acetyl-galactosamine was identified to be present in the O6 O-repeating unit structure, an UDP-N-acetylglucosamine-4-epimerase should be encoded within the wb_{O6} gene cluster. However, the deduced amino acid sequence of ORF6 shows a higher similarity to UDP-glucose-4- and UDP-galactose-4-epimerases than to the UDP-N-acetylglucosamine-4-epimerase of *E. coli* O55 (AF461121) (Wang et al., 2002) (see section V.3.1.2, Tab. 6).

To correlate the genetic and structural analyses (Grozdanov et al., 2003; Fig. 32), the

cloned O6 antigen gene cluster of the smooth *E. coli* strain 536 was compared with that of the semi-rough strain Nissle 1917. It was shown that the entire wb_{O6} gene cluster of strain 536 was able to restore full-length O6 side chain synthesis and to complement the semi-rough phenotype of *E. coli* strain Nissle 1917. Therefore, it was concluded that the predicted mutation is located in this strain‘s O6 antigen gene cluster. Following sequence annotation, a C to A transition within $wzy_{\text{Nissle 1917}}$ was identified which results in a premature stop codon (TAA). Since the rest of the nucleotide sequence of *wzy* in the *E. coli* strains Nissle 1917 and 536 was identical, the strain Nissle 1917 was complemented with the entire wzy_{536} as well as with the shorter form representing the size of the strain Nissle 1917-specific *wzy* gene and it was proven that only the intact wzy_{536} is functional and able to restore the O6 side chain synthesis in *E. coli* Nissle 1917.

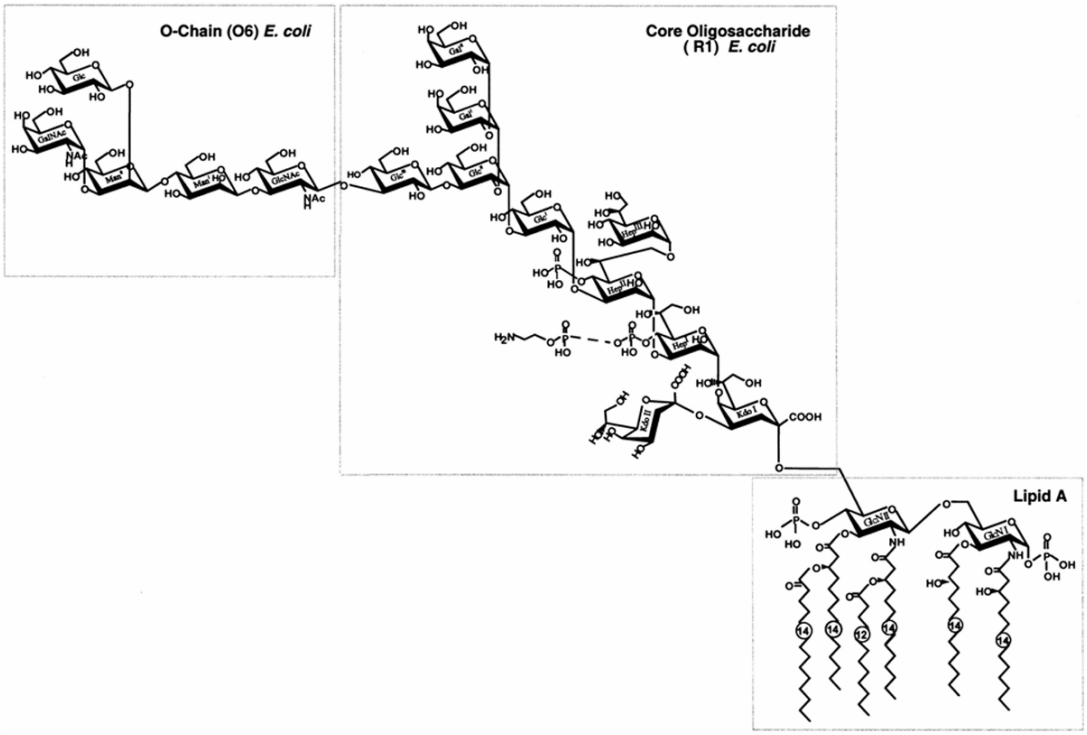

Fig. 32: Structure of the complete semi-rough LPS from *E. coli* Nissle 1917 containing O-antigen-core oligosaccharide and lipid A. Incomplete substitution with Etn-*P* is shown by dashed lines.

Although chromosomally complemented derivatives of strain Nissle 1917 which carry one copy of wzy_{536} in the chromosomal attachment site of the bacteriophage λ, expressed less LPS side chains than strain 536, the results demonstrated that expression of a functional O6-specific *wzy* gene results in polymerization of multiple

repeating units within the O6 side chains of strain Nissle 1917. The different promoter::wzy_{536} fusions may not be in accordance with the "optimal arrangement" of genes and the optimal distance between promoter and *wzy* translational start as it is in the *wb** operon. In order to clarify if wzy_{536} alone is able to restore fully the O6-antigen polymerization when introduced in $wb_{\text{Nissle 1917,}}$ allelic exchange was performed. The results (Fig. 27) indicated that the entire wb^*_{536} cluster is necessary to restore the O6 side chain polymerization in the same degree as in the *E. coli* strain 536 and CFT073, whereas the $wzy_{\text{Nissle 1917}}$ gene is sufficient to cause a semi-rough phenotype when introduced into $wb^*_{536.}$ Weaker expression of repeating units in all Nissle 1917 strains complemented with wzy_{536} alone, compared to strains complemented with the plasmid-encoded *wb** determinant of strain 536, may be indicative of other factors which impair proper *wb** expression in strain Nissle 1917.

Since the O6 serotype is widely distributed among wild type (pathogenic and non-pathogenic) *E. coli* strains, we studied the importance of full length O6 LPS side chains for serum resistance of different derivatives of *E. coli* strain Nissle 1917 (Fig. 24b) as this trait represents an important biological feature of this therapeutically used *E. coli* strain with respect to its biosafety. The results showed that the O6 antigen expression contributes to serum resistance of these strains. Serum resistance of *E. coli* Nissle 1917 was quantitatively related to the amounts of O6 antigen. This example underlines the impact of point mutations for evolution of E*nterobacteriaceae* in addition to DNA rearrangements as well as to acquisition and deletion of large genetic determinants. Recent studies showed that the O antigen chain length is critical for the virulence properties of *Salmonella* (Murray et al., 2002), thus supporting the presumed contribution of the Nissle 1971 semi-rough phenotype to the biological safety and non-pathogenicity of this strain. The O antigen synthesis and chain-length regulation is also known to be important for the *Shigella flexneri* invasion and virulence (Hong and Payne, 1997).

4. Structural analysis and sequence comparison of GEI $I_{\text{Nissle 1917}}$

The GEI $I_{\text{Nissle 1917}}$ was identified by screening for a cosmid clone coding for microcins. The same clone encoded F1C fimbriae, and it was concluded that since two fitness factors which are not part of the *E. coli* K-12 core genome are clustered together, this cosmid contains genetic information which might represent a part of a genomic island. This island was subsequently sequenced and exhibits genetic

organization which is very similar to that of the *serX*-associated island of UPEC strain CFT073. The PAI III_{536} contains the S-fimbriae-encoding determinant, as well as the entire *iro* gene cluster and some microcin-related sequences. Thus, GEI $I_{Nissle\ 1917}$ and PAI III_{536} exhibit some common structural features and represent examples of one and the same "S-fimbrial" island family (Dobrindt et al., 2002; Fig. 10). Nevertheless, they differ significantly in their composition and in the fitness factors that are encoded (e.g., microcin). In addition, PAI III_{536} is integrated into the *thrW* tRNA gene, whereas the Nissle 1917-specific *thrW* gene is probably an integration site for another genomic island, since the tRNA screening revealed the same gene located downstream of *thrW* as in the *E. coli* strain CFT073 genome (*c0391*).

GEI $I_{Nissle\ 1917}$ represents a fitness island, since adhesion- , microcin- , and iron uptake-conferring determinants are present, but not any toxin-encoding genes. The microcin expression is probably involved in the antagonistic action of *E. coli* strain Nissle 1917, contributing to the successful competition with other bacteria present in the intestine during colonization and is therefore believed to be important for its probiotic effect.

5. Structural analysis and sequence comparison of GEI $II_{Nissle\ 1917}$

The identification of an aerobactin-encoding cosmid clone led to the discovery and characterization of GEI $II_{Nissle\ 1917}$. This island comprises a DNA region of about 100 kb and is integrated into the *pheV* tRNA gene, that is immediately followed by a P4-like integrase-encoding gene. GEI $II_{Nissle\ 1917}$ contains the determinants coding for aerobactin (*iuc*), a serine protease (*sat*), an Ag43-homologue (*sap*), an adhesin (*iha*), and the K5 capsule (*kps*), thus representing an island responsible for the expression of several important fitness traits. Many putative ORFs coding for hypothetical proteins with unknown functions are also located on GEI $II_{Nissle\ 1917}$, together with a surprisingly high number (21) of mobility-associated DNA elements, such as transposases and IS elements. This suggests that GEI $II_{Nissle\ 1917}$ has been acquired lately during the evolution of *E. coli* strain Nissle 1917 and is probably unstable, due to possible deletion events and rearrangements.

The expression of the aerobactin iron uptake system is important for the fitness of strain Nissle 1917 in the light of the limited availability of iron in the human body. Thus, together with the expression of other iron uptake systems (encoded by *ent, fec, ybt, chu*) it contributes to the successful survival and colonization. The aerobactin

gene cluster is widely distributed among pathogenic and non-pathogenic *Escherichia coli* and *Shigella spp.,* and is located on plasmids or on genomic islands of different structure, composition, and chromosomal localization (Braun et al., 1983; Purdy et al., 2001; McDougall et al., 1984; Vokes et al., 1999; Welch et al., 2002) . The *iha* gene codes for Iha, a putative non-fimbrial adherence-conferring molecule, which has initially been identified as putative siderophore receptor molecule present in the genomes of EHEC strains as well as in the genome of UPEC strain CFT073 (Tarr et al., 2000; Schmidt et al., 2001). It has been demonstrated that Iha is a non-functional adhesin in strain CFT073, but functional in the investigated EHEC strains.

The serine protease-encoding gene *sat* is located between the *iuc* gene cluster and *iha* in GEI $II_{Nissle\ 1917}$. The secretion of the Sat autotransporter protein and its protease activity were demonstrated. Interestingly, although this protein exhibits a cytopathic effect, it has been shown that its inactivation did not attenuate virulence of *E. coli* strain CFT073 (Guyer et al., 2002). That is why it cannot be considered a true virulence- , but a fitness factor. Another putative autotransporter gene (*sap*), is also located on GEI $II_{Nissle\ 1917}$ which codes for antigen 43-homologue. It is well known that Ag43 expression is important for biofilm formation of *E. coli* (Danese et al., 2000; Kjaergaard et al., 2000; Michaelis, K., and U. Dobrindt, personal communication). Therefore, Sap expression may contribute to cell-cell interactions and aggregation of *E. coli* strain Nissle 1917. The K5 capsule-encoding *kps* gene cluster is also located on GEI $II_{Nissle\ 1917}$ (Fig. 12) and consequently three putative adhesion factors are encoded on GEI $II_{Nissle\ 1917}$ (*iha, sap, kps*).

The *pheV*-associated GEI $II_{Nissle\ 1917}$ is very similar in its structure and DNA content to the *pheV*-associated island of *E. coli* strain CFT073, but also exhibits some important differences with respect to the genetic determinants located there, gene order, stability and presence of mobile genetic elements (Fig. 12). Whereas the K5-specific gene cluster is present in GEI $II_{Nissle\ 1917}$, the K2(?)-specific gene cluster is part of the *pheV* island of CFT073 (*pheV*-associated GEI $_{CFT073}$). Although the genes *kpsF, E, D, U* (region 1) and *kpsM, T* (region 3) are highly homologous and identically organized in GEI $II_{Nissle\ 1917}$ and *pheV*-associated GEI $_{CFT073}$, the two *kps* clusters differ in region 2 (*kfi* genes) exhibit no sequence homology. The localization of both *kps* gene clusters with respect to the DNA context in the “CFT073 backbone“ is identical in both strains.

Another striking difference between the two islands is the organization and DNA

content of the *iuc/sat/iha*-region. The genomic islands are often able to delete entirely or partially from the chromosome, due to homologous recombination or with the participation of different mobile genetic elements (Turner et al., 2001). In GEI $II_{Nissle\ 1917}$ these three determinants are located within a possibly deletable DNA region that is flanked by two IS*2* elements with opposite orientation and that also contains three IS/transposase-related ORFs and two putative ORFs of unknown function (IS*2*-*iuc*-*sat*-ORFs-*iha*-IS*2*; Fig. 15a; see section VIII, Tab. 3). The spontaneous deletion of this DNA region was shown by exclusion PCR and sequence analysis of the resulting PCR product (Fig. 16). Such a deletion event presumably results in aerobactin/Sat/Iha-negative Nissle 1917 derivative, whose fitness and colonization capacity is likely to be reduced. On other hand, the deletion of *sat* would abolish the cytotoxic activity of the Sat protease. In comparison, in *pheV*-associated GEI $_{CFT073}$ only one copy of the IS*2* elements is present and the gene order is different (*iha*-IS*2*-*sat*-*iuc*; Fig. 13). The *sat* and *iuc* genes have a reversed orientation compared to GEI $II_{Nissle\ 1917}$ and the five ORFs located downstream of *iuc* are not present in *pheV*-associated GEI $_{CFT073.}$ In this way, the *iuc/sat/iha* genes are stable, since they are not flanked by IS*2*. One can speculate that the immobilization of *iuc* and *sat* in *pheV*-associated GEI $_{CFT073}$ might have played a role for the evolution of *E. coli* strain CFT073, increasing its pathogenic potential.

In the DNA region between *pheV* and *iha* in *pheV*-associated GEI $_{CFT073,}$ the α-haemolysin-encoding *hly* gene cluster and the P-fimbriae-encoding *pap* operon are located, together with other ORFs coding for proteins of unknown or putative function. The *hly* and *pap* determinants comprise a 30 kb region, which presumably is of crucial importance for the virulence properties of *E. coli* strain CFT073 (Hull et al., 1994; Johanson et al., 1992; Brauner et al., 1995; Hacker et al., 1986; Linggood and Ingram, 1982). Interestingly, only remnants of the *pap* operon are present in the same DNA context in GEI $II_{Nissle\ 1917.}$ Instead of the 30 kb *hly/pap* region only a fragment of *papA* (*papA'*) and the intact *papI* gene are present. Immediately upstream of *papA'* a transposon-like element is present. This Tn*10*-like structure (Fig. 15b; see section VIII. Tab. 3) consists of two IS*10* elements in opposite orientation, flanking a transposase-encoding gene (*yedA* transposase of Tn*10*). The transposase gene is adjacent to remnant DNA sequences of the *papX* gene. It is tempting to speculate that the intact *pap* gene cluster had initially been located on GEI $II_{Nissle\ 1917.}$ Later in the evolution of strain Nissle 1917 it has been disrupted due to insertion of IS*10*/Tn*10*.

Having the similar organization and DNA content of GEI $II_{Nissle\ 1917}$ and *pheV*-associated GEI $_{CFT073}$ in mind, one cannot exclude the possibility that the *hly* gene cluster has also been deleted from GEI $II_{Nissle\ 1917}$ due to the insertion of the Tn*10*-like element. Since this insertion is responsible for the inactivation of the P-fimbrial and probably of the α-haemolysin expression, it represents an important evolutionary event, dramatically reducing the virulence capacity of the ancestral proto-Nissle 1917 strain. The 4 kb Tn*10*-like structure is putative mobile genetic element which may still contribute to the dynamic of genome plasticity of strain Nissle 1917.

Interestingly, the *pheV*-associated island PAI V_{536} of UPEC strain 536 contains the *kps* gene cluster, coding for the K15 capsule expression (Schneider, G., unpublished data). Immediately downstream of *pheV* identical P4 integrase-encoding gene (*c3556*) as in GEI $II_{Nissle\ 1917}$ and *pheV*-associated GEI $_{CFT073}$ is located. Nevertheless, GEI $II_{Nissle\ 1917}$ and PAI V_{536} exhibit no sequence similarity with exception of the conserved *kps* region 1 and region 2 genes. Thus, although GEI $II_{Nissle\ 1917}$ and PAI V_{536} might be considered members of one and the same family of *pheV*-associated capsule-encoding islands of *E. coli* spp. (including *pheV*-associated GEI $_{CFT073}$), they seem to be evolutionary more distant from each other, than GEI $II_{Nissle\ 1917}$ and *pheV*-associated GEI $_{CFT073}$. It is noteworthy that a putative *pap*-like fimbrial gene cluster (*pix*) is present on PAI V_{536}, although it exhibits homology to *pap* only on protein, but not on DNA level, and most likely it is not functional (Lugering et al.,, 2003; Schneider, G., unpublished data).

6. Initial characterization of determinants involved in biofilm formation of *E. coli* strain Nissle 1917

The biofilm formation of *E. coli* is due to the expression of specific genetic determinants which contribute to adhesion to inert surfaces and multicellular behaviour of bacteria. The exact mechanism contributing to and importance of biofilm formation for colonization of the host or survival in the environment need to be further investigated. It is well known that type 1 and curli fimbrial adhesion expression is important for the adherence to abiotic surfaces (Cookson et al., 2002). Therefore we revealed the complete sequence of the type 1 fimbrial determinant of strain Nissle 1917. It is already known that strain Nissle 1917 exhibits an altered type 1 fimbriation due to a truncation of the *fimB* recombinase gene due to an insertion

element. This *fimB* inactivation results in an abolished expression of type 1 fimbriae under shaking growth conditions in batch cultures, but does not influence the expression under static conditions growth (Stentebjerg-Olesen et al., 1999). Five different copies of *fimBE*-like genes have been identified in the genome of *E. coli* strain CFT073, suggesting that phase variation by inversion of a DNA region may occur at other regions than the type 1 fimbrial determinant in strain Nissle 1917. The manner of the type 1 fimbrial expression in *E. coli* Nissle 1917 is probably of significance for the biofilm formation on abiotic surfaces and especially on eukaryotic cell surfaces. It is also known that variations in the FimH adhesin amino acid sequence influence this process (Boddicker et al., 2002).

Expression of curli fimbriae and cellulose is of crucial importance for biofilm formation and multicellular behavior of *E. coli* spp. (Römling et al., 1998; Solano et al., 2002). The *E. coli* strain Nissle 1917 expresses the so-called rdar morphotype forming typical tight, dry–looking colonies on agar plates since it co-expresses curli fimbriae and cellulose in a temperature-independent manner,. The rdar colonies are able to absorb the dye congo red,, adhere better to plastic surfaces and glass, and form a pellicle with a tight bacterial network at the air-liquid surface when grown as a static culture in rich medium at room temperature. It is not known if this phenotype contributes to the successful colonization of the host, but it probably plays a role for the survival of *E. coli* outside the host and may be associated with the probiotic character of strain Nissle 1917. Inactivation of the cellulose biosynthesis-encoding $bcs_{\text{Nissle 1917}}$ gene cluster abolished the rdar morphotype, confirming the critical role of the cellulose biosynthesis for biofilm/rdar morphotype formation. Interestingly, the rdar morphotype is only rarely distributed among pathogenic and non-pathogenic *E. coli* isolates. The UPEC strain 536 expresses curli fimbriae and cellulose only at 30°C (bdar morphotype), whereas strain CFT073 does not express curli fimbriae even at 30°C (saw morphotype: smooth and white). Thus, it is tempting to speculate that strain Nissle 1917 is more adapted for survival outside the host than pathogenic isolates. The fact that the *bcs* and *csg* operons are widely distributed among non-pathogenic and pathogenic *E. coli* strains, that they are part of the core genome but their expression differs dramatically, mirrors (together with the identified frame-shift mutation in the *wzy* gene) a process of accumulation of vertically acquired DNA regions and sequence alterations, which contribute to the evolution of the non-pathogenic, successful colonizing *E. coli* strain Nissle 1917.

Interestingly, a spontaneous loss of the rdar morphotype of Nissle 1917, when cultivated on congo red agar plates has been observed (Nagy, G., Grozdanov, L., Dobrindt, U., unpublished observation; Fig. 30).

In addition, the expression of several other determinants located on the chromosome may contribute to biofilm formation, e. g. that of Ag43 homologues (e.g., *sap*, located on GEI $II_{Nissle\ 1917}$)(Kjaergaard et al., 2000). Since the expression of the Ag43-like homologues is upregulated upon inactivation of the transcriptional antiterminator RfaH protein, a mutant strain Nissle 1917*rfaH*- was created. Increased cell-to-cell interaction, as well as cell-surface aggregation was observed in the strain 536*rfaH*- and Nissle 1917*rfaH*- during cultivation in microfermentors (Dobrindt et al., unpublished data).

7. The genome of the *E. coli* strain Nissle 1917

In this study, the organization of the fitness determinants of the *E. coli* strain Nissle 1917 was revealed. These data were combined with two whole-genome approaches: PFGE and PCR-based tRNA screening. Thus, an integrated picture of the *E. coli* strain Nissle 1917 genome was created, allowing to speculate on the evolution of this strain (Fig. 33).

All together, more than 350 kb of the *E. coli* strain Nissle 1917 genome were sequenced (Tab. 9), including the known fitness determinants. Three tRNA genes-associated genomic islands were identified and completely characterized, on which most of the fitness determinants are clustered together. Several new putative fitness determinants were identified within these islands (*sat, iha, shiA, iro*) as well as many putative ORFs coding for hypothetical proteins of unknown functions (Tab. 10).

Table 9: Compilation of the fitness determinants of strain Nissle 1917 analyzed in this study.

Features, traits encoded	Construct	Determinants	Size [kb]
Type 1 fimbriae	pCos3YC6	*fim*	36.4
R1 core LPS	pCos8YA1	*wa**, *selC, int*	47.5
GEI $III_{Nissle1917}$: autotransporters	pCos1YA7	*argW*	37.9
GEI $II_{Nissle1917}$: aerobactin, serine protease, put. adhesin, K5 capsule	pCos9YB4	*shiA, iuc, sat*	41.4
	pCos2YE4	*sat, iha*	44.1
	pCos2RF2	*kps, sap*	39
GEI $I_{Nissle1917}$: F1C fimbriae, microcin	pCos3YE4	*foc, mcm/mch, iro*	48.2
	pCos2RA4	put. ORFs	38.9
O6 side-chain LPS	pGBW1917	wb_{O6}	11.8
Cellulose biosynthesis	pGBC1917	*bcs*	10.9

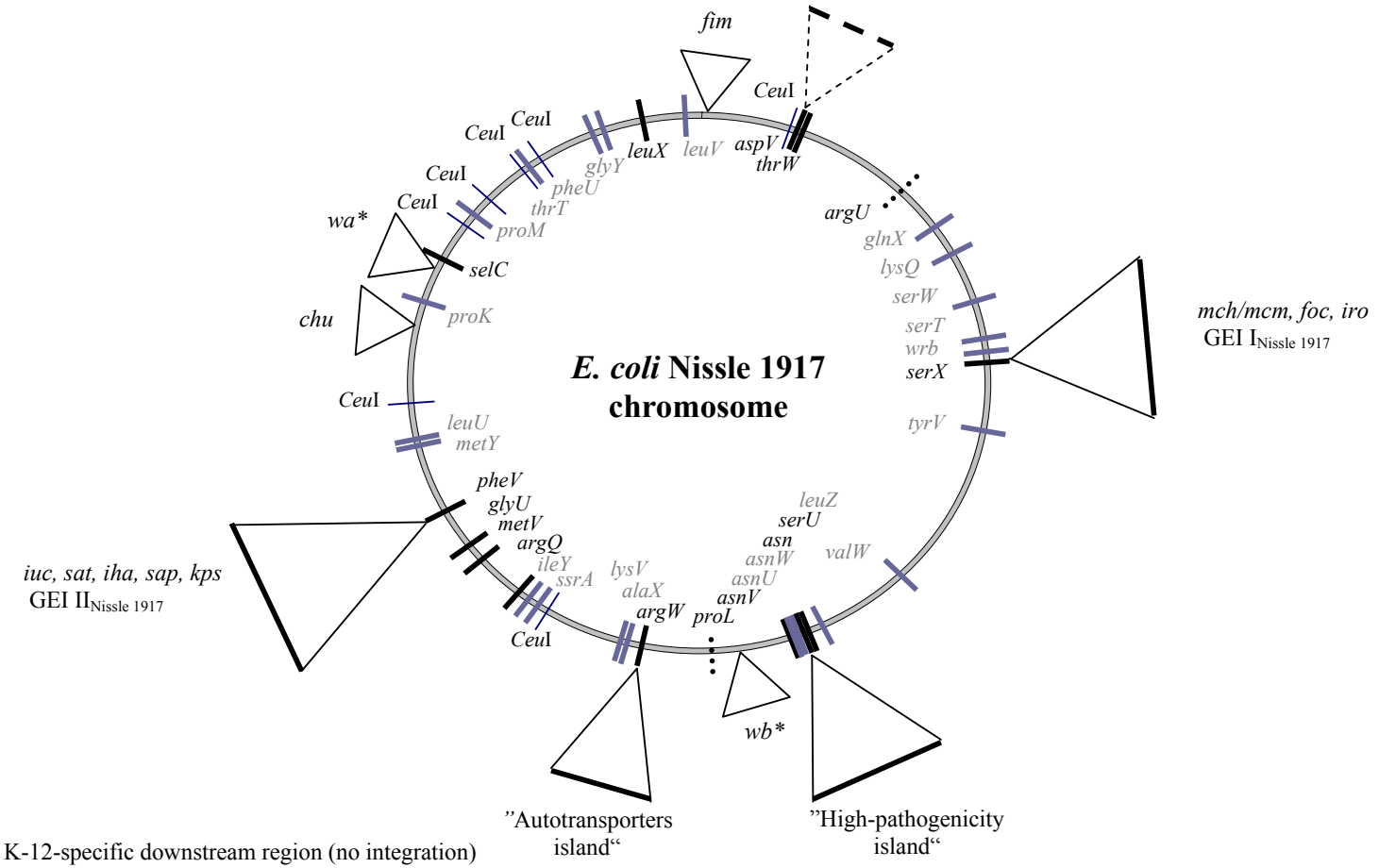

Fig. 33: Integrated map of the *E. coli* strain Nissle 1917chromosome. The position of the putative insertion sites and DNA content of the investigated genomic islands (GEIs $_{\text{Nissle 1917}}$) and islets is depicted . The localization of the tRNA genes screened for the presence of non-K-12 DNA sequences as well as the positions of the *Ceu*I restriction sites are shown . The presence of putative *aspV*-associated genomic island is indicated with interrupted lines.

Table 10: Number and characteristics of the ORFs identified in the *E. coli* strain Nissle 1917-specific DNA sequences characterized in this study.

cosmid/ plasmid/ island	ORFs/deleted known genes or pseudogenes	tRNA genes	autotransporters	cell structure-related	hypothetical protein; unknown	regulation	Tn/IS	fimbriae/ adhesion	nucleotide metabolism; DNA/RNA restr., modification; methylases	iron uptake	microcin-related	enzymes; methabolism	transport	LPS/capsule synthesis
pCos8YA1	43/2	1	-	-	7	3	1	-	11	-	-	5	3	12
pBWB1917	9/1	-	-	-	-	-	-	-	-	-	-	-	-	9
pCos1YA7	33	1	1	1	13	1	-	5	1	-	-	6	4	-
GEI $I_{Nissle\ 1917}$	68/1	-	-	-	20	-	6	11	1	5	10	14	1	-
GEI $II_{Nissle\ 1917}$	82/3	-	2	2	26	2	21	5	-	5	-	4	1	14
pCos3YC6	31	-	-	-	5	3	4	8	2	-	-	6	3	-
pGBC1917	6	-	-	-	-	-	-	-	-	-	-	6	-	-
all together:	272/6	2	3	3	71	9	32	29	15	10	10	41	12	35

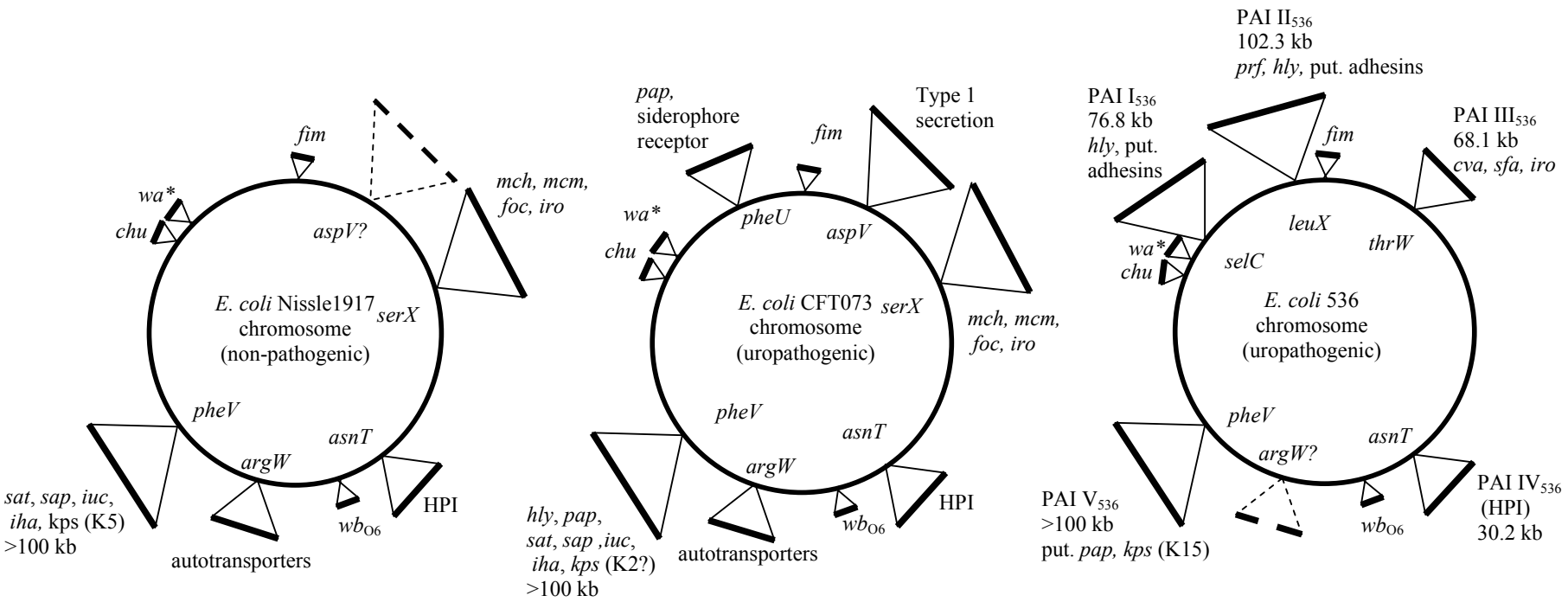

Fig. 34: Comparative genomic maps of the non-pathogenic and pathogenic *E. coli* O6 strains Nissle 1917, CFT073 and 536. The chromosomal localization of the major genomic islands and islets is depicted. The interrupted lines indicate the presence of islands whose DNA content is still not known.

The genomic organization of *E. coli* Nissle 1917 has been compared with that of the very well studied uropathogenic strains 536 (Dobrindt et al., 2002) and the completely sequenced strain CFT073 (Welch et al., 2002; Fig. 34). The strain 536 was isolated from a patient with pyelonephritis (Blum et al., 1994) and the strain CFT073 was isolated from the blood of a patient with acute pyelonephritis (Mobley et al., 1990), whereas the strain Nissle 1917 is a faecal isolate. Interestingly, the strain Nissle 1917 shares striking similarities concerning the chromosomal structure and organization of the fitness determinants (Fig. 34) with the strain CFT073. Having the PFGE data in mind, it is tempting to speculate that these two strains have a common clonal origin. Although the genetic structure of the investigated genomic islands on which the important fitness-conferring factors are encoded is very similar, the DNA content differs: no α-haemolysin and P-related fimbriae are encoded on the chromosome of strain Nissle 1917. On the contrary, the *hly* and *pap* gene clusters are located on the *pheV*-associated GEI $_{CFT073}$ and an additional *pap*-encoding island is integrated next to the *pheU* tRNA gene in the genome of strain CFT073. The *hly* and *pap* genes seem to be replaced by a transposon-like genetic structure in GEI $II_{Nissle\ 1917,}$ whereas the entire *pheU*-associated GEI_{CFT073} island seems to be absent from the genomes of the strains Nissle 1917 and 536. The GEI $I_{Nissle\ 1917}$ is almost completely identical to *serX*-associated GEI_{CFT073} island, and probably considerably contributes to the fitness of both strains during colonization. From the results of the PCR-based tRNA screening it is evident that most of the tRNA-associated foreign DNA elements present in the genome of strain CFT073 are also present in the genome of strain Nissle 1917. Despite of the similarity in the localization and structure of the horizontally acquired foreign DNA between the *E. coli* strains Nissle 1917 and CFT073 revealed by the tRNA-screening and the sequence analysis, one should have in mind that the DNA content of the genomic islands is not the same. For instance, highly homologous P4-like prophage integrases are present downstream of *selC* tRNA-encoding gene in the genomes of *E. coli* strains EDL933, CFT073 and 536, but the *selC*-assotiated islands in those strains differ completely. The same is true for the *pheV*-associated islands of those strains.

The second biggest genomic island of strain CFT073 located at *aspV* is at least partially present in strain Nissle 1917 (since the gene located immediately downstream of *aspV* is identical in both strains), but absent in strain 536. In spite of these similarities between the two strains, some differences were identified by screening the sequence context of tRNA genes, such as the absence of the *pheU*-associated island in the genome of strain Nissle 1917.

It is remarkable that the organization and structure of the pathogenicity-related determinants of UPEC strains CFT073 and 536 differ considerably with regard to their localization and DNA content. The analogous "island pairs" *pheV*-associated GEI $_{CFT073}$ - PAI V_{536} and *serX*-associated GEI_{CFT073} - PAI III_{536} have different chromosomal integration sites and also differ with respect to the presence of some additional fitness determinants in the strain CFT073 (*iuc, mch/mcm, sat, iha*; Fig. 34) which are absent in *E. coli* strain 536. The functional integrity of the virulence factors expressed from both UPEC strains, which are absent in the non-pathogenic Nissle 1917, is evident: α-haemolysin (encoded on PAI $I_{536,}$ PAI II_{536}, and *pheV*-associated GEI $_{CFT073}$), P-related fimbriae (encoded on PAI $II_{536,}$ *pheV*-associated GEI $_{CFT073,}$ and *pheU*-associated GEI_{CFT073}), and serum resistance-conferring full-length O6 LPS side-chain expression (*galF*-island). Analogous islands coding for adhesins of the S-fimbrial family are present in all three strains (*serX*-located islands of strains CFT073, Nissle 1917 and PAI III_{536} of strain 536). Since there is no clear evidence that S- or F1C-fimbrial expression alone is contributing to virulence, one could consider these adhesins fitness factors playing a role for colonization. One should also not exclude the possibility that some of the putative ORFs are coding for hypothetical proteins which are also important for fitness or virulence properties of the investigated strains.

It is notworthy that the non-pathogenic *E. coli* strain Nissle 1917 strain possesses a surprisingly large set of iron uptake systems (*iuc, ybt, fec, ent, chu, iro*) compared to the UPEC strains, which might significantly contribute to its fitness.

To further substantiate the analysis of the genome structure of strain Nissle 1917, the overall genome content was analyzed by DNA-DNA hybridization experiments with *E. coli* K-12-specific arrays, as well as with the "*E. coli* Pathoarray", thus enabling detection of virulence- and fitness-associated genes of pathogenic *E. coli* from different

pathogroups (Dobrindt, U., and Plaschke, B., personal communication). The results demonstrate again that e.g., adhesins, iron uptake systems and proteases do not necessarily have to be considered as virulence-associated factors but can also contribute to the fitness and adaptability of bacteria. Although many DNA regions which belong to the flexible gene pool of pathogenic *E. coli* or at least homologoues thereof can be detected in the genome of non-pathogenic strain Nissle 1917, those coding for important virulence factors of uropathogenic *E. coli* are absent (protein toxin- or P fimbriae-encoding genes).

Having in mind the data on the genomic organization of the *E. coli* strains Nissle 1917, CFT073 and 536 it is tempting to speculate on the evolution of uropathogenicity and "non-pathogenicity" of these *E. coli* strains (Fig. 35). It seems that strain Nissle 1917 acquired a set of important fitness determinants during evolution, which guarantees successful survival and colonization of the human body and most likely represents the genetic background for the probiotic action of this *E. coli* strain. At the same time no important virulence factors were identified, suggesting that either "by good fortune" such genetic factors were not horizontally transfered to the strain Nissle 1917, or they were present but subsequently deleted from the chromosome during adaptation as a non-virulent successful colonizer of the human intestine. Additionally, the strain Nissle 1917 exhibits some phenotypic features which are not typical for the UPEC isolates 536 and CFT073: the temperature-independent rdar morphotype, a non-FimB dependent type 1 fimbrial switch, the presence of two plasmids, a semi-rough LPS resulting in serum sensitivity. These features might mirror the process of evolutionary withdrawing from the pathogenic lineages, probably due to adaptation to the non-pathogenic lifestyle or survival in the environment outside the human body. The striking similarities of the genome structure between the strain Nissle 1917 and the virulent UPEC strain CFT073 suggest that these strains have a common evolutionary origin from an ancestral strain, but at some time point have been separated and adapted to different lifestyles (Fig. 35).

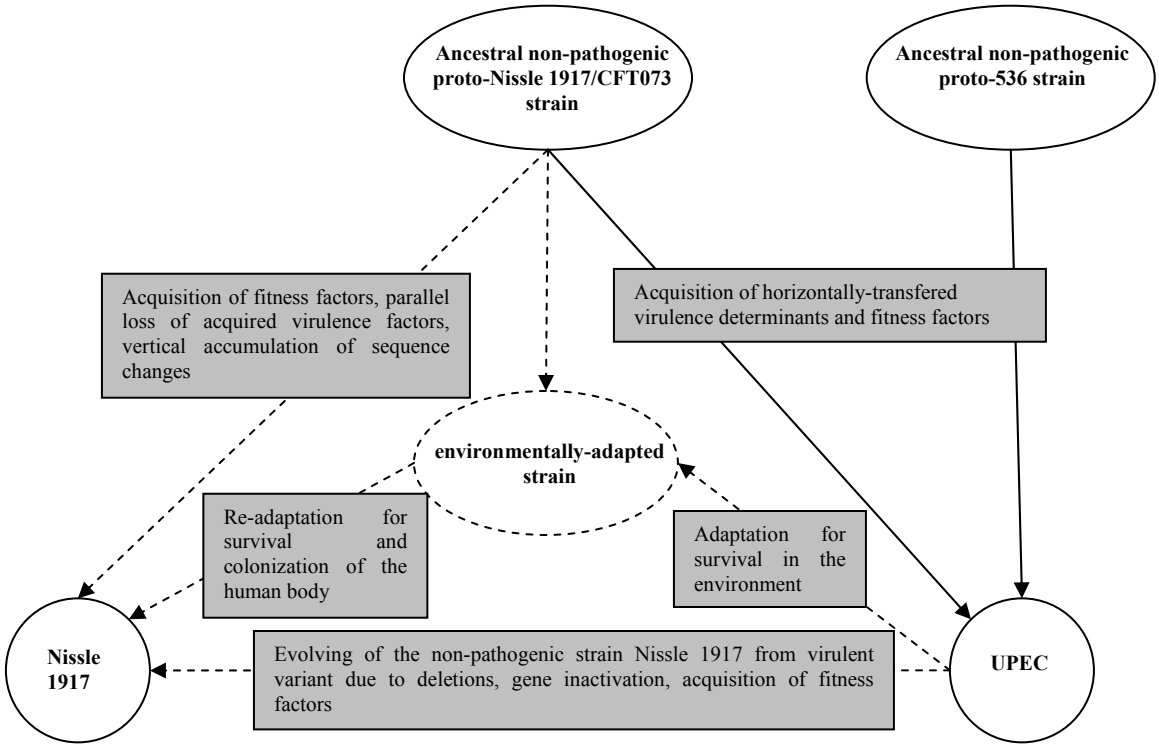

Fig. 35: Schematic representation of the presumed evolutionary relationships between the investigated *E. coli* O6 strains. The interrupted arrows represent possible ways for evolution of the non-pathogenic strain Nissle 1917. The interrupted circle indicates the putative presence of an environmentally-adapted proto-Nissle 1917 variant.

The *E. coli* strains 536 and CFT073 differ significantly in their structural organization of their virulence determinants, but at the end of their, probably independent and separate, evolution is a similar virulence capacity and a type of disease they cause.

Taken all together, the results of this study show that the *E. coli* strain Nissle 1917 exhibits a specific combination of phenotypic features due to acquisition of foreign DNA elements, loss of genetic information and minor genetic events (e. g., point mutations), which contribute to its probiotic nature and distinguish this strain from UPEC O6 strains as well as from other non-pathogenic *E. coli* isolates studied so far.

VII. References

Aktories, K. 1997. Rho properties: targets for bacterial toxins. Trends Microbiol. **5:**282-288.

Alm, R. A., Ling, L.-S. L., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., deJonge, B. L., Carmel, G., Tummino, P. J., Caruso, A., Uria-Nickelsen, M., Mills, D. M., Ives, C., Gibson, R., Merberg, D., Mills, S. D., Jiang, Q., Taylor, D. E., Vovis, G. F. and Trust, T. J. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature **397:**176-180.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipmann. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Res. **25:**3389–3402.

Amor, K., D. E. Heinrichs, E. Fridrich, K. Ziebell, R. P. Johnson, and C. Whitfield. 2000. Distribution of Core Oligosaccharide Types in Lipopolysaccharides from *Escherichia coli.* Infect. Immun. **68:**116–1124.

Andreu, A., A. E. Stapleton, C. Fennell, H. A. Lockman, M. Xercavins, F. Fernandez, and W. E. Stamm. 1997. Urovirulence determinants in *Escherichia coli* strains causing prostatitis. J. Infect. Dis. **176:**464-469.

Appelmelk, B. J., Y.-Q. An, T. A. M. Hekker, L. G. Thijs, D. M. MacLaren, and J. De Graaf. 1994. Frequencies of lipopolysaccharide core types in *Escherichia coli* strains from bacteraemic patients. Microbiology **140:**1119-1124.

Bach, S., C. Buchrieser, M. Prentice, A. Guyioule, T. Msadek, and E. Carniel. 1999. The high-pathogenicity island of *Yersinia enterocolitica* Ye8081 undergoes low-frequency deletion but not precise excision suggesting recent stabilization in the genome. Infect. Immun. **67:**5091-5099.

Bacheller, C. D., and J. M. Bernstein. Urinary tract infections. Med. Clin. North Am. **81:**719-730.

Barnett, B. J., and D. S. Stephens. 1997. Urinary tract infections: an overview. Am. J. Med. Sci. **341:**245-249.

Barua, S., T. Yamashino, T. Hasegawa, K. Yokoyama, K. Torii, and M. Ohta. 2002. Involvement of surface polysaccharides in the organic acids resistance of Shiga-producing *Escherichia coli* O157:H7. Mol. Microbiol. **43:**629-640.

Baumler, A. J., R. M. Tsolis, A. W. van der Velden, I. Stojiljkovic, S. Anic, and F. Heffron. 1996. Identification of a new iron regulated locus of *Salmonella typhi.* Gene **183:**207-213.

Berger, H., J. Hacker, A. Juarez, C. Hughes, and W. Goebel. 1982. Cloning of the chromosomal determinants encoding haemolysin production and mannose-resistant hemagglutinanion in *Escherichia coli.* J. Bacteriol. **152:**1241-1247.

Bergthorsson, U., and H. Ochman. 1998. Distribution of chromosome length variation in natural isolates of *Escherichia coli* K-12. Mol. Biol. Evol. **15:**6-16.

Berlyn, M. K. B. 1998. Linkage map of *Escherichia coli* K-12, edition 10: the traditional map. Microbiol. and Mol. Biol. Rev. **62:**814-984.

Bettelheim, K. A. 1997. *Escherichia coli* in the normal flora of humans and animals, p. 85-109. *In* M. Sussman (ed.), *Escherichia coli* – Mechanism of virulence. Cambridge University Press, Cambridge.

Blanc-Potard, A.-B., F. Solomon, J. Kayser, and E. A. Groisman. 1999. The SPI-3 pathogenicity island of *Salmonella enterica.* J. Bacteriol. **181:**998-1004.

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

Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschäpe, and J. Hacker. 1994. Excision of large DNA elements termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606-614.

Blum, G., M. Ott, A. Lischewski, A. Ritter, I. Imrich, H. Tschäpe, and J. Hacker. 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606-614.

Blum, G., R. Marre, and J. Hacker. Properties of *Escherichia coli* strains of serotype O6. Infection. **23:** 243- 236.

Blum, G., V. Falbo, A. Caprioli, and J. Hacker. 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606-614.

Blum-Oehler, G., S. Oswald, K. Eiteljörge, U. Sonnenborn, J. Schulze, W. Kruis, and J. Hacker. 2003. Development of strain-specific PCR reactions for the detection of the probiotic *Escherichia coli* strain Nissle 1917 in fecal samples. Res. Microbiol. **154:**59-66.

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

Boyd, E. F., and D. L. Hartl. 1997. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. J. Bacteriol. **180:**1159-1165.

Braun, V., R. Gross, W. Köster, and L. Zimmermann. 1983. Plasmid and chromosomal mutants in the iron (III)-aerobactin transport system of *Escherichia coli*. Use of Streptonigrin for selection. Mol. Gen. Genet. **192:**131-139.

Brauner, A., M. Katouli, and C. G. Ostenson. 1995. P-fimbriation and haemolysin production are the most important virulence factors in diabetic patients with *Escherichia coli* bacteremia: a multivariate statistical analysis of seven virulence factors. J. Infect. **31:**27-31.

Brosius, J., R. L. Cate, and A. P. Perlmutter. 1982. Precise location of two promoters for the β-lactamase gene of pBR322. J**.** Biol. Chem. **257:**9205-9210.

Buchrieser, C., C. Rusinok, L. Frangeul, E. Couve, A. Billaut, F. Kunst, E. Carniel, and P. Glaser. 1999. The 102-kilobase *pgm* locus of *Yersinia pestis*:

sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. Infect. Immun. **67:**606-614.

Buchrieser, C., R. Brosch, S. Bach, A. Guiyoule, and E. Carniel. 1998. The high pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three *asn* tRNA genes. Mol. Microbiol. **30:**965-978.

Burns, S. M., and S. I. Hull. 1998. Comparison of Loss of Serum Resistance by Defined Lipopolysaccharide Mutants and an Acapsular Mutant of Uropathogenic *Escherichia coli* O75:K5. Infect. Immun. **66:**4244–4523.

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

Cheetham, B. F., and M. E. Katz. 1995. A role of bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbio. **18:**108-201.

chromosomal and plasmid-encoded lipopolysaccharide genes on invasion and serum resistance. Mol. Microbiol. **24:**779-792.

Coimbra, R. S., F. Grimont, P. Lenormand, P. Burguiere, L. Beutin, and P. A. Grimont. 1998. Identification of *E. coli* serogroups by restriction of the amplified O-antigen gene cluster. Res. Microbiol. **151:**639-54.

Comayras, C., C. Tasca, S. Y. Peres, B. Ducommun, E. Oswald, and J. de Rycke. 1997. *Escherichia coli* cytolethal distending toxin blocks the HeLa cycle at G2/M transition by preventing cdc2 protein kinase dephosphorylation and activation. Infect. Immun. **65:**5088-5059.

Cookson, A. L., W. A. Cooley, and M. J. Woodward. 2002. The role of type 1 and curli fimbriae of Shiga toxin-producing *Escherichia coli* in adherence to abiotic surfaces. Int. J. Med. Microbiol. **292:**195-205.

Danese, P. N., L. A. Pratt, S. L. Dove, and R. Kolter. 2000. Antigen 43 from *Escherichia coli* induces inter- and intraspecies cell aggregation and changes in colony morphology of *Pseudomonas fluorescens.* J. Bacteriol. **182:**4789-4796.

Danese, P. N., L. A. Pratt, S. L. Dove, and R. Kolter. 2000. Antigen 43 from *Escherichia coli* induces inter- and intraspecies cell aggregation and changes in colony morphology of *Pseudomonas fluorescens.* J. Bacteriol. **182:**4789-4796.

Danese, P. N., L. A. Pratt, S. L. Dove, and R. Kolter. 2000. The outer membrane protein, antigen 43, mediates cell-to-cell interactions within *Escherichia coli* biofilms. Mol. Microbiol. **37:**424-432.

Danese, P. N., L. A. Pratt, S. L. Dove, and R. Kolter. 2000. The outer membrane protein, antigen 43, mediates cell-to-cell interactions within *Escherichia coli* biofilms. Mol. Microbiol. **37:**424-432.

de Louvois, J. 1994. Acute bacterial meningitis in the newborn. J. Antimicrob. Chemoter. **34 Suppl A:**61-73.

Diederich, L., L. J. Rasmussen, and W. Messer. 1992. New cloning vectors for integration into the λ attachment site *attB* of the *Escherichia coli* chromosome. Plasmid **28:**14-24.

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

Dobrindt, U., F. Agerer, K. Michaelis, A. Janka, C. Buchrieser, M. Samuelson, C. Svanborg, G. Gottschalk, H. Karch, and J. Hacker. 2003. Analysis of Genome

plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. J. Bacteriol. **185:**1831-1840.

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

Dobrindt, U., G. Blum-Oehler, T. Hartsch, G. Gottschalk, E. Z. Ron, R. Fünfstuck, and J. Hacker. 2001. S-fimbria-encoding determinant *sfa*I is located on pathogenicity island III_{536} of uropathogenic *Escherichia coli* strain 536. Infect. Immun. **69:**4248-4256.

Donnenberg, M. S., and J. B. Kaper. 1991. Construction of an *eae* deletion mutant of enteropathogenic Escherichia coli by using a positive-selection suicide vector. Infect. Immun. **59:**4310-4317.

Dutta, P. R., R. Capello, F. Navarro-Garcia, and J. P. Cataro. 2002. Functional comparison of serine protease autotransporters of *Enterobacteriaceae.* Infect. Immun. **70:**7105-7113.

Elwell, C., K. Chao, K. Patel, and L. Dreyfus. 2001. *Escherichia coli* CdtB mediates cytolethal distending toxin cell cycle arrest. Infect. Immun. **69:**3418-3422.

Ferrieres, L., and D. J. Clarke. 2003. The RcsC sensor kinase is required for normal biofilm formation in *Escherichia coli* K-12 and controls the expression of a regulator in response to growth on a solid surface. Mol. Microbiol. **50:**1665-1682.

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

Fleischmann, R. D., Adams, M.D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus infuenzae* Rd. Science **269:**496-512.

Gally, D. L., T. J. Rucker, I. C. Blomfield. 1994. The leucine–responsive regulatory protein binds to the *fim* switch to the control phase variation of type 1 fimbrial expression in *Escherichia coli.* J. Bacteriol. **176:**5665-5672.

Goldman, R. C., K. A. Joiner, and L. Leive. 1984. Serum-resistant mutants of *Escherichia coli* O111 contain increased lipopolysaccharide, lack an O antigen–containing capsule, and cover more of their lipid A core with O antigen. J. Bacteriol. **159:**877–882.

Gophna, U., M. Barlev, R. Seijffers, T. A. Oelschlaeger, J. Hacker and E. Z. Ron. 2001. Curli fibers mediate internalization of *Escherichia coli* by eukaryotic cells. Infect. Immun. **69:**2659-2665.

Grimberg, J., S. Maguire, and L. Belluscio. 1989. A simple method for the preparation of plasmid and chromosomal DNA. Nucl. Acids Res. **21:**8893.

Grossman, N., A. A. Lindberg, S. B. Svenson, K. A. Joiner, and L. Leive. 1991. Structural aspects of LPS: role in evasion of host defense mechanism, p. 143–149. *In* E. Z. Ron and S. Rotten (ed.), Microbial surface components and toxins in relation to pathogenesis. Plenum Press, New York.

Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., and U. Dobrindt. 2004. Analysis of the genome structure of probiotic *Escherichia coli* strain Nissle 1917. *Submitted.*

Grozdanov, L., Zähringer, U., Blum-Oehler, G., Brade, L., Henne, A., Knirel, Y. A., Schombel, U., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., Rietschel, E. T., and U. Dobrindt. 2003. A single nucleotide exchange in the *wzy* gene is responsible for the semirough O6 lipopolysaccharide phenotype and serum sensitivity of the *Escherichia coli* strain Nissle 1917. J. Bacteriol. **184:**5912-5925.

Guerinot, M. L. 1994. Microbial iron transport. Annu. Rev. Microbiol. **48:**743-772.

Guijarro, J. I., J. E. Gonzàlez-Pastor, F Baleux, J. L. San Millàn, M. A. Castilla, M. Rico, F. Moreno, and D. Delepierre. 1995. Chemical structure and translation inhibition studies of the antibiotic microcin C7. J. Biol. Chem. **270:** 23520-23532.

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

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

Hacker, J., and E. Carniel. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. EMBO Reports **21:**376-381.

Hacker, J., and J. B. Kaper. 1999. The concept of pathogenicity islands. In: Pathogenicity islands and other mobile virulence elements. ed. J. B. Kaper, J. Hacker, pp 1-11. Washington: ASM.

Hacker, J., H. Hof, L. Emödy, and W. Goebel. 1986. Influence of cloned hemolysin genes, S-fimbriae and serum resistance on pathogenicity in different animal models. Microb. Pathogen. **1:**533-547.

Hacker, J., S. Knapp, and W. Goebel. 1983. Spontaneous deletions and flanking regions of the chromosomally inherited hemolysin determinant of an *Escherichia coli* O6 strain. J. Bacteriol. **154:**1145-1152.

Hanna, A., Berg, M., Stout, V., and A. Razatos. 2003. Role of capsular colanic acid in adhesion of uropathogenic *Escherichia coli.* Appl. Environ. Microbiol. **69:**4474-4481.

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

Harrington, C. A., C. Rosenow, and J. Retief. 2000.Monitoring gene expression using DNA microarrays. Curr. Opinion Microbiol. **3:**285-291.

Hartley, C. L., C. S. Neumann, and M. H. Richmond. 1979. Adhesion of commensal bacteria to the large intestine wall in humans. Infect. Immun. **23:**128-132.

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

Heinrichs, D. E., J. A. Yethon, and C. Whitfield. 1998. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. Mol. Microbiol. **30 (2):**221-232.

Henderson, I. R., F. Navarro-Garcia, and J. P. Nataro. 1998. The great escape: structure and function of the autotransporter proteins. Trends Microbiol. **6:**370-378.

Hentschel, U., M. Steinert, and J. Hacker. 2000. Common molecular mechanisms of symbiosis and pathogenesis. Trends Microbiol. **8:**226-230.

Herias, M. V., T. Mivedt, L. A. Hanson, and A. E. Wold. 1997. *Escherichia coli* K5 capsule expression enhances colonization of the large intestine in the gnotobiotic rat. Infect. Immun. **65:**531-536.

Holland, I. B., Blight, M. A., and B. Kenny. 1990. The mechanism of secretion of hemolysin and other polypeptides from Gram-negative bacteria. J.Bioenerg. Biomembr. **22:**473-491.

Holst, O. 1999. Chemical structure of the core region of lipopolysaccharides, p. 115-154. *In* H. Brade, S. M. Opal, S. N. Vogel, and D. C. Morrison (eds.), Endotoxin in health and disease. Marcel Dekker, New York.

Hong, M., and S. M. Payne. 1997. Effect of mutations in *Shigella flexneri*

Hou, Y.-M. 1999. Transfer RNAs and pathogenicity islands. TIBS **24:**295-298.

Hughes, C., T. Phillips, and P. Roberts. 1982. Serum resistance among *Escherichia coli* strains causing urinary tract infection in relation to O type and the carriage of haemolysin, colicin, and antibiotic resistance determinants. Infect. Immun. **35:**270–275.

Hull, R. A., B. Nowicki, A. Kaul, R. Runyan, C. Svanborg, and S. I. Hull. 1994. Effect of *pap* copy number and receptor specificity on virulence of fimbriated *Escherichia coli* in a murine urinary tract colonization model. Microb. Pathog. **17:**79-86.

Hull, S. 1997. *Escherichia coli* lipopolysaccharide in pathogenesis and virulence, p. 145–167. *In* M. Sussman (ed.) *Escherichia coli*: mechanisms of virulence. Cambridge University Press, Cambridge, United Kingdom.

Ingersoll, M. A., J. E. Moss, Y. Weinrauch, P. E. Fischer, E. A. Groisman, and A. Zychlinsky. 2003. The ShiA protein encoded by the *Shigella flexneri* SHI-2 pathogenicity island attenuates inflamation. Cell. Microbiol. **5:**797-807.

Jacques, M. 1996. Role of lipo-oligosaccharides and lipopolysaccharides in bacterial adherence. Trends Microbiol. **4:**408-410.

Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H., Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J., Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S., Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y., Yu J. 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K-12 and O157. Nucleic Acids Res. **30:**4432-4441.

Johanson, I., R. Lindstedt, and C. Svanborg. 1992. Roles of the *pap*- and *prs*-encoded adhesins in *Escherichia coli* adherence to human uroepithelial cells. Infect. Immun. **60:**3416-3422.

Karch, H., S. Schubert, D. Zhang, W. Zhang, H. Schmidt, T. Ölschläger, and J. Hacker. 1999. A genomic island, termed high-pathogenicity island, is present in

certain non-O157 Shiga toxin-producing *Eschrichia coli* clonal lineages. Infect. Immun. **67:**5994-6001.

Karch, H., S. Schubert, D. Zhang, W. Zhang, H. Schmidt, T. Ölschläger, and J. Hacker. 1999. A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing *Eschrichia coli* clonal lineages. Infect. Immun. **67:**5994-6001.

Kauffmann, F. 1965. Supplement to the Kauffmann-White scheme. 8. Acta Pathol. Microbiol. Scand. **64:**362-366.

Kim, K. S. 2002. Strategy of Escherichia coli for crossing the blood-brain barrier. J. Infect. Dis. **186:**220-224.

Kjaergaard, K., M. A. Schembri, C. Ramos, S. Molin, and P. Klemm. 2000. Antigen 43 facilitates formation of multispecies biofilms. Environ. Microbiol. **2:**695-702.

Klumpp, D. J., A. C. Weiser, S. Sengupta, S. G. Forrestal, R. A. Batler, and A. J. Schaeffer. 2001. Uropathogenic *Escherichia coli* potentiates type 1 pilus-induced apoptosis by suppressing NF-kappaB. Infect. Immun. **69:**6689-6695.

Korhonen, K. T. 2000. Adherence of enteric bacteria to onto mammalian extracellular matrix. Test-tube artefact or a virulence function? Adv. Exp. Med. Biol. **485:**193-200.

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

Kruis, W., P. Fric, M. Stolte. 2001. The Mutaflor study group. Maintenance of remission in ulcerative colitis is equally effective with *Escherichia coli* Nissle 1917 and with standard mesalamine. Gastroenterology. **120:Supl** A127.

Kupfer, M. 1999. Molekularbiologische Untersuchungen zur *leuX*-abhängigen Expression von virulenzgenen bei pathogenen *Escherichia coli* Stämmen. Master thesis.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.

Lawley, T. D., V. Burland, and T. E. Taylor. 2000. Analysis of the complete nucleotide sequence of the tetracycline-resistance transposon *Tn10.* Plasmid **43:** 235-239.

Lawrence, J. G., and H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Nat. Acad. Sci. USA, **95:**9413-9417.

Lewis, A. L., N. Gadura, M. Greene, R. Saby, and N. D. F. Grindley. 2001. The basis of asymmetry in IS*2* transposition. Mol. Microbiol. **42:**887-901.

Lidin-Janson, G., B. Kaijser, K. Lincoln, S. Olling, and H. Wedel. 1978. The homogeneity of the faecal coliform flora of normal school-girls, characterized by serological and biochemical properties. Med. Microbiol. Immunol. **164:**247-253.

Lindsay, J. A., A. Ruzin, H. F. Ross, N. Kurepina, and R. P. Novick. 1998. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus.* Mol. Microbiol. **29:**527-543.

Linggood, M. A., and P. L. Ingram. 1982. The role of haemolysin in the virulence of *Escherichia coli* for mice. J. Med. Microbiol. **15:**23-30.

Lodnikova-Zadnikova, R., U. Sonnenborn. 1997. Effect of preventive administration of nonpathogenic *Escherichia coli* strain on the colonization of the intestine with microbial pathogens in newborn infants. Biol. Neonate **71:**224-232.

Low, D., V. David, D. Lark, G. Schoolnik, and S. Falkow. 1984. Gene cluster governing the production of hemolysin and mannose- resistant hemagglutination are closely linked in *Escherichia coli* serotype O4 and O6 isolates from urinary tract infections. Infect. Immun. **43:**353-358.

Luck, S. N., S. A. Turner, K. Rajakumar, H. Sakellaris, and B. Adler. 2001. Ferric dicitrate transport system (Fec) of *Shigella flexneri* 2a YSH6000 is located on a novel pathogenicity island carrying multiple antibiotic resistance genes. Infect. Immun. **69:**6012-6021.

Lugering, A., I. Benz, S. Knochenhauer, M. Ruffing, and M. A. Schmidt. 2003. The Pix pilus adhesin of the uropathogenic *Escherichia coli* strain X2194 (O2:K(-):H6) is related to Pap pili but exhibits truncated regulatory region. Microbiology **149:**1387-1397.

Lukowski, S., A. R. Hull, and S. I. Hull. 1996. Identification of the O antigen (*rfc*) gene in *Escherichia coli* O4 by insertional mutagenesis using a non polar chloramphenicol resistance cassette. J. Bacteriol. **178:**240–247.

Lyla, J. Meklerson-Watson, C. K. Rode, L. Zhang**, B. Foxman and C. A. Bloch.** 2000. Integrated genomic map from uropathogenic *Escherichia coli* J96. Infect. Immun. **68:**5933-5942.

Malchow, H. 1997. Crohn`s disease and *E. coli*: A new approach in therapy to maintain remission of colonic Crohn`s disease? J. Clin. Gastroenterol. **15:**653-658.

Malchow, H., U. Sonnenborn, R. Greinwald, A. Körner. 1995. Colonization of adults by a non-pathogenic *Escherichia coli* strain. Gastroenterology. **108:Suppl** A869.

Marolda, C. L., M. F. Feldman, and M. A. Valvano. 1999. Genetic organization of the O7-specific lipopolysaccharide biosynthesis cluster of *Escherichia coli* VW187 (O7:K1). Microbiology **145:**2485-95.

Marshall, A., and J. Hodgson. 1998. DNA chips: an array of possibilities. Nat. Biotechnol. **16:**27-31.

Martinez, J. J., M. A. Mulvay, J. D. Scilling, J. S. Pinkner, and S. J. Hultgren. 2000. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. EMBO J. **19:**2803-2812.

Maurelli, T. A., R. E. Fernandez, C. A. Bloch, C. K. Rode, and A. Fasano. 1998. "Black holes" and bacterial pathogenicity: A large genomic deletion that enhances the virulence of *Shigella spp.* and enteroinvasive *Escherichia coli.* Proc. Natl. Acad. Sci. USA, **95:**3943-3948.

McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg, and J. B. Kaper. 1995. A genetic locus of enterocyte effacement conserved among diverse bacterial pathogens. Proc. Natl. Acad. Sci. USA, **92:**1664-1668.

McDougall, A., and J. B. Mekalanos. 1984. Plasmid- and chromosome coded aerobactin synthesis in enteric bacteria: insertion sequences flank operon in plasmid-mediated systems. J. Bacteriol. **159:**300-305.

Middendorf, B., G. Blum-Oehler, U. Dobrindt, I. Mühldorfer, S. Salge, and J. Hacker. 2001. The pathogenicity islands of the uropathogenic *Escherichia coli* strain 536: island probing of PAI II_{536}. J. Infect. Dis. **183:**17-20.

Miller, J. F., J. J. Mekalanos, and S. Falkow. 1988. Coordinate regulation and sensory transduction in the control of bacterial virulence. Science **243:**916-922.

Mobley, H. L., D. Island, and G. Massad. 1994. Virulence determinants of uropathogenic *Escherichia coli* and *Proteus mirabilis.* Kidney Int. Suppl **47:** S129-S136.

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

Möllenbrink, M., E. Bruckschen. 1994. Mikrobiologosche Behandlung der chronischen Obstipation. Med. Klin. **89:**587-593.

Moss, J. E., Cardozo, T. J., Zychlinsky, A. and Groisman, E. A. 1999. The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri.* Mol. Microbiol. **33:**74-83.

Mühldhofer, I., Blum, G., Donohue-Rolfe, A., Heier, H., Ölschläger, T., Tschäpe, H., Wallner, U., and J. Hacker. 1996. Characterization of *Escherichia coli* strains isolated from environmental water habitats and from stool samples of healthy volunteers. Res. Microbiol. **147:**625-635.

Murray, G. L., S. R. Attridge, and R. Morona. 2003. Regulation of *Salmonella typhimurium* lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second Wzz. Mol. Microbiol. **47:**1395-1406.

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

Nagy, G., U. Dobrindt, M. Kupfer, L. Emödy, H. Karch, and J. Hacker. 2000. Expression of hemin receptor molecule ChuA is influenced by RfaH in uropathogenic *Escherichia coli* strain 536. Infect. Immun. **69:**1924-1928.

Nissle, A. 1918. Die antagonistische Behandlung chronischer Darmstörungen mir Kolibakterien. Med. Klein. **2:**29-30.

Nissle, A. 1925. Weiteres über Grundlagen und Praxis der Mutaflorbehandlung. DMW **4:**1809-1813.

Nowicki, B., A. Labigne, S. Moseley, R. Hull, S. Hull, and J. Moulds. 1990. The Dr hemagglutinin, afimbrial adhesins AFA-I and AFA-II, and F1845 fimbriae of uropathogenic and diarrhea-associated *Escherichia coli* belong to a family of hemagglutinins with Dr receptor recognition. Infect. Immun. **58:**279-281.

Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299-304.

Oefner, P., S. P. Hunicke-Smith, L. Chiang, F. Dietrich, J. Mulligan, and R. W. Davis. 1996. Efficient random subcloning of DNA sheared in a recirculating point-sink floe system. Nucleic Acids Res. **24:**3879–3886.

Ohlsen, K., W. Ziebuhr, W. Reichardt, W. Witte, F. Götz, and J. Hacker. 1999. Mobile elements, phages, and genomic islands of *Staphylococci* and *Streptococci.* In: Pathogenicity islands and other mobile virulence elements, ed. J. B. Kaper, J. Hacker, pp265-287. Washington: ASM.

Ölschläger, T. A., D. Zhang, S. Schubert, E. Carniel, W. Rabsch, H. Karch, and J. Hacker. 2003. The high-pathogenicity island is absent in human pathogens of *Salmonella enterica* subspecies I but present in isolates of subspecies III and VI. J. Bacteriol. **185:**1107-1111.

Olsen, A., A. Arnqvist, M. Hammar, and S. Normak. 1993. environmental regulation of curli production in *Escherichia coli.* Infect. Agents. Dis. **2:**272-274.

Ørskov, I., F. Ørskov, B. Jann, and K. Jann. 1977. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. Bacteriol. Rev. **41:**667-710.

Ott, M., L. Bender, G. Blum, M. Schmittroth, M. Achtmann, H. Tschäpe, and J. Hacker. 1991. Virulence patterns and long-range genetic mapping of extraintestinal *Escherichia coli* K1, K5, and K100 isolates: use of pulse-field gel elctrophoresis. Infect. Immun. **59:**2664-2672.

Paton, A. W., and J. C. Paton. 1998. Molecular characterization of the locus encoding biosynthesis of the lipopolysaccharide O antigen of *Escherichia coli* serotype O113. Infect. Immun. **67:**5930-5937.

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

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

Porat, R., R. Mosseri, E. Kaplan, M. A. Johns, and S. Shibolet. 1992. Distribution of polysaccharide side chains of lipopolysaccharide determine resistance of *Escherichia coli* to the bactericidal activity of serum. J. Infect. Dis. **165:**953–956.

Purdy, G. E. and S. M. Payne. 2001. The SHI-3 iron transport island of *Shigella boydii* 0-1392 carries the genes for aerobactin synthesis and transport. J. Bacteriol. **183:**4716-4182.

Ranger, F., S. Niklen, A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. USA **74:**5463-5467.

Reeves, P. R. 1995. Role of O–antigen variation in the immune response. Trends Microbiol. **3:**381–386.

Rietschel, E. T., L. Brade, B. Linder, and U. Zähringer. 1992. Biochemistry of lipopolysaccharides. *In* Morrison, D. C. and Ryan, J. L. (eds.) Bacterial endotoxic Lipopolysaccharides. CRC Press, Boca Raton, FL.

Rodriguez, E., **C. Gaggero, and M. Lavina.** 1999. The structural gene for microcin H47 encodes a peptide precursor with antibiotic activity.

Römling, U., and M. Rohde. 1999. Flagella modulate the multicellular behaviour of *Salmonella typhimurium* on the community level. FEMS Microbiol. Lett. **180:**91-102.

Römling, U., W. D. Sierralta, K. Eriksson, and S. Normark. 1998. Multicellular and aggregative behavior of *Salmonella typhimurium* strains is controlled by mutations in the *agfD* promoter. Mol. Microbiol. **28:**249-264.

Römling, U., Z. Bian, M. Hammar, W. D. Sierralta, and S. Normark. 1998. Curli fibers are highly conserved between *Salmonella Typhimurium* and *Escherichia coli* with respect to operon structure and regulation. J. Bacteriol. **180:** 722-731.

Rubin, R. H. 1990. Urinary tract infections. *In* Medicine. Scientific American Inc.

Sambrook, S., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Cold spring Harbor, New York.

Schmidt, H., W. L.Zhang, U. Hemmrich, S. Jelacic, W. Brunder, P. I. Tarr, U. Dobrindt, J. Hacker, and H. Karch. 2001. Identification and characterization of a novel genomic island integrated at *selC* in locus of enterocyte effacement-negative, Shiga toxin-producing *Escherichia coli.* Infect Immun. **66:**6863-6873.

Schubert, S., A. Raki, D. Fischer, J. Sorsa, and J. Heesmann. 1999. Characterization of the integration site of *Yersinia* high-pathogenicity island in *Escherichia coli.* FEMS Microbiol. Lett. **179:**409-414.

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

Smyth, C. J., M. B. Marron, J. M Twohig, and S. G. Smith. 1996. Fimbrial adhesins: similarities and variations in the structure and biogenesis. FEMS Immunol. Med. Microbiol. **16:**127-139.

Sobel, J. D. 1997. Vaginitis. N. Engl. J. Med. **337:**1896-1903.

Sokurenko, E. V., H. S. Courtney, D. E. Ohman, P. Klemm, and D. L. Hasty. 1994. FimH family of type 1 fimbrial adhesins: functional heterogenicity due to minor sequence variations among *fimH* genes. J. Bacteriol. **177:**3680-3686.

Solano, C., B. Garcia, J. Valle, C. Berasain, J.-M. Ghigo, C. Gamazo, and I. Lasa. 2002. genetic analysis of *salmonella enteritidis* biofilm formation: critical role of cellulose. Mol. Microbiol. **43:**793-808.

Sonnenborn, U., and R. Greinwald. 1991. Beziehungen zwischen Wirtsorganismus und Darmflora unter besonderer Berücksichtigung von Physiologie und Funktion der normalen *Escherichia coli* Flora, p. 55-68. 2nd ed. Schattauer, Stuttgart/New York.

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

Staden, R., K. F. Beal, and J. K. Bonfield. 2000. The Staden package 1998. Methods Mol. Biol. **132:**115–130.

Stentebjerg-Oelsen, B., T. Chakraborty, and P. Klemm. 1999. Type 1 Fimbriation and phase switching in a natural *Escherichia coli fimB* null strain, Nissle 1917. J. Bacteriol. **181:**7470-7478.

Stoodley, P., K. Sauer, D. G. Davies, and J. W. Costerton. 2002. Biofilms as complex differentiated communities. Annu. Rev. Microbiol. **56:**187-209.

Sullivan, J. T., and C. W. Ronson. 1998. Evolution of *Rhizobia* by acquisition of a 500-kb symbiosis island that integrates into *phe-* tRNA gene. Proc. Natl. Acad. Sci. USA, **95:**5245-5149.

Svanborg, C., and G. Godaly. 1997. Bacterial virulence in urinary tract infection. Infect. Dis. Clin. North Am. **11:**513-529.

Svanborg–Eden, C., L. Hagberg, R. Hull, S. Hull, K.-E. Magnusson, and L. Öhman. 1987. Bacterial virulence versus host resistance in the urinary tracts of mice. Infect. Immun. **55:**1224–1232.

Swenson, D. L., N. O. Bukanov, D. E. Berg, and R. A. Welch. 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. Infect. Immun. **64:**3736-3743.

Tao, H., C. Bausch, C. Richmond, F. R. Blattner, and T. Conway. 1999. Functional genomics: Expression analysis of *E. coli* growing on minimal and rich media. J. Bacteriol. **181:**6425-6440.

Tarr, P. I., L. M. Schoening, Y.-L. Yea, T. R. Ward, S. Jelacic, and T. S. Whittam. 2000. Acquisition of the *rfb-gnd* cluster in evolution of *Escherichia coli* O55 and O157. J. Bacteriol. **181:**6183-6191.

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

Tauschek, M., R. A. Strugnel, and R. M. Robins-Browne. 2002. Characterization and evidence of mobilization of the LEE pathogenicity island of rabbit-specific strains of enteropathogenic *Escherichia coli.* Mol. Microbiol. **44:** 1533-1550.

Tomb, J.-F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J .D., Kelley, J. M., Karp, P. D., Smith, H. O., Fraser, C. M. and Venter, J. C. 1997. The complete genome sequence of the gastric pathogen Helicobacter pylori. Nature **388:**539-547.

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

Toth, I., **Herault, F., Beutin, L., and E. Oswald.** 2003. Production of cytolethal distending toxins by pathogenic *Escherichia coli* strains isolated from human and animal sources: establishment of the existence of a new *cdt* variant (type IV). J. Clin. Microbiol. **41:**4285-4291.

Tsai, C. M., and C. E. Frasch. 1982. A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. Anal. Biochem. **119:**115–119.

Turner, S. A., S. N. Luck, H. Sakellaris, K. Rajakumar, and B. Adler. 2001. Nested deletions of the SRL pathogenicity island of *Shigella flexneri* 2a. J. Bacteriol. **183:**5535-5543.

Valvano, M. A. 1992. Pathogenicity and molecular genetics of the O-specific side-chain lipopolysaccharides of *Escherichia coli* . Can J. Microbiol. **38:**711–719.

Van Bogele, R., E. Schiller, J. Thomas, and F. Neidhardt. 1999. Diagnosis of cellular states of microorganisms using proteomics. Electrophoresis **20:**2149-2159.

Vizàn, J. L., C. Hernàndez-Chico, I. Del Castillo, and F. Moreno. 1991. The peptide antibiotic microcin B17 induces double strand cleavage of DNA mediated by gyrase. EMBO J. **10:**467-476.

Vokes, S. A., S. A. Reeves, A. G. Torres, and S. M. Payne. 1999. The aerobactin iron uptake transport system genes in *Shigella flexneri* are present within a pathogenicity island. Mol. Microbiol. **33:**63-73.

Wagner, W., Vogel, M., and W. Goebel. 1983. Transport of hemolysin across the outer membrane of *Escherichia coli* requires two functions. J. Bacteriol. **154:**200-210.

Wang, L., and P. R. Reeves. 1998. Organization of *Escherichia coli* O157 O antigen cluster and identification of its specific genes. Infect. Immun. **66:**2580-2584.

Wang, L., H. Curd, W. Qu, and P. R. Reeves. 1998. Sequencing of *Escherichia coli* O111 O–antigen Gene Cluster and Identification of O111–specific genes. J. Clin. Microbiol. **36:**3182-3187.

Wang, L., H. Curd, W. Qu, and P. R. Reeves. 1998. Sequencing of *Escherichia coli* O111 O–antigen Gene Cluster and Identification of O111–specific genes. J. Clin. Microbiol. **36:**3182-3187.

Wang, L., S. Huskic, A. Cisterne, D. Rothemund, and P. R. Reeves. 2002. The O-antigen gene cluster of *Escherichia coli* O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. J. Bacteriol. **184:**2620-2625.

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

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

Wyckoff, E. E., D. Duncan, A. G. Torres, M. Mills, K. Maase, and S. M. Payne. 1998. Structure of the *Shigella dysenteriae* haem transport locus and its phylogenetic distribution in enteric bacteria. Mol. Microbiol. **28:**1139-1152.

Yang, C.-C., and J. Konisky. 1984. Colicin V-treated *Escherichia coli* does not generate membrane potential. J. Bacteriol. **158:**757-759.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene. **33:**103-119.

Zingler, G., G. Blum, U. Falkenhagen, I. Ørskov, F. Ørskov, J. Hacker, and M. Ott. 1993. Clonal differentiation of uropathogenic *Escherichia coli* isolates of serotype O6:K5 by fimbrial antigen typing and DNA long-range mapping techniques. Med. Microbiol. Immunol. **182:**13-14.

Zingler, G., M. Ott, G. Blum, U. Falkenhagen, G. Naumann, W. Sokolowska-Köhler and J. Hacker. 1992. Clonal analysis of *Escherichia coli* serotype O6 strains from urinary tract infections. Micr. Path. **12:**299-310.

Zogaj, X., M. Nimtz, M. Rohde, W. Bokranz, and U. Römling. 2001. The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. Mol. Microbiol. **39:**1452-1463.

VIII. Supplements

Table 1: Primers for amplification of PAI III_{536}-specific sequences (Dobrindt et al., 2001).

PCR number	Position	Gene name	Primer	Sequence (5‘ > 3‘)	PCR product (bp)
1		*thrW*	*thrW*	CCACCGTGATTCACGTTCG	
	1404-1422	SfX integrase	III.1	AGGCACTATCACGCAGTGG	1455
2	1404-1422	SfX integrase	III.2	CCACTGCGTGATAGTGCCT	
	3583-3601	IS100 integrase	III.3	AGCCCATCGCTATTGCCAG	2198
3	3583-3601	IS100 integrase	III.4	TCTGGCAATAGCGATGGGC	
	5606-5624	*pifA*	III.5	CCACACGCAGACAACCTTC	2042
4	5606-5624	*pifA*	III.6	GAAGGTTGTCTGCGTGTGG	
	7700-7018	DAHP-Synthetase	III.7	GGCGAGCATATCGTATCCC	2112
5	7700-7018	DAHP-Synthetase	III.8	GGGATACGATATGCTCGCC	
	9875-9893	LEE, IS tnpA	III.9	TGCTGACGAAATGCCGGAG	2193
6	9875-9893	LEE, IS tnpA	III.10	ACTCCGGCATTTCGTCAGC	
	11908-11926	ColV	III.11	TCATTCCCGTCACCTCCAG	2052
7	11908-11926	ColV	III.12	TCTGGAGGTGACGGGAATG	
	14064-14082	unknown/*int* *Shigella*	III.13	CTGCTGAACAAAACGCCGG	2174
8	14064-14082	unknown/*int* *Shigella*	III.14	GCGCGCTTTACTGTGTCTG	
	15701-15719	*sfaC*	III.15	GGTATGGTTCAGCGCTCTC	1760
9	15701-15719	*sfaC*	III.16	GAGAGCGCTGAACCATACC	
	17301-17319	*sfaA*	III.17	CTTTAAAGGTGGCGTCGGC	2571
10	17301-17319	*sfaA*	III.18	GCCGACGCCACCTTTAAAG	
	19423-19441	*sfaF*	III.19	ACCACGGGCCTGATTACTC	1965
11	19423-19441	*sfaF*	III.20	TGAGTAATCAGGCCCGTGG	
	21597-21615	*sfaG*	III.21	ACTGCCGCAGGAGGTTAAC	2401
12	21597-21615	*sfaG*	III.22	GTTAACCTCCTGCGGCAGT	
	23526-23544	*sfaH*	III.23	TCACTGACTGGACAGCACC	1739
13	23526-23544	*sfaH*	III.24	GGTGCTGTCCAGTCAGTGA	
	25825-25843	*iroN*	III.25	GTGCAAGAGTGAGCCTCTG	1674
14	25825-25843	*iroN*	III.26	CAGAGGCTCACTCTTGCAC	
	27959-27977	*iroE*	III.27	TAGCGCCTGAAGCGGTTTG	2194
15	27959-27977	*iroE*	III.28	ATGCCGGAGTTACCGCATC	
	30188-30206	*iroD*	III.29	GCGAAGCTGAGTCGCTGAA	2249
16	30188-30206	*iroD*	III.30	TTCAGCGACTCAGCTTCGC	
	32242-32260	*iroC*	III.31	CCAGCGACACAAGACGATG	2072
17	32242-32260	*iroC*	III.32	CATCGTCTTGTGTCGCTGG	
	34415-34433	*iroB*	III.33	CCACAAATCGTCTTCGGCC	2192
18	34415-34433	*iroB*	III.34	GGCCGAAGACGATTTGTGG	
	36505-36523	IS 4	III.35	GATGCAACTGAGCAGGCTG	2108
19	40206-40224	HmuR	HmuR.1	AGGCAGCGACCAGTCATTC	
	42067-420085		HmuR.2	AGTTCTGCCTGAAAGCGGC	1880
20	42166-42184	NMA1686	NMA1686.1	TTGTCTGACAGCATCGGGG	
	42900-42919		NMA1686.2	TCGGGAGAAGCTTAAGGTTG	752
21	43217-43235	Protein B. halodurans	Bachalo.1	ATGGTTTCCTTGGCGCAGG	

Table 1 continued

	44323-44341		Bachalo.2	CACCTGATTGCCACCAGAC	1130
22	37574-37592	IS-ORFs 9-11	III-22.1	TACGCTGCTTGCCAGCTTG	
	39412-39430		III-22.2	TCTGTCCTGATGTACGCCC	1866
23	44212-44230	IS*100*-Tnp	III-23.1	ATGGGGCGGTTTGGAGACA	
	46321-46339		III-23.2	TTGCACGCAGTTATGCCGG	2127
24	46321-46339	IS fragments 17-19	III-24.1	CCGGCATAACTGCGTGCAA	
	48430-48448		III-24.2	TGCAAGCACTCAGCAGTGC	2127
25	48430-48448	*iciA*	III-25.1	GCACTGCTGAGTGCTTGCA	
	49203-49221		III-25.2	GTTTTACAGAGCGGCCCGA	791
26	49415-49442	*lysU*-like	III-26.1	ATTCCCAGACCTGCAGTCG	
	50759-50777		III-26.2	TAAGACAACTTCGCGCCCG	1353
27	50759-50777	ORF22	III-27.1	CGGGCGCGAAGTTGTCTTA	
	52075-52093		III-27.2	GGTCTGTATCGAAAGGCGG	1334
28	52075-52093	*cadA*	III-28.1	CCGCCTTTCGATACAGACC	
	54063-54081		III-28.2	GCACTGAACTGGCTGGTTG	2006
29	54063-54081	*cadB*	III-29.1	CAACCAGCCAGTTCAGTGC	
	55409-55427		III-29.2	TCTTCCTGCAAGTCTGGCG	1364
30	59322-59340	*sap*	III-30.1	TCTGGATGTGCTGAGCGGA	
	61211-61229		III-30.2	ATGTTACTGCCGGCTGCGG	1907
31	65528-65547	unknown/*int Shigella*	III-31.1	TATTGAACAGGCCAGGGAAG	
	67207-67225		III-31.2	GCATCTGGTGGCTATGGAAT	1698
32	70163-70182	Haemoglobin protease	III-32.1	ACGGTATCCGACTTCTGCAC	
	71943-71962		III-32.2	TTGGTTATGGACGGCTCAGC	1801
33	73105-73124	Haemoglobin protease/IS*O*-IS*1*	III-33.1	AAGAGGGCAGGATGCTATGA	
	74985-75005		III-33.2	ATTTTGCCGGTCAGGTATTTC	1900

Table 2: Characteristics of the ORFs located the GEI $I_{Nissle1917}$.

Putative ORF	nt positions [bp]	Accession no.	DNA homology to	Similar of the encoded protein	Identity of the aa sequence (%)	Function
ORF1	776-1264	AE016758	partial *c1176*	Hypothetical protein	163/163 (100%) NP_753101	unknown
ORF2	2275-2775	AE016758	*c1179* (*Escherichia coli* K-12 B3128 pseudogene)	D-galactarate dehydratase	145/165 (87%) NP_755750	putative enzyme
ORF3	2857-3318	AE016758	partial *c1180*	D-galactarate dehydratase	77/95 (81%) NP_755750	putative enyzme
ORF4	3278-3782	AE016758	*c1183*	D-galactarate dehydratase	123/174 (70%) NP_755750	unknown
ORF5	4556-5785	AE016758	*c1186*	Putative beta-ketoacyl-ACP synthase	409/409 (100%) NP_753105	putative enzyme; Fatty acid biosynthesis: Fatty acid and phosphatidic acid biosynthesis
ORF6	5782-6513	AE016758	*c1187*	3-oxoacyl-[acyl-carrier protein] reductase	229/243 (94%) NP_753106	enzyme; Fatty acid and phosphatidic acid biosynthesis
ORF7	6513-6977	AE016758	*c1188*	Conserved hypothetical protein	145/154 (94%) NP_753107	unknown
ORF8	6974-8152	AE016758	*c1189*	Putative 3-oxoacyl-[ACP] synthase	362/392 (92%) NP_753108	putative enzyme; Fatty acid biosynthesis: Fatty acid and phosphatidic acid biosynthesis
ORF9	8145-8729	AE016758	*c1190*	Conserved hypothetical protein	184/194 (94%) NP_753109	unknown
ORF10	8736-11044	AE016758	*c1191*	Conserved hypothetical protein	687/769 (89%) NP_753110	unknown

Table 2 continued

ORF11	11013-11618	AE016758	*c1192*	Conserved hypothetical protein	191/201 (95%) NP_753111	unknown
ORF12	11615-12037	AE016758	*c1193*	Hypothetical protein	139/140 (99%), NP_753112	unknown
ORF13	12041-13717	AE016758	*c1194*	Putative enzyme	540/540 (100%) NP_753113	putative enzyme
ORF14	14048-15406	AE016758	*c1197*	Putative enzyme	452/452 (100%), NP_753116	putative enzyme; Fatty acid biosynthesis: Fatty acid and phosphatidic acid biosynthesis
ORF15	15403-15984	AE016758	*c1198*	Conserved hypothetical protein	142/193 (73%) NP_753117	unknown
ORF16	15989-16240	AE016758	*c1199*	Putative acyl carrier protein	83/83 (100%) NP_753118	putative carrier; Fatty acid biosynthesis: Fatty acid and phosphatidic acid biosynthesis
ORF17	16252-16509	AE016758	*c1200*	Putative acyl carrier protein	85/85 (100%) NP_753119	putative carrier; Fatty acid biosynthesis: Fatty acid and phosphatidic acid biosynthesis
ORF18	16484-17323	AE016758	*c1201*	Putative phospholipid biosynthesis acyltransferase	224/279 (80%) NP_753120	putative enzyme
ORF19	17302-18024	AE016758	*c1202*	Conserved hypothetical protein	240/240 (100%) NP_753121	unknown
ORF20	18065-19123	AE016758	*c1203*	Putative O-methyltransferase	352/352 (100%) NP_753122	putative enzyme
ORF21	19188-19577	AE016758	*c1204*	Conserved hypothetical protein	129/129 (100%) NP_753123	unknown
ORF22	20468-22869	AE016758	partial *c1205*	Hypothetical protein	786/799 (98%) NP_753124	unknown
ORF23	22828-25632	AE016758	*c1206* *c1207*	Putative member of ShlA/HecA/FhaA exoprotein family	788/930 (84%) NP_752286	putative enzyme
ORF24	25469-26077	AE016758	partial to *c1207*	Hypothetical protein	53/53 (100%) NP_752286	unknown

Table 2 continued

ORF25	26153-26530	AE016758	*c1208*	Hypothetical protein	75/104 (72%) NP_753127	unknown
ORF26	27395-27814	AE016758	partial *c1209* *c1210* *c1211*	Hypothetical protein	135/139 (97%) NP_755482	unknown
ORF27	27909-28481	AE016758	partial *c1211* *c1212*	Hypothetical protein	154/201 (76%) NP_755482	unknown
ORF28	29010-29213	AE016758	partial *c1213*	Hypothetical protein	67/67 (100%) NP_753130	unknown
ORF29	29374-29664	AE016758	partial *c1214*	Cea protein	35/52 (67%) NP_753131	putative
ORF30	30287-30496	AE016758	*c1215*	Entry exclusion protein 2	69/69 (100%) NP_753132	putative
ORF31	30656-30856		partial *c1216* *c1217*	Hypothetical protein	38/39 (97%) NP_753133; 19/19 (100%) NP_753134	unknown
ORF32	31185-31688	AE016758	*c1218*	Hypothetical protein	167/167 (100%) NP_753135	unknown
ORF33	31660-32061	AE016758	partial *c1219*	Putative Transposase	133/133 (100%) NP_753135	putative transposase
ORF34	31968-32204	AE016758	partial *c1219*	Putative Transposase	78/78 (100%) NP_753135	putative transposase
ORF35	32605-32793	AF302690	*ORF6* PAI III_{536}	Hypothetical protein	62/62 (100%) CAD66177	unknown
ORF36	33354-34388	AF302690	*ORF7* PAI III_{536}	Hypothetical protein	330/344 (95%) CAC43413	unknown
ORF37	32265-36206	AE016758	*c1223* *c1224*	Transposase of insertion element IS*3*	281/284 (98%) CAC43414	IS, phage, Tn; Transposon-related functions
ORF38	37234-37353	AJ009631	*mchX*	Protein (MchX)	18/39 (46%) O86199	required for microcin H47 production. possibly involved in a regulatory loop modulating its own expression and that of *mchB* and *mchB*.

Table 2 continued

ORF39	37422-37631	AJ009631	*mchI*	Microcin H47 immunity protein (MchI)	51/69 (73%) CAA66217	protects microcin H47 producing cell against microcin H47
ORF40	37648-37875	AJ009631	*mchB*	(MchB) protein	AAN79685	putative, microcin production
ORF41	38147-39697	AE016758	*mchC*	(MchC) protein	516/516 (100%) NP_753143	putative, microcin production
ORF42	39723-40175	AE016758	*mchD*	(MchD) protein	150/150 (100%) NP_753144	putative, microcin production
ORF43	40361-41599	AE016758	*mchE*	Microcin H47 secretion protein (MchE)	394/413 (95%) NP_753145	putative, microcin secretion
ORF44	41595-43691	AE016758	*mchF*	Probable microcin H47 secretion ATP-binding protein (MchF)	671/698 (96%) NP_753146	putative, microcin production
ORF45	43727-43948	AF302690	*ORF11* PAI III_{536}	Hypothetical protein	58/73 (79%) CAC43416	putative, microcin production
ORF46	43945-44223	AF302690	*ORF12* PAI III_{536}	microcin V bacteriocin	29/70 (41%) AAC16357	putative, microcin production
ORF47	44398-45084	AF302690	partial *ORF13* PAI III_{536}	Hypothetical protein	221/228 (96%) CAC43418	putative, microcin production
ORF48	45184-45654	AF302690	*ORF14* PAI III_{536}	Hypothetical protein	152/156 (97%) CAC43419	unknown
ORF49	47698-47919	AF302690	*sfaC*	regulatory protein (SfaC)	73/73 (100%) CAC16949	involved in regulation of *Shigella* fimbriae expression; putative F1C and S fimbrial switch regulatory protein
ORF50	48338-48446	AE016758	*sfaB*	(SfaB) protein	36/36 (100%) CAC43422	involved in regulation of *Shigella* fimbriae expression; putative F1C and S fimbrial switch regulatory protein
ORF51	49039-49581	AE016758	*focA*	F1C major fimbrial subunit precursor (FocA)	180/180 (100%) NP_753153	structural component; surface structures
ORF52	49667-50191	AE016758	*sfaD*	Putative minor subunit precursor (SfaD)	173/173 (100%) NP_753154	structural component; surface structures
ORF53	50232-50927	AE016758	*focC*	F1C periplasmic chaperone (FocC)	231/231 (100%) NP_753155	structural component; surface structures

Table 2 continued

ORF54	50997-53627	AE016758	*focD*	F1C fimbrial usher (FocD)	876/876 (100%) NP_753156	membrane; outer membrane constituents
ORF55	53640-54167	AE016758	*focF*	F1C minor fimbrial subunit F precursor (FocF)	175/175 (100%) NP_753157	structural component; surface structures
ORF56	54189-54692	AE016758	*focG*	F1C minor fimbrial subunit protein G precursor (FocG)	167/167 (100%) NP_753158	structural component; surface structures
ORF57	54754-55653	AE016758	*focH*	F1C Putative fimbrial adhesin precursor (FocH)	299/299 (100%) NP_753159	structural component; surface structure
ORF58	55957-56697	AE016758	*c1246*	Hypothetical protein	246/246 (100%) NP_753160	unknown
ORF59	56895-57396	AE016758	*focX*	Putative Regulatory protein (FocX)	166/166 (100%) NP_753161	putative regulator
ORF60	57721-58758	AE016758	*c1248*	Hypothetical protein	345/345 (100%) NP_753162	putative transposase related
ORF61	58988-61165	AE016758	*iroN*	Siderophore receptor (IroN)	711/725 (98%) NP_753164	membrane; transport of small molecules
ORF62	61210-62166	AE016758	*iroE*	IroE protein	302/318 (94%) NP_753165	putative exported protein
ORF63	62251-63480	AE016758	*iroD*	IroD protein, ferric enterochelin esterase	398/409 (97%) NP_753166	putative esterase
ORF64	63584-67369	AE016759	*iroC*	ATP binding cassette (ABC) transporter homolog (IroC)	1150/1245 (92%), NP_753167	putative transport
ORF65	67383-68498	AE016759	*iroB*	Putative glucosyltransferase (IroB)	359/371 (96%) NP_753168	putative sugar transferase
ORF66	69513-69796	AE016759	partial to *c1256*	Hypothetical protein	94/94 (100%) AAN78568	putative transposase related
ORF67	70063-70310	AE016759	partial to *c1256, c1257*	Hypothetical protein	75/78 (96%) CAD66180	putative IS, Tn related
ORF68	70098-70460	AE016759	partial to *c1257*	Hypothetical conserved protein	75/78 (96%) CAD66180	putative IS, Tn related
ORF69	71335-71769	AE016759	*c1259*	Hypothetical protein	100/118 (84%) NP_753173	unknown

Table 2 continued

tRNA	71977-72053	AE016759	*c5534*	tRNA *arg*	-	tRNA
ORF 70	72332-72790	AE016759	*c1260*	Putative Transposase	151/152 (99%) NP_753174	putative enzyme; transposases insertion sequence associated
ORF 71	72802-73695	AE016759	*ORF35* (PAI III_{536})	Hypothetical protein (PAI III)	265/297 (89%) CAD66183	unknown
ORF 72	74041-76011	AE016759	*c1265*	Outer membrane heme/hemoglobin receptor, membrane; Transport of small molecules:Cations	645/656 (98%) NP_753179	putative membrane-related; transport of small molecules:cations
ORF 73	76030-76821	AE016759	*c1266*	Hypothetical protein	263/263 (100%) NP_753180	unknown
ORF 74	77085-78284	AE016759	partial to *c1267*	Hypothetical protein	386/399 (96%) NP_753181	unknown
ORF 75	79235-80371	AE016759	*c1269*	Hypothetical protein	354/378 (93%) NP_753183	unknown
ORF 76	80407-80494	AE016759	*c1270*	Hypothetical protein	180/180 (100%) NP_753184	unknown
ORF 77	81888-82247	AE016759	-	Hypothetical protein	30/118 (25%) P02985	unknown
ORF 78	82456-82785	AE016759	-	Hypothetical protein	33/87 (37%) NP_294168	unknown
ORF 79	82799-83104	AE016759	-	Hypothetical protein	-	unknown
ORF 80	83880-84791	AE016759	*c1272*	Hypothetical protein (YeeP)	290/290 (100%) NP_753186	putative structure
ORF 81	84966-88241	AE016759	*c1273*	Antigen 43 precursor	1064/1091 (97%) NP_753187	membrane; outer membrane constituents

Table 2 continued

ORF 82	88356-90878	AE016759	*c1274*	Hypothetical protein (YeeR)	749/829 (90%) NP_753188	unknown
ORF 83	90954-91409	AE016759	*c1275*	Hypothetical protein	151/151 (100%) NP_753189	unknown
ORF 84	91820-92641	AE016759	*c1278*	Hypothetical protein (yafZ)	271/271 (100%) NP_753193	unknown
ORF 85	92898-93857	AE016759	*c1281*	Hypothetical protein	319/319 (100%) NP_753195	unknown
ORF 86	94172-94816	AE016759	*c1285*	Hypothetical protein	214/214 (100%) NP_753198	unknown
ORF 87	94758-95240	AE016759	partial to *c1285, c1286, c1287* (entire),	Hypothetical protein	151/160 (94%) AAL67343	unknown
ORF 88	95185-95667	AE016759	partial to *c1287*, entire *c1288*	Hypothetical protein (YeeV)	126/126 (100%) NP_753201	unknown
ORF 89	95664-96152	AE016759	*c1289*	Unknown protein encoded within prophage	149/162 (91%) NP_753202	phage or prophage related
ORF 90	95935-96369	AE016759	partial to *1289, c1290*	Hypothetical protein	77/94 (81%) AAD32187	unknown
tRNA	97477-97534	AE016759	*c5535*	tRNA *serX*	-	tRNA

Table 3: Characteristics of the ORFs located the GEI II_{Nisle} 1917.

Putative ORF	nt positions [bp]	Accession no.	DNA homology to	Similar of the encoded protein	Identity of the aa sequence (%)	Function	Putative ORF
tRNA	1-76	AE016766	tRNA *pheV*	-	-	-	tRNA
ORF1	274-1539	AE016766	*c3556*	1263/1263 (100%)	Prophage P4 integrase	421/421 (100%) NP_755431	IS, phage, Tn; Phage-related functions and prophages
ORF2	1765-2838	AE016766	*c3557*	1069/1074 (99%)	ShiA homolog	344/347 (99%) NP_755432	unknown
ORF3	2891-3211	AE016766	*c3558*	222/231 (96%)	Hypothetical protein	75/76 (98%) NP_756355	unknown
ORF4	3368-3643	J01729 M25736	*insA*	276/276 (100%)	Hypothetical protein (InsA)	91/91 (100%) CAA11401	putative IS*10*-related
ORF5	3562-4065	J01729 M25736	*insB*	501/504 (99%)	Hypothetical protein (InsB)	166/167 (99%) AAK97138	putative IS*10*-related
ORF6	4531-4947	AE016766	*ΔpapX* (last 135 nt lacking)	398/417 (95%)	PapX protein	131/138 (94%) NP_755457	not functional; surface structures, Pap-fimbriae synthesis
ORF7	5088-6296	AP000342	*yedA*	1206/1209 (99%)	Transposase of Tn*10*	402/402 (100%) AAB28848	putative transposon-related
8	6312-6528	AE016766	*ΔpapX* (first 335 nt lacking)	208/217 (95%)	PapX protein	64/72 (88%) NP_755457	not functional; surface structures, pap-fimbriae synthesis
ORF9	6800-7303	J01729 M25736	*insB*	501/504 (99%)	Hypothetical protein (InsB)	166/167 (99%) AAK97138	putative IS*10*-related
ORF10	7222-7497	J01729 M25736	*insA*	276/276 (100%)	Hypothetical protein (InsA)	91/91 (100%) CAA11401	putative IS*10*-related
11	7967-8127	AE016766	*ΔpapA* lacking 149 nt from the beginning	151/161 (93%)	PapA fimbrial protein precursor	47/53 (88%) NP_755467	not functional; surface structures, pap-fimbriae synthesis
ORF12	9020-9253	AJ494981	*papI (prfI)*	230/234 (98%)	Transcription regulator (PapI)	74/77 (96%) RGECPI	transcription regulation

Table 3 continued

ORF13	9904-10515	AE016766	*c3594* *c3595*	1292/1310 (98%) 350/358 (97%)	Putative transposase IS*629*	284/296 (95%) NP_757061	putative enzyme; Transposases
ORF14	10704-11030	AE016766	*c3596*	323/327 (98%)	Hypothetical protein in IS	81/108 (75%) CAC37925	Insertion Sequence Associated
ORF15	11770-12516	AE016766	*c3597*	735/747 (98%)	Putative Transposase	224/248 (90%) AAF60967	Insertion Sequence Associated
ORF16	12575-13071	AE016766	partial *c3598*	314/315 (99%)	Hypothetical protein	90/104 (86%) AAC61723	unknown
ORF17	13068-13328	AE016766	*c3599*	257/257 (100%)	Hypothetical protein	86/86 (100%) NP_755474	unknown
ORF18	13370-13930	AE016766	*c3600*	559/561 (99%)	Hypothetical protein	184/186 (98%) NP_755475	unknown
ORF19	13940-14398	AE016766	*c3601*	457/459 (99%)	Hypothetical protein	150/152 (98%) NP_755476	unknown
ORF20	15679-16335	AE016766	partial *c3602*	646/657 (98%)	IS*2*-related hypothetical protein (transposase)	215/218 (98%) NP_755486	putative IS*2*-related
21	15016-16346	AB079602	IS*2*	1294/1331 (97%)	IS*2*-repeat region; Insertion element IS*2*A hypothetical 48.2 kDa protein; belongs to transposase family 8.	283/304 (93%) P51026	Transposable element
ORF22	16448-17641	AE016766	*shiF*	1189/1194 (99%)	ShiF protein	370/397 (93%) NP_709453	putative membrane protein
ORF23	17720-19501	AE016766	*iucA*	1772/1782 (99%)	IucA protein	583/593 (98%) AAD44746	putative aerobactin siderophore synthetase
ORF24	19502-20449	AE016766	*iucB*	941/948 (99%)	IucB protein	313/315 (99%) NP_755501	putative siderophore biosynthesis
ORF25	20449-22191	AE016766	*iucC*	1707/1740 (98%)	IucC protein	576/580 (99%) NP_755500	putative siderophore biosynthesis
ORF26	22188-23465	ECU90207	*iucD*	1268/1268 (100%)	L-lysine 6-monooxigenase (lysine N6-hydroxylase) (IucD)	425/425 (100%) P11295	putative siderophore biosynthesis
ORF27	23547-25748	ECDF13RE	*iutA*	2188/2202 (99%)	Ferric aerobactin receptor precursor (IutA)	701/733 (95%) P14542	putative siderophore biosynthesis

Tabke 3 continued

ORF28	26692-30579	AE016766	*sat*	3854/3888 (99%)	Secreted autotransporter toxin (Sat)	1258/1295 (97%) NP_755494	serine protease
ORF29	31176-32276	AJ278144	*ORF13*	1096/1098 (99%)	Hypothetical protein IS*30*-related (InsI); putative transposase	350/366 (95%) P37246	IS, phage, transposon-related
ORF30	33665-34483	AE016766	*c3575*	818/819 (99%)	Transposase insF for insertion sequence IS*3*A/B/C/D/E/Fa	251/272 (92%), NP_755450	IS, phage, transposon-related
ORF31	35250-35933	AF336309	77-626 nt homologous to 588-1137 nt of pYVe8081	441/550 (80%)	Hypothetical protein	71/191 (37%) NP_052891	unknown; possible transposase
ORF32	36476-36898	AF386526	*CP0118*	406/418 (97%)	IS*1353* putative transposase-like protein	135/139 (97%) AAL72503	IS, phage, transposon-related
ORF33	38092-38364	AF081285	*r8*	272/273 (99%)	neurotensin receptor R8	89/90 (98%) AAC61726	putative
ORF34	38873-39121	AF081285	*r7*	248/249 (99%)	maltopentaose forming amylase R7	59/82 (71%) AAC61727	enzyme, sugar methabolism
ORF35	40343-42433	AE016766	*iha*	2085/2091 (99%)	Putative receptor/adhesin (Iha)	667/696 (95%) NP_309387	putative membrane; ell envelope: Outer membrane constituents
ORF36	42933-43589	AE016766	partial *c3612*	646/657 (98%)	IS*2*-related hypothetical protein (transposase)	215/218 (98%) NP_755486	putative IS*2*-related
37	42922-44252	AB079602	IS*2*	1294/1331 (97%)	IS*2*-repeat region; Insertion element IS*2* hypothetical 48.2 kDa protein; belongs to transposase family 8.	283/304 (93%) P51026	Transposable element
ORF38	44313-45302	AE016766	*c3630*	989/989 (100%)	Hypothetical protein YjhS precursor	329/329 (100%) NP_755505	unknown
ORF39	45392-45760	AE016766	partial *c3631*	319/343 (93%)	Hypothetical transcriptional regulator (YhcK)	70/114 (61%) NP_755846	putative regulator
ORF40	46169-47104	AE016766	partial *c3632*	913/939 (97%)	Hypothetical protein	268/318 (84%) NP_755508	unknown
ORF41	47154-48263	AE016766	*c3634*	1076/1110 (96%)	Hypothetical protein YjhT precursor	368/369 (99%) NP_755509	unknown

Table 3 continued

ORF42	48276-48986	AE016766	*c3635*	446/453 (98%)	Hypothetical protein YjhA precursor	226/236 (95%) NP_755510	unknown
ORF43	49032-49862	AE016766	*c3636*	815/831 (98%)	Hypothetical protein	268/269 (99%) NP_755511	unknown
ORF44	49866-51338	AE016766	*c3637*	1473/1473 (100%)	Putative sialic acid transporter	474/490 (96%) NP_755512	transport; murein sacculus, peptidoglycan
ORF45	51387-52262	AE016766	*c3638*	876/876 (100%)	Hypothetical protein (YhcI)	279/291 (95%) NP_755513	putative regulator
ORF46	52296-53213	AE016766	*c3639*	889/897 (99%)	N-acetylneuraminate lyase subunit	305/305 (100%) NP_755514	enzyme; surface polysaccharides
ORF47	54296-54646	AE016766	partial *c3641*	351/351 (100%)	Unknown in ISEc8	116/116 (100%) NP_755516	Insertion sequence associated
ORF48	54677-55777	AE016766	*c3643*	1092/1101 (99%)	Unknown in ISEc8	325/366 (88%) NP_755518	Insertion sequence associated
ORF49	55774-57312	AE016766	partial *c3645*	1539/1539 (100%)	Unknown protein encoded by ISEc8 within prophage	512/512 (100%) NP_755520	Insertion sequence associated
ORF50	60185-60481	AE016766	partial *c3650*	282/293 (96%)	Hypothetical protein	no significant match	unknown
ORF51	60770-61342	AE016766	partial *c3651*	573/573 (100%)	Hypothetical protein	177/190 (93%) NP_755526	unknown
ORF52	61463-62953	AE016766	partial *c3652*	1472/1491 (98%)	Hypothetical protein (YfjI)	461/496 (92%) NP_755527	unknown
ORF53	64285-65157	AE016766	*c3654*	873/873 (100%)	Hypothetical protein (YeeP)	290/290 (100%) NP_755529	putative structure
ORF54	65485-68613	AE016766	*c3655 (sap)*	3024/3129 (96%)	Antigen 43 precursor (Sap)	960/1042 (92%) NP_755530	membrane; outer membrane constituents
ORF55	68744-69835	AE016766	partial *c3664*	1088/1092 (99%)	Hypothetical protein (YeeR)	320/352 (90%) NP_753188	unknown
ORF56	69911-70366	AE016766	*c3665*	456/456 (100%)	Hypothetical protein	151/151 (100%) NP_755540	unknown
ORF57	70779-71597	AE016766	*c3668*	818/819 (99%)	Hypothetical protein (YafZ)	257/272 (94%) NP_755543	unknown

Table 3 continued

ORF58	71652-72137	AE016766	*c3669*	485/486 (99%)	Hypothetical protein (YfjX)	148/161 (91%) NP_755544	unknown
ORF59	71706-72617	AE016766	*c3670*	909/912 (99%)	Hypothetical protein	300/303 (99%) NP_755545	unknown
ORF60	72626-73354	AE016766	*c3672*	677/683 (99%)	Hypothetical protein	228/242 (94%) NP_755547	unknown
ORF61	72932-73576	AE016766	*c3674*	629/645 (97%)	Hypothetical protein	209/214 (97%) NP_755549	unknown
ORF62	73518-73994	AF453442	*RorfE*	452/477 (94%)	Hypothetical protein	145/158 (91%) AAL57576	unknown
ORF63	74457-74945	AE016766	*c3678*	486/489 (99%)	Conserved hypothetical protein	150/162 (92%) NP_755553	unknown
ORF64	74728-75159	AE016766	partial *c3678, c3679*	428/432 (99%)	Hypothetical protein	80/110 (72%) AAD32187	unknown
ORF65	75241-76089	AJ488511	*z1226*	259/278 (93%)	Hypothetical protein	259/278 (93%) CAD33792	unknown
ORF66	76158-76553	AE016766	*c3681*	395/396 (99%)	Hypothetical protein	130/131 (99%) NP_755556	unknown
ORF67	76546-77286	AE016766	partial *c3682*	741/742 (99%)	Hypothetical protein	243/243 (100%) NP_755557	unknown
ORF68	78877-79860	X95264	*kpsF*	984/984 (100%)	KpsF protein	327/327 (100%) CAA64561	K5 capsule biosynthesis
ORF69	79932-81080	X74567	*kpsE*	1149/1149 (100%)	Capsule polysaccharide export inner-membrane protein (KpsE)	351/382 (91%) P42214	K5 capsule membrane export
ORF70	81104-82780	X74567	*kpsD*	1674/1677 (99%)	KpsD protein	557/558 (99%) NP_755563	K5 capsule biosynthesis
ORF71	82790-83530	S76943	*kpsU*	741/741 (100%)	3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase) (CMP-2-keto-3-deoxyoctulosonic acid synthetase) (CKS) (KpsU)	246/246 (100%) P42216	K5 capsule biosynthesis
ORF72	83527-85554	X74567	*kpsC*	2007/2028 (98%)	Capsule polysaccharide export protein (KpsC)	612/675 (90%) P42217	K5 capsule export

Table 3 continued

ORF73	85589-86758	X74567	*kpsS*	1169/1170 (99%)	Capsule polysaccharide export protein (KpsS)	388/389 (99%) P42218	K5 capsule export
ORF74	87097-88275	X77617	*kfiD*	1171/1179 (99%)	UDP-glucose 6-dehydrogenase (UDPGDH) (KfiD)	378/392 (96%) Q47329	K5 capsule biosynthesis
ORF75	88317-89879	X77617	*kfiC*	1539/1563 (98%)	putative glycosyltransferase (KfiC)	519/520 (99%) S70198	K5 capsule biosynthesis
ORF76	90212-90484	X77617	*ORF* (*kfi*-region)	273/273 (100%)	hypothetical protein kfi-region	72/90 (80%) S70194	unknown
ORF77	90575-90877	X77617	*ORF* (*kfi*-region)	303/303 (100%)	hypothetical 11.5 KDa protein	100/100 (100%) S70191	unknown
ORF78	91171-92862	X77617	*kfiB*	1659/1692 (98%)	KfiB protein	518/563 (92%) S70196	K5 capsule biosynthesis
ORF79	93294-94010	X77617	*kfiA*	696/717 (97%)	KfiA protein	238/238 (100%) S70195	K5 capsule biosynthesis
ORF80	94995-95669	X53819	*kpsT*	673/675 (99%)	polysialic acid transport ATP-binding protein (KpsT)	211/224 (94%) P24586	K5 capsule biosynthesis
ORF81	95666-96442	X53819	*kpsM*	777/777 (100%)	polysialic acid transport protein (KpsM)	246/258 (95%) P24584	K5 capsule biosynthesis

Table 4: Characteristics of the ORFs located in the sequence context of *argW* in strain Nissle 1917.

Putative ORF	nt positions [bp]	Accession no.	DNA homology to	Similar protein	Identity of the aa sequence (%) accession no.	Function
ORF1	700-1575	AE016766	*c3638*	Hypothetical protein (YhcI)	278/291 (95%) NP_755513	putative regulator
ORF2	1624-3096	AE016766	*c3637*	Putative sialic acid transporter	473/490 (96%) NP_755512	transport; murein sacculus, peptidoglycan
ORF3	3100-3930	AE016766	*c3636*	Hypothetical protein	268/269 (99%) NP_755511	unknown
ORF4	3976-4686	AE016766	1-113 and 259-711 nt homologous to *c3635*	Hypothetical protein (YjhA) precursor	225/236 (95%) NP_755510	unknown
ORF5	4688-5808	AE016766	*c3634*	Hypothetical protein (YjhT) precursor	368/369 (99%) NP_755509	unknown
ORF6	5858-6793	AE016766	*c3633*	Hypothetical protein	267/318 (83%) NP_755508	unknown
ORF7	6829-7563	AE016766	1-666 nt homologous to *c3632* and *c3631*	putative GntR-family transcriptional regulator [*Salmonella enterica* subsp. enterica serovar Typhi]	120/227 (52%) NP_462250	putative regulator
ORF8	7202-7570	AE016766	1-343 nt homologous to *c3631* and *c3630*	Hypothetical protein; Hypothetical protein (YjhS) precursor	69/110 (62%) NP_462250	unknown
ORF9	7660-8160	AE016764	203-501 nt homologous to *c2876*	Hypothetical adenine-specific methylase (YfcB)	84/128 (65%) AAN82078	putative methylase
ORF10	8275-8826	AE016764	*c2877*	Hypothetical protein (YfcN)	167/167 (100%) P77458	unknown
ORF11	9146-10003	AE016764	*c2878*	Hypothetical protein (YfcO) precursor	274/285 (96%) AAN81328	unknown

Table 4 continued

ORF12	10005-10532	AE016764	*c2879*	Hypothetical fimbrial-like protein (YfcP) precursor	167/175 (95%) NP_754761	putative structure
ORF13	10529-11008	AE016764	*c2880*	Hypothetical fimbrial-like protein (YfcQ) precursor	159/159 (100%) NP_754762	putative structure
ORF14	11005-11550	AE016764	*c2881*	Hypothetical protein (YfcR) fimbrial precursor	167/167 (100%) NP_754763	putative structure
ORF15	11525-11277	AE016764	*c2882*	Hypothetical fimbrial chaperone (YfcS) precursor	228/250 (91%) NP_754764	putative factor
ORF16	12297-14951	AE016764	*c2883*	Putative outer membrane protein (YfcU)	830/884 (93%) NP_754765	putative membrane
ORF17	150026-15592	AE016764	*c2884*	Hypothetical fimbrial-like protein (YfcV) precursor	136/156 (87%) NP_754766	putative structure
ORF18	16159-16644	AE016764	*c2785*	Phosphohistidine phosphatase (SixA)	161/161 (100%) NP_754767	enzyme; methabolism
ORF19	16847-18991	AE016764	*c2886*	Putative fatty oxidation complex alpha subunit; Enoyl-CoA hydratase; 3-hydroxyacyl-CoA dehydrogenase; 3-hydroxybutyryl-CoA epimerase	675/714 (94%) NP_754768	putative enzyme
ORF20	18991-20301	AE016764	*c2887*	Probable 3-ketoacyl-CoA thiolase	425/436 (97%) NP_754769	putative enzyme
ORF21	21131-22477	AE016764	*c2889*	Long-chain fatty acid transport protein precursor	448/448 (100%) NP_754771	enzyme; Transport of small molecules: Carbohydrates, organic acids, alcohols
ORF22	22539-23294	AE016764	*c2890*	Lipoprotein precursor (VacJ)	251/251 (100%) NP_754772	membrane; Macromolecule synthesis, modification: Lipoprotein
ORF23	23588-24520	AE016764	*c2892*	Hypothetical protein (YfdC)	283/310 (91%) NP_754773	putative transport
tRNA	24596-24670	AE016764	*argW*	-	-	tRNA-Arg
ORF24	25108-25638	AE016764	*c2893*	Hypothetical protein	164/176 (93%) NP_754775	unknown
ORF25	25686-31691	AE016764	1-1024 and 1085-5940 nt homologous to *c2894, c2895*	putative autotransporter protein [*Yersinia pestis* CO92] (YapH)	624/2029 (30%) CAC89847	unknown

Table 4 continued

ORF26	31616-32254	AE016760	7-639 nt homologous to *c1857*	Hypothetical protein (YdcO)	183/199 (91%) NP_753758	putative transport
ORF27	32253-32789	AE016760	*c1858*	Hypothetical protein (YdcN)	178/178 (100%) NP_753759	unknown
ORF28	33162-34823	AE016760	*c1859*	Putative protease (YdcP) precursor	553/553 (100%), NP_753760	putative enzyme
ORF29	34915-35145	AE016760	*c1860*	Hypothetical protein (YncJ) precursor	76/76 (100%) NP_753761	putative
ORF30	35639-36604	AE016760 (*c1863*) AE000500 (*yhjF*)	1-682 nt homologous to *c1863;* 679-966 nt homologous to *yjhF*	Hypothetical protein (YdcR)	229/237 (96%) NP_753764	unknown
ORF31	36284-37006	AE000500	34-723 nt homologous to *yjhF*	Putative transport system permease	168/208(80%) AAC77252	putative transport
ORF32	36744-37217	AE000500	*yjhF*	Hypothetical protein	3-197 aa 37/65 (56%) AAC77252; 382-474 aa 31/31 (100%) P39358	unknown

Table 5: Characteristics of the *wa** gene cluster and flanking regions.

Putative ORF	nt positions [bp]	Accession no.	DNA homology to	Similar of the encoded protein	Identity of the aa sequence (%)	Function
ORF1	24-662	AE016768	*waaC*	lipopolysaccharide heptosyltransferase-1 (WaaC)	213/213 (100%),	enzyme; macromolecule metabolism: lipopolysaccharide
ORF2	735-1988	AE016768	*waaL*	lipid A-core, surface polymer ligase (WaaL)	388/417 (93%)	enzyme, R1 core LPS synthesis
ORF3	2034-3017	AE016768	*waaV*	putative beta1,3-glucosyltransferase (WaaV)	327/327 (100%)	enzyme; macromolecule metabolism: lipopolysaccharide
ORF4	3099-4127	AE016769	*waaW*	UDP-galactose:(galactosyl) LPS alpha1,2-galactosyltransferase (WaaW)	342/342 (100%)	enzyme; macromolecule metabolism: lipopolysaccharide
ORF5	4153-4845	AE016769	*waaY*	lipopolysaccharide core biosynthesis protein (WaaY)	217/217 (100%)	enzyme; macromolecule metabolism: lipopolysaccharide
ORF6	4855-5850	AE016769	*waaT*	lipopolysaccharide 1,2-glucosyltransferase (WaaT)	331/331 (100%)	enzyme; surface polysaccharides and antigens
ORF7	5867-6883	AE016769	*waaO*	lipopolysaccharide 1,3-galactosyltransferase (WaaO)	327/327 (100%)	enzyme; surface polysaccharides and antigens
ORF8	6899-7696	AE016769	*waaP*	lipopolysaccharide core biosynthesis protein (WaaP)	251/265 (94%)	enzyme; macromolecule metabolism: lipopolysaccharide
ORF9	7689-8813	AE016769	*waaG*	glucosyltransferase I (WaaG)	374/374 (100%)	enzyme; macromolecule metabolism: lipopolysaccharide
ORF10	8810-9832	AE016769	*waaQ*	lipopolysaccharide core biosynthesis glycosyl transferase (WaaQ)	316/316 (100%)	enzyme; macromolecule metabolism: lipopolysaccharide

Table 5 continued

ORF11	10281-11558	AE016769	*kdtA*	3-deoxy-D-manno-octulosonic-acid transferase (KdtA)	367/387 (94%)	enzyme; surface polysaccharides and antigens
ORF12	11398-12045	AE016769	*kdtB*	putative phosphopantetheine adenylyltransferase (KdtB)	159/159 (100%),	enzyme; macromolecule metabolism: lipopolysaccharide
ORF13	12084-12893	AE016769	*mutM*	formamidopyrimidine-DNA glycosylase (MutM)	256/269 (95%)	enzyme; DNA - replication, repair, restriction/modification
ORF14	13632-14276	AE016769	*radC*	DNA repair protein (RadC)	214/214 (100%)	enzyme; DNA - replication, repair, restriction/modification
ORF15		AE016769	*dfp*	DNA/pantothenate metabolism flavoprotein (Dfp)	378/430 (87%)	phenotype; DNA - replication, repair, restriction/modification
ORF16	15673-16125	AE016769	*dut*	deoxyuridine 5'-triphosphate nucleotidohydrolase Dut	151/151 (100%)	enzyme; 2'-deoxyribonucleotide metabolism
ORF17	16193-16828	AE016769	*ttk*	Ttk protein	201/212 (94%)	putative regulator
ORF18	16871-17509	AE016769	*pyrE*	orotate phosphoribosyltransferase (PyrE)	213/213 (100%)	enzyme; pyrimidine ribonucleotide biosynthesis
ORF19	17607-18291	AE016769	*rph*	ribonuclease PH	207/228 (90%)	enzyme; degradation of RNA
ORF20	18418-19297	AE016769	*yicC*	protein (yicC)	153/169 (90%)	unknown
ORF21	19492-20325	AE016769	*dinD*	DNA-damage-inducible protein D (Din D)	278/278 (100%)	phenotype; DNA - replication, repair, restriction/modification
ORF22	20566-21234	AE016769	*yicG*	hypothetical protein (yicG)	163/223 (73%)	unknown
ORF23	21237-22922	AE016769	*yicF*	hypothetical DNA ligase-like protein (yicF)	561/562 (99%)	putative

Table 5 continued

ORF24	23174-23794	AE016769	*gmk*	guanylate kinase (Gmk)	75/91 (82%)	enzyme; purine ribonucleotide biosynthesis
ORF25	23852-24127	AE016769	*rpoZ*	DNA-directed RNA polymerase omega chain	688/702 (98%)	enzyme; RNA synthesis, modification, DNA transcription
ORF26	24146-26251	AE016769	*spoT*	guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase (spoT)	688/702 (98%)	enzyme; global regulatory functions
ORF27	26261-26947	AE016769	*spoU*	tRNA (guanosine-2'-O-)-methyltransferase	229/229 (100%)	putative methyltransferase-related
ORF28	26956-29037	AE016769	*recG*	ATP-dependent DNA helicase (RecG)	638/693 (92%)	enzyme; DNA - replication, repair, restriction/modification
ORF29	29047-30276	AE016769	*gltS*	sodium/glutamate symport carrier protein (Get S)	338/401 (84%)	transport; transport of small molecules: amino acids, amines
ORF30	30556-31944	AE016769	*yicE*	putative purine permease (YicE)	449/463 (96%)	putative transport of purine
ORF31	32068-33774	AE016769	*yicH*	hypothetical protein (YicH)	546/546 (100%)	unknown
ORF32	33821-34501	AE016769	*c4481*	conserved hypothetical protein	226/226 (100%)	unknown
ORF33	34482-35414	AE016769	*c4482*	hypothetical protein (YajF)	274/310 (88%)	putative regulator
ORF34	34462-36322	AE016769	*c4483*	putative aldolase	286/286 (100%)	putative enzyme
ORF35	36403-37257	AE016769	*c4484*	putative aldolase	275/283 (97%)	putative enzyme
ORF36	37266-38357	AE016769	*c4485*	Putative phosphotransferase system (PTS) enzyme-ii fructose	350/363 (96%)	putative enzyme; ; degradation of small molecules: carbon compounds
ORF37	38382-38696	AE016769	*c4486*	PTS system, fructose-like-2 IIB component 1	104/104 (100%)	putative enzyme; degradation of small molecules: carbon compounds

Table 5 continued

ORF38	38714-39184	AE016769	*c4487*	putative phosphotransferase system (PTS)	156/156 (100%)	putative; enzyme; degradation of small molecules: carbon compounds
ORF39	39211-40764	AE016769	*c4488*	putative transcriptional Antiterminator	509/517 (98%)	putative regulation
ORF40	41050-43365	AE016769	*yicI*	putative family 31 glucosidase (YicI)	746/772 (96%)	putative enzyme; sugar methabolism
ORF41	43378-44808	AE016769	*yicJ*	hypothetical symporter (YicJ)	462/477 (96%)	putative transport
	45050-45144	AE016769	*selC*	-	-	tRNA
ORF42	45445-46653	AE016769	*intC,* 8-251 and 269-1059 88 % homologous to *intC* from *E. coli* CFT073	putative prophage integrase (Int C)	375/390 (96%)	IS, phage, Tn

Table 6: Characteristics of *fim*$_{\text{Nissle 1917}}$ gene cluster and flanking sequences.

Putative ORF	nt positions [bp]	Accession no.	DNA homology to	Similar of the encoded protein	Identity of the aa sequence (%)	Function
ORF1	473-1303	AE000500	*yjhH* (partial)	putative lyase/synthase (YjhH9	263/277 (94%) NP_418718	putative enzyme; not classified
ORF2	1257-2042	AE000500	*yjhI*	putative regulator (YjhI)	261/262 (99%) NP_418719	putative regulator; not classified
ORF3	2348-3127	AE000500	*sgcR*	putative DeoR-type transcriptional regulator (SgcR)	259/260 (99%) NP_418720	putative regulator; not classified
ORF4	3147-3776	AE000500	*sgcE*	putative epimerase (SgcE)	197/210 (93%) NP_418721	putative enzyme; not classified
ORF5	3788-4219	AE000500	*sgcA*	putative PTS system enzyme II A component (SgcA)	141/143 (98%) NP_418722	putative transport; not classified
ORF6	4353-5156	AE000501	*sgcQ*	putative nucleoside triphosphatase (SgcQ)	268/268 (100%) NP_418723	putative enzyme; not classified
ORF7	5172-6482	AE000501	*sgcC*	putative PTS system enzyme IIC component (SgcC)	418/437 (95%) NP_418724	putative transport; not classified
ORF8	6772-7920	AE000501	*sgcX*	putative lyase/synthase (SgcX)	368/383 (96%) NP_418725	putative enzyme; not classified
ORF9	8680-9423	AE000501	*yjhP*	putative methyltransferase (YjhP)	238/248 (95%) NP_418726	putative enzyme; not classified
ORF10	9479-10026	AE000501	*yjhQ*	orf, hypothetical protein (YjhQ)	163/172 (94%) NP_418727	orf; unknown
ΔORF11	10839-11635	AE000498	5-797 nt homologous to *b4285*	putative transposase	253/254 (99%) CAD48134	IS, phage, Tn; not classified
ORF12	12169-13185	AE000501	*yjhR*	putative frameshift suppressor (YjhR)	332/338 (98%) NP_418728	phenotype; not classified
ORF13	13771-14748	AE000501	*yjhS*	orf, hypothetical protein (YjhS)	315/326 (96%) NP_418729	orf; unknown
ORF14	14813-16026	AE000501	*yjhT*	orf, hypothetical protein (YjhT)	398/404 (98%) NP_418730	orf; unknown

Table 6 continued

ORF15	15941-16663	AE000501	*yjhA*	orf, hypothetical protein (YjhA)	240/241 (99%) NP_418731	orf; unknown
16	18298-20211	AF188737	put. IS	Unknown	406/406 (100%) AAF60967; 89/166 (53%) NP_754342	insertion element
ORF17	18342-19559	AF188737	put. transposase	transposase	406/406 (100%) AAF60967	IS element-related
ORF18	19581-20096	AF188737	put. transposase	unknown	172/172 (100%) AAF60966	IS element-related
ORF19	21126-21674	AE000502	*fimE*	recombinase involved in phase variation; regulator for *fimA* (FimE)	183/183 (100%) NP_418733	regulator; surface structures
ORF20	22156-22703	D13186	*fimA*	major subunit of type 1 fimbriae (FimA)	154/182 (84%) AAG35675	structural component; surface structures
ORF21	22661-23305	AP002569	*fimI*	fimbrial protein (FimI)	213/215 (99%) NP_418735	structural component; surface structures
ORF22	23344-24067	AE000502	*fimC*	periplasmic chaperone, required for type 1 fimbriae (FimC)	238/241 (98%) BAB38698	structural component; surface structures
ORF23	24137-26770	AE000502	*fimD*	outer membrane protein; export and assembly of type 1 fimbriae (FimD)	842/878 (95%) BAB38699	membrane; outer membrane constituents
ORF24	26783-27310	AE000502	*fimF*	fimbrial morphology (FimF)	172/176 (97%) BAB38700	structural component; surface structures
ORF25	27326-27826	AP002569	*fimG*	fimbrial morphology (FimG)	151/167 (90%) NP_757247	structural component; surface structures
ORF26	27885-28748	AP002569	*fimH*	minor fimbrial subunit, D-mannose specific adhesin (FimH)	286/288 (99%) NP_418740	structural component; surface structures
ORF27	28928-30268	AE000503	*gntP*	gluconate transport system permease 3 (GntP)	413/447 (92%) NP_313307	transport; transport of small molecules: carbohydrates, organic acids, alcohols
ORF28	30608-31789	AE000502	*uxuA*	mannonate hydrolase (UxuA)	394/394 (100%) NP_757250	enzyme; degradation of small molecules: carbon compounds
ORF29	31873-33330	AE000503	*uxuB*	D-mannonate oxidoreductase (UxuB)	486/486 (100%) NP_290939	enzyme; degradation of small molecules: carbon compounds

Table 6 continued

ORF30	33548-34318	AE000503	*uxuR*	regulator for *uxu* operon (UxuR)	257/257 (100%) NP_313310	regulator; degradation of small molecules: carbon compounds
ORF31	34462-35248	AE000503	*yjiC*	orf, hypothetical protein (YjiC)	208/229 (90%) NP_313311	orf; unknown

Table 7: Characteristics of the ORFs located in the $bcs_{\text{Nissle 1917}}$ gene cluster.

Putative ORF	nt positions [bp]	Accession no.	DNA homology to	Similarity of the encoded protein	Identity of the aa sequence (%) accession no.	Function
ORF1	237-3710	AE016768	*bcsC/yhjL*	Cellulose synthase operon protein C (YhjL)	1122/1157(96%) NP_756204	Putative cellulose biosynthesis-related
ORF2	3692-4798	AE016768	*bcsZ/yhjM*	Putative cellulose biosynthesis-related (YjhM)	367/368 (99%) NP_756205	Putative cellulose biosynthesis-related
ORF3	4805-7144	AE016768	*bcsB/yhjN*	Cellulose synthase regulatory subunit (YjhN)	735/779 (94%) NP_756206	Cyclic di-GMP binding protein precursor
ORF4	7155-9773	AE016768	*bcsA/yhjO*	Cellulose synthase catalytic subunit [UDP-forming] (YjhO)	860/872 (98%) NP_756207	Putative cellulose biosynthesis-related
ORF5	9770-10522	AE016768	*yhjQ*	Hypothetical protein (YhjQ)	227/227 (100%) NP_756208	Putative cellulose biosynthesis-related
ORF6	10534-10722	AE016768	*yhjR*	Hypothetical protein (YhjR)	62/62 (100%) NP_756209	Putative regulator of cellulose biosynthesis-related; putative cellulose synthase

Table 8: Detection of the presence of *E. coli* Nissle 1917-specifc DNA regions in the genomes of different *E. coli* strains by PCR. For positive control reaction *E. coli* Nissle 1917 genomic DNA was used as a template; for negative control reaction *E. coli* MG1655 genomic DNA was used as a template.
" –" – negative PCR reaction; "+"- positive PCR reaction; "(+)" – presence of PCR product of correct size, but weaker compare to the positive control.

Strain/PCR	*mcmD*	*focG*	*iroB*	*iuc* (*iucB* – *iucC*)	*sat*	*iha*	GEI III (ORF23-ORF24)	HPI 3 R (*irp1* – *ybtT*)	HPI 4 L (*irp2*)	GEI I left (Z4866)	GEI I right (c1274)	GEI II left (*shiA*)	GEI II right (*kfiC*)	IS*2*-left (IS*2*–*iuc shiF*)	IS*2*-right (IS*2*–*iha*)
ECOR7 (A)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR1 (A)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR23 (A)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR25 (A)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR28 (B1)	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
ECOR32 (B1)	+	-	-	-	-	-	-	-	-	+	-	+	-	-	-
ECOR42 (E)	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-
ECOR59 (B2)	-	-	-	+	-	-	-	+	+	-	-	+	-	-	-
F18	-	-	+	+	-	-	-	+	+	-	-	+	-	-	-
M1/5	-	-	-	+	+	+	-	+	+	-	-	+	+	+	+
M1/6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M1/32	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
M1/49	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M2/24	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
M2/39	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
M2/40	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
M2/43	-	-	-	+	-	+	-	-	-	-	-	+	-	-	-
M3/6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M3/15	-	-	-	-	-	+	-	-	-	-	+	+	-	-	-
M3/26	-	+	-	+	+	+	-	+	+	+	+	+	-	-	+
M1/13	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
M1/43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M1/14	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
M1/15	-	-	-	-	-	+	-	-	-	+	+	-	-	-	-
M1/30	-	-	+	+	+	-	-	+	+	-	+	+	-	-	-
M1/37	-	-	+	+	-	-	-	-	-	+	-	+	-	-	-
M2/16	-	-	+	+	-	+	-	+	+	-	-	+	-	-	-
M2/37	+	-	+	+	-	-	-	-	-	+	(+)	+	-	-	-
M2/46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M3/27	-	-	-	+	+	+	-	+	+	+	+	+	-	-	+
M3/29	-	+	+	+	+	+	-	-	-	+	+	+	-	-	-
M3/22	-	-	+	+	-	+	-	-	-	+	+	+	-	-	-
M3/36	-	-	+	+	+	-	-	-	-	+	+	+	-	-	-
M2/17	-	-	+	+	-	-	-	-	-	-	+	+	-	-	-
M1/18	+	+	+	+	+	+	-	-	-	+	+		+	+	-
M2/31	-	-	+	-	-	-	-	-	-	+	+	+	-	-	-
M1/36	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
M2/1	-	-	-	+	+	+	-	-	-	+	+	+	-	-	+
M2/47	+	-	-	+	-	+	-	-	-	-	+	-	-	-	-
M3/3	+	-	+	+	-	+	-	-	-	+	+	+		-	-
Nissle 1917	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MG1655	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RZ411	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
RZ412	-	-	+	+	+	+	+	+	+	-	-	+	-	-	-
RZ418	+	+	+	+	-	+	-	+	+	+	+	+	-	-	-
RZ422	-	-	+	+	-	-	-	+	+	+	+	+	-	-	-
RZ423	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ424	+	+	+	+	-	+	-	+	+	+	+	+	+	-	-
RZ430	+	+	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ436	+	-	+	-	-	-	-	+	+	-	+	+	-	-	-
RZ446	+	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ447	+	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ448	+	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ449	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+
RZ451	-	-	+	+	-	-	-	+	+	-	+	+	-	-	-
RZ454	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ458	+	+	+	+	+	+	+	+	+	-	+	+	-	-	-
RZ460	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ461	+	-	+	-	-	-	-	-	+	+	+	+	-	-	-

Table 8 continued

RZ462	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ463	-	-	+	-	-	-	-	-	+	+	+	+	-	-	-
RZ467	+	+	+	+	+	+	-	+	+	+	+	+	-	-	-
RZ470	-	-	-	+	-	+	-	+	+	-	+	+	-	-	-
RZ479	-	-	-	-	-	-	-	+	+	+	-	+	-	-	-
RZ484	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ485	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ486	-	+	-	+	+	+	-	+	+	+	-	+	-	-	-
RZ496	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ501	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+
RZ502	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ504	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ505	+	+	+	-	-	-	+	+	+	+	+	+	-	-	-
RZ507	+	+	+	+	-	+	-	+	+	+	+	+	-	-	-
RZ532	+	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ536	+	-	+	-	-	-	-	+	+	+	+	+	-	-	-
RZ536/1	+	-	+	-	-	-	-	+	+	+	+	+	-		
RZ537	+	+	+	+	+	+	-	+	+	+	+	+	-	-	-
RZ439	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+
RZ442	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
RZ471	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
RZ525	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
RZ498	-	-	+	+	+	+	-	+	+	+	+	+	+	+	+
RZ495	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-
RZ440	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+
RZ441	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+
RZ468	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+
RZ475	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ526	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ450	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-
RZ414	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ419	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ420	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ521	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ522	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
RZ500	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
RZ466	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ452	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ453	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
RZ465	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
RZ443	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ512	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
RZ513	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
RZ533	+	+	+	+	+	+	-	+	+	+	+	+	+	-	-
RZ421	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ429	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
RZ477	+	-	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ519	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
RZ523	+	+	+	+	-	+	+	+	+	+	+	+	+	-	-
RZ499	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
RZ524	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

1. Sequence determination of *csg* promoter – *csgD* region of *E. coli* strains Nissle 1917, Nissle 1917 CR- and 536

E. coli Nissle 1917:

```
   1 attctgccgc cacaatccag cgtaaataac gtttcatggc tttatcgcct
gaggttatcg tttgcccagg aaaccgcttg 80
    81 tgtccggttt tttacggcta tcttcttgaa aagattataa agatgcgttt
taaccgtatt ttcgctgatg aacaacgaac 160
   161 gagcgatctc gttattagac gcgccgatac gcagcttatt caggatctct
ttttcccgat gagtaaggag ggctgattcc 240
   241 gtgctgttat aacgatagtt acctgaatgc gtaatcaggt aactggcaag
cttttgcgta aagtagcatt cgccgcgcag 320
   321 gacgccttgc aacccattga caacacgttc ttgatcctcc atggcataaa
aaacgccgtt gatatgaggc cagttttcaa 400
   401 tgtcgcggta cgggtaatct tcaggcgtat ttagcaacaa tattttgata
ttgttgtttt tcctgctcaa agtatcctgc 480
   481 caataatgga taagcttttt atccgcttcc atcatatcca gaagaataat
agagcctgaa gatatatcgt ccagagaacg 560
```

```
   561 ttgaatatta tgtaattttc ctgtaattgc cagcgattgt ttaaggtgct
gcaagagagc tgtcgcctgc aaagaaggtt 640
   641 tagtgatcaa caataatgta tgaccatgaa tactatggac ttcattaaac
atgatgaaac cccgcttttt ttattgatcg 720
   721 cacacctgac agctgcctct aaaatagaag caccagaagt actgacagat
gttgcactgc tgtgtgtagt aataaatcag 800
   801 ccctaaatgg gtaaaatata aaactaatgg attacatctg atttcaatct
agccattaca aatcttaaat caagtgttaa 880
   881 acatgtaact aaatgtaact cgttatatta aaatgttaac cttaaggttt
tattaagttt ataaatgata gaaaagttgt 960
   961 acatttggtt tttattgcac aattttaaaa aatcatacaa atggtgctaa
cttactaata atgcatataa aaaatatttc 1040
  1041 gctgtagtcc tttcgtcatg aaaaacgttc ttgtttttc tccataccac
cgtggacaat tttttactgc aaaaagacga 1120
  1121 ggtttgtcac ggcttgtgcg caagacatat cgcagcaatc agcgacgggc
aagaagaatg actgtctggt gatttttgat 1200
  1201 agcggaaaac ggagatttaa aagaaaacaa aatatttttt tgcgtagata
acagcgtatt tacgtgggtt ttaatacttt 1280
  1281 ggtatgaaca aaaaaagaaa aatacaacgt gcgggtgagt tattaaaaat
atttccgcag acatactttc catcgtaacg 1360
  1361 cagcgttaac aaaatacagg ttgcgttaac aaccaagttg aaatgattta
atttcttaaa tgtacgacca ggtccagggt 1440
  1441 gacaacatga aaaacaaatt gttatttatg atgttaacaa tactgggtgc
g                                1491
```

E. coli Nissle 1917 CR-:

```
     1 attctgccgc cacaatccag cgtaaataac gtttcatggc tttatcgcct
gaggttatcg tttgcccagg aaaccgcttg 80
    81 tgtccggttt tttacggcta tcttcttgaa aagattataa agatgcgttt
taaccgtatt ttcgctgatg aacaacgaac 160
   161 gagcgatctc gttattagac gcgccgatac gcagcttatt caggatctct
ttttcccgat gagtaaggag ggctgattcc 240
   241 gtgctgttat aacgatagtt acctgaatgc gtaatcaggt aactggcaag
cttttgcgta aagtagcatt cgccgcgcag 320
   321 gacgccttgc aacccattga caacacgttc ttgatcctcc atggcataaa
aaacgccgtt gatatgaggc cagttttcaa 400
   401 tgtcgcggta cgggtaatct tcaggcgtat ttagcaacaa tattttgata
ttgttgtttt tcctgctcaa agtatcctgc 480
   481 caataatgga taagcttttt atccgcttcc atcatatcca gaagaataat
agagcctgaa gatatatcgt ccagagaacg 560
   561 ttgaatatta tgtaattttc ctgtaattgc cagcgattgt ttaaggtgct
gcaagagagc tgtcgcctgc aaagaaggtt 640
   641 tagtgatcaa caataatgta tgaccatgaa tactatggac ttcattaaac
atgatgaaac cccgcttttt ttattgatcg 720
   721 cacacctgac agctgcctct aaaatagaag caccagaagt actgacagat
gttgcactgc tgtgtgtagt aataaatcag 800
   801 ccctaaatgg gtaaaatata aaactaatgg attacatctg atttcaatct
agccattaca aatcttaaat caagtgttaa 880
   881 acatgtaact aaatgtaact cgttatatta aaatgttaac cttaaggttt
tattaagttt ataaatgata gaaaagttgt 960
   961 acatttggtt tttattgcac aattttaaaa aatcatacaa atggtgctaa
cttactaata atgcatataa aaaatatttc 1040
  1041 gctgtagtcc tttcgtcatg aaaaacgttc ttgtttttc tccataccac
cgtggacaat tttttactgc aaaaagacga 1120
```

```
  1121 ggtttgtcac ggcttgtgcg caagacatat cgcagcaatc agcgacgggc
aagaagaatg actgtctggt gatttttgat 1200
  1201 agcggaaaac ggagatttaa aagaaaacaa aatatttttt tgcgtagata
acagcgtatt tacgtgggtt ttaatacttt 1280
  1281 ggtatgaaca aaaaaagaaa aatacaacgt gcgggtgagt tattaaaaat
atttccgcag acatactttc catcgtaacg 1360
  1361 cagcgttaac aaaatacagg ttgcgttaac aaccaagttg aaatgattta
atttcttaaa tgtacgacca ggtccagggt 1440
  1441 gacaacatga aaaacaaatt gttatttatg atgttaacaa tactgggtgc
g                                1491
```

E. coli 536:

```
     1 attctgccgc cacaatccag cgtaaataac gtttcatggc tttatcgcct
gaggttatcg tttgcccagg aaaccgcttg 80
    81 tgtccggttt tttacggcta tcttcttgaa aagattataa agatgcgttt
taaccgtatt ttcgctgatg aacaacgaac 160
   161 gagcgatctc gttattagac gcgccgatac gcagcttatt caggatctct
ttttcccgat gagtaaggag ggctgattcc 240
   241 gtgctgttat aacgatagtt acctgaatgc gtaatcaggt agctggcaag
cttttgcgta aagtagcatt cgccgcgcag 320
   321 gacgccttgc aacccattga caacacgttc ttgatcctcc atggcataaa
aaacgccgtt gatatgaggc cagttttcaa 400
   401 tgtcgcggta cgggtaatct tcaggcgtat ttagcaacaa tattttgata
ttgttgtttt tcctgctcaa agtatcctgc 480
   481 caataatgga taagcttttt atccgcttcc atcatatcca gaagaataat
agagcctgaa gatatatcgt ccagagaacg 560
   561 ttgaatatta tgtaattttc ctgtaattgc cagcgattgt ttaaggtgct
gcaagagagc tgtcgcctgc aaagaaggtt 640
   641 tagtgatcaa caataatgta tgaccatgaa tactatggac ttcattaaac
atgatgaaac cccgcttttt ttattgatcg 720
   721 cacacctgac agctgcctct aaaatagaag caccagaagt actgacagat
gttgcactgc tgtgtgtagt aataaatcag 800
   801 ccctaaatgg gtaaaatata aaactaatgg attatatctg atttcaatct
agccattaca aatcttaaat caagtgttaa 880
   881 acatgtaact aaatgtaact cgttatatta aaatgttaat ctcaaggttt
tattaagttt ataaatgata gaaaagttgt 960
   961 acatttggtt tttattgcac aattttaaaa aatcatacaa atagtgctaa
cttactaata atgcatataa aaaatatttc 1040
  1041 gctgtagtcc tttcgtcatg aaaaacgttc ttgtttttc tccataccac
cgtggacaat tttttactgc aaaaagacga 1120
  1121 ggtttgtcac ggcttgtgcg caagacatat cgcagcaatc agcgacgggc
aagaagaatg actgtctggt gatttttgat 1200
  1201 agcggaaaac ggagatttaa aagaaaacaa aatatttttt tgcgtagata
acagcgtatt tacgtgggtt ttaatacttt 1280
  1281 ggtatgaaca aaaaaagaaa aatacaacgc gcgggtgagt tattaaaaat
atttccgcag acatactttc catcgtaacg 1360
  1361 cagcgttaac aaaatacagg ttgcgttaac aaccaagttg aaatgattta
atttcttaaa tgtacgacca ggtccagggt 1440
  1441 gacaacatga aaaacaaatt gttatttatg atgttaacaa tactgggtgc
g                                1491
```

2. Acronyms and abbreviations

A	amper
aa	amino acid
Ap/Ap^R	ampicillin/ampicillin resistence
APS	NH_4-persulfate
bp	base pair
°C	grad celsius
cat	chloramphenicol-acetyl-transferase
Da	dalton
dATP	desoxyadenosin-5'-triphosphate
dCTP	desoxycytosin-5'-triphosphate
dGTP	desoxyguanosin-5'-triphosphate
dH_2O	distilled water
DNA	desoxyribonucleic acid
dNTP	desoxynucleotide
dTTP	desoxythymidin-5'-triphosphate
E.	*Escherichia*
EDTA	ethylene-diamin-tetraacetate
et al.	et alili (and others)
EtBr	ethidium bromide
EtOH	ethanol
Fig.	figure
g	gram
GEI	genomic island
h	hour
IPTG	isopropyl-β-D-thiogalactopyranosid
kb	kilobase (pairs)
Km/Km^R	kanamycin/kanamycin resistence
l	liter
LB	Luria Bertani broth
LPS	lipopolysaccharide
m	meter; milli (1×10^{-3})
M	molar, mega (1×10^{6})

mg	milligram
min	minute
ml	milliliter
OD	optical density
ON	overnight
ORF	open reading frame
ori	origin of replication
PAA	polyacrylamid
PAI	pathogenicity island
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RNA	ribonucleic acid
RNase	ribonuclease
RT	room temperature
SDS	sodium-dodecyl-sulfate
sec	second
Sm/Sm^R	streptomycin/streptomycin resistence
SSC	standart saline citrate
t	time
Tab.	table
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TCA	trichloroacetic acid
TE	Tris-EDTA
TEMED	N,N,N`,N`-tetramethyldiamin
Tris	Tris-(hydroxymethyl)-aminomethan
U	enzyme unit (1 U=1 µmol substrate × min^{-1})
UV	ultraviolet irradiation
V	volt
v/v	volume/volume
W	watt
w/v	weight/volume
WT	wild type
X-gal	5-bromo-4-chloro-3-indolyl-β-glucoside

YT	yeast triptone
µ	micro (1×10^{-6})
µg	microgram
µl	microliter

Publications and poster presentations at international meetings:

Publications:

1. **Grozdanov, L., Zähringer, U., Blum-Oehler, G., Brade, L., Henne, A., Knirel, Y. A., Schombel, U., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., Rietschel, E. T., and U. Dobrindt.** 2003. A single nucleotide exchange in the *wzy* gene is responsible for the semirough O6 lipopolysaccharide phenotype and serum sensitivity of the *Escherichia coli* strain Nissle 1917. J. Bacteriol. **184:**5912-5925.
2. **Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., and U. Dobrindt.** 2003. Analysis of the genome structure of probiotic *Escherichia coli* strain Nissle 1917. *Manuscript submitted.*

Poster presentations at international meetings:

1. **A genomic analysis of the probiotic *E. coli* strain DSM6601.** Grozdanov, L., Blum-Oehler, G., Dobrindt , U., and Hacker, J. Meeting of the European Graduate College "Gene regulation in and by microbial pathogens", 16–20 October 2000, Schloß Zeilitzheim, Germany.

2. **Comparative genomic analysis of the probiotic *E. coli* strain DSM6601.** Grozdanov, L., Blum-Oehler, G., Dobrindt, U., Gottschalk, G., and Hacker, J. International Workshop "Microbial-Host Interactions; Approaches and Molecular Tools", June 10-16, 2001, Umeå, Sweden.

3. **Molecular characterization of the O6 lipopolysaccharide-encoding genetic determinant of the *E. coli* strain DSM6601.** Grozdanov, L., Dobrindt, U., Blum-Oehler, G., Henne, A., Gottschalk, G., Schulze, J., Sonnenborn, U., and Hacker, J. H-48. ASM general Meeting, 2002, Salt Lake City, USA.

Curriculum Vitae

Name: Lubomir Assenov Grozdanov
Date of birth: 18. October 1975
Place of birth: Sofia, Bulgaria

Education

1993 Graduated from Highschool ”Prof. Dr. Assen Zlatarov“, Sofia, Bulgaria

1993 – 1998 Graduation from Sofia University „St. Kliment Ohridski“ as Master of Science, major subjects: Molecular Biology and Microbiology

1998 Master Thesis (“Molecular taxonomy analysis of *Bacillus sphaericus* strains of different environmental origin”) at the Department of Genetics, Institute of Microbiology, Bulgarian Academy of Science and successfully passed Master Thesis defence

1998 – September 1999 Participation in the European TEMPUS Project “Second Master Degree in Environmental Biotechnology“

1999 Second Master Thesis at the Institute for Microbiology and Genetics, Georg–August Universität, Göttingen, Germany (“Physiological role of the 2-phosphoglycolate phosphatase for the autotrophic growth of *Ralstonia eutropha*”) and successfully passed Master Thesis defence

since September 1999 Ph.D. thesis at the Institute for Molecular Biology of Infectious Diseases, University of Würzburg, Germany. Title of the Ph.D. project: “Comparative genomic analysis of the the probiotic *Escherichia coli* strain Nissle 1917 (O6:K5:H1)”. Scientific

supervisors: Prof. Dr. Dr. h. c. Jörg Hacker

2000 – 2003 Associated member of the European Graduate College "Gene regulation in and by microbial pathogens" of the Universities of Würzburg and Umeå (Sweden)

2000 – 2004 Ph.D. scholarship of the Bavarian Research Foundation (Bayerische Forschungsstiftung)

Lebenslauf

Name: Lubomir Assenov Grozdanov
Geburtsdatum: 18.10.1975
Geburtsort: Sofia, Bulgarien

Ausbildung

1993	Erlangung der Hochschulreife am ”Prof. Dr. Assen Zlatarov“ - Gymnasium, Sofia, Bulgarien
1993 – 1998	Erlangung des Magistergrades der Biologischen Fakultät an der “St. Kliment Ochridski“ Universität, Sofia (Bulgarien)
1998	Diplomarbeit (“Molekular- und Taxomanalyse des Stammes *Bacillus sphaericus*”), Institut für Mikrobiologie, Bulgarische Akademie der Wissenschaften, Sofia (Bulgarien)
1998 – September 1999	Teilnahme am Magisterprogramm im Rahmen des TEMPUS Projektes “Second Master Degree in Environmental Biotechnology“ (Nr. 11454/96) der Europäischen Union
1999	Diplomarbeit am Institut für Mikrobiologie und Genetik, Georg–August Universität, Göttingen (“2-Phosphoglykolat-phosphatase: Physiologische Rolle für das autotrophe Wachstum von *Ralstonia eutropha*”); Magistergrad in Umweltbiotechnik
seit September 1999	Dissertation am Institut für Molekulare Infektionsbiologie der Julius-Maximilians-Universität zu Würzburg unter Anleitung von Prof. Dr. Dr. h. c. J. Hacker (“Untersuchungen zur Genomorganisation und zur Fitness des apathogener *Escherichia coli* Stammes Nissle 1917 (O6:K5:H1)“)

2000 – 2003	Mitglied des Europäischen Graduiertenkollegs “Gene regulation in and by microbial pathogens”, Universität Würzburg und Universität Umeå (Schweden)
2000 – 2004	Promotionsstipendiat der Bayerischen Forschungsstiftung