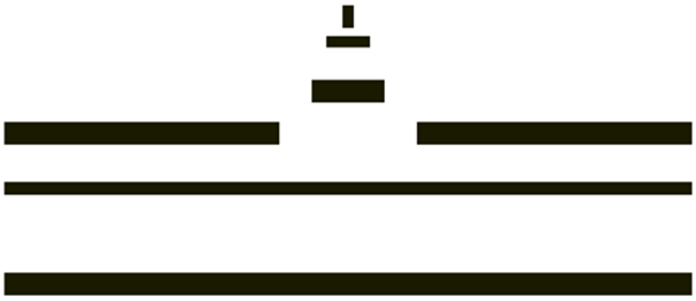
Biology

Analysis of adaptation strategies of uropathogenic *Escherichia coli* during growth in the bladder



Doctoral thesis for submission to a doctoral degree at the
Faculty of Natural Science
of the Westfälische Wilhelms-Universität Münster

submitted by

Birgit Ewert
from
Haselünne, Germany

Münster, 2015

Submitted on: .....

Office stamp

Dean:	Prof. Dr. Michael Weber
Primary Supervisor:	Prof. Dr. Ulrich Dobrindt
Supervisor (second):	Prof. Dr. Eva Liebau

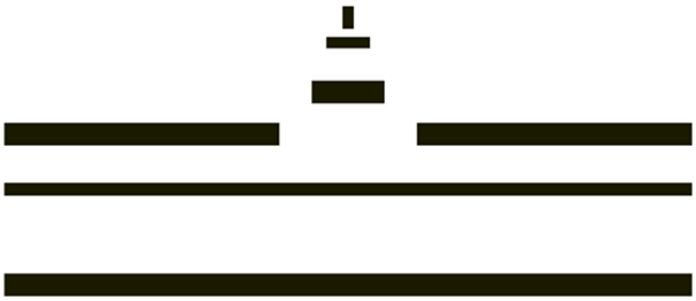
Date of Public Defense:	21.05.2015
Date of Receipt of Certificates:	05.06.2015

Only the one who knows his goal will find the path.

Laotse

Biologie

Untersuchung von Anpassungsstrategien uropathogener *Escherichia coli* bei Wachstum in der Harnblase



Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich Biologie
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von

Birgit Ewert
aus
Haselünne

Münster, 2015

Eingereicht am: .....
Bürostempel

Dekan: Prof. Dr. Michael Weber
Erster Gutachter: Prof. Dr. Ulrich Dobrindt
Zweite Gutachterin: Prof. Dr. Eva Liebau

Tag des Promotionskolloquiums: 21.05.2015
Doktorurkunde ausgehändigt am: 05.06.2015

Affidavit

I hereby declare that I have produced the presented thesis by myself and without unpermitted help, that all sources and aids used are indicated, and that this dissertation has not been presented elsewhere as an examination paper.

Münster, ...
(signature)

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorgelegte Dissertation selbst und ohne unerlaubte Hilfe angefertigt, alle in Anspruch genommenen Quellen und Hilfsmittel in der Dissertation angegeben habe und die Dissertation nicht bereits anderweitig als Prüfungsarbeit vorgelegen hat.

Münster, ...
(Unterschrift)

Acknowledgements

First of all, I would like to thank my supervisor Prof. Dr. Ulrich Dobrindt for the opportunity to work on this interesting project. Thanks for your intensive support, the countless discussions and for sharing your great knowledge in the field of *E. coli* research.

Secondly I´d like to thank Prof. Dr. Eva Liebau and Prof. Dr. M. Alexander Schmidt for being part of my thesis committee and for the takeover of the second and third expertise.

Thanks to Prof. Dr. Claudia Acquisti for supporting me in the field of evolutionary biology and for the fruitful discussions on my thesis. In this context I would also like to thank Robert Fürst for providing me insights into the complex world of R and supporting me in all questions about bioinformatics.

I´d also like to thank Prof. Dr. Michael Lalk and Kirsten Dörries for the performance of the extracellular metabolome analysis.

Dr. Michael Berger, many thanks for your constant helpfulness, the straight forward ideas that pushed the project in the right direction and for the proof-reading of my thesis. I´m proud of being your favorite PhD student ;-).

Big thanks are dedicated to Olena Mantel for her perfect practical help and technical support of my project at any time. Here I´d also like to mention the whole ICB crew for the very nice working atmosphere and the constant support, encouragement and help. Special thanks to Vanja for support with RT-PCR analyses, Chrissy for help with R, and Nadine & Marina^2 for their general helpfulness.

Apart from the scientific interaction I´m grateful for all the wonderful and funny moments we shared in the lab and besides work. Special thanks go to Sadrick for his daily visits, the maintenance of a fair mandarine trade and the frequent laugh breaks that nobody else understood ;-). Further special thanks go to Chrissy for the unforgettable adventures that we experienced and above all survived – needs to be continued! Many thanks go as well to Knödel, Marina^2, the Italian fraction Nadine & Vanja, Sarah, Roswitha, Johannes, Gabi, Karin, Matt, Mr. Wilson and all of the other current and former members of the working group. Without you, it would have made only as half as much fun!

Special thanks go to my whole family, especially my sisters, who always supported me at every stage of my life. Particular thanks for the beloved sugary motivation packages ;-).

Last but not least, thank you Michael, for your continuous support, encouragement and your talent to make me laugh at any time!

Table of contents

List of Figures ..... **VI**

List of Tables ..... **IX**

List of Abbreviations ..... **X**

Summary **1**

Zusammenfassung **2**

1 Introduction **4**

1.1 *Escherichia coli* – a bacterial species of clinical relevance 4

1.2 Urinary tract infections 5

1.3 Uropathogenic *E. coli* – the main causative of UTI 7

1.3.1 Virulence factors 8

1.4 Asymptomatic bacteriuria (ABU) 10

1.4.1 *E. coli* 83972 11

1.5 Genome plasticity and bacterial adaptation 11

1.6 Two-component systems 13

1.6.1 TCS BarA/UvrY 16

1.6.2 Role of the carbon storage regulator system 18

1.7 Metabolism 20

1.7.1 Glycolysis 20

1.7.2 Gluconeogenesis 21

1.7.3 Pentose Phosphate shunt 22

1.7.4 Tricarboxylic Acid Cycle (TCA) 22

1.7.5 Mixed Acid Fermentation 22

1.8 Aims of the study 23

2 Material **25**

2.1 Bacterial strains 25

2.2 Plasmids 27

2.3 Oligonucleotides 27

2.4 Chemicals, enzymes and kits 29

2.5 Buffers and solutions 30

2.6 Media and agar plates 30

2.6.1 Media 30

2.6.2 Agar plates 31

2.7 Antibiotics 32

2.8 Antibodies 32
2.9 DNA / protein size markers 32
2.10 Technical equipment 33

3 Methods 35
3.1 Bacterial cultures 35
3.1.1 Cultivation 35
3.1.2 Bacterial preservation 35
3.1.3 Bacterial growth 35
3.2 Molecular biological methods - DNA 35
3.2.1 Extraction of genomic DNA 35
3.2.2 Photometric determination of nucleic acid concentration 36
3.2.3 Polymerase chain reaction 36
3.2.4 Analysis of PCR fragments 38
3.2.5 Gel extraction 39
3.2.6 Ethanol precipitation of DNA 39
3.2.7 PCR clean-up 39
3.2.8 DNA sequencing 39
3.2.8.1 PCR purification by EXOSAP 40
3.2.8.2 Sequencing protocol 40
3.2.9 Preparation of electrocompetent cells 40
3.2.10 Transformation of electrocompetent cells 41
3.2.11 Isolation of plasmids 41
3.2.12 Enzymatic hydrolysis of DNA 41
3.2.13 Ligation of DNA fragments 42
3.2.14 Colony PCR 42
3.2.15 Dot blot 42
3.2.15.1 Vacuum Blotting 42
3.2.15.2 Probe labeling 43
3.2.15.3 Hybridization and detection 43
3.2.16 Southern blot 44
3.2.16.1 Capillary Blotting 44
3.2.16.2 Probe labeling, hybridization and detection 44
3.3 Construction of *uvrY* allelic variants 45
3.3.1 Site-specific recombination by recombineering according to Datsenko & Wanner 45
3.3.2 Genetic modification by a counter-selection system 46
3.3.3 P1 mediated transduction 48

3.4 Molecular biological methods - RNA 49
3.4.1 RNA Isolation 49
3.4.1.1 RNA isolation from urine 49
3.4.1.2 RNA isolation from LB 49
3.4.2 DNA digestion 49
3.4.3 RNA quality check 50
3.4.4 Reverse transcription (RT) for cDNA synthesis 50
3.4.5 Quantitative real-time PCR 51
3.5 Phenotypic characterization 52
3.5.1 Growth kinetics 52
3.5.2 Biofilm assay 52
3.5.3 Motility assay 53
3.5.4 Congo Red staining assay 53
3.5.5 Calcoflour assay 53
3.5.6 *In vitro* competition assay 53
3.5.7 Sub-cultivation assay 54
3.6 Protein biochemical methods 54
3.6.1 Isolation of proteins – whole cell lysate 54
3.6.1.1 BCA quantification of protein concentration 54
3.6.2 Tricine-SDS-Page 55
3.6.3 Coomassie staining 56
3.6.4 Western blot 56
3.6.4.1 Transfer of proteins - `semi-dry blot` 56
3.6.4.2 Immune detection 57
3.6.4.3 ECL detection 57
3.7 Metabolomics 58
3.7.1 Metabolome sample preparation 58
3.8 *In silico* analysis 59
3.9 Statistics 59

4 Results 60
4.1 Extracellular metabolome 60
4.2 Limitations for growth in urine 65
4.3 Stress response to growth in urine 66
4.4 Sequence variation in different groups of *E. coli* isolates 69
4.4.1 Amplification and sequencing of *barA/uvrY* 69
4.4.2 BarA/ UvrY sequence analysis 74

4.5 Phylogenetic classification 85
4.6 Test of selection – McDonald-Kreitman Test 87
4.7 SNP combination analysis 90
4.8 Phenotypic characterization 91
4.8.1 Biofilm 91
4.8.2 Synthesis of curli fimbriae and cellulose expression 93
4.8.3 Motility assay 95
4.8.4 Growth characteristics 97
4.9 Expression of *csrB* 97
4.10 Impact analysis of non-synonymous SNPs in *uvrY* – allelic variants 100
4.10.1 Selection of *uvrY* alleles 100
4.10.2 Construction of *uvrY* allelic variants 101
4.10.3 Phenotypic assays 103
4.10.3.1 Biofilm 103
4.10.3.2 Expression of curli, cellulose & swarming motility 103
4.10.3.3 Growth characteristics 106
4.10.3.4 Competition assays 107
4.10.3.5 Sub-cultivation assay 112
4.10.4 Real-Time PCR 113

5 Discussion 117
5.1 Basic metabolomics vs. adapted metabolome 117
5.2 Carbon – a limiting factor for growth in urine 122
5.3 Stress response is growth medium-dependent 124
5.4 Sequence context variation of *barA* and *uvrY* 127
5.5 Sequence variability of *barA* and *uvrY* 128
5.6 Correlation of SNPs and phylogeny 131
5.7 Detection of adaptive evolution for *barA* and *uvrY* 131
5.8 Impact of individual *barA*/*uvrY* alleles on bacterial phenotypes 133
5.8.1 Biofilm formation 133
5.8.2 Expression of components of the extracellular matrix 134
5.8.3 Motility 134
5.8.4 Growth 135
5.9 Gene expression of *csrB* 136
5.10 Selection of *uvrY* allelic variants 136
5.11 Phenotypic variations suggest “loss-of-function” as an adaptive trait 138
5.11.1 Biofilm formation 138

5.11.2 Expression of components of the extracellular matrix 139
5.11.3 Motility 140
5.11.4 Growth kinetics 141
5.12 Competition assay 142
5.13 Sub-cultivation 143
5.14 Mutated *uvrY* exhibits impact on the Csr system 144
5.14.1 Comparison of the expression of components of the Csr system under standard laboratory conditions 144
5.14.2 Comparison of the expression of components of the Csr system in urine 146

6 Conclusions and outlook 148

7 Appendix i
7.1 Supplements i
7.1.1 Amplification and sequencing of *barA/uvrY* i
7.1.2 Evolutionary rates ii
7.1.2.1 UvrY ii
7.1.2.2 BarA iii
7.1.3 Biofilm xi
7.1.4 Summary of the phenotypic assays xiii
7.2 References xvi
7.3 Publications xxxiii
7.4 Presentations xxxiii
7.5 Curriculum vitae xxxiv

List of Figures

Figure 1: Pathogenesis of urinary tract infection caused by uropathogenic *E. coli.* 6
Figure 2: Virulence factors of UPEC. 8
Figure 3: Impact of genome plasticity on adaptation of pathogenic bacteria. 12
Figure 4: Domain organization and signaling in two-component systems (TCSs). 14
Figure 5: Domain organization of BarA and UvrY. 17
Figure 6: Regulatory interactions between the BarA/UvrY TCS and the Csr system *in E. coli.* 19
Figure 7: Metabolic and biosynthetic pathways in *E. coli*. 21
Figure 8: DNA and protein markers used in this study. 33
Figure 9: Strategy for inactivation of chromosomal genes using PCR products (modified according to Datsenko & Wanner, 2000). 45
Figure 10: Schematic overview of the allelic exchange of *uvrY* variants into *E. coli* MG1655 background. 47
Figure 11: Schematic overview of a semi-dry blot application (modified according to BioRad). 57
Figure 12: Strain-specific changes of extracellular metabolites in pooled human urine upon growth of *E. coli* strains Nissle 1917, CFT073 and 83972. 61
Figure 13: Connection of central metabolic pathways in *E. coli.* 62
Figure 14: Concentrations of extracellular metabolites in pooled human urine measured during growth of *E. coli* strains 83972, CFT073 and Nissle 1917 for 24 h. 63
Figure 15: Kinetics of extracellular metabolites over a time period of 24 h. 64
Figure 16: Growth kinetics of *E. coli* strains 83972, CFT073 and Nissle 1917 in pooled human urine supplemented with different carbon and nitrogen sources at 37 °C. 66
Figure 17: Time-resolved RpoS expression of *E. coli* 83972 detected via Western blot. 67
Figure 18: Flowchart for the identification process of strain specific *barA*/ *uvrY* alleles. 70
Figure 19: Dot Blot approach (A) and PCR products (B) obtained during the screening for *barA*. 71
Figure 20: The deduced arrangement of IS elements detected in *barA*. 72
Figure 21: The deduced arrangement of IS elements detected in *barA*. 73
Figure 22: Graphical summary of the sensor histidine kinase BarA secondary structure and identified mutations in the four groups of isolates. 77
Figure 23: Graphical summary of the response regulator UvrY secondary structure and identified mutations in the four groups of isolates. 78
Figure 24: Graphical summary of the ribosome recycling factor *frr* secondary structure and the identified mutations in the four groups of isolates. 81
Figure 25: Minimum Spanning Tree (MST) based on occurring SNPs in *barA* portraying the distributions of ABU, HA-ABU, UTI and fecal *E. coli* isolates. 83

Figure 26: Minimum Spanning Tree (MST) based on occurring SNPs in *uvrY* (A) and *frr* (B) portraying the distribution of ABU, HA-ABU, UTI and fecal *E. coli* isolates. 84
Figure 27: Minimum Spanning Tree (MST) based on occurring SNPs in seven housekeeping genes portraying the distribution of ABU, HA-ABU, UTI and fecal *E. coli* isolates. 86
Figure 28: Conservation scores calculated for BarA, UvrY, Frr and the whole proteome of *E. coli* MG1655. 89
Figure 29: Box-Whisker-plot for the biofilm formation. 92
Figure 30: Phenotypes of the tested strains on Congo Red (CR) and Calcoflour (CF) agar plates. 94
Figure 31: Percentile distribution of non-motile, medium motile and highly motile strains on urine and LB swarming agar. 95
Figure 32: Swarming ability on LB and urine. 96
Figure 33: Growth characteristics of *E. coli* strains in (A) LB medium under aerobic conditions and (B) pooled human urine under static conditions at 37 °C. 98
Figure 34: Real Time-based quantification of *csrB* transcripts in pooled human urine. 99
Figure 35: Genomic organization of *uvrY* and its conserved domains. 100
Figure 36: Biofilm formation in LB and pooled human urine. 103
Figure 37: Curli fimbriae/cellulose expression and swarming motility of *E. coli* MG1655$^{(Zeo)}$ and the allelic variant strains. 105
Figure 38: Growth kinetics of the MG1655^{Zeo} wild type and the variant strains in (A) LB medium under aerobic conditions and (B) pooled human urine under static conditions at 37 °C. 107
Figure 39: Competition of the *E. coli* MG1655WT^{Zeo} vs. the different allelic variants in LB for 24 hours. 109
Figure 40: Competition of the *E. coli* MG1655WT^{Zeo} vs. the different allelic variants in urine for 24 hours. 110
Figure 41: Results of the competition assays in (A) LB and (B) pooled human urine. 111
Figure 42: Scheme of sub-cultivation assay. 112
Figure 43: Fitness of the variant strains after 96 h of sub-cultivation in (A) LB or (B) urine. 113
Figure 44: Real Time PCR-based quantification of transcriptional levels of *csrB, csrC, csrD, csrA* and *uvrY* of the allelic variants grown in LB. 115
Figure 45: Real Time PCR-based quantification of transcriptional levels of *csrB, csrC, csrD, csrA* and *uvrY* of the allelic variants grown in pooled human urine. 116
Figure 46: Decarboxylation reaction of lysine to cadaverine catalyzed by the lysine decarboxylase. 121
Figure 47: Binding of RNA polymerase to DNA 124
Figure 48: Schematic demonstration of the regulatory interactions between UvrY and the Csr components CsrB and CsrA and its effects on motility. 140

Figure 49: Model of the Csr regulatory circuitry (Yakhnin *et al.*, 2011). 145
Figure 50: The construction of *uvrY* based on the sequencing alignment. i
Figure 51: Biofilm formation of all isolates tested upon growth at 37 °C in LB. xi
Figure 52: Biofilm formation of all isolates tested upon growth at 37 °C in pooled human urine. xii

List of Tables

Table 1: *E. coli* pathotypes. 5
Table 2: Virulence factors of uropathogenic *E. coli* and their functions. 9
Table 3: Bacterial strains used in this study. 25
Table 4: Plasmids used in this study. 27
Table 5: oligonucleotides used in this study. 27
Table 6: Antibiotics used in this study. 32
Table 7: Primary antibody used in this study. 32
Table 8: Secondary antibody used in the study. 32
Table 9: Technical equipment used in this study. 33
Table 10: Pipetting instructions for a standard PCR using a Taq DNA Polymerase. 37
Table 11: Pipetting instructions for a PCR using the Phusion High-Fidelity DNA polymerase. 37
Table 12: qRT-PCR setup. 51
Table 13: Composition of the different gels for Tricine-SDS-PAGE 55
Table 14: Strain collection at the beginning of the study. 69
Table 15: From the study excluded *E. coli* strains. 74
Table 16: sequence variation found in total for *barA*, *uvrY* and *frr*. 75
Table 17: ns SNP found in the four groups of isolates and calculated to an average of ns SNP per isolate. 76
Table 18: Non-synonymous SNPs detected in *barA*, *uvrY* and *frr*. 79
Table 19: Results of the McDonald-Kreitman test for *uvrY*, *barA* and *frr*. 88
Table 20: Detected ns SNP combinations regarding the nucleotide positions in *uvrY* and *barA*. 90
Table 21: Selected *uvrY* alleles for the exchange of native *uvrY* of *E. coli* MG1655. 101
Table 22: Derivative of *E. coli* MG1655 with a zeocine resistance cassette and the constructed *uvrY* allelic variants used in this study. 102
Table 23: Summary of the results of the CR/CF plates and the swarm agar plates. 105
Table 24: Sigma factors and their functions in *E. coli*. 125
Table 25: Evolutionary rates of UvrY on the protein level. ii
Table 26: Evolutionary rates for BarA on the protein level. The evolutionary score is shown for each position of the protein. iii
Table 27: Summary of the results for the phenotypic assays for the tested strain panel (see 4.7). xiii

List of Abbreviations

(v/v)	volume per volume
(w/v)	weight per volume
^{1}H-NMR	proton nuclear magnetic resonance
A	absorbance/ Ampere
aa	amino acid
ABU	asymptomatic bacteriuria
Ag43	antigen 43
Amp	ampicillin
APEC	avian pathogenic *E. coli*
APS	ammonium persulfate
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
bp	base pairs
BR	biological replicates
BSA	bovine serum albumine
CA	catalytic and ATP-binding
c-di-GMP	cyclic diguanylate monophosphate
cDNA	complementary DNA
CEACAM	carcinoembryonic antigen related adhesion molecules
CF	Calcoflour
CFU	colony forming unit
Cm	chloramphenicol
CNF-1	cytotoxic necrotizing factor 1
CO_2	carbondioxide
CR	Congo Red
Csr	carbon storage regulator
CTAB	cetyl trimethylammonium bromide
DAEC	diffusely adherent *E. coli*
DAF	decay accelerating factor
dH_2O	deionized water
DHp	histidine phospotransfer domain
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleosidtriphosphate
DTT	dithiothreitol
E. coli	*Escherichia coli*
e.g.	*exempli gratia*
EAEC	enteroaggregative *E. coli*
ECOR	*E. coli* group of reference strain
EDTA	ethylenediaminetetraacetic acid
EHEC	enterohaemorrhagic *E. coli*
EIEC	enteroinvasive *E. coli*
EMP	Embden-Meyerhof-Parnas
EPEC	enteropathogenic *E. coli*
ETEC	enterotoxigenic *E. coli*
et al.	*et alii*
EXOSAP	exonuclease I and shrimp alkaline phosphatase
ExPEC	extraintestinal pathogenic *E. coli*
$FADH_2$	flavine adenine dinucleotide
FE	fast evolving
FHL	formate hydrogenlyase

FI	fecal isoltes
GC	gas chromatography
gDNA	genomic DNA
GEI	genomic island
H	flagellar
H_2	dihydrogen
HA-ABU	health-care associated ABU
HAMP	histidine kinases, adenyl cyclases, methyl binding proteins, phosphatases
hATPase_c	histidine kinase-type ATPase catalytic
HGT	horizontal gene transfer
HisKA	histidine kinase A
HK	histidine kinase
HlyA	α-hemolysin
HPK	histidine protein kinase
HPt	histidine phosphotransfer
HRP	horseradish peroxidase
HTH	helix-turn-helix
Ig	immunoglobulin
IL	interleukin
INDEL	insertion/deletion
IPEC	intestinal pathogenic *E. coli*
IS	insertion sequence
K	capsule
kb	kilo bases
LB	lysogeny broth
LPS	lipopolysaccharides
MLST	multi locus sequence typing
MNEC	neonatal meningitis-associated *E. coli*
mRNA	messenger RNA
MS	mass spectrometry
MST	minimum spanning tree
MW	molecular weight
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
ncRNA	non coding RNA
NH_4Cl	ammonium chloride
NI	neutrality index
NO	nitric oxide
no.	number
ns	non-synonymous
nt	nucleotide
Ntr	nitrogen regulated response
O	LPS
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PAGE	polyacrylamide gel electrophoresis
PAIs	pathogenic island
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDH	pyruvate dehydrogenase
PEP	phosphoenolpyruvate
PFL	pyruvate-formate-lyase

PGA	poly-β-1-6-N-glucosamine
PMN	polymorphonuclear leukocyte
PMSF	phenylmethanesulfonylfluoride
ppGpp	guanosine tetraphosphate
PVC	polyvinyl chloride
PVDF	polyvinylidene fluoride
qRT-PCR	quantitative real-time PCR
RBS	ribosome binding site
Rec	receiver
RIN	RNA integrity number
RNA	ribonucleic acid
RNAP	RNA polymerase
RNI	reactive nitrogen intermediates
rpm	revolutions per min
RpoS	RNA polymerase, sigma S factor
RR	response regulator
RT	room temperature
RT	reverse transcriptase
rxn	reaction
SAP	shrimp alkaline phosphatase
sat	secreted autotransporter toxin
SD	standard deviation
SDS	sodium dodecyl sulfate
SDS	sodium dodecyl sulfate
SE	slow evolving
SEM	standard error of the mean
SEPEC	sepsis-associated *E. coli*
SNP	single nucleotide polymorphism
spp.	species pluralis
ssp.	subspecies
ST	sequence types
sv.	serovar
SVT	sequence variation type
syn	synonymous
t	time
TAE	Tris acetate EDTA
TBE	Tris boric acid EDTA
TBS	Tris-buffered saline
TBS-T	TBS-Tween
TCA	tricarboxylic acid cycle
TCS	two-component system
TE	Tris EDTA
TEMED	tetramethylethylenediamine
TM	transmitter domain
U	unit
UPEC	uropathogenic *E. coli*
UTI	urinary tract infection
UV	ultra violet
VF	virulence factors
vol	volumes
vs	versus
WT	wild type
Zeo	zeocine

Units

°C	degree Celsius
g	gram
x g	gravitational force (9.81 m s^{-2})
h	hour(s)
kb	kilo base pairs
l	liter
M	molar
min	min
sec	second
V	volt
Ω	Ohm

Prefixes

µ	micro
c	centi
m	milli
n	nano
p	pico

IUPAC amino acid code

DNA

Nucleotide code	Base
A	adenine
C	cytosine
T (U)	thymine (or uracil)
G	guanine

Protein

Amino acid code	Three letter code	amino acid
A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartate
E	Glu	glutamic Acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

Summary

Urinary tract infections (UTI) represent one of the most common bacterial infectious diseases which affect millions of people per year and constitute a major health problem. They are mainly caused by bacteria of the species *Escherichia coli*. The most frequent form of UTI is asymptomatic bacteriuria (ABU) that is characterized by the absence of clinical symptoms in spite of large numbers of bacteria in the urine. The molecular mechanisms that allow ABU-associated *E. coli* strains to colonize the human urinary tract without provoking a significant host response are so far poorly understood. Virulence attenuation by genome reduction seems to be an evolutionary principle for the establishment of a commensal life style.

In the present study, we investigated evolutionary strategies that allow *E. coli* strains to successfully colonize the urinary tract. A special focus was placed on the investigation of metabolic constraints in urine. Therefore the extracellular metabolome of three closely related *E. coli* strains grown in urine was determined. The analyses revealed that *E. coli* is able to utilize a variety of nutrients for the production of precursor metabolites, reducing power and energy for further metabolic processes.

Further studies were directed to detect nutritional limitations during growth in urine. We identified carbon as a limiting factor for growth in urine. Carbon is one of the most important components of living organisms and therefore of essential importance.

Since the two-component system (TCS) BarA/UvrY was already shown to be frequently mutated in ABU *E. coli* strains upon prolonged growth in the bladder and as this TCS is known to play an important role in the regulation of carbon metabolism via the carbon storage regulator (Csr) system, we investigated this TCS in more detail. DNA sequence comparisons of the genes encoding the TCS BarA/UvrY from different *E. coli* isolates from the urinary tract or feces provided evidence that *uvrY* is subjected to positive selection during growth in the urinary tract. Afterwards the function of selected *uvrY* allelic variants containing non-synonymous SNPs was investigated in an isogenic strain background. Besides other suitable phenotypic assays, we focused on the potential growth advantages for strains harbouring the *uvrY* allelic variants in urine. The studies revealed that mutations in *uvrY* result in a growth advantage in urine which is most likely due to a more efficient utilization of the specific composition of available nutrients in urine. Therefore, we propose that a loss-of-function of the TCS BarA/UvrY is resulting in a selective advantage for cells harboring such a mutation, which explains the detected positive selection pressure at least for *uvrY*. Thus our results are in agreement with the hypothesis that genome reduction is an important strategy of efficient adaptation to the host environment, in this case the urinary tract.

Overall, the findings contribute to our understanding of adaptational mechanisms that are applied by ABU *E. coli* strains to successfully colonize the urinary tract.

Zusammenfassung

Harnwegsinfektionen (HWI) sind eine der häufigsten bakteriellen Infektionskrankheiten, welche jährlich Millionen von Menschen betreffen und ein bedeutendes gesundheitliches Problem darstellen. Sie werden hauptsächlich durch Bakterien der Spezies *Escherichia coli* verursacht. Die häufigste Form von HWI stellt die asymptomatische Bakteriurie (ABU) dar, welche trotz einer hohen Bakterienzahl im Urin durch die Abwesenheit von klinischen Symptomen charakterisiert ist. Die molekularen Mechanismen, die es ABU-assoziierten *E. coli* Stämmen ermöglichen den menschlichen Harnweg zu kolonisieren, ohne eine signifikante Wirtsantwort hervorzurufen, sind bisher weitgehend unbekannt. Eine Anpassung der Virulenz durch Genomreduzierung scheint ein evolutionäres Prinzip zur Etablierung eines kommensalen Lebensstils darzustellen.

In der vorliegenden Arbeit wurden evolutionäre Strategien untersucht, die *E. coli* Stämmen eine erfolgreiche Kolonisierung der Harnwege ermöglichen. Ein besonderer Schwerpunkt stellte hierbei die Untersuchung metabolischer Einschränkungen in Urin dar. Daher wurde das extrazelluläre Metabolom von drei eng verwandten *E. coli* Stämmen nach Anzucht in Urin bestimmt. Die Untersuchungen ergaben, dass *E. coli* in der Lage ist eine Vielzahl an Nährstoffen für die Herstellung von Vorläufer-Metaboliten, Reduktionsäquivalenten und Energie für weitere metabolische Prozesse zu verwenden.

Weitere Studien dienten der Ermittlung von nährstoffbedingten Einschränkungen in Urin. Hierbei wurde Kohlenstoff als limitierender Faktor für Wachstum in Urin identifiziert. Kohlenstoff stellt eine der wichtigsten Komponenten in lebenden Organismen dar und ist daher von essentieller Bedeutung. Frühere Studien haben bereits gezeigt, dass das Zweikomponentensystem (TCS) BarA/UvrY in ABU-assoziierten *E. coli* Stämmen bei längerem Wachstum in der Harnblase häufig mutiert ist. Da dieses Zweikomponentensytem eine wichtige Rolle in der Regulation des Kohlenstoff-Stoffwechsels über das Kohlenstoff-Speicher-Regulator (Csr)-System spielt, wurde dieses im Detail untersucht. Vergleiche der TCS BarA/UvrY-kodierenden DNA-Sequenzen von verschiedenen *E. coli* Isolaten aus Harnwegs- oder Fäkal-Proben ergaben, dass *uvrY* einer positiven Selektion während des Wachstums in den Harnwegen unterliegt. Anschließend wurde die Funktion ausgewählter *uvrY* Allelvarianten, welche nicht-synonyme SNPs aufwiesen, in einem isogenen Stammhintergrund untersucht. Neben geeigneten phänotypischen Tests konzentrierten wir uns auf die potenziellen Wachstumsvorteile in Urin für Stämme mit *uvrY* Allelvarianten. Die Untersuchungen ergaben, dass Mutationen in *uvrY* zu einem Wachstumsvorteil in Urin führten, welcher höchstwahrscheinlich auf eine effizientere Verwertung von verfügbaren Nährstoffen in Urin zurückzuführen ist. Aufgrund dessen vermuten wir, dass ein Funktionsverlust des TCS BarA/UvrY zu einem selektiven Vorteil für Zellen führt, die derartige Mutationen aufweisen, was auch die positive Selektion, zumindest für *uvrY,* erklärt. Somit stimmen

unsere Ergebnisse mit der Hypothese überein, dass Genomreduzierung eine wichtige Strategie für eine effiziente Anpassung an die Wirtsumgebung, in diesem Fall, den Harnwegstrakt, darstellt.
Gesamt gesehen tragen die Forschungsergebnisse zu einem besseren Verständnis von Anpassungsmechanismen bei, die von ABU *E. coli* Stämmen angewandt werden, um den Harnwegstrakt erfolgreich zu kolonisieren.

1 Introduction

Urinary tract infection (UTI) is one of the most common bacterial infectious diseases worldwide and represents a major health problem with high economic and medical burden (Foxman, 2002, 2010). It is estimated that nearly one third of all women will have at least one episode of UTI requiring antimicrobial therapy by the age of 24 years and 50 % of all women will experience one UTI during their lifetime (Barnett & Stephens, 1997; Foxman, 2002; Kunin, 1994). Emerging antibiotic resistances and the fact that UTI is still a major cause of death and increased morbidity in hospitalized patients even strengthen the urgent need for alternative UTI therapeutic approaches (Roos *et al.*, 2006; Sundén *et al.*, 2006; WHO, 2002). *E. coli* strains causing asymptomatic bacteriuria (ABU) represent a promising prophylactic agent in the post-antibiotic era.

1.1 *Escherichia coli* – a bacterial species of clinical relevance

The Gram-negative γ-proteobacterium *Escherichia coli* (*E. coli*) has initially been described by Theodor Escherich in 1886 (Escherich, 1886). Nowadays *E. coli* is one of the most widely studied prokaryotic model organisms in research and an important species in the field of biotechnology. *E. coli* is one of the first colonizers of the gastrointestinal tract of humans and animals after birth and usually inhabits the lower intestine as an innocuous commensal coexisting with the host (Johnson, 1991; Kaper *et al.*, 2004; Long & Swenson, 1977). Its beneficial symbiotic relationship with its host is based on the fact that *E. coli* promotes the stability of the luminal microbial flora, maintains the normal intestinal homeostasis and prevents the invasion of potential pathogenic bacteria (Kamada *et al.*, 2013; Yan & Polk, 2004). The proportion of *E. coli* is with 0.1 % -0.5 % of the total gut flora relatively low compared to other major commensal bacteria (Heizmann *et al.*, 1999; Katouli, 2010). Like most of the members of the family *Enterobacteriaceae*, *E. coli* is a facultative anaerobic, peritrichous flagellated, rod-shaped bacterium with a size of about 1.1-1.5 µm x 2.0-6.0 µm (Hahn *et al.*, 2013). The species *E. coli* is very versatile with respect to genotypic and phenotypic characteristics. A discrimination of the various *E. coli* types can be performed based on the presence of certain surface antigens, the O (lipopolysaccharides, LPS) and H (flagellar) antigen as well as the K antigen (capsule), which define the serotypes (Kaper *et al.*, 2004; Kauffmann, 1965).

Although *E. coli* normally exists in a mutual relationship with its host, different pathogenic variants evolved that acquired specific virulence factors enabling them to adapt to new niches. Besides being able to colonize different sites of the body these pathogenic variants can cause a broad spectrum of diseases that are divided according to their clinical symptoms in enteric/diarrhoeal diseases, UTIs and sepsis/meningitis. Up to now several pathotypes implicated in these infections have been described which are broadly classified in two categories: intestinal and extraintestinal *E. coli* (Table 1). Among

the intestinal pathogenic *E. coli* (IPEC), also designated as diarrheagenic *E. coli*, six well-characterized pathotypes are described: enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and the diffusely adherent *E. coli* (DAEC) (Croxen & Finlay, 2010; Johnson & Russo, 2002; Kaper *et al.*, 2004). *E. coli* isolates from infections outside of the intestinal tract are referred to as extraintestinal pathogenic *E. coli* (ExPEC). The most common extraintestinal infections are UTIs caused by uropathogenic *E. coli* strains (UPEC). Further ExPEC subgroups are the neonatal meningitis-associated *E. coli* (MNEC), sepsis-associated *E. coli* (SEPEC) and avian pathogenic *E. coli* (APEC) that cause systemic infection in poultry (Croxen & Finlay, 2010; Dho-Moulin & Fairbrother, 1999; Kaper *et al.*, 2004). The IPEC strains are considered to be obligate pathogens whereas ExPEC strains are classified as facultative pathogenic *E. coli* (Köhler & Dobrindt, 2011).

Table 1: *E. coli* pathotypes.

Intestinal Pathogenic *E. coli* (IPEC)	Extraintestinal Pathogenic *E. coli* (ExPEC)
Enteropathogenic *E. coli* (EPEC)	Uropathogenic *E. coli* (UPEC)
Enterohaemorrhagic *E. coli* (EHEC)	Meningitis-associated *E. coli* (NMEC)
Enterotoxigenic *E. coli* (ETEC)	Sepsis-associated *E. coli* (SEPEC)
Enteroaggregative *E. coli* (EAEC)	Avian pathogenic *E. coli* (APEC)*
Enteroinvasive *E. coli* (EIEC)	
Diffusely adherent *E. coli* (DAEC)	

*animal pathotype, causes extraintestinal infections in poultry.

Phylogenetic analyses revealed that the *E. coli* population according to the ECOR (*E. coli* group of reference strains) classification can mainly be categorized into the phylogenetic groups A, B1, B2 and D whereby further groups (C-I to C-V, E, F) and cryptic clades exist (Clermont *et al.*, 2011, 2000, 2013; Herzer *et al.*, 1990; Ochman & Selander, 1984; Tenaillon *et al.*, 2010a). Group A and B1 typically harbor commensal *E. coli* strains whereas groups B2 and to a lesser extent group D include the majority of ExPEC strains. IPEC strains are mainly assigned to group A, B1 and D (Boyd & Hartl, 1998; Johnson & Russo, 2002, 2005; Tenaillon *et al.*, 2010b).

1.2 Urinary tract infections

The urinary tract consists of the urethra, the bladder, the kidneys and the ureters that connect the bladder with the kidneys (Figure 1). Due to proper host defense mechanisms the urinary tract is usually sterile (Ali *et al.*, 2009). However, bacteria prevalently colonize the urinary tract resulting in a UTI. UTI is defined as a mucosal infection of the urinary tract in the presence of significant numbers of pathogenic bacteria in the urinary system (>10^5 colony forming units/ml urine). UTIs are mainly caused by pathogenic bacteria, mostly *E. coli*, or in rare cases, by viruses or parasites (Gasser, 2011;

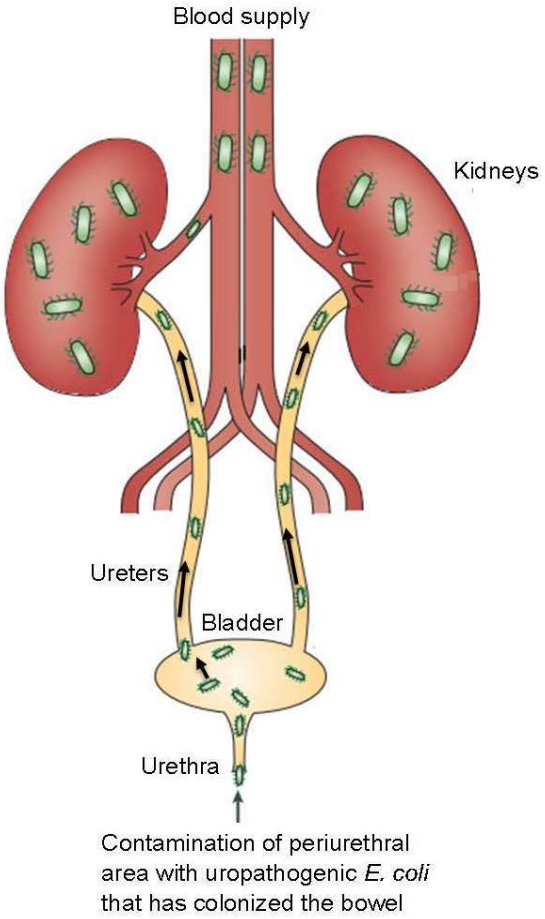


Figure 1: Pathogenesis of urinary tract infection caused by uropathogenic *E. coli.*

Different stages of urinary tract infection from the bladder to the kidneys and the bloodstream (modified according to Kaper *et al.* (2004)).

Hautmann & Geschwend, 2014). A UTI can occur in any part of the urinary tract, the lower (urethra, bladder) or the upper (ureters, kidneys) part and almost always progresses as a so-called ascending infection. Thereby pathogens pass through the periurethral area and ascend the urethra to the bladder or even disseminate to the kidney if the infection is left untreated (Kaper *et al.*, 2004, Figure 1). Only very rarely microorganisms enter the kidneys from the bloodstream, which is depicted as hematogenic or bloodborne infection (Gasser, 2011). According to their site of infection UTIs are classified into different disease categories: urethritis (urethra), cystis (bladder) and pyelonephritis (kidney). All of these disease categories are accompanied by a broad spectrum of symptoms (symptomatic infection) like dysuria with a burning pain during urination, bloody or cloudy urine, pain of the lower or upper abdomen and even fever or chills (Johnson, 1991). Acute pyelonephritis, a severe infection, is the most serious form of UTI with the possibility of causing irreversible kidney damage leading to renal failure, dissemination to the blood and even death (Katouli, 2010; Roos *et al.*, 2006).

From the medical perspective UTIs are divided into uncomplicated and complicated infections. Complicated infections are usually restricted to people with special risk factors for a complicated progress of the infection, e.g. individuals that are diagnosed with genitourinary tracts that have anatomical or functional abnormalities. In contrast to that, UTIs in patients with a structurally normal genitourinary tract and no prior instrumentation are considered to be uncomplicated (Foxman, 2002).

The risk to suffer from UTI is much higher in women than in men which is based on the anatomical constitution in women having a shorter distance from the urethra to the bladder or the bowel (Foxman, 2010; Lee & Neild, 2007). It is estimated that nearly 33 % of women will have had at least one episode of UTI requiring antimicrobial therapy by the age of 24 years and 50 % of all women will experience one UTI during their lifetime (Barnett & Stephens, 1997; Foxman, 2002; Kunin, 1994). Additionally, the recurrence rate was reported to be very high (Hooton, 2001). Besides women there are also other risk groups for UTI, for example infants, elderly, pregnant women, patients with diabetes or with urologic abnormalities or patients with spinal cord injuries and/or catheters (Foxman, 2002). Besides these anatomical and sex-dependent reasons for an increased risk to suffer from UTI the susceptibility is further influenced by factors like the vaginal flora, age, diet or sexual intercourse (Foxman & Frerichs, 1985; Stamey & Sexton, 1975).

In relative numbers UTI is the most common bacterial infectious disease accounting for roughly 25 % of all infections in industrialized countries. Moreover, UTI followed by catheterization is considered to be the most frequent nosocomial infection, occasionally leading to bacteraemia and death in hospitalized patients (Foxman, 2002; WHO, 2002).

The high incidence of UTI represents an enormous clinical and economical burden with a high impact on public health. The estimated health-care costs of UTI are significant, at about $ 2.3 billion per year (Foxman, 2002, 2010).

1.3 Uropathogenic *E. coli* – the main causative of UTI

The main causative agents for community and hospital acquired UTI are uropathogenic *E. coli* (UPEC). They are responsible for more than 80 % of all UTIs (Croxen & Finlay, 2010). Although *E. coli* is the predominant infecting agent of UTI a number of other bacteria are also found to cause UTI including organisms like *Klebsiella* spp., *Enterococcus* spp., *Streptococcus* spp., *Staphylococcus* spp. or *Proteus* spp. and *Pseudomonas* spp. (Hooton, 1999; Ronald, 2003; White, 1987).

UPEC can be distinguished from commensal *E. coli* by the presence of additional genetic material encoding a number of different virulence factors (VFs). These VFs provide UPEC with an enhanced ability to successfully persist in extra-intestinal sites and to colonize the urinary tract to cause

infection. VFs are often encoded by flexible genetic elements, such as the so-called pathogenicity islands (PAIs) representing large genomic regions (10-200 kb). PAIs usually contain multiple virulence genes and are characterized by a different GC content than the rest of the chromosome (Emody *et al.*, 2003; Hacker & Kaper, 2000). Moreover PAIs are often flanked by mobile genetic elements like bacteriophages, insertion sequences or transposons and are inserted near to tRNA genes suggesting that PAIs have evolved from mobile genetic elements as a result of horizontal gene transfer (Dobrindt *et al.*, 2002).

1.3.1 Virulence factors

Most virulence factors can be grouped as adhesins, flagella, extracellular polysaccharides, toxins or iron uptake systems. VFs can be subdivided in VFs that are (i) located on the surface of the cell and those which are (ii) secreted and exported to the site of action (Emody *et al.*, 2003). An overview of the best characterized VFs occurring in UPEC is given in Figure 2. The detailed functions of the VFs are listed in Table 2.

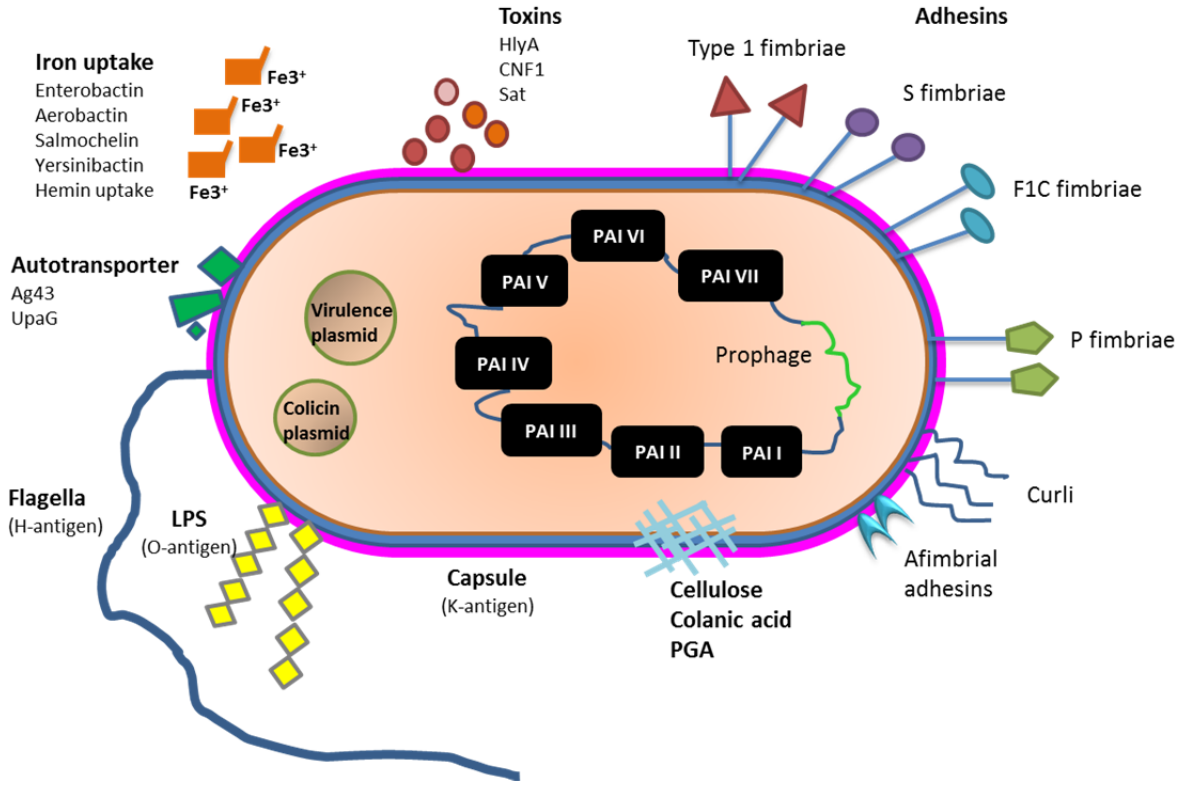


Figure 2: Virulence factors of UPEC.

Different VFs that facilitate pathogenesis of UTI are shown (LPS: Lipopolysaccharid; HlyA:α-hemolysin; CNF-1: cytotoxic necrotizing factor 1; Sat: secreted autotransporter toxin, PGA: poly-β-1-6-N-glucosamine; Ag43: antigen 43) (modified according to Oelschlaeger & Fünfstück, 2006).

Table 2: Virulence factors of uropathogenic *E. coli* and their functions.

Summarized according to Bien *et al.* (2012); Croxen & Finlay (2010); Donnenberg (2002); Johnson (1991); Kaper *et al.* (2004).

Virulence Factor	Type (coding gene)	Function
Adhesins	Type 1 fimbriae (*fim*)	➢ Binding to mannosylated glycoproteins of the hosts´uroepithelium ➢ Invasion ➢ Biofilm formation
	P fimbriae (*pap*)	➢ Adhesion to glycosphingolipids carrying a Gal-α-(1,4)-Gal moiety on renal epithelial cells ➢ Induction of inflammatory response of the host
	F1C fimbriae (*foc*)	➢ Adhesion to uroepithelial/renal cells by recognition of galactosyl/globotriaosyl ceramides
	S fimbriae (*sfa*)	➢ Adhesion to uroepithelial cells by recognition of sialyl galactosides ➢ Hemagglutination of erythrocytes
	Afa/ Dr fimbriae (*afa*/*dra*)	➢ Adhesion to epithelial cells, polymorphonuclear leukocytes (PMNs), type IV collagen, decay-accelerating factor (DAF) and carcinoembryonic antigen related adhesion molecules (CEACAM)
	Curli (*csg*)	➢ Binding to proteins of extracellular matrix like fibronection, laminin or various types of collagen
Flagella	Flagella (*fli*, *flg*)	➢ Mediate motility and chemotaxis ➢ Support bacterial dissemination in the host ➢ Invasion into renal epithelial cells ➢ Induction of interleukin-8 (IL-8) release
Extracellular poly-saccharides	Lipopolysaccharide (LPS) (*rfb*)	➢ Serum resistance ➢ Induction of inflammatory response (IL-1, IL-6, tumor necrosis factor (TNF) α)
	Capsule (*kps*, *neu*)	➢ Serum resistance ➢ Molecular mimicry
	Colanic acid (*wca*) Cellulose (*bcs*) PGA -> poly-ß-1-,6-N-glucosamine (*pga*)	➢ Components of the extracellular matrix ➢ Biofilm formation ➢ Adhesion
Toxins	α-hemolysin (*hlyA*)	➢ Lysis of erythrocytes ➢ Induction of apoptosis and exfoliation of bladder cells ➢ Induction of inflammatory response
	CNF-1 (*cnf1*)	➢ Induces actin cytoskeleton reorganisation ➢ Apoptosis ➢ Facilitates bacterial invasion
	Autotransporter toxin Sat (*sat*)	➢ Cytopathic effect on bladder/kidney cells ➢ Tissue damage ➢ Activation of immune response
Iron acquisition systems	Enterobactin (*ent*/ *feb*) Salmochelin (*i*) Aerobactin (*iuc*/ *iut*) Yersiniabactin (*irp*/ *fyu*) Haeme-receptor (*chu*)	➢ Iron uptake ➢ Invasion and internalisation into uroepithelial cells ➢ Adhesion

Virulence Factor	Type (coding gene)	Function
Autotrans-porter	Antigen 43 (AG43) (*agn*43)	➢ Diffuse adhesion ➢ Autoaggregation ➢ Biofilm formation
	Uropathogenic autotransporter UpaG (*upaG*)	➢ Adhesion to bladder epithelial cells and extracellular matrix ➢ Biofilm formation ➢ Cell aggregation

The above listed VFs present in UPEC strains facilitate the infection of different sites of the urinary tract and at least partially contribute to the complex pathogenesis of UTI (Donnenberg, 2002). Although a lot of UPEC strains carrying the described VFs can cause symptomatic infection there is another type of UTI, the asymptomatic infection, which is described to be the most common form of UTI.

1.4 Asymptomatic bacteriuria (ABU)

Asymptomatic bacteriuria (ABU), the most common type of UTI, is defined by the presence of a significant number of bacteria in urine with the absence of clinical symptoms and pathology in the patient. The patients may carry ABU-causing strains for months or even years without developing significant symptoms (Lindberg *et al.*, 1978; Lindberg & Winberg, 1976) although a low mucosal host immune response resulting in increased urine leukocytes and cytokine levels can often be observed (Agace *et al.*, 1993; Kass, 1956). This bacterial carriage in the urinary bladder resembles commensalism at other mucosal sites. The quantitative criteria for ABU are defined as > 10^5 CFUs/ml in clear voided mid-stream urine or > 10^2 CFUs/ml in urine of catheterized people (Kass, 1956; Nicolle *et al.*, 2005; Warren *et al.*, 1982). In medical terms an ABU is diagnosed when these laboratory criteria are met for the same bacterial species in two subsequent urine cultures of patients without typical symptoms of an infection.

ABU is mostly caused by a clonal bacterial population resulting in a monoculture rather than a complex microflora (Roos *et al.*, 2006; Zdziarski *et al.*, 2008). With respect to ABU, *E. coli* is the most common organism isolated from ABU patients. Other bacterial species that are found in ABU samples include other members of the *Enterobacteriaceae*, *Pseudomonas aeruginosa*, *Enterococcus* spp., and group B streptococci. The infecting organisms can be correlated with the patient's environment: *E. coli* is more likely to be found in healthy persons whereas a multi-drug resistant polymicrobial flora is frequently associated with nursing home residents. *Enterococcus* species and Gram-negative bacilli are more often found in men (Warren *et al.*, 1982).

The exact mechanisms that allow ABU-associated *E. coli* strains to cause a persistent infection in the bladder without provoking a host response are not yet fully understood. However, a lot of studies focused on the investigation of ABU strains, especially of *E. coli* 83972, to investigate the differences between symptomatic and asymptomatic UTI in terms of pathogenesis and genetic background of the colonizing bacteria.

1.4.1 *E. coli* 83972

E. coli strain 83972 is a well-characterized prototype ABU strain originally isolated from a Swedish schoolgirl who carried the strain for at least three years without renal deterioration or clinical symptoms of UTI (Lindberg *et al.*, 1975). Phylogenetically strain 83972 is assigned to the ECOR group B2 and the clonal sequence type ST73, which also includes virulent UPEC strains like CFT073, isolated from the bloodstream of a women with acute pyelonephritis (Mobley *et al.*, 1990), and the non-pathogenic fecal isolate Nissle 1917 (Zdziarski *et al.*, 2008). Comparative genomic approaches confirmed a close relatedness of *E. coli* 83972 to UPEC CFT073 and strain Nissle 1917, all of them strains with different environmental origins and dispositions (Hancock *et al.*, 2010; Vejborg *et al.*, 2010). Despite of the genetic relationship to virulent UPEC strains *E. coli* 83972 does not cause symptomatic infections. This is due to the genetic characteristics of *E. coli* 83972. *E. coli* 83972 carries a lot of classical virulence-associated genes but fails to express them (Dobrindt *et al.*, 2003; Hull *et al.*, 1999). The loss of functional UPEC VFs like type 1, F1C, and P fimbriae was shown to be due to multiple point mutations and deletions indicative of genome reduction as a mechanism to achieve long-term colonization in the host (Klemm *et al.*, 2006; Zdziarski *et al.*, 2008). This kind of virulence attenuation is not particular for *E. coli* 83972. Molecular epidemiology revealed that > 60 % of ABU strains carry virulence genes, even though they lack virulence-associated phenotypes (Dobrindt *et al.*, 2003; Roos *et al.*, 2006; Salvador *et al.*, 2012).

Adaptation strategies like alterations in essential virulence genes including single base mutations, deletions or DNA rearrangements might be essential for mucosal pathogens to achieve long-term persistence and to adapt to the host environment, consistent with a reduced activation of the immune response (Dobrindt *et al.*, 2010).

1.5 Genome plasticity and bacterial adaptation

As stated by Bennett (2004) *"The genetic content and the ability to control the expression of distinct genes, both temporally and in response to the environment, determines whether an organism can survive changing conditions and can compete for the resources it needs to reproduce"*. *E. coli* as a very versatile species is known to exhibit a very dynamic genome structure. The genome of *E. coli,*

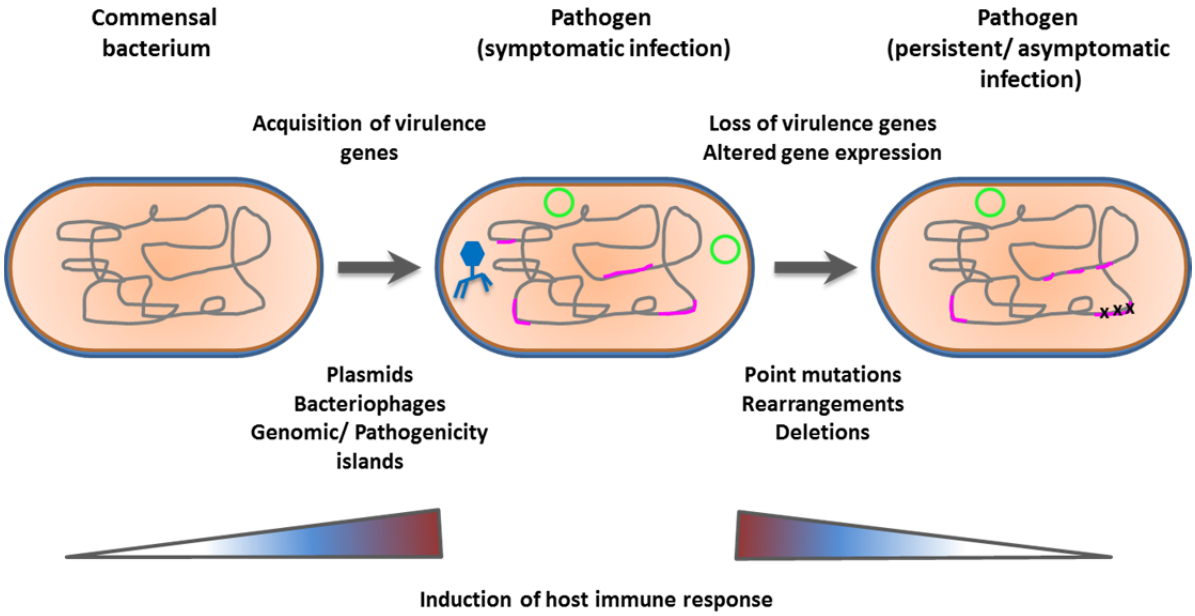


Figure 3: Impact of genome plasticity on adaptation of pathogenic bacteria.

Different adaption strategies to acquire or lose genes can alter the characteristics of a bacterium developing from a commensal to a pathogenic strain or vice versa (modified according to (Dobrindt *et al.*, 2010)).

which roughly contains 4.6 -5.5 x 10^6 bp, can be divided into (i) the core genome which is present in all *E. coli* strains, containing essential housekeeping genes and defining basic metabolic functions, and into (ii) the so-called flexible/dispensable genome which comprises strain-specific genes contributing to diverse phenotypes and conferring selective advantages for the adaptation to specific environmental conditions (Medini *et al.*, 2005). Genome flexibility, also depicted as genome plasticity, is achieved by the ability of bacteria to acquire or lose genetic material from other microorganisms to adapt their genomes to challenging environmental conditions (Figure 3). Therefore genome plasticity is a very important tool for the adaptation to new ecological niches or host conditions (Dobrindt *et al.*, 2010).

The flexible gene pool comprises mobile elements, like plasmids, bacteriophages, transposons, insertion sequence (IS) elements or genomic islands (GEIs) (Dobrindt *et al.*, 2004). The most important mechanisms that enhance genome flexibility are the horizontal gene transfer (HGT), homologous recombination events or genome reduction resulting in a genomic rearrangement. Besides these characteristics intra-strain mechanisms like point mutations or deletions contribute to the process of adaptation (Leimbach *et al.*, 2013).

In the context of long-term adaptation of *E. coli* to growth and survival in the urinary tract these processes play a major role. Geno- and phenotypic analysis of *E. coli* 83972 revealed that this strain may have arisen from a virulent ancestral variant by genome reduction (Vejborg *et al.*, 2010). This assumption is in accordance with the observed expression of non-functional VFs due to different mutations resulting in a reduced activation of the immune response (Klemm *et al.*, 2007, 2006).

These genomic alterations apparently provided a colonization advantage for *E. coli* 83972. This ABU strain thus obviously used genome flexibility as a tool to evolve towards commensalism and thereby becoming a successful colonizer of the human urinary tract.

Besides the direct regulation of virulence-associated genes like type 1, F1C and P fimbriae in strain 83972, which play a major role in the outcome of an infection, the modulation of regulatory systems that are involved in urovirulence is important for an efficient adaptation to challenging environmental conditions. One major regulator system of *E. coli* is the two-component system (TCS) BarA/UvrY. Re-sequencing studies of *E. coli* 83972 after therapeutic bladder colonization showed that components of the BarA/UvrY TCS are frequently mutated upon prolonged growth in the bladder. This observation suggests that the BarA/UvrY TCS is a possible mutational hotspot under these conditions, allowing for successful bacterial colonization of the urinary tract (Zdziarski *et al.*, 2010).

1.6 Two-component systems

Bacteria are often challenged by drastic environmental changes like changes in nutrient availability or diverse stress conditions in their natural environment. In order to successfully compete and survive bacteria need to rapidly adapt to fluctuating environmental conditions by adjusting their gene expression patterns to environmental cues (Mitrophanov & Groisman, 2008). In order to sense and react to environmental changes bacteria developed TCSs. TCSs are key mediators of bacterial signal transduction enabling the microbes to quickly and adequately act in response to external stimuli (Gao & Stock, 2009; Hoch, 2000; Jung *et al.*, 2012; Stock *et al.*, 2000). A TCS functions as a sensory and response regulatory system implying sensing of various chemical/physical signals and regulation of gene product levels (Mitrophanov & Groisman, 2008). A prototypic, simple TCS consists of a membrane-associated histidine sensor kinase (HK) responsible for the external signal recognition and a cytoplasmic response regulator (RR), which converts the signal into gene activation or other cellular processes (Hoch, 2000) (Figure 4 A). The signal transduction is achieved via a phosphorelay including phosphorylation and dephosphorylation at histidine and aspartate residues of conserved protein domains (West & Stock, 2001).

The transmembrane sensor histidine kinase (HK), also called Histidine Protein Kinase (HPK), is an integral membrane protein and typically consists of two parts with different functions and structures: (i) a N-terminal periplasmic sensing region that is coupled to (ii) a C-terminal cytoplasmic kinase domain (Grebe & Stock, 1999) (Figure 4). The N-terminal sensing region is variable in sequence which can be explained by the huge diversity of signals that it has to sense from the environment. In contrast the C-terminal kinase core domain is highly conserved. The kinase core domain is also

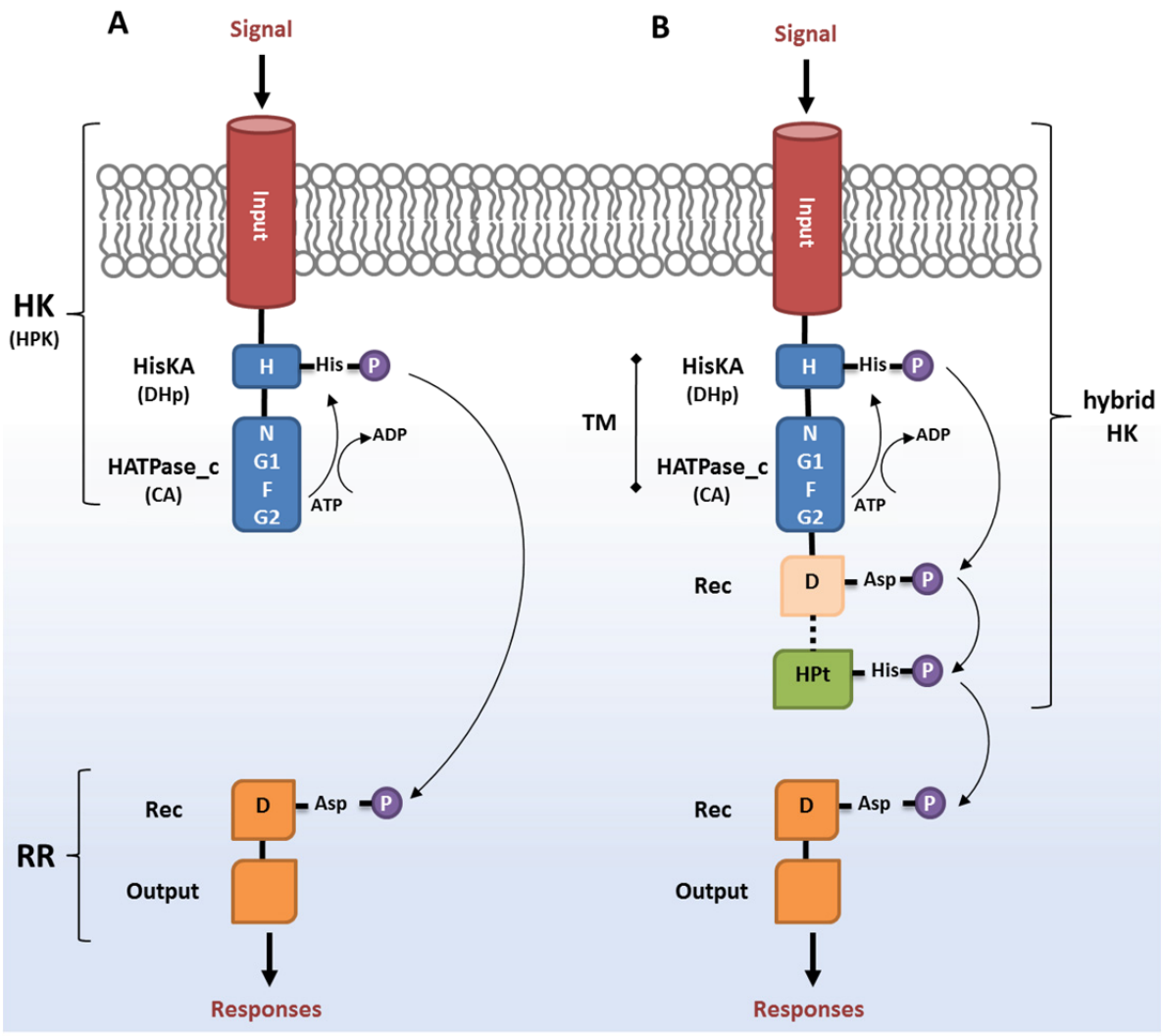


Figure 4: Domain organization and signaling in two-component systems (TCSs).

(A) The prototypical TCS consists of a dimeric transmembrane sensor histidine kinase (HK) and a cytoplasmic response regulator (RR). A phosphoryl residue is transferred between the highly conserved kinase core (HisKA and HATPase_c), representing the transmitter domain (TM), and the receiver domain (REC). The kinase core contains several homology boxes, H, N, G1, F and G2 which are conserved across HKs. **(B)** A phosphorelay scheme is utilized by hybrid HKs involving additional Rec and histidine phosphotransfer (HPt) domains for multiple phosphotransfer events. The intermediate HPt domain can either be an independent protein or linked to the HK (Gao & Stock, 2009).

designated as transmitter domain (TM). The TM harbors the so-called homology boxes H, N, G1, F and G2, which share unique sequence motifs and serve as conserved sequence fingerprints to define the protein family (Grebe & Stock, 1999; Jung *et al.*, 2012; West & Stock, 2001). In addition, this domain can be divided in two parts: the H-box containing the dimerization and histidine phospotransfer domain (DHp), also referred to as His kinase A (HisKA) domain, and the histidine kinase-type ATPase catalytic (HATPase_c) domain, also known as **c**atalytic and ATP-binding (CA) domain (Jung *et al.*, 2012).

Upon signal detection the HK undergoes ATP-dependent autophosphorylation at the histidine (His) residue within the DHp domain (Figure 4). Catalyzed by the RR the phosphoryl group is subsequently

transferred to an aspartate (Asp) side chain in the receiver (Rec) domain of the downstream, cognate RR itself. Upon phosphorylation the RR is in its active state.

RRs are typically the terminal component of the phosphotransfer pathways and provide the central, intracellular switch regulating the adaptive output responses, in most of the cases an alteration in gene expression (Robinson *et al.*, 2000; West & Stock, 2001). RRs usually have a two-domain structure consisting of a N-terminal regulatory domain, often called receiver (Rec) domain, linked to a variable C-terminal effector domain, also designated as output domain (Figure 4). The regulatory domain actively catalyzes the phosphotransfer from the HK to its own Asp residue. Besides this interaction with the HK the regulatory domain catalyzes autodephosphorylation to limit the active state and regulates the activity of the associated effector domain in a phosphorylation-dependent manner (West & Stock, 2001). The phosphorylation of the RR induces conformational changes enabling the RR to elicit the response, e.g. act as a transcriptional regulator either in activating or repressing target gene transcription (Sahu *et al.*, 2003).

TCSs in general exhibit a modular architecture of conserved components. Besides the simplest phosphotransfer involving a single HK and a single RR (Figure 4 A) more complex signal transduction pathways involving multiple phosphotransfer steps, designated as phosphorelay, are possible (Figure 4 B). These expanded His-Asp-His-Asp phosphorelays facilitate a greater versatility in signaling strategies and bear multiple points of regulation, thus enabling a better output fine-tuning than a prototypical TCSs (Jung *et al.*, 2012; West & Stock, 2001). The transmitter domain of the hybrid HKs is fused to another Rec domain which at least transfers the phosphoryl group to a **H**is-containing **p**hospho**t**ransfer (HPt) domain, which is either located within the hybrid HK or forms a separate protein. Subsequently the HPt domain transfers the phosphoryl group to the Rec domain of the corresponding RR. The HKs exhibiting this kind of domain organization are depicted as hybrid kinases, containing multiple phosphoacceptor and phosphodonor sites (Robinson *et al.*, 2000). It has been estimated that nearly 20 % of all HKs are encoded as hybrid proteins (Grebe & Stock, 1999).

Surveys of representative genomes revealed 62 two-component proteins for *E. coli*, all of them being involved in the regulation of different processes like chemotaxis, osmoregulation, metabolism and transport (Mizuno, 1997). Further different TCSs were reported to form networks of functional interactions, such as cross- or cascade-regulation of important cellular functions such as RpoS regulation or flagella synthesis. These aspects underline the important role of TCSs in cell physiology (Oshima *et al.*, 2002; Yamamoto *et al.*, 2005).

Although TCS are widespread in bacteria and archaea no phosphotransfer systems have been identified in mammals yet making them an interesting candidate for the development of new antimicrobial agents (Robinson *et al.*, 2000; Stock *et al.*, 2000).

1.6.1 TCS BarA/UvrY

One important TCS of *E. coli* with global regulatory function is the TCS BarA/UvrY (Baker *et al.*, 2002; Pernestig *et al.*, 2003; Romeo *et al.*, 1993). This TCS is a hybrid histidine kinase system that is transferring signals via phosphorelay (Figure 4 B). The HK BarA and the RR UvrY itself are encoded by genes that are located on distal chromosomal regions (Oshima *et al.*, 2002).

The HK BarA is encoded by the bacterial adaptive response gene, *barA*. BarA consists of 918 amino acids, belongs to the family of tripartite histidine kinases and exhibits the typical architecture of a hybrid histidine kinase (Figure 4 B) (Ishige *et al.*, 1994; Nagasawa *et al.*, 1992; Sahu *et al.*, 2003). Besides two transmembrane domains the structure includes a HAMP domain which was named this way based on its occurrence in **h**istidine kinases, **a**denyl cyclases, **m**ethyl binding proteins and **p**hosphatases (Figure 5 A). The HAMP domain is usually located at the C-terminal end of the last transmembrane (TM) segment and represents a linker region between the extra- and intracellular part of the HK. In addition, it plays an important role in the signal transmission from the TM regions to the kinase core (Hulko *et al.*, 2006). In combination with the HisKA and the HATPase_c domains this region is depicted as the primary transmitter domain relaying the phosphoryl group to the primary receiver domain and subsequently to the secondary transmitter domain containing the HPt domain, which mediates the final transfer of a phosphoryl group to its cognate RR, UvrY. Homologues of BarA were identified in other bacterial species, such as *Salmonella* (BarA), *Erwinia* (ExpS), *Vibrio* (VarS), *Legionella* (LetS), and *Pseudomonas* (GacS) species (Altier *et al.*, 2000; Chavez *et al.*, 2010; Corbell & Loper, 1995; Hammer *et al.*, 2002; Lenz *et al.*, 2005). BarA was found to be transcriptionally activated in UPEC after P-pili mediated adherence to host cells. In addition it was shown to be required for the activation of siderophore systems (Zhang & Normark, 1996). Moreover, BarA could activate the transcription of the sigma factor RpoS which is a conserved stress regulator playing a critical role in survival under stress conditions (Hengge-Aronis, 1999; Mukhopadhyay *et al.*, 2000). These findings suggest that BarA plays an important role in the regulation of genes that are important for the colonization and the survival during UTI or growth in urine.

Upon signal detection BarA activates the RR UvrY which was shown to be the cognate RR for BarA by Pernestig *et al.* (2001).

UvrY consists of a N-terminal receiver (Rec) domain harboring the Asp residue and a C-terminal helix-turn-helix (HTH) DNA-binding output domain which is characteristic for members of the LuxR/FixJ family of response regulators (Goodier & Ahmer, 2001) (Figure 5 B). This domain is also designated LuxR_c like domain (Figure 23). UvrY consists of 218 amino acids and is encoded by the *uvrY* gene. It is co-transcribed with the downstream *uvrC* gene, which encodes a DNA repair enzyme (Moolenaar *et al.*, 1987). Similar to BarA various homologues for UvrY have been identified in other bacterial species like SirA in *Salmonella*, ExpA in *Erwinia*, VarA in *Vibrio*, GacA in *Pseudomonas* (Goodier &

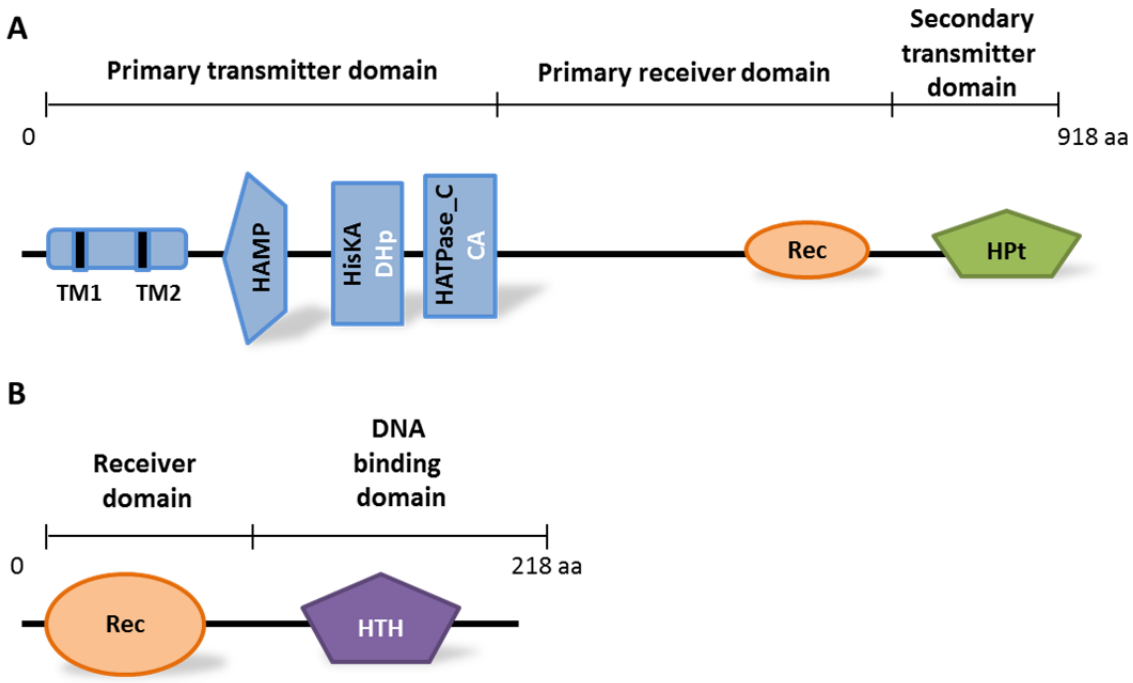


Figure 5: Domain organization of BarA and UvrY.

(A) The histidine kinase BarA with its two transmembrane domains (TM1/2) and the HAMP domain (**H**istidine kinase, **A**denyl cyclase, **M**ethyl binding proteins, **P**hosphatases) as well as the HisKA (histidine kinase A; also referred to as DHp) and the HATPase_c (histidine kinase-type ATPase catalytic, also referred to as CA) domain, followed by a Rec and a HPt (His-containing phosphotransfer) domain. **(B)** UvrY consists of two main parts, the Rec domain and the DNA binding domain containing a HTH (helix-turn-helix) motif.

Ahmer, 2001; Pernestig *et al.*, 2001). Together with the identified homologues for BarA they build the analogue TCSs to BarA/UvrY in other species

The induction of the BarA/UvrY TCS is dependent on several environmental conditions and physiological signals. Mondragón *et al.* (2006) could show that a pH lower than 5.5 does not allow the activation of this TCS while formate and acetate, both representing end products of glucose metabolism, as well as short-chain fatty acids provide a stimulatory signal for BarA/UvrY (Chavez *et al.*, 2010). The stimulation by end products of glucose metabolism makes sense in the light that BarA/UvrY is highly activated with transition into the stationary phase when glucose is depleted and the bacteria have to switch their metabolism (Mondragón *et al.*, 2006). However, the nature of physiological stimuli is still under investigation. Besides the activation of BarA/UvrY a lot of studies focused on the impact of this TCS on cellular or metabolic processes. BarA/UvrY in *E. coli* has a pleiotropic effect and has been connected with different metabolic processes and the regulation of different phenotypic traits including biofilm formation, oxidative stress response, and sigma S (RpoS) expression (Mitra *et al.*, 2013; Pernestig *et al.*, 2003; Suzuki *et al.*, 2002). It is suggested to play a major role in the establishment of early infections by promoting biofilm formation, persistence and virulence of UPEC (Mitra *et al.*, 2013; Palaniyandi *et al.*, 2012; Tomenius *et al.*, 2006). Moreover BarA/UvrY as well as their homologues show a clear connection to genes involved in bacterial

virulence (Altier *et al.*, 2000; Eriksson *et al.*, 1998; Teplitski *et al.*, 2003; Tomenius *et al.*, 2006). Interestingly mutations in the BarA/UvrY system led to attenuated virulence *in vivo* and *in vitro* as shown for *E. coli* CFT073 and an APEC strain suggesting an adaptational mechanism to the host environment (Herren *et al.*, 2006; Palaniyandi *et al.*, 2012).

The regulation of the above mentioned features is often done in concert with the so-called carbon storage regulator (Csr) system whose expression is positively controlled by BarA/UvrY (Suzuki *et al.*, 2002). One major impact of the Csr system is the regulation of the carbon metabolism in *E. coli* (Baker *et al.*, 2002).

1.6.2 Role of the carbon storage regulator system

The carbon storage regulator (Csr) system represents a global posttranscriptional regulatory network that controls various global responses in bacterial metabolism and physiology, e.g. affects stationary-phase gene expression and especially the central carbohydrate metabolism of *E. coli* on a wide scale. Further profound effects of this system on multicellular behavior of *E. coli* were shown (Romeo, 1998; Suzuki *et al.*, 2002). The system consists of the non-coding RNAs CsrB and CsrC, the protein CsrD as well as the RNA binding protein CsrA (Figure 6 A). The signaling circuitry that connects BarA/UvrY with the Csr system was shown by Suzuki *et al.* (2002). As schematically shown in Figure 6 A, UvrY functions as a transcriptional regulator and directly activates the gene expression of CsrB and CsrC whose decay is facilitated by binding of the protein CsrD that converts the non-coding RNAs (ncRNAs) into a substrate for RNase E degradation (Suzuki *et al.*, 2006). CsrB and CsrC, both RNA molecules with a short lifetime of about 1.5 - 4 minutes, contain multiple CsrA binding sites and form a globular ribonucleoprotein complex with approximately 18 CsrA polypeptides for CsrB or rather nine for CsrC. Consequently they function by sequestering CsrA and thereby antagonize its effects on gene expression (Babitzke & Romeo, 2007; Liu *et al.*, 1997; Romeo, 1998; Weilbacher *et al.*, 2003). CsrA, a homodimeric RNA binding protein of 61 amino acid length, facilitates the decay of specific mRNA transcripts by binding to the ribosome binding site of target transcripts and thereby blocking the ribosome access to the mRNA followed by subsequent mRNA degradation (Baker *et al.*, 2002). Interestingly, CsrA directly modulates the synthesis of CsrB and CsrC trough UvrY, implicating an autoregulatory mechanism of CsrA to control its own activity (Gudapaty *et al.*, 2001).

CsrA was shown to control a broad range of phenotypes as well as many metabolic functions in *E. coli* (Figure 6 B). It represses gluconeogenesis, glycogen biosynthesis and catabolism, the synthesis of the second messenger cyclic diguanylate monophosphate (c-di-GMP) and biofilm formation. In contrast it positively regulates glycolysis, motility, acetate metabolism, virulence, oxidative stress response, quorum sensing and pathogenesis (Jackson *et al.*, 2002; Liu *et al.*, 1995; Romeo *et al.*, 1993; Sabnis *et al.*, 1995; Timmermans & Van Melderen, 2010; Wei *et al.*, 2001).

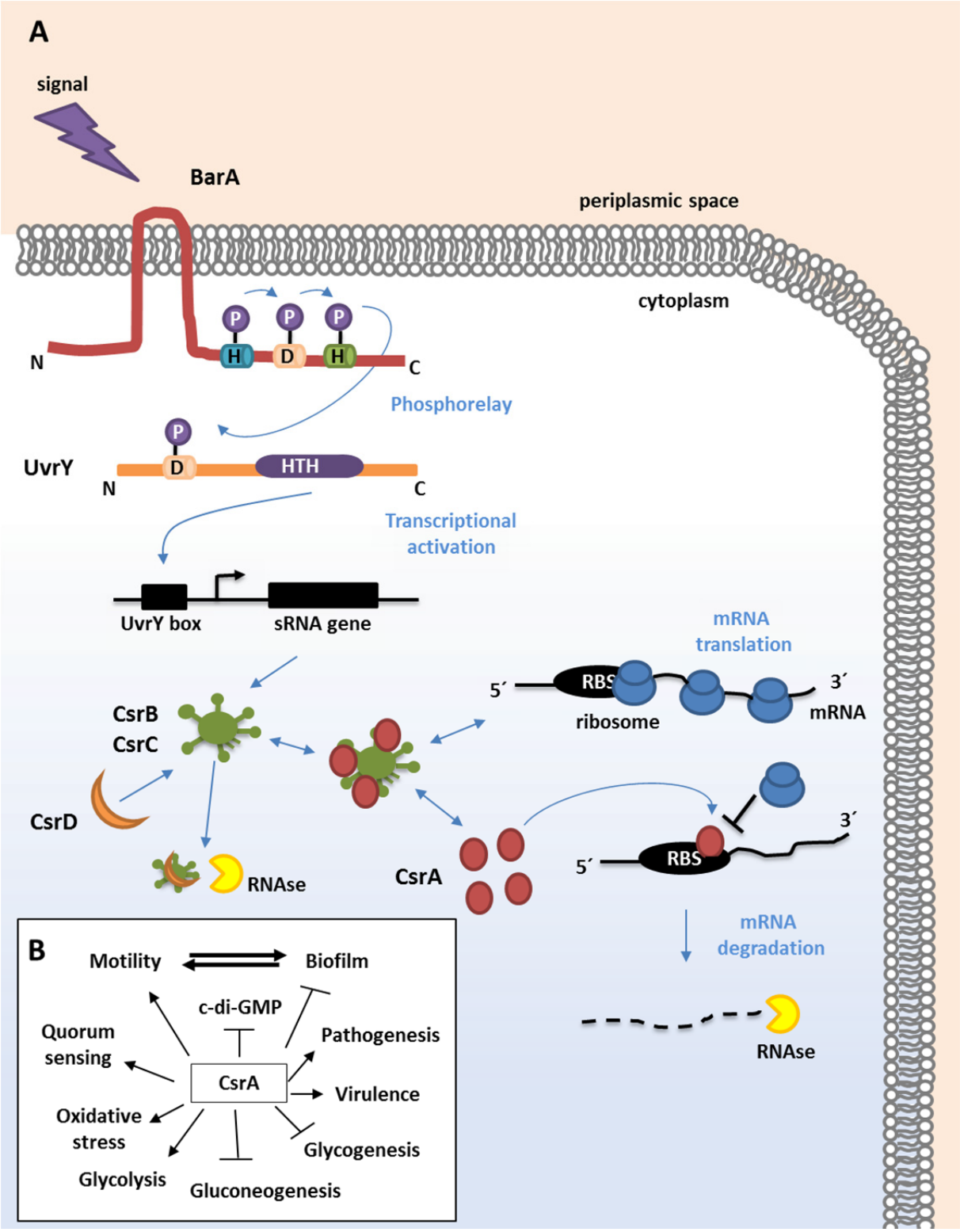


Figure 6: Regulatory interactions between the BarA/UvrY TCS and the Csr system *in E. coli.*

(A) Upon signal transduction BarA is autophosphorylated and activates UvrY. UvrY activates transcription of the ncRNAs CsrB and CsrC which sequester the RNA binding protein CsrA. CsrA itself binds near to the ribosomal binding site (RBS) of target mRNAs thereby inhibiting ribosomal binding resulting in mRNA degradation. Decay of CsrB and CsrC is mediated by binding of the protein CsrD. **(B)** CsrA as the global component of the Csr system activates/represses a multitude of phenotypic traits as well as metabolic pathways. Symbols indicate positive (→) or negative (--I) regulation (modified according to Lapouge *et al.* (2008), Timmermans & Van Melderen (2010) and Heroven *et al.* (2012)).

Taken together the Csr and BarA/UvrY systems affect a lot of biological and metabolic pathways in *E. coli* and control numerous functions which are important during the interaction with the host. These facts even strengthen the relevance of Csr and BarA/UvrY in the context of bacterial survival in the host under challenging environmental conditions.

To provide a better understanding of the affected metabolic pathways, a short overview of the most important carbon metabolic pathways will be given in the following chapter.

1.7 Metabolism

In order to sustain life, to grow and to reproduce all living organisms require certain nutrients which provide necessary chemical elements for the *de novo* synthesis of cellular components and the production of biologically utilizable energy (Gottschalk, 1986). The metabolism of an organism can be divided into catabolism, the breakdown of chemical compounds for the release of energy, and anabolism, the synthesis of complex compounds requiring energy.
The central metabolism of *E. coli* can be divided into different pathways as follows.

1.7.1 Glycolysis

Glycolysis, also known as Embden-Meyerhof-Parnas (EMP) pathway, named after its discoverers, is one of the main metabolic pathways in carbon metabolism. It is widespread amongst bacteria and suggested to be one of the most ancient known metabolic pathways occurring either in aerobic or anaerobic organisms (Romano & Conway, 1996). During glycolysis glucose is converted to two molecules pyruvate via ten enzyme-catalyzed reactions (Figure 7 A). The net yield of this process accounts for two molecules adenosine triphosphate (ATP) and two molecules nicotinamide adenine dinucleotide (NADH). NADH represents an important coenzyme essential for redox reactions in different metabolic pathways. The glycolysis can be divided into two parts: (i) the preparatory or investment phase where energy is consumed to convert glucose into glyceraldehyde 3-phosphate and the (ii) pay-off phase, comprising the conversion to pyruvate.
Under aerobic conditions 50 % the available glucose is used for the production of ATP, the remaining 50 % are converted into cellular material (Gottschalk, 1986). A general scheme of the biosynthesis of cellular components from glucose illustrating its importance in the metabolism of *E. coli* is shown in Figure 7 B.

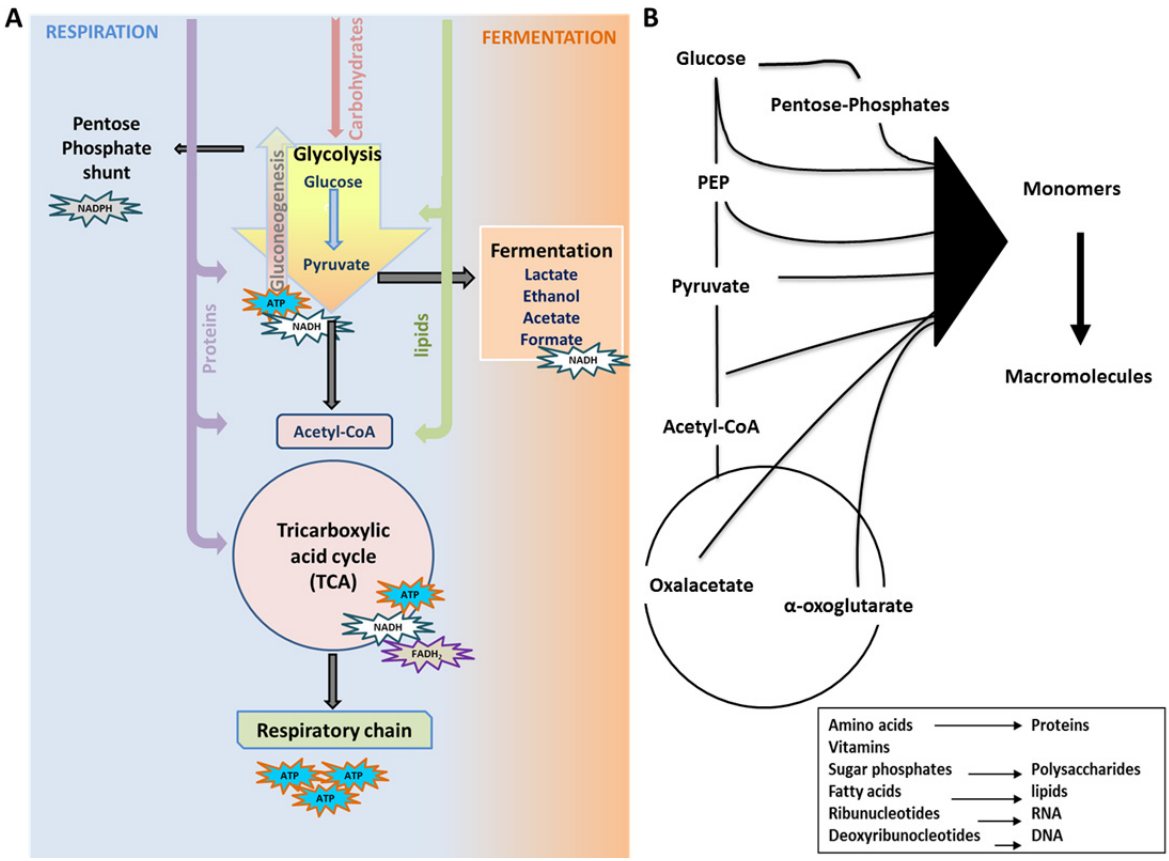


Figure 7: Metabolic and biosynthetic pathways in *E. coli*.

(A) Schematic overview of the central metabolism including the entry points of different chemical compounds like proteins, carbohydrates and lipids which are catabolized. The metabolism is divided into a respiratory and a fermentative part, dependent on the availability of oxygen (O_2). The main goal of the metabolism is the provision of cellular material for biosynthetic reactions and the provision of energy. **(B)** A general scheme showing the biosynthesis of different macromolecules from glucose as a basic substrate.

1.7.2 Gluconeogenesis

The gluconeogenesis also represents an essiential part of the carbon metabolism. It basically depicts the reverse reaction of the glycolysis although partly catalyzed by other enzymes (Figure 7 A). This pathway becomes relevant if *E. coli* grows on substrates other than hexoses or pentoses which can be utilized via the EMP pathway. Therefore the aim of the gluconeogenesis is the synthesis of glucose from non-carbohydrate substrates like pyruvate, lactate, acetate, glycerol or glucogenic amino acids. Since gluconeogenesis is a very energy consuming process this pathway is activated strictly on demand. The basic precursors for the gluconeogenesis are oxaloacetate and pyruvate which are converted to phosphoenolpyruvate (PEP). Based on PEP glucose can be rebuilt and again fed into the glycolysis.

1.7.3 Pentose Phosphate shunt

The pentose phosphate shunt starting with glucose as a substrate is another essential pathway for the provision of precursors for the nucleotide biosynthesis and the redox reagent nicotinamide adenine dinucleotide phosphate (NADPH) for several synthetic processes (Figure 7 A). It is estimated that about 20 % of the degraded glucose is used in this pathway depending on the metabolic state of the bacterium (Fuhrer *et al.*, 2005; Gottschalk, 1986). The pathway can be divided into an oxidative, irreversible portion and into a reductive, reversible part, whereas the two parts of the pathways do not have to run necessarily consecutively. However, the oxidative part includes the degradation of glucose to ribulose-5-phopsphate and NADPH where the first product can be used as a basic module for the synthesis of nucleosides and the latter as a redox equivalent in anabolic pathways. The regenerative path includes the conversion of ribulose-5-phosphate into fructose-6-phosphate and glyceraldehyde-3-phosphate. Moreover, it allows the transformation of various carbohydrates into another. These resulting intermediates and end products can be, again, fed into the glycolysis.

1.7.4 Tricarboxylic Acid Cycle (TCA)

The Tricarboxylic acid cycle (TCA), also referred to as citric acid or Krebs cycle, is a catabolic pathway of aerobic respiration. A lot of different metabolic intermediates from carbohydrate, lipid, and protein metabolism are fed into this cycle (Figure 7 A). Besides the generation of precursors for e.g. the synthesis of certain amino acids or pyrmidine the cycle provides mainly energy in form of ATP as well as the reducing agents NADH and $FADH_2$ (flavine adenine dinucleotide) through the oxidation of acetate into carbon dioxide. NADH and $FADH_2$ which are generated by the TCA cycle are directly fed into the respiratory chain and the connected oxidative phosphorylation (electron transport) that drives the generation of chemical energy in the form of ATP.

The net yield of the glycolysis, the TCA and the respiratory chain results in 38 ATP molecules per molecule glucose. Therefore these pathways are absolutely essential to provide *E. coli* with an appropriate level of energy and cellular material which is needed to survive.

1.7.5 Mixed Acid Fermentation

Under anaerobic conditions pyruvate is further degraded via the mixed acid fermentation into intermediates and end products like lactate, ethanol, formate or acetate (Figure 7 A). Fermentation is a metabolic pathway which leads to the reoxidation of NADH to NAD^+ (Clark, 1989). This has important relevance due to the fact that NAD^+ is needed in the glycolysis or the Entner-Doudoroff pathway, another possible metabolic pathway of *E. coli* to breakdown sugars to pyruvate. In this way NAD^+ is “recycled”.

In addition to the above mentioned metabolic pathways *E. coli* possesses several other pathways, which are alltogether important for the generation of both energy and precursors of cellular components. These pathways do not run independently from each other, but rather build a metabolic network. Therefore, a complex regulation that senses both, the environment and the metabolic state of the bacterium and that allows for the optimal supply by activating the most suitable pathways is necessary.

1.8 Aims of the study

In order to successfully survive and persist during the colonization of diverse ecological niches bacteria have to adapt to a lot of different environments and selective pressures in the favored host niche (Dobrindt *et al.*, 2010). The urinary tract represents one of the more challenging environmental conditions for *E. coli*. The mechanisms how long-term persistence of UPEC is achieved are so far poorly understood, but there is increasing evidence that bacterial genome plasticity plays an important role for the adaptation to the host niche (Dobrindt *et al.*, 2010). Different genomic approaches could show that bacterial growth in the urinary tract triggers virulence attenuation by genome reduction which includes mechanisms like single base mutations, deletions or DNA rearrangements (Klemm *et al.*, 2006; Roos *et al.*, 2006). These adaptations could illustrate an evolutionary principle for the establishment of a commensal life style of *E. coli* strains causing asymptomatic infections. Re-sequencing studies of ABU strains after therapeutic bladder colonization revealed frequent mutations of the genes coding for the TCS BarA/UvrY upon prolonged growth in the bladder (Zdziarski *et al.*, 2010). This TCS is an important regulatory system that is crucial for the efficient coordination of central metabolic pathways and affects a multitude of important cellular functions (Baker *et al.*, 2002; Pernestig *et al.*, 2003; Romeo *et al.*, 1993) .

The overall aim of the study was to investigate the importance of the TCS BarA/UvrY for adaptation of *E. coli* to long-term growth in the urinary tract.

Due to the apparent importance of the global TCS BarA/UvrY one key question of the study was, if this TCS might represent a mutational hot spot during long-term growth in the bladder. In order to address this question both, *barA* and *uvrY*, were analyzed by sequencing to determine nucleotide sequence variability in UPEC, ABU, health-care associated ABU (HA-ABU) and fecal *E. coli* isolates. Afterwards the phenotypic traits affected by selected *uvrY* allelic variants were investigated focusing on potential growth advantages in urine. Gene expression studies were used to further characterize the impact of the UvrY variants on the expression of components of the Csr system as well as on CsrA-dependent phenotypes.

In order to learn more about the specific adaptation mechanisms of *E. coli* to the urinary tract, a second project of this study was to investigate the metabolic constraints in urine. Accordingly, the question was addressed if and to what extent the genetically very closely related *E. coli* strains CFT073, 83972 and Nissle 1917 differ with regard to their metabolic activity in pooled human urine. For this purpose, the differences in the extracellular metabolome of the three strains during growth in pooled human urine were determined.

This thesis aims at a better understanding of adaptational strategies employed by *E. coli* strains upon growth in urine. The results can contribute to the long-term objective to develop alternative therapeutic strategies against UTI.

2 Material

2.1 Bacterial strains

All bacterial strains used in this study and their respective properties are listed in Table 3.

Table 3: Bacterial strains used in this study.

***E. coli* strain**	**Description/ Genotype**	**Reference**
83972	asymptomatic bacteriuria isolate	Lindberg & Winberg, 1976
CFT073	wild type, pyelonephritis isolate; O6:K2:H1	Mobley *et al.*, 1990
DH5α	*F-, endA1, hsdR17, supE44, thi-1, recA1, gyrA96, relA1, Δ(argF-lac) U169, λ-, φ80lacZΔM15*	Bethesda Research Laboratories
MG1655 (# 539)	wild type, K-12 isolate; OR:H48:K-, *F-, λ-, ilvG-, rfb-50, rph-1*	Blattner *et al.*, 1997
MG1655 Δ*uvrY*	derivative of MG1655, exchange of native *uvrY* by the *sacB-cat* cassette of pCVD422_cat via homologous recombination	This study
MG1655 *uvrY*::*uvrY* 1972	derivative of MG1655, exchange of native *uvrY* by the *uvrY* allele of strain # 1972 via homologous recombination	This study
MG1655 *uvrY*::*uvrY* 1976	derivative of MG1655, exchange of native *uvrY* by the *uvrY* allele of strain # 1976 via homologous recombination	This study
MG1655 *uvrY*::*uvrY* 1980	derivative of MG1655, exchange of native *uvrY* by the *uvrY* allele of strain # 1980 via homologous recombination	This study
MG1655 *uvrY*::*uvrY* 1992	derivative of MG1655, exchange of native *uvrY* by the *uvrY* allele of strain # 1992 via homologous recombination	This study
MG1655 *uvrY*::*uvrY* 2000	derivative of MG1655, exchange of native *uvrY* by the *uvrY* allele of strain # 2000 via homologous recombination	This study
MG1655 *uvrY*::*uvrY* 2031	derivative of MG1655, exchange of native *uvrY* by the *uvrY* allele of strain # 2031 via homologous recombination	This study
MG1655 *uvrY*::*uvrY* 83972	derivative of MG1655, exchange of native *uvrY* by the *uvrY* allele of strain 83972 via homologous recombination	This study
MG1655 *uvrY*::*uvrY* MG1655	derivative of MG1655, re-insertion of native *uvrY* by the native *uvrY* of strain MG1655 via homologous recombination	This study
MG1655^{Zeo}	derivative of MG1655, insertion of the zeocine resistance cassette by phage P1 transduction; Zeo^{R}	This study

***E. coli* strain**	**Description/ Genotype**	**Reference**
Nissle1917 (DSM6601, EcN)	wild type, fecal isolate; O6:K5:H1	Nissle, 1918
SW102 λ pir	Recombineering strain EL250 *gal +, λ pir lysogen*	Warming *et al.* (2005)
# 255	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 256	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 1963	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 1965 - 1969	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 1972	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 1976 - 1978	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 1980 - 1985	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 1987 - 2011	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 2014 - 2021	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 2023 - 2024	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 2026	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 2028	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 2030 - 2035	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
# 2037	asymptomatic bacteriuria isolate	IfH Strain collection (Münster)
IMI 961	asymptomatic bacteriuria isolate	IMIB strain collection (Würzburg)
IMI 963 - 970	asymptomatic bacteriuria isolate	IMIB strain collection (Würzburg)
# 2038	fecal isolate	IfH Strain collection (Münster)
# 2040 - 2042	fecal isolate	IfH Strain collection (Münster)
# 2044 - 2045	fecal isolate	IfH Strain collection (Münster)
# 2048 - 2054	fecal isolate	IfH Strain collection (Münster)
# 2056 - 2057	fecal isolate	IfH Strain collection (Münster)
# 2059 - 2060	fecal isolate	IfH Strain collection (Münster)
# 2063 - 2069	fecal isolate	IfH Strain collection (Münster)
# 2071	fecal isolate	IfH Strain collection (Münster)
# 2075	fecal isolate	IfH Strain collection (Münster)
IMI 917	clinical isolate from urinary tract infections	IMIB strain collection (Würzburg)
IMI 924	clinical isolate from urinary tract infections	IMIB strain collection (Würzburg)
IMI 926	clinical isolate from urinary tract infections	IMIB strain collection (Würzburg)
IMI 928 - 929	clinical isolates from urinary tract infections	IMIB strain collection (Würzburg)
# 516	clinical isolate from urinary tract infections	IfH Strain collection (Münster)
# 587	clinical isolate from urinary tract infections	IfH Strain collection (Münster)
# 960	clinical isolate from urinary tract infections	IfH Strain collection (Münster)
# 1336 - 1354	clinical isolates from urinary tract infections	IfH Strain collection (Münster)
# 1517 - 1531	clinical isolates from urinary tract infections	IfH Strain collection (Münster)

E. coli strain	Description/ Genotype	Reference
# 1535 - 1540	clinical isolates from urinary tract infections	IfH Strain collection (Münster)
# 1291 - 1318	clinical isolates from catheterized patients	IfH Strain collection (Münster)
# 1320 - 1321	clinical isolates from catheterized patients	IfH Strain collection (Münster)
# 1532 - 1534	clinical isolates from catheterized patients	IfH Strain collection (Münster)

2.2 Plasmids

All plasmids used in this study are listed below (Table 4).

Table 4: Plasmids used in this study.

Plasmid	Description	Reference
pCVD422	*oriRγ*, *mob*RP4, *sacB, bla*	Donnenberg & Kaper, 1991
pCVD422_cat	derivative of pCVD422, insertion of *cat* cassette downstream of *sacB*; *bla, cat*	This study
pKD3	*oriRγ*, *bla*, *cat*-cassette flanked by FRT sites	Datsenko & Wanner, 2000
pKD46	*repA101* (ts), *ara*Bp_*γ-ß*-exo (*λ*-Red recombinase under the control of *araB* promoter), *bla*	Datsenko & Wanner, 2000

2.3 Oligonucleotides

All synthetic oligonucleotides (primers) used in this study for PCR, qRT-PCR or *λ*-Red based recombination (Datsenko & Wanner, 2000) were purchased from Sigma-Genosys (Steinheim, Germany). Their respective sequences as well as the applications are listed in Table 5. Homologous sequence regions for λ-Red mediated recombination are written in capital letters.

Table 5: oligonucleotides used in this study.

Primer	Primer sequence (5´ → 3´)	Application	Product size
E.c. uvrY Fw 1	gagaaaaatcgaatacccacca	Amplification and sequencing of *uvrY*	
E.c. uvrY Fw 2	gaaaaatcgaatacccaccatt		
E.c. uvrY Fw 3	aaaaccctttaccagcgaaa		
E.c. uvrY Fw 4	gatacgacgcattctggaaga		
E.c. uvrY Rev 1	tttcaggtctttcgctttgc		
E.c. uvrY Rev 2	gataaccgtaccaccagcatc		
E.c. uvrY Rev 3	gcatgactttcgctggtaaa		
E.c. uvrY Rev 4	ccaggtgagtcagctcaaca		

Primer	Primer sequence (5´ → 3´)	Application	Product size
E.c. barA Fw 1	tgcgtcgttttgcagagtag	Amplification and sequencing of *barA*	
E.c. barA Fw 2	gtttcctcgccagctcact		
E.c. barA Fw 3	gaagcggcgcgtattaaatc		
E.c. barA Fw 4	tatttcccgccgtcatgg		
E.c. barA Fw 5	gcgtgtctgctgaaaccatt		
E.c. barA Fw 6	atttagcgcgcgatatgct		
E.c. barA Rev 1	gcatccggcataaacacag		
E.c. barA Rev 2	gcgctaaatcggtttttcct		
E.c. barA Rev 3	gcgccgataagtttcaggtt		
E.c. barA Rev 4	tgtggaaccagaaagttgaacc		
E.c. barA Rev 5	ggaatactttccagaatcagcttacc		
E.c. barA Rev 6	attaagcgccagccaaaaat		
E.c. barA Rev 7	cgcaggctgtagttggtcat		
E.c. barA Fw 1b	ccattggcgagactttctca		
E.c. barA Rev 7b	tggagttccgttatgggaca		
CsrA_RT_Fw	cgagttggtgagaccctcat	qRT-PCR	161 bp
CsrA_RT_Rev	ctggactgctgggatttttc		
CsrB_F	gatgacacttctgaaggacacacc		224 bp
CsrB_R	tgcggccaatcgttcatcct		
CsrC_Fw	gaggcgaagacagaggattg		73 bp
CsrC_Rev	tccgtgttgattccatttcc		
47_csrD_RT_fw	acccgggctactcgtatctt		107 bp
48_csrD_RT_Rv	tcacaaagcagcgtatccag		
frr_left	ggcaagcgtaacggtagaag		208 bp
frr_right	cttgttctgcttcaccacga		
gapA_F	gttgtcgctgaagcaactgg		170 bp
gapA_R	agcgttggaaacgatgtcct		
uvrY RT Fw	tttgccagtttgtctgaacg		104 bp
uvrY RT Rv	caccgttttcggactgagat		
3´uvrY_F	cagttttgatttaaacgtgg	Control primer for PCR screening of positive MG1655 Δ*uvrY::sacBcat* recombinants	540 bp
3´uvrY_R	gatagagtttgatgtagttg		
5´uvrY_F	gctgattttcatttttgctg		401 bp
5´uvrY_R	ctttacacattttaggtcttg		

Primer	Primer sequence (5´ → 3´)	Application	Product size
cat_in_pCVD442_Fw	ATGGATGACGGATGGCTGGCCGCTGTATGAATCCCGCCTGAAGGGAAAGCTgtaggctggagctgcttcg	Amplification of the *cat* cassette with homologous sequences for recombination into pCVD422 to generate pCVD422_cat	1102 bp
cat_in_pCVD442_Rv	TGGTAATGACTCCAACTTATTGATAGTGTTTTATGTTCAGATAATGCCCGCCTCCTTAGTTcctattccgaag		
Pins1	taggcccgtagtctgcaaat	Amplification of *sacB-cat* along with Pins2	2889 bp
Pins2	gtaggctggagctgcttcg	Amplification of *cat* cassette from pKD3	1001 bp
Pins3	cctccttagttcctattccgaag		
uvrY::sacB-cat_for	ACGAATGACTAACTATCAGTAGCGTTATCCCTATTTCTGGAGATATTCCTtaggcccgtagtctgcaaat	Amplification of *sacB-cat* from pCVD422_cat with homologous sequences to the up- and downstream regions of *uvrY* region in MG1655; replacement of *uvrY* with *sacB-cat*	2989 bp
uvrY::sacB-cat_rev	TGGCTGGCTGGTTACGGTTTTTAAAAACGCTTTTGCGTCAAACTGATCACgtaggctggagctgcttcg		

2.4 Chemicals, enzymes and kits

All chemicals and enzymes used in this study were purchased from the following providers: Affymetrix (High Wycombe, UK), Applichem (Darmstadt, Germany), Applied Science (Penzberg, Germany), BioRad (München, Germany), Dianova (Hamburg, Germany), Difco (Augsburg, Germany), GE Healthcare Life Sciences (Freiburg, Germany), Life Technologies (Darmstadt, Germany), Macherey-Nagel (Düren, Germany), MBI Fermentas (St. Leon-Rot, Germany), Merck-Millipore (Darmstadt, Germany), New England Biolabs (Frankfurt am Main, Germany), Peqlab (Erlangen, Germany), Promea (Mannheim, Germany), Qiagen (Hilden, Germany), Roche Diagnostics (Mannheim, Germany), Roth (Karlsruhe, Germany), Serva (Heidelberg, Germany), Sigma-Aldrich (Taufkirchen, Germany), Thermo Scientific (Schwerte, Germany) and VWR (Darmstadt, Germany).

The following commercial kits were used:

- BigDye® Terminator v3.1 Cycle Sequencing Kit, Life Technologies
- Gene Images AlkPhos Direct Labelling and Detection System, GE Healthcare Life Sciences
- Micro BCA™ Protein assay kit, Thermo Scientific
- NucleoSpin® Gel and PCR clean up kit, Macherey-Nagel
- NucleoSpin® Plasmid kit, Macherey-Nagel
- RNeasy Mini kit, Qiagen
- SuperScript™ Reverse III Transcriptase, Life Technologies

2.5 Buffers and solutions

If available, the manufacturers´ supplied buffers and solutions were used for the respective reactions. All self-prepared buffers and solutions are documented in the respective method. Sterile distilled (dH_2O) water was used for the preparation of all buffers and solutions if not stated otherwise.

2.6 Media and agar plates

Before usage all media were autoclaved for 20 min at 121 °C if not stated otherwise. Supplements for media and plates were added after being sterile filtered. Heat-unstable supplements were added after cooling to a temperature of 50 °C.

2.6.1 Media

Lysogeny Broth (LB) medium	Tryptone	10 g
	Yeast extract	5 g
	NaCl	5 g
	dH_20	ad 1 l
dYT	Tryptone	16 g
	Yeast extract	10 g
	NaCl	5 g
	dH_20	ad 1 l
Urine	Human urine was collected from several healthy volunteers with no recent history of UTI or antibiotic treatment. Male and female urine was pooled at the rate of 1:1. Fresh urine was sterile filtered by using Millipore Express™ Plus 0.22 µm (Merck-Millipore) and frozen at -80 °C. Sterile urine was stored at 4 °C no longer than seven days.	

2.6.2 Agar plates

Lysogeny Broth (LB) agar	Tryptone	10 g
	Yeast extract	5 g
	NaCl	5 g
	Agar	15 g
	dH_20	ad 1 l
dYT agar	Tryptone	16 g
	Yeast extract	10 g
	NaCl	5 g
	Agar	16 g
	dH_20	ad 1 l
LB motility agar plates	LB medium (Lennox; Roth)	20 g
	Agar (Difco)	3 g
	dH_20	ad 1 l
Urine motility agar plates	sterile pooled human urine	800 ml
	1.5 % (w/v) Agar (Difco)	200 ml
Congo Red agar plates	Tryptone	8 g
	Yeast extract	4 g
	Agar	15 g
	CongoRed	32 mg
	Brilliant blue R-250 (Roth)	16 mg
	dH_20	ad 800 ml
Calcoflour plates	Tryptone	8 g
	Yeast extract	4 g
	Agar	15 g
	Fluorescent brightener 28 (Sigma)	128 mg
	dH_20	ad 800 ml

2.7 Antibiotics

If required, media and plates were supplemented with the antibiotics listed in Table 6 using the indicated concentrations. Stock solutions were sterile filtered and stored at -20 °C until use.

Table 6: Antibiotics used in this study.

Antibiotic	Stock concentration	Solvent	Working concentration
Ampicillin (Amp)	100 mg ml^{-1}	dH_2O	100 µg ml^{-1}
Chloramphenicol (Cm)	25 mg ml^{-1}	EtOH	25 µg ml^{-1}
Zeocin (Zeo)	100 mg ml^{-1}	dH_2O	50 µg ml^{-1}

2.8 Antibodies

All antibodies used in the study are shown in Table 7 and
Table 8.

Table 7: Primary antibody used in this study.

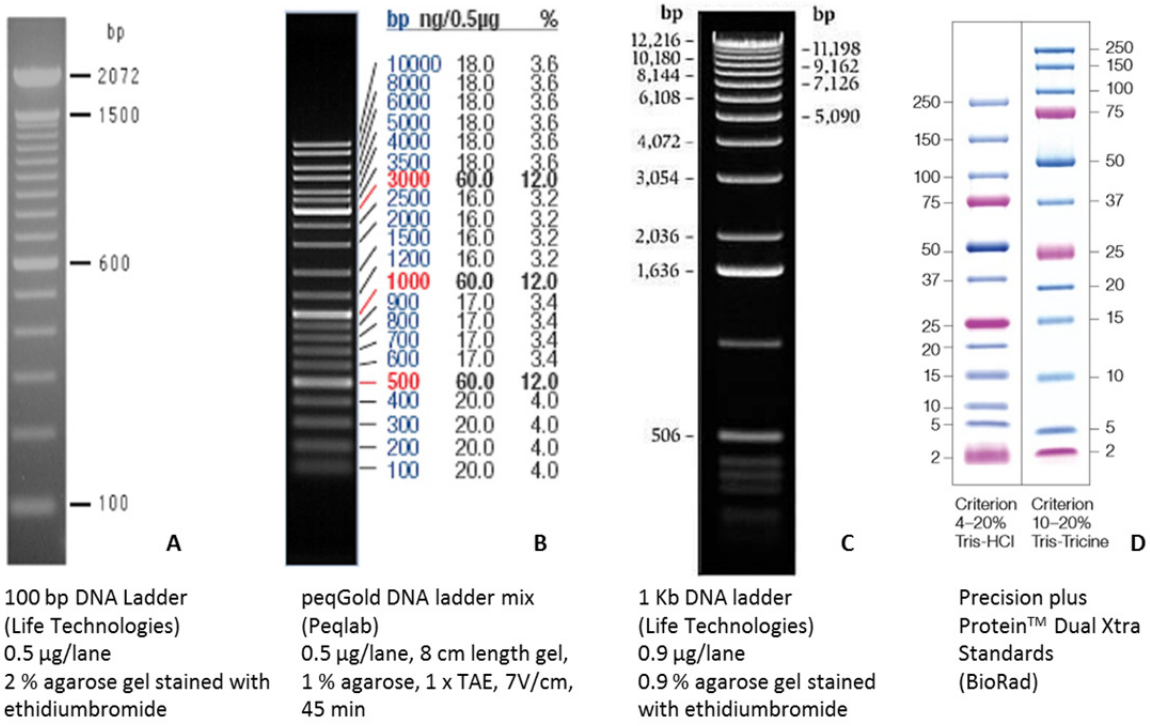
Primary antibody	Host	Stock concentration	Working concentration	Source
α-Rpos (polyclonal)	rabbit	-	1:10000	R. Hengge (Berlin)

Table 8: Secondary antibody used in the study.

Secondary antibody	Host	Stock concentration	Working concentration	Source
Peroxidase-conjugated AffinityPure Goat Anti-Rabbit IgG	Goat	0.8 mg ml^{-1}	1:16000	Dianova (Hamburg, Germany)

2.9 DNA / protein size markers

To determine the size of DNA fragments in agarose gels the 100 bp DNA ladder (Life Technologies), the PeqGold DNA ladder mix (Peqlab) and the 1 Kb DNA ladder (Life Technologies) were used. For Western Blot applications the Precision Plus Protein™ Dual Xtra Standards (BioRad) was used (Figure 8).



100 bp DNA Ladder (Life Technologies) 0.5 µg/lane 2 % agarose gel stained with ethidiumbromide

peqGold DNA ladder mix (Peqlab) 0.5 µg/lane, 8 cm length gel, 1 % agarose, 1 x TAE, 7V/cm, 45 min

1 Kb DNA ladder (Life Technologies) 0.9 µg/lane 0.9 % agarose gel stained with ethidiumbromide

Precision plus Protein™ Dual Xtra Standards (BioRad)

Figure 8: DNA and protein markers used in this study.

(A) 100 bp DNA ladder from Life Technologies. **(B)** DNA ladder mix from Peqlab, **(C)** 1 Kb DNA ladder supplied by Life Technologies and **(D)** the Precision plus Protein™ Dual Xtra Standards from BioRad.

2.10 Technical equipment

All supply materials used in this study were obtained from Agilent Technologies (Waldbronn, Germany), Becton Dickinson (Heidelberg, Germany), BioRad (München, Germany, Diagonal (Münster, Germany), Eppendorf (Hamburg, Germany), Greiner Bio-One (Frickenhausen, Germany), Peqlab (Erlangen, Germany), Merck-Millipore (Darmstadt, Germany), Miltenyi Biotec (Bergisch Gladbach, Germany), MP Biomedicals (Eschwege, Germany), Sarstedt (Nümbrecht, Germany), Tecan (Crailsheim, Germany), Thermo Scientific (Schwerte, Germany), VWR (Darmstadt, Germany) and Whatman (Dassel, Germany). Technical equipment used in this work is listed in Table 9.

Table 9: Technical equipment used in this study.

Equipment	Manufacturer (model)
Autoclave	Systec (Systec VX-150)
Centrifuges	Eppendorf (5417 R, 5424, 5415 D, 5804 R); Hettich (Rotanta 460RS); Biozym (Sprout)
Chemiluminescence documentation	BioRad (ChemiDoc™ MP System)
Clean bench	The Baker Company (SterilGard)
Colony Counter	Don Whitley Scientific (ProtoCol 2)

Equipment	Manufacturer (model)
Electrophoresis system	BioRad ((Mini) Sub-cell GT Cell)
Electroporator	BioRad (Micropulser)
Freezer	Liebherr (Profi line, comfort, premium no frost -20 °C); Thermo Scientific (HeraFreeze -80°C); Sanyo (VIP series -80 °C)
Gel documentation	Biozym (AlphaImager® EP)
Homogenizer	Peqlab (Precellys)
Ice machine	Scotsman (AF100)
Incubator (37 °C)	Heraeus (Function Line)
Lab rocker	Heidolph (Duomax 1030)
Laboratory scales	Sartorius; Ohaus (Explorer)
Magnetic stirrer	Heidolph (MR3001); VWR (VMS-C7)
Micro syringe	Hamilton (705 LT SYR)
Micropipettes	Eppendorf (Research plus, Reference)
Microwave oven	AEG (Micromat)
PCR thermo cycler	BioRad (T100™ Thermal cycler, T personal); Eppendorf (Mastercycler® personal); Applied Biosystems (GeneAmp® PCR System 9700)
pH meter	Schott-Instruments (CG 842)
Power supplies	Biozym (Consort E143); BioRad (Power Pac 300); Peqlab (PeqPower 250 V)
Protein transfer system	Roth (Semi-dry blotter); Peqlab (Semi-dry blotter)
Real Time PCR Cycler	BioRad (CFX96 Real-Time System)
Sequencer	Applied Biosystems (ABI PRISM 3130xl Genetic Analyzer)
Shaking incubator	Infors (HT Multitron); IKA (KS 4000i control)
Spectrophotometer	Amersham Biosciences (Ultrospec 2100 pro); Thermo Scientific (Nanodrop 2000)
Speedvac	Eppendorf (Concentrator plus)
TapeStation	Agilent (TapeStation 2200)
Tecan reader	Tecan (Infinite F200)
Thermoblocks	Bioer (Mixing Block MB-102); Peqlab (Thriller)
UV-crosslinker	UVP (CL-1000 ultraviolet crosslinker)
Vacuum pump	Vacuubrand (MZ 2 NT)
Video printer	Sony (Digital Graphic Printer UP-D897)
Vortexer	Merck (Eurolab); Scientific Industries (Vortex-Genie 2)
Water bath	Memmert (WNE 22)

3 Methods

3.1 Bacterial cultures

3.1.1 Cultivation

All bacterial strains used in this study were routinely grown in LB medium or on LB agar plates. If appropriate, ampicillin, chloramphenicol or zeocine were added to the growth medium at concentrations of 100 µg ml^{-1}, 25 µg ml^{-1} or 50 µg ml^{-1}, respectively. Cultures were usually incubated at 37 °C if not stated otherwise. Liquid cultures were inoculated with single bacterial clones from LB agar plates and incubated overnight under shaking conditions (180 rpm).

3.1.2 Bacterial preservation

For long-term storage a stock culture was generated for every bacterial strain. For this, an aliquot of a bacterial culture was mixed with glycerol to a final concentration of 25 % (v/v). Finally the glycerol stock was frozen at -80 °C.

3.1.3 Bacterial growth

Bacterial growth in liquid cultures was determined by means of photometric measurement of the optical density at a wavelength of 595 nm / 600 nm.

3.2 Molecular biological methods - DNA

3.2.1 Extraction of genomic DNA

The extraction of genomic DNA was done by a modified protocol according to Wilson (2001). For this purpose, strains were cultivated in 3 ml LB medium overnight. On the next day cells from 1.5 ml culture were harvested by centrifugation (2 min, 6000 x g) to form a compact pellet which was resuspended in 561 µl TE buffer. 30 µl 10 % (w/v) SDS solution and 4 µl proteinase K (20 mg ml^{-1}) were added to lyse the cells. The suspension was incubated for 30 min at 65 °C. For the removal of RNases RNase A (Sigma) and RNase T1 (Sigma-Aldrich) were added to final concentrations of 10 µg ml^{-1} and 25 U ml^{-1}, respectively. The mixture was incubated for one hour at 37 °C. Following this 100 µl 5 M NaCl was added. The next step was the addition of 75 µl CTAB solution (10 % (w/v) CTAB in 0.7 M NaCl) which resulted in a milky-turbid suspension which was incubated at 65 °C. As

soon as the lysate cleared off it was mixed thoroughly with an equal volume of chloroform/isoamylalcohol (ratio of 24:1) and the mixture was centrifuged (5 min, 16.000 x g). The aqueous, viscous upper phase was transferred to a fresh reaction tube and mixed with 0.6 volume isopropanol in order to precipitate the genomic DNA. The mixture was centrifuged (5 min, 16.000 x g) and the DNA was washed once with 70 % (v/v) ethanol, dried for 15 min at 37 °C and afterwards solved in 100 µl dH_20.

TE buffer	Tris-HCL (pH 7.5)	10 mM
	EDTA	0.1 mM

3.2.2 Photometric determination of nucleic acid concentration

The photometric determination of the DNA or RNA content of a sample can be made due to the property of nucleic acids to absorb light of a wavelength of 260 nm. Measuring the absorption was carried out at wavelengths of 230 nm, 260 nm and 280 nm using the NanoDrop 2000 (Thermo Scientific). According to Sambrock *et al.* (1989), the absorbance at 260 nm is a measure of the nucleic acid concentration, while the absorbance at 280 nm indicates the protein content of the solution. The ratio A260 / A280 can be considered as a measure of the purity of the DNA, which in the case of pure DNA/RNA should be between 1.8 and 2.0. The ratio A_{260}/A_{230} indicates the degree of salinity in the sample and should have a value of 2.0. An $A_{260} = 1$ corresponds to a DNA concentration of 50 µg ml^{-1} or a RNA concentration of 40 µg ml^{-1}, respectively. Therefore, the following equation results for determining the concentration of double-stranded DNA (or RNA and single stranded DNA):

$$\text{Concentration } (\mu g\ ml^{-1}) = A_{260} * \text{dilution factor} * 50\ (40)\ \mu g\ ml^{-1}$$

3.2.3 Polymerase chain reaction

The polymerase chain reaction is an efficient method for the amplification of any DNA fragments *in vitro*. Besides of desoxyribonucleotide-triphopshates (dNTPs) and the DNA polymerase that are needed for the synthesis, synthetic oligonucleotides, called primers, are required which anneal in opposite directions to the DNA strand. The primer annealing is possible due to a previous heat denaturation of the template in its single strands so that the primers can subsequently bind to the template at their specific annealing temperature which depends on the sequence of the primer itself. These primers are used for the heat-resistant DNA polymerase as the starting point for the synthesis of the respective complementary strand from 5' to 3' direction (5`→ 3`), the so-called elongation. The dNTPs serve as a substrate for the generation of the new strand. By multiple repetitions of these steps an exponential amplification of the desired DNA region is achieved.

PCR was performed according to an adapted program based on the specific sequences to be amplified and the reaction additives (depending on the supplier). The thermal cycling profile was designed according to the elongation temperature and the annealing temperature of the individual primers and the length of the expected amplification product, respectively. For each reaction a negative control without template was carried.

The standard PCR protocol was performed using a Taq polymerase such as GoTaq® Green Master Mix (Promega) or REDTaq® Ready Mix (Sigma-Aldrich) that do not have a proof-reading function. The Phusion High-Fidelity DNA polymerase (Thermo Scientific) was used to amplify DNA fragments that were used in downstream applications like sequencing or cloning. For those purposes a proof-reading activity was required that prevents misincorporations of nucleotides during extension.

The reaction mixtures for Taq polymerase (Table 10) or the proof-reading polymerase (Table 11) were prepared according to the following pipetting schemes:

Table 10: Pipetting instructions for a standard PCR using a Taq DNA Polymerase.

Component	Final concentration	25 µl rxn
GoTaq® Green Master mix, 2 x or REDTaq® ReadyMix™, 2 x	1 x	12.5 µl
Primer forward (10 µM)	0.1 – 1.0 µM	0.25 µl – 2.5 µl
Primer reverse (10 µM)	0.1 – 1.0 µM	0.25 µl – 2.5 µl
DNA template	100 ng	1 µl
Nuclease free water	N.A.	ad 25 µl

The thermal cycling profile for GoTaq® or REDTaq®:

1.	Initial denaturation	95 °C/ 94 °C	2 min/ 5 min	
2.	Denaturation	95 °C/ 94 °C	30 sec/ 1 min	25-35 x
3.	Annealing	57°C - 62 °C	30 sec	
4.	Elongation	72 °C (1min/1 kb)	30 sec - 3 min	
5.	Final elongation	72 °C	10 min/ 5 min	

Table 11: Pipetting instructions for a PCR using the Phusion High-Fidelity DNA polymerase.

Component	Final concentration	20 µl rxn
5 x Phusion HF buffer	1 x	4 µl
dNTPs (10 mM)	0.2 mM	0.4 µl
Primer forward (10 µM)	0.5 µM	1 µl
Primer reverse (10 µM)	0.5 µM	1 µl
Polymerase (2 U μl^{-1})	0.02 U μl^{-1}	0.2 µl

Component	Final concentration	20 µl rxn
DMSO, optional	*3* % (v/v)	-
DNA template	10 - 100 ng	1 µl
Nuclease free water	N.A.	ad 20 µl

The thermal profile for the proof-reading PCR:

1.	Initial denaturation	98 °C	2 min	
2.	Denaturation	98 °C	10 sec	25 x
3.	Annealing	57°C - 62 °C	30 sec	
4.	Elongation	72 °C (1min/1 kb)	30 sec - 3 min	
5.	Final elongation	72 °C	10 min	

3.2.4 Analysis of PCR fragments

The separation of DNA fragments was carried out by means of horizontal gel electrophoresis, a method to separate and analyze nucleic acids according to their size. The phosphate groups in the DNA backbone carry negatively charged oxygen atoms – giving the DNA molecule an overall negative charge. This leads to a migration of nucleic acids in the agarose matrix towards the anode. That way the DNA is separated depending on the molecular mass and confirmation. To determine the size of the DNA fragments in the agarose gel a suitable molecular size marker (Figure 8) was loaded in each gel electrophoresis.

Agarose gels with concentrations between 0.8 % (w/v) and 1.5 % (w/v) prepared in 1 x TAE were used depending on the size of the fragments to be separated for either analytical or preparative purposes. The samples were mixed with an appropriate volume of 6 x DNA loading dye and separated using an electric voltage of 6 V/cm. Afterwards, the gels were stained with ethidium bromide (0.5 µg ml^{-1}) and washed with dH_20. The DNA in the gel was visualized using UV-light at a wavelength of 302 nm and results were documented with a CCD camera. For preparative purposes, the DNA was analyzed with UV-light at a wavelength of 365 nm to avoid DNA damages and cut using a scalpel.

TAE (1 x)	Tris	2 M
	EDTA	50 mM
	Glacial acetic acid	5.71 % (v/v)
	pH 8.3	

DNA loading dye (6 x)	Tris-HCL (pH 7.6)	10 mM
	Bromphenol blue	0.03 % (w/v)
	Xylencyanol	0.03 % (w/v)
	Glycerol	60 % (v/v)
	EDTA	60 mM

3.2.5 Gel extraction

DNA extraction from agarose gels was performed using the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) according to the manufacturer's instructions. The extracted DNA was finally taken up in 25 µl dH_20 and stored until use at -20 ° C.

3.2.6 Ethanol precipitation of DNA

The concentration and purification of nucleic acids was carried out by means of an alcohol precipitation protocol. For this purpose the DNA solution was mixed with 1/10 vol. of 3 M sodium acetate (pH 5.2), 2.5 vol. of ice-cold 100 % ethanole, 1 µl glycogen (20 mg ml^{-1}) and incubated for 1 h at -80 °C. The precipitated DNA was pelleted by centrifugation for 30 min at 16.000 x g and 4 °C. The supernatant was removed and the pellet was washed with 1 ml ice-cold 80 % (v/v) ethanol to remove residues of sodium acetate. This was followed by another centrifugation step for 5 min at 16.000 x g and 4 ° C. The alcohol was withdrawn with a pipette; the DNA pellet was dried at 50 °C for 5 min and finally resolved in 20 µl dH_20.

3.2.7 PCR clean-up

In some instances, the PCR product was purified directly after amplification using a commercial kit. For this purpose the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) was used. The kit allows an efficient purification of PCR products from amplification reactions to remove excessive nucleotides, primers and salts which could interfere with subsequent reactions. The clean-up was done according to the manufacturer's instructions.

3.2.8 DNA sequencing

The nucleotide sequences of genomic DNA or plasmid constructs (see 3.2.11) were determined by Sanger sequencing (Sanger *et al.*, 1977) using the BigDye® Terminator v3.1 Cycle Sequencing kit (Life Technologies). To optimize the results PCR products were purified before sequencing by means of **Exo**nuclease I (New England Biolabs) and **S**hrimp **A**lkaline **P**hosphatase (Affymetrix) treatment (**EXOSAP**).

3.2.8.1 PCR purification by EXOSAP

5 µl of PCR product were mixed with 1 µl SAP (1 U μl^{-1}), 0.05 µl Exonuclease I (20 U μl^{-1}) and 0.95 µl TE buffer (pH 7.5). The mixture was incubated in a PCR cycler according to the following protocol: 45 min at 37 °C, 10 min at 80 °C and a final step of 4 °C.

TE buffer	Tris	10 mM
	EDTA	10 mM
	pH 7.5	

3.2.8.2 Sequencing protocol

The sequencing-PCR mix for one sample was:

0.5 µl	BigDye® Terminator v3.1 Ready Reaction Mix (kit component)
1.8 µl	5 x ABI Sequencing buffer (kit component)
2.0 µl	Primer (10 µM)
2.0 µl	PCR product (or 200 ng plasmid DNA)
3.7 µl	dH_20

The thermal cycler, for the sequencing PCR, was programmed as follows: a denaturation step at 95 °C followed by an annealing step for 5 sec at the optimal temperature for the respective primer. Elongation was performed for 4 min at 60 °C. The profile was repeated for 25 cycles. Sequencing products were purified by ethanol precipitation (see 3.2.6) and resuspended in 20 µl HiDi™-Formamide (Life Technologies). The analysis of the sequences was done in an ABI PRISM 3130xl Genetic Analyzer sequencer (Life Technologies) at the Institute of Hygiene in Münster (Germany).

3.2.9 Preparation of electrocompetent cells

220 ml fresh dYT medium was inoculated with an *E. coli* overnight culture in a ratio of 1:220 and grown at 37 °C under shaking (180 rpm) until an OD_{600} of 0.4 -0.55 was reached. Afterwards, the flask containing the culture was transferred on ice and incubated for 20 min. The suspension was centrifuged (20 min, 2700 x g, 4 °C) and the pellet was washed three times with ice- cold dH_20. The cells were resuspended in 45 ml ice-cold 10 % (v/v) glycerol and washed another time. Finally the pellet was resuspended in 1 ml ice-cold 10 % (v/v) glycerol. Aliquots were tested for arcing by electroporation (see 3.2.10). If negative, aliquots of 80 µl were frozen and stored at -80°C until use. Electrocompetent cells containing the temperature-sensitive plasmid pKD46, which were used for the transformation of PCR products according to (Datsenko & Wanner, 2000) (see 3.3.1) were grown in 200 ml dYT under appropriate selective pressure at 30 °C. After reaching an OD_{600} of

0.4-0.6 6 ml 10 % (v/v) arabinose was added to induce the expression of the λ-Red-recombinase encoding on the plasmid pKD46. Subsequently the culture was incubated at 37 °C for 1 h. The remaining procedure was performed as described above.

3.2.10 Transformation of electrocompetent cells

Transformation is a process by which exogenous genetic material is transferred into a bacterial cell. The basic principle used in this work is the electroporation, a method to make cell membranes temporarily permeable to inject DNA into prokaryotic cells.

For this purpose an aliquot of electrocompetent cells was thawed on ice and mixed with either plasmid DNA, PCR or a ligation product. The mixture was transferred into a pre-cooled electroporation cuvette and the electroporation was carried out at 1.8 kV. Subsequently the cells were taken up in pre-warmed dYT medium and incubated at 37 °C under shaking (180 rpm) for 1 h. Appropriate volumes were plated on respective selective agar plates and incubated overnight. Transformation efficiency was checked the next day.

3.2.11 Isolation of plasmids

The isolation of plasmid DNA was carried out using the NucleoSpin®Plasmid kit (Macherey-Nagel). This method is based on the modified technique for alkaline lysis described by Birnboim & Doly (1979). The separation of plasmid DNA from chromosomal DNA and cell proteins was performed by anion-exchange chromatography on a NucleoSpin®–column. The preparation was done according to the manufacturer´s instructions. Plasmid DNA was eluted in 50 µl buffer AE and the concentration was measured using the Nanodrop 2000 (Thermo Scientific). The plasmids were stored at -20 °C until use.

3.2.12 Enzymatic hydrolysis of DNA

The enzymatic hydrolysis of DNA was done by restriction endonucleases using 1 unit (U) of enzyme for the digestion of 1 µg DNA. The final volume of the reaction mixture including DNA, enzyme and the manufacturer´s recommended buffer varied between 20-50 µl. The reaction mixture was incubated 37 °C depending of the specific requirements of the enzyme (stated by supplier). When appropriate, the reaction mix was stopped by heat inactivation (20 min, at 65 °C) and purified by means of NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel). Afterwards the digestion was checked by gel electrophoresis.

3.2.13 Ligation of DNA fragments

The ligation of DNA fragments with linearized vectors was performed using the T4 DNA ligase (New England Biolabs). A prerequisite for the ligation is the compatibility of the ends of the DNA fragment and the vector. In order to create blunt or sticky ends for ligation, both components required to be digested with restriction enzymes. Insert and vector were mixed in molar ratios of 1:1 - 3:1 depending on the size of the insert. Reactions were performed overnight at 16 °C in a final volume of 20 µl containing 2 µl 10 x ligation buffer and 1 µl T4 ligase.

3.2.14 Colony PCR

For a quick screening of the presence or absence of DNA inserts in plasmids a colony PCR was performed meaning that some colony material of individual transformants was directly added to the PCR reaction mixture (see 3.2.3, Table 10). The cells were lysed during the initial heating and the plasmid DNA served as a template for the amplification. The presence or absence of the amplicon was checked by electrophoresis on an agarose gel (see 3.2.4).

3.2.15 Dot blot

The Dot blot technique is used to determine the relative amount of target sequences in samples that are immobilized onto a membrane. By means of probe hybridization the target sequence can be detected. In this study, the technique was used to check if *barA* is present in the samples where the amplification of *barA* via PCR failed.

3.2.15.1 Vacuum Blotting

The Dot blot was done according to a modified protocol of Brown (2001). For this purpose 2 µg DNA of the strains to be examined were mixed with 40 µl 1 M NaOH and 2 µl 0.5 M EDTA (pH 8.0) in a final volume of 100 µl. The mixture was incubated at 100 °C for 10 min and cooled on ice afterwards. In the meantime the blotting equipment was prepared. A positively charged nylon membrane (Hybond N+, pore size: 0.45 µm; GE Healthcare Life Sciences) and a Whatman paper (GE Healthcare Life Sciences) with the appropriate size were equilibrated in dH_2O. The dot blot manifold was closed and a vacuum was applied. Prior to sample application, each well was washed with 500 µl dH_2O. Following this the DNA of the samples was immobilized on the membrane by applying a vacuum. After immobilization of the DNA on the membrane the wells were rinsed with 500 µl 0.4 M NaOH. The membrane was subsequently washed in 2 x SSC and placed on a piece of Whatman paper to remove the remaining SSC buffer. After drying the membrane was put in a plastic wrap and placed

with the DNA-side facing downwards on an UV transilluminator for 90 sec at 302 nm in order to crosslink the DNA to the membrane.

3.2.15.2 Probe labeling

The probe labeling was done using the Gene Images AlkPhos Direct Labeling and Detection System provided by GE Healthcare Life Sciences according to the manufacturer's instructions. This system allows the labeling of probe DNA with a specially developed thermostable alkaline phosphatase enzyme by means of a special cross linker solution. The CDP-Star™ chemiluminescent detection reagents utilize the probe-bound alkaline phosphatase to catalyze the decomposition of a stabilized dioxetane substrate which leads to the emission of chemiluminescence. For labelling of the probe - which was a previously amplified and purified fragment of *barA* – 10 ng DNA per ml hybridization buffer in a final volume of 10 µl in dH_2O were denatured for 10 min at 90 °C and cooled for 5 min on ice. Afterwards 10 µl reaction buffer, 2 µl labeling reagent and 10 µl cross-linker solution were added. The mixture was incubated for 30 min at 37 °C and then added to the hybridization solution.

3.2.15.3 Hybridization and detection

The hybridization of the membrane was performed overnight at 42 °C in 10-15 ml hybridization solution. The nylon membrane was pre-incubated at 42 °C in the hybridization solution for 1 hr. On the next day the membrane was washed twice with wash solution I for 20 min at 42 °C and for 5 min at room temperature (RT) with wash solution II. The membrane was placed on Whatman paper to remove the remaining wash solution, and then incubated for 2-5 min in 30-40 µl/cm^2 of the detection solution. The membrane was packed in saran wrap afterwards and the emitted chemiluminescence was detected using the ChemiDoc™ MP System (BioRad).

20 x SSC	Na_3Citrate	0.3 M
	NaCl	3 M
	pH 7.0	
Hybridization buffer	NaCl	0.5 M
	Blocking reagent	4 % (w/v)
Wash solution I	Urea	2 M
	SDS	0.1 % (w/v)
	Sodium phosphate	0.5 M
	NaCl	150 mM
	$MgCl_2$	1 mM
	Blocking reagent	0.2 % (w/v)
Wash solution II (20 x)	Tris base	1 M
	NaCl	2 M

1 x wash solution II was prepared by diluting the 20 x wash solution II in dH_2O and $MgCl_2$ was added in a final concentration of 2 mM. This working solution had to be prepared freshly before use.

3.2.16 Southern blot

The molecular biological technique of transferring fragments of DNA from agarose gels to a filter membrane and the subsequent detection of the fragments via probe hybridization is referred to as Southern Blot (Southern, 1975).

3.2.16.1 Capillary Blotting

For the Southern Blot approach 10 µg chromosomal DNA were restricted with an appropriate endonuclease restriction enzyme, resulting in fragments ranging from 1 to 5 kb DNA containing the target gene. Following this enzymatic hydrolysis the DNA fragments were separated by horizontal gel electrophoresis. The gel was denatured afterwards for 40 min at RT in DNA solution I and the buffer was changed after 20 min. This was followed by a neutralization step for 40 min in DNA solution II. Meanwhile, a nylon membrane (Hybond N+, pore size: 0.45 µm; GE Healthcare Life Sciences) and six Whatman papers (GE Healthcare Life Sciences) were cut into the appropriate size of the gel. The gel then was placed on the platform of a capillary transfer system, sitting over a reservoir filled with 20 x SSC. The platform was continuously fed with buffer via a long Whatman paper piece, which dipped into the buffer. The buffer was slowly carried over to the platform and passed to the overlying Whatman paper pieces. First the gel and the membrane were placed on top. Subsequently a layer of Whatman papers was applied followed by a stack of paper towels (~5 cm). The platform was then loaded with a weight of about 500 g to ensure the capillary forces to work. The DNA transfer was performed overnight. After blotting the membrane was incubated in 6 x SSC for 5 min and air-dried for 30 min. Afterwards the DNA was fixed on the membrane by UV-light using a UV-crosslinker.

3.2.16.2 Probe labeling, hybridization and detection

The probe labeling was done according to the method described in section 3.2.15.2. The only modification of the protocol was the inclusion of a pre-incubation of the membrane for 15 min at 55 °C. The main hybridization step was also done at 55 °C overnight. The hybridization and detection of the membrane was exactly done as described in section 3.2.15.3.

DNA solution I	NaCl	1.5 M
	NaOH	0.5 M
DNA solution II	NaCl	1.5 M
	Tris	0.5 M

3.3 Construction of *uvrY* allelic variants

One aim of the study was the investigation of the impact of non-synonymous SNPs in *uvrY* on the functionality of the protein. Therefore different allelic variants of *uvrY* were cloned into a homogenous strain background (*E. coli* MG1655). The methods used to construct the *uvrY* variants are described below.

3.3.1 Site-specific recombination by recombineering according to Datsenko & Wanner

Modification of the *E. coli* chromosome or a plasmid was done using linear DNA fragments as described by Datsenko & Wanner (2000). The protocol was modified according to the application required for this work. However, the principle of the method remained the same. The method relies on the replacement of sequences with desired genetic information by means of recombineering (**recombi**nation-mediated engi**neering**) (Ellis *et al.*, 2001). Datsenko and Wanner refer to the chromosomal exchange of an antibiotic marker, but in general any selectable marker or screenable sequence can be replaced. To achieve recombination the sequence to be exchanged is generated by PCR using primers with homology extensions to the flanking regions of the target sequence. The recombination itself is mediated by the bacteriophage λ homologous recombination proteins (γ, β, exo), called Red proteins, that facilitate recombination and protect linear DNA from degradation by the host RecBCD exonuclease V (Figure 9).

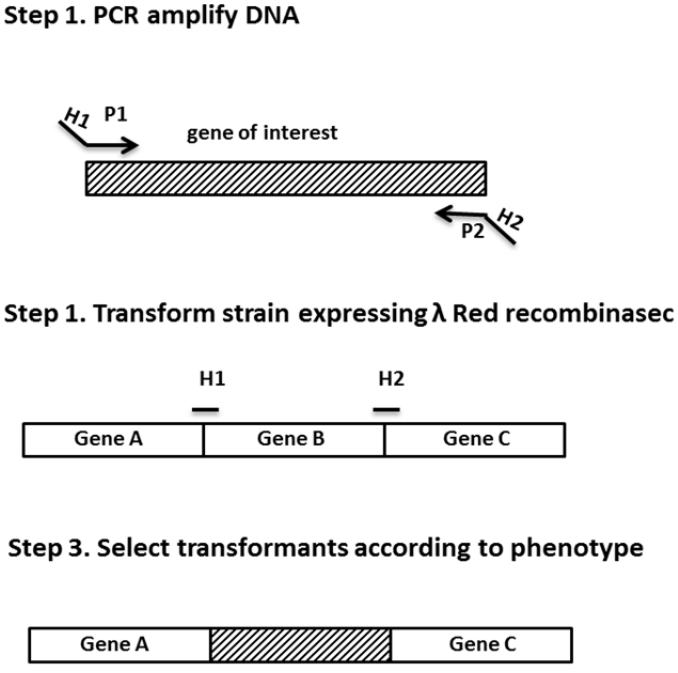


Figure 9: Strategy for inactivation of chromosomal genes using PCR products (modified according to Datsenko & Wanner, 2000).

To undergo homologous recombination the DNA fragment had to be amplified with primers generating the overhanging extensions (~ 50 nt) homologous to the 5´ and 3´ regions of the target chromosomal or plasmid sequence. The PCR product was purified and respective strains were transformed with 100 ng of this PCR product. The strains harbored either the plasmid pKD46, that encoded for the arabinose inducible genetic recombination protein machinery, or encoded chromosomally the recombination proteins, which were expressed from a stably integrated defective λ prophage, such as strain SW102λpir (Warming *et al.*, 2005). The transformants were selected according to their expected phenotype and the presence of the desired genetic alteration was verified by PCR and sequencing.

3.3.2 Genetic modification by a counter-selection system

In order to perform unmarked allelic exchange of desired genetic information within the chromosome of a wild type strain, a counter selection approach was used by means of a two-step process of positive and negative selection. According to Li *et al.* (2013) this is a standard way to enable genetic modification and engineering of bacterial genomes involving homologous recombination methods (see 3.3.1). For this purpose a dual counter-selection system using the chloramphenicol resistance gene *cat* and the *sacB* gene from *Bacillus subtilis* was employed. The expression of *cat* results in resistance of the bacteria to chloramphenicol, a broad spectrum antibiotic inhibiting bacterial protein synthesis. The structural gene *sacB* codes for the exoenzyme levansucrose that converts sucrose to levan, a toxic product for *E. coli*. This way levansucrase activity is lethal to most of Gram-negative bacteria, leading to sucrose sensitivity (Gay *et al.*, 1983; Steinmetz *et al.*, 1983). Thus a selection either for chloramphenicol-resistance or against sucrose sensitivity is possible.

The exchange of different allelic variants of the target gene *uvrY* in the chromosome of the wild type *E. coli* strain MG1655 was carried out in several steps (Figure 10). In the first step the *cat* cassette of pKD3 was amplified using primers Pins2 and Pins3 or cat_in_pCVD422_Fw/Rv (see Table 5), with homologous regions of the vector pCVD422 (Figure 10, step 1). This way the resistance gene could be joined precisely downstream of *sacB* via homologous recombination (see 3.3.1). After integration of *cat* into pCVD422 (pCVD422_cat) the entire *sacB-cat* cassette was amplified via standard PCR (see 3.2.3). The reaction was performed using primers uvrY::sacB-cat_for/rev with~ 50 nt extensions being homologous to the up- and downstream region of *uvrY* (Figure 10, step 2). The respective PCR products were purified (see 3.2.7) and electrocompetent cells of *E. coli* MG1655 carrying the pKD46 helper plasmid (Figure 10, step 3) were transformed with 100 ng PCR product to enable homologous recombination. After electroporation, the mixture was plated on LB-agar containing 25 µg ml^{-1} chloramphenicol to select for chloramphenicol-resistant recombinants. The plates were incubated at

37 °C. An overnight culture was done of several transformants on the next day. To check for successful recombination events, the transformants were tested both for the presence of the *sacB-cat* cassette by PCR and sequencing as well as for sensitivity to sucrose. Therefore the overnight cultures were diluted to 10^{-2} and 10^{-3} and plated on LB-agar containing 5 % (w/v) sucrose but no NaCl. The plates were incubated at 30 °C overnight. If recombination took place the chloramphenicol resistant transformants should have been sensitive to sucrose. Electrocompetent cells were prepared of the intermediate strain MG1655 Δ*uvrY::sacB-cat*. To replace the *sacB-cat* cassette by the desired

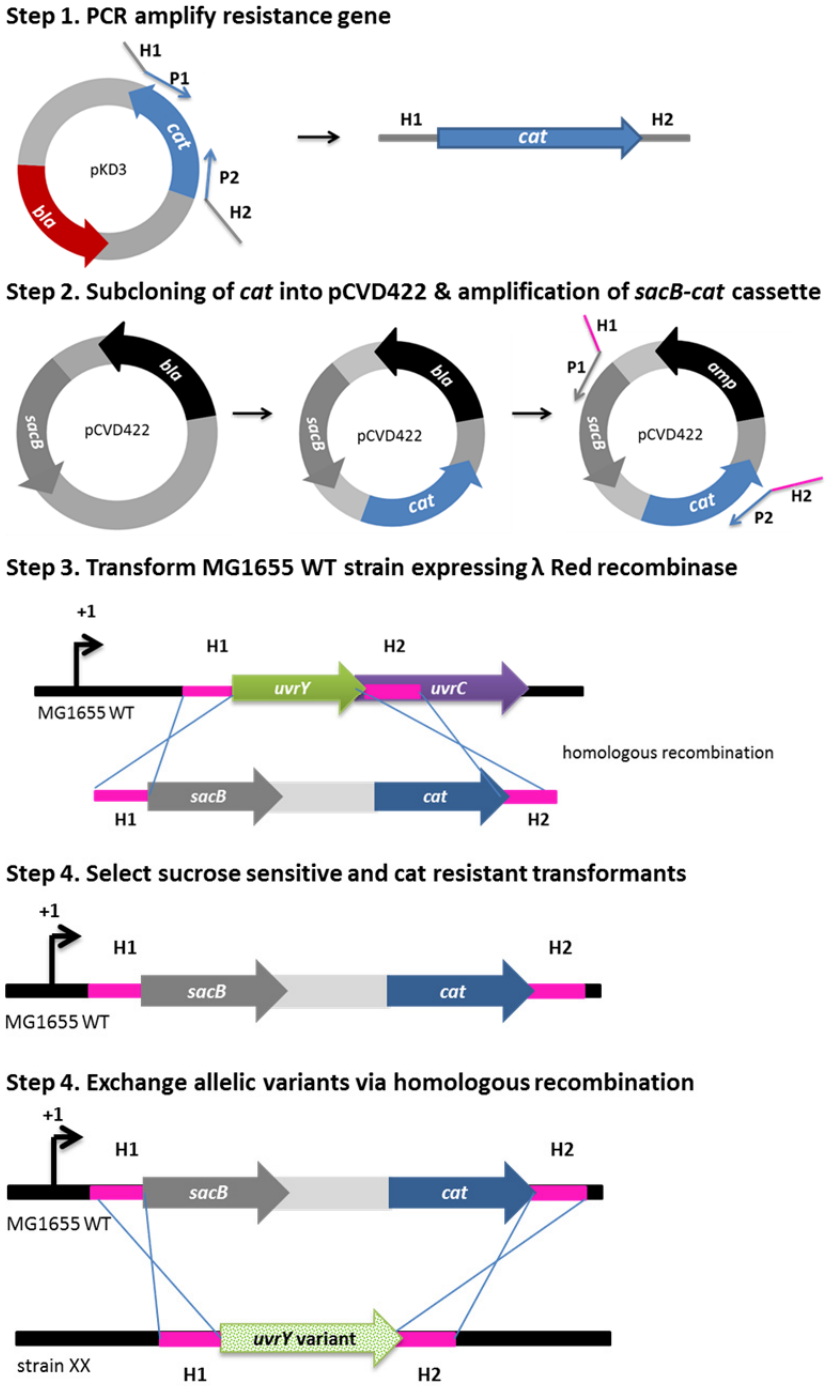


Figure 10: Schematic overview of the allelic exchange of *uvrY* variants into *E. coli* MG1655 background.
H1 and H2 refer to the homology extensions or regions. P1 and P2 refer to priming sites.

uvrY alleles the intermediate strain was transformed again with pKD46 and electrocompetent cells were prepared. The desired *uvrY* allelic variants were amplified with primers containing homologous regions as described before. *E. coli* MG1655 Δ*uvrY::sacB-cat* + pKD46 cells were transformed with the PCR products. The plates were then incubated at 30 °C. The following day the plates were checked for recombinants. The final recombinants should be resistant to sucrose but sensitive to chloramphenicol.

3.3.3 P1 mediated transduction

Bacteriophage-mediated generalized transduction describes the transfer of any number of genetic traits from one bacterial cell (donor) to another (recipient). After infection of a cell bacteriophage P1 is able to package up to 100 kb of bacterial DNA in its capsid. Upon infection of another strain the phage delivers the bacterial DNA into the recipient which will incorporate the bacterial fragments into its chromosome via genetic recombination.

The procedure of transduction is a two-step process. In the first step the donor strain containing the marker to be transferred, needs to be lysed. For this purpose 50 µl of an overnight culture of the donor strain *E. coli* MG1655 *lacZ-zeo* were inoculated in 5 ml LGC and incubated at 37 °C under shaking (180 rpm). After 1 h 0.1 ml of another P1-lysate (*E. coli* CSH50 [provided by M. Berger, Institute of Hygiene, Münster]) was added and growth was continued for approximately 2-3 h until the cultures were completely lysed. After lysis 0.1 ml chloroform was added and the culture was vortexed for 30 sec. The lysate was then transferred into a new 15 ml tube and centrifuged for 10 min at 6000 x g. 0.1 ml of chloroform was again added and the lysed solution was stored at 4 °C until use. The second step involves the transduction of the zeocine cassette into the *E. coli* recipient strain MG1655. Therefore, 2 ml of an overnight culture of the recipient strain was pelleted by centrifugation (5 min, 6000 x g). The pellet was resuspended in 1 ml LGC containing 10 mM $MgSO_4$ and 200 µl were transferred into a new glass tube. Next 0.1 ml P1-lysate was added. After an incubation time of exactly 25 min at RT 100 µl 1 M Na_3Citrate were added. The mixture was then transferred to a new glass tube containing 1 ml LB and incubated for 1 h at 37 °C under shaking (500 rpm). The sample was centrifuged (5 min, 6000 x g) and the supernatant was discarded. After pellet resuspension in 0.1 ml LB containing 10 mM Na_3Citrate, appropriate volumes were plated on selective medium with 50 µg ml^{-1} zeocine. Colonies were re-streaked two times on selective medium to clear P1 completely.

Having the zeocine resistant wild type *E. coli* strain MG1655^{Zeo} enabled us to test the different allelic *uvrY* variants against the wild type strain.

3.4 Molecular biological methods - RNA

3.4.1 RNA Isolation

Total RNA was isolated using the RNeasy Mini Kit (Qiagen). All steps were performed at RT.

3.4.1.1 RNA isolation from urine

The RNA isolation from bacteria grown in urine was done according to the protocol `*Enzymatic Lysis of Bacteria´* with slight modifications (see below). The protocol is recommended for Gram-negative bacteria with short generation times that are grown in minimal media.

For RNA isolation bacteria were grown, depending on the individual mid-log phase of each strain analyzed, in 20 ml fresh pooled human urine without agitation until an optical density OD_{600} between 0.08-0.16 was reached. 5 ml of the urine culture were taken and centrifuged for 5 min at 4 °C and 3000 x g. The supernatant was discarded and the pellet was resuspended in 2 ml PBS/RNA Protect (Qiagen) (1:1). Samples were incubated for 5 min and centrifuged at 5000 x g, 4 °C for 10 min. The pellets were resuspended in 200 µl of lysozyme-containing TE buffer (50 mg ml^{-1}) and incubated at 37 °C for 5 min under shaking (180 rpm). 700 µl of buffer RLT containing β-mercaptoethanol (10 µl ml^{-1}) was added and later complemented with 500 µl ethanol. The subsequent steps of the protocol were consistent with those of the protocol supplied with the RNeasy Mini kit (Qiagen).

Following the isolation procedure the concentration of RNA was determined with the Nanodrop 2000 (see 3.2.2).

3.4.1.2 RNA isolation from LB

For bacteria grown in LB the protocol `*Enzymatic Lysis and Proteinase K Digestion of Bacteria´* was used. This protocol involves enzymatic lysis in combination with proteinase K digestion and is recommended for Gram-negative bacteria grown in complex media.

For RNA isolation, bacteria were grown for 2.5 h at 180 rpm in fresh LB medium until the mid-exponential growth phase was reached. The respective volume, that contained 2.5 $*10^8$ cells at that time, was calculated and centrifuged for 5 min at 4 °C and 3000 x g. The further protocol was exactly followed as described by the manufacturer.

The RNA concentration was measured as described (see 3.2.2).

3.4.2 DNA digestion

Contaminating DNA was removed from total RNA preparations by DNase I digestion. 10-50 µg RNA in a final volume of 44 µl were mixed with 5 µl 10 x incubation buffer and 1 µl DNase I (Applied Science, Penzberg, Germany). Samples were incubated for 1 h at 37 °C. The reaction was stopped by adding

2 µl of 0.2 M EDTA (pH 8.0) and heating to 75 °C for 10 min. Subsequently the RNA was cleaned up again using the RNeasy Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The purified RNA was finally eluted from the column in 25 µl nuclease-free water.
Concentration and sample purity of the resulting RNA was determined using the Nanodrop 2000 (see 3.2.2).
To verify the complete removal of DNA from the RNA samples, a control PCR with primers binding to the coding sequence of *gapA* was performed using 100 ng of DNAse-treated RNA. As a control 100 ng of a genomic DNA sample were used. The DNase digestion was considered successful if it did not result in a product. In the case of positive amplification the RNA was again applied to DNase I digestion.

3.4.3 RNA quality check

A recommended method for the measurement of RNA quality is to check the RNA integrity by automated capillary gel-electrophoresis. The RNA integrity number (RIN), which is determined based on the ratio of ribosomal RNAs and its degradation products, is a powerful tool in RNA integrity measurement and represents a quality label that allows for an objective and standardized assessment of RNA (Schroeder *et al.*, 2006). The RIN value is a numerical number between 1 and 10. Totally degraded RNA is assigned a value of 1, completely intact RNA has a value of the 10 (Fleige & Pfaffl, 2006). The RIN was determined by using the Agilent 2200 Tapestation (Lab9o1, Agilent Technologies) and the Agilent Technologies R6K Screen Tape system according to the supplied protocol. RIN values > 7.0 were considered to meet the requirements for expression analyses via cDNA synthesis (Imbeaud *et al.*, 2005; Schroeder *et al.*, 2006).

3.4.4 Reverse transcription (RT) for cDNA synthesis

The cDNA (complementary DNA) synthesis is the reverse transcription of RNA into cDNA by means of a RNA-dependent DNA polymerase, the so-called reverse transcriptase. Reverse transcriptases are enzymes of retroviruses that are capable of rewriting RNA into the corresponding cDNA sequence.
The cDNA synthesis was done with the Superscript III Reverse Transcription kit provided by Life Technologies. 2 µg of total RNA was mixed with 250 ng random hexamer primer (GE Healthcare Life Sciences) and 1 µl 10 mM dNTP mix (Peqlab, Erlangen, Germany) to a final volume of 13 µl. Primer annealing was carried out an 65 °C for 5 min. After 5 min cooling, the following reverse transcription mixture was added to the samples:

4 µl	5 x First strand buffer (kit component)
1 µl	0.1 M DTT (kit component)

1 µl RNase Out recombinant RNase inhibitor (40 U/µl; Invitrogen)

1 µl Superscript III reverse transcriptase (kit component)

The cDNA synthesis was performed at 52 °C for 60 min, followed by heat inactivation of the transcriptase at 70 °C for 15 min.

3.4.5 Quantitative real-time PCR

The quantitative real-time PCR (qRT-PCR) was used to amplify single genes of interest and simultaneously quantify their gene expression levels. The basic principle of this method is the polymerase chain reaction that amplifies gene transcripts in the presence of the fluorescent dye EvaGreen (BioRad). EvaGreen intercalates into double stranded DNA and emits signals documented by the optical camera within the CFX96 RealTime System (BioRad). The fluorescent molecule therefore reports an increase in amount of DNA with a proportional increase in fluorescent signal. The calculation of transcript levels was automatically computed by means of the BioRad CFX Manager™ software. To assess gene expression levels the threshold cycle (C_t) was taken into consideration. This is the cycle number at which enough amplified product accumulated to yield a detectable fluorescent signal.

The primers designed for the qRT-PCR approach had to meet two main criteria: a product length in the range from 100-250 nucleotides and an annealing temperature between 57 °C – 59 °C. Before starting quantitative real-time PCR experiments the assays had to be optimized according to the BioRad guidelines.

The qRT-PCR was done using 3 ng cDNA per sample each. The reaction mix was prepared for one reaction as follows:

Table 12: qRT-PCR setup.

Component	Final concentration	10 µl rxn
SsoFast EvaGreen	1 x	5 µl
Primer forward (10 µM)	400 nM	0.4 µl
Primer reverse (10 µM)	400 nM	0.4 µl
cDNA template	3 ng	X µl
RNase/DNase free water	N.A.	Ad 10 µl

With the following cycling conditions:

1.	95 °C	30 sec	
2.	95 °C	5 sec	35 x
3.	57 °C	5 sec	
4.	65°C – 95 °C	5 sec	(melting curve)

All PCR reactions were done in triplicates, both biological and technical. Housekeeping genes *gapA (*glyceraldehyde phosphate dehydrogenase) and *frr* (ribosome recycling factor) were carried as internal controls to compare transcript levels from different RNA preparations.

3.5 Phenotypic characterization

3.5.1 Growth kinetics

Bacterial growth was measured in LB medium (180 rpm) and in pooled human urine under static conditions both at 37 °C. If doing growth kinetics in an Erlenmeyer glass flask the optical density (OD) was measured at determined times at a wavelength of 600 nm against the blank value. For kinetics performed in the Tecan Infinite F-200 (Tecan) reader the growth was precisely measured at a wavelength of 595 nm every 18 min in LB or rather 30 min in urine over a period of 24 h.

In order to check if bacterial growth is dependent on different organic compounds growth was measured in urine under addition of various amino acids as well as different carbon sources with a final concentration of 0.1 % (w/v). Additionally growth under addition of ammonium chloride (NH_4Cl) as an inorganic compound with a final concentration of 0.2 % (w/v) was tested. The kinetics were measured as described above in Erlenmeyer glass flasks.

3.5.2 Biofilm assay

To monitor the biofilm formation of bacterial strains, overnight cultures were diluted in fresh LB medium or pooled human urine at a ratio of 1:150 and incubated for 16 h at 37 °C either under shaking (180 rpm) or static conditions. 12.5 µl of the culture were diluted in 125 µl of the desired medium, pipetted into the cavity of a PVC-microtiter plate and incubated statically for 48 h (LB medium) or 72 h (urine) at a growth temperature of 37 °C. In order to remove the planktonic cells the plates were washed three times with 200 µl phosphate buffered saline (PBS) per cavity and dried afterwards for 20 min at 80 °C. 175 µl of 1 % (w/v) crystal-violet solution was added to each well and incubated for 1 h at RT. The staining solution was discarded and the plates were washed three times with dH_2O. To quantify the extent of biofilm formation, the wells were destained with 200 µl destaining solution (80 % (v/v) ethanol, 20 % (v/v) acetone), which was added to each well for 30 min. The content of each well was mixed briefly by pipetting and 100 µl of the solution were transferred into a fresh optically clear flat-bottom 96 well plate. Optical density was measured at a wavelength of 570 nm.

PBS	NaCl	140 mM
	KCl	2.7 mM
	$Na2HPO_4$	48.1 mM
	$KH2PO_4$	41.5 mM
	pH 7.4	
Destaining solution	Ethanol	80 % (v/v)
	Acetone	20 % (v/v)

3.5.3 Motility assay

The swarming motility of different bacterial strains was measured on swarm motility agar plates (see 2.6.2). Therefore bacterial material of a single colony was stabbed into the middle of the agar plate by means of a sterile toothpick. Plates were incubated for 16 h at 37 °C. The extent of motility was assessed by measuring the diameter of the migration zone.

3.5.4 Congo Red staining assay

To prove the expression of curli fimbriae and the production of cellulose bacteria were spotted on Congo Red agar plates (see 2.5) and incubated at RT, 30 °C and 37 °C for 96 h. Red, dry and rough colonies (*rdar* morphotype) indicate expression of curli and cellulose. The *rdar*-positive *E. coli* strain Nissle 1917 served as a control and was loaded on each plate.

3.5.5 Calcoflour assay

The synthesis of cellulose was tested by cultivation of the bacterial strains on calcoflour agar plates (see 2.5). Plates were examined under UV-light (λ = 254 nm). Cellulose-positive strains were identified by the emission of fluorescence.

3.5.6 *In vitro* competition assay

To check the fitness of *E. coli* MG1655 strains with different *uvrY* allelic variants in comparison to *E. coli* MG1655^{Zeo}, carrying a zeocine resistance cassette (wild type), a competition assay was performed. Therefore the different strains were grown individually overnight in LB medium or pooled human urine under shaking (180 rpm) or static conditions. On the next day, overnight cultures of *E. coli* MG1655^{Zeo} and one of the different competitor *uvrY* allelic variant *E. coli* strains were mixed in a ratio of 1:1 to a final volume of 20 ml (urine) or rather 25 ml (LB) with an initial optical density (OD_{600}) of 0.02 in 100-ml Erlenmeyer glass flasks. At determined times (0, 1, 2, 3, 5, 7, 24 h) 100 µl of the mixed cultures were withdrawn and diluted serially in PBS. Appropriate dilution factors were chosen for the respective medium. The serial dilution samples were plated on petri dishes containing LB agar

or LB agar supplemented with 50 µg ml^{-1} zeocine and incubated overnight at 37 °C. On the next day the colony forming units (CFUs) per ml were determined and the CFU ratio between wild type and variant strain was calculated.

3.5.7 Sub-cultivation assay

To determine the long-term effects of the different *uvrY* allelic variants with regard to their competitive behavior, a sub-cultivation assay was performed. Therefore an aliquot of the mixed cultures, which were prepared according to the protocol described in chapter 3.5.6, was transferred from the previous culture into fresh medium (LB or urine) every 24 h. The initial OD_{600} was readjusted to 0.02 every time. The sub-cultivation experiment was done for 96 h in total. To estimate the amount of CFUs/ml 100 µl of the mixed culture were serially diluted every day and plated on (selective) media as described above.

3.6 Protein biochemical methods

3.6.1 Isolation of proteins – whole cell lysate

For the isolation of total cell proteins from bacteria, cells were harvested by centrifugation for 30 min at 6000 x g and 4 °C. For this a culture volume corresponding to an OD_{600} of 20 was used. The supernatant was discarded and the pellet was washed once in resuspension buffer and centrifuged again for 10 min (16.000 x g, 4°C). The pellet was resuspended in 500 µl resuspension buffer and homogenized using the Precellys homogenizer (Peqlab). The homogenization was carried out two times for 20 sec at 6800 rpm. Between the two homogenization steps the tube containing the homogenized cells was placed on ice. After a short centrifugation step at 4 °C to pellet the cell debris, the supernatant of about 200 µl was transferred to a new tube and the total protein concentration was determined (see 3.6.1.1). Afterwards, 10-20 µg of protein were subjected to SDS-PAGE (see 3.6.2) or applied to alternative downstream processes (see 3.6.2 - 3.6.4).

Resuspension buffer	Tris-HCL pH 8.0	100 mM
	Complete Proteinase Inhibitor	1 x
	Phenylmethanesulfonylfluoride (PMSF)	1 Mm

3.6.1.1 BCA quantification of protein concentration

The total protein concentration was determined using the Micro BCA™ Protein Assay Kit (Thermo Scientific) according to the manufacturer´s instructions. For this purpose, protein standards with

defined concentrations of bovine serum albumin (BSA) were prepared to generate a calibration curve. Protein samples were diluted in a ratio of 1:25 in dH_2O and 150 µl sample and standards were pipetted in a 96-well plate. 150 µl working reagent was added to each well and the plate was incubated at 37 °C for 2 h. Afterwards the absorption was measured with a spectrophotometer at a wavelength of 570 nm. The protein concentration of the sample was determined based on the calibration curve.

3.6.2 Tricine-SDS-Page

The discontinuous ´**s**odium **d**odecyl **s**ulfate **p**oly**a**crylamide **g**el **e**lectrophoresis´ (**SDS-PAGE**) described by Laemmli (1970) is used for the electrophoretic separation of proteins according to their molecular weight. A modified version of the Laemmli method is the tricine-SDS-PAGE according to Schägger & von Jagow (1987) that allows a better resolution of small proteins in the range between 5-20 kDa.
Handcast 1.0 mm thick polyacrylamide gels (8 x 7 cm) were run in the Mini-PROTEAN Tetracell system (BioRad) for vertical gel electrophoresis. The gels consisted of, from top to bottom, a stacking gel, a spacer gel and a separation gel. The composition calculated for four polyacrylamide gels is given below (Table 13). Before loading the proteins samples were mixed with 4 x protein loading buffer and boiled at 100 °C for 10 min. The separation of the proteins was started at 30 V for 30 min followed by 100 V for approximately 1 h. The run was performed in a two buffer system consisting of an anode buffer and a cathode buffer. As a size marker the Precision Plus Protein Dual Xtra standards provided by BioRad was used (see 2.9). Afterwards, the gels were stained (see 3.6.3) or subjected to Western blotting (see 3.6.4).

Table 13: Composition of the different gels for Tricine-SDS-PAGE

Component	Separation gel (15 %)	Spacer gel (10 %)	Stacking gel (4 %)
Rotiphorese 30 (Roth)	7.5 ml	2 ml	800 µl
3 x gel buffer	5 ml	2 ml	1.5 ml
50 % (v/v) glycerol	2.5 ml	-	-
dH_2O	-	2 ml	3.7 ml
10 % (w/v) APS	145 µl	50 µl	70 µl
TEMED	25 µl	10 µl	10 µl

Cathode buffer (10 x)	Tris	1 M
	Tricine	1 M
	SDS	1 % (w/v)
Anode buffer (10 x)	Tris	2 M
	pH 8.9	

Gel buffer (3 x)	Tris	3 M
	HCL	1 M
	SDS	0.3 % (w/v)
Protein loading buffer (4 x)	Tris/HCL (pH 8.0)	60 mM
	SDS	10 % (w/v)
	β-mercaptoethanol	5 % (v/v)
	Glycine	25 % (v/v)
	Bromphenol blue	0.25 % (w/v)

3.6.3 Coomassie staining

To visualize the protein bands after separation by SDS-PAGE, the gels were fixed for 10 min in fixation solution and subsequently stained for 4 h in Coomassie staining solution. Afterwards the gels were washed in destaining solution for several times, wrapped in a plastic sleeve and documented in the ChemiDoc™ MP System (BioRad).

Staining solution	Fixation solution	1 x
	Coomassie brilliant blue R250	0.25 % (w/v)
Fixation solution	dH_2O	40 % (v/v)
	Acetic acid	10 % (v/v)
	Methanol	50 % (v/v)
Destaining solution	dH_2O	67.5 % (v/v)
	Acetic acid	7.5 % (v/v)
	Methanol	25 % (v/v)

3.6.4 Western blot

Western blot is the transfer of proteins onto a carrier membrane after SDS-PAGE (Burnette, 1981). This way immobilized proteins are accessible for immunological interaction analysis with specific antibodies.

3.6.4.1 Transfer of proteins - `semi-dry blot`

The transfer of proteins was carried out using the semi-dry method (Kyhse-Andersen, 1984). By means of this method the proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane (Roth). For this purpose the membrane was activated in methanol for 30 sec and briefly equilibrated in transfer buffer. Subsequently, the membrane was placed on a stack of three Whatman papers, previously soaked in transfer buffer, in a semi-dry blotter device (Roth). Attention had to be taken that there were not any air bubbles between the individual layers to ensure a proper transfer of the proteins. The gel, which was also equilibrated in transfer buffer, was placed on top

the membrane followed by another stack of three transfer buffer soaked Whatman papers (Figure 11). The protein transfer was carried out at 1.5 mA/cm^2 for 68 min.

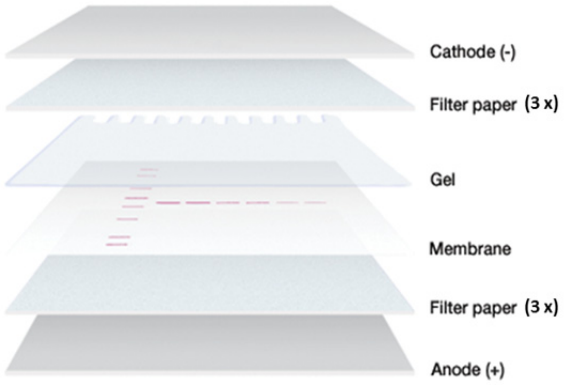


Figure 11: Schematic overview of a semi-dry blot application (modified according to BioRad).

Transfer buffer (1 x)	Tris	25 mM
	Glycine	192 mM
	Methanol	20 % (v/v)

3.6.4.2 Immune detection

Following the transfer of the proteins the membrane was initially blocked for 1 h at RT or overnight at 4 °C in blocking buffer to saturate unspecific binding sites. After blocking, the membrane was washed three times for 5 min in TBS-T and incubated with the respective primary antibody for 1 h under rotation. The primary antibody was diluted in TBS-T with a dilution factor of 1:10.000 according to the suppliers´ recommendations. Before addition of the secondary antibody the membrane was washed the Horseradish Peroxidase (HRP) - conjugated anti-rabbit IgG + IgM (Dianova) was used in a dilution of 1:16.000. The incubation time was 1 h under shaking at RT. Afterwards the membrane was washed again three times in TBS-T for 5 min.

TBS	Tris	20 mM
	NaCl	150 mM
	pH 7.6	
TBS-T	TBS	
	Tween-20	0.05 % (v/v)
Blocking buffer	TBS	
	Skim milk powder	5 % (w/v)

3.6.4.3 ECL detection

The detection of the secondary HRP-bound antibody was carried out by the generation of chemiluminescence using the Clarity^{TM} Western ECL substrate from BioRad. For this purpose, equal volumes of substrate kit components 1 and 2 were mixed and added to the protein side of the

membrane for about 5 min (0.1 ml solution /cm^2 of membrane). After incubation at RT, the liquid was removed and the membrane was placed free of air bubbles between two transparent plastic sheets. The emitted chemiluminescence, which resulted from the oxidation of luminol by hydrogen peroxide catalyzed by the antibody bound HRP, was detected using the ChemiDoc™ MP System (BioRad) containing a light-sensitive CCD camera system.

3.7 Metabolomics

Metabolomics are a very new and emergent technology that deals with the investigation of metabolites that occur as products of metabolic processes in organisms. Thus, metabolomics is a powerful tool to gain insights in the physiological state of a cell in response to different environmental conditions. Since one aim of the study was to investigate differences between the metabolic footprints of the *E. coli* strains 83972, CFT073 and Nissle1917 the extracellular metabolites were analyzed in pooled human urine. The exact procedure is described below.

3.7.1 Metabolome sample preparation

To analyze the extracellular metabolome 400 ml of fresh, sterile filtered pooled human urine were inoculated in a 1 l Erlenmeyer glass flask with an *E. coli* overnight culture in a ratio of 1:100. The culture was grown at 37 °C under static conditions. Samples of the culture were obtained at the following time points: 0, 1, 2, 3, 4, 5, 6, 7, 8, 24 h. Therefore 10 ml culture volume was withdrawn at each time point with a sterile, pre-cooled glass pipette. To prevent any changes in the metabolite levels the samples were sterile filtered immediately after removal by using a Millex-GP Syringe Filter Unit (Millipore). After filtration the samples were frozen on dry ice.

Further analysis of the extracellular metabolome was kindly conducted by Kirsten Dörries in the laboratory of Dr. Michael Lalk (Institute of Biochemistry, Section Metabolomics, University of Greifswald, Germany). The methods used for the analysis of the metabolic footprints were the **G**as **c**hromatography–**m**ass **s**pectrometry (GC-MS) and the **P**roton **N**uclear **M**agnetic **R**esonance (^{1}H-NMR) spectroscopy. GC-MS is an analytical method that combines the separation of chemical mixtures into pure chemicals by means of gas chromatography (GC) with the identification and quantification of those chemicals by means of mass spectrometry (MS). In contrast, the ^{1}H-NMR spectroscopy is used to determine the structure of organic compounds. The combination of both methods allows the successful determination of the metabolome.

3.8 *In silico* analysis

Standard comparison of nucleotide and protein sequences were performed using the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) of the National Centre for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The alignment of nucleotide and proteins sequences was performed with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/), Ridom SeqSphere (http://www.ridom.de/ seqsphere /index.shtml) or CLC Bio (http://www.clcbio.com/). The CLC Bio software was also used for the graphical visualization of nucleotide sequences. As further sources of information on nucleotide and protein sequences the databases KEGG (Ogata *et al.*, 1999; http://www.genome.jp/kegg/), EcoCyc (Keseler *et al.*, 2005; http://ecocyc.org/), Colibri (http://genolist.pasteur.fr/Colibri/) and xBase (Chaudhuri *et al.*, 2008; http://www.xbase.ac.uk/colibase/) were employed. The selection of primers used for PCR and sequencing reactions was done with the freely available program Primer 3 (Rozen & Skaletsky, 2000). Melting temperatures were calculated by means of the Biomath Calculator (Rychlik & Rhoads, 1989). Restriction sites for DNA sequences were investigated with the NEBcutter V2.0 (Vincze *et al.*, 2003). The analysis of nucleotide polymorphism from aligned DNA sequence data was done with DNASp5 V5.10.1 (Rozas & Rozas, 1995; http://www.ub.edu/dnasp/).

3.9 Statistics

The statistical analysis of the data was performed with the GraphPad Prism 5 software (GraphPad Software, California, USA) and the freely available statistic software R (http://www.r-project.org/).

4 Results

I. Urine – a challenging niche for bacteria

According to several studies the urinary tract is a challenging environment forcing bacteria to adapt to this special niche (Hull & Hull, 1997; Snyder *et al.*, 2004; Zdziarski *et al.*, 2010). Although many studies focused on genome or transcriptome data from *E. coli* strains grown under different conditions ranging from rich medium to human urine (Hancock *et al.*, 2008; Snyder *et al.*, 2004; Zdziarski *et al.*, 2010) less is known about the metabolome from *E. coli* strains living in the urinary tract. The aim of the study was the analysis of the extracellular metabolome of different bacterial strains grown in human urine. For this purpose a comparative approach of the extracellular metabolome of the extra-intestinal *E. coli* strain 83972, the gut commensal strain *E. coli* Nissle 1917 and the highly virulent uropathogenic strain CFT073 was performed upon growth in pooled human urine. A comparative metabolome assay should clarify if metabolic traits differ between these strains that exhibit only minor genetic variations (Hancock *et al.*, 2008; Vejborg *et al.*, 2010).

4.1 Extracellular metabolome

The time-resolved uptake and secretion pattern of different extracellular metabolites of the three strains is shown in Figure 12 in form of a color coded chart $\log_2$ (x). A total amount of 44 different metabolites were quantified in a time-dependent fashion by GC-MS and ^{1}H-NMR spectroscopy whereby these metabolites belong to different groups of substrates. The following substrate groups could be identified: sugars, sugar acids, sugar alcohols, purine/pyrmidine bases, nucleosides, amino acids, organic acids and other metabolites. The heat map (Figure 12) clearly illustrates that a vast majority of the measured metabolites were taken up or secreted at a comparable level by all the strains at the same time points. However, marked differences could be observed for some metabolites, namely cadaverine, formate and the amino acids tryptophan, lysine and glutamic acid (Figure 12; marked with a red arrow).

The measured extracellular metabolome in general provided data that can be coupled to the metabolic state of bacteria during growth in urine. A lot of extracellular metabolites were quantified that could be connected to various central metabolic pathways which *E. coli* can use to generate energy and metabolic precursors to be used as a starting point for diverse biosynthetic reactions (Romano & Conway, 1996). An overview of central metabolic pathways that could be linked with selected measured extracellular metabolites is shown in Figure 13. One of the most important metabolic pathways is the glycolysis, also known as Embden-Meyerhof-Parnas (EMP) pathway that

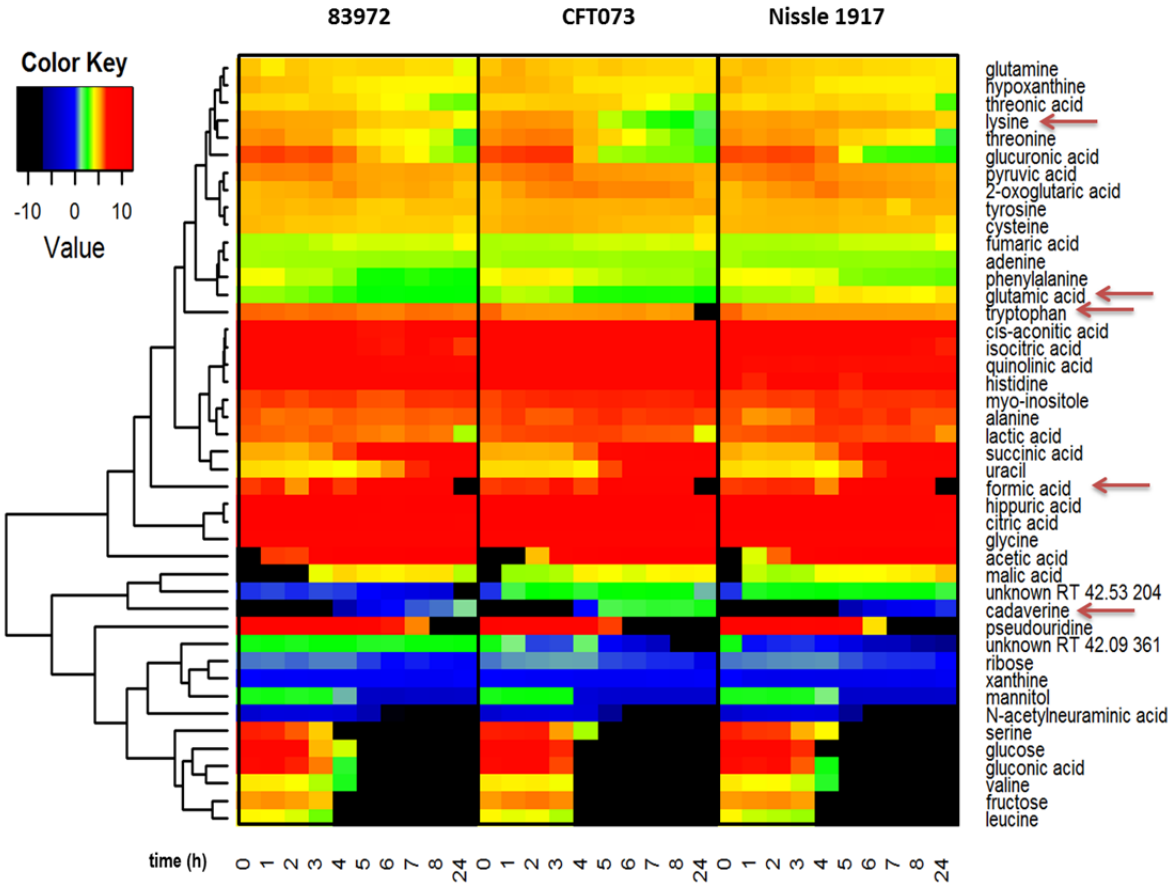


Figure 12: Strain-specific changes of extracellular metabolites in pooled human urine upon growth of *E. coli* strains Nissle 1917, CFT073 and 83972.

Hierarchical clustering of different uptake or secretion patterns during growth in pooled human urine at different time points (0 – 24 h). Uptake or secretion patterns are presented according to the color key (log 2-fold, mean values of three experiments). Complete substrate consumption is indicated by black color.

converts glucose into pyruvate. The kinetics for the glucose levels during growth in urine indicate that all of the strains catabolized the available glucose during exponential growth (Figure 14 A). However, with the transition into the stationary phase after about four hours the glucose levels decreased dramatically resulting in a complete depletion of the medium. The strain-specific growth kinetics are shown in Figure 14 L.

Another measured metabolite was pyruvate, a degradation product of the glycolysis (Figure 13; Figure 14 B). Pyruvate can also be found as a carbohydrate breakdown product in the alternative Enter-Doudoroff pathway (Entner & Doudoroff, 1952). Further it serves as an intermediate for the mixed acid fermentations where acetate, lactate or formate are generated. The kinetics for these metabolites are shown in Figure 14 C-E. An accumulation of pyruvate (Figure 14 B) can be observed during exponential growth for all strains while concentrations in urine decreased after three hours suggesting an augmented uptake again. In contrast, acetate (Figure 14 C) started to accumulate in the medium with the transition into the stationary phase after four hours. The results for formate and lactate are shown in Figure 14 D /E. The formate concentration (Figure 14 D) showed a pyramid-

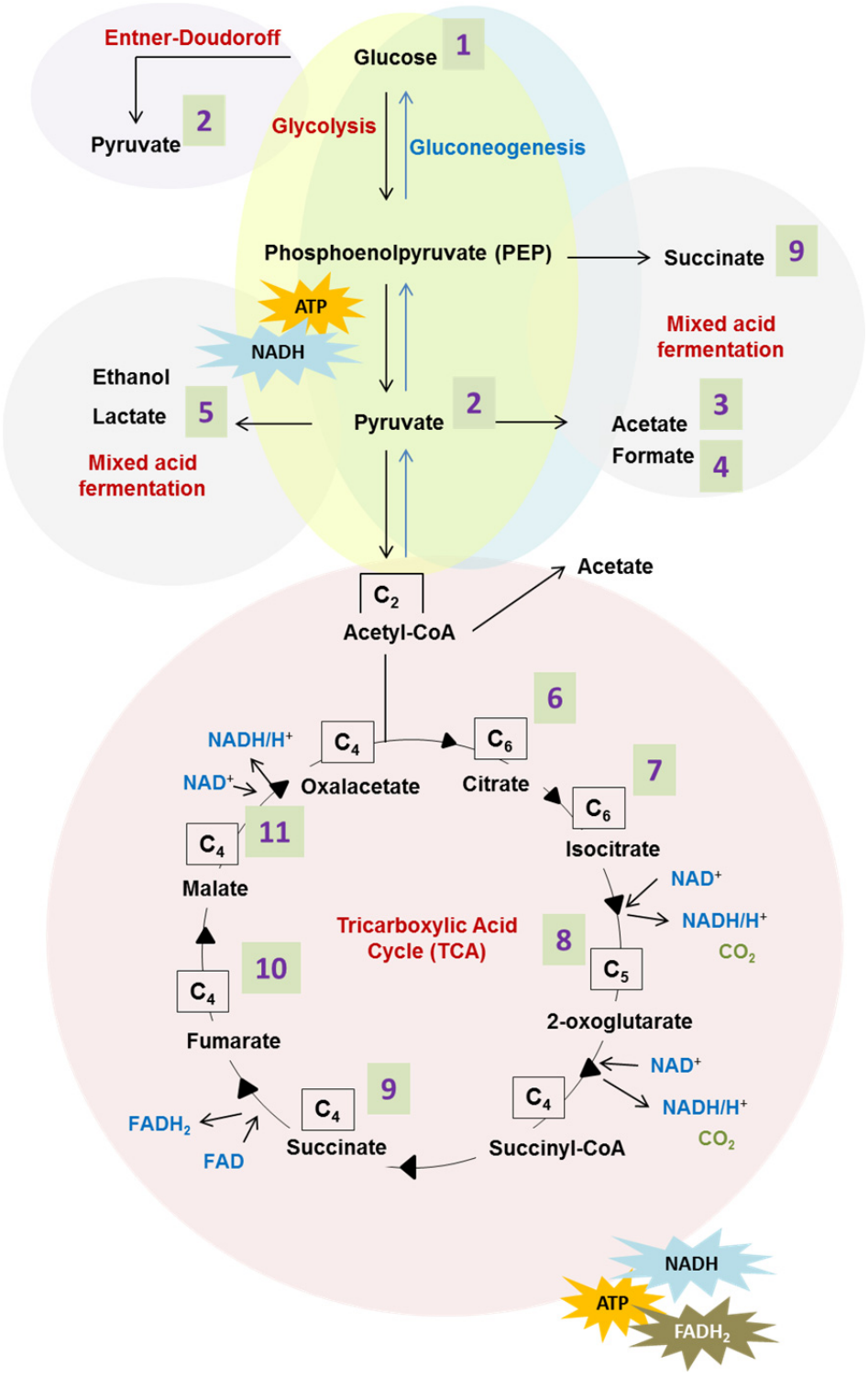


Figure 13: Connection of central metabolic pathways in *E. coli.*

The central metabolism of *E. coli* can be divided in different pathways like the glycolysis, gluconeogenesis, mixed acid fermentation or the tricarboxylic acid cycle (TCA). All of the pathways provide important precursors for the biosynthesis of cellular components and in particular they produce energy (ATP/NADH/$FADH_2$) for other energy-consuming processes. Green shaded numbers designate metabolites analyzed in Figure 14.

like kinetic for all strains. The accumulation of this metabolite which started after four hours was followed by a dramatic decline in the stationary phase. As already mentioned above, strain-specific

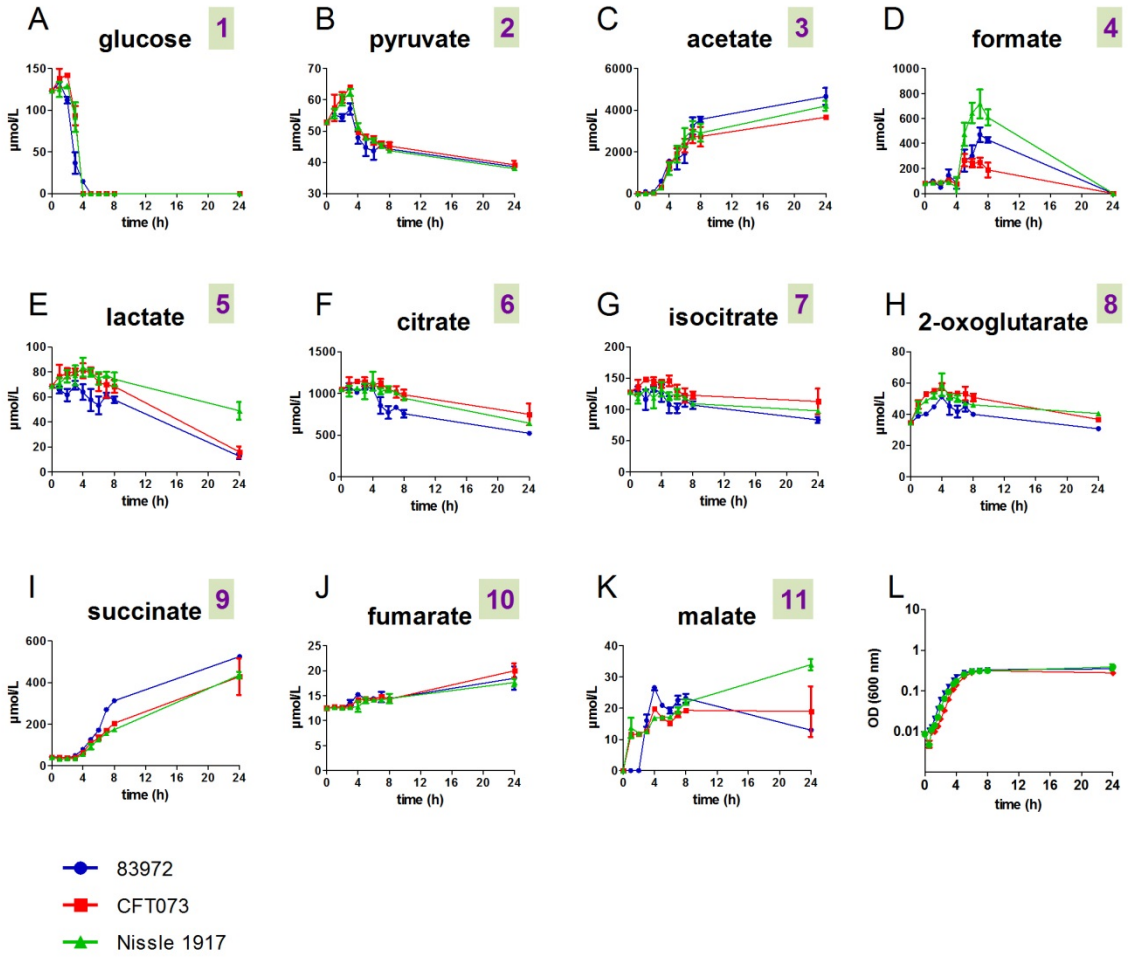


Figure 14: Concentrations of extracellular metabolites in pooled human urine measured during growth of *E. coli* strains 83972, CFT073 and Nissle 1917 for 24 h.

Extracellular metabolites are plotted in a time-resolved fashion according to their concentrations. A connection to central metabolic pathways is indicated by the green shaded numbers (1-11) as shown in Figure 13. (L) Growth kinetics for the *E. coli* strains during growth in urine.

differences regarding the kinetics for formate synthesis and - uptake could be identified whereby *E. coli* Nissle 1917 showed the highest secretion and uptake pattern. In contrast *E. coli* CFT073 showed the lowest secretion and uptake. Completely different kinetics were observed for lactate (Figure 14 E). The lactate concentration constantly decreased after eight hours for the three strains.

The substrates involved in the tricarboxylic acid cycle (TCA), the major energy generating pathway (Busby *et al.*, 1998) are shown in Figure 14 (F-K). The metabolites citrate (Figure 14 F), isocitrate (Figure 14 G) and 2-oxoglutarate (Figure 14 H) showed a time-dependent slight decrease while an increase was observed for succinate (Figure 14 I), fumarate (Figure 14 J) and malate (Figure 14 K). This increase was most pronounced for succinate. Despite the changes of the kinetics for the single metabolites only a few non-significant alterations could be observed comparing the different kinetics for a selected metabolite between the strains.

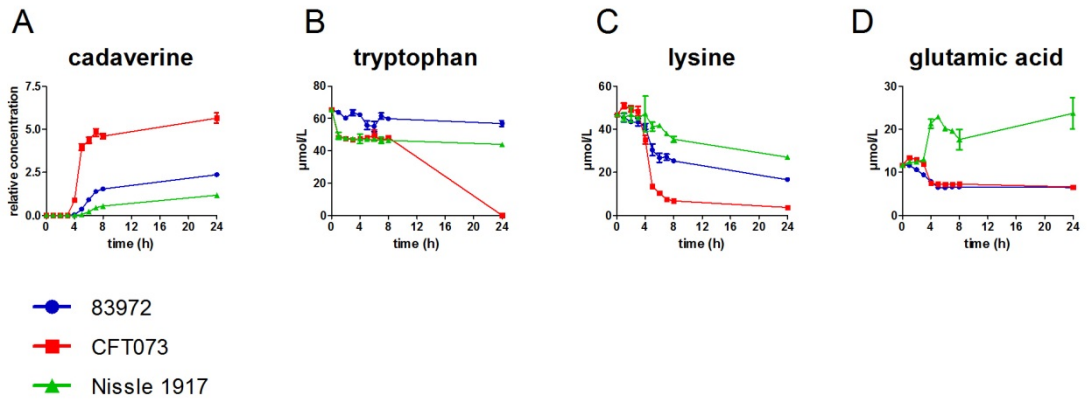


Figure 15: Kinetics of extracellular metabolites over a time period of 24 h.

Time-resolved uptake and secretion pattern are shown for (A) cadaverine, (B) tryptophan, (C) lysine and (D) glutamic acid. Values for cadaverine are given as relative concentrations referred to internal standards. The other metabolites are given in absolute concentrations (µmol/L). Error bars represent the standard deviation (SD) of three biological replicates.

However, conspicuous changes, as already described above, were also identified for the measured extracellular metabolites cadaverine and the amino acids tryptophan, lysine and glutamic acid between the strains as shown in Figure 15. The cadaverine concentrations (Figure 15 A) were quite different for the strains. While the UPEC strain CFT073 accumulated the highest amount of cadaverine after transition into the stationary phase a relative slight increase of the cadaverine levels could be observed for *E. coli* 83972 and *E. coli* Nissle 1917.

Marked inconsistencies between the strains regarding the uptake or production of metabolites were also identified for some amino acids including tryptophan. Although *E. coli* Nissle 1917 showed an increased uptake of tryptophan in the first hour of growth, the concentration remained nearly constantly over time for both *E. coli* Nissle 1917 and *E. coli* 83972 (Figure 15 B). In contrast the values markedly decreased after eight hours for UPEC strain CFT073 up to zero after 24 hours. A further difference could be observed for the amino acid lysine (Figure 15 C), which is also assigned to the group of proteinogenic amino acids. This amino acid seemed to be taken up most efficiently by *E. coli* CFT073, followed by *E. coli* 83972 and *E. coli* Nissle 1917 which showed the lowest uptake of the three examined strains. Interestingly, the decrease of lysine proceeded in a concentration-dependent manner in parallel to the increase of cadaverine for all *E. coli* strains. The situation was rather different for glutamic acid (Figure 15 D), another proteinogenic amino acid. Glutamic acid was secreted by *E. coli* Nissle 1917 in an increasing manner whilst its concentration decreased similar with time in *E. coli* 83972 and *E. coli* CFT073.

4.2 Limitations for growth in urine

Data from transcriptome analysis of *E. coli* CFT073 suggested that *E. coli* mainly uses peptides as a main carbon source for growth in the urinary tract (Alteri *et al.*, 2009). Therefore the TCA cycle and the gluconeogenesis are required for optimal growth during UTI.

To investigate if the bacteria suffer from carbon or nitrogen limitation during growth in urine we measured the growth kinetics in the presence of the 20 amino acids and three different carbon sources, namely glucose, fructose and ribose. In contrast to the tested carbohydrates the degradation products of the amino acids could serve (i) as a carbon and (ii) as a nitrogen source. Therefore we additionally tested growth in urine supplemented with an inorganic nitrogen source (ammonium chloride (NH_4Cl)) to test if growth in urine can also be promoted by an inorganic-nitrogen source. The growth curves of the three *E. coli* strains in pooled human urine complemented with selected amino acids, whose uptake and secretion pattern has been shown to be altered by extracellular metabolome analysis (Figure 15 B/C/D), glucose and NH_4Cl are shown in Figure 16 A-C. The kinetics for the strains grown in urine supplemented with the respective substrates are indicated with a plus (+). In general, we could not observe any improvement of the growth kinetics for the three tested *E. coli* strains grown under addition of various amino acids as shown for tryptophan (Figure 16 A) and lysine (Figure 16 B) if compared with the strains grown under normal conditions. However, for glutamic acid (Figure 16 C), a difference could be observed between strains grown in urine with (+) and without addition of this amino acid. Strains that were grown without glutamic acid exhibited higher growth rates and reached higher cell densities than the strains grown under addition of this amino acid.

In contrast to growth in urine upon the addition of amino acids the results for the addition of sugars as a carbon source, here exemplarily shown for glucose (Figure 16 D), indicated that the availability of carbon seems to be a limiting factor for growth in urine. This can be concluded because all of the tested strains showed higher growth rates and markedly higher final ODs upon the addition of the different sugars. These results are in accordance with the findings from the metabolome analysis which showed that glucose and other sugars were depleted from the medium very quickly, suggesting that they are preferred carbon sources in urine. Since amino acids could be used as a carbon and a nitrogen source, an inorganic nitrogen source, namely ammonium chloride (NH_4Cl) was added to check if nitrogen is a limiting factor for growth in urine, as well. The results for growth upon addition of NH_4Cl are shown in Figure 16 E. Obviously the addition of nitrogen did not alter the growth kinetics of the strains. Therefore it can be concluded that carbon, but not nitrogen is the main limiting factor for growth in urine.

Independent from the addition of different substrates we could always observe the highest growth rates for *E. coli* 83972 indicating that this strain is well-adapted to growth in urine.

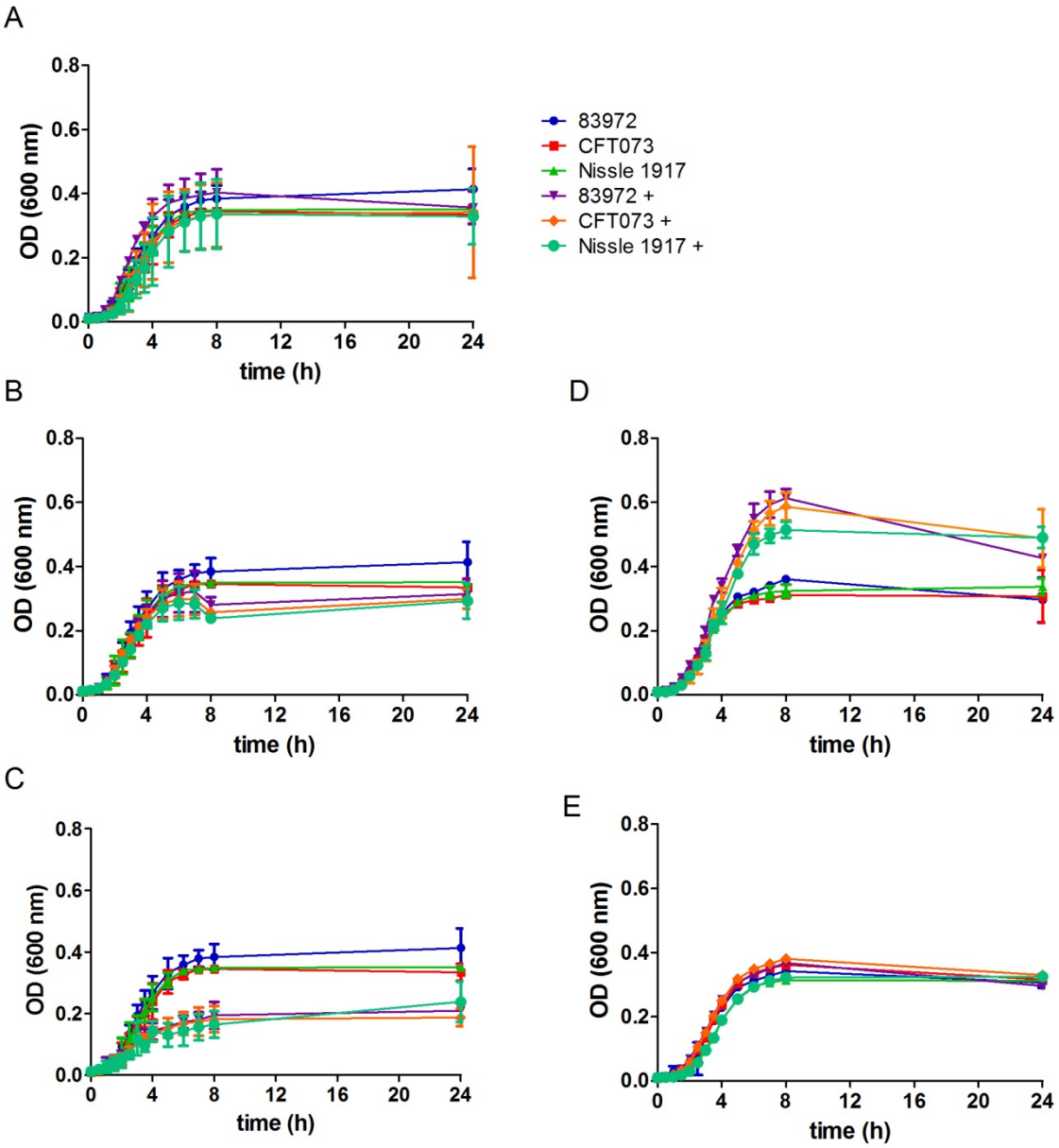


Figure 16: Growth kinetics of *E. coli* strains 83972, CFT073 and Nissle 1917 in pooled human urine supplemented with different carbon and nitrogen sources at 37 °C.

(A) Tryptophan, (B) lysine and (C) glutamic acid were added to a final concentration of 0.1 % (w/v). (D) Growth under addition of glucose (0.1 % (w/v)). (E) NH_4Cl was added to test for nitrogen utilization in a final concentration of 0.2 % (w/v).

4.3 Stress response to growth in urine

E. coli has to face a lot of challenging fluctuations during UTI or growth in urine. Especially with the transition into the stationary phase the stress increases due to nutrient depletion and suboptimal external factors like osmotic or oxidative stress or variations in osmolarity and pH (Babior, 2000; Battesti *et al.*, 2011). Human urine which is known to be of high osmolarity, moderately oxygenated

as well as iron- and nutrient limited and should therefore constitute a highly stressful growth medium for bacteria (Aubron *et al.*, 2012; Gordon & Riley, 1992; Hull & Hull, 1997; Snyder *et al.*, 2004). To deal with these stressful conditions the alternative sigma factor RpoS, a major regulator of central stress response, is up-regulated during stationary phase (Dong *et al.*, 2008; Jishage & Ishihama, 1995; Lange & Hengge-Aronis, 1994). Due to its important role as a regulator for about 400 genes (Dong *et al.*, 2008; Patten *et al.*, 2004; Schellhorn *et al.*, 1998) that are required for adaptation to stationary phase we wanted to test, if the levels of RpoS expression differ between growth in pooled human urine and LB. For this purpose a Western blot approach was chosen to identify the expression levels on a time scale from one to 24 hours of growth. The results for the immuno detection are exemplarily shown for *E. coli* 83972 in Figure 17 A. As expected, the level of RpoS which has a molecular weight of 37.8 kDa increased with prolonged growth in LB. The signal reached its highest intensity after 24 hours of growth. The corresponding Coomassie stained SDS gel is shown in Figure 17 B excluding that alterations on the immuno blot were due to different amounts of total protein on the gel.

Surprisingly the expression level of RpoS differed in pooled human urine. Although comparable to LB an increase of the signal was first detected after five hours of growth, the signal intensity decreased again afterwards. The expression level after 24 hours resembled the ones that were measured after three hours of growth. The corresponding Coomassie gel is shown for the urine samples (Figure 17 B) confirming that the same amount of total protein was loaded for each time point on the SDS gel.

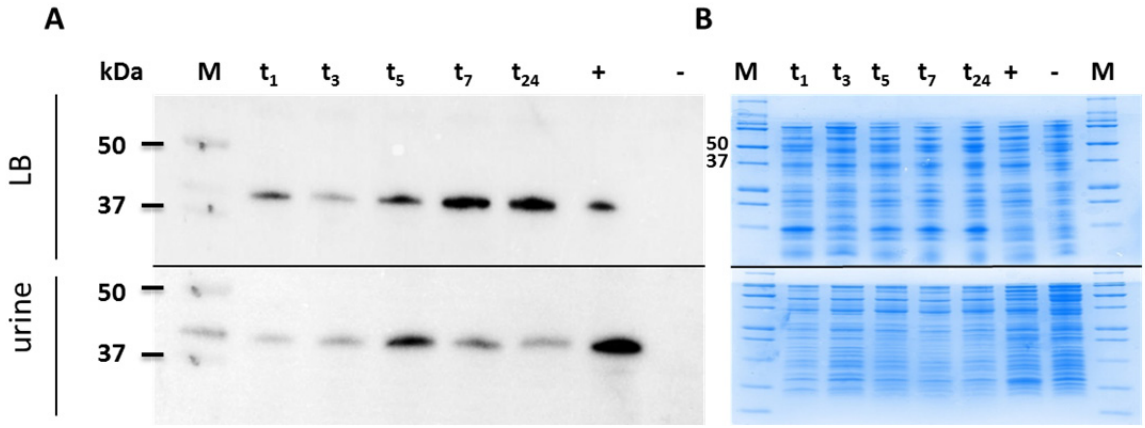


Figure 17: Time-resolved RpoS expression of *E. coli* 83972 detected via Western blot.

Immuno blotting was performed with bacterial samples obtained at different time points during growth in (A) LB and pooled human urine. The corresponding Coomassie stained SDS gels are shown in (B). 15 µg total protein were loaded per lane. UPEC strain 536 and 536Δ*rpos* were used as positive and negative controls, respectively.

Finally, we could observe different switches of RpoS expression during growth in LB and urine. While an increase of RpoS expression in LB appeared after five hours with a continuous increase

afterwards, we observed a dramatic increase of RpoS levels until time point 5 and a decrease afterwards again. The data suggest that different regulatory processes may be involved dependent on the growth medium or changing environmental conditions. A major regulatory system, which has also been found to be frequently mutated upon prolonged growth in the bladder, is the TCS BarA/UvrY (Zdziarski *et al.*, 2010). This TCS is an important regulator for carbon metabolism. Due to the fact that the data of either the extracellular metabolome analysis or the growth kinetics revealed that (i) *E. coli* has to face stressful conditions during growth in urine and that (ii) carbon, one of the major bioelements required for organic compounds (Gottschalk, 1988), seemed to be the limiting factor for growth in urine we focused in the course of the further study on the detailed investigation of this TCS.

II. BarA/UvrY TCS – a key factor for the adaptation to growth in urine

As shown in previous studies the BarA/UvrY TCS is an important regulator being crucial for the efficient coordination of central metabolic pathways (Baker *et al.*, 2002; Pernestig *et al.*, 2003; Romeo *et al.*, 1993). It plays an important role in controlling regulatory networks affecting a multitude of cellular functions, e.g. regulation of carbon metabolism, biofilm formation or motility and adhesion (Jackson *et al.*, 2002; Pernestig *et al.*, 2003; Suzuki *et al.*, 2002; Wei *et al.*, 2001). Previous re-sequencing studies of ABU strains after therapeutic bladder colonization revealed that the BarA/UvrY TCS frequently accumulates mutations upon prolonged growth in the bladder (Zdziarski *et al.*, 2010). Mutations in this TCS could therefore constitute a general adaptation strategy to the urinary tract of humans which led us to the hypothesis that the TCS might represent a mutational hot spot and that the *barA* and *uvrY genes* are under positive selection pressure during long-term growth in the urinary bladder. In order to prove this hypothesis the sequence variation in different *E. coli* isolates from different environments was compared.

4.4 Sequence variation in different groups of *E. coli* isolates

4.4.1 Amplification and sequencing of *barA/uvrY*

To test the hypothesis that *barA* and *uvrY* undergo adaptive evolution upon growth in the urinary tract, we have amplified both genes by PCR and sequenced the PCR products subsequently by Sanger sequencing in a strain panel initially comprising 213 *E. coli* strains. The tested strains are a collection from humans with ABU, symptomatic UTI, or health-care associated ABU (HA-ABU) as well as isolates from the feces of healthy individuals (FI) that were used as a commensal control group (Table 14).

Table 14: Strain collection at the beginning of the study.

ABU (n=85)	HA-ABU (n=34)	UTI (n=55)	FI (n=39)
# 255 - 256	# 1291 - 1321	IMI 917 - 924, 926, 928, 929	# 2038 - 2076
# 1963 - 1972	# 1532 - 1534	# 358, 516, 587, 960	
# 1974 - 2024		# 1336 - 1354, 1517 – 1531	
# 2026 - 2037		# 1535 - 1540	
IMI 961 - 970			

The identification of the strain-specific *barA* and *uvrY* alleles was carried out according to a defined working plan that is shown in Figure 18. All alleles that could not be identified on the basis of this scheme were excluded from the study.

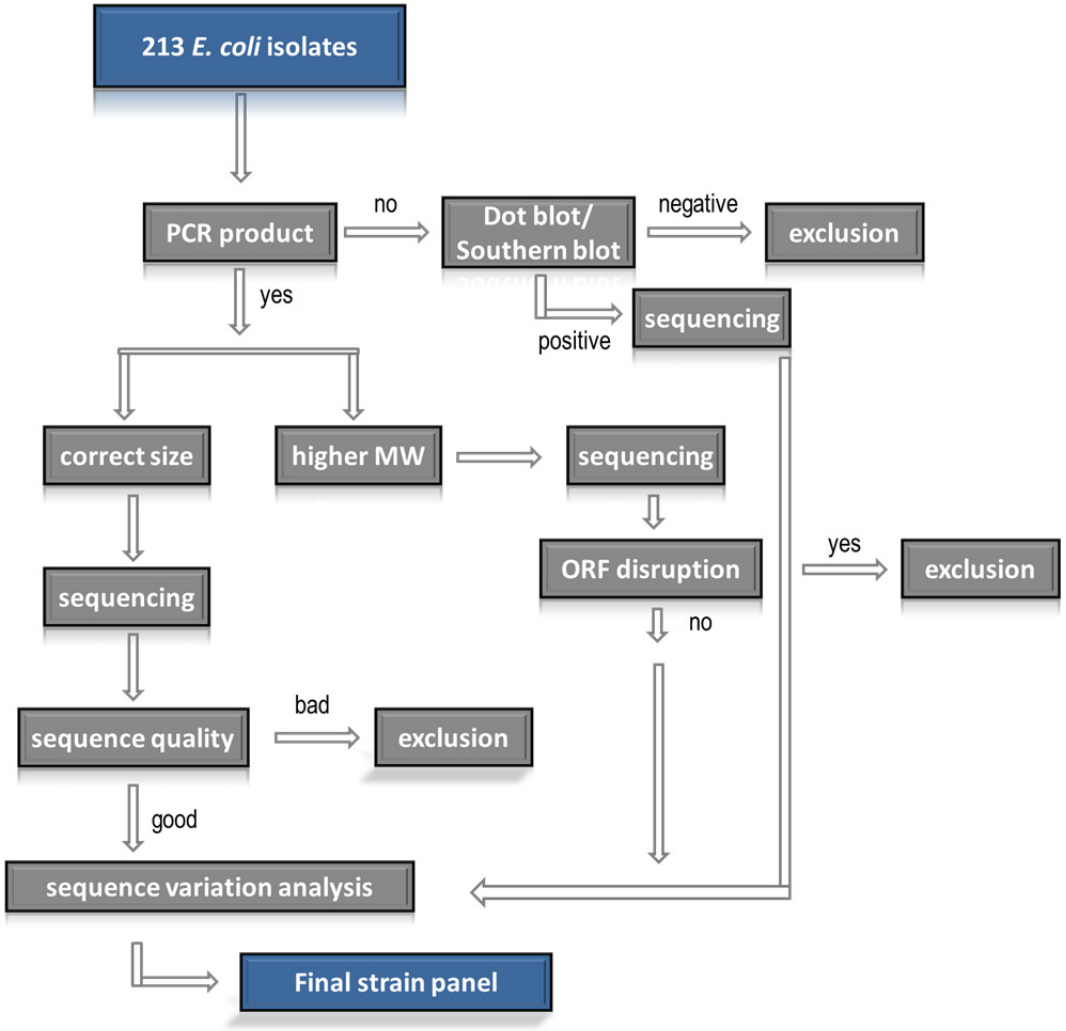


Figure 18: Flowchart for the identification process of strain specific *barA*/ *uvrY* alleles.

After optimization of the PCR conditions the amplification of *barA* could be successfully performed for 211 of the 213 *E. coli* strains. Since the amplification was unsuccessful for two fecal samples (# 2061, # 2062), these were examined according to the working plan by Dot blot and Southern blot. The results for the Dot blot are shown in Figure 19 A. The signal strength appeared to be the same for all samples suggesting that (i) the hybridization conditions were not stringent enough or (ii) that the probe was unspecific. According to the working plan mentioned above, these samples were excluded from the study.

Nine of the 211 *barA* amplicons showed specific products of higher molecular weight compared to the positive control on the agarose gel (2900 bp) as exemplarily shown in Figure 19 B. Four of those nine amplicons were assigned to strains belonging to the ABU group and five to fecal isolates. The *barA* sequence alignment against the reference strain *E. coli* MG1655 (NC_00913) revealed that various insertion sequence (IS) elements (IS1, IS3, IS4) were introduced in *barA* in eight of nine cases.

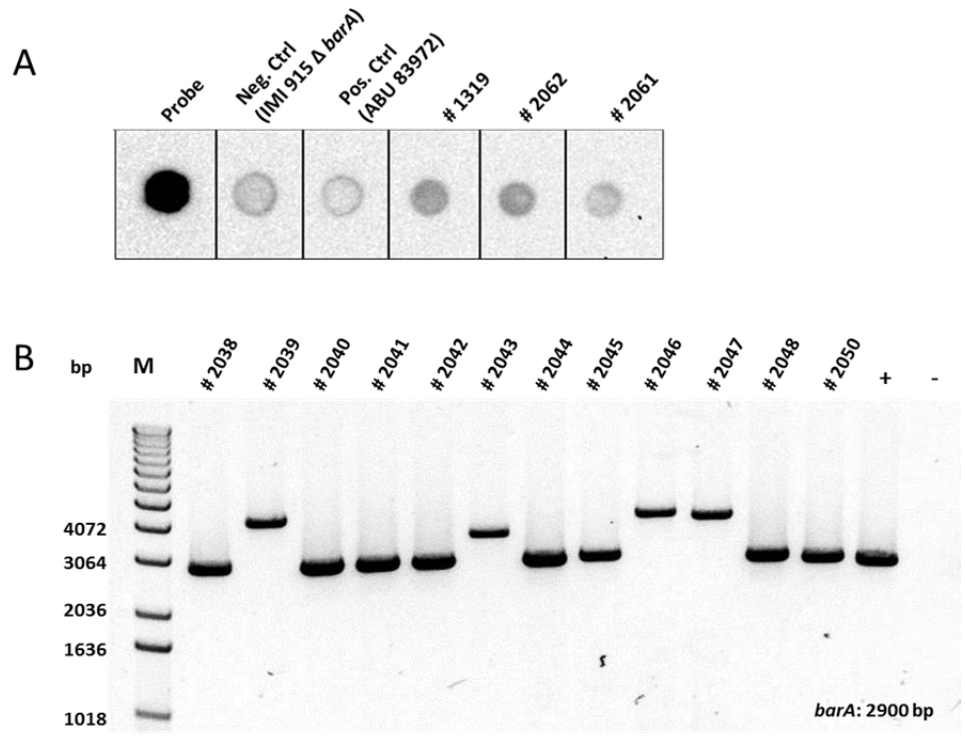


Figure 19: Dot Blot approach (A) and PCR products (B) obtained during the screening for *barA*.

(A) For the Dot blot *E. coli* strain IMI 915Δ*barA* served as a negative control whereas *E. coli* 83972 was used as a positive control. 10 ng of the labeled probe and 2 µg treated DNA/sample were applied to the membrane. (B) PCR of *barA* for selected samples. The regular size of the PCR product was 2900 bp. Genomic DNA (gDNA) of UPEC strain CFT073 served as a positive control.

For one fecal sample (# 2039) the exact nature of the sequence variation could not be determined. The detailed graphical overview of the insertion sequence elements in *barA* is shown for the respective strains in Figure 20 and Figure 21. For some strains only parts of IS elements were found to be inserted as in the case for ABU strains (A) # 1986, (B) # 2013, (C) # 2022, (D) # 2029, and the fecal strains (F) # 2046 and (G) # 2047. In some other strains the entire insertion element IS1 including down- and upstream regions was found to be integrated into the genomic context of *barA.* This was the case for the fecal strains # 2043 (E) and # 2070 (H). Though in most of the cases the main part of *barA* could be covered by sequencing, in some of the samples the start and the end of the coding region could not be identified. In addition some small fragments, as in the case for strain # 2043 (E), could not be identified exactly by genome mapping. The construction of a visual summary, as already mentioned, was not possible for sample # 2039 due to the fact that *barA* was not covered by the contigs.

Based on the fact that the IS elements disrupted the *barA* open reading frame (ORF), these mutant alleles were excluded from the study.

On the basis of the bad quality of the obtained nucleotide sequences from the sequencing procedure nine further samples for *barA* had to be excluded from the study.

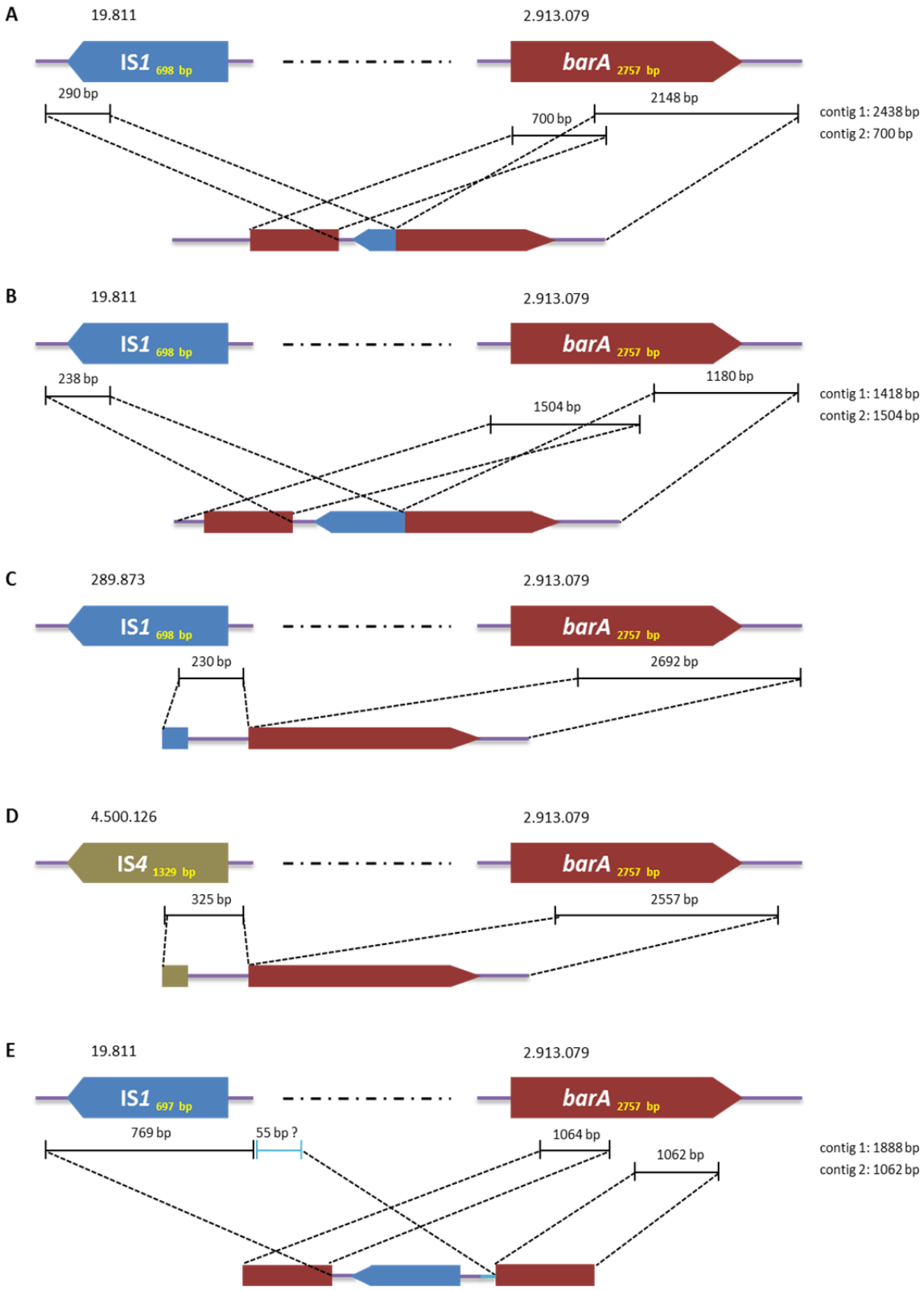


Figure 20: The deduced arrangement of IS elements detected in *barA*.

Positions of the single IS elements are shown in the genomic context of the respective *E. coli* strains. The deduced order is shown for *E. coli* strain (A) **#** 1986, (B) **#** 2013, (C) **#** 2022, (D) **#** 2029 and (E) **#** 2043.

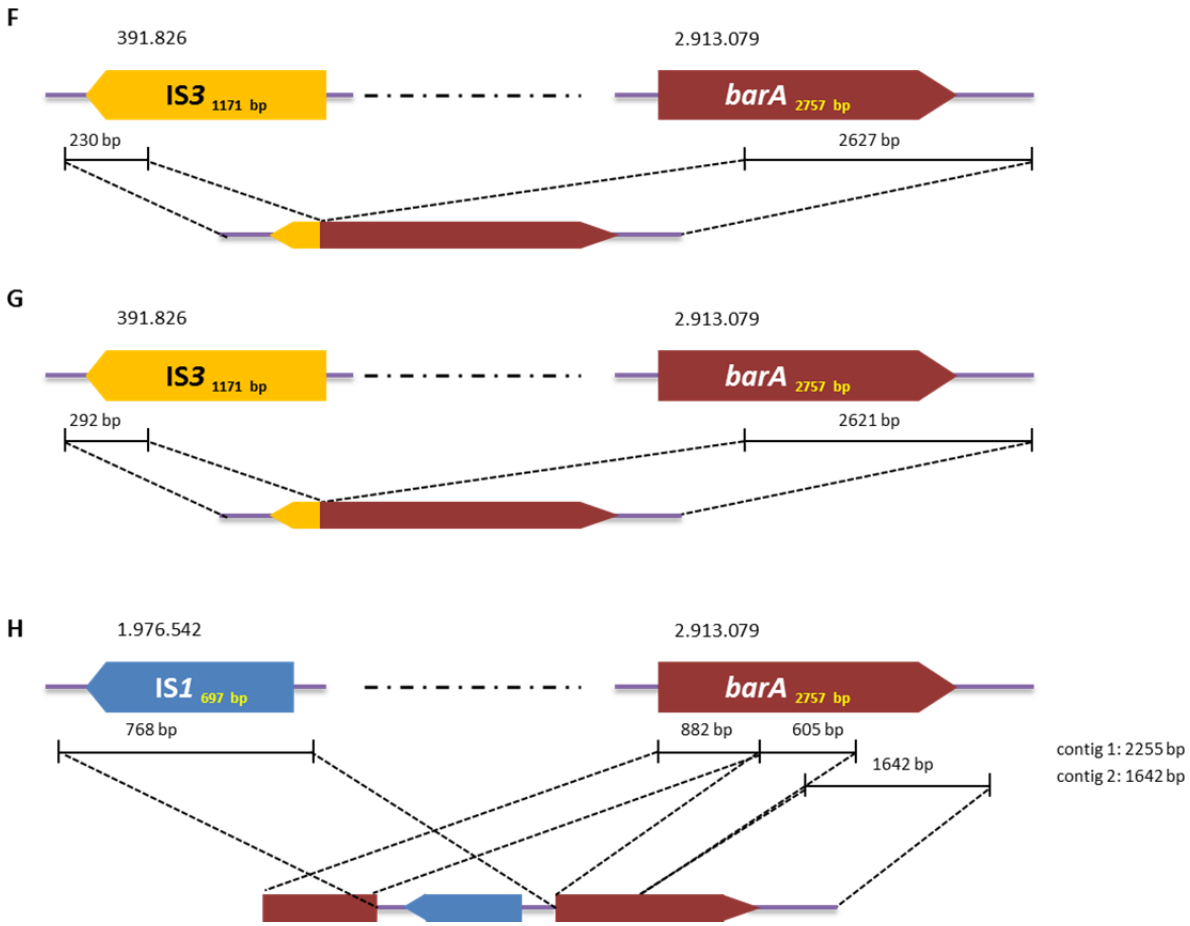


Figure 21: The deduced arrangement of IS elements detected in *barA*.

Positions of the single IS elements are shown in the genomic context of the respective *E. coli* strains. The deduced order is shown for *E. coli* (F) **#** 2046, (G) **#** 2047 and (H) **#** 2070.

For *uvrY* the same screening method was used as for *barA* (Figure 18). Five strains were excluded from the study. For two of them the sequencing reaction failed. As in the case for *barA*, three strains were identified that produced PCR products for *uvrY* (854 bp) with a higher molecular weight. Two of these were fecal isolates (# 2055, # 2076) and one was an ABU isolate (# 1964). Sequence alignments showed that the strains # 1964 and # 2055 were proven to harbor the insertion element IS5 while for *E. coli* # 2076 the insertion of IS1 was detected. The detailed graphical overview for the samples showing a higher fragment size is shown in the appendix (see 7.1.1).

Based on the ORF-disruptive nature of these IS elements the samples # 1964, # 2055 and # 2076 were rejected from the study, as well.

Interestingly IMI 917 and IMI 923, both UTI isolates, also showed a distinct band of higher molecular weight for *uvrY*. The sequences for both samples were extracted from Illumina whole genome data. In these strains, *uvrY* was not disrupted by an insertion of a mobile genetic element. The analysis of the draft genome sequences indicated than an IS element must be inserted between *uvrY* and ORF

yecF which is located upstream of *uvrY*. In the draft genomes of strains IMI 917 and IMI 923 the inverted repeat sequence GGGAAGGTGCGATTTAAT could be identified at the end of the two contigs which included either *uvrY* and its upstream or *yecF* and its upstream sequence context supporting the idea that an IS element or transposon must be inserted between *uvrY* and *yecF*. This insertion thus does not result in an inactivation of the ORF, but most likely affects the appropriate expression of the protein.

Finally only those samples were included in the studies that were successfully Sanger sequenced for both *uvrY* and *barA*. All strains that were excluded from the study are listed in Table 15. A total number of 188 strains were finally further investigated.

Table 15: From the study excluded *E. coli* strains.

Strain	Group	Reason for exclusion
IMI 962	ABU	Sequencing quality of *barA* not sufficient
# 1319	HA-ABU	Sequencing quality of *barA* not sufficient
# 1964	ABU	Insertion element found in *uvrY*
# 1970	ABU	Sequencing quality of *barA* not sufficient
# 1971	ABU	Sequencing quality of *barA* not sufficient
# 1974	ABU	Sequencing quality of *barA* not sufficient
# 1975	ABU	Sequencing quality of *barA* not sufficient
# 1979	ABU	Sequencing quality of *barA* not sufficient
# 1986	ABU	Insertion element found in *barA*
#2012	ABU	Sequencing for *uvrY* failed
# 2013	ABU	Insertion element found in *barA*
# 2022	ABU	Insertion element found in *barA*
# 2029	ABU	Insertion element found in *barA*
# 2036	ABU	Sequencing for *uvrY* failed
# 2039	Fecal isolate	Insertion element found in *barA*
# 2043	Fecal isolate	Insertion element found in *barA*
# 2046	Fecal isolate	Insertion element found in *barA*
# 2047	Fecal isolate	Insertion element found in *barA*
# 2055	Fecal isolate	Insertion element found in *uvrY*
# 2058	Fecal isolate	Sequencing quality of *barA* not sufficient
# 2061	Fecal isolate	Amplification of *barA* negative
# 2062	Fecal isolate	Amplification of *barA* negative
# 2070	Fecal isolate	Insertion element found in *barA*
# 2076	Fecal isolate	Insertion element found in *uvrY*
# 368	UTI	Sequencing quality of *barA* not sufficient

4.4.2 BarA/ UvrY sequence analysis

The analysis of the sequence variation in *barA* and *uvrY* for the tested *E. coli* isolates was done using the Ridom SeqSphere software. The *barA* and *uvrY* alleles of *E. coli* MG1655 (NC_000913) were used as reference sequences. With the help of the software, the following DNA sequence variations were detected: (i) synonymous SNPs, (ii) non-synonymous SNPs, (iii) deletions and (iv) insertions. As shown

in Table 16 the most common variation in comparison with the reference sequence was the occurrence of synonymous (syn) SNPs that alter the nucleotide sequence but not the amino acid of the encoded protein. Therefore those kinds of SNPs are also referred to as silent mutations. Another kind of mutation that was frequently detected and that we mainly focused on in the study is the appearance of non-synonymous (ns) SNPs (also described as missense SNPs) leading to a change of the amino acid sequence. These SNPs therefore might have an effect on the function of the protein. Because of this the ns SNPs were investigated in detail during the course of the study. A special sub-group of SNPs that was also categorized as ns SNPs were the non-sense SNPs. These result in a stop codon thus breaking the function of the protein. These variations were rarely detectable in both *barA* and *uvrY*. Another variation identified was the appearance of insertions or deletions of nucleotides if compared to the reference sequence. The outcome of INDELs (insertion/deletion) can vary between having no effect on the function of the protein and leading to a frame shift.

The sequencing of the PCR products of 188 *E. coli* samples revealed the following results for the analysis of the gene coding for the sensor histidine kinase BarA as shown in Table 16: a total of 179 mutations were found in *barA*. Those included 154 syn SNPs, 22 non-synonymous SNPs, two non-synonymous mutations resulting in a premature stop codon and one deletion (1 nucleotide).

Table 16: sequence variation found in total for *barA*, *uvrY* and *frr*.

Variation	*barA*	*uvrY*	*frr*
synonymous	154	26	9
non-synonymous	22	19	3
non-sense	2	2	0
insertion	0	0	0
deletion	1	2	0
	179	49	12

The distribution of the detected variations in *barA* for the four different groups of isolates is shown in Figure 22. The general structure of *barA* with its conserved domains according to the NCBI database is shown at the top of the graphic in (A). Listed below (B) are the identified mutations found for each group of isolates (ABU, HA-ABU, FI and UTI). The identified mutations were combined for the relevant groups and show the distribution of the detected SNPs from all isolates. The red bars indicate ns SNP positions that have not been found in other isolate groups thus being individual for the respective group. The yellow bars display ns SNP locations that are found in other groups, as well, hence not being special for this collection of isolates. The light blue bar depicts an INDEL. Positions marked with asterisks denote the occurrence of stop codons due to a non-sense SNP. As can be seen

by comparing the identified sequence variations in each group (Figure 22 B), most ns SNPs apparently occur in the ABU and UTI isolates. As shown in Table 17 a total number of 16 ns SNPs (including the INDEL) was found for the ABU group. The number of ns SNPs detected in the UTI isolates accounts for eight. In contrast to this the HA-ABU strains exhibit a number of five ns SNPs in total and for the fecal isolates seven ns SNPs could be detected. The exact positions of the ns SNPs are shown in Table 18. Even though the number of isolates tested was not the same for each group (ABU (n=72), HA-ABU (n=33), FI (n=29) and UTI (n=54)) it can anyhow be stated that the amount of non-synonymous variation tends to be highest in ABU and fecal isolates if normalized to the amount of strains tested (Table 17). With regard to the amount of SNPs per strain it can be summarized that the majority of strains only exhibited one ns SNP in the entire *barA* gene (data not shown). One *E. coli* strain belonging to the group of UTI isolates was found to harbor three ns SNPs, whereby two of the three ns SNPs were found in isolates of the other groups, as well. In addition to the above mentioned observations a further striking feature of the SNP distribution could be detected. Apparently 40 % of the ns SNPs occurred in functional regions of the protein. A similar pattern can be observed for the synonymous SNPs, whose distributions are shown in Figure 22 C. Although the syn SNPs spread over the entire gene a partial clustering can be detected. For example a clear accumulation of syn SNPs can be identified for the receiver domain (Rec). In contrast to that the histidine kinase A (HisKa) domain seems to be very conserved and insusceptible for even synonymous changes.

Table 17: ns SNP found in the four groups of isolates and calculated to an average of ns SNP per isolate.

Group	*barA*		*uvrY*	
	In total	**∅ per isolate**	***In total***	**∅ per isolate**
ABU (n=72)	16	0.22	15	0.21
HA-ABU (n=33)	5	0.15	1	0.03
FI (n=29)	7	0.24	3	0.1
UTI (n=54)	8	0.15	7	0.13

The summary of the SNP analysis for *uvrY* is graphically shown in Figure 23. The construction and arrangement of the graphic is analogous to the one for *barA*. Conserved regions of *uvrY* are illustrated according to the NCBI database. The comparative analysis of the *uvrY* alleles revealed a number of 49 mutations in total. Among these as 26 silent mutations, 19 non-synonymous SNPs, two nonsense SNPs as well as two INDELs were identified (Table 16). For ABU isolates a total number of 15 ns SNPs including one nonsense and one INDEL could be detected (Table 17). The next highest number of ns SNPs was found in the UTI group (six ns SNPs and one non-sense mutation) followed by the fecal isolates showing three ns SNPs. Among these one INDEL was found. The group with the lowest amount of ns SNPs was the HA-ABU collection. Here only one variation could be detected. The

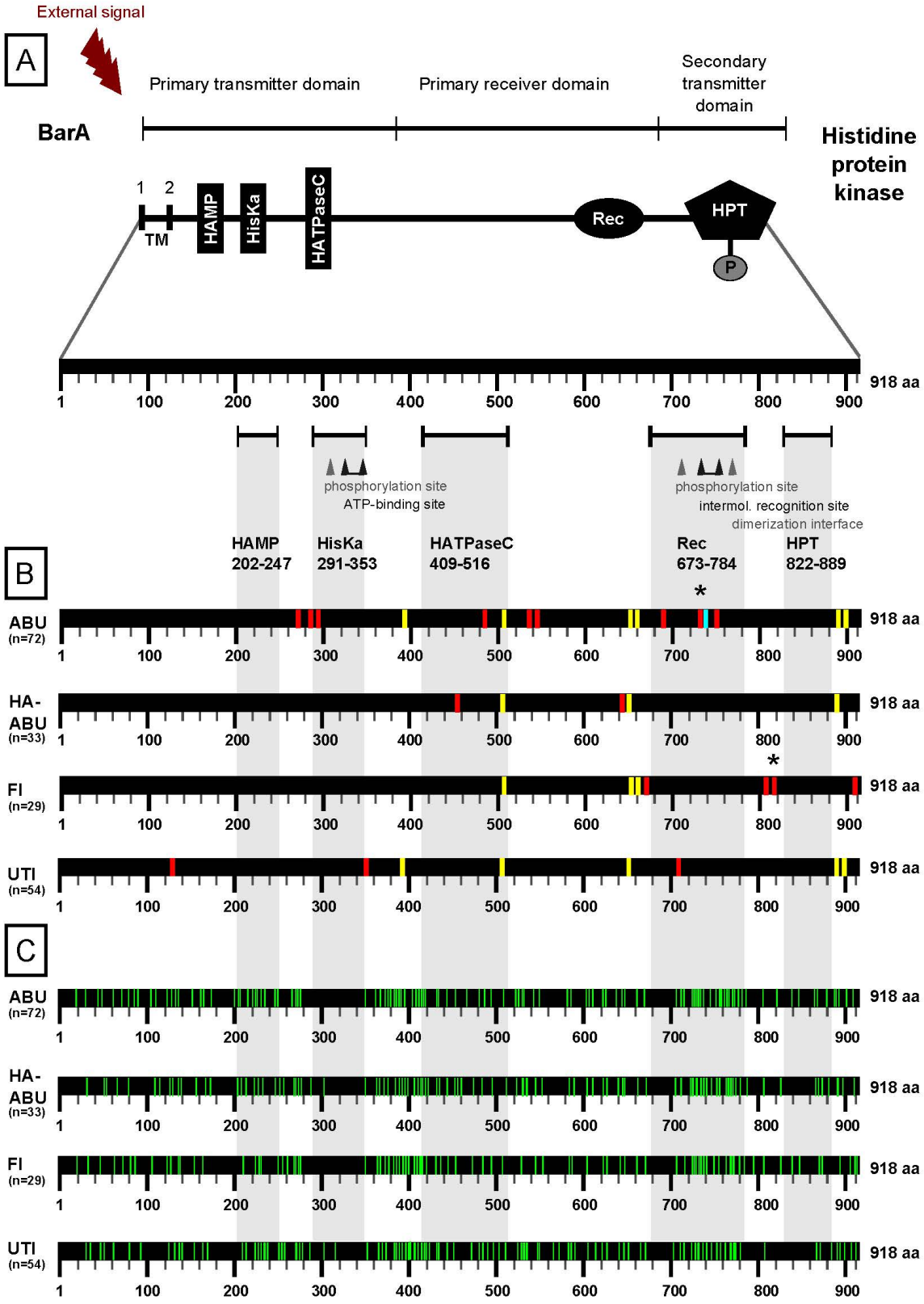


Figure 22: Graphical summary of the sensor histidine kinase BarA secondary structure and identified mutations in the four groups of isolates.

BarA conserved domains are shown in (A). The distribution of the non-synonymous SNPs in the proteins compared between the four isolate groups ABU, HA-ABU, UTI and FI is shown in (B) whereas the location of syn SNPs is shown in (C). Yellow bars indicate the same position of the ns SNPs between the groups, blue bars show an insertion or deletion (INDEL)), red bars depict different ns SNPs and green ones the syn SNPs. Asterisks indicate the position of ns SNPs resulting in a stop codon.

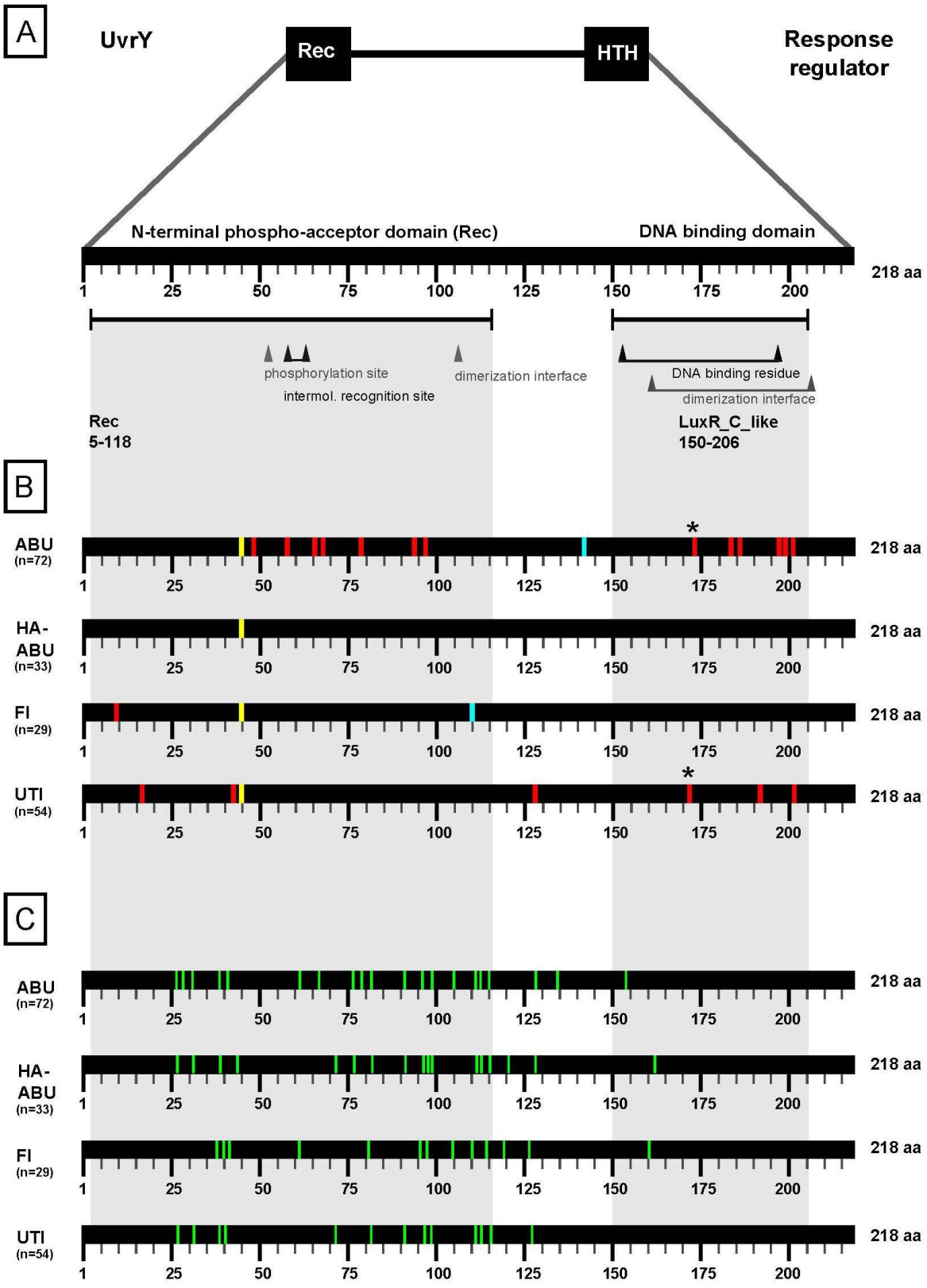


Figure 23: Graphical summary of the response regulator UvrY secondary structure and identified mutations in the four groups of isolates.

UvrY conserved domains are shown in (A). The distribution of the non-synonymous SNPs in the proteins compared between the four isolate groups ABU, HA-ABU, UTI and FI is shown in (B). The one of synonymous SNPs in (C). Yellow bars indicate the same position of the ns SNPs between the groups, blue bars show an insertion or deletion (INDEL), red bars depict different ns SNPs and green ones the syn SNPs. Asteriks indicate the position of ns SNPs resulting in a stop codon.

SNP distribution was similar to *barA*. The majority of ns SNPs appeared in isolates belonging to the ABU and UTI group (Figure 23 B). Normalization of the number of SNPs on the amount of tested samples per group confirmed this assumption (Table 17). In conclusion, the sequence variation is - independently of the group size - highest in ABU and UTI isolates. As for *barA* the average numberof ns SNPs was one per strain and only one UTI isolates showed three ns SNPs. The majority of ns SNPs found in *uvrY*, namely 95 %, was again primarily located in functional domains as observed for *barA*. The exact positons of the ns SNPs are shown in Table 18. Regarding the distribution of synonymous SNPs it can be summarized that the majority of the SNPs seem to accumulate in the N-terminal part of the protein. The C-terminal region only exhibits a few synonymous SNPs for all groups except for the UTI isolates which did not show any synonymous SNPs in the C-terminal region.

Table 18: Non-synonymous SNPs detected in *barA*, *uvrY* and *frr*.

The different SNP variants are assigned to the groups in which they were detected (1: ABU, 2: HA-ABU, 3: FI, 4: UTI). In addition the exact position of the SNPs and the change in the nucleotide/amino acid (AA) sequence is shown. The reference strain (Ref) is always compared to the variant strain

Gene	No	SNP variant	Group	Nucleotide position	Base change (Ref:Variant)	AA change (Ref:Variant)	Codon	Codon effect
barA	1	ns	4	398	C:T	T:M	133	missense
	2	ns	1	808	G:A	E:K	270	missense
	3	ns	1	857	C:T	A:V	286	missense
	4	ns	1	872	G:A	R:H	291	missense
	5	ns	4	1053	C:A	F:L	351	missense
	6	ns	1,4	1193	T:A	V:E	398	missense
	7	ns	2	1376	G:T	R:L	459	missense
	8	ns	1	1445	C:A	T:N	482	missense
	9	ns	1,2,3,4	1514	A:T	Q:L	505	missense
	10	ns	1	1616	T:C	L:P	539	missense
	11	ns	1	1618	G:C	A:P	540	missense
	12	ns	2	1939	A:G	T:A	647	missense
	13	ns	1,2,3,4	1949	G:A	C:Y	650	missense
	14	ns	1,3	1979	C:T	V:A	660	missense
	15	ns	3	1999	G:C	A:P	667	missense
	16	ns	1	2075	A:T	Q:L	692	missense
	17	ns	4	2110	G:A	V:I	704	missense
	18	ns	1	2203	C:T	Q:*	735	nonsense
	19	deletion	1	2211	Δg			frameshift
	20	ns	1	2258	C:T	A:V	753	missense
	21	ns	3	2425	A:C	T:P	809	missense
	22	ns	3	2452	C:T	Q:*	818	nonsense
	23	ns	1,2,4	2663	G:A	S:N	888	missense
	24	ns	1,4	2674	G:A	E:K	892	missense
	25	ns	3	2718	G:A	M:I	906	missense
uvrY	1	ns	3	25	G:C	D:H	9	missense
	2	ns	4	47	G:A	G:E	16	missense
	3	ns	4	130	C:T	R:W	44	missense
	4	ns	1,2,3,4	133	A:G	T:A	45	missense
	5	ns	1	139	G:T	A:S	47	missense
	6	ns	1	173	C:T	P:L	58	missense
	7	ns	1	194	C:A	A:E	65	missense

Gene	No	SNP variant	Group	Nucleotide position	Base change (Ref:Variant)	AA change (Ref:Variant)	Codon	Codon effect
	8	ns	1	202	A:G	K:E	68	missense
	9	ns	1	239	T:C	M:T	80	missense
	10	ns	1	284	A:C	Q:P	95	missense
	11	ns	1	286	G:A	A:T	96	missense
	12	deletion	3	331	Δgtcgtgagtgc			frameshift
	13	ns	4	380	C:A	S:Y	127	missense
	14	deletion	1	422	Δa			frameshift
	15	ns	3	505	G:T	E:*	169	nonsense
	16	ns	1	517	C:T	Q:*	173	nonsense
	17	ns	1	551	A:G	Y:C	184	missense
	18	ns	1	556	T:G	Y:D	186	missense
	19	ns	3	575	T:C	L:P	192	missense
	20	ns	1	589	G:A	D:N	197	missense
	21	ns	1	592	G:A	V:I	198	missense
	22	ns	1	599	T:C	L:P	200	missense
	23	ns	4	617	G:A	R:H	206	missense
frr	1	ns	1	287	G:A	S:N	96	missense
	2	ns	1	460	C:T	R:C	154	missense
	3	ns	1	530	A:G	D:G	177	missense

In order to estimate the extent of sequence variability of the analyzed *barA* and *uvrY* genes we also sequenced the housekeeping gene *frr* (ribosome recycling factor) in all isolates to be able to compare the sequence variability. The graphical overview analogous to *barA* and *uvrY* is shown in Figure 24. A total of 12 SNPs were detectable in *frr* for the analyzed isolates (Table 16) whereby three of them accounted for ns SNPs. Interestingly all of them exclusively appeared in ABU isolates. The low number of ns SNPs corroborates that specifically in the ABU and UTI isolates *barA* and *uvrY* exhibited a significantly greater nucleotide sequence variability than *frr* in comparison to the HA-ABU or fecal isolates, thus underlining that these genes may be subjected to positive selection upon growth in the human urinary tract. The nine identified syn SNPs were distributed among the entire gene. A striking clustering, as for *barA* or *uvrY*, could not be observed. The detailed information on the positions of the detected ns SNPs is listed in Table 18 A.

Thus the general assumption of a housekeeping gene being more conserved than other genes can be confirmed if comparing *frr* with *barA* and *uvrY*.

To further illustrate the divergence of the appearing SNPs and the connection between the four examined groups a graphical visualization was done for *barA* (Figure 25), *uvrY* (Figure 26 A) and *frr* (Figure 26 B) by creating a Minimum Spanning Tree (MST) based on all detected SNPs in the respective gene. The isolates in the MSTs are represented by their sequence variation types (SVTs) which are noted in the circles. The sequence variation type was determined based on the nucleotides at the different SNP positions which were identified during the SNP screening (4.4.2). The nucleotides found in the individual strains were assigned to different numbers which finally define the sequence

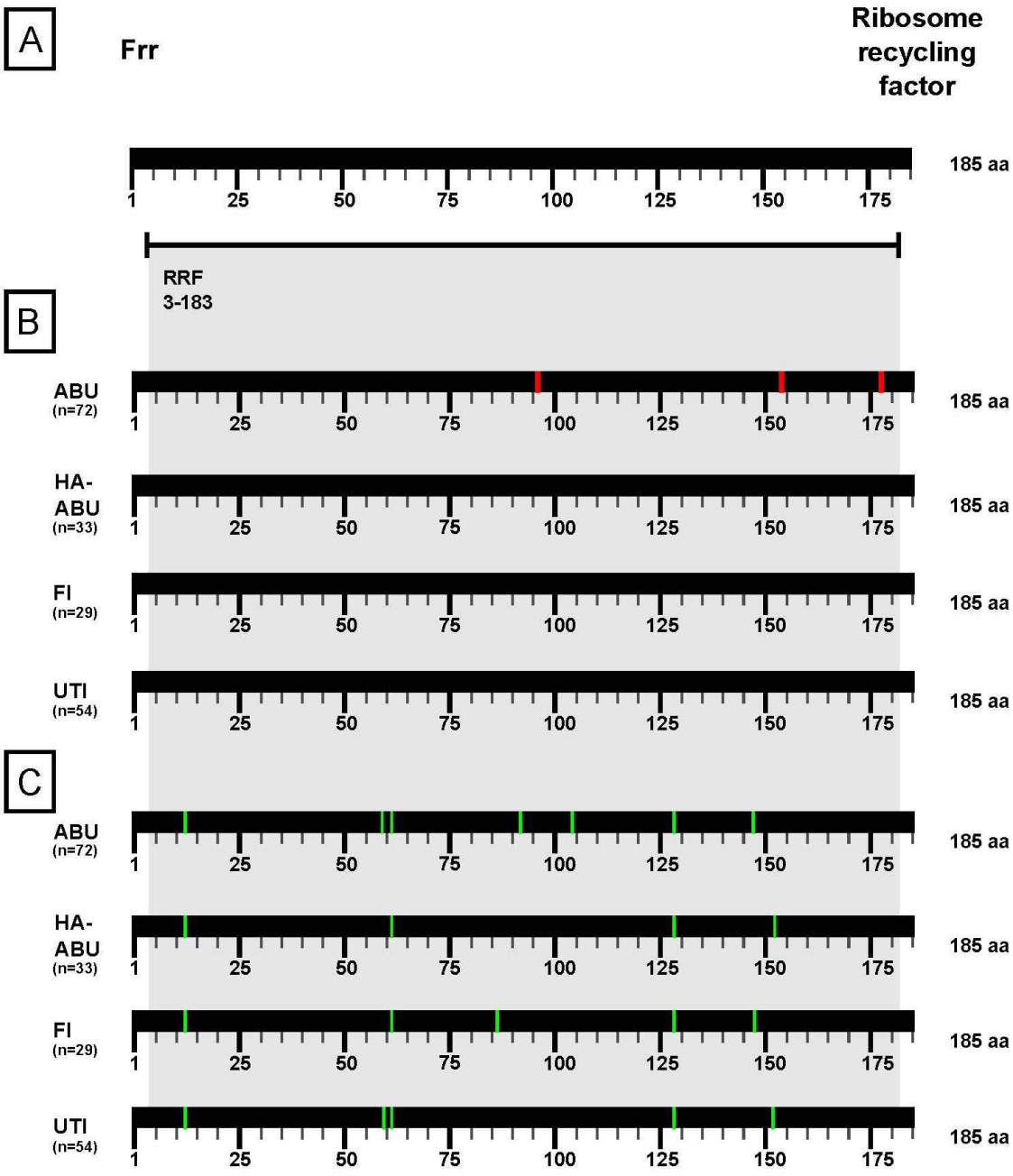


Figure 24: Graphical summary of the ribosome recycling factor *frr* secondary structure and the identified mutations in the four groups of isolates.

Frr conserved domains are shown in (A). The distribution of the non-synonymous SNPs in the proteins compared between the four isolate groups ABU, HA-ABU, UTI and FI is shown in (B). The allocation of synonymous SNPs in (C). Red bars depict different ns SNPs. Green bars represent the syn SNPs appearing in the protein.

variation type (SVT), a procedure similar to the MLST classification. The allocation of ABU, HA-ABU, UTI and fecal isolates is indicated by the pie charts whereby different colors depict the four groups. Additionally the size of the pie chart mirrors the number of strains allocated to the individual SVTs.

The MST of *barA* (Figure 25) clearly shows the appearance of a huge number of different SNPs (see also Table 16) that are distributed among all of the four examined groups. A number of 85 different SVTs were identified among the strains. The most common SVTs were 3, 14, 34 and 39 whereby SVT 14 and 34 harbored isolates of all groups. In contrast SVT 3 and 39 could not be assigned to HA-ABU isolates. The graphic further nicely illustrates that distinct SVTs were identified which represent a special SNP profile being characteristic for individual groups.

A similar phenomenon can be observed for *uvrY* (Figure 26 A). The SNP variation is not as high as for *barA*, but nevertheless the graphic representation shows that 40 different sequence variation types were found in *uvrY*. The most shared SVTs are e.g. 2, 3, 11 or 20 which are distributed among all of the examined groups. This implies that at least the vast majority of the strains from the different groups carry the same SNP combination for *uvrY*. Comparable to *barA* there is also a multitude of SVTs which seem to be individual for different groups, e.g like SVT 16 which is characteristic for ABU isolates or the SVT 26 being present only in HA-ABU isolates.

The situation for *frr* is rather different (Figure 26 B). A total number of 15 different SVTs were detected which is much lower than the number of SVTs found in *barA* or *uvrY*. Most of the tested isolates belong to SVT 1, 3 or 5 which is mirrored by the size of the pie chart. These SVTs are distributed among the four different groups. Another striking feature which is also underlined by the detected number of SVTs is the aspect that *frr* did not show such a marked SNP variation as *uvrY* or *barA*. This fact once again emphasizes that *frr*, due to its nature as a housekeeping gene, is more conserved than *barA* or *uvrY*.

Taken together the majority of the strains as mirrored by the sizes of the pie charts belong either for *barA* or *uvrY* and *frr* to the same SVTs. This implies that a discrimination of the strains based on their SVTs cannot be performed.

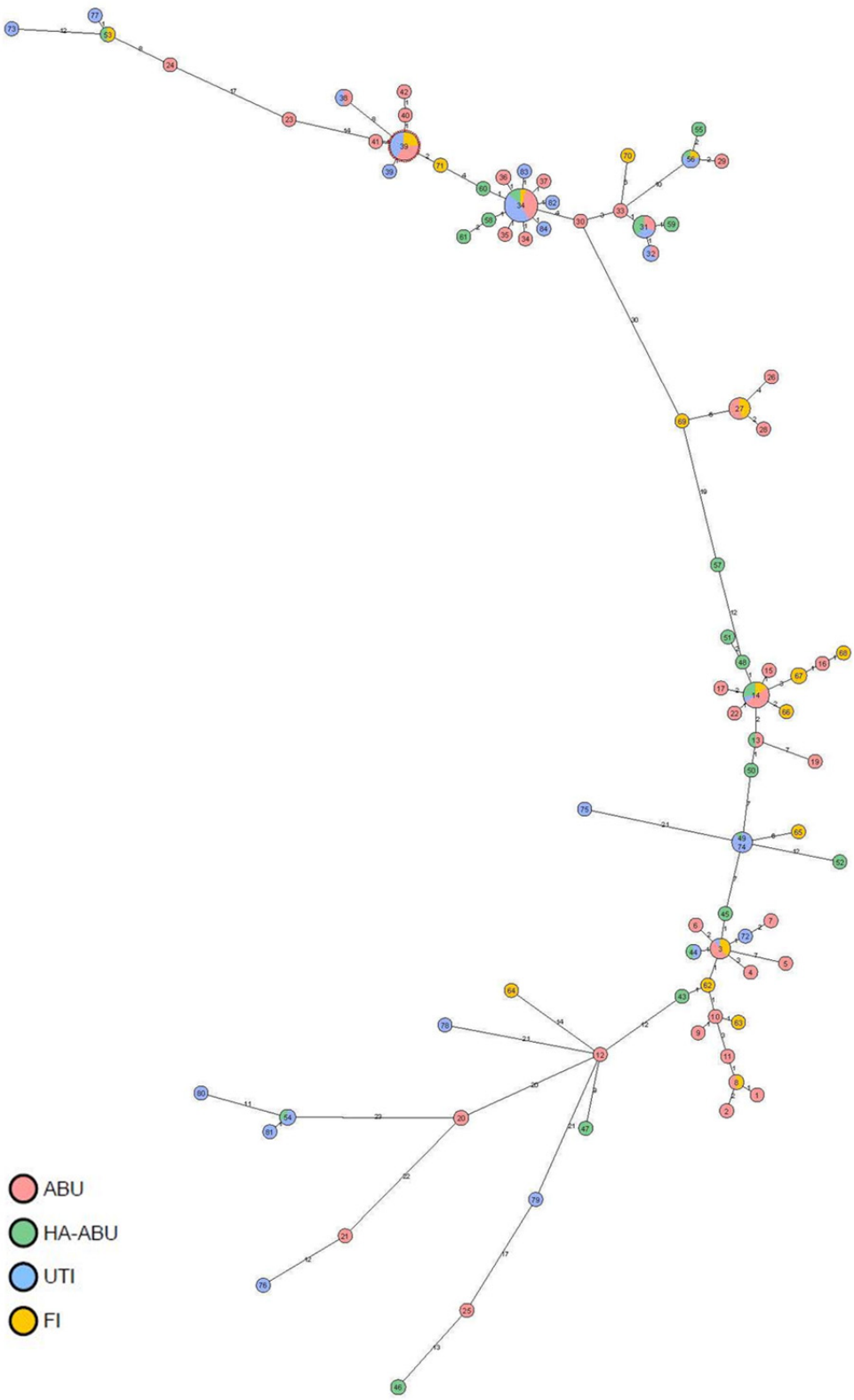


Figure 25: Minimum Spanning Tree (MST) based on occurring SNPs in *barA* portraying the distributions of ABU, HA-ABU, UTI and fecal *E. coli* isolates.

Isolates are represented by their sequence variation types (SVT) named in the circles. The allocation of ABU, HA-ABU, UTI and fecal isolates is indicated by pie charts whereby different colors depict the four groups. The size of the pie chart mirrors the number of strains allocated to the individual SVTs.

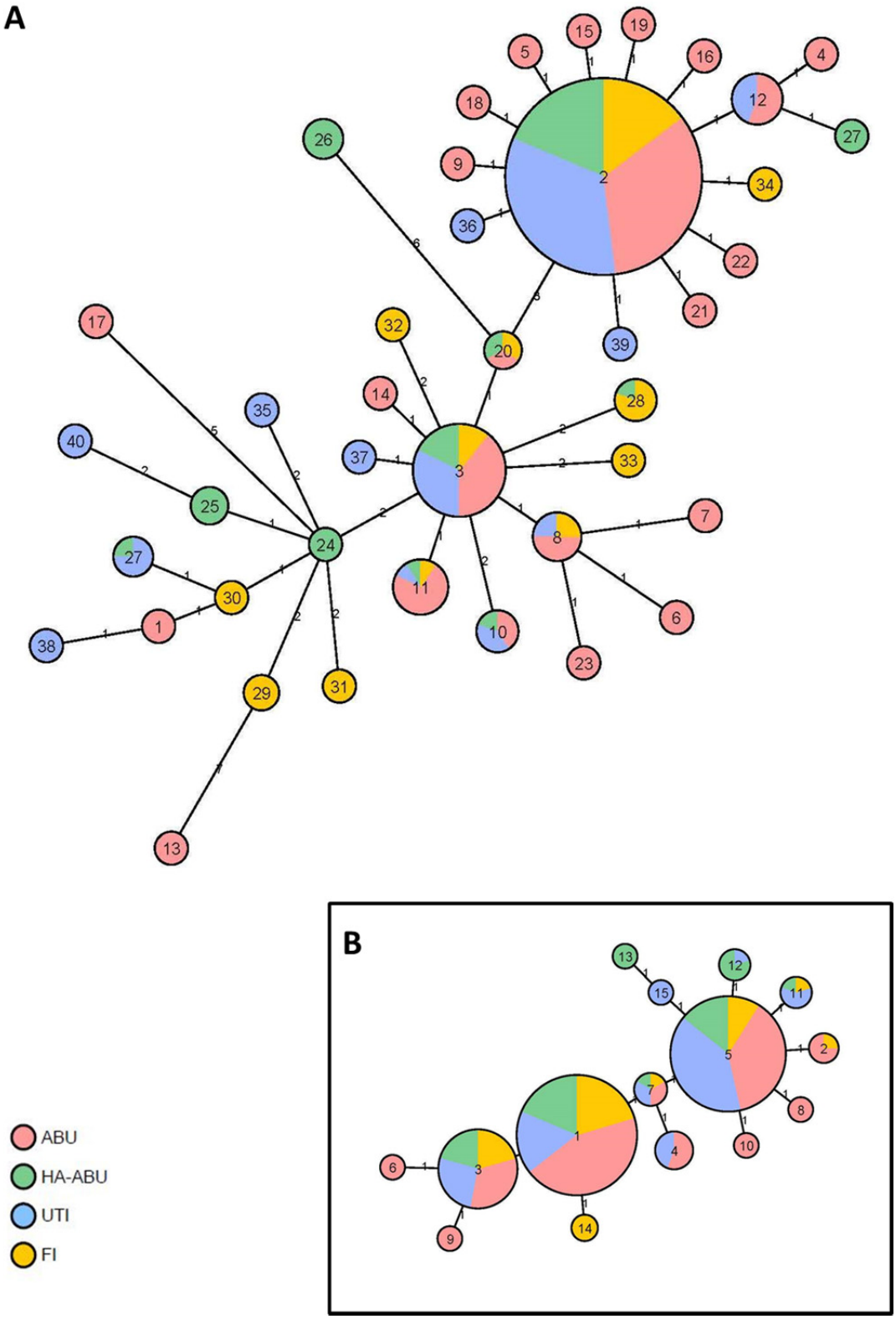


Figure 26: Minimum Spanning Tree (MST) based on occurring SNPs in *uvrY* (A) and *frr* (B) portraying the distribution of ABU, HA-ABU, UTI and fecal *E. coli* isolates.

Isolates are represented by their sequence variation types (SVT) named in the circles. The allocation of ABU, HA-ABU, UTI and fecal isolates is indicated by pie charts whereby different colors depict the four groups. The size of the pie chart mirrors the number of strains allocated to the individual SVTs**.**

4.5 Phylogenetic classification

In order to investigate if the four different *E. coli* groups, ABU, HA-ABU, FI and UTI, can be associated with a specific phylogenetic background a Minimum Spanning Tree based on given multi locus sequence typing (MLST) data of seven housekeeping genes namely *adk* (adenylate kinase), *fumC* (fumarate hydratase), *gyrB* (DNA gyrase), *icd* (isocitrate/isopropylmalate dehydrogenase), *mdh* (malate dehydrogenase), *purA* (adenylosuccinate dehydrogenase) and *recA* (DNA strand exchange and recombination protein with protease and nuclease activity) (Wirth *et al.*, 2006; http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/documents/primersColi_html) was created for the examined strain panel of 188 isolates (Figure 27). MLST, which has been introduced by Maiden *et al.* (1998), is a very well-established method for the phylogenetic classification of bacterial isolates. Based on the MLST profiles, *E. coli* strains can be assigned to their sequence type as well as to a phylogenetic group according to the ECOR (*E. coli* group of reference strains) phylogenetic classification which was established by Ochman and Selander (1984). As stated by Clermont *et al.* (2000) *E. coli* strains can mainly be classified into four phylogenetic groups (A, B1, B2, D) whereby A and B1 are typical ECOR groups for commensal strains and groups B2 and D include the majority of ExPEC strains (Boyd & Hartl, 1998; Tenaillon *et al.*, 2010b). The MLST results for the tested strain panel (Figure 27) showed that the isolates could be assigned to six phylogenetic lineages (A, B1, B2, ABD, AxB1 and D) whereby each of this phylogroup included ABU, UTI, HA-ABU and fecal isolates. The most prominent phylogroup was B2 which included a high proportion of UTI isolates and the dominant clonal complexes CC 73, CC 14 and CC 95. The second most common phylogroup was B1 which mainly comprised avirulent and commensal strains. In contrast, phylogroup A harbored mostly ABU or HA-ABU isolates. Although there was a partial correlation between the four types of isolates and their phylogenetic background as mentioned above, strains of all four groups were present in each phylogroup.

These results show that all groups share a general phylogenetic background and that a discrimination of the isolates based on the MLST analysis into special groups or branches was not possible.

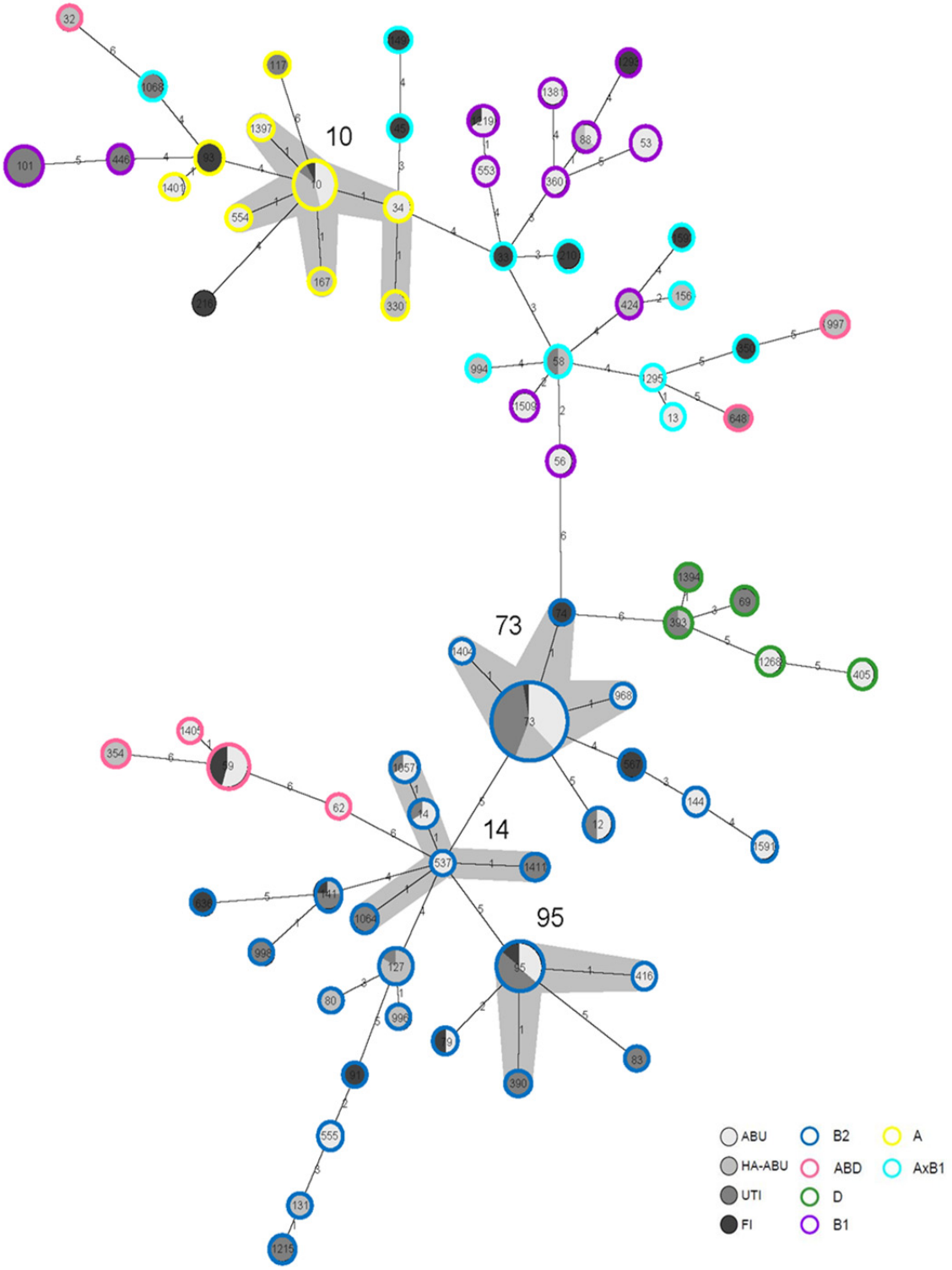


Figure 27: Minimum Spanning Tree (MST) based on occurring SNPs in seven housekeeping genes portraying the distribution of ABU, HA-ABU, UTI and fecal *E. coli* isolates.

Isolates are represented by their sequence types (ST) named in the circles. The allocation of ABU, HA-ABU, UTI and fecal isolates is indicated by pie charts whereby different colors depict the four groups. The size of the pie chart mirrors the number of strains allocated to the individual STs**.**

4.6 Test of selection – McDonald-Kreitman Test

To quantitatively assess the role of adaptive evolution in shaping the patterns of variation observed a mathematical test was applied. The fact that the population is exposed to constant fluctuations in the bladder like micturition or changing nutrient conditions makes the choice of a mathematical test complicated. The frequency of urination - which varies depending on the age of the individual - leads to a decrease of the bacterial population followed by an increase of the residual culture. As a consequence of this physiological situation the bacterial population in the bladder is exposed to a bottleneck effect. The mathematical test of selection that meets the problem the best and therefore is most suitable for this scenario is the McDonald-Kreitman test of neutral evolution. This classical test of adaptive evolution assumes that most evolution is random without selective effects. It was initially used and described by McDonald & Kreitman (1991). The test aims to investigate the signature of natural selection at the DNA level by comparing the amount of variation within a species (polymorphisms) to the divergence between species (fixed substitutions) at two types of sites. Thereby it is assumed that one of the sites evolves neutrally. This one is used as a reference to detect selection at potentially functional sites (Egea *et al.*, 2008; Messer & Petrov, 2013). Practically the test calculates two ratios, on the one hand the ratio of fixed replacement differences to fixed synonymous differences and the ratio of polymorphic replacement differences to fixed synonymous differences:

$$\frac{ns\ fixed}{syn\ fixed} = \frac{ns\ polym}{syn\ polym}$$

The ratios are indicators of evolutionary rates. If they are equal, the hypothesis of neutral evolution does not have to be rejected; evolution is proven to occur by neutral processes.

In order to perform the McDonald-Kreitman test and to estimate the interspecies divergence for *barA, uvrY* and *frr* in the four *E. coli* populations (ABU, HA-ABU, UTI, FI) we have used other enterobacterial species namely *Salmonella enterica* ssp. *enterica* sv. *Typhimurium* as an outgroup. These bacteria are able to cause gastroenteritis but they are not adapted to growth in the urinary tract. Based on the differences of their natural habitats *Salmonella enterica Typhimurium* should merely be distantly related to *E. coli* isolates analyzed in this study on the molecular level. Since a calculation of the McDonald-Kreitman test parameters with phylogenetically close organisms is not possible the choice of a suitable outgroup is a critical step of the test design. Consequently performing the test has not been possible using species like *Shigella* or *E. coli* K-12. The results for the McDonald-Kreitman test, the type of substitutions identified, the neutrality index (NI: denotes the degree of deviation of the levels of amino acid polymorphism according to the neutral model) and the associated statistics are shown in Table 19. The McDonald-Kreitman test determined a

signature which was compatible with non-neutral evolution for *barA* or *uvrY*. Whereas no such a signature was detected for the control housekeeping gene *frr*. However, the scenarios for *uvrY* and *barA* are different. In *barA* there was no sign of positive selection. Compared to the number of fixed substitutions, the amount of polymorphic synonymous variation is in excess in each of the four *E. coli* groups, suggesting a situation where purifying selection limits sequence variation to the same extent in each of the four groups analyzed. In *uvrY* instead, we detected a signature compatible with positive selection in the UTI and ABU strains. While the ratio of fixed ns/syn substitution was constant between the four strain sets (about 1/11), indicating similar long-term selective constraints in the four groups of isolates, the polymorphic differences varied substantially between the groups. While for the HA-ABU and FI isolates the Fishers test indicated a scenario compatible with neutrality,

Table 19: Results of the McDonald-Kreitman test for *uvrY*, *barA* and *frr*.

Salmonella enterica Typhimurium was used as an outgroup (OG). The Neutrality Index (NI) and the Fishers exact test p value are also shown in the table. Blue numbers indicate the approximate down calculated ratios.

Gene	OG	Group	Substitutions	Fixed	Polymorphic	Neutrality Index	Fishers exact test P value
barA	*Salmonella enterica* subsp. *enterica* serovar *Typhimurium*	ABU	non-synonymous	107 1	15 1	0.483	0.012911*
			synonymous	424 4	123 8		
		HA-ABU	non-synonymous	107 1	5 1	0.171	0.000004***
			synonymous	420 4	115 23		
		UTI	non-synonymous	107 1	8 1	0.255	0.000053***
			synonymous	419 4	123 15		
		Fecal	non-synonymous	107 1	7 1	0.274	0.000311**
			synonymous	427 4	102 15		
uvrY		ABU	non-synonymous	7 1	14 1	7.7	0.000075***
			synonymous	77 11	20 1.4		
		HA-ABU	non-synonymous	7 1	1 1	0.643	1
			synonymous	81 11	18 18		
		UTI	non-synonymous	7 1	7 1	5.786	0.005369**
			synonymous	81 11	14 2		
		Fecal	non-synonymous	7 1	2 1	2.229	0.595916
			synonymous	78 11	10 5		
frr		ABU	non-synonymous	12 1	4 1	3.167	0.199502
			synonymous	57 5	6 2		
		HA-ABU	non-synonymous	12 1	2 1	1.867	0.608352
			synonymous	56 5	5 3		
		UTI	non-synonymous	12 1	1 1	0.778	1
			synonymous	56 5	6 6		
		Fecal	non-synonymous	12 1	1 1	0.764	1
			synonymous	55 5	6 6		

in the ABU and UTI groups the polymorphic sites showed an excess of non-synonymous mutations. This is suggestive of positive selection.

The fact that *barA* and *uvrY*, as suggested by the McDonald-Kreitman test, are obviously more susceptible to sequence variation than *frr* was also supported by the investigation of the conservation scores. Conservation scores are calculated by comparing the target gene/protein with the entire database containing all available target sequences of any species. This way the conservation rate can be calculated providing the opportunity to directly compare different genes/proteins at this level. The results of the calculated conservation levels for the proteins BarA, UvrY and Frr are shown as a non-normalized boxplot distribution in Figure 28. BarA and UvrY proteins had similar distributions of the conservation rate with values of 1.59 (± 0.70) and 1.63 (± 0.97), while Frr showed a lower value of 1.32 (± 1.04) and therefore tends to be more conserved, as expected. The same can be observed for the comparison with the entire proteome of *E. coli* MG1655 (NC_000913) with a mean value of 1.49 (± 1.08).

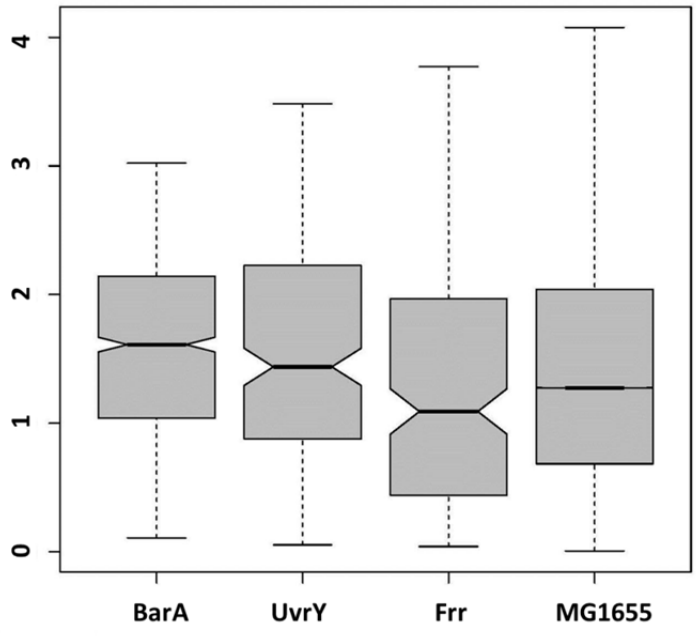


Figure 28: Conservation scores calculated for BarA, UvrY, Frr and the whole proteome of *E. coli* MG1655.

In summary both, the McDonald-Kreitman test and conservation scores suggested that (i) positive selection pressure is acting on *uvrY* during growth in the bladder and that (ii) UvrY and BarA are less conserved compared to the Frr protein. The latter aspect supports our initial hypothesis that the TCS BarA/UvrY is more tolerant to adaptive evolution.

In addition to the degree of conservation the evolutionary rates of the individual positions of the protein sequences for BarA and UvrY were calculated. The results are presented in the appendix (see 7.1.2).

4.7 SNP combination analysis

In order to gain further insights of the effect of the ns SNPs in *barA*/*uvrY* detected via SNP analysis (Table 18) in different groups of isolates we have analyzed various *E. coli* strains of the initial strain panel regarding different phenotypic traits. The phenotypic diversity should be assessed using different phenotypic assays which are described in the following chapters. Due to the high number of primarily analyzed *E.coli* strains a SNP combination analysis based on the variation in both *uvrY* and *barA* was performed. Thus it was ensured that one representative strain of any possible SNP combination was selected for further experiments. The tested strains were chosen randomly. By exclusive consideration of the ns SNPs the analysis revealed a number of 47 possible ns SNP combinations. The results are shown in Table 20.

Table 20: Detected ns SNP combinations regarding the nucleotide positions in *uvrY* and *barA*.

No	Group	*E. coli* strain	*uvrY*			*barA*		
			1	2	3	1	2	3
1		# 255	133					
2		# 1963	133			2075		
3		# 1967	133			872	1514	
4		# 1969	133	422		2663		
5		# 1972	133	517				
6		# 1976	173					
7		# 1978	239					
8		# 1980				1514		
9		# 1983	133	599				
10		# 1984				808		
11		# 1990	133			2663		
12		# 1991	133			2663	2674	
13		# 1992	139	551		1979		
14		# 1994	194					
15	**ABU**	# 1999	133	286		1514	1949	
16		# 2000	133	589		1514		
17		# 2005				2211		
18		# 2006	133			1445		
19		# 2007	133			1514		
20		# 2011	133			1616		
21		# 2017	133			2203		
22		# 2020				1618		
23		# 2021	133			2258		
24		# 2027	133	284		1514		
25		# 2028	133	202		1514		
26		# 2032	133	592				
27		# 2033				857		
28		# 2035	133			1193	1514	
29		IMI 961	556					

No	Group	*E. coli* strain	*uvrY* 1	2	3	*barA* 1	2	3
30		IMI 969	133			1514	1949	
31	**HA-ABU**	# 1311	133			1376	1939	
32	**FI**	# 2051	133			1999		
33		# 2053	133			1979		
34		# 2056	331					
35		# 2063				2425	2452	
36		# 2067	133			2718		
37		# 2068	25	133		1514	1949	
38	**UTI**	IMI 929	133	380	505			
39		# 516	133			1949		
40		# 1339	133			1193		
41		# 1345	130	133		1514		
42		# 1348	133			1053	1514	1949
43		# 1354	47					
44		# 1528	133	575				
45		# 1529				1514	2110	
46		# 1535	133	617				
47		# 1539	133			398		

4.8 Phenotypic characterization

To investigate whether the detected *uvrY*/ *barA* alleles have an impact on the phenotypes of the bacteria direct genotype-phenotype correlation studies were applied on selected bacterial strains listed in Table 20. The phenotypic characterization of bacteria, which includes a broad spectrum of different assays with either descriptive or discriminative power, aims to assess different strain features and constitutes an essential part for their discrimination into different species and subspecies in general. Therefore a classification of a bacterium is not possible without detailed studies on its morphology as well as its physiological and/or pathogenic features (Garrity *et al.*, 2001).

4.8.1 Biofilm

Biofilms are multicellular bacterial communities that attach to different surfaces which are held together by extracellular substances (Costerton *et al.*, 1987, 1995). Microbes living in biofilms are marked by outstanding characteristics like extraordinary resistance to antimicrobial treatment or the immune response of the host (Sauer, 2003). Especially in the medical field biofilms have attracted a lot of attention recently because they are associated with various persistent and chronic bacterial infections (López *et al.*, 2010). The ability to form biofilm was compared for the strains of the different groups in LB and pooled human urine. For this purpose static cultures were grown in microtiter plates and stained with crystal violet (for details see 3.5.2). The results of the biofilm

formation test for the four groups of isolates are shown in Figure 29. The boxplot is a graphical representation of the key values from the summary statistics which illustrates the minimum and maximum values as well as the median and the 25th and 75th percentile. The whiskers used in the boxplot show the data between the 10th and 90th percentile. The boxplot distribution led to the following results: the groups showed similar mean values for the biofilm formation in urine and LB. The values in urine were 0.017 $\pm$ 0.042 for ABU isolates, 0.002 for the HA-ABU strain (only one isolate), 0.004 $\pm$ 0.006 for FI and 0.009 $\pm$ 0.008 for bacterial strains from UTI, respectively. For the biofilm formation in LB values of 0.09 $\pm$ 0.2 for ABU, 0.2 for the HA-ABU isolate followed by 0.06 $\pm$ 0.1 for FI and 0.1 $\pm$ 0.1 for UTI isolates were obtained, respectively. Significant differences between the four isolate groups with respect to biofilm formation ability could neither be detected in urine nor LB medium according to the statistical analysis using Kruskal-Wallis test (Kruskal & Wallis, 2012; Theodorsson-Norheim, 1986). Whereas significant differences ($p < 0.001$) were observed when the biofilm formation capability of each group was compared between the two media thus leading to the conclusion that the biofilm formation ability depends on the bacterial culture medium employed.

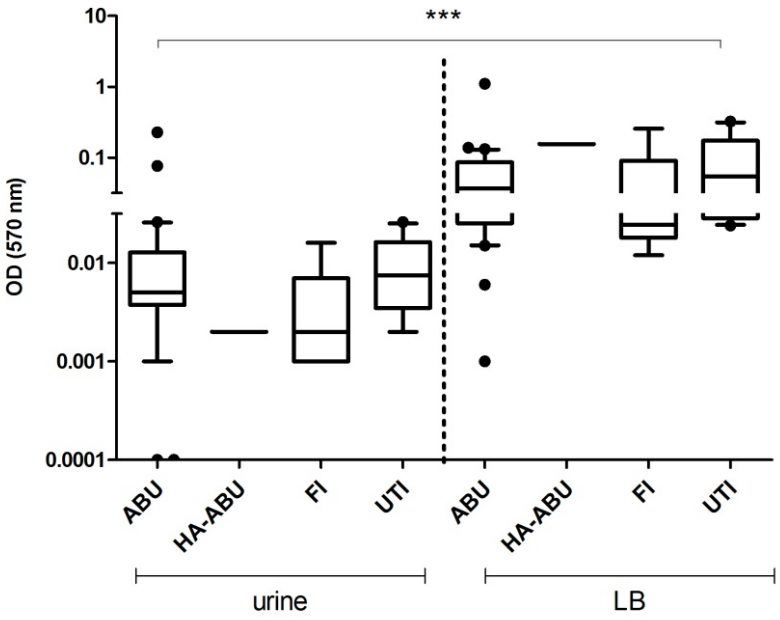


Figure 29: Box-Whisker-plot for the biofilm formation.

Box-Whisker-plots are shown for the four groups in pooled human urine and LB medium measured at 570 nm. Statistics were performed using the Kruskil-Wallis test. Outliers are depicted as dots (●).

The individual results of the biofilm formation assay for each of the tested strains are shown in the appendix (see 7.1.3). The comparison of the different media showed that the biofilm formation of the strains in urine with a few exceptions was very weak.

4.8.2 Synthesis of curli fimbriae and cellulose expression

Since biofilm formation is associated with the expression of curli fimbria and various exopolysaccharides like cellulose or colanic acid (Danese *et al.*, 2000; Uhlich *et al.*, 2006), the ability of the strains to express curli fimbria and/or cellulose was tested. The synthesis of curli fimbriae was analyzed by incubation on Congo Red (CR) agar plates for 96 h at RT, 30 °C and 37 °C. The different phenotypes reflected different expression patterns. The co-expression of curli fimbriae and cellulose led to red, dry and rough colonies designated *rdar* morphotype. Expression of only curli but no cellulose resulted in brown colonies on the CR plates whereas the expression of only cellulose led to pink colonies. Strains that were deficient in both, cellulose and curli expression exhibited the so-called *saw* morphotype characterized by smooth and white colonies (Bokranz *et al.*, 2005; Uhlich *et al.*, 2006). As an alternative approach, the synthesis of cellulose was tested by growing the strains on Calcoflour (CF) agar plates and testing them for fluorescence under UV light. Since the fluorescent dye calcoflour is bound by cellulose fluorescent colonies served as an indicator for cellulose production.

The results for the tested strains on CR are shown in Figure 30 in direct comparison to the results of the CF plates. A definitive assignment of the strains to the four defined morphotypes (*rdar*, *saw*, brown and pink) was challenging for some cases. *E. coli* strain # 1967, for example, showed a combination of different morphotypes at RT: although the main part of the colony was pink, a white-coloured outer ring could be detected. Nevertheless, for the majority of the strains the results for the analysis of curli and/or cellulose expression using CR plates correlated well with those of cellulose expression tested on CF plates. Strains that exhibited either *rdar* or pink colonies usually showed a fluorescent signal using UV light, while strains with a *saw* or brown morphotype did not produce a fluorescencent signal when grown on CF plates.

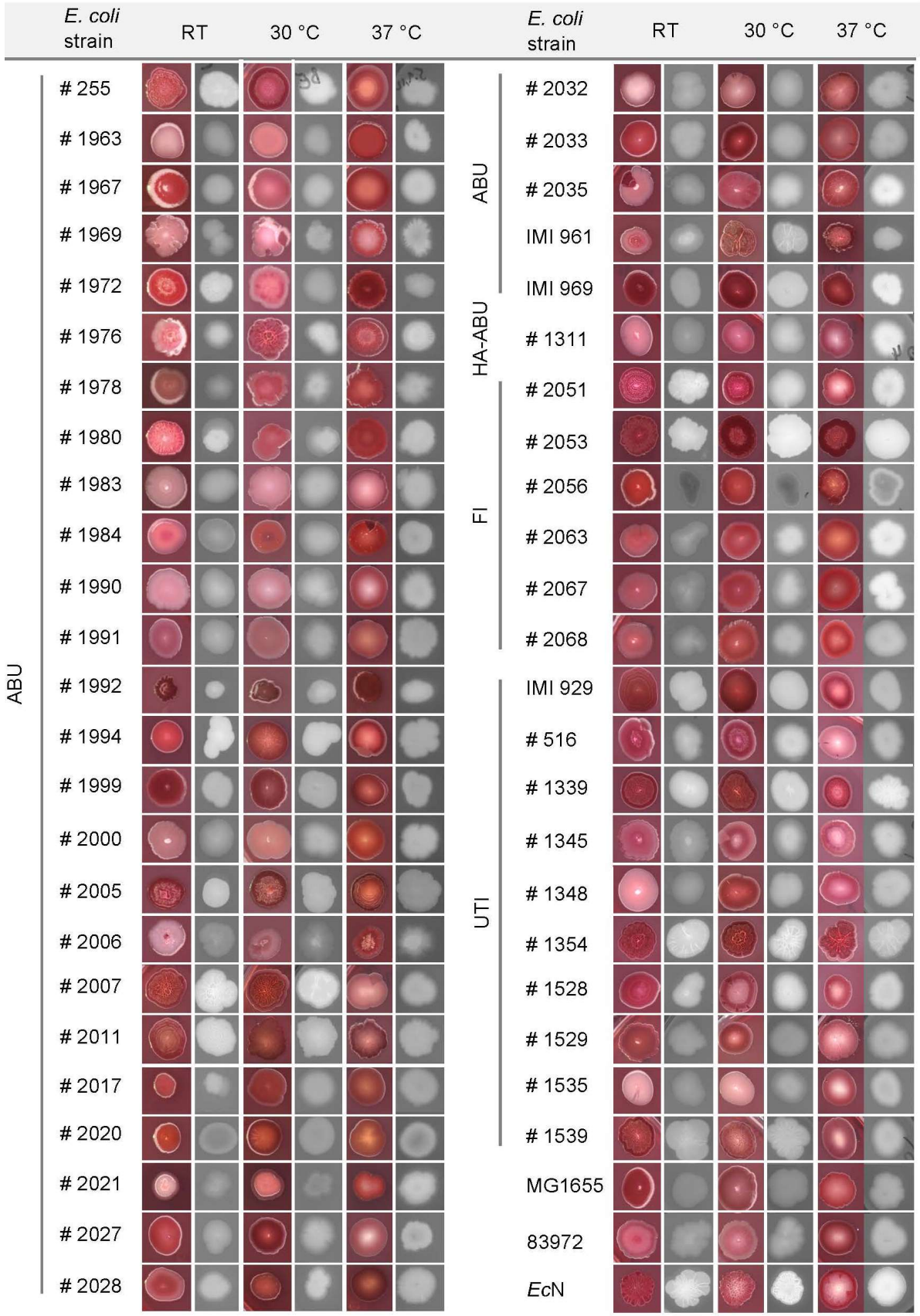


Figure 30: Phenotypes of the tested strains on Congo Red (CR) and Calcoflour (CF) agar plates.

All isolates were incubated for 96 h at RT, 30 °C and 37 °C. Results obtained by phenotypic analysis with CR are presented in red, whereas the results for the phenotypes with CF are shown in grey. *E. coli* strain Nissle 1917 (*Ec*N) was used as a positive control for the *rdar* morphotype on each plate.

4.8.3 Motility assay

The swarming ability of the strains was tested on LB motility agar plates as well as on urine motility agar plates (see 2.6). For this purpose bacterial material was stabbed on the plate and incubated overnight at 37 °C. After 16 h the migration zone was measured to assess the swarming properties of the strains. A characterization of the strains with regard to their swarming abilities was done according to a subjectively defined scheme as shown in Figure 32. Strains with swarming zones $\leq$ 0.5 cm were considered to be non-motile (-), whereas bacteria with migration zones between 0.5 cm and 4 cm were considered as medium motile (+). High motility (++) was assigned for bacteria with migration zones $\geq$ 4 cm. The results of the swarm motility assay are shown in Figure 32. Motility on LB was directly compared to motility on urine plates for all strains taken together. The analysis revealed that more than half of the strains (53 %) were highly motile on LB plates, while 33 % and 14 % of the bacteria were shown to be non-motile and medium motile, respectively. With 49 % most of the strains incubated on urine plates were assigned to medium motility, followed by a proportion of 33 % non-motile strains and strains with a high motility, which accounted for the smallest fraction with 18 % of the total. The detailed distribution of the different motility properties for the four groups is shown in Figure 31. Most ABU and HA-ABU strains exhibited medium motility in urine, while FI and UTI isolates were mainly non-motile. In LB the majority of all strains of the tested groups was highly motile.

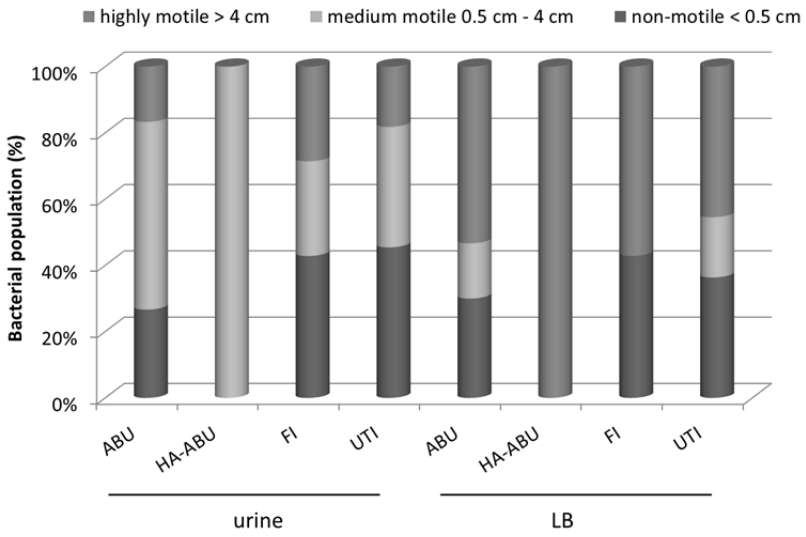


Figure 31: Percentile distribution of non-motile, medium motile and highly motile strains on urine and LB swarming agar.

The detailed analysis of the swarming characteristics for each of the four tested group is shown for the two tested media, LB and urine.

A table summarizing the results obtained with the phenotypic assays is shown in the Appendix (Table 27).

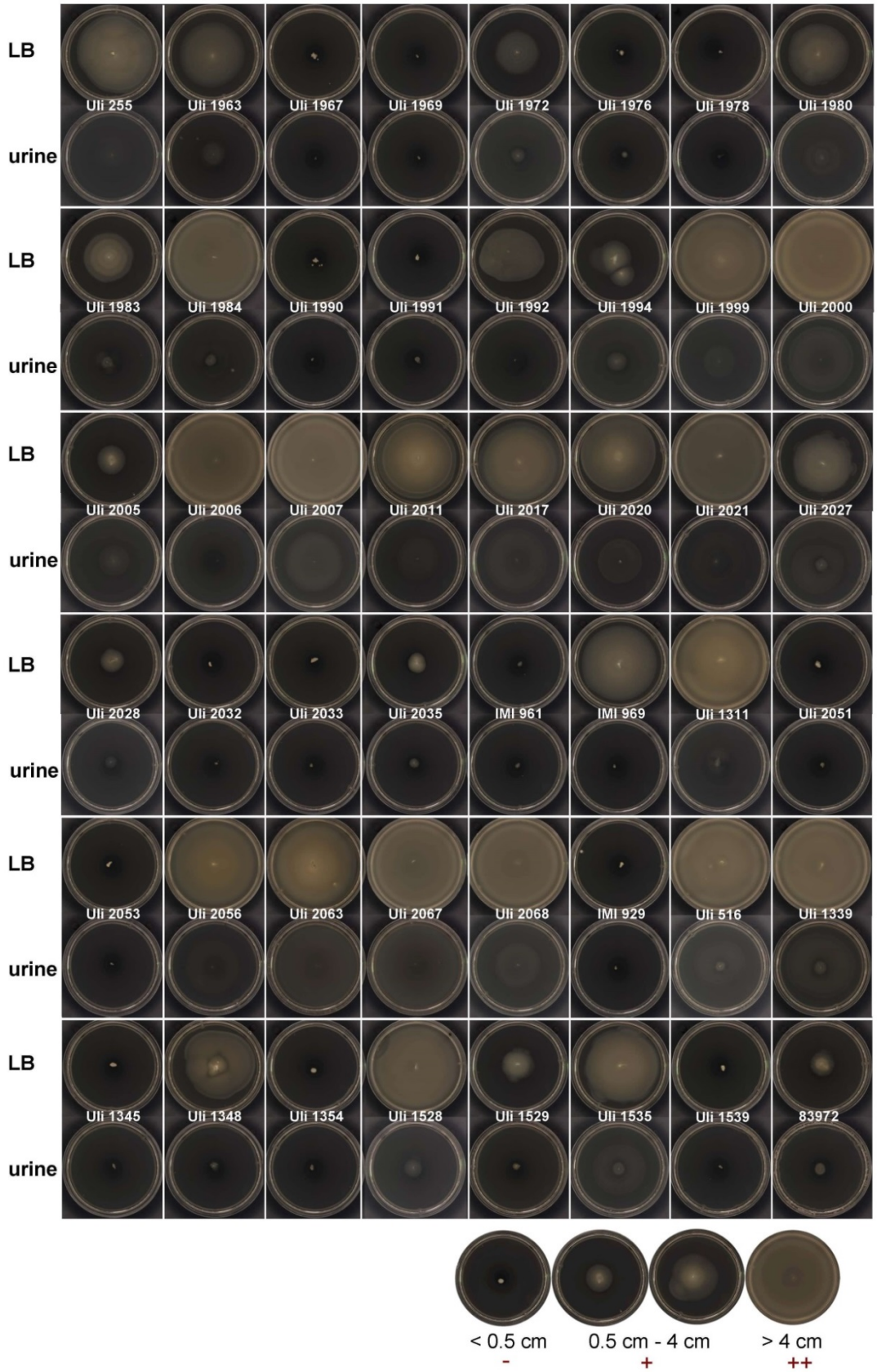


Figure 32: Swarming ability on LB and urine.

Migration zones of the tested *E. coli* strains on LB motility agar plates and urine motility agar plates incubated for 16 h overnight at 37 °C. The classification of motility was performed according to the diameter of the migration zone as shown in the bottom.

4.8.4 Growth characteristics

The different *E. coli* strains were also compared with respect to their growth kinetics. Growth was monitored - both in LB medium and pooled human urine - over a period of 24 h at 37 °C. LB cultures were incubated under shaking while growth in urine was performed under static conditions.

No striking differences could be detected between the studied strains when grown in LB. Figure 33 A shows the growth curves for selected strains as representatives for all investigated strains. Most of the strains showed a comparable growth behavior with entering the stationary phase after about four hours and reaching an average final OD_{595} of about 0.6. Only for some strains, such as strain # 2017, differences like a lower final end OD_{595} or a shortened lag phase as for strains # 1983/ # 1999, were observed.

In contrast to growth in LB, growth in urine revealed remarkable differences between the various strains as can be seen for selected strains in Figure 33 B. Initially, all strains showed exponential growth with similar growth kinetics and reached a plateau after about two to four hours. For the majority of the strains the exponential phase was followed by a longer lag phase, which in turn was followed by an increase of the bacterial population again, whereas some strains such as # 2017 did not show this second cell division phase and remained in stationary phase. Consequently, a high variance of the final ODs was detected ranging from 0.09 up to 0.30 thus implying that the strains did not grow as well as in LB (compared to Figure 33 A).

4.9 Expression of *csrB*

In order to evaluate a potential impact of the different mutations found in the SNP screening the transcript levels of *csrB* were analyzed upon growth in pooled human urine. Since transcriptional activation of *csrB* is directly controlled by *uvrY* (Suzuki *et al.*, 2002; Weilbacher *et al.*, 2003) *csrB* was a suitable target for the measurement of any effects due to the ns SNPs. The strains, including *E. coli* MG1655 as a reference, were grown statically at 37 °C until the individual mid-log phase was reached and RNA was extracted according to 3.4.1.1. Due to a shift of the cycle threshold values between the two biological replicates per sample tested the analysis was representatively done based on one biological sample and its technical triplicates.

In general, *csrB* expression was highly variable among the tested strains as shown in Figure 34. Five of the 30 ABU isolates as well as the HA-ABU strain showed higher transcript level of *csrB* compared to the calibrator strain *E. coli* MG1655, with the maximal value of 1.7 for strain # 1984. However, the majority of the ABU strains showed a down-regulation of the *csrB* expression. Since *csrB* is a direct

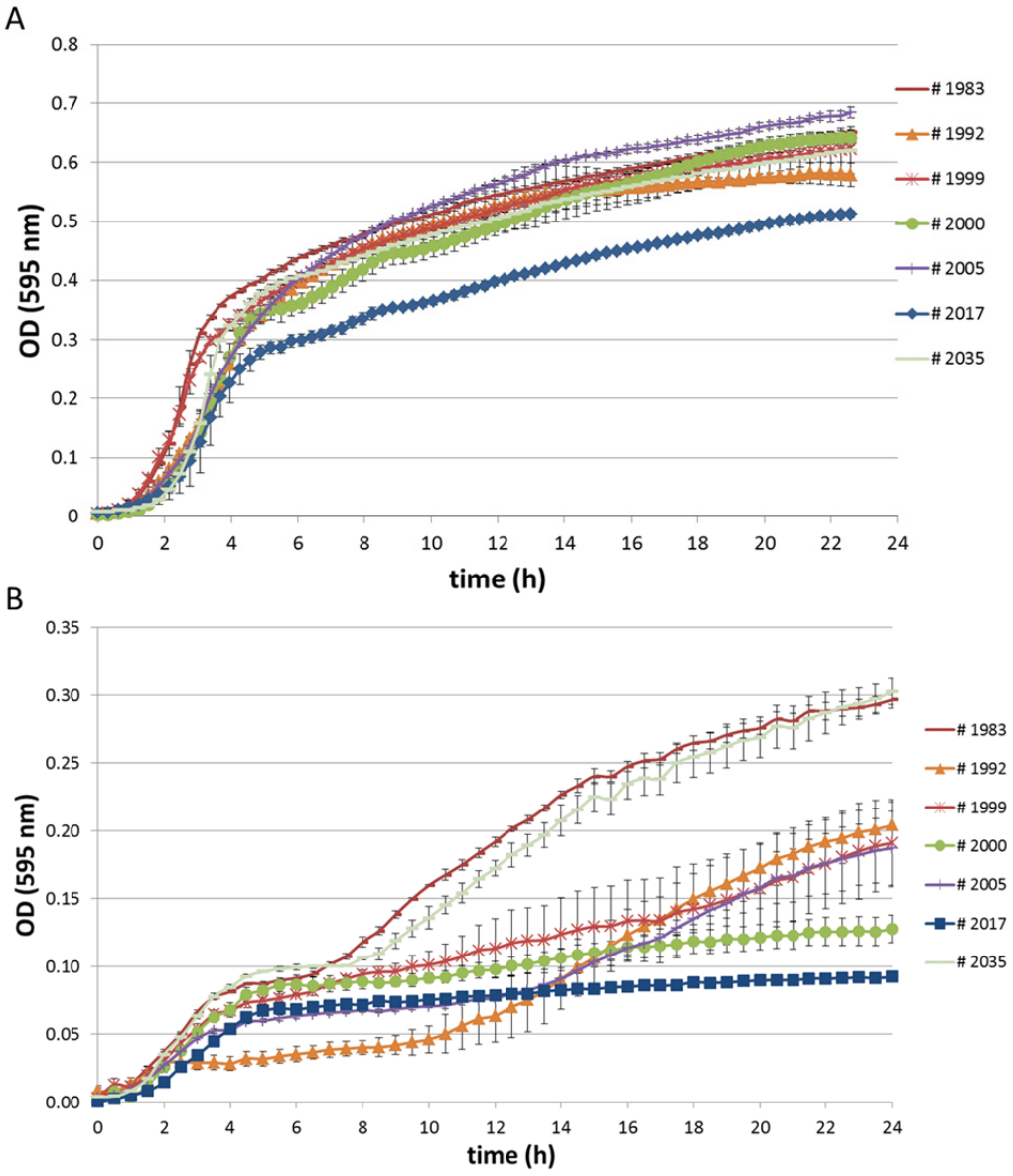


Figure 33: Growth characteristics of *E. coli* strains in (A) LB medium under aerobic conditions and (B) pooled human urine under static conditions at 37 °C.

Error bars represent the standard deviations of at least 3 biological replicates. Values are represented non-logarithmically. Growth was monitored by measuring the OD_{595} in a Tecan infinite F200 reader over a period of 24 h.

downstream target of *uvrY* the tested ns SNPs might have a potential impact on the transcript levels. Half of the UTI isolates exhibited an up-regulation of *csrB* transcription while the other half was down-regulated. Interestingly, a marked increase of the transcription level with a 6.2 fold up-regulation was detected for the UTI isolate # 1348, harboring a ns SNP in *uvrY* at nucleotide position 133 and three ns SNPs in *barA* at nucleotide positions 1053, 1514 and 1949.

However, it is important to mention that the different isolates display a heterogeneous genetic background. Consequently, in this instance, linking these expression results to the SNPs appearing in *uvrY* (or *barA*) was not possible.

In order to evaluate any possible impact of the SNPs found in the initial screening a more straightforward approach had to be employed. Therefore, we focused on *uvrY* which was proven to be under positive selection during growth in the bladder (see 4.6).

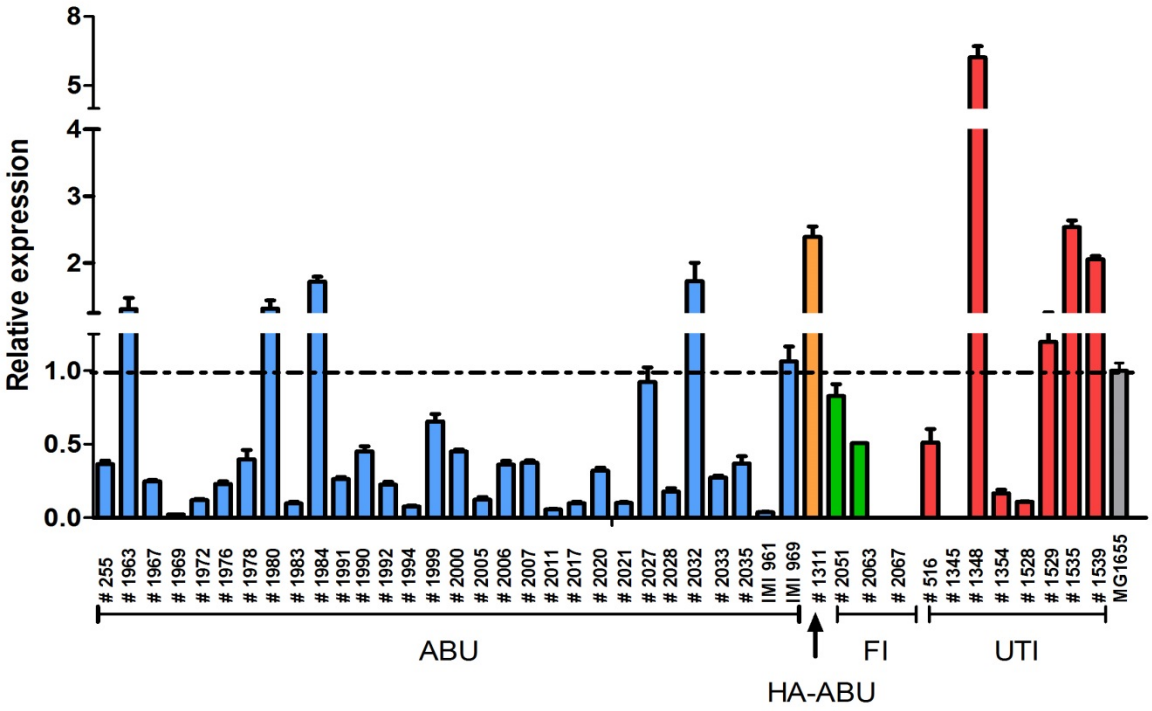


Figure 34: Real Time-based quantification of *csrB* transcripts in pooled human urine.

Gene expression was compared to that of *E. coli* MG1655. Gene expression was calculated according to the ΔC_t method. ABU strains are shown in blue, HA-ABU in orange, green depicts fecal isolates and red the UTI strains. The calibrator strain MG1655 is shown in grey. Data from three technical replicates are shown. Error bars represent the standard error of the mean (SEM).

4.10 Impact analysis of non-synonymous SNPs in *uvrY* – allelic variants

To be able to directly trace back observed phenotypic changes to any ns SNPs, we exchanged the *uvrY* with allelic variants that were found in the SNP screening into a genetically homogeneous strain background using the laboratory strain *E. coli* MG1655. For this purpose we markerless replaced the wild type *uvrY* first by a *sacB-cat* cassette and exchanged this again by the desired *uvrY* allele (see 3.3.2). The final mutant strains were then subjected to different functional assays in order to test the impact of the SNPs on protein function *in vitro*.

4.10.1 Selection of *uvrY* alleles

We selected *uvrY* allelic variants with various non-synonymous SNPs for the subsequent construction of the mutants. Based on the hypothesis that positive selection is acting on *E. coli* during long-term growth in the bladder which forces the adaptation to the host, only ABU isolates were considered for the selection of allelic variants. A criterion for the selection of a variant was the appearance of as many as possible ns SNPs at different positions of the protein. This should ensure a stronger expression of the phenotype. Since *E. coli* 83972 represents a well-characterized reference strain that is well-adapted to growth in urine we decided to carry this allele as well in the study. The seven finally selected *uvrY* alleles are listed in Table 21. The table includes the entire information of the respective strains including the different kinds of SNPs, their codon effects and the exact positions in the corresponding protein. Additionally the SNPs were grouped based on their respective conservation scores (see 7.1.2.1) into slow evolving (SE) and fast evolving (FE) positions. For a better overview a graphical illustration of *uvrY* and its conserved domains in the genomic context is shown very generally in Figure 35.

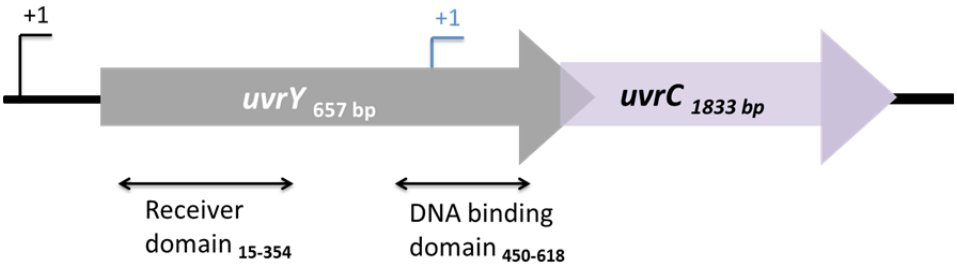


Figure 35: Genomic organization of *uvrY* and its conserved domains.

UvrY is shown in its genomic context. The transcriptional start sites are marked with +1. The conserved domains are indicated by arrows.

The selection process resulted in seven different allelic variants that should be exchanged with the native *uvrY* of *E. coli* MG1655. Among these variants were also two alleles that contained frequently

occurring synonymous SNPs (*E. coli* strain # 1980/ # 2031). These variants should serve as a negative control to be able to demonstrate that potential effects are only due to non-synonymous SNPs.

Table 21: Selected *uvrY* alleles for the exchange of native *uvrY* of *E. coli* MG1655.

SNPs and their codon effects as well as the position in the protein and the corresponding conservation scores are shown. SNP positions are further grouped on the basis of their conservation scores into fast evolving (FE) or slow evolving (SE) sites.

No	*E. coli* strain	Nucleotide position (codon, AA change Ref:Variant)	Codon effect	Protein region	Conservation score (evolutionary rate: FE,SE)
1	**# 1972**	123 (41; K:K)	silent	Rec domain	1.916
		133 (45; T:A)	missense	Rec domain	1.676 (FE)
		291 (97; G:G)	silent	Rec domain	-1.534
		384 (128; D:D)	silent	Inter-functional region	1.392
		517 (173; Q:*)	nonsense	DNA binding domain	1.362 (FE)
2	**# 1976**	117 (39; A:A)	silent	Rec domain	-1.093
		173 (58; P:L)	missense	Rec domain (intermolecular recognition site)	-1.44 (SE)
3	**# 1992**	93 (31; G:G)	silent	Rec domain	-1.154
		139 (47)	missense	Rec domain	1.818 (FE)
		186 (62; G:G)	silent	Rec domain	-1.62
		201 (67; R:R)	silent	Rec domain	-0.1144
		246 (82; T:T)	silent	Rec domain	-1.425
		315 (105; G:G)	silent	Rec domain	-1.23
		345 (115; I:I)	silent	Rec domain	-0.3461
		402 (134; A:A)	silent	Inter-functional region	1.917
		551 (184; Y:C)	missense	DNA binding domain (DNA binding residue)	-0.7821 (SE)
4	**# 2000**	123 (41; K:K)	silent	Rec domain	1.916
		133 (45; T:A)	missense	Rec domain	1.676 (FE)
		291 (97; G:G)	silent	Rec domain	-1.534
		384 (128; D:D)	silent	Inter-functional region	1.392
		589 (197; D:N)	missense	DNA binding domain (dimerization interface)	0.1268 (FE)
5	**83972**	123 (41; K:K)	silent	Rec domain	1.916
		133 (45; T:A)	missense	Rec domain	1.676 (FE)
		291 (97; G:G)	silent	Rec domain	-1.534
		384 (128; D:D)	silent	Inter-functional region	1.392
6	**# 1980**	117 (39; A:A)	silent	Rec domain	-1.093
7	**# 2031**	384 (128; D:D)	silent	Inter-functional region	1.392

4.10.2 Construction of *uvrY* allelic variants

In order to construct the *uvrY* allelic variants in the wild type strain *E. coli* MG1655 the native *uvrY* first had to be deleted. For this purpose *uvrY* was replaced by the *sacB-cat* cassette from plasmid pCVD422_cat. The exchange was performed via homologous recombination (see 3.3.1). Prior to the amplification of the entire *sacB-cat* cassette the *cat* gene was amplified from pKD3 and subsequently

inserted into pCVD422 via homologous recombination, resulting in the modified vector pCVD422_*cat* (see 3.3.2). Successful replacement of *uvrY* by *sacB-cat* was verified by PCR. The resulting transformant *E. coli* MG1655 Δ*uvrY* was tested for sensitivity to sucrose and resistance to chloramphenicol (see 3.3.2). Following this, the *sacB-cat* cassette of MG1655 Δ*uvrY* was changed by the desired *uvrY* alleles (Table 21) via homologous recombination. The counter selection system using *sacB-cat* provided the opportunity to check finally for the right recombinants being resistant to sucrose but sensitive to chloramphenicol again. The application of this system offered the advantage that the final recombinants were markerless. Thus, observed phenotypic changes can be directly attributed to the *uvrY* allele. In the end the recombinants were verified by sequencing. As a control to the selected *uvrY* variants the native *uvrY* was exchanged by the wild type *uvrY*, as well. This variant served as a control that should exclude any possible effects due to technical reasons. To be able to compare the effect of different *uvrY* alleles to the wild type in further analyses, the introduction of a selection marker into wild type was required. For this purpose a zeocine resistance cassette was transduced via P1 transduction (see 3.3.3) into wild type *E. coli* MG1655. A summary of the constructed *uvrY* allelic variants is shown in Table 22.

Table 22: Derivative of *E. coli* MG1655 with a zeocine resistance cassette and the constructed *uvrY* allelic variants used in this study.

No	*E. coli* strain (*uvrY* allelic variant)	characteristic
0	MG1655 WT^{Zeo}	Wild type strain containing a selection marker (Zeo^{R})
1	MG1655 Δ*uvrY*	*uvrY* complete deletion mutant containing a *sacB-cat* cassette
2	MG1655 *uvrY*::*uvrY* 1972	Allelic variant containing *uvrY* allele of strain # 1972
3	MG1655 *uvrY*::*uvrY* 1976	Allelic variant containing *uvrY* allele of strain # 1976
4	MG1655 *uvrY*::*uvrY* 1992	Allelic variant containing *uvrY* allele of strain # 1992
5	MG1655 *uvrY*::*uvrY* 2000	Allelic variant containing *uvrY* allele of strain # 2000
6	MG1655 *uvrY*::*uvrY* 83972	Allelic variant containing *uvrY* allele of strain 83972
7	MG1655 *uvrY*::*uvrY* MG1655	Control strain, reconstitution of native *uvrY*
8	MG1655 *uvrY*::*uvrY* 1980	Allelic variant containing *uvrY* allele of strain # 1980
9	MG1655 *uvrY*::*uvrY* 2031	Allelic variant containing *uvrY* allele of strain # 2031

Afterwards, to test the hypothesis that certain SNPs within *uvrY* are advantageous for the strains, their fitness was compared to that of the wild type strain.

4.10.3 Phenotypic assays

4.10.3.1 Biofilm

The ability to form biofilm was tested according to the method described in section 3.5.2 for the allelic variants in LB and pooled human urine. The results are shown in Figure 36. Interestingly, all of the tested strains showed the tendency to produce higher biofilms in pooled human urine compared to LB. This was also true for the wild type *E. coli* MG1655^{Zeo}. Looking at the results for *E. coli* MG1655 tested for biofilm formation in 4.8.1, the results obtained are vice versa. According to 4.8.1 the wild type showed higher biofilm formation in LB compared with urine. However, the standard deviations were pretty high. Therefore a marked difference between biofilm formation in LB and urine could not be detected.

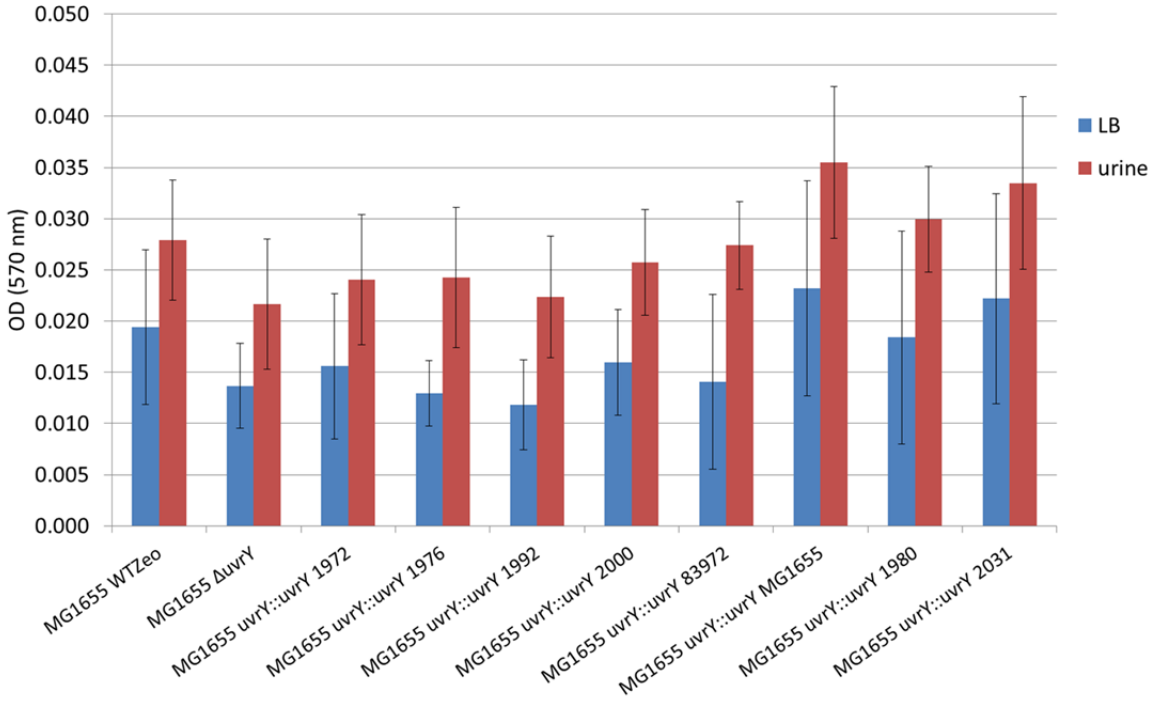


Figure 36: Biofilm formation in LB and pooled human urine.

The biofilm forming ability of the wild type strain and the allelic variants is shown. The error bars represents the SD of three biological replicates.

4.10.3.2 Expression of curli, cellulose & swarming motility

The expression of curli fimbriae and cellulose was tested for the wild type *E. coli* strain MG1655 as well as for the modified wild type strain MG1655^{Zeo} and the corresponding mutants analogous to the setup described above (see 4.8.2). As can be seen in Figure 37 and Table 23 the wild type strain, the wild type carrying the zeocine resistance and the allelic variants exhibited the same characteristics on CR and CF plates for two of the three tested temperatures, namely RT and 30 °C. At these temperatures all strains appeared as brown morphotype implying that only curli fimbriae but not

cellulose are expressed. This conclusion is supported by the results of the corresponding CF plates, where bacteria did not show any fluorescence. However, in some cases a little fluorescent spot was visible in the center of the colony indicating that the colonies did not always exhibit a homogeneous morphology. Therefore, it can be concluded that a combined expression of curli and cellulose was restricted to bacterial cells located in the center of the colonies. Interestingly, at 37 °C all of the tested strains except for the allelic variants containing the syn SNPs [no. 8/9], displayed a pink/brown intermediate phenotype on CR plates. Allelic variants no. 8 and 9 instead remained with the brown morphotype. For all of the allelic variants a light, but at least increased signal in fluorescence could be observed on CF plates compared to both, the wild type strain as well as the zeocine-resistant wild type strain. In conclusion the *uvrY* allelic variants showed a stronger curli and cellulose expression than the wild type strains at 37 °C.

The motility of the strains was tested for both, LB and urine according to 3.5.3. The classification of motility was performed according to the radius of the migration zone with zones $\leq$ 0.5 cm as non-motile (-), migration zones between 0.5 cm and 4 cm being medium motile (+) or highly motile if the swarming zone was bigger than $\geq$ 4 cm. As shown in Figure 37 and Table 23 it can be summarized that all of the strains exhibit medium motility (+) on LB swarm agar plates. The same can be observed for the majority of the strains on urine swarm agar plates. Only *E. coli* strain MG1655 *uvrY*::*uvrY* 83972 and the two variants containing the syn SNPs exhibited reduced biofilm forming ability and were classified as non-motile. Due to the fact that the defined classification of medium motility contains migration zones between 0.5 cm and 4.0 cm, a differentiated consideration of the swarming ability based on this classification was not possible.

Taking a more detailed look at the plates the following can be concluded for the swarming on LB: *E. coli* strain MG1655 *uvrY*::*uvrY* 83972 showed the smallest migration zones with an average value of 0.6 cm suggesting that this strain is almost non-motile. Therefore the allele of the ABU strain 83972 shows a poor swarming ability on LB, which is a well described phenotype of *E. coli* strain 83972 (Roos *et al.*, 2006). The strain with the highest swarming zones on LB was MG1655 *uvrY*::*uvrY* 1992.

Regarding the swarming on urine it can be mentioned that (i) all of the strains show low values near to non-motility and (ii) the strain MG1655 *uvrY*::*uvrY* 83972 as well as the two variants containing the syn SNPs [strain no. 8 and no. 9; Figure 37] are considered as non-motile.

The individual results are also summarized for each strain in the Appendix (see 7.1.3).

Finally the data demonstrate that nature is capable of modulating the swarming ability dependent on the environment.

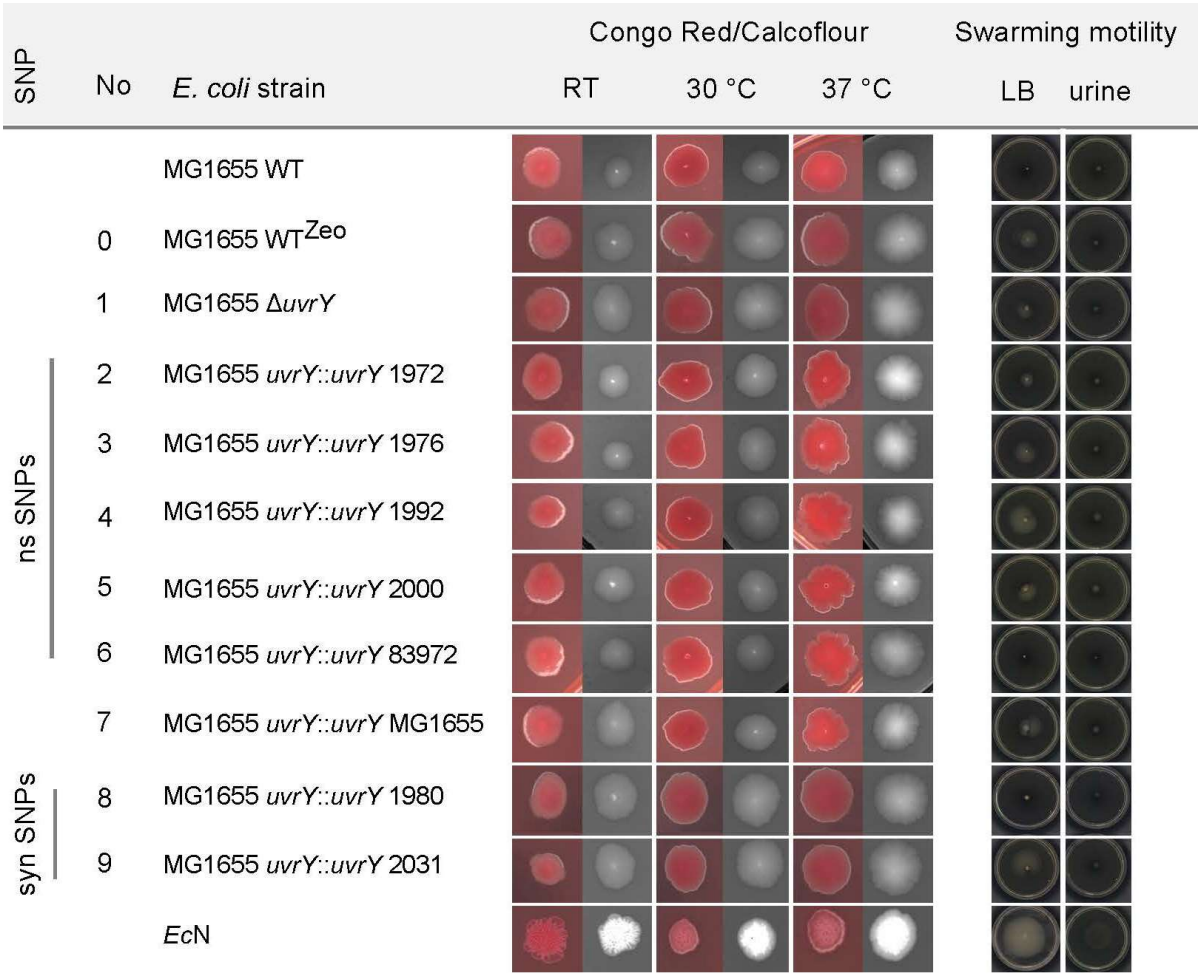


Figure 37: Curli fimbriae/cellulose expression and swarming motility of *E. coli* MG1655$^{(Zeo)}$ and the allelic variant strains.

Expression of curli and cellulose was investigated by incubation on Congo Red (CR) and Calcoflour (CF) agar plates. Swarming motility was analyzed by determining the migration zones on either LB or urine swarming agar plates. As a positive control for curli and cellulose expression *E. coli* Nissle 1917 (*Ec*N) was used.

Table 23: Summary of the results of the CR/CF plates and the swarm agar plates.

No	*E. coli* strain	CR morphotypes			CF phenotypes			Motility	
		RT	30 °C	37 °C	RT	30 °C	37 °C	LB	urine
	MG1655 WT	b	b	p/b	-	-	-	+	+
0	MG1655 WT^{Zeo}	b	b	p/b	-	-	-	+	+
1	MG1655 Δ*uvrY*	b	b	p/b	-	-	(+)	+	+
2	MG1655 *uvrY*::*uvrY* 1972	b	b	p/b	-	-	(+)	+	+
3	MG1655 *uvrY*::*uvrY* 1976	b	b	p/b	-	-	(+)	+	+
4	MG1655 *uvrY*::*uvrY* 1992	b	b	p/b	-	-	(+)	+	+
5	MG1655 *uvrY*::*uvrY* 2000	b	b	p/b	-	-	(+)	+	+
6	MG1655 *uvrY*::*uvrY* 83972	b	b	p/b	-	-	(+)	+	-
7	MG1655 *uvrY*::*uvrY* MG1655	b	b	p/b	-	-	(+)	+	+

No	*E. coli* strain	CR morphotypes			CF phenotypes			Motility	
		RT	30 °C	37 °C	RT	30 °C	37 °C	LB	urine
8	MG1655 *uvrY*::*uvrY* 1980	b	b	b	-	-	-	+	-
9	MG1655 *uvrY*::*uvrY* 2031	b	b	b	-	-	-	+	-
	Nissle 1917 (*Ec*N)	*rdar*	*rdar*	*rdar*	+	+	+	++	+

4.10.3.3 Growth characteristics

Growth properties of the wild type strain MG1655^{Zeo} and the created *uvrY* allelic variants were checked by cultivation in the Tecan infinite F200 reader according to 3.5.1. In LB all strains grew well with respect to growth properties and the final optical density (Figure 38 A). A transition into the stationary phase was observable after four hours. For the different *uvrY* alleles we did not observe any differences regarding their growth characteristics in LB. The growth behavior when grown in urine, however, was rather different (Figure 38 B). In contrast to LB striking differences were observable. The individual growth phases did not resemble the standard characteristics of a growth curve, in fact all strains showed a modified growth curve. This curve included the transition into a first stationary phase after approximately one to two hours of growth followed by a second increase of bacterial cell density after 12 hours. The final OD_{595} values were between 0.1 and 0.15, depending on the strain. Interestingly this pattern was observable for all tested variants including the wild type strain. Since it is known that *uvrY* is involved in the carbon storage regulation (Pernestig *et al.*, 2003; Suzuki *et al.*, 2002) and that nutrients are limited in urine, which forces bacteria to adjust their metabolism, the observed growth differences in urine provided first evidence for altered adaptation due to different *uvrY* variants.

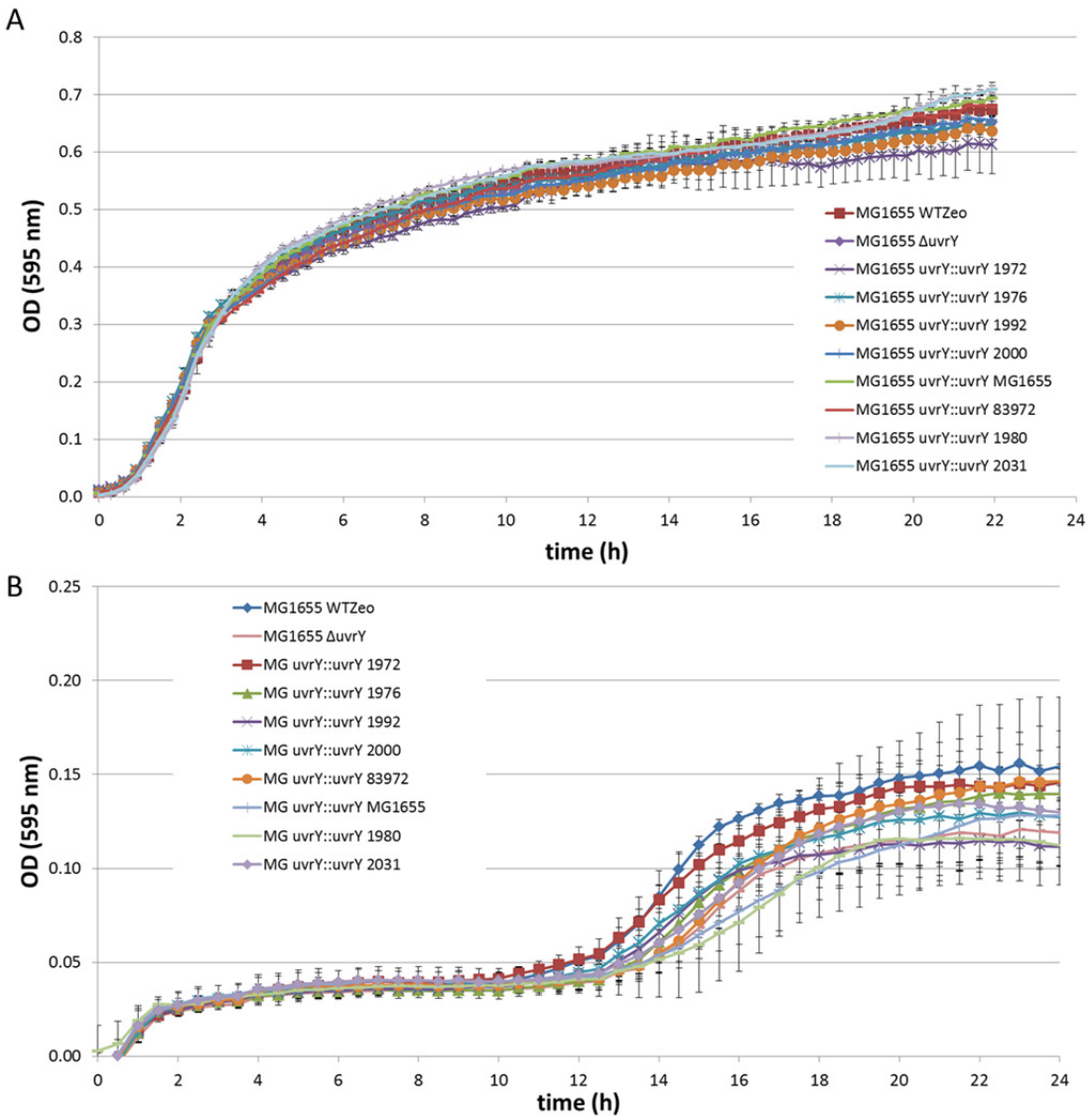


Figure 38: Growth kinetics of the MG1655^{Zeo} wild type and the variant strains in (A) LB medium under aerobic conditions and (B) pooled human urine under static conditions at 37 °C.

Error bars represent the standard deviations of at least three biological replicates. Values are presented on a non-logarithmic scale.

4.10.3.4 Competition assays

Fitness differences often cannot be monitored very well in single growth curves (Figure 38). Due to this competition experiments were carried out in LB and pooled human urine. Competition assays are the most direct approach to show, if a mutation affects the fitness compared to a wild type strain. For this purpose the wild type control strain MG1655 WT^{Zeo} and one of the mutant strains were used in pair-wise competition. Appropriate culture volumes were plated at defined times on selective and plain agar plates (see 3.5.6). Based on the CFUs/ml the percentage ratio representing the relative fitness of the strains was calculated.

To get an overview of the general fitness and to see in which stage of the competition differences can be observed, a competition experiment was done over a time scale of 24 hours. The first graph (Figure 39) shows the results of the competition experiments for all tested strains performed in LB. All of the graphs clearly illustrate that the comparative fitness changed during growth. In most of the cases (Figure 39 B-F, H-I) the wild type strain *E. coli* MG1655WT^{Zeo} revealed a fitness advantage in the beginning of the assay. For the majority of the samples this advantage of the WT strain was nevertheless followed by an out-competition by the competitor strain which tended to have a higher fitness in the end (Figure 39 A-E). The ratio for the competition of MG1655WT^{Zeo} vs. MG1655 *uvrY*::*uvrY* 83972 (Figure 39 F) was after 24 hours nearly 50 %. The ratios for the control strain MG1655 *uvrY*::*uvrY* MG1655 (Figure 39 G) and the allelic variants MG1655 *uvrY*::*uvrY* 2031 as well as MG1655 *uvrY*::*uvrY* 1980 (Figure 39 H/I) were even below 50 %.
Thus the data suggest that (i) the allelic variants containing a ns SNP (B-D) had an effect on the fitness and in contrast (ii) the syn SNPs did not show an advantage for the mutant strains compared to the *E. coli* strain MG1655WT^{Zeo}.

The results for the competition in pooled human urine are shown in Figure 40. Regarding the fitness a similar pattern can be observed as for the competition in LB. The fitness seemed to change over time, as well, but the wild type strain *E. coli* MG1655WT^{Zeo} did not reveal any striking fitness advantages at the beginning of the experiment except during the competition with MG1655 *uvrY*::*uvrY* 1976 (Figure 40 C) and MG1655 *uvrY*::*uvrY* 2031 (Figure 40 I). In this case a clear advantage could be observed. Nevertheless some of the allelic variants (Figure 40 B-C, E-G) showed an out-competion of the *E. coli* MG1655WT^{Zeo} strain after 24 hours of growth. The other variants, namely MG1655 *uvrY*::*uvrY* 1992 (Figure 40 D), MG1655 *uvrY*::*uvrY* 1980 (Figure 40 H) and MG1655 *uvrY*::*uvrY* 2031 (Figure 40 I) showed a decreased fitness after 24 hours with percental values below 50 %. The *uvrY* deletion mutant MG1655 Δ*uvrY* showed with 49 % a lower fitness than the WT strain, as well. However, the differences were only marginal. In summary we observed that the results of a fitness advantage were not as clear as in LB, but nevertheless it seemed that the variants were somehow fitter after 24 hours of competition compared to the WT strain.
Based on these findings we wanted to address the question, if the determined differences can be reproduced and if they are significant. Therefore we investigated a possible fitness advantage of the mutants in more detail. For this purpose a total of five biological replicates were tested for their fitness at time point 24. To test for statistical significance the paired Wilcoxon test was used (Wilcoxon, 1945). The results for LB (A) and urine (B) are shown in Figure 41. The statistical analysis underlines the findings that have been described before.

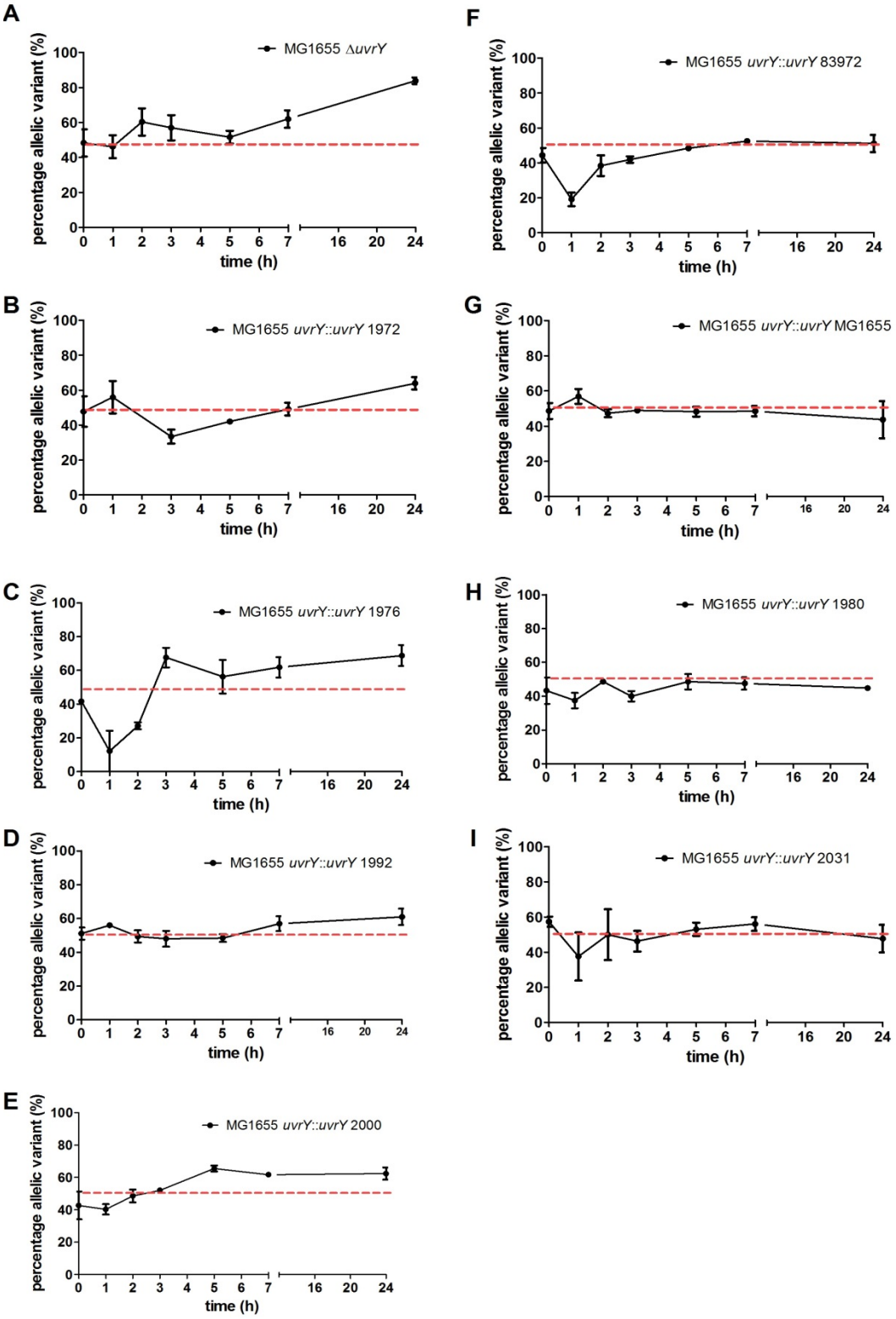


Figure 39: Competition of the *E. coli* MG1655WT^{Zeo} vs. the different allelic variants in LB for 24 hours.

The graphs show the percentage values of the allelic variants in competition with the *E. coli* MG1655WT^{Zeo} grown in LB. The results are shown for (A) MG1655 Δ*uvrY*, (B) MG1655 *uvrY*::*uvrY* 1972, (C) MG1655 *uvrY*::*uvrY* 1976, (D) MG1655 *uvrY*::*uvrY* 1992, (E) MG1655 *uvrY*::*uvrY* 2000, (F) MG1655 *uvrY*::*uvrY* 83972, (G) MG1655 *uvrY*::*uvrY* MG1655, (H) MG1655 *uvrY*::*uvrY* 1980 and (I) MG1655 *uvrY*::*uvrY* 2031. Error bars represent the standard error of the mean (SEM) from three biological replicates. At least two biological replicates were tested for MG1655 *uvrY*::*uvrY* 1976, MG1655 *uvrY*::*uvrY* 1992, MG1655 *uvrY*::*uvrY* 2000 and MG1655 *uvrY*::*uvrY* 1980 at t_0, t_2, t_3 and t_7. t_2 was not tested for MG1655 *uvrY*::*uvrY* 1972.

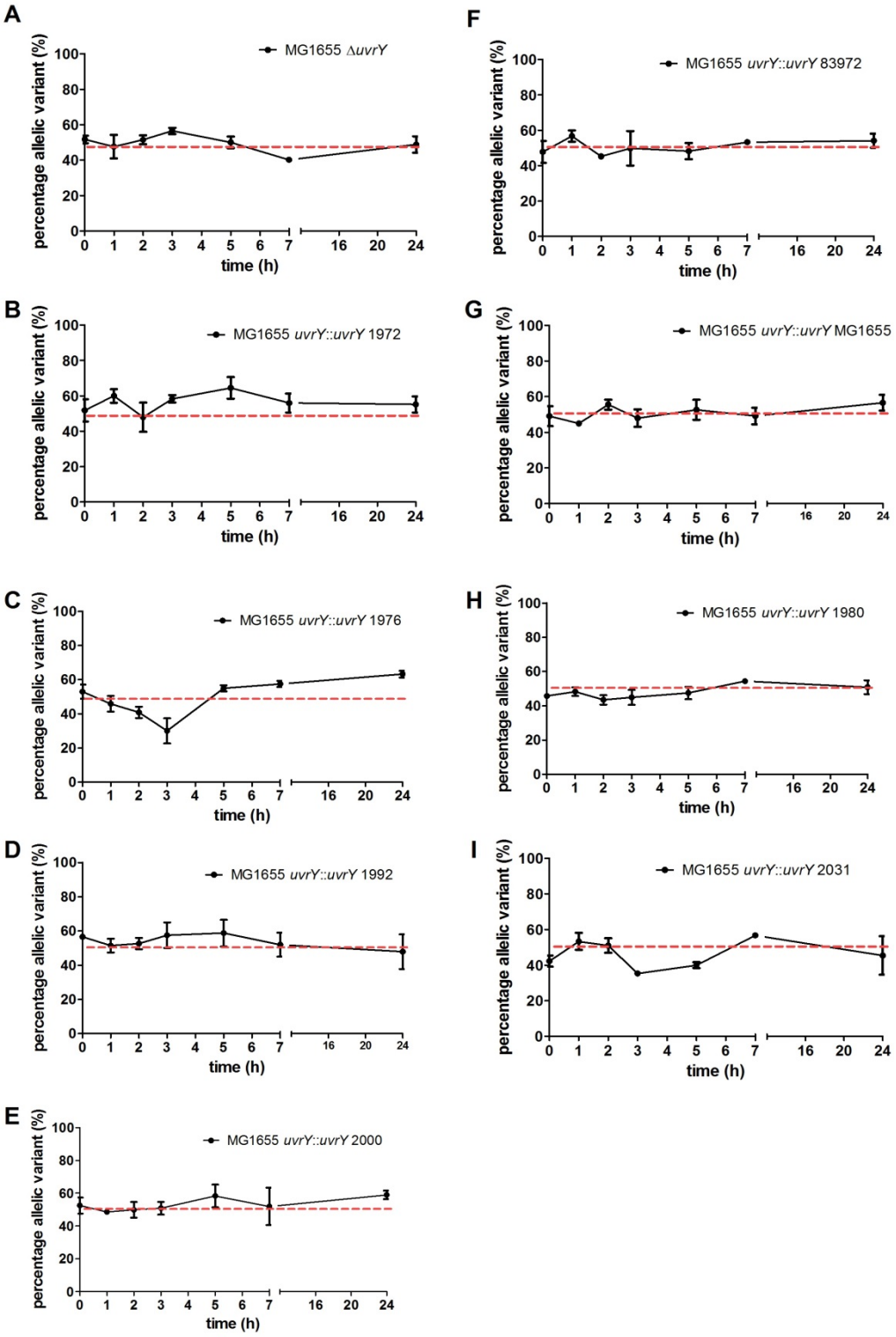


Figure 40: Competition of the *E. coli* MG1655WT^{Zeo} vs. the different allelic variants in urine for 24 hours.

The graphs show the percentage values of the allelic variants in competition with the *E. coli* MG1655WT^{Zeo} grown in urine. The results are shown for (A) MG1655 Δ*uvrY,* (B) MG1655 *uvrY*::*uvrY* 1972, (C) MG1655 *uvrY*::*uvrY* 1976, (D) MG1655 *uvrY*::*uvrY* 1992, (E) MG1655 *uvrY*::*uvrY* 2000, (F) MG1655 *uvrY*::*uvrY* 83972, (G) MG1655 *uvrY*::*uvrY* MG1655, (H) MG1655 *uvrY*::*uvrY* 1980 and (I) MG1655 *uvrY*::*uvrY* 2031. Error bars represent the standard error of the mean (SEM) from three biological replicates. For MG1655 *uvrY*::*uvrY* 1992 only two biological replicates were tested at t_1. The same is true for MG1655 *uvrY*::*uvrY* MG1655 at t_0 and for MG1655 *uvrY*::*uvrY* 1980 at t_7 and t_{24}.

All of the *uvrY* allelic variants containing ns SNPs as well as the *uvrY* deletion mutant MG1655 Δ*uvrY* exhibited a higher fitness during growth in LB (Figure 41 A) compared to the variants with the synonymous SNPs, namely MG1655 Δ*uvrY*::*uvrY* 1980 and MG1655 Δ*uvrY*::*uvrY* 2031, and the control strain MG1655 Δ*uvrY*::*uvrY* MG1655.

Nearly the same could be observed for the competition in urine (Figure 41 B). All variants except for the MG1655 Δ*uvrY*::*uvrY* 83972, the control strain MG1655 Δ*uvrY*::*uvrY* MG1655 and the ones containing the syn SNPs [strains no. 8 and no. 9, Figure 41], were predominant in the bacterial population after 24 hours of growth in urine which is confirmed by the statistics. These data indicate that the introduced mutations were advantageous for the adaptation to growth in urine. Interestingly the *uvrY* deletion mutant MG1655 Δ*uvrY* showed a very significant value with $p < 0.01$, as well. Based on the results for the *uvrY* deletion mutant MG1655 Δ*uvrY* one can speculate that an evolution towards loss of *uvrY* is beneficial for the adaptation to growth in the urinary tract.

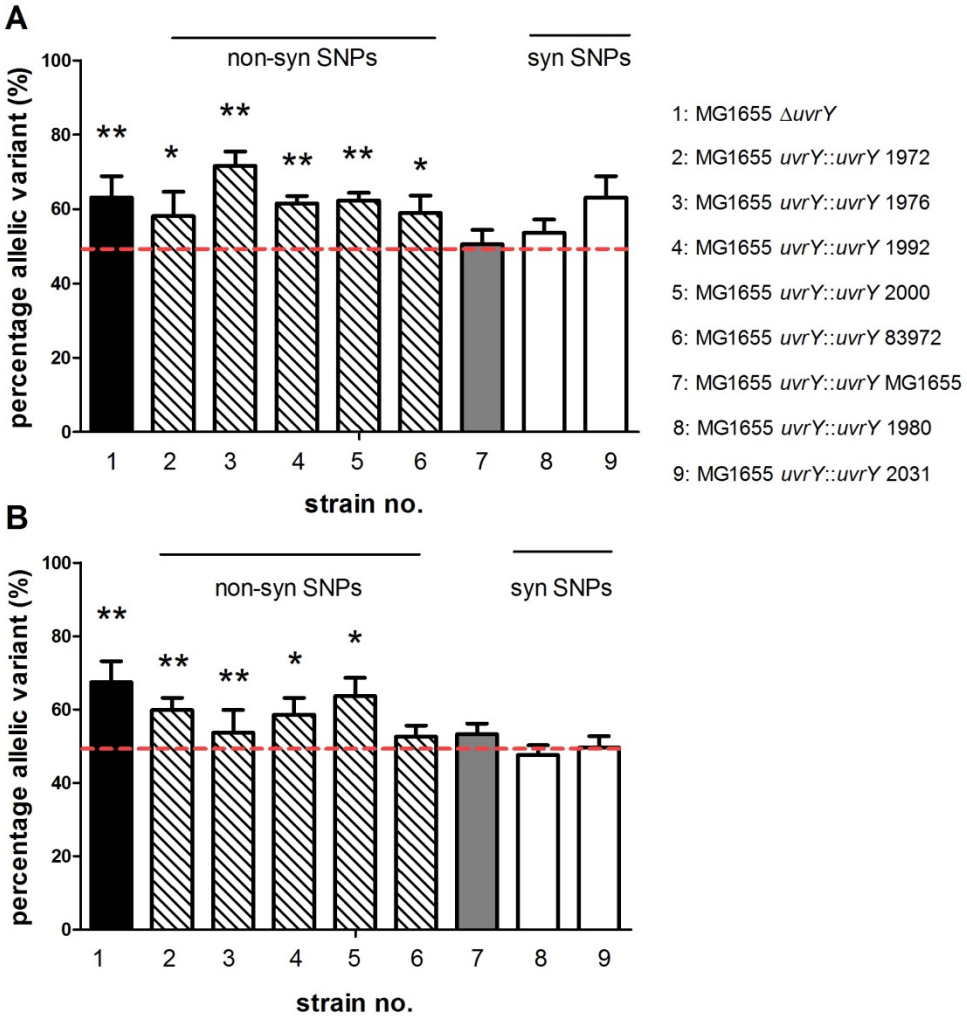


Figure 41: Results of the competition assays in (A) LB and (B) pooled human urine.

The ratio of variant strain (%) compared to the wild type MG1655^{Zeo} is shown for each allelic variant after 24 h of competition. The bars represent the different variants according to the numeration in Table 23. Error bars represent the SEM from five biological replicates (BR) (except for variant 7 in LB, only 4 BR were done). Significant values were evaluated using the paired Wilcoxon test with $p < 0.01$ (**) and $p < 0.05$ (*).

Based on these findings a sub-cultivation assay was done to test if the *uvrY* allelic variants were able to retain the proven fitness advantage and to test if they could even increase their fitness during long-term growth.

4.10.3.5 Sub-cultivation assay

The sub-cultivation was performed according to 3.5.7 over a time period of 96 hours. For this purpose an aliquot of a mixed culture containing the wild type and the competitor strain in a 1:1 ratio was inoculated in fresh medium every 24 hours as shown in Figure 42. An appropriate culture volume was withdrawn and plated to be able calculate the percentage ratios representing the fitness of the strains on the basis of the CFUs/ml.

The results of the sub-cultivation assay after 96 hours for LB and urine are shown in Figure 43. Based on the results of the competition assays in LB a higher fitness would be expected for the allelic variants 1 to 6. Looking at the results (Figure 43 A) this can be confirmed. Surprisingly the allelic variants containing a syn SNP [strains no.8 and no. 9] were significantly fitter than the wild type strain, as well.

The results for the sub-cultivation in urine could confirm the results of the competition assay (Figure 43 B). The variants no. 1, 2, 4 and 5 remained significantly different from the wild type strain even after 96 hours of competition. Variant number 3 showed the same characteristics as in the competition assay (Figure 41 B), however not statistically significant in this experiment. The fitness of the remaining variants can be reproduced and consequently confirmed by long-term sub-cultivation in urine.

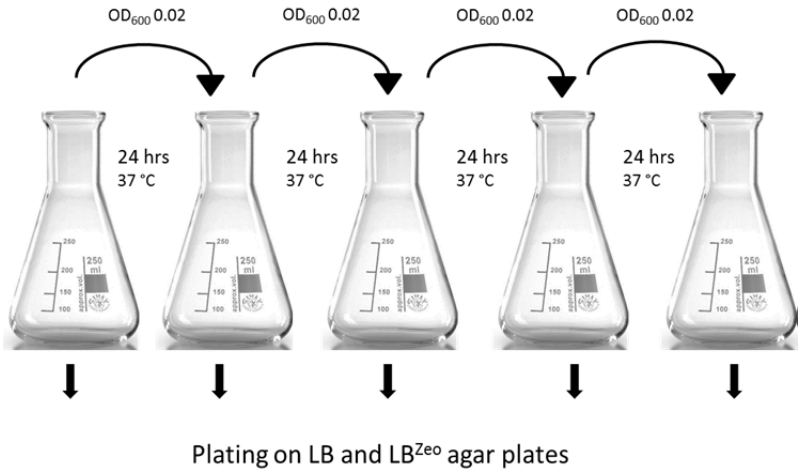


Figure 42: Scheme of sub-cultivation assay.

A fresh culture volume was inoculated every 24 h to an OD_{600} of 0.02. An aliquot was plated to be able to calculate the ratio of wild type and competitor strains at the different time points.

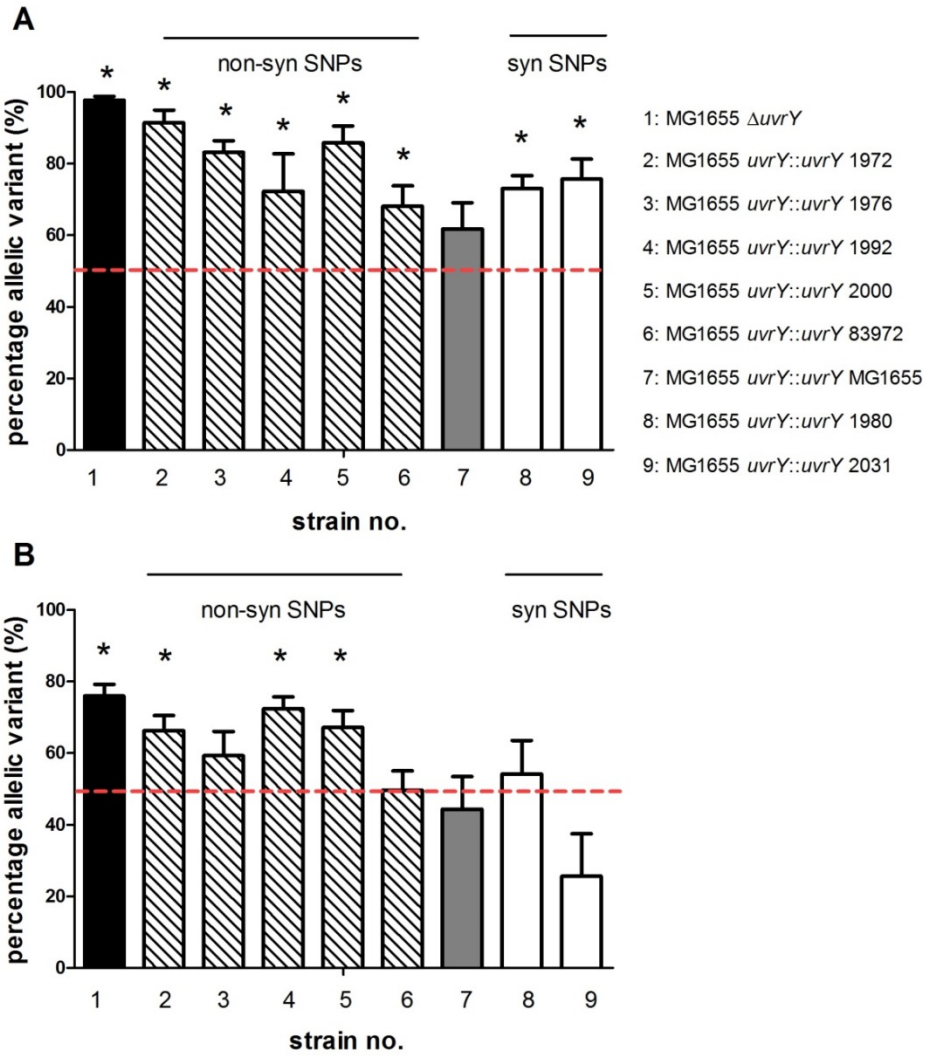


Figure 43: Fitness of the variant strains after 96 h of sub-cultivation in (A) LB or (B) urine.

Error bars represent the SEM of at least 3 biological replicates except for variant no. 9. For this one only two biological replicates were available. Significant values were evaluated using the paired Wilcoxon test with $p < 0.01$ (**) and $p < 0.05$ (*).

4.10.4 Real-Time PCR

In order to analyze a potential impact of the different *uvrY* allelic variants on the transcription levels of genes involved in the carbon storage regulation (Csr) system, we compared the relative expression levels of *csrB*, *csrC*, *csrD*, *csrA* and *uvrY* in a set of isogenic strains carrying the different *uvrY* alleles upon growth in LB and pooled human urine. The *E. coli* strain MG1655 *uvrY*::*uvrY* MG1655 served as a reference and the expression levels were normalized to the housekeeping gene *frr*.

The data for LB (Figure 44) suggest a direct impact of the ns SNPs of variant no. 2 to no. 6 on the expression level of *csrB*, which was significantly down-regulated in each case (Figure 44 A). The *csrB* transcriptional levels did not differ significantly in between the reference strain [no. 1] and variant

no. 7 to no. 8, suggesting that the *uvrY* allelic variant of MG1655 *uvrY*::*uvrY* 83972 as well as the variant no. 8 containing the ns SNP did not have an effect on the direct expression of *csrB*. Interestingly, we observed a significant 2-fold up-regulation of *csrB* in MG1655 *uvrY*::*uvrY* 2031 [no. 9], which carries a syn SNP variant of *uvrY*.

Regarding the expression levels of *csrC* (Figure 44 B) and *csrD* (Figure 44 C) no significant differences were observed when compared to the reference strain *E. coli* MG1655 *uvrY*::*uvrY* MG1655, except for allelic variant no. 5 that showed a down-regulation of *csrD*.

As shown in Figure 44 D, CsrA was significantly down-regulated in the *uvrY* deletion mutant [no. 2], as well as in MG1655 *uvrY*::*uvrY* 1976 [no. 4]. In contrast to that, a significant up-regulation was observed in one variant harboring a syn SNP, namely MG1655 *uvrY*::*uvrY* 2031 [no. 9].

As expected no *uvrY* was detected for the *uvrY* deletion mutant MG1655 Δ*uvrY* [no. 2]. For the remaining strains no significant differences were observed.

The data for the expression levels of *csrB*, *csrC*, *csrD*, *csrA* and *uvrY* in pooled human urine are shown in Figure 45. As in LB, all of the tested allelic variants showed a significant decrease in the transcription level of *csrB* when compared to the reference strain (Figure 45 A). Interestingly, this was also the case for the allelic variants no. 7 to no. 8 that did not show significant down-regulation of *csrB* in LB.

For *csrC* a significant down-regulation was detected in the *uvrY* deletion mutant [no. 2], in MG1655 *uvrY*::*uvrY* 1976 [no. 4], as well as in both strains containing the syn SNPs [no. 8/9] (Figure 45 B).

As shown in Figure 45 C a significant up-regulation of *csrD* was observed for variant MG1655 *uvrY*::*uvrY* 1972 [no. 3], whereas a down-regulation was shown for allelic variant no. 4.

Interestingly *csrA* was down-regulated in all allelic variants, except for the variant carrying the *E. coli* 83972 *uvrY* allele [no. 7] and variant no. 8 that contains a *uvrY* allele with a syn SNP (Figure 45 D).

As expected no signals for *uvrY* were detected in the *uvrY* deletion mutant MG1655 Δ*uvrY* [no. 2] (Figure 45 E) while the other strains showed no altered expression levels.

Taken together, the amino acid exchanges in UvrY obviously negativeley affect the expression level of the direct downstream target *csrB* in LB as well as in urine. CsrA was found to be strongly up-regulated in urine, which fits to the model for the regulation of carbon metabolism as described by Pernestig *et al.* (2003).

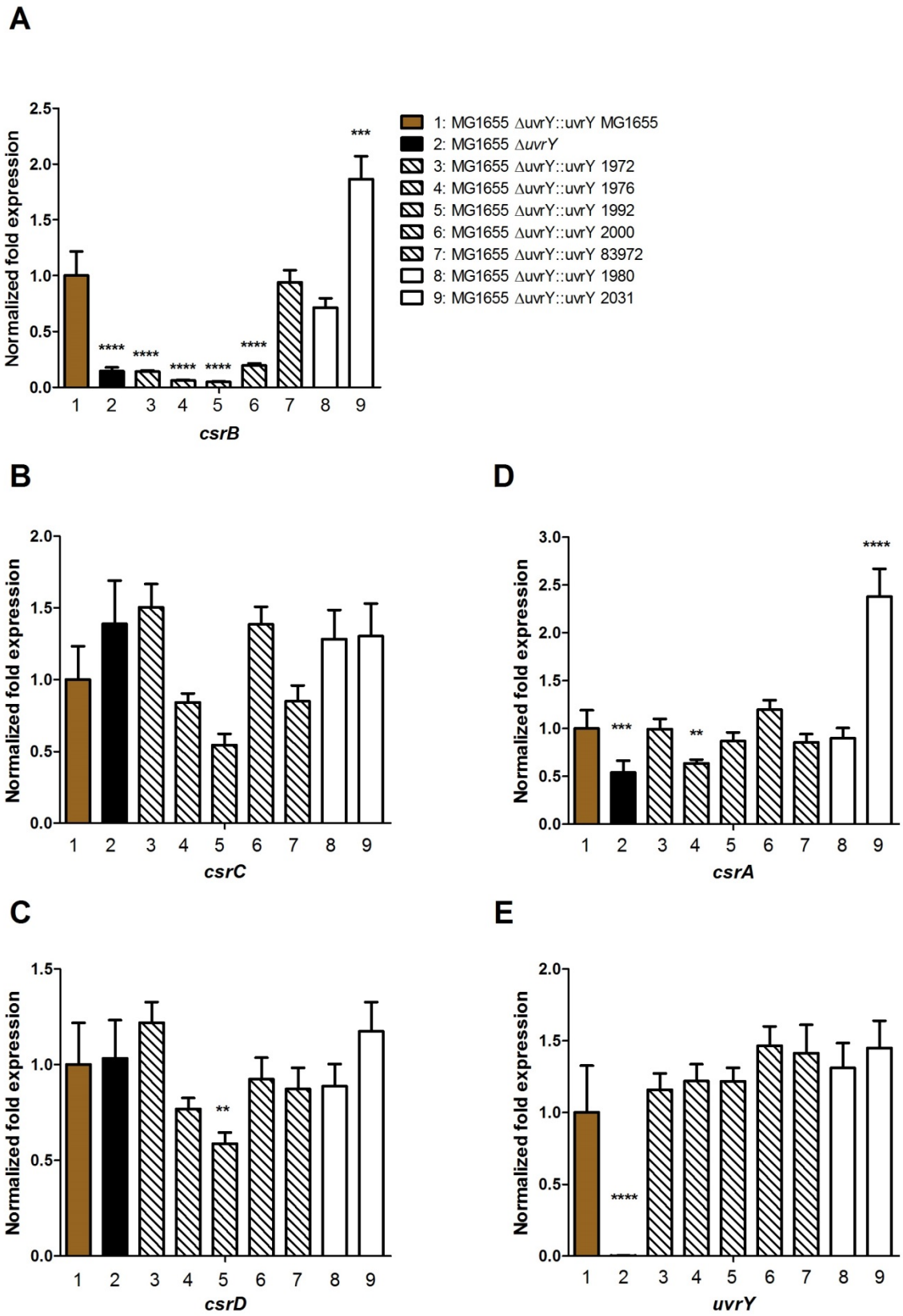


Figure 44: Real Time PCR-based quantification of transcriptional levels of *csrB, csrC, csrD, csrA* and *uvrY* of the allelic variants grown in LB.

Gene expression levels were determined according to the $\Delta\Delta C_t$ method using *frr* as a reference gene. Relative transcription of (A) *csrB*, (B) *csrC*, (C) *csrD*, (D) *csrA* and (E) *uvrY* is shown. Error bars represent SEM of three biological replicates. Significant values were evaluated using a one-way ANOVA with $p < 0.0001$ (****), $p < 0.001$ (***), $p < 0.01$ (***) and $p < 0.05$ (*).

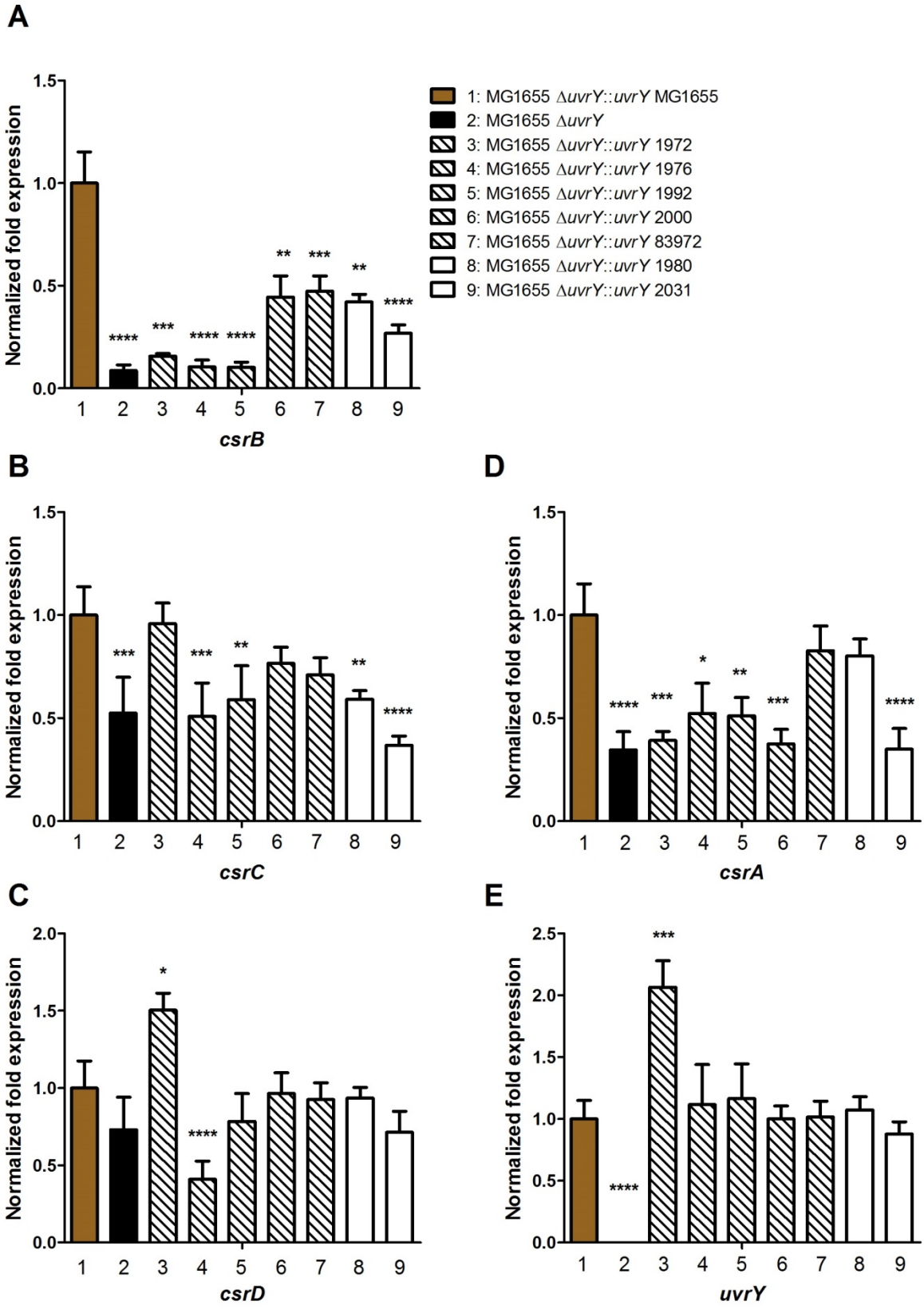


Figure 45: Real Time PCR-based quantification of transcriptional levels of *csrB, csrC, csrD, csrA* and *uvrY* of the allelic variants grown in pooled human urine.

Gene expression levels were determined according to the $\Delta\Delta C_t$ method using *frr* as a reference gene. Relative transcription of (A) *csrB*, (B) *csrC*, (C) *csrD*, (D) *csrA* and (E) *uvrY* is shown. Error bars represent SEM of three biological replicates. Significant values were evaluated using a one-way ANOVA with $p < 0.0001$ (****), $p < 0.001$ (***), $p < 0.01$ (***) and $p < 0.05$ (*).

5 Discussion

In the last decades a lot of studies have been carried out in order to characterize pathogenic *E. coli* inhabiting the urinary tract. These studies often focused on the investigation of virulence factors enabling the bacteria to successfully colonize this habitat (Johnson, 1991; Kaper *et al.*, 2004). Recently another type of *E. coli* that is capable to colonize the bladder without causing the typically associated symptoms (**a**symptomatic **b**acteri**u**ria, ABU) received increased interest, because (i) ABU is the most frequent form of UTI that is affecting millions of people per year and therefore representing a major health problem and (ii) the prototype ABU *E. coli* strain 83972 was successfully used as a prophylactic agent to treat chronic UTI patients (Hull *et al.*, 2000; Sundén *et al.*, 2010). Against the background of emerging antibiotic resistances in UPEC the latter approach is a very promising alternative therapy for the management of problematic UTIs and therefore of high medical interest. In order to learn more about the mechanisms that allow ABU *E. coli* strains to inhabit the challenging environment of the urinary tract, several comparative genomic and transcriptomic approaches have been carried out in the last years (Hancock *et al.*, 2008; Klemm *et al.*, 2006; Snyder *et al.*, 2004). The obtained data revealed that especially modifications in gene expression or genomic content are necessary for an efficient and symptom free colonization of the urinary tract. Thus, the ability of the bacteria to acquire or lose genetic material, the so-called bacterial “genome plasticity” plays a major role in the adaptation to this ecological niche (Dobrindt *et al.*, 2003; Zdziarski *et al.*, 2010, 2008). The genetic differences between *E. coli* isolates causing UTI or ABU, especially modifications that result in altered metabolic activities or gene regulation might therefore help to explain how ABU strains evolved towards a commensal lifestyle.

The aim of the first part of the thesis was to determine differences in metabolic activities of the *E. coli* strains CFT073, 83972 and Nissle 1917 during growth in urine. The second part of the study dealt with the identification and comparative functional analysis of allelic variants of the BarA/UvrY TCS in different clinically relevant *E. coli* isolates

I. Urine – a challenging niche for bacteria

5.1 Basic metabolomics vs. adapted metabolome

In order to be able to successfully colonize the urinary tract bacteria have to efficiently utilize the available nutrients for growth. Urine is regarded as a very challenging growth medium. This is for instance due to the fact that its composition is very variable depending on the fluid intake, exercise

or diet of the host. In addition, the pH of urine is reported to vary between 4.5 and 8, which represents a further substantial difficulty for bacterial growth (Brooks & Keevil, 1997). Most importantly, urine is a nutrient limited environment, e.g. for iron or neutral sugars. Even though urine contains urea, available nitrogen in the form of amino acids and small peptides may be limited (Anfora *et al.*, 2008; Brooks & Keevil, 1997; Snyder *et al.*, 2004). Moreover, it is assumed that the urinary tract is of high osmolarity which results in osmotic stress (Snyder *et al.*, 2004; Zdziarski *et al.*, 2010). In order to deal with these difficulties *E. coli* must exhibit substantial catabolic flexibility and the ability to utilize mixed substrates (Ihssen & Egli, 2005). And indeed, the analysis of the extracellular metabolome showed that *E. coli* is able to utilize different nutrients for the production of precursor metabolites, reducing power, and energy required for the biosynthesis of new cellular components.

Interestingly, the time-resolved extracellular metabolome analysis showed a similar pattern for the majority of measured metabolites for the three tested *E. coli* strains CFT073, 83972 and Nissle 1917. This might not be surprising, as the three tested strains are genetically closely related and it is suggested that *E. coli* strains 83972, Nissle 1917 and CFT073 have originated from the same ancestor (Hancock *et al.*, 2010; Klemm *et al.*, 2007; Vejborg *et al.*, 2010). Nevertheless, we found minor differences in the extracellular metabolome.

Using a pathway-dependent interpretation of the extracellular metabolome, active metabolic routes within bacterial cells could be identified. Regarding the central carbon metabolism the glycolysis represents a central pathway to degrade carbohydrates for the production of ATP and NADH. *E. coli* is characterized by the ability to grow on a variety of different carbon sources (Orth *et al.*, 2011). Identified metabolites that are fed into the glycolysis are for example glucose or ribose. As described for *E. coli* glucose represents a preferred carbon source (Chubukov *et al.*, 2014). This can be confirmed by our analysis showing that glucose was already completely depleted from urine after about four hours of growth. This was also the case for other sugars. This uptake pattern was observed for all of the strains. In contrast the utilization of ribose, a pentose sugar, showed different kinetics compared to glucose although the kinetics per se were similar between the strains. The uptake of ribose started after four hours, but the extracellular level decreased only slightly and ribose was not completely depleted from the medium as observed for other carbon sources. Due to these observations it may be concluded that ribose is not a preferred or primary carbon source for *E. coli*. The slight decrease of the extracellular ribose might also be explained by the growth phases of the culture. The stationary phase is reached after four hours implying a reduced cell division which comes along with lower needs of nutrients. Regarding the utilization of carbon sources a sequential

uptake pattern in general might be explained by diauxic growth meaning a strictly consecutive consumption of carbon sources and catabolite repression. The latter is part of regulatory circuits which sense the presence of different carbon sources and induce a carbon-specific regulation including the transport of preferred and the transport-inhibition of non-preferred carbon sources into the cell (Brückner & Titgemeyer, 2002; Chubukov *et al.*, 2014). However, growth in urine seems to be carbon limited due to the fact that the vast majority of identified carbon sources are directly catabolized and completely depleted from the medium. These conclusions are in accordance with findings from Zdziarski *et al.* (2010) who detected an up-regulation of genes involved in sugar uptake and degradation by transcriptome analysis of re-isolates from *E. coli* strain 83972 inhabiting the human urinary bladder over longer periods suggesting adaptational mechanisms to carbon starvation, as well. In contrast Snyder *et al.* (2004) who investigated the transcriptome of UPEC strain CFT073 grown in urine *in vivo* and *in vitro* could not find an up-regulation of genes normally induced during carbon starvatio. Thus, they concluded that *E. coli* CFT073 is not limited for carbon sources during acute infection of the urinary tract.
However, our results indicate that (i) carbon is a limited source during growth in urine and that (ii) certain metabolic behaviours described for *E. coli*, e.g. the sequential utilization of sugars, also function during growth in urine. This implies the functionality of general regulatory mechanisms during growth in urine. The measurability of differences in the metabolome highlights the functionality of the assays as part of the investigation.

Besides sugars or sugar-related metabolites other metabolites were identified that mirror the metabolic activities of the three strains during growth in urine. Pyruvate, a downstream product of the glycolysis and the main switch between aerobic and anaerobic metabolism, showed for all strains the same pattern: an increased excretion during exponential phase followed by an uptake in parallel with the transition into the stationary phase when sugars were depleted (see Figure 14 B). These observations might underline the important role of pyruvate as a central metabolite being necessary for many metabolic pathways. Further metabolites that were found during the study could be assigned to the TCA cycle, e.g citrate, isocitrate or 2-oxoglutarate as well as to the mixed acid fermentation, e.g. acetate, formate or lactate with all of them showing the same excretion pattern for all strains except for fomate (see Figure 14 D). Besides the striking kinetics with a rapid increase and decrease the highest excretion and reuse of the extracellular formate was observed for *E. coli* Nissle 1917 followed by 83972 and CFT073. The latter *E. coli* strain showed the most moderate secretion and uptake pattern. Formate is derived from pyruvate in the mixed-acid fermentation of *Enterobacteriaceae*, whereby as much as one-third of the carbon atoms of the carbohydrate substrate are converted to formate. The conversion of pyruvate to formate is catalyzed by the

pyruvate-formate-lyase (PFL), the counterpart to the pyruvate dehydrogenase (PDH), which oxidatively decarboxylates pyruvate under respiratory conditions (Sawers, 2005). Formate is either transported out of the cell to prevent acidification of the cytoplasm or, if the pH of the medium drops below approximately 6.8 and if exogenous electron acceptors are absent, taken up again and further degraded via the so-called formate hydrogenlyase (FHL) complex to carbon dioxide (CO_2) and dihydrogen (H_2) (Rossmann *et al.*, 1991; Stephenson & Stickland, 1932). The observable pyramid-like accumulation and decrease of formate in urine might therefore be explained by a possible acidification of urine which reaches a threshold level after about 6 hours resulting in a rapid uptake of formate into the cell which is subsequently degraded via the FHL pathway. The reasons for the different kinetics between the strains remain to be further investigated.

The occurrence of diverse intermediates found in different metabolic pathways that are used both under aerobic and anaerobic conditions suggest that the oxygenation during the *in vitro* experimental procedure was neither strictly aerobic nor anaerobic. This assumption corresponds with *in vivo* transcriptomic data from *E. coli* CFT073 grown in the urinary tract showing that different genes indicative of anaerobiosis were either down- or up-regulated (Snyder *et al.*, 2004).

Further metabolites that play important roles for the organism and that were quantified during the extracellular metabolome analysis were amino acids. Although *E. coli* is able to synthesize all amino acids required for growth it also utilizes freely available amino acids present in the environment (Smith, 1992). Amino acids serve as a suitable source for carbon, nitrogen or energy that can be utilized for growth (Gottschalk, 1986). As reported from transcriptomic analyses of *E. coli* strains grown in urine gene expression required for the amino acid transport and utilization were significantly up-regulated indicating the importance of amino acids as essential nutrients in urine (Snyder *et al.*, 2004; Zdziarski *et al.*, 2010). In the broader sense the results obtained in this study confirm these findings due to the fact that several amino acids were completely catabolized from the medium already after four hours of growth implying a well-functioning and efficient amino acid uptake system. Moreover it seems that some amino acids were more preferred regarding their uptake and utilization than others. Serine, valine, leucine or lysine for example seem to be primary carbon or nitrogen sources because they were utilized very quickly until complete depletion of the medium. On the contrary amino acids like alanine, glycine, histidine or tyrosine were not utilized as strongly as the afore mentioned ones. In this context it needs to be mentioned that these amino acids did not show the same starting concentrations which might explain why differences were not as strong observable as for other amino acids with similar initial concentrations. In any case, the concentration of the majority of amino acids changed in a similar way in the three strains tested except for the amino acids lysine, glutamic acid and tryptophan (see Figure 15 B-D).

Lysine was taken up to by the strains in different amounts after about four hours whereby the UPEC strain CFT073 showed the highest uptake pattern (see Figure 15 C). In parallel to the decrease of lysine an increase of cadaverine could be observed with highest extracellular concentrations for *E. coli* CFT073, as well. Cadaverine, a polyamine, represents the decarboxylation product of lysine (Figure 46). Therefore reduction of extracellular lysine should be accompanied with an increase of cadaverine which could be observed for all strains. Interestingly, cadaverine is reported to mediate resistance of UPEC strains against nitrosative stress. Nitrosative stress is caused by nitric oxide (NO) and reactive nitrogen intermediates (RNI), key players of innate immunity, that accumulate during inflammation (Bower & Mulvey, 2006). NO, a signalling and defense molecule in biological systems with bactericidal effects, was reported to be increased 30- to 50-fold in the bladder of patients with cystitis (Lundberg *et al.*, 1996). Furthermore bacteria themselves were also shown to produce NO as an obligatory intermediate thereby also contributing to high levels of NO and its derivatives in infected bladders (Bower & Mulvey, 2006; Corker & Poole, 2003). Based on these findings the higher uptake of lysine and the stronger excretion of cadaverine by UPEC strain CFT073 may be due to the nature of UPEC strains which may have evolved better mechanisms to resist the damaging effects of NO and its derivatives (Bower & Mulvey, 2006).

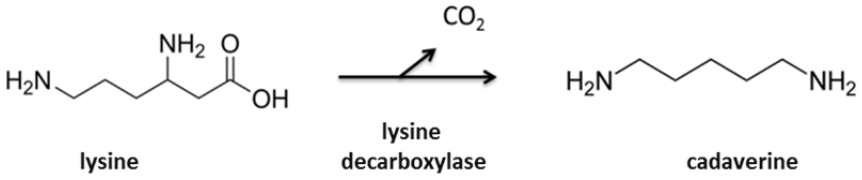


Figure 46: Decarboxylation reaction of lysine to cadaverine catalyzed by the lysine decarboxylase.

The results regarding the extracellular levels of glutamic acid (see Figure 15 D) reveal the following kinetics for *E. coli* 83972 and CFT073: It was slightly taken up during the exponential phase until a constant extracellular level remained after four hours implying no further uptake or excretion. Interestingly, the results for *E. coli* Nissle 1917 were the other way around: In parallel to the detected decrease of glutamic acid for *E. coli* 83972 and *E. coli* CFT073, Nissle 1917 showed a drastic increase of the extracellular concentration which reached a peak around four to five hours, followed by a slight decrease and a repeated continuous increase of glutamic acid up to 24 hours. Glutamic acid represents the focal point of nitrogen assimilation which is an essential process for growth in *E. coli* by serving, together with glutamine, as nitrogen donors for biosynthetic reactions (Helling, 1998; Tyler, 1978). Due to the fact that most cellular nitrogen enters the metabolism through glutamate it might be speculated that *E. coli* strains 83972 and CFT073 exhibit an increased need for nitrogen reflected in a higher uptake of glutamic acid.

Despite some minor changes the different extracellular levels of tryptophan were shown to remain practically constant over time for *E. coli* strains 83972 and Nissle 1917 (see Figure 15 B). In contrast *E. coli* CFT073 showed a dramatic uptake starting after about 8 hours resulting in a complete depletion of the medium after 24 hours. If this is a characteristic feature of *E. coli* CFT073 grown in urine cannot be answered here because no comparable studies are available during growth in other media.

The direct comparative metabolomic approach which was carried out in this study between the three *E. coli* strains CFT073, 83972 and Nissle 1917 provides the first attempt to shed light into different adaptations to the urinary tract on a metabolomic level. The results of the analysis showed that the strains did not differ significantly for the vast majority of measured extracellular metabolites implying no markedly strain-specific metabolic footprint. With regard to their different natural niches one could speculate that in particular *E. coli* 83972 would show other secretion or uptake patterns of distinct metabolites based on the fact that it is regarded as a well-adapted strain to the urinary tract (Klemm *et al.*, 2007; Roos, Ulett, *et al.*, 2006). But, as already mentioned above, the comparable behaviour might be reflected by the similar genetic content of the investigated isolates. In the end, this might lead to similar responses to particular environments which are mirrored in the metabolic activity, as well. However, a few differences could be observed for some metabolites. Whether these differences might be correlated with different adaptations of the strains to growth in urine cannot be answered here. To investigate this in more detail comparative extracellular metabolome analyses have to be carried out in other media to conclude if differences are due to adaptation to growth in urine or if they are strain-specific. In order to achieve a better understanding of different adaptational mechanisms of *E. coli* strains from different ecological niches additional investigations such as comparative genomics, transcriptomics, proteomics or fluxomics have to be carried out. The combination of these data would help to understand the mechanisms or regulatory interactions that facilitate successful survival in ecological niches such as the urinary tract.

5.2 Carbon – a limiting factor for growth in urine

As suggested by the extracellular metabolome analysis and as implied by other studies growth in urine seems to be carbon limited (Zdziarski *et al.*, 2010). The dry mass carbon - with 50 % - accounts for the main elemental component of *E. coli* followed by oxygen (20 %), nitrogen (14 %), hydrogen (8 %), phosphorus (3 %), potassium (2 %) and sulfur (1 %) (Neidhardt *et al.*, 1990). The distribution of these elements clearly illustrates the importance of carbon as a fundamental requirement for growth. In order to check whether carbon might be a limiting factor for growth in urine we cultivated

the three *E. coli* strains CFT073, 83972 and Nissle 1917 in urine with the addition of glucose, fructose or ribose and the 20 standard amino acids and measured their growth kinetics in comparison to growth in pure urine. Based on the fact that amino acids could serve either as a carbon or a nitrogen source we further carried out growth kinetics in urine complemented with NH_4Cl, an inorganic component which could be utilized as a nitrogen source only.

For the majority of added amino acids no significant changes in growth behaviour could be observed compared to the strains that were cultivated in pure urine. Growth curves are exemplarily shown for the amino acids tryptophan and lysine (see Figure 16 A-B), which exhibited different utilization patterns for the three strains (see Figure 15 B-D). Interestingly, a poor growth profile was detected for all of the strains grown in urine under addition of glutamic acid (see Figure 16 C) implying a deleterious effect of elevated glutamic acid concentrations. The reason for the impaired growth still needs to be elucidated. Therefore increased availability of amino acids does not improve the growth and might even disturb the equilibrium of the metabolic network resulting in reduced bacterial growth. The initial question of putative carbon or nitrogen starvation might be answered by the growth kinetics performed under addition of glucose (see Figure 16 D) and NH_4Cl (see Figure 16 E). Growth was markedly improved under addition of glucose for all strains. Thus, it can be concluded that *E. coli* suffers from carbon starvation during growth in urine. Nitrogen obviously does not represent a limiting factor. If this would be true *E. coli* would have shown improved growth upon addition of NH_4Cl, as well. Maybe most of the nitrogen is assimilated from the amino acids which are sufficiently present in urine.

In order to prove the hypothesis of carbon or nitrogen starvation during growth in urine it is not sufficient to carry out growth curves under addition of the questionable nutrients. The obtained data in this study merely represent first hints on the essential requirements for growth in urine and putative starvation situations. To achieve deeper insights and to determine limiting nutrients it would be reasonable to analyze the genetic regulatory networks behind and to measure distinct signal molecules that are involved in carbon or nitrogen starvation. Possible targets to identify carbon starvation are the carbon starvation response (*cst*) genes and the so-called *pex* (postexponential) genes which have been shown to accumulate under carbon starvation (Groat *et al.*, 1986; Groat & Matin, 1986). During nitrogen starvation *E. coli* activates the nitrogen regulated response (Ntr) resulting in the expression of ~ 100 genes to facilitate nitrogen scavenging from alternative sources (Brown *et al.*, 2014). Different molecular control parameters in this context might be the master transcription regulator NtrC (nitrogen regulator protein C), which is part of the NtrB/NtrC TCS and known to activate transcription of genes/operons whose products minimize the reduction of growth under nitrogen-limiting conditions or the nitrogen assimilatory gene *glnA* which

is activated by NtrB/NtrC (Zimmer *et al.*, 2000). Guanosine tetraphosphate (ppGpp), a signal molecule, which is synthesized during nitrogen starvation might be a further target (Villadsen & Michelsen, 1977).

5.3 Stress response is growth medium-dependent

Since urine is known as a challenging medium in which bacteria constantly face changes like nutrient starvation, variations in osmolarity or pH, we wanted to investigate the stress response of *E. coli* 83972 upon growth in urine and LB focusing on the measurement of RpoS (**R**NA **po**lymerase, sigma **S** factor; σ^{38}), the master regulator of the general stress response (Hengge, 2009). RpoS is a 37.8 kDa protein which belongs to the group of sigma (σ) factors. Sigma factors are subunits of the RNA polymerase (RNAP) that control the promoter selectivity of the RNAP and allow efficient promoter recognition and transcription initiation (Chandrangsu & Helmann, 2014). Sigma factors bind to specific promoter regions (see Figure 47 A) and facilitate binding of the RNAP thereby initiating transcription at the target promoter. The RNAP itself consists of a core factor which is composed of four subunits (α [RpoA], α [RpoA], ß [RpoB], ß´ [RpoC]) forming together with the sigma factor the RNAP holoenzyme (see Figure 47 B).

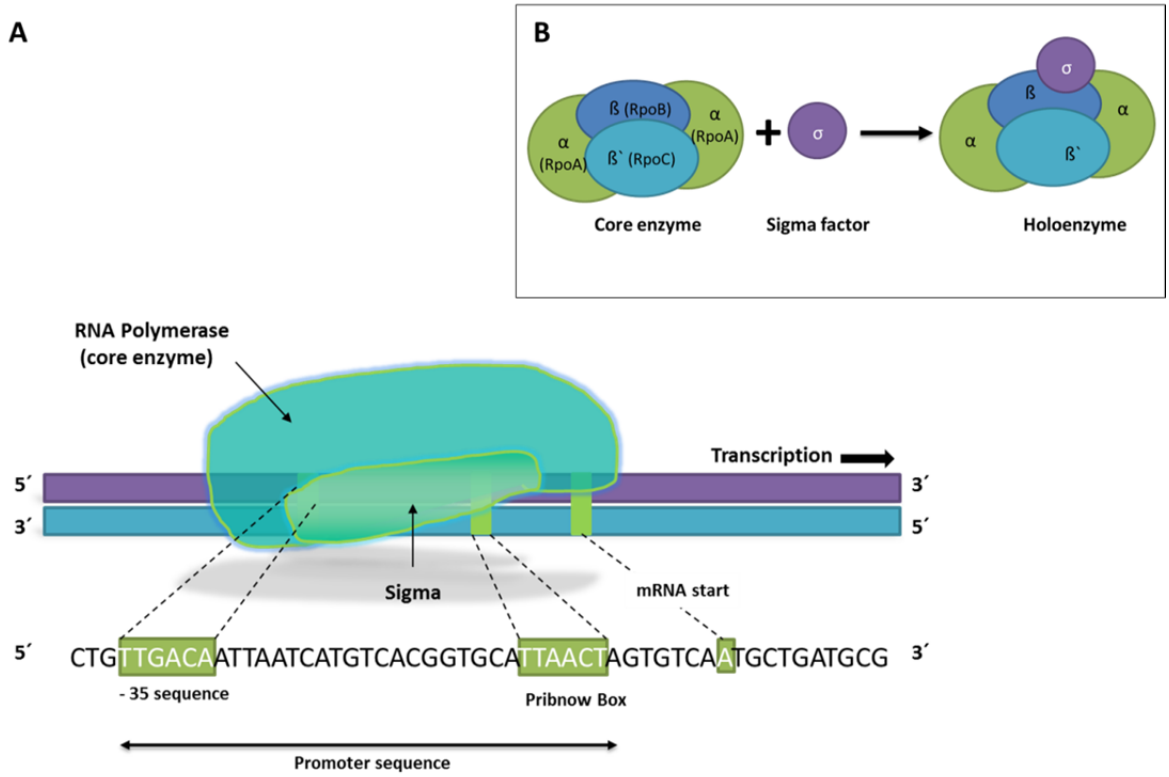


Figure 47: Binding of RNA polymerase to DNA

(A) The sigma factor which exhibits high affinity to the - 35 sequence and the Pribnow box binds to the promoter and facilitates binding of the RNAP. Subsequently the RNAP holoenzyme initiates transcription. (B) The RNA polymerase consists of the core enzyme which is composed of four subunits (α, α, ß, ß´) and forms together with the sigma factor the RNAP holoenzyme.

Besides the primary sigma factor RpoD (σ^{70}), which is responsible for the transcription of housekeeping genes, six further sigma factors can be expressed in *E. coli* that control specific sets of genes which supplement the ones controlled by σ^{70}. The six sigma factors respond to different stressors like nitrogen depletion (σ^{54}), heat shock (σ^{32}), extracytoplasmic stress (σ^{24}), the need for flagellin (σ^{28}), starvation (σ^{38}) and the need for iron transport (σ^{19}) (Loewen *et al.*, 1998). An overview of the sigma factors is given in Table 24. RpoS is known to accumulate under conditions of nutrient starvation and stress. Once induced, the cell is resistant to a wide range of stress and starvation treatments therefore RpoS is essential for an adequate adaptation to stationary phase by regulation of genes required for this adaptation (Battesti *et al.*, 2011; Hengge, 2009). It is described to control approximately 10 % of all *E. coli* genes (Weber *et al.*, 2005).

Table 24: Sigma factors and their functions in *E. coli*.

Sigma	Gene	Expression conditions/ Functions
σ^{70}	*rpoD*	Normal conditions; housekeeping sigma factor; σ^{70} targets promoters that are essential for normal growth
σ^{32}	*rpoH*	Heat stress; σ^{32} controls the heat shock response
σ^{54}	*rpoN*	Nitrogen starvation; σ^{54}controls expression of nitrogen-related genes
σ^{28}	*rpoF*	Flagella expression; responsible for initiation of transcription of a number of genes involved in motility and flagella synthesis
σ^{38}	*rpoS*	General stress response; σ^{38} transcribes genes involved in stress responses
σ^{19}	*fecI*	Iron transport; σ^{19} regulates the expression of genes involved in the transport of ferric citrate
σ^{24}	*rpoE*	Stress of cellular envelope; specializing in responses to the effects of heat shock and other stresses on membrane and periplasmic proteins

The data obtained in this study for growth in LB revealed that RpoS was slightly expressed during the exponential phase followed by an RpoS accumulation upon transition into the stationary phase (t_5). The signal further substantially increased with prolonged growth reaching its maximum intensity after 7 to 24 hours (see Figure 17 A). These data are in accordance with previous results showing that (i) RpoS is expressed at a basal rate during exponential phase implying a low signal which is due to efficient proteolytic degradation of RpoS and in contrast (ii) RpoS is highly expressed during stationary phase when nutrients are depleted and the stress increases (Dong *et al.*, 2008; Jishage & Ishihama, 1995; Lange & Hengge-Aronis, 1994). Therefore it can be concluded that the cells are sufficiently supplied with nutrients in the first hours of growth. After 5 hours the nutrients have already started to become depleted resulting in an up-regulation of the stress response. Besides nutrient deprivation factors like acid stress, low or high temperature and osmolarity as well as oxidative stress lead to an up-regulation of RpoS expression (Battesti *et al.*, 2011; Becker *et al.*, 1999; Chung *et al.*, 2006). The RpoS expression reached its maximum after 7 hours.

The results for growth in pooled human urine were rather different (see Figure 17 A). Although an increase of the RpoS signal was detected with the highest signal intensity after 5 hours of growth, the signal intensity decreased afterwards. These results are quite surprising, because it cannot be assumed that the stress decreased with prolonged growth in urine. However, different conclusions can be drawn from the results: based on the fact that RpoS was most strongly expressed after 5 hours it can be suggested that nutrients were depleted much earlier in urine than in LB. This is supported by (i) the growth kinetics for different strains in LB and urine shown in this study (Figure 33) as well as by (ii) the literature showing better growth for diverse ExPEC strains in LB than in urine implying a better supply with nutrients (Aubron *et al.*, 2012). Moreover, the complex regulatory network which is connected with expression of RpoS might be different regarding the mechanisms controlling the level of RpoS in LB. An important aspect in this context is the promoter selectivity control of the RNAP by replacement of σ subunits on the core enzyme (Jishage & Ishihama, 1995). It is well known that transcription of the genes expressed during exponential phase is controlled by σ^{70}, while σ^{38} controls the transcription of a huge amount of genes required in stationary phase (Loewen & Hengge-Aronis, 1994). Loewen *et al.* (1998) and Kawakami *et al.* (1979) could show that the intracellular level of σ^{70} remains constant during exponential and stationary phase growth, while the levels of σ^{38} starts to accumulate during stationary phase. The fact that σ^{70} and σ^{38} are closely related and that they even recognize the same promoter sequences implies that both subunits compete for the RNAP to bind in order to reprogram the global gene expression (Hengge, 2009). Although there are a lot of different factors known which affect the competition between σ^{70} and σ^{38} and which facilitate the σ^{70} or σ^{38} specific gene expression, it might be speculated that σ^{70} binds the RNAP more efficiently during ongoing stationary phase in urine, thus leading to a decreased level of RpoS. This might also be due to the fact that RpoS levels in general are known to be lower than those of RpoD, at least in LB. Another factor which might influence RpoS levels is the ATP-driven complex ClpXP protease that in concert with the response regulator RssB proteolytically degrades RpoS (Hengge-Aronis, 2002; Petersen *et al.*, 2007). It might be speculated that the protease becomes more active again in urine with ongoing stationary phase. Further studies report of several nutrient starvation conditions under which RpoS is mutated or down-regulated implying that RpoS is not always beneficial especially with respect to overcoming of growth limitations. This aspect might play an important role during growth in urine as a very challenging medium. Maybe survival and growth is more important than resting in a protected status. However, these findings also imply that recovery mechanisms which return cells to growth also play an important role in the complex regulation of response to environmental stimuli (Battesti *et al.*, 2011; Notley-mcrobb *et al.*, 2002).

Nevertheless, the reason why the RpoS signal decreases in the course of stationary growth in urine remains elusive. In order to shed light into the stress responses which are activated during growth in urine and LB it would make sense to measure both RpoS and RpoD levels in parallel. Preliminary data (not shown) could confirm that the RpoD expression level remained constantly during the exponential and stationary growth phase, at least in urine, as already described by Loewen *et al.* (1998) and Kawakami *et al.* (1979). However, further investigations need to be performed to provide detailed insights into the competition between RpoS and RpoD in the employed media. Additionally the measurement of RpoC, a subunit of the core enzyme, could serve as an internal standard for expression normalization.

II. BarA/UvrY TCS – a key factor for the adaptation to growth in urine

In order to detect adaptive traits promoting growth in the urinary tract, we investigated re-isolates of ABU strain 83972 obtained from patients suffering from chronic or recurrent UTI who were deliberately colonized with *E. coli* strain 83972 at the University of Lund (Sweden). Whole genome sequencing of ABU *E. coli* 83972 re-isolates from this study revealed that various chromosomal loci repeatedly accumulated mutations upon prolonged growth in the bladder. Besides metabolic and virulence-related genes the genes coding for the TCS BarA/UvrY were also detected to be frequently mutated (Zdziarski *et al.*, 2010). The TCS BarA/UvrY has already been described to be an important factor for the regulation of various cellular functions like the regulation of carbon metabolism, biofilm formation, motility or adhesion (Jackson *et al.*, 2002; Pernestig *et al.*, 2003; Suzuki *et al.*, 2002; Wei *et al.*, 2001). Due to this global regulatory role of the TCS, we hypothesized that the accumulation of mutations in *barA/uvrY* might be of advantage for long-term growth in the bladder and that the genes coding for this TCS might therefore represent a mutational hot spot. If this was the case *barA* and *uvrY* should be in general under selective pressure when *E. coli* is growing in the urinary tract, but not in isolates from other niches. In order to address this question, we amplified *barA* and *uvrY* from 213 *E. coli* isolates originating from the urinary tract of patients or from the feces of healthy individuals. Subsequently we sequenced and compared the *barA* and *uvrY* alleles of the different isolates.

5.4 Sequence context variation of *barA* and *uvrY*

Already the amplification of the *barA* and *uvrY* alleles showed surprising results as the molecular weight of *barA* (2900 bp) or *uvrY* (854 bp)-specific PCR products was higher than expected for some ABU, UTI and fecal isolates. Subsequent sequence alignments with the reference alleles *barA* and

uvrY of *E. coli* MG1655 revealed the insertion of IS elements or fragments of IS elements into or in the vicinity of both genes. The inserted sequences could be assigned to IS elements IS*1*, IS*3*, IS*4* and IS*5* (Figure 20; Figure 21; Figure 50). IS elements are widespread, small autonomous genetic mobile elements encoding their own transposases for the transposition into many different sites of the bacterial genome (Ooka *et al.*, 2009). In the majority of the strains containing these mobile genetic elements the *barA* or *uvrY* coding sequences were disrupted by the insertion, resulting in gene inactivation. Therefore the corresponding *E. coli* isolates were excluded from the study. Interestingly, the situation was rather different for two UTI isolates, IMI 917 and IMI 923. The analysis of available draft genome sequences showed that flexible elements were not inserted into the ORF of *uvrY* but into its upstream region, which might lead to an impaired UvrY expression by either inactivation or deregulation of the *uvrY* promoter or contrarily to an even stronger activation of the gene (Siguier *et al.*, 2014). The exact nature of the mobile genetic elements inserted in IMI 917 and IMI 923 could not be clarified, because these genomes have not been closed and the genomic localization of redundant mobile genetic elements could not yet be exactly determined. However, the detection of inverted repeat sequences, which are characteristic for several mobile genetic elements, upstream of *uvrY* in the draft genomes led to the conclusion that either an IS element or a transposon containing accessory genes must be inserted directly upstream of *uvrY.*

Nowadays 500 different IS elements are known for *E. coli* which are (i) often present in more than a single copy and are (ii) described to play an important role in bacterial evolution. Moreover they contribute to genome diversification by insertion of mutations, inverting or deleting DNA segments, sequestering genes and influencing neighboring gene expression (Dyda *et al.*, 2012; Mahillon & Chandler, 1998; Ooka *et al.*, 2009). As we detected insertion elements in the analyzed *E. coli* isolates, it is possible that in specific ecological niches not only modulations of the protein function by point mutations, but a complete loss-of-function of BarA or UvrY could be of advantage for *E. coli*.

5.5 Sequence variability of *barA* and *uvrY*

The sequence analysis of the *barA* and *uvrY* alleles revealed the occurrence of different types of sequence variations with syn SNPs as the most prominent ones followed by ns SNPs as well as non-sense SNPs or deletions. The total number of detected syn and ns SNPs was higher for *barA* than for *uvrY*, which may be at least partially explained by the approximately four times bigger size of *barA* (Table 16). In contrast to *barA* and *uvrY* the housekeeping gene *frr* showed only few sequence variations, as expected for a housekeeping gene that is well conserved within a bacterial species. Thus, the *frr* encoding sequence served as a valuable control in the attempts to assess the nucleotide sequence variability of *barA* and *uvrY* in the different sets of *E. coli* strains investigated in this study.

Regarding the different groups of isolates chosen we expected a higher selection pressure for ABU and UTI isolates compared to HA-ABU or FI isolates. This assumption is due to several reasons: ABU strains exhibit an extended colonization time *in vivo*, therefore they are much more subjected to selection pressure of the urinary tract. UTI strains have to face the full immune response of the host also representing a selection pressure. In contrast HA-ABU isolates, which are received during hospital stays, are rapidly detected and treated with antibiotics. Thus, they do not exhibit a prolonged persistence in the urinary tract with relatively short time for selection pressure to act on these isolates. Due to the fact that fecal isolates colonize a completely different niche, these strains should not face any selection pressure specific for the urinary tract. Looking at the results the hypothesis of higher selection pressure in ABU and UTI isolates can be confirmed since by trend more of the ns SNPs occurred in *uvrY* and *barA* in these two groups compared to the HA-ABU or fecal isolates.

The average number of ns SNPs per isolate was comparable for ABU and fecal isolates in the case of *barA*, whereas the average number of ns SNPs per isolate for *uvrY* was highest for ABU followed by UTI strains (Table 17). This indicates that in the urinary tract indeed a selection pressure acts on the TCS BarA/UvrY, leading to a more frequent appearance of mutations, although much stronger for the response regulator component UvrY than for the sensor kinase BarA.

Interestingly, 40 % of the ns SNPs detected in *barA* were located or clustered in functional regions which might have a stronger impact on the functionality of the transcribed protein. One affected region was the HATPase_c domain which is responsible for the transfer of a phosphoryl group from ATP to the His residue in the HisKA domain and therefore required for functionality (Gao & Stock, 2009) (Figure 22 A/B). In addition, a non-sense SNP leading to a premature stop in the transcription of the protein was detected in the Rec domain of BarA in an ABU isolate. This mutation in *barA* has a pivotal impact on the response regulator UvrY, which due to the disruption of the phosphorelay cannot be activated anymore. Although not located in a functional region a further stop codon was detected in one fecal isolate. Intriguingly, and in contrast to *barA*, the occurrence of ns SNPs within *uvrY* was markedly restricted to gene regions coding for functional domains of the protein: 95 % of the ns SNPs were detected in functional regions of the deduced UvrY amino acid sequence (Figure 23 A/B). A further striking feature was the appearance of non-sense SNPs affecting the DNA binding domain of UvrY. These non-sense SNPs were restricted to the group of ABU and UTI isolates. The resulting premature stop codons should therefore also have a significant impact on the regulation of downstream targets of the BarA/UvrY TCS.

The obtained data revealed that (i) ABU and UTI isolates harbored the highest amount of ns SNPs at least for *uvrY* and that (ii) large fractions of these SNPs were located in important functional regions of the UvrY protein and that (iii) stop codons in functional regions of UvrY were only found in ABU

and UTI isolates. Moreover, the fact that the coding sequences of *barA* and *uvrY* were much more affected than the coding sequence of the housekeeping gene *frr* (Figure 24 A/B) indicates that the different *barA* and *uvrY* alleles are resulting from selection and not simply from an overall increased mutation rate in the urinary tract. These observations further support the hypothesis that selective pressure is acting on the TCS during growth in the urinary tract and confirm the initial hypothesis of the *barA* and *uvrY* genes representing a mutational hot spot in the urinary tract.

Regarding the distribution of the identified syn SNPs in both proteins it can be stated that *barA* exhibited a more even distribution of syn SNPs although special domains showed a slight clustering of syn changes or even insusceptibility to syn SNPs like the HisKa domain (Figure 22 C). In contrast to *barA* the syn SNPs found in *uvrY* appeared to accumulate in the N-terminal part of the protein (Figure 23 C). This observation is interesting, because although syn SNPs in the N-terminal part of a protein do not alter the amino acid sequence, they can have a dramatic influence on gene expression levels (Kudla *et al.*, 2009). However, even though it might be interesting in the future to see to which extent these syn SNPs may play a role in controlling protein abundance, we focused in this study on the influence of the identified ns SNPs on protein function.

In order to illustrate a putative relationship between the four groups of isolates a Minimum Spanning Tree (MST) was created. MSTs represent a widely used graph-theoretical method for pattern recognition and reflection of possible relationships between arbitrary points (Dussert *et al.*, 1987). The MST created in this study is based on the occurrence of all identified SNPs, which are used to cluster the different isolates (representing the arbitrary points) into groups. The different MSTs for *barA* (Figure 25), *uvrY* (Figure 26 A) and *frr* (Figure 26 B) clearly demonstrated that the general sequence variation is much higher in *barA* than in *uvrY* or *frr*. This is attributed to an obviously higher sequence variability of *barA* if compared to *frr*. This is also confirmed by the calculation of the SNP rate per nucleotide under consideration of all detected SNPs. This calculation resulted in a rate of 0.07 SNPs/nucleotide for *uvrY*, 0.06 SNPs/nucleotide for *barA* and 0.022 SNPs/nucleotide for *frr*. This once again underlines that genes encoding the TCS BarA/UvrY exhibit a higher sequence variability compared to *frr*.

Although some sequence variation types (SVTs) seemed to be more prominent in either *barA* or *uvrY* a strict assignment of isolates from the four different groups to a special SVT could not be achieved. This might have multiple reasons. Due to the fact that no information is available of how long the tested isolates persisted in the patients we cannot estimate if the detected SNPs are already fixed in the strains or if they might represent recent mutations that have not been stabilized by selection yet. Another explanation is the fact that bacterial adaptation is driven by individual host environments as

shown by Zdziarski *et al.* (2010). This might also contribute to the diversity of SVTs found in *barA* and *uvrY*.

5.6 Correlation of SNPs and phylogeny

In order to investigate the phylogenetic background of the four *E. coli* groups and to analyze if the four groups of isolates could be assigned to particular phylogroups a MST based on available MLST sequences was calculated. The results shown in Figure 27 clearly demonstrate that the strains can be assigned to six different phylogenetic lineages (A, B1, B2, ABD, AxB1 and D). Each phylogroup included ABU, UTI, HA-ABU and fecal isolates. Therefore the four groups exhibited a phylogenetic heterogeneity and a direct correlation between strains from a special phylogenetic group and certain groups of isolates cannot be made. This is not surprising, as although several phylogroups were described to contain predominantly strains of a certain niche - such as phylogroup A and B1 which are reported to contain mainly commensal isolates as also partly shown in the study - a strict assignment of *E. coli* isolates based on their ecological background into specific phylogroups is not possible (Bailey *et al.*, 2010; Clermont *et al.*, 2000). This phenomenon is in accordance with the literature reporting that for example ABU isolates often share the same phylogenetic background as UTI isolates and therefore cannot be assigned to a distinct phylogenetic lineage (Dobrindt *et al.*, 2003; Mabbett *et al.*, 2009; Salvador *et al.*, 2012; Zdziarski *et al.*, 2008).

5.7 Detection of adaptive evolution for *barA* and *uvrY*

Selection pressure forces bacteria to quickly adapt to existing environments which affects bacterial genome-pattern and molecular variation. To quantitatively assess the evolutionary forces acting on the genes coding for the TCS BarA/UvrY during growth in the urinary tract the McDonald-Kreitman test, one of the most powerful and extensively used tests to detect signatures of neutral evolution, was used (Egea *et al.*, 2008; McDonald & Kreitman, 1991). By means of this test the rate of possible positive selection should be inferred from polymorphism and divergence data. The basis of the test is represented by different types of mutations that are compared on an interspecies (fixed substitutions) and intraspecies (polymorphic substitutions) level whereby *Salmonella* was used as an outgroup in the population genetic sense. The ratio of fixed replacement differences to fixed syn differences is compared to the ratio of polymorphic replacement differences to fixed syn differences. The results can be interpreted in the light of neutral evolution. In order to estimate the adaptive evolution in *barA* and *uvrY,* the conserved gene *frr* was used again as a control with respect to evolutionary forces.

Due to the fact that no information is available on how long the investigated *E. coli* isolates inhabited the different individuals the results need to be analyzed with regard to a short evolutionary time scale. This is also due to the situation in the urinary tract as a niche, where the population size changes repeatedly upon micturition. Based on these facts the divergence between populations was used as a validation baseline to evaluate short term evolution.

The results showed, as indicated by the Fishers exact test P value, that sequence variability of both genes, *barA* and *uvrY*, in contrast to the control housekeeping gene *frr*, was not compatible with neutral evolution (Table 19). For *barA* an excess of polymorphic syn SNPs was detected compared to the number of fixed substitutions. This is suggestive of purifying selection that limits the sequence variation in all of the groups. Aside from similar long-term selective constraints in *uvrY* - as indicated by the fixed substitutions – the polymorphic differences varied substantially in the four groups with ABU and UTI isolates showing an excess of ns mutations indicating positive selection. In contrast the HA-ABU and fecal isolates showed values compatible with neutral evolution. These observations are clearly indicative of branch variation (Chen *et al.*, 2006). The aspect of positive selection is additionally underlined by the neutrality index (NI) values which are quite high for *uvrY* in ABU and UTI isolates (7.7 and 5.7). The NI quantifies the direction and degree of departure from neutrality (Rand & Kann, 1996). Values > 1 suggest an excess of amino acid variation within a species which was true for *uvrY* in ABU and UTI isolates if compared to the other groups.

Finally the results indicate that *uvrY* is underlying positive selection in the bladder. Positive selection obviously plays an important role in adaptive evolution. Charlesworth and Eyre-Walker (2006) who investigated adaptive evolution in enteric bacteria could even show that even 50 % of amino acid substitutions is driven by positive selection. Furthermore Petersen *et al.* (2007) reported that a lot of *E. coli* genes involved in interaction with other bacteria, phages or the hosts´ immune system are under positive selection. These findings might indicate that especially genes responsible for environmental interactions are targets of positive selection which would support our findings for *uvrY*, as part of a global regulatory TCS that is responsible for the processing of environmental stimuli, being under positive selection.

Another aspect shown by the McDonald-Kreitman test was the higher susceptibility of *uvrY* and *barA* for sequence variation compared to *frr*. This finding could be validated by the investigation of the conservation scores which were calculated by comparing the target sequence with all available target sequences of any species documented in the entire database. The calculated conservation levels clearly demonstrated that BarA and UvrY protein sequences are much more flexible compared to Frr which is also in accordance with the findings described in chapter 4.4.2 .

According to our findings *uvrY* represents a less conserved but rather flexible gene. Since *uvrY* was shown to underlie positive selection pressure the focus was placed on the investigation of *uvrY* in the further course of the study. However, focusing on UvrY does not mean that BarA is a less important component of the TCS. It may well be that bacteria also exhibit BarA-dependent adaptations. This remains to be further investigated.

5.8 Impact of individual *barA*/*uvrY* alleles on bacterial phenotypes

Since components of the Csr sytem as well as the TCS BarA/UvrY are known to affect a variety of phenotypic traits like biofilm formation or flagella expression an important aspect of the study was to address the question, if distinct phenotypes can be correlated with specific *barA/uvrY* alleles (Jackson *et al.*, 2002; Mitra *et al.*, 2013; Timmermans & Van Melderen, 2010; Wang *et al.*, 2005; Wei *et al.*, 2001). We tested 47 isolates of the initial strain panel, each representing one of the identified ns SNP combinations detected with regard to *barA* and *uvrY,* for several relevant phenotypic traits like biofilm formation, synthesis of curli and cellulose expression as well as motility and growth characteristics.

5.8.1 Biofilm formation

Many persistent and chronic UTIs are believed to be associated with biofilm formation. Therefore, the ability to form biofilms is commonly considered to be a virulence-associated trait of UPEC (Costerton *et al.*, 1999; Ferrières *et al.*, 2007). In addition, the formation of biofilms is clinically relevant because it is known that a lot of UPEC and ABU strains are able to produce biofilms on medical devices, e.g. catheters. Once embedded in a biofilm bacteria show extraordinary resistance to antimicrobial treatment or the immune defense responses of the host and are thus able to persist over longer times at the site of infection (Costerton *et al.*, 1999; López *et al.*, 2010; Parsek & Singh, 2003; Sauer, 2003). Several studies have shown that *E. coli* ABU strains are producing more biofilms in urine compared to UTI strains (Ferrières *et al.*, 2007; Hancock *et al.*, 2007). Therefore we tested selected isolates of this study for biofilm formation in pooled human urine and LB medium. Although some ABU and UTI strains showed increased biofilm formation in LB or urine when compared to *E. coli* MG1655 (Figure 51; Figure 52), our results - in contrast to previously published data - did not show any significant differences for biofilm production in urine or LB. We also did not observe striking differences in between the four groups comprising ABU, HA-ABU, UTI or fecal strains (Figure 29). Thus, we cannot confirm that biofilm formation in either urine or LB is exclusively linked to ABU strains as stated by Ferrières *et al.* (2007). When comparing the biofilm producing capacity of the individual strains in LB and urine, cultivation in LB clearly resulted in stronger biofilm formation. This

is on the one hand surprising, because the ability to form biofilms is described to be enhanced in nutrient-poor media (Beloin *et al.*, 2008). On the other hand the bacteria are reaching far lower overall numbers when cultivated in urine, i.e. the observed stronger biofilm formation in LB might be simply resulting from mass effect. To address this question in the future, it might be necessary to normalize the amount of biofilm to the cell numbers in these assays.

5.8.2 Expression of components of the extracellular matrix

Since many bacterial extracellular structures like curli fimbriae or cellulose are required for adhesion and represent important components of the extracellular matrix of biofilms, we investigated the synthesis of curli fimbriae and cellulose expression of the selected *E. coli* isolates (Danese *et al.*, 2000; Prigent-Combaret *et al.*, 2000; Uhlich *et al.*, 2006). In the four different groups of isolates no consistent pattern regarding the synthesis of curli or the expression of cellulose could be identified. The observed phenotypes varied in between and among the groups, which makes it impossible to assign specific phenotypes to the individual groups of isolates (Figure 30). However, for the majority of tested isolates the results obtained on the CR plates correlated very well with the results of the CF plates. If an individual strain exhibited the *rdar* or pink morphotype, also cellulose expression was detected on the CF plates indicating that the assay is functional.

5.8.3 Motility

In the urinary tract, bacteria are exposed to strong hydrodynamic forces due to the flow of urine. Therefore active motility is a further important factor that enhances the capability of bacteria to interact with surfaces and subsequently to colonize the environment (Donlan, 2002; Verstraeten *et al.*, 2008). But motility is not only important with respect to biofilm formation. Motility enables bacterial dissemination to novel host niches like swarming up the urinary tract, to escape from detrimental environments and to move towards nutrients or other stimuli (Wei *et al.*, 2001). Moreover, motility, which is mediated by rotating flagella, is important during the infection cycle of bacterial pathogens (Ottemann & Miller, 1997). The biosynthesis of flagella involves > 50 genes and is regulated by several global regulatory circuits with FlhDC representing the master regulatory transcription factor that governs the hierarchical expression of flagella genes (Patrick & Kearns, 2012; Verstraeten *et al.*, 2008). CsrA was also shown to activate *flhDC* expression (Wei *et al.*, 2001). Therefore the strains were analyzed regarding their swarming motility on LB and urine swarm agar plates. The four different groups of isolates showed again a mixed phenotypic behaviour for individual strains either for swarming on LB or urine agar plates (Figure 32). However, a general tendency for better swarming on LB plates could be observed (Figure 31). *E. coli* 83972, which was also tested in the assay, showed a weakly motile, almost non-motile phenotype in the assay, which is

in accordance with previous observations (Hancock & Klemm, 2007). However, if the observed phenotypes directly correlate with mutations in the TCS BarA/UvrY that might affect motility via CsrA levels and subsequent activation of *flhDC* expression cannot be answered here due to the heterogeneity of the genetic backgrounds. In addition, CsrA represents one out of numerous global regulatory factors that regulate *flhDC* expression (Wei *et al.*, 2001). This complex regulatory mechanism can therefore not be reduced to one component that was shown to be involved in the complex regulation of *flhDC* expression.

5.8.4 Growth

Another phenotypic trait that was tested for the strains was their growth behaviour in pooled human urine and LB medium. *E. coli* ABU strains are well-adapted to growth in urine and the ability to grow in urine is most likely an important trait to efficiently colonize the bladder (Roos *et al.*, 2006). As exemplarily shown in the results growth in LB resulted in a typical bacterial growth curve with similar growth characteristics for the majority of the strains (Figure 33 A). However, some strains showed divergent growth kinetics with shortened lag phases or lower final ODs. In general, there was no difference in growth between the tested strains of the four different groups of isolates. However, the kinetics for growth in urine were rather different (Figure 33 B) with a lot of isolates showing an extended second lag phase after the first log phase which was followed by an increase of the bacterial population. Another set of isolates exhibited impaired or only poor growth in urine with no second increase of the bacterial population. These observations are very interesting and have not been described in the literature so far. It might be speculated that the two-phased growth behaviour (diauxic growth) represents a bacterial response that is due to the utilization of different available nutrients. The ability to efficiently switch between different carbon sources has also been proposed to be an advantage during growth by Pernestig *et al.* (2003). The detected mutations in BarA/UvrY might therefore have a pivotal impact on phenotypic traits like the tested growth behaviour.

However, a critical aspect of the performed phenotypic characterization, which diminishes the explanatory power of the phenotypic tests, is the heterogeneous genetic background of the tested isolates. The observed phenotypes can not therefore be directly attributed to the mutations detected in the BarA/UvrY TCS.

The analyzed phenotypic traits are often a result of complex regulatory networks involving a multitude of several other factors than BarA/UvrY. Therefore the measurement of a direct downstream target which is exclusively regulated by the TCS BarA/UvrY was analyzed in order to learn more about the impact of detected ns SNPs in the TCS BarA/UvrY on the functionality of the

proteins. For this purpose the expression levels of the ncRNA CsrB were measured by RT-PCR in the investigated set of strains. CsrB represents a direct downstream target of UvrY, which acts as a transcriptional regulator and directly activates the expression of *csrB* (Suzuki *et al.*, 2002).

5.9 Gene expression of *csrB*

The gene expression level of *csrB* in pooled human urine was measured by RT-PCR for selected isolates that represent the different ns SNP combinations in BarA/UvrY. Since the response regulator UvrY acts as a transcriptional activator of *csrB* transcription, *csrB* mRNA levels should be a measure of UvrY functionality (Suzuki *et al.*, 2002; Weilbacher *et al.*, 2003). Interestingly, when compared to *E. coli* strain MG1655 *csrB* expression was down-regulated in the majority of strains belonging to the ABU or fecal group. In contrast *csrB* levels were mostly up-regulated in UTI strains as well as in the tested HA-ABU strain. Therefore it will be interesting to further investigate whether strains of individual pathotypes can be characterized by a specific *csrB* expression pattern. However, due to the fact that the analysis was only based on one biological replicate per strain, the measurement can only be regarded as an indicator of possible *csrB* expression trends in the different strains tested. To answer the question of possible impacts of the detected ns SNPs on expression levels of *csrB* further RT-PCR experiments need to be carried out.

5.10 Selection of *uvrY* allelic variants

In order to learn more about the biological role of the detected ns SNPs in the TCS BarA/UvrY, we decided to analyze the function of representative alleles in a homogeneous genetic background. We focused on the analysis of *uvrY* alleles, as this component of the TCS was according to our previous investigation subjected to positive selection pressure during long-term growth in the bladder. We replaced the native *uvrY* allele of *E. coli* MG1655 with different *uvrY* alleles that were identified in our screen, including the *uvrY* allele of ABU *E. coli* 83972. Thus, our experimental procedure provided the genetic homogeneity that allows us to directly study the phenotype associated with a defined *uvrY* allele. The exchange of desired *uvrY* allelic variants was done by means of a dual counter selection approach involving homologous recombination (Li *et al.*, 2013). The selection marker, which was first used to replace with the chromosomal *uvrY* allele of *E. coli* MG1655, was the *sacB-cat* cassette which allows in a first step for selection of chloramphenicol-resistance. The *sacB-cat* cassette of positive transformants was subsequently markerless replaced by an *uvrY* allele, using the counter-selectable marker *sacB* (Figure 10). This construction procedure results in perfect allelic exchanges, i.e. the genetic modifications are strictly limited to *uvrY*.

Since *uvrY* alleles of ABU strains were found to be under positive selection pressure in the urinary tract and ABU isolates are in general longer exposed to the bladder, the selection of *uvrY* alleles to be tested was done under exclusive consideration of ABU strains. The *uvrY* alleles were selected according to several criteria. First, (i) all alleles that contain the highest amount of ns SNPs were chosen, as each ns SNP may effect protein function. Secondly, (ii) alleles with ns SNPs occurring in distinct functional domains of the protein were chosen in order to be able to learn more about the impact of these SNPs on the functionality of functional domains of UvrY. Third, (iii) at least one ns SNP per allele should be ABU-specific to be able to exclusively investigate the impact of ns SNPs that were shown to underlie positive selection. These criteria resulted in the selection of six different *uvrY* alleles to be tested plus the allele of *E. coli* 83972. The ns SNP detected in this *uvrY* allele is not restricted to ABU strains implying that this allele did not fulfill the mentioned criteria. However, since this ABU strain is very well-characterized and shown to be adapted to groth in urine we also chose this allele for the investigations. The exact positions of the SNPs and their classification into slow or fast evolving positions according to their conservation scores are shown in Table 21. The more conserved the residue (negative values) is at a position of the protein, the more important is this position for maintaining the structure or function of a protein (Valdar, 2002). The first allele from *E. coli* strain # 1972 contained one ns SNP located in the Rec domain of the UvrY protein and one non-sense SNP located in the DNA binding domain. The effect of this allele was of special interest because the stop codon led to the disruption of the DNA binding domain of UvrY which most likely prevents the appropriate transcriptional regulation of downstream targets. The second allele (# 1976) contained a ns SNP in the Rec domain located in an intermolecular recognition site (IRSs). The third allele (#1992) contained a ns SNP in the Rec domain and an additional ns SNP at a DNA binding residue of the DNA binding domain, which is responsible for the DNA-protein binding (Ofran *et al.*, 2007). The fourth *uvrY* allele (# 2000) contained ns SNPs in the same functional regions as allele no. 3, whereby the ns SNP in the DNA binding domain was located in the dimerization interface, which is involved in the dimerization and the subsequent conformational change of the RR. Due to the fact that the RR undergoes conformational changes upon phosphorylation which involves the interaction of numerous residues, it is possible that these ns SNPs have pivotal impacts on protein function (Stock *et al.*, 2000). The fifth allele (ABU prototype strain 83972) contained one ns SNP located in the Rec domain. Alleles six and seven were control alleles from *E. coli* strains # 1980 and # 2031, both containing one randomly chosen syn SNP. In addition, we re-introduced the native *uvrY* allele of *E. coli* MG1655 into *E. coli* MG1655 *uvrY*::*sacB-cat* as a technical control. The constructed *uvrY* allelic variants as well as the reference *E. coli* MG1655^{Zeo} wild type strain that was transduced with a zeocine marker and a complete *uvrY* deletion mutant that was also tested in the experiments are shown in Table 22.

In order to detect ns SNP-dependent phenotypes, the listed *E. coli* strains were subsequently tested in several relevant phenotypic assays that were described to be affected by the BarA/UvrY TCS and Csr system.

5.11 Phenotypic variations suggest "loss-of-function" as an adaptive trait

5.11.1 Biofilm formation

As already described in chapter 5.8 biofilms are considered to be a relevant virulence-associated trait of UPEC and constitute a major problem with regard to medical settings. A multitude of different factors are involved in the molecular mechanisms underlying the ability to colonize surfaces and produce biofilms (Beloin *et al.*, 2008). Interestingly, Mitra *et al.* (2013) showed that deletion of *uvrY* suppressed the biofilm formation in UPEC CFT073 whereas deletion of *csrA* did not result in decreased biofilm formation. In contrast overexpression of CsrA diminished the ability to form biofilms. The connection between UvrY and CsrA regulating biofilm formation was shown by Wang *et al.* (2005, 2004). They reported that CsrA post-transcriptionally represses the *pgaA* mRNA transcript responsible for the synthesis of a polysaccharide adhesion (PGA) required for biofilm formation. Thus, deletion of *uvrY* leads to a higher abundance of CsrA in the cell, which destabilizes the *pgaA* mRNA and consequently prohibited biofilm formation. However, when comparing the *uvrY* allelic variants with the wild type *E. coli* MG1655^{Zeo} strain, we did not observe marked differences between the individual allelic variants and the reference strains regarding their biofilm formation capacities both, in LB and pooled human urine. This was surprising, especially with respect to the *uvrY* deletion mutant MG1655 Δ*uvrY*. Therefore, at least for *E. coli* MG1655 our results do not support the published data. Therefore, the previously reported reduction in biofilm formation for cells lacking *uvrY* is obviously a background-specific phenotype, i.e. the result of the deregulation of a specific downstream target of UvrY *in E. coli* CFT073. A further striking feature of the test was the fact that all of the allelic variants, including the *uvrY* deletion mutant, showed by trend a marginal stronger biofilm formation in urine than in LB (Figure 36). This was quite unexpected, as the initially tested 47 *E. coli* wild type isolates in general showed the opposite. However, the data fit well with the literature that describes biofilm formation to be enhanced in nutrient-poor media (Beloin *et al.*, 2008). Reisner *et al.* (2006) even showed that biofilms are significantly dependent on the medium composition with LB serving as the worst medium for biofilm formation if compared to a minimal medium and to a porcine mucus-derived solution resembling a natural *in vivo* situation.

5.11.2 Expression of components of the extracellular matrix

The next phenotypic trait which was tested was the expression of major components of the extracellular matrix. The expression of curli fimbriae, for example, was reported to play a major role in biofilm formation (Bokranz *et al.*, 2005; Prigent-Combaret *et al.*, 2000). Curli are thin aggregative fimbriae that are expressed by more than 50 % of UPEC and facilitate the binding to proteins of the extracellular matrix, thereby acting as colonization factors (Olsén *et al.*, 1989). Besides these cell-surface interactions they also mediate cell-cell interactions, as also described for cellulose, a second major component of the extracellular matrix (Römling, 2002). Therefore we also tested for expression of cellulose.

The results for curli synthesis obtained in this study showed a quite uniform picture for nearly all of the tested isolates. All of the strains showed expression of curli but not of cellulose at RT and 30 °C (Figure 37; Table 23). At 37 °C no curli, but cellulose was expressed by the strains carrying *uvrY* allelic variants and by the *uvrY* deletion mutant of *E. coli* MG1655. In contrast the wild type strains as well as the strains containing allelic *uvrY* variants with syn SNPs showed expression of only curli but not of cellulose at 37 °C. However, a clear assignment to one phenotype was sometimes very challenging and it will be necessary to use a more sensitive method to more precisely quantify curli and cellulose expression in the future. The results obtained in the CR assay were further supported by the calcoflour assay, indicating that cellulose expression is indeed dependent on UvrY. In detail no fluorescent signal was observed for all of the strains at RT or 30 °C correlating with the brown phenotype whereas a signal was detected at 37 °C for the *uvrY* allelic variants no. 1 to no. 7. These results indicate that (i) the intermediate pink/brown phenotype observed at 37 °C for the *E. coli* MG1655$^{(Zeo)}$ wild type strain(s) resembles probably rather the brown morphotype if combined with the absence of a fluorescent signal on the CF plates. In contrast the allelic variants no. 1 to no. 7 are more likely to be assigned to the pink morphotype due to the positive signal on the CF plates. One puzzling aspect of the study was that the technical control strain MG1655 *uvrY*::*uvrY* MG1655 [no. 7] also exhibited a fluorescent signal on the CF plates at 37 °C. However, both phenotypic tests are subjective and sometimes hard to interpret as already mentioned above additional and more quantitative tests will be necessary to confirm the results. However, aside from the technical control the results were quite interesting because a temperature-dependent expression of cellulose at 37 °C was exclusively observed for the allelic variants with ns SNPs and the complete *uvrY* deletion mutant. This suggests that the capacity to express cellulose at body temperature might result from the *uvrY*-dependent adaptation to the urinary tract. A further conclusion is that there is apparently an evolutionary pressure towards loss-of-function of UvrY, because all of the allelic variants containing the ns SNPs showed the same phenotypes as the *uvrY* complete deletion mutant MG1655 Δ*uvrY*. Bokranz *et al.* (2005) found out that fecal *E. coli* isolates expressed curli and cellulose predominantly

at 28 °C and 37 °C whereas UTI isolates showed this concomitant expression pattern mostly at 30 °C. Based on these findings it might be speculated that the expression of cellulose at 37 °C might also represent an important adaptation strategy for long-term colonization in the bladder.

5.11.3 Motility

A further tested phenotypic trait was the motility of the strains. As already described in chapter 5.8 CsrA was shown to positively regulate the expression of the master regulatory transcription factor FlhDC that in turn controls the expression of various flagella genes and thereby promotes motility. Thus, UvrY indirectly regulates motility via CsrA (Figure 48).

All of the tested strains exhibited medium motility (+) on LB swarm agar plates (Figure 37; Table 23) indicating no regulatory differences between the different *uvrY* alleles.

In contrast motility was impaired on urine swarm agar plates exclusively for *E. coli* strain MG1655 *uvrY*::*uvrY* 83972 and the two variants containing the syn SNPs [no. 8/9]. The results obtained for the allelic variants with the ns SNPs and the *uvrY* deletion mutant MG1655 Δ*uvrY* can be explained by the regulatory interactions between UvrY and the Csr system (Figure 48). Lower abundance or functionality of UvrY, which might be correlated with the mutations, leads to a lower expression of CsrB and a higher abundance of CsrA resulting in increased motility according to the findings described by Wei *et al.* (2001). In contrast, normal or increased functionality or abundance of UvrY leads to an accumulation of CsrB sequestering CsrA and thereby inhibiting the positive effects on motility. Assuming that the syn SNPs do not affect *csrB* expression levels this might explain the findings for the allelic variants containing the syn SNPs in urine. Surprisingly, the allelic variant MG1655 *uvrY*::*uvrY* 83972 also showed decreased motility on urine which cannot be correlated with the afore mentioned regulatory interactions if it is assumed that the ns SNP also leads to impaired function of UvrY. However, the results for *E. coli* MG1655 *uvrY*::*uvrY* 83972 correlate very well with the literature characterizing *E. coli* 83972 as non-motile. Last but not least, motility could also represent an adaptive trait of strains inhabiting the urinary tract for longer periods. This might be

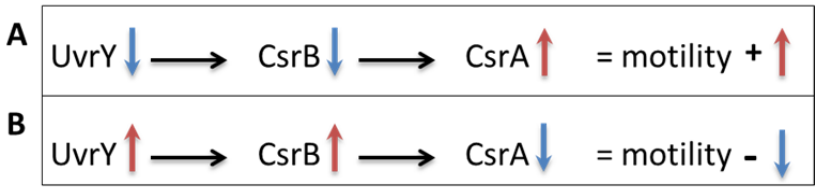


Figure 48: Schematic demonstration of the regulatory interactions between UvrY and the Csr components CsrB and CsrA and its effects on motility.

(A) Lower expression of *uvrY* results in decreased transcriptional activation of *csrB*. CsrB therefore cannot build a complex with CsrA which in turn positively activates motility via *flhDC* expression (Wei *et al.*, 2001). (B) Functional UvrY activates transcription of *csrB*. CsrB sequesters CsrA thereby inhibiting positive regulation of motility. Blue arrows indicate a decreased expression and a negative phenotype and red arrows vice versa.

explained by the fact that motility helps the bacteria to counteract hydrodynamic forces present in the bladder. The aspect that all of the tested variants showed motility on the urine agar plates except for the variants containing the *uvrY* allele of *E. coli* 83972 and the variants containing the syn SNPs might underline this idea.

5.11.4 Growth kinetics

Next, we compared the growth characteristics of the strains carrying the different *uvrY* alleles with the wild type strain MG1655^{Zeo} in pooled human urine and LB (Figure 38).

The growth properties of the wild type strain and the strains with *uvrY* allelic variant in LB (Figure 38 A) were comparable (i) between the strains and (ii) similar to those detected for the 47 *E. coli* isolates tested before (Figure 33 A). In detail a typical bacterial growth curve could be observed for all of the strains.

Interestingly, the growth kinetics in urine (Figure 38 B) showed for all of the tested strains a similar pattern with a prolonged second lag phase and a subsequent second growth phase, which was in part already observed for some of the strains from the initial strain panel before (Figure 33 B). This growth behaviour of the bacterial population was quite interesting and seemed to be a special feature of *in vitro* growth in urine that has not been described so far. As already discussed in chapter 5.8, this two-phased growth pattern might reflect a bacterial response to the changing nutrient availability in urine which requires efficient switching between different metabolic pathways as proposed by Pernestig *et al.* (2003). They could show that *uvrY* knockout mutants of UPEC strains had marked competition advantages compared to the isogenic wild type strain when grown in LB or a gluconeogenic medium. In contrast, growth in a glycolytic medium led to a clear growth advantage of the wild type, which suggested that a functional BarA/UvrY TCS was essential for the efficient switch between glycolytic and gluconeogenic carbon sources. With respect to our findings we cannot support this statement at all. Since urine is a very nutrient limited medium and it is assumed that amino acids and small peptides represent the main nutrient source, we propose that an optimal adjustment of the metabolism, like elevated gluconeogenic pathways, is required for optimal growth in urine. The positive selection of ns SNPs resulting in a loss-of-function of UvrY might therefore favor gluconeogenesis and thus promotes a better competitiveness under this conditions. The fact that gluconeogenesis represents a very important metabolic pathway used during growth in urine is underlined (i) by our findings of carbon being a limited source in urine and further (ii) supported by others stating the gluconeogenesis and the TCA cycle are required for fitness of *E. coli* during urinary tract infections (Alteri *et al.*, 2009).

Since the individual cultures of strains with *uvrY* allelic variants exhibited no apparent difference in growth kinetics compared to the wild type strain MG1655^{Zeo}, competition assays were carried out in LB and pooled human urine. This method has more discriminatory power in order to monitor putative fitness defects or advantages for the allelic variants in comparison to the wild type allele.

5.12 Competition assay

Since the well-adapted *E. coli* ABU strain 83972 as well as further tested ABU strains were shown to successfully outcompete UPEC strains in human urine (Roos, Nielsen, *et al.*, 2006; Roos, Ulett, *et al.*, 2006), the effect of the *uvrY* allelic variants containing the SNPs as well as the *uvrY* deletion mutant should be elucidated in pair-wise competition experiments with the wild type strain *E. coli* MG1655 WT^{Zeo} in LB and pooled human urine.

The results for LB clearly indicated that fitness of the different tested strains in comparison to the wild type strain changed during the different phases of growth (Figure 39.) Interestingly, all of the strains carrying the *uvrY* alleles with the ns SNPs were able to outcompete the wild type strain already after 24 hours of growth in LB to various extents, except for *E. coli* MG1655 *uvrY*::*uvrY* 83972. This strain and the wild type strain were equally competitive and existed after 24 hours of growth in equal proportions in the population. In individual cases, even a takeover of the culture was observed when the allelic variant was the minority in the initial inoculum LB (Figure 39 C). These results are in accordance with previously published data that also showed a growth competition advantage for *uvrY* knockout UPEC strains in comparison to the respective wild type strain in LB (Pernestig *et al.*, 2003; Tomenius *et al.*, 2006). In contrast to the ns SNPs a takeover of the cultures in LB after 24 hours was observed for the wild type when in competition with the allelic variants containing the syn SNPs.

We observed the same trend for the competiton experiments in urine, whereby the *uvrY* deletion mutant MG1655 Δ*uvrY* was not as fit as in LB (Figure 40). A more detailed investigation of the competitive behaviour after 24 hours of growth confirmed these results (Figure 41). Thus, the tested allelic variants containing ns SNPs of *uvrY* exhibited a competitive growth advantage in both, LB and pooled human urine. Moreover, the fact that the *uvrY* deletion mutant showed a similar phenotype as the allelic ns SNP variants suggests a loss-of-function as a result of these *uvrY* mutations. Therefore, *E. coli* ABU strains were obviously subjected to an evolutionary pressure that favors bacterial cells with abolished UvrY function. That growth competition advantages in comparison to UPEC isolates are an important phenotype of ABU isolates was also shown by others (Roos, Ulett, *et al.*, 2006). These authors showed that a competitive advantage was resulting from a higher growth rate of the ABU strain 83972 in urine when compared to different UPEC strains. The exchange of the

uvrY alleles in our study did not result in higher growth rates in LB or urine, as can be seen from the individual growth curves of the strains shown in Figure 38. Therefore a simple outgrowth of the wild type strain due to shorter doubling times of the strains carrying allelic variants of *uvrY* in the competition experiments appears to be very unlikely, especially as the strains carrying ns SNPs apparently took over at later stages of growth in the competition experiments (Figure 39; Figure 40). In order to completely exclude this possibility, a detailed analysis of cell shape and of viable cell numbers per OD unit over the course of the experiment would be necessary. However, based on the above described observations and due to the fact that the TCS BarA/UvrY is involved in the regulation of a variety of different metabolic pathways, we speculate that mutations in *uvrY* simply lead to a better utilization of diverse nutrients during growth in LB or urine. The same conclusion regarding the importance of an efficient utilization of available nutrients was drawn by Tomenius *et al.* (2006) who infected a monkey bladder *in vivo* with a UPEC wild type and a *uvrY* deletion derivative strain. Since in their experiments the wild type strain outcompeted the *uvrY* deletion mutant *in vivo* they attributed the characteristic of better utilization of nutrients to the wild type strain with an intact UvrY protein. However, *in vitro* competition in pooled human urine over 8 days showed different results dependent on the urine charge used with the mutant strain sometimes being outcompeted and sometimes not. Moreover, if transferring the competitors after 7 days from urine to LB, a clear advantage was shown again for the *uvrY* deletion mutant with takeover rates of nearly 100 % (Tomenius *et al.*, 2006). These results fit very well with our observations.

In summary, our results do not support the idea that intact UvrY leads to growth advantages in LB or urine. We rather propose that mutations of *uvrY* towards loss-of-function represent an adaptational mechanism of bacteria for efficient growth in urine. This hypothesis is further strengthened by the results of the sub-cultivation experiments done in both, LB and urine over a period of four days.

5.13 Sub-cultivation

The sub-cultivation experiments in LB revealed that all of the strains carrying allelic variants with ns SNPs in *uvrY* showed takeover rates of nearly 100 %, similar to the *uvrY* deletion mutant MG1655 Δ*uvrY* (Figure 43 A). Surprisingly, the variants with the syn SNPs also showed a significantly increased fitness. Due to the fact that the *uvrY* variant carrying strains have not been re-sequenced after the experiments it cannot be excluded that other mutations might have arisen that could lead to the observed phenotypes. However, this appears extremely unlikely, as all strains carrying the allelic variants of uvrY show similar phenotypes. Anyhow, the takeover of the culture for the variants with the syn SNPs were not as strong as observed for the majority of the strains with the ns SNPs.

Therefore it might still be concluded that the strains carrying the ns SNPs show a higher competitiveness than those with the syn SNPs.

With respect to the results of the sub-cultivation experiments in urine it can be stated that except for the *uvrY* allele of *E. coli* 83972 the same tendency was observed as in LB, suggesting that the takeover in competition experiments is a general phenotype mediated by the *uvrY* alleles and not limited to growth in rich medium (Figure 43 B). Since the strain carrying the *uvrY* allele of *E. coli* 83972 did not show a significant takeover in urine it might be speculated that this strain exhibits other *uvrY*-independent adaptational strategies for growth in urine which allow a successful long-term colonization of the bladder. However, in contrast to LB the allelic variants with the syn SNPs did not show any competitive advantage in urine, limiting the competitive advantages of theses alleles to a nutrient rich environment.

Consequently, the sub-cultivation experiments also underlined the above described conclusions that adapted UvrY response regulators lead to improved fitness of the strains in LB and urine.

5.14 Mutated *uvrY* exhibits impact on the Csr system

5.14.1 Comparison of the expression of components of the Csr system under standard laboratory conditions

In order to learn more about the *in vivo* function of the *uvrY* allelic variants, we measured the expression levels of the main components of the Csr sytem which were known downstream targets of UvrY. To get a comprehensive picture of the effect of the *uvrY* alleles, we first analyzed the expression levels of the Csr genes when the respective strains are grown in LB and compared the mRNA levels with the isogenic strain MG1655 *uvrY*::*uvrY* MG1655 and MG1655 Δ*uvrY* by means of RT-PCR.

The observed down-regulation of *csrB* for the *uvrY* allelic variants grown in LB suggests that the ns SNPs in UvrY negatively affect the function of the protein as transcriptional activator of *csrB* expression (Figure 44 A). An interesting exception was the *uvrY* allelic variant MG1655 *uvrY*::*uvrY* 83972, that did not result in down-regulation of *csrB*. Previous studies showed a strongly reduced *csrB* expression in strains with deletions in either *uvrY*, *barA* or *csrA* when grown in rich medium (Gudapaty *et al.*, 2001; Jonas & Melefors, 2009; Jonas *et al.*, 2006; Suzuki *et al.*, 2002). Due to the fact that the down-regulation of *csrB* mRNA in the strains carrying the allelic variants of *uvrY* is comparable to the level of down-regulation in the *uvrY* deletion mutant, the ns SNPs in *uvrY* most likely result in a loss-of-function as transcriptional activator. Surprisingly, the allelic variant

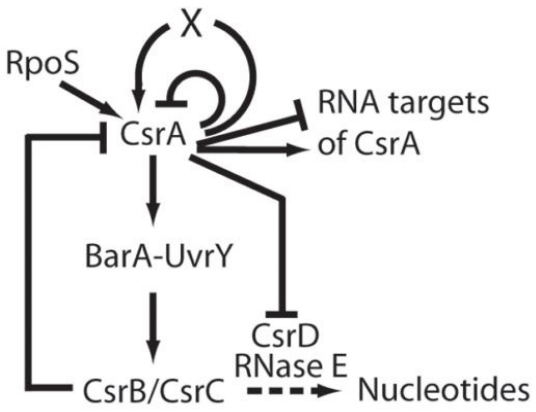


Figure 49: Model of the Csr regulatory circuitry (Yakhnin *et al.*, 2011).

CsrA represses expression of target mRNAs, including its own transcript, by inhibiting translation initiation. It is also capable of activating expression of mRNAs by unknown mechanisms. RpoS activates transcription of *csrA*. CsrA indirectly activates transcription of its RNA antagonists CsrB and CsrC via the BarA/UvrY TCS. Phosphorylated UvrY directly activates transcription of csrB and csrC. CsrB and CsrC contain several CsrA binding sites such that each sRNA is capable of sequestering several CsrA dimers. CsrA represses expression of CsrD, a protein that is required for degradation of CsrB and CsrC by RNase E. Finally, CsrA indirectly activates its own transcription by an unknown mechanism.

MG1655 *uvrY*::*uvrY* 2031 containing a syn SNP also showed a significant up-regulation of *csrB* transcript levels. As this cannot be attributed to an altered function of the protein, this syn SNP most likely affects the level of UvrY expression. Notably, we also detected the highest *uvrY* mRNA levels for the strain harboring the *uvrY* 2031 allele in LB (Figure 44 E), suggesting that this syn SNP might enhance *uvrY* promoter activity, or more likely the stability of the *uvrY* mRNA. An additional more efficient translation of this allele could of course also contribute to increased UvrY levels in this strain. As CsrA expression is also controlled by a negative feedback loop, the observed elevated levels of *csrA* mRNA in MG1655 *uvrY*::*uvrY* 2031 [no.9] might simply reflect the requirement to produce more CsrA protein in order to quench the excess of free *csrB* transcripts (Figure 44 D; Figure 49).

Regarding the expression levels of *csrC* no significant differences could be observed for all of the tested strains compared to the reference strain (Figure 44 B). This is a strong indicator that *csrC* expression is in addition to that of *uvrY*, regulated by other, so far unknown factors. CsrC levels were also found to be less dependent on UvrY than CsrB levels as shown by others (Weilbacher *et al.*, 2003). Thus, it is not suprising that we did not detect significant differences in CsrC levels under standard laboratory conditions.

As for *csrC*, no significant differences in *csrD* transcript levels were observed in the majority of the strains (Figure 44 C). CsrD is a protein involved in the degradation of CsrB and CsrC (Suzuki *et al.*, 2006). *E. coli* MG1655 *uvrY*::*uvrY* 1992 [no. 5] showed a significant down-regulation of *csrD* transcripts, which might be simply a response to the overall low *csrB*/C levels in this background. The fact that – even though not statistically significant – higher amounts of *csrD* mRNA levels were detected in MG1655 *uvrY*::*uvrY* 2031, the strain with the highest *csrB*/*C* transcript levels, further supports this hypothesis.

A significant up-regulation of *csrA* mRNA was detected for the allelic variant MG1655 *uvrY*::*uvrY* 2031 [no. 9], which is discussed above (Figure 44 D). A significant down-regulation of *csrA* mRNA was

observed for the *uvrY* deletion mutant and the allelic variant strain MG1655 *uvrY*::*uvrY* 1976 [no.4]. Since the regulation of *csrA* expression has been shown to occur by a variety of mechanisms, the known regulatory interactions of the Csr system are once more shown in a schematic overview in Figure 49. CsrA regulation involves several mechanisms: (i) CsrA binds to its own leader transcript thereby reperessing its own translation, (ii) CsrA activates transcription of its own gene by a so far unknown mechanism and (iii) CsrA indirectly activates the transcription of *csrB*/*csrC* via the TCS BarA/UvrY resulting in the stabilization of its own antagonists (Yakhnin *et al.*, 2011). This complex, partially autoregulatory mechanism was suggested to ensure optimal CsrA levels in *E. coli*. The reduced *csrA* mRNA levels in the *uvrY* mutant is therefore not surprising and fits well to previous observations of this system. Furthermore, the reduced *csrA* mRNA levels in the strain harboring the *uvrY* allele 1976 [no. 4] in combination with the expression pattern of the other tested genes that are very similar to the expression patterns observed in the *uvrY* deletion mutant lend additional support to the hypothesis that the ns SNP in *uvrY* allele 1976 results in a loss-of-function of the protein.

With respect to UvrY expression, the tested strains showed no significant differences in comparison to the reference strain which was expected, because so far no auto-regulatory mechanism was described for *uvrY* expression.

5.14.2 Comparison of the expression of components of the Csr system in urine

Next, we analyzed the effect of the *uvrY* alleles on the Csr system when the strains were grown in urine. The results for urine differed partly from the ones observed in LB (Figure 45). A direct impact of UvrY on *csrB* expression could be observed for all of the tested allelic variants, including the *uvrY* deletion mutant MG1655 Δ*uvrY* [no. 1] (Figure 45 A). Interestingly, the allelic variants MG1655 *uvrY*::*uvrY* 1976 [no. 4] and MG1655 *uvrY*::*uvrY* 1992 [no. 5], both containing ns SNPs at slow evolving and therefore highly conserved sites (Table 21), exhibited either in LB or urine the strongest down-regulation of *csrB* transcript levels. This might indicate that ns SNPs occurring in slow evolving positions of UvrY have the strongest impacts on protein function, at least with respect to its function in transcriptional activation of *csrB* transcription. The fact that the allelic variants containing the syn SNPs in *uvrY* [no. 8/9] showed a significant decrease of *csrB* levels might be a hint that syn SNPs can also effect the transcription levels in a growth-condition-dependent manner and most likely by affecting UvrY levels on a posttranscriptional level (compared to the results obtained for the MG1655 *uvrY*::*uvrY* 2031 under standard laboratory conditions).

The expression levels of *csrC* were varying to different extents in between the different allelic variants with some of them showing a significant down-regulation, but none exhibiting increased expression of *csrC* (Figure 45 B). With regard to the fact that transcription of *csrC* is also activated by

UvrY the partial down-regulation of *csrC* transcript levels that was not observed when the cells were grown in LB, fits to the expression levels observed for *csrB*.

Expression of *csrD* showed a significant up-regulation in strain MG1655 *uvrY*::*uvrY* 1972 [no. 3], which might be explained by a requirement for a more efficient degradation of CsrC as *csrA* mRNA and consequently most likely CsrA protein levels are reduced in this background (Figure 45 C). In contrast to that transcript levels of *csrD* in *E. coli* MG1655 *uvrY*::*uvrY* 1976 [no. 4] were down-regulated, which could be due to the overall reduced CsrB and CsrC levels in this strain.

CsrA mRNA levels were reduced for all of the strains except for MG1655 *uvrY*::*uvrY* 83972 [no. 7] and MG1655 *uvrY*::*uvrY* 1980 [no. 8] (Figure 45 D). Since only *csrA* mRNA levels were measured and CsrA expression is also regulated at the posttranscriptional level, the results have to be interpreted carefully here and CsrA concentrations will have to be determined directly in the future in order to be able to assess the biological significance of these differences on mRNA level. Due to the fact that CsrA was described to inhibit gluconeogenesis, which, as we suggest, represents a very important pathways in urine for energy production, it might be possible, that CsrA levels are indeed low in the strains containing the allelic variants of *uvrY* in comparison to the reference strain.

Regarding *uvrY* no signal was detected in the *uvrY* deletion mutant, as expected (Figure 45 E). Interestingly, we could observe an up-regulation of *uvrY* mRNA levels in *E. coli* MG1655 *uvrY*::*uvrY* 1972. Since this *uvrY* variant contained a stop codon it might be speculated that an autoregulatory circuit exists also for *uvrY* expression. Increased *uvrY* transcription in this strain background may represent a means to compensate for the loss of UvrY function by increased expression of this mutated gene.

Taken together, we were able to show a clear impact of the *uvrY* allelic variants on the expression levels of different components of the Csr system, both under standard laboratory growth conditions and in urine. The general effects were stronger in pooled human urine than in LB and are indicative of a loss-of-function of the *uvrY* alleles containing ns SNPs. The fact that urine is a carbon limited medium and that gluconeogenesis is most likely the most important pathway for energy production of *E. coli* in such an environment fits well with the observation that the BarA/UvrY system, that indirectly inhibits this pathway via the Csr system, is often found to be mutated in isolates adapted to the urinary tract. Mutations in *uvrY* might be simply less detrimental than the well-known pleiotropic consequences of mutations in the Csr system, which might explain why the mutations do not occur directly in the genes of the Csr system.

6 Conclusions and outlook

In the present study, we investigated adaptational strategies of *E. coli* strains promoting growth in urine. A special emphasis was put on the TCS BarA/UvrY, which was shown to be involved in the regulation of a multitude of cellular functions and metabolic pathways.

An initial metabolomic analyses, as well as the assessment of growth characteristics of *E. coli* strains in urine suggested that carbon is the major limiting nutrient. Sequence analysis of the genes encoding the TCS BarA/UvrY, which is involved in the regulation of carbon metabolism via the Csr system, indicated an increased genetic variability for *barA* and *uvrY*. By analyzing the DNA sequence of both genes in *E. coli* isolates from urine and feces, we were able to show that *uvrY* is subjected to positive selection pressure during growth in the urinary tract. Different relevant phenotypic assays with isogenic *E. coli* strains carrying different *uvrY* allelic variants showed that ns SNPs provided growth advantages in urine. These results suggest that an intact UvrY response regulator seems to be detrimental for growth of *E. coli* in the urinary tract. Moreover, we provide evidence that expression of *csrB*, encoding a component of the Csr system, is usually down-regulated in the presence of the tested ns SNPs in *uvrY* when compared to an isogenic *E. coli* strain carrying the wild type allele.

Taken together, our results indicate that the described non-synonymous mutations in *uvrY* are indeed resulting in a growth advantage in urine and that this growth advantage is most likely due to a more efficient utilization of the spectrum of available nutrients in urine.

In order to clarify the mechanistic basis of this growth advantage, more detailed studies with media containing defined combinations of nutrients that mimic the contents of urine will be necessary. Another very important question that has to be addressed is, to which extent oxygen availability plays a role in the urine growth advantage of *E. coli* strains with compromised BarA/UvrY TCS. A further study under anaerobic conditions should be performed, as fermentative growth in the absence of oxygen is more likely to resemble the natural situation of the urinary tract. One might therefore speculate that the growth advantage of the strains carrying the allelic variants of *uvrY* might be even more significant under such conditions. Moreover, it will be interesting to characterize the effects of the ns SNPs on the functionality of UvrY on the molecular level in order to clarify the question, if the observed positive selective pressure on *uvrY* in the urinary tract is simply directed towards a loss-of-function of UvrY or rather towards a modification of its activity. Last but not least it will be of course necessary to answer the question, if the observed *in vitro* growth advantages mediated by the allelic variants of *uvrY* are also observed *in vivo*, for which a suitable animal model will have to be employed.

7 Appendix

7.1 Supplements

7.1.1 Amplification and sequencing of *barA/uvrY*

The approximate positions of the chromosomal insertions that were found in the samples for *uvrY* are shown in the genomic context for each strain.

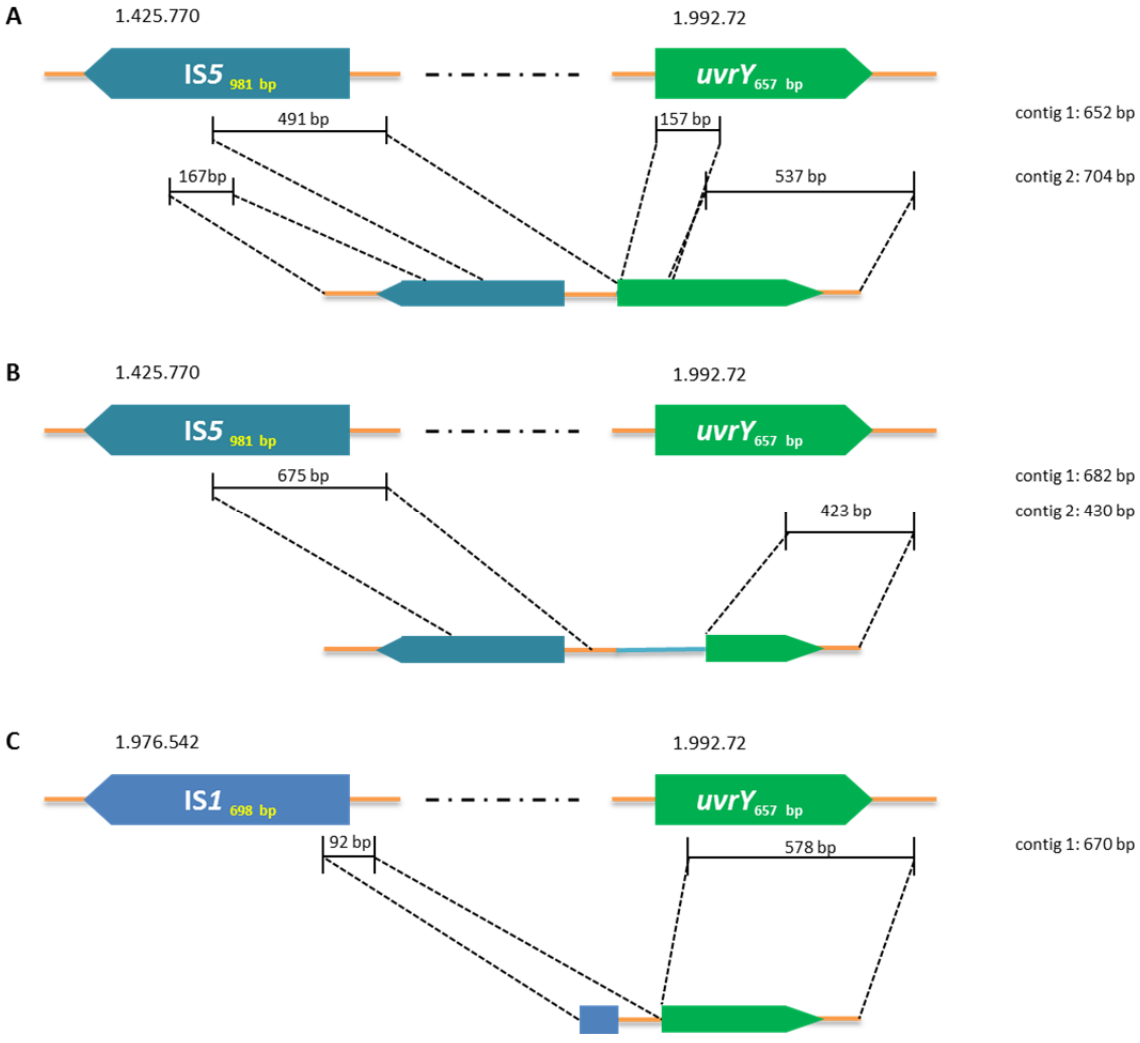


Figure 50: The construction of *uvrY* based on the sequencing alignment.

Positions of the single elements are shown in the genomic context. The deduced order for *uvrY* is shown for the *E. coli* strains (A) # 2055, (B) # 1964 and (C) # 2076.

7.1.2 Evolutionary rates

7.1.2.1 UvrY

Table 25: Evolutionary rates of UvrY on the protein level.

The single positions with the respective amino acid and the according evolutionary rate ist shown. The lower the value the more conserved the position.

#POS	SEQ	SCORE	#POS	SEQ	SCORE	#POS	SEQ	SCORE
1	M	-1.551	40	V	-0.1796	79	I	-0.7569
2	I	-0.5752	41	K	1.916	80	M	-0.6844
3	N	-0.04314	42	W	1.917	81	L	-0.8217
4	V	-1.163	43	C	-0.06477	82	T	-1.425
5	L	-0.9482	44	R	1.917	83	V	-1.189
6	L	-0.6484	45	T	1.676	84	H	-0.3476
7	V	-1.135	46	N	1.912	85	T	0.002751
8	D	-1.377	47	A	1.818	86	E	-0.2603
9	D	-1.543	48	V	-0.5662	87	N	-0.3619
10	H	-1.275	49	D	-1.166	88	P	0.9504
11	E	0.1314	50	V	-0.8376	89	L	-0.5606
12	L	0.1034	51	V	-0.566	90	P	-0.771
13	V	-0.758	52	L	-0.8459	91	A	-0.2067
14	R	-0.8923	53	M	-1.134	92	K	0.6194
15	A	0.6	54	D	-1.62	93	V	-1.154
16	G	-0.2044	55	M	-0.5005	94	M	-0.7013
17	I	-0.5824	56	S	-0.8439	95	Q	-0.05368
18	R	0.5224	57	M	-1.353	96	A	0.3507
19	R	1.545	58	P	-1.44	97	G	-1.534
20	I	0.1194	59	G	0.2181	98	A	-1.485
21	L	-1.044	60	I	-0.3427	99	A	-0.7256
22	E	0.6639	61	G	-0.7746	100	G	-1.508
23	D	1.813	62	G	-1.62	101	Y	-1.012
24	I	0.255	63	L	-0.1375	102	L	-0.5777
25	K	0.596	64	E	-0.1148	103	S	-0.5112
26	G	-0.25	65	A	-0.5741	104	K	-1.621
27	I	-0.2168	66	T	-0.6161	105	G	-1.23
28	K	-0.02169	67	R	-0.1144	106	A	-0.5685
29	V	-0.4116	68	K	0.6151	107	A	0.05041
30	V	-0.8924	69	I	-0.6753	108	P	0.4948
31	G	0.3746	70	A	-0.9443	109	Q	0.6342
32	E	0.3537	71	R	0.969	110	E	-0.6511
33	A	-0.8148	72	S	1.745	111	V	-0.5341
34	S	1.908	73	T	0.8649	112	V	0.4922
35	C	-0.645	74	A	0.5927	113	S	-0.2306
36	G	-0.8395	75	D	1.207	114	A	-0.9908
37	E	1.915	76	V	0.1213	115	I	-0.3461
38	D	0.3572	77	K	-1.266	116	R	0.2517
39	A	-1.093	78	I	-1.079	117	S	-0.4016

#POS	SEQ	SCORE
118	V	0.2329
119	Y	0.3063
120	S	-0.4365
121	G	-1.191
122	Q	0.7834
123	R	1.911
124	Y	1.912
125	I	0.3025
126	A	0.6195
127	S	1.87
128	D	1.392
129	I	1.034
130	A	-0.07946
131	Q	1.911
132	Q	0.6748
133	M	0.1785
134	A	1.917
135	L	1.891
136	S	1.872
137	Q	1.825
138	I	1.231
139	E	0.4708
140	P	1.916
141	E	1.916
142	K	1.847
143	T	1.917
144	E	1.914
145	S	1.914
146	P	0.629
147	F	0.4644
148	A	1.079
149	S	1.398
150	L	-0.8138
151	S	-0.8654
152	E	0.8395
153	R	-0.2874
154	E	-1.081
155	L	0.7246
156	Q	0.4712
157	I	-0.4758
158	M	-0.6885
159	L	1.794
160	M	1.914
161	I	-0.1819
162	T	0.2817
163	K	1.909
164	G	-0.08929
165	Q	0.6324
166	K	-0.699
167	V	-1.089
168	N	-0.2312
169	E	0.3805
170	I	-1.218
171	S	-0.8103
172	E	0.9799
173	Q	1.362
174	L	-0.7297
175	N	0.7347
176	L	0.05815
177	S	-0.7425
178	P	-0.07401
179	K	-0.6944
180	T	-1.093
181	V	-0.9614
182	N	-0.6952
183	S	-0.7165
184	Y	-0.7821
185	R	-0.7406
186	Y	-0.3219
187	R	-0.1759
188	M	-1.014
189	F	-0.9581
190	S	0.7021
191	K	-1.254
192	L	-0.7804
193	N	0.2956
194	I	-0.2191
195	H	1.619
196	G	0.02815
197	D	0.1268
198	V	0.02489
199	E	0.2385
200	L	-0.07009
201	T	0.4442
202	H	-0.6761
203	L	0.9642
204	A	-1.289
205	I	-0.1694
206	R	-0.8088
207	H	-0.5249
208	G	-0.388
209	L	-0.9035
210	C	-0.4451
211	N	0.6018
212	A	-0.5162
213	E	1.445
214	T	0.7336
215	L	0.4811
216	S	-0.2428
217	S	0.5605
218	Q	-0.521

7.1.2.2 BarA

Table 26: Evolutionary rates for BarA on the protein level. The evolutionary score is shown for each position of the protein.

#POS	SEQ	SCORE
1	M	0.3605
2	T	0.9168
3	N	-1.047
4	Y	-0.1395
5	S	-1.256
6	L	-0.275
7	R	-0.7123
8	A	-0.4681
9	R	0.7438
10	M	-0.7581
11	M	-0.1652
12	I	-0.04684
13	L	-0.6407
14	I	0.3662
15	L	-0.6242

#POS	SEQ	SCORE
16	A	-0.1468
17	P	0.8685
18	T	-0.7128
19	V	-0.02514
20	L	-0.065
21	I	-0.9231
22	G	-0.2933
23	L	1.118
24	L	0.06137
25	L	-0.7774
26	S	-0.7511
27	I	0.1148
28	F	1.093
29	F	-0.03488
30	V	0.3294
31	V	0.6489
32	H	0.9758
33	R	0.8693
34	Y	0.7798
35	N	-0.4104
36	D	0.8352
37	L	-0.4189
38	Q	0.272
39	R	0.7457
40	Q	-0.476
41	L	0.6073
42	E	0.3924
43	D	0.07671
44	A	0.5577
45	G	0.0261
46	A	1.203
47	S	0.7167
48	I	0.6554
49	I	-1.228
50	E	0.5888
51	P	1.104
52	L	-0.4677
53	A	-0.5739
54	V	0.8684
55	S	0.5358
56	T	0.1702
57	E	-1.078
58	Y	0.9728
59	G	0.2326
60	M	1.05
61	S	0.8154
62	L	0.8227
63	Q	0.5835
64	N	0.7278
65	R	0.9974
66	E	0.6824
67	S	0.4469
68	I	1.14
69	G	1.148
70	Q	0.5416
71	L	-0.1669
72	I	1.165
73	S	-0.0253
74	V	0.1053
75	L	0.3364
76	H	0.8701
77	R	-0.7034
78	R	-0.5351
79	H	1.423
80	S	-0.09059
81	D	0.435
82	I	0.3654
83	V	-0.5703
84	R	0.3637
85	A	1.01
86	I	0.5085
87	S	-0.4114
88	V	1.212
89	Y	1.523
90	D	0.3116
91	E	1.355
92	N	0.7815
93	N	0.4402
94	R	0.2785
95	L	0.06328
96	F	-0.4988
97	V	1.222
98	T	0.3139
99	S	0.5641
100	N	1.4
101	F	0.7427
102	H	0.3933
103	L	-0.1037
104	D	0.4389
105	P	-0.2274
106	S	1.504
107	S	0.08255
108	M	0.1665
109	Q	1.796
110	L	0.4118
111	G	0.6347
112	S	0.5553
113	N	1.019
114	V	1.492
115	P	1.561
116	F	1.299
117	P	-0.0192
118	R	1.402
119	Q	0.3976
120	L	0.445
121	T	0.6541
122	V	-0.18
123	T	-0.04061
124	R	0.3336
125	D	1.152
126	G	-0.9309
127	D	1.394
128	I	0.109
129	M	1.239
130	I	-0.2533
131	L	1.198
132	R	0.3667
133	T	0.2005
134	P	-0.1747
135	I	0.3219
136	I	-0.1072
137	S	-0.2777
138	E	0.7422
139	S	-0.7313
140	Y	-0.1069
141	S	0.1105
142	P	-0.1713
143	D	0.2013
144	E	0.531
145	S	-0.5549
146	P	0.3493
147	S	-0.6384
148	S	-1.045
149	D	0.04309
150	A	-0.9422
151	K	0.1978
152	N	0.003254
153	S	0.6917

#POS	SEQ	SCORE
154	Q	-0.2714
155	N	-0.8221
156	M	-0.5554
157	L	0.7312
158	G	0.1013
159	Y	-0.3954
160	I	-0.8127
161	A	-0.09453
162	L	-0.59
163	E	0.5558
164	L	-0.3166
165	D	-0.3142
166	L	0.6262
167	K	-0.1825
168	S	0.6773
169	V	0.2335
170	R	-0.1603
171	L	0.4218
172	Q	-0.0102
173	Q	0.03619
174	Y	0.2412
175	K	-0.3757
176	E	0.7316
177	I	0.6487
178	F	0.6122
179	I	1.492
180	S	-0.3698
181	S	-0.1383
182	V	1.095
183	M	-0.036
184	M	-0.1994
185	L	0.8812
186	F	-0.4549
187	C	1.711
188	I	0.3443
189	G	-0.02913
190	I	0.4733
191	A	0.3834
192	L	1.418
193	I	-0.2332
194	F	1.485
195	G	0.8598
196	W	-0.1954
197	R	1.135
198	L	-0.2522
199	M	0.5951

#POS	SEQ	SCORE
200	R	-1.032
201	D	-0.08604
202	V	1.45
203	T	-0.557
204	G	0.03047
205	P	0.527
206	I	0.1457
207	R	0.4852
208	N	0.6296
209	M	0.9296
210	V	1.215
211	N	0.7786
212	T	0.5362
213	V	1.451
214	D	-0.08354
215	R	0.913
216	I	0.9384
217	R	1.512
218	R	1.476
219	G	0.6375
220	Q	1.477
221	L	0.3956
222	D	0.6323
223	S	0.7185
224	R	1.66
225	V	0.9637
226	E	1.604
227	G	0.7741
228	F	0.8467
229	M	-0.1909
230	L	1.439
231	G	0.7708
232	E	0.4843
233	L	1.246
234	D	0.3582
235	M	1.122
236	L	0.3512
237	K	1.538
238	N	1.818
239	G	0.1232
240	I	-0.007677
241	N	-0.6357
242	S	1.543
243	M	-0.3101
244	A	0.6051
245	M	0.339

#POS	SEQ	SCORE
246	S	1.08
247	L	-0.1087
248	A	-0.2674
249	A	1.554
250	Y	1.261
251	H	0.4726
252	E	1.1
253	E	0.8114
254	M	1.127
255	Q	1.521
256	H	1.812
257	N	0.1949
258	I	0.5251
259	D	0.4293
260	Q	0.332
261	A	0.457
262	T	0.00992
263	S	1.899
264	D	0.9608
265	L	-0.3005
266	R	0.9067
267	E	0.2501
268	T	-0.2356
269	L	0.3903
270	E	0.2503
271	Q	0.6027
272	M	1.127
273	E	-0.03902
274	I	0.9366
275	Q	-0.4364
276	N	0.07098
277	V	0.3372
278	E	-0.6023
279	L	-0.09505
280	D	-0.1917
281	L	0.8052
282	A	-1.37
283	K	-0.9222
284	K	-0.6231
285	R	0.2676
286	A	-1.623
287	Q	-0.7872
288	E	0.1598
289	A	-1.303
290	A	-0.4998
291	R	-0.297

#POS	SEQ	SCORE
292	I	-0.5763
293	K	-1.49
294	S	-1.219
295	E	-0.531
296	F	-1.709
297	L	-1.045
298	A	-2.06
299	N	-1.51
300	M	-1.429
301	S	-2.017
302	H	-2.051
303	E	-1.919
304	L	-1.681
305	R	-1.831
306	T	-1.867
307	P	-1.845
308	L	-0.7834
309	N	-0.9753
310	G	-1.043
311	V	-1.352
312	I	0.2719
313	G	-1.656
314	F	-0.7291
315	T	-0.9721
316	R	-0.5623
317	L	-0.8505
318	T	-1.124
319	L	1.052
320	K	-1.219
321	T	-0.9779
322	E	0.341
323	L	-0.1608
324	T	0.7061
325	P	1.744
326	T	0.7155
327	Q	-1.058
328	R	-0.02543
329	D	-0.2983
330	H	1.144
331	L	-1.243
332	N	-0.09357
333	T	-1.064
334	I	-0.7561
335	E	0.9473
336	R	1.271
337	S	-1.671

#POS	SEQ	SCORE
338	A	-0.9718
339	N	0.2651
340	N	-0.3075
341	L	-1.522
342	L	-0.4891
343	A	-0.2483
344	I	-1.156
345	I	-1.599
346	N	-1.923
347	D	-1.619
348	V	-1.241
349	L	-1.898
350	D	-1.412
351	F	-0.4842
352	S	-1.931
353	K	-1.389
354	L	-1.551
355	E	-1.704
356	A	-0.9442
357	G	-1.292
358	K	-0.18
359	L	-0.004974
360	I	0.4477
361	L	-1.17
362	E	-0.5181
363	S	1.474
364	I	0.6493
365	P	0.8459
366	F	-1.302
367	P	-0.6097
368	L	-0.3554
369	R	-0.3785
370	S	1.493
371	T	0.2534
372	L	-0.7522
373	D	0.7005
374	E	-0.3509
375	V	-1.035
376	V	1.008
377	T	0.0611
378	L	0.8932
379	L	-0.06247
380	A	-0.2501
381	H	0.863
382	S	1.894
383	S	-0.6966

#POS	SEQ	SCORE
384	H	1.558
385	D	1.032
386	K	-1.771
387	G	0.8767
388	L	-0.925
389	E	0.3069
390	L	-0.7643
391	T	-0.0959
392	L	1.665
393	N	0.4658
394	I	-0.7536
395	K	0.9055
396	S	1.433
397	D	-0.7043
398	V	-0.9014
399	P	-1.265
400	D	1.794
401	N	1.694
402	V	0.3553
403	I	1.558
404	G	-1.117
405	D	-2.09
406	P	-0.2777
407	L	0.8508
408	R	-1.576
409	L	-1.318
410	Q	-0.1507
411	Q	-2.04
412	I	-1.507
413	I	-0.2142
414	T	-0.7305
415	N	-2.001
416	L	-1.521
417	V	-1.393
418	G	-0.9331
419	N	-2.106
420	A	-1.862
421	I	-0.9221
422	K	-1.762
423	F	-1.276
424	T	-1.934
425	E	0.6398
426	N	-0.05101
427	G	-1.927
428	N	0.7642
429	I	-1.528

#POS	SEQ	SCORE
430	D	0.5182
431	I	-0.4573
432	L	0.4661
433	V	-0.6464
434	E	0.7522
435	K	1.485
436	R	1.367
437	A	1.027
438	L	1.191
439	S	1.382
440	N	1.462
441	T	1.581
442	K	1.584
443	V	0.1516
444	Q	1.552
445	I	-0.4488
446	E	1.351
447	V	-0.8144
448	Q	0.8815
449	I	-1.63
450	R	0.2091
451	D	-2.09
452	T	-1.59
453	G	-1.93
454	I	0.02951
455	G	-1.926
456	I	-1.79
457	P	-1.022
458	E	1.435
459	R	0.2815
460	D	0.8497
461	Q	-0.4413
462	S	1.532
463	R	0.5254
464	L	-0.7856
465	F	-1.919
466	Q	-0.8884
467	A	-1.328
468	F	-1.641
469	R	-0.03194
470	Q	-1.588
471	A	-0.9414
472	D	-1.249
473	A	1.425
474	S	-0.5284
475	I	-0.4568
476	S	-1.053
477	R	-1.57
478	R	-0.2319
479	H	0.3183
480	G	-0.7953
481	G	-2.035
482	T	-1.862
483	G	-1.931
484	L	-1.948
485	G	-2.035
486	L	-1.937
487	V	-1.017
488	I	-2.069
489	T	-0.3987
490	Q	-0.9579
491	K	0.02523
492	L	-1.416
493	V	-1.363
494	N	-0.4844
495	E	-0.1833
496	M	-1.897
497	G	0.065
498	G	-2.035
499	D	0.1132
500	I	-1.479
501	S	0.7615
502	F	-1.26
503	H	1.172
504	S	-1.825
505	Q	1.534
506	P	0.623
507	N	-0.629
508	R	0.02149
509	G	-1.921
510	S	-1.483
511	T	-1.095
512	F	-1.611
513	W	1.214
514	F	0.1035
515	H	0.5391
516	I	-0.5913
517	N	-0.7342
518	L	-0.7356
519	D	1.488
520	L	0.3084
521	N	0.7671
522	P	1.096
523	N	1.119
524	I	1.571
525	I	1.977
526	I	1.865
527	E	1.649
528	G	1.091
529	P	1.165
530	S	1.93
531	T	1.732
532	Q	0.9287
533	C	0.1873
534	L	-0.6605
535	A	0.9763
536	G	-0.7703
537	K	1.255
538	R	-0.0875
539	L	0.308
540	A	0.6892
541	Y	-0.1127
542	V	-0.5671
543	E	0.3578
544	P	1.193
545	N	-0.2972
546	S	1.654
547	A	1.262
548	A	-0.689
549	A	0.1691
550	Q	0.4329
551	C	0.2257
552	T	0.9625
553	L	1.071
554	D	0.6737
555	I	1.361
556	L	0.3728
557	S	1.394
558	E	0.1224
559	T	0.743
560	P	-0.2612
561	L	-1.087
562	E	1.296
563	V	-0.001025
564	V	0.3886
565	Y	0.09813
566	S	0.5597
567	P	0.924

#POS	SEQ	SCORE
568	T	0.5962
569	F	0.8577
570	S	0.5922
571	A	-0.3093
572	L	0.02107
573	P	0.7908
574	P	0.858
575	A	0.8707
576	H	0.7305
577	Y	1.626
578	D	1.459
579	M	-0.1091
580	M	0.203
581	L	-0.2288
582	L	0.18
583	G	1.399
584	I	1.279
585	A	1.561
586	V	0.6739
587	T	1.687
588	F	0.5683
589	R	0.6902
590	E	0.6955
591	P	1.324
592	L	1.717
593	T	0.1594
594	M	1.334
595	Q	1.736
596	H	1.414
597	E	1.604
598	R	0.7663
599	L	1.815
600	A	0.952
601	K	0.9028
602	A	1.567
603	V	1.669
604	S	0.2854
605	M	1.458
606	T	1.546
607	D	1.539
608	F	1.874
609	L	1.407
610	M	1.081
611	L	1.616
612	A	0.2233
613	L	1.155
614	P	1.581
615	C	1.277
616	H	-0.7812
617	A	-0.8006
618	Q	1.509
619	V	-0.3405
620	N	0.811
621	A	0.9075
622	E	-0.1393
623	K	0.9508
624	L	1.099
625	K	1.793
626	Q	0.8485
627	D	0.3626
628	G	1.397
629	I	1.258
630	G	1.704
631	A	1.509
632	C	1.049
633	L	0.7844
634	L	1.135
635	K	-0.2236
636	P	0.2724
637	L	0.7664
638	T	0.4656
639	P	0.03893
640	T	-0.6485
641	R	1.807
642	L	1.188
643	L	2.03
644	P	1.675
645	A	1.334
646	L	-0.5477
647	T	0.6437
648	E	1.415
649	F	0.683
650	C	1.687
651	H	1.872
652	H	1.795
653	K	0.3236
654	Q	0.2625
655	N	0.8139
656	T	0.4496
657	L	1.139
658	L	0.02686
659	P	-0.1406
660	V	0.6184
661	T	1.375
662	D	1.605
663	E	1.082
664	S	1.445
665	K	0.9509
666	L	1.466
667	A	0.8442
668	M	0.3931
669	T	0.944
670	V	-1.012
671	M	-1.674
672	A	-0.522
673	V	-1.279
674	D	-1.55
675	D	-1.75
676	N	-1.946
677	P	0.8653
678	A	-0.5687
679	N	-1.355
680	L	-0.8683
681	K	0.4261
682	L	-0.7348
683	I	-0.3339
684	G	1.045
685	A	0.6851
686	L	-0.4016
687	L	-1.475
688	E	0.5799
689	D	-0.565
690	M	0.4256
691	V	-1.346
692	Q	0.5974
693	H	1.177
694	V	-1.783
695	E	1.227
696	L	1.07
697	C	-0.965
698	D	0.6697
699	S	-0.9608
700	G	-1.226
701	H	1.766
702	Q	-0.8154
703	A	-1.144
704	V	-0.5768
705	E	1.164

#POS	SEQ	SCORE
706	R	1.117
707	A	0.6977
708	K	-0.42
709	Q	0.465
710	M	0.2773
711	P	1.482
712	F	-0.8631
713	D	-1.599
714	L	-1.092
715	I	-0.9759
716	L	-1.296
717	M	-1.422
718	D	-2.066
719	I	0.256
720	Q	-0.619
721	M	-1.716
722	P	-1.894
723	D	0.9944
724	M	-0.532
725	D	-1.419
726	G	-1.855
727	I	1.483
728	R	-1.081
729	A	-0.6394
730	C	-1.498
731	E	-0.271
732	L	0.2201
733	I	-1.682
734	H	-1.608
735	Q	0.3499
736	L	1.018
737	P	-0.3631
738	H	0.2284
739	Q	-0.01054
740	Q	0.1971
741	Q	0.553
742	T	-0.1691
743	P	-1.78
744	V	-1.083
745	I	-0.8696
746	A	-1.586
747	V	-0.5004
748	T	-1.817
749	A	-1.759
750	H	-0.8765
751	A	-0.4121
752	M	0.6674
753	A	0.08814
754	G	0.3949
755	Q	-0.3555
756	K	0.3718
757	E	0.5384
758	K	0.5267
759	L	-0.239
760	L	0.04956
761	G	0.2959
762	A	-0.5276
763	G	-1.753
764	M	-1.294
765	S	-0.8244
766	D	0.4071
767	Y	-0.2131
768	L	-0.6144
769	A	-1.072
770	K	-2.033
771	P	-1.869
772	I	-0.8106
773	E	-0.09083
774	E	1.511
775	E	-0.2359
776	R	0.9559
777	L	-1.654
778	H	0.9845
779	N	0.8386
780	L	0.5024
781	L	-0.7061
782	L	0.9373
783	R	1.215
784	Y	0.3181
785	K	0.7862
786	P	0.1867
787	G	1.283
788	S	0.5322
789	G	1.133
790	I	1.036
791	S	-0.4494
792	S	0.609
793	R	0.9964
794	V	-0.4653
795	V	0.4532
796	T	0.3086
797	P	-0.2847
798	E	0.4346
799	V	0.1644
800	N	-0.4612
801	E	0.3533
802	I	0.2874
803	V	-0.9329
804	V	0.3406
805	N	0.2155
806	P	0.1989
807	N	0.2845
808	A	-0.4667
809	T	-0.1247
810	L	-0.5437
811	D	-1.156
812	W	0.2681
813	Q	-0.4496
814	L	0.4094
815	A	-0.1878
816	L	-1.453
817	R	-1.067
818	Q	0.1236
819	A	-0.7272
820	A	-1.145
821	G	-1.471
822	K	-0.9528
823	T	-0.1681
824	D	-1.146
825	L	-0.4448
826	A	-0.9674
827	R	-0.7779
828	D	-1.156
829	M	-1.167
830	L	-1.453
831	Q	0.02959
832	M	-0.8751
833	L	-0.5437
834	L	-0.4109
835	D	0.136
836	F	-0.9191
837	L	-0.4448
838	P	-0.4477
839	E	-0.8277
840	V	0.4119
841	R	-0.4462
842	N	0.3072
843	K	-0.04016

#POS	SEQ	SCORE
844	V	-0.5553
845	E	0.2258
846	E	0.3037
847	Q	-0.1868
848	L	-0.1936
849	V	-0.616
850	G	-0.2909
851	E	-0.585
852	N	-0.6918
853	P	-0.04367
854	E	-0.0959
855	G	-0.4784
856	L	-1.184
857	V	-0.7682
858	D	-0.6862
859	L	-0.1936
860	I	-1.321
861	H	-1.382
862	K	-0.6686
863	L	-1.184
864	H	-1.382
865	G	-1.203
866	S	-0.6571
867	C	-0.1879
868	G	-0.4299
869	Y	-1.188
870	S	-0.5373
871	G	-1.203
872	V	-1.306
873	P	-1.243
874	R	-0.556
875	M	-0.5734
876	K	-0.6686
877	N	-0.7348
878	L	-0.09904
879	C	-1.196
880	Q	-1.337
881	L	-0.1936
882	I	-0.2245
883	E	-1.26
884	Q	-0.2087
885	Q	-0.3518
886	L	-1.184
887	R	-0.6686
888	S	-0.2149
889	G	-0.8404

#POS	SEQ	SCORE
890	T	-0.8404
891	K	-0.8404
892	E	-0.8404
893	E	-0.8404
894	D	-0.8404
895	L	-0.8404
896	E	-0.8404
897	P	-0.8404
898	E	-0.8404
899	L	-0.8404
900	L	-0.8404
901	E	-0.8404
902	L	-0.8404
903	L	-0.8404
904	D	-0.8404
905	E	-0.8404
906	M	-0.8404
907	D	-0.8404
908	N	-0.8404
909	V	-0.8404
910	A	-0.8404
911	R	-0.8404
912	E	-0.8404
913	A	-0.8404
914	S	-0.8404
915	K	-0.8404
916	I	-0.8404
917	L	-0.8404
918	G	-0.8404

7.1.3 Biofilm

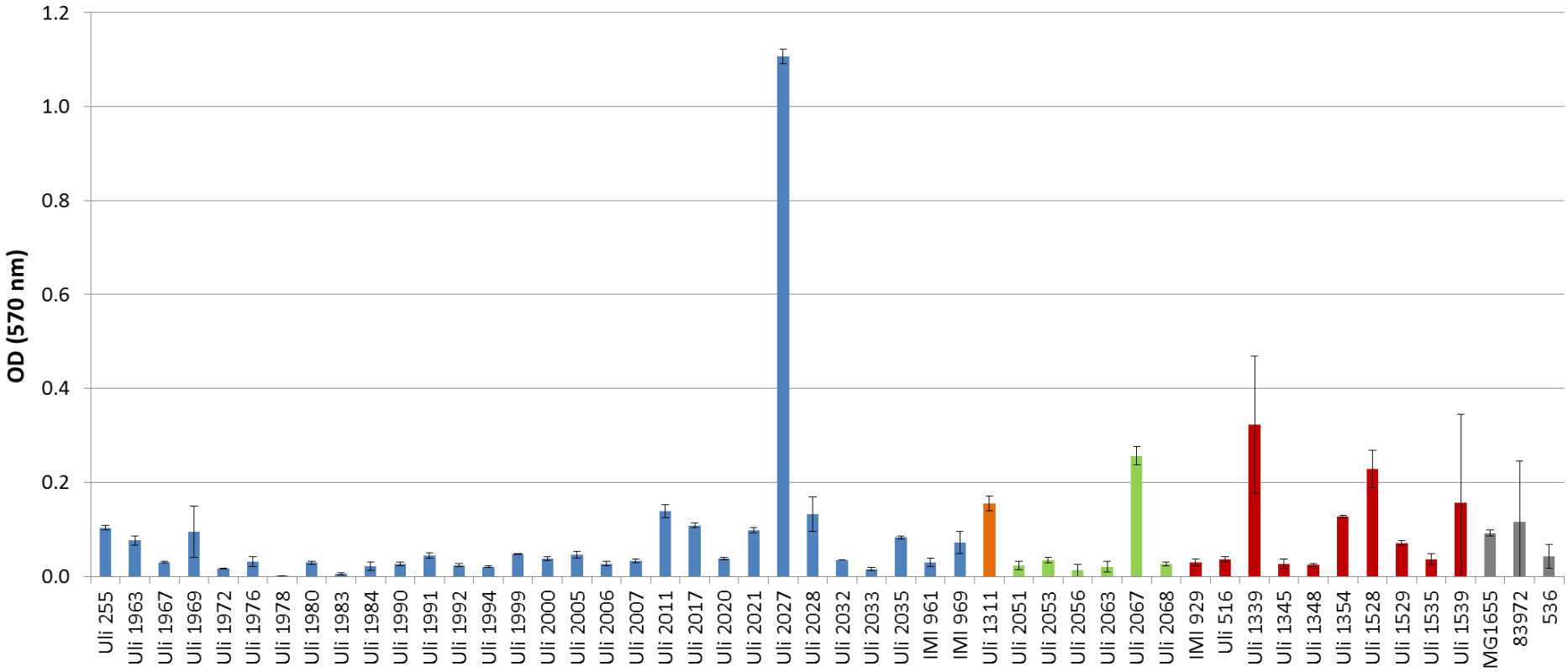


Figure 51: Biofilm formation of all isolates tested upon growth at 37 °C in LB.

Blue bars indicate ABU isolates, HA-ABU is orange. Fecal isolates are colored green and UTI isolates are marked in red. The grey bars are *E. coli* MG1655, ABU 8972 and UPEC strain 536, respectively. Error bars represent the standard deviation of results from at least three independent experiments.

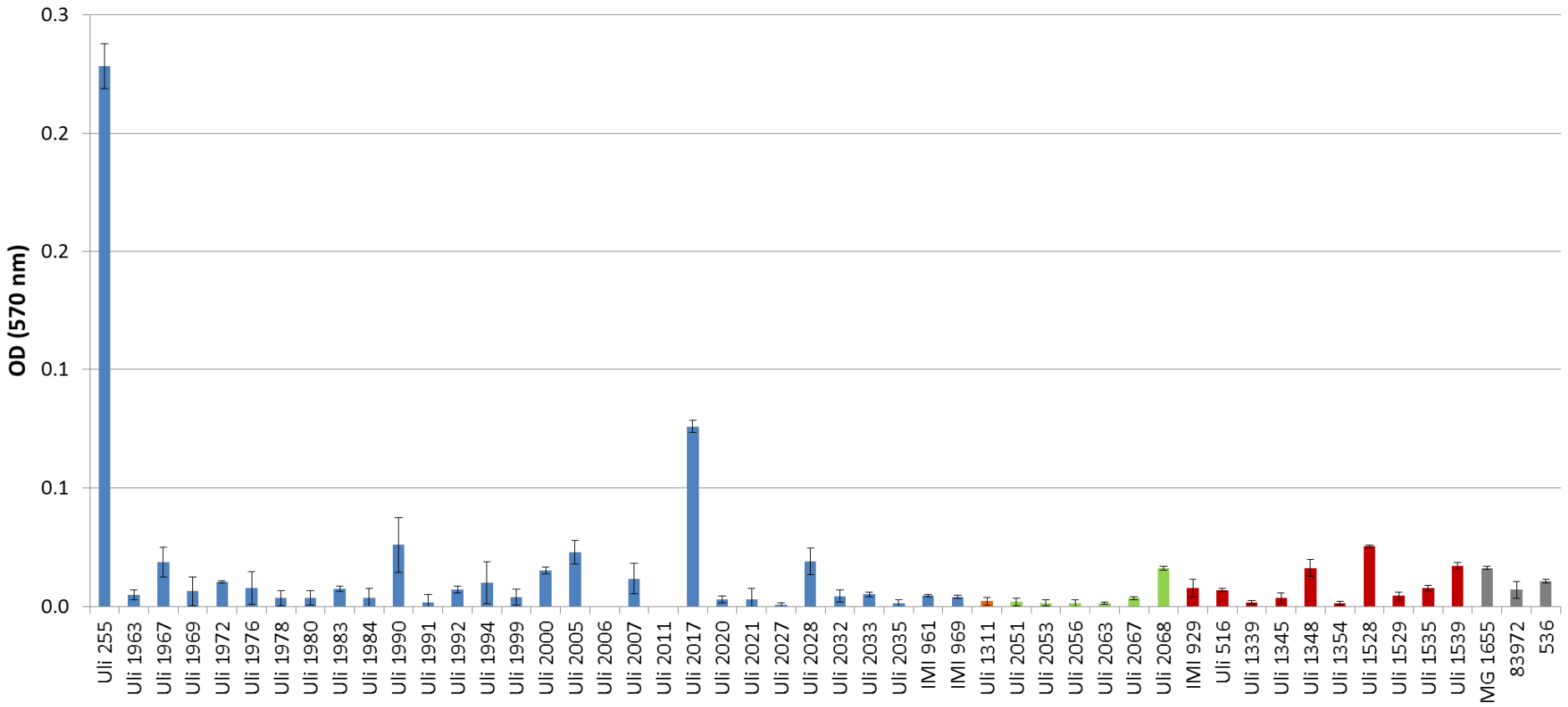


Figure 52: Biofilm formation of all isolates tested upon growth at 37 °C in pooled human urine.

Blue bars indicate ABU isolates, HA-ABU is orange. Fecal isolates are colored green and UTI isolates are marked in red. The grey bars are *E. coli* MG1655, ABU 8972 and UPEC strain 536, respectively. Error bars represent the standard deviation of results from at least three independent experiments.

7.1.4 Summary of the phenotypic assays

Table 27: Summary of the results for the phenotypic assays for the tested strain panel (see 4.7).

Motility is indicated by the following signs: (-): non-motile, (+): medium motile and (++): highly motile. The morphotypes for the strains on the Congo Red (CR) plates are depicted as *rdar* (red, dry and rough), brown, pink and *saw* (smooth and white). The results for the cellulose expression are shown in the column for the Calcoflour (CF) assay (plus or minus). The yellow marked positions indicate that the results are not in accordance with the ones from the CF agar. Signs in brackets depict an inconclusive outcome.

Group	*E. coli* strain	Motility		CR morphotypes			CF phenotypes		
		LB	Urine	RT	30 °C	37 °C	RT	30 °C	37 °C
	# 255	++	+	*rdar*	*rdar*	brown	+	+	-
	# 1963	++	+	*saw*	*saw*	brown	-	-	-
	# 1967	-	-	pink/*saw*	pink	brown	-	-	-
	# 1969	-	-	*saw*	*saw*	pink	-	-	(+)
	# 1972	+	+	*rdar*	pink	brown/ *rdar*	+	-	-
	# 1976	-	+	*rdar*	*rdar*	*rdar*	+	+	+
	# 1978	-	-	brown	pink	pink	-	+	+
	# 1980	++	+	*rdar*	pink	pink/brown	+	+	+
	# 1983	++	+	*saw*	*saw*	*saw*	-	-	-
	# 1984	++	+	pink	brown	brown	-	-	-
	# 1990	-	-	*saw*	*saw*	pink	-	-	+
	# 1991	-	-	*saw*	pink	brown	-	(+)	(+)
ABU	# 1992	++	+	*rdar*	*rdar*	*rdar*/brown	+	+	+
	# 1994	+	+	pink	*rdar*	brown	+	+	-
	# 1999	++	+	brown	brown	brown	-	-	-
	# 2000	++	++	*saw*	*saw*	brown/pink	-	-	(+)
	# 2005	+	+	*rdar*	*rdar*	*rdar*	+	+	+
	# 2006	++	++	*saw*/pink	*saw*	*rdar*	-	-	+
	# 2007	++	++	*rdar*	*rdar*	*saw*	+	+	-
	# 2011	++	+	*rdar*	*rdar*	brown	+	+	-
	# 2017	++	++	brown	brown	brown	-	-	-
	# 2020	++	+	brown	brown	brown	-	-	-
	# 2021	++	++	*saw*	*saw*/pink	pink	-	-	+
	# 2027	++	+	brown	brown	pink	-	-	+
	# 2028	+	+	pink	brown	brown	+	(+)	-

Group	*E. coli* strain	Motility		CR morphotypes			CF phenotypes		
		LB	Urine	RT	30 °C	37 °C	RT	30 °C	37 °C
	# 2032	-	-	*saw*	*saw*	brown	-	-	-
	# 2033	-	-	brown	brown	brown	-	-	-
	# 2035	+	+	*saw*	pink	pink	-	+	+
	IMI 961	-	-	brown	brown	brown	-	(+)	(+)
	IMI 969	++	+	*rdar/saw*	*rdar*	*rdar*	+	+	+
HA-ABU	# 1311	++	+	*saw*	*saw*	pink	-	-	+
FI	# 2051	-	-	*rdar*	*rdar*	pink/ *saw*	+	+	+
	# 2053	-	-	*rdar*	*rdar*	*rdar*	+	+	+
	# 2056	++	+	brown	brown	brown	-	-	-
	# 2063	++	++	pink	brown	pink	(-)	+	(+)
	# 2067	++	++	*saw*	pink	pink/brown	-	+	(+)
	# 2068	++	+	*saw*	brown	pink/brown	-	(+)	+
UTI	IMI 929	-	-	*rdar*	brown	pink	-	(+)	+
	# 516	++	+	*rdar*	*rdar*	pink	+	+	+
	# 1339	++	++	*rdar*	*rdar*	*rdar*	+	+	+
	# 1345	-	-	pink/brown	pink	pink	-	+	+
	# 1348	++	+	*saw*	brown	pink	-	-	+
	# 1354	-	-	*rdar*	*rdar*	*rdar*	+	+	+
	# 1528	++	+	pink	pink	pink	+	+	+
	# 1529	+	-	brown	brown	*saw*/brown	-	-	-
	# 1535	++	++	*saw*	*saw*	pink	-	-	-
	# 1539	-	+	*rdar*	*rdar*	pink	+	+	+
K-12	MG1655	+	+	brown	brown	pink/brown	-	-	-
	*Ec*N (Nissle 1917)	++	+	*rdar*	*rdar*	*rdar*	+	+	+
ABU	83972	+	+	pink	pink	pink/brown	-	+	+
K-12 modified	MG1655 WT^{Zeo}	+	+	brown	brown	pink/brown	-	-	-
***uvrY* allelic variants**	MG1655Δ*uvrY*	+	+	brown	brown	pink/brown	-	-	(+)
	MG1655Δ*uvrY*:*uvrY* 1972	+	+	brown	brown	pink/brown	-	-	(+)
	MG1655Δ*uvrY*:*uvrY* 1976	+	+	brown	brown	pink/brown	-	-	(+)
	MG1655Δ*uvrY*:*uvrY* 1992	+	+	brown	brown	pink/brown	-	-	(+)
	MG1655Δ*uvrY*:*uvrY* 2000	+	+	brown	brown	pink/brown	-	-	(+)
	MG1655Δ*uvrY*:*uvrY* 83972	+	-	brown	brown	pink/brown	-	-	(+)

Group	*E. coli* strain	Motility		CR morphotypes			CF phenotypes		
		LB	Urine	RT	30 °C	37 °C	RT	30 °C	37 °C
	MG1655Δ*uvrY*:*uvrY* MG1655	+	+	brown	brown	pink/brown	-	-	(+)
	MG1655Δ*uvrY*:*uvrY* 1980	+	-	brown	brown	brown	-	-	-
	MG1655Δ*uvrY*:*uvrY* 2031	+	-	brown	brown	brown	-	-	-

7.2 References

Agace, W. W., Hedges, S. R., Ceska, M., & Svanborg, C. (1993). Interleukin-8 and the neutrophil response to mucosal gram-negative infection. *Journal of Clinical Investigation*, *92*(2), 780–785.

Ali, A. S. M., Townes, C. L., Hall, J., & Pickard, R. S. (2009). Maintaining a sterile urinary tract: the role of antimicrobial peptides. *The Journal of Urology*, *182*(1), 21–8.

Alteri, C. J., Smith, S. N., & Mobley, H. L. T. (2009). Fitness of Escherichia coli during urinary tract infection requires gluconeogenesis and the TCA cycle. *PLoS Pathogens*, *5*(5), e1000448.

Altier, C., Suyemoto, M., Ruiz, a I., Burnham, K. D., & Maurer, R. (2000). Characterization of two novel regulatory genes affecting Salmonella invasion gene expression. *Molecular Microbiology*, *35*(3), 635–46.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–10.

Anfora, A. T., Halladin, D. K., Haugen, B. J., & Welch, R. a. (2008). Uropathogenic Escherichia coli CFT073 is adapted to acetatogenic growth but does not require acetate during murine urinary tract infection. *Infection and Immunity*, *76*(12), 5760–7.

Aubron, C., Glodt, J., Matar, C., Huet, O., Borderie, D., Dobrindt, U., … Bouvet, O. (2012). Variation in endogenous oxidative stress in Escherichia coli natural isolates during growth in urine. *BMC Microbiology*, *12*(1), 120.

Babior, B. M. (2000). Phagocytes and oxidative stress. *The American Journal of Medicine*, *109*(1), 33–44.

Babitzke, P., & Romeo, T. (2007). CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current Opinion in Microbiology*, *10*(2), 156–63.

Bailey, J. K., Pinyon, J. L., Anantham, S., & Hall, R. M. (2010). Distribution of Human Commensal Escherichia coli Phylogenetic Groups. *Journal of Clinical Microbiology*, *48*(9), 3455–3456.

Baker, C. S., Morozov, I., Suzuki, K., Romeo, T., & Babitzke, P. (2002). CsrA regulates glycogen biosynthesis by preventing translation of glgC in Escherichia coli. *Molecular Microbiology*, *44*(6), 1599–610.

Barnett, B. J., & Stephens, D. S. (1997). Urinary tract infection: an overview. *The American Journal of the Medical Sciences*, *314*(4), 245–9.

Battesti, A., Majdalani, N., & Gottesman, S. (2011). The RpoS-mediated general stress response in Escherichia coli. *Annual Review of Microbiology*, *65*, 189–213.

Becker, G., Klauck, E., & Hengge-Aronis, R. (1999). Regulation of RpoS proteolysis in Escherichia coli: the response regulator RssB is a recognition factor that interacts with the turnover element in RpoS. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(11), 6439–44.

Beloin, C., Roux, A., & Ghigo, J. M. (2008). Escherichia coli biofilms. *Current Topics in Microbiology and Immunology*, *322*, 249–89.

Bennett, P. M. (2004). Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods in Molecular Biology (Clifton, N.J.)*, *266*, 71–113.

Bien, J., Sokolova, O., & Bozko, P. (2012). Role of Uropathogenic Escherichia coli Virulence Factors in Development of Urinary Tract Infection and Kidney Damage. *International Journal of Nephrology*, *2012*, 681473.

Birnboim, H. C., & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, *7*(6), 1513–23.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., … Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science (New York, N.Y.)*, *277*(5331), 1453–62.

Bokranz, W., Wang, X., Tschäpe, H., & Römling, U. (2005). Expression of cellulose and curli fimbriae by Escherichia coli isolated from the gastrointestinal tract. *Journal of Medical Microbiology*, *54*(Pt 12), 1171–82.

Bower, J. M., & Mulvey, M. A. (2006). Polyamine-mediated resistance of uropathogenic Escherichia coli to nitrosative stress. *Journal of Bacteriology*, *188*(3), 928–33.

Boyd, E. F., & Hartl, D. L. (1998). Chromosomal regions specific to pathogenic isolates of Escherichia coli have a phylogenetically clustered distribution. *Journal of Bacteriology*, *180*(5), 1159–65.

Brooks, T., & Keevil, C. W. (1997). A simple artificial urine for the growth of urinary pathogens. *Letters in Applied Microbiology*, *24*(3), 203–6.

Brown, D. R., Barton, G., Pan, Z., Buck, M., & Wigneshweraraj, S. (2014). Nitrogen stress response and stringent response are coupled in Escherichia coli. *Nature Communications*, *5*, 4115.

Brown, T. (2001). Dot and slot blotting of DNA. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et al.]*, *Chapter 2*, Unit2.9B.

Brückner, R., & Titgemeyer, F. (2002). Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiology Letters*, *209*(2), 141–8.

Burnette, W. N. (1981). “Western blotting”: electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Analytical Biochemistry*, *112*(2), 195–203.

Busby, S. J. W., Thomas, C. M., & Brown, N. L. (Eds.). (1998). *Molecular Microbiology*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Chandrangsu, P., & Helmann, J. D. (2014). Sigma Factors in Gene Expression.

Charlesworth, J., & Eyre-Walker, A. (2006). The rate of adaptive evolution in enteric bacteria. *Molecular Biology and Evolution*, *23*(7), 1348–56.

Chaudhuri, R. R., Loman, N. J., Snyder, L. A. S., Bailey, C. M., Stekel, D. J., & Pallen, M. J. (2008). xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Research*, *36*(Database issue), D543–6.

Chavez, R. G., Alvarez, A. F., Romeo, T., & Georgellis, D. (2010). The physiological stimulus for the BarA sensor kinase. *Journal of Bacteriology*, *192*(7), 2009–12.

Chen, S. L., Hung, C.-S., Xu, J., Reigstad, C. S., Magrini, V., Sabo, A., … Gordon, J. I. (2006). Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(15), 5977–82.

Chubukov, V., Gerosa, L., Kochanowski, K., & Sauer, U. (2014). Coordination of microbial metabolism. *Nature Reviews. Microbiology*, *12*(5), 327–40.

Chung, H. J., Bang, W., & Drake, M. A. (2006). Stress Response of Escherichia coli. *Comprehensive Reviews in Food Science and Food Safety*, *5*(3), 52–64.

Clark, D. P. (1989). The fermentation pathways of Escherichia coli. *FEMS Microbiology Reviews*, *5*(3), 223–34.

Clermont, O., Bonacorsi, S., & Bingen, E. (2000). Rapid and simple determination of the Escherichia coli phylogenetic group. *Applied and Environmental Microbiology*, *66*(10), 4555–8.

Clermont, O., Christenson, J. K., Denamur, E., & Gordon, D. M. (2013). The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports*, *5*(1), 58–65.

Clermont, O., Olier, M., Hoede, C., Diancourt, L., Brisse, S., Keroudean, M., … Denamur, E. (2011). Animal and human pathogenic Escherichia coli strains share common genetic backgrounds. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *11*(3), 654–62.

Corbell, N., & Loper, J. E. (1995). A global regulator of secondary metabolite production in Pseudomonas fluorescens Pf-5. *Journal of Bacteriology*, *177*(21), 6230–6.

Corker, H., & Poole, R. K. (2003). Nitric oxide formation by Escherichia coli. Dependence on nitrite reductase, the NO-sensing regulator Fnr, and flavohemoglobin Hmp. *The Journal of Biological Chemistry*, *278*(34), 31584–92.

Costerton, J. W., Cheng, K. J., Geesey, G. G., Ladd, T. I., Nickel, J. C., Dasgupta, M., & Marrie, T. J. (1987). Bacterial biofilms in nature and disease. *Annual Review of Microbiology*, *41*, 435–64.

Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R., & Lappin-Scott, H. M. (1995). Microbial biofilms. *Annual Review of Microbiology*, *49*, 711–45.

Costerton, J. W., Stewart, P. S., & Greenberg, E. P. (1999). Bacterial biofilms: a common cause of persistent infections. *Science (New York, N.Y.)*, *284*(5418), 1318–22.

Croxen, M. A., & Finlay, B. B. (2010). Molecular mechanisms of Escherichia coli pathogenicity. *Nature Reviews. Microbiology*, *8*(1), 26–38.

Danese, P. N., Pratt, L. A., & Kolter, R. (2000). Exopolysaccharide production is required for development of Escherichia coli K-12 biofilm architecture. *Journal of Bacteriology*, *182*(12), 3593–6.

Datsenko, K. a, & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(12), 6640–5.

Dho-Moulin, M., & Fairbrother, J. M. (1999). Avian pathogenic Escherichia coli (APEC). *Veterinary Research*, *30*(2-3), 299–316.

Dobrindt, U., Agerer, F., Michaelis, K., Janka, A., Buchrieser, C., Samuelson, M., … Karch, H. (2003). Analysis of Genome Plasticity in Pathogenic and Commensal Escherichia coli Isolates by Use of DNA Arrays, *185*(6), 1831–1840.

Dobrindt, U., Blum-Oehler, G., Nagy, G., Schneider, G., Johann, A., Gottschalk, G., & Hacker, J. (2002). Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic Escherichia coli strain 536. *Infection and Immunity*, *70*(11), 6365–72.

Dobrindt, U., Hochhut, B., Hentschel, & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews. Microbiology*, *2*(5), 414–24.

Dobrindt, U., Zdziarski, J., Salvador, E., & Hacker, J. (2010). Bacterial genome plasticity and its impact on adaptation during persistent infection. *International Journal of Medical Microbiology : IJMM*, *300*(6), 363–6.

Dong, T., Kirchhof, M. G., & Schellhorn, H. E. (2008). RpoS regulation of gene expression during exponential growth of Escherichia coli K12. *Molecular Genetics and Genomics : MGG*, *279*(3), 267–77.

Donlan, R. M. (2002). Biofilms: microbial life on surfaces. *Emerging Infectious Diseases*, *8*(9), 881–90.

Donnenberg, M. S. (2002). *Escherichia coli: Virulence Mechanisms of a Versatile Pathogen*. Elsevier.

Donnenberg, M. S., & Kaper, J. B. (1991). Construction of an eae deletion mutant of enteropathogenic Escherichia coli by using a positive-selection suicide vector. *Infection and Immunity*, *59*(12), 4310–7.

Dussert, C., Rasigni, M., Palmari, J., Rasigni, G., Llebaria, A., & Marty, F. (1987). Minimal spanning tree analysis of biological structures. *Journal of Theoretical Biology*, *125*(3), 317–323.

Dyda, F., Chandler, M., & Hickman, A. B. (2012). The emerging diversity of transpososome architectures. *Quarterly Reviews of Biophysics*, *45*(04), 493–521.

Egea, R., Casillas, S., & Barbadilla, A. (2008). Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, *36*(Web Server issue), W157–62.

Ellis, H. M., Yu, D., DiTizio, T., & Court, D. L. (2001). High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(12), 6742–6.

Emody, L., Kerényi, M., & Nagy, G. (2003). Virulence factors of uropathogenic Escherichia coli. *International Journal of Antimicrobial Agents*, *22 Suppl 2*, 29–33.

Entner, N., & Doudoroff, M. (1952). Glucose and gluconic acid oxidation of Pseudomonas saccharophila. *The Journal of Biological Chemistry*, *196*(2), 853–62.

Eriksson, A. R., Andersson, R. A., Pirhonen, M., & Palva, E. T. (1998). Two-component regulators involved in the global control of virulence in Erwinia carotovora subsp. carotovora. *Molecular Plant-Microbe Interactions : MPMI*, *11*(8), 743–52.

Escherich, T. (1886). *Die Darmbakterien des Säuglings und ihre Beziehungen zur Physiologie der Verdauung, von Dr Theodor Escherich,...*

Ferrières, L., Hancock, V., & Klemm, P. (2007). Biofilm exclusion of uropathogenic bacteria by selected asymptomatic bacteriuria Escherichia coli strains. *Microbiology (Reading, England)*, *153*(Pt 6), 1711–9.

Fleige, S., & Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine*, *27*(2-3), 126–39.

Foxman, B. (2002). Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Disease-a-Month : DM*, *49*(2), 53–70.

Foxman, B. (2010). The epidemiology of urinary tract infection. *Nature Reviews. Urology*, *7*(12), 653–60.

Foxman, B., & Frerichs, R. R. (1985). Epidemiology of urinary tract infection: II. Diet, clothing, and urination habits. *American Journal of Public Health*, *75*(11), 1314–7.

Fuhrer, T., Fischer, E., & Sauer, U. (2005). Experimental identification and quantification of glucose metabolism in seven bacterial species. *Journal of Bacteriology*, *187*(5), 1581–90.

Gao, R., & Stock, A. M. (2009). Biological insights from structures of two-component proteins. *Annual Review of Microbiology*, *63*, 133–54.

Garrity, G., Boone, D. R., & Castenholz, R. W. (2001). *Bergey's manual of determinative bacteriology*.

Gasser, T. (2011). *Basiswissen Urologie*. Springer-Verlag.

Gay, P., Le Coq, D., Steinmetz, M., Ferrari, E., & Hoch, J. A. (1983). Cloning structural gene sacB, which codes for exoenzyme levansucrase of Bacillus subtilis: expression of the gene in Escherichia coli. *Journal of Bacteriology*, *153*(3), 1424–31.

Goodier, R. I., & Ahmer, B. M. M. (2001). SirA Orthologs Affect both Motility and Virulence. *Journal of Bacteriology*, *183*(7), 2249–2258.

Gordon, D. M., & Riley, M. a. (1992). A theoretical and experimental analysis of bacterial growth in the bladder. *Molecular Microbiology*, *6*(4), 555–62.

Gottschalk, G. (1986). *Bacterial Metabolism*. New York, NY: Springer New York.

Grebe, T. W., & Stock, J. B. (1999). *The histidine protein kinase superfamily. Advances in microbial physiology* (Vol. 41).

Groat, R. G., & Matin, A. (1986). Synthesis of unique proteins at the onset of carbon starvation in Escherichia coli. *Journal of Industrial Microbiology*, *1*(2), 69–73.

Groat, R. G., Schultz, J. E., Zychlinsky, E., Bockman, A., & Matin, A. (1986). Starvation proteins in Escherichia coli: kinetics of synthesis and role in starvation survival. *Journal of Bacteriology*, *168*(2), 486–93.

Gudapaty, S., Suzuki, K., Wang, X. I. N., & Babitzke, P. (2001). Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli, *183*(20), 6017–6027.

Hacker, J., & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annual Review of Microbiology*, *54*, 641–79.

Hahn, H., Falke, D., Kaufmann, S. H. E., & Ullmann, U. (2013). *Medizinische Mikrobiologie und Infektiologie* (Vol. 21). Springer Berlin Heidelberg.

Hammer, B. K., Tateda, E. S., & Swanson, M. S. (2002). A two-component regulator induces the transmission phenotype of stationary-phase Legionella pneumophila. *Molecular Microbiology*, *44*(1), 107–18.

Hancock, V., Ferrières, L., & Klemm, P. (2007). Biofilm formation by asymptomatic and virulent urinary tract infectious Escherichia coli strains. *FEMS Microbiology Letters*, *267*(1), 30–7.

Hancock, V., & Klemm, P. (2007). Global gene expression profiling of asymptomatic bacteriuria Escherichia coli during biofilm growth in human urine. *Infection and Immunity*, *75*(2), 966–76.

Hancock, V., Seshasayee, A. S., Ussery, D. W., Luscombe, N. M., & Klemm, P. (2008). Transcriptomics and adaptive genomics of the asymptomatic bacteriuria Escherichia coli strain 83972. *Molecular Genetics and Genomics : MGG*, *279*(5), 523–34.

Hancock, V., Vejborg, R. M., & Klemm, P. (2010). Functional genomics of probiotic Escherichia coli Nissle 1917 and 83972, and UPEC strain CFT073: comparison of transcriptomes, growth and biofilm formation. *Molecular Genetics and Genomics : MGG*, *284*(6), 437–54.

Hautmann, R., & Geschwend, J. E. (2014). *Urologie*. Springer-Verlag.

Heizmann, W. R., Döller, P. C., Kropp, S., & Bleich, S. (1999). *Kurzlehrbuch medizinische Mikrobiologie und Immunologie: zur Vorbereitung auf das 1. Staatsexamen ; mit 61 Tabellen*. Schattauer Verlag.

Helling, R. B. (1998). Pathway choice in glutamate synthesis in Escherichia coli. *Journal of Bacteriology*, *180*(17), 4571–5.

Hengge, R. (2009). Proteolysis of sigmaS (RpoS) and the general stress response in Escherichia coli. *Research in Microbiology*, *160*(9), 667–76.

Hengge-Aronis, R. (1999). Interplay of global regulators and cell physiology in the general stress response of Escherichia coli. *Current Opinion in Microbiology*, *2*(2), 148–152.

Hengge-Aronis, R. (2002). Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiology and Molecular Biology Reviews : MMBR*, *66*(3), 373–95, table of contents.

Heroven, A. K., Böhme, K., & Dersch, P. (2012). The Csr/Rsm system of Yersinia and related pathogens: a post-transcriptional strategy for managing virulence. *RNA Biology*, *9*(4), 379–91.

Herren, C. D., Mitra, A., Palaniyandi, S. K., Coleman, A., Elankumaran, S., & Mukhopadhyay, S. (2006). The BarA-UvrY two-component system regulates virulence in avian pathogenic Escherichia coli O78:K80:H9. *Infection and Immunity*, *74*(8), 4900–9.

Herzer, P. J., Inouye, S., Inouye, M., & Whittam, T. S. (1990). Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of Escherichia coli. *Journal of Bacteriology*, *172*(11), 6175–81.

Hoch, J. A. (2000). Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology*, *3*(2), 165–70.

Hooton, T. (1999). Practice guidelines for urinary tract infection in the era of managed care. *International Journal of Antimicrobial Agents*, *11*(3-4), 241–245.

Hooton, T. M. (2001). Recurrent urinary tract infection in women. *International Journal of Antimicrobial Agents*, *17*(4), 259–68.

Hulko, M., Berndt, F., Gruber, M., Linder, J. U., Truffault, V., Schultz, A., … Coles, M. (2006). The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell*, *126*(5), 929–40.

Hull, R. a, Rudy, D. C., Donovan, W. H., Wieser, I. E., Stewart, C., & Darouiche, R. O. (1999). Virulence properties of Escherichia coli 83972, a prototype strain associated with asymptomatic bacteriuria. *Infection and Immunity*, *67*(1), 429–32.

Hull, R. A., & Hull, S. I. (1997). Nutritional requirements for growth of uropathogenic Escherichia coli in human urine. *Infection and Immunity*, *65*(5), 1960–1.

Hull, R., Rudy, D., Donovan, W., Svanborg, C., Wieser, I., Stewart, C., & Darouiche, R. (2000). Urinary tract infection prophylaxis using Escherichia coli 83972 in spinal cord injured patients. *The Journal of Urology*, *163*(3), 872–7.

Ihssen, J., & Egli, T. (2005). Global physiological analysis of carbon- and energy-limited growing Escherichia coli confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. *Environmental Microbiology*, *7*(10), 1568–81.

Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., … Auffray, C. (2005). Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Research*, *33*(6), e56.

Ishige, K., Nagasawa, S., Tokishita, S., & Mizuno, T. (1994). A novel device of bacterial signal transducers. *The EMBO Journal*, *13*(21), 5195–202.

Jackson, D. W., Suzuki, K., Oakford, L., Simecka, J. W., Hart, M. E., & Romeo, T. (2002). Biofilm Formation and Dispersal under the Influence of the Global Regulator CsrA of Escherichia coli.

Jishage, M., & Ishihama, a. (1995). Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of sigma 70 and sigma 38. *Journal of Bacteriology*, *177*(23), 6832–5.

Johnson, J. R. (1991). Virulence factors in Escherichia coli urinary tract infection. *Clinical Microbiology Reviews*, *4*(1), 80–128.

Johnson, J. R., & Russo, T. A. (2002). Extraintestinal pathogenic Escherichia coli : "The other bad E coli ." *Journal of Laboratory and Clinical Medicine*, *139*(3), 155–162.

Johnson, J. R., & Russo, T. A. (2005). Molecular epidemiology of extraintestinal pathogenic (uropathogenic) Escherichia coli. *International Journal of Medical Microbiology : IJMM*, *295*(6-7), 383–404.

Jonas, K., & Melefors, O. (2009). The Escherichia coli CsrB and CsrC small RNAs are strongly induced during growth in nutrient-poor medium. *FEMS Microbiology Letters*, *297*(1), 80–6.

Jonas, K., Tomenius, H., Römling, U., Georgellis, D., & Melefors, O. (2006). Identification of YhdA as a regulator of the Escherichia coli carbon storage regulation system. *FEMS Microbiology Letters*, *264*(2), 232–7.

Jung, K., Fried, L., Behr, S., & Heermann, R. (2012). Histidine kinases and response regulators in networks. *Current Opinion in Microbiology*, *15*(2), 118–24.

Kamada, N., Chen, G. Y., Inohara, N., & Núñez, G. (2013). Control of pathogens and pathobionts by the gut microbiota. *Nature Immunology*, *14*(7), 685–90.

Kaper, J. B., Nataro, J. P., & Mobley, H. L. (2004). Pathogenic Escherichia coli. *Nature Reviews. Microbiology*, *2*(2), 123–40.

Kass, E. H. (1956). Asymptomatic infections of the urinary tract. *Transactions of the Association of American Physicians*, *69*, 56–64.

Katouli, M. (2010). Population structure of gut Escherichia coli and its role in development of extra-intestinal infections. *Iranian Journal of Microbiology*, *2*(2), 59–72.

Kauffmann, F. (1965). [Supplement to the Kauffmann-White scheme. 8.]. *Acta. Pathol. Microbiol. Scand.* **64**(3): 362–366

Kawakami, K., Saitoh, T., & Ishihama, A. (1979). Biosynthesis of RNA polymerase in Escherichia coli. *MGG Molecular & General Genetics*, *174*(2), 107–116.

Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., … Karp, P. D. (2005). EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Research*, *33*(Database issue), D334–7.

Klemm, P., Hancock, V., & Schembri, M. A. (2007). Mellowing out: adaptation to commensalism by Escherichia coli asymptomatic bacteriuria strain 83972. *Infection and Immunity*, *75*(8), 3688–95.

Klemm, P., Roos, V., Ulett, G. C., Svanborg, C., & Schembri, M. A. (2006). Molecular Characterization of the Escherichia coli Asymptomatic Bacteriuria Strain 83972 : the Taming of a Pathogen, *74*(1), 781–785.

Köhler, C.-D., & Dobrindt, U. (2011). What defines extraintestinal pathogenic Escherichia coli? *International Journal of Medical Microbiology : IJMM*, *301*(8), 642–7.

Kruskal, W. H., & Wallis, W. A. (2012). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. *Science (New York, N.Y.)*, *324*(5924), 255–8.

Kunin, C. M. (1994). Urinary tract infections in females. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *18*(1), 1–10; quiz 11–2.

Kyhse-Andersen, J. (1984). Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *Journal of Biochemical and Biophysical Methods*, *10*(3-4), 203–9.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, *227*(5259), 680–685.

Lange, R., & Hengge-Aronis, R. (1994). The cellular concentration of the sigma S subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. *Genes & Development*, *8*(13), 1600–1612.

Lapouge, K., Schubert, M., Allain, F. H.-T., & Haas, D. (2008). Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. *Molecular Microbiology*, *67*(2), 241–53.

Lee, J. B. L., & Neild, G. H. (2007). Urinary tract infection. *Medicine*, *35*(8), 423–428.

Leimbach, A., Hacker, J., & Dobrindt, U. (2013). E . coli as an All-Rounder : The Thin Line Between Commensalism and Pathogenicity.

Lenz, D. H., Miller, M. B., Zhu, J., Kulkarni, R. V, & Bassler, B. L. (2005). CsrA and three redundant small RNAs regulate quorum sensing in Vibrio cholerae. *Molecular Microbiology*, *58*(4), 1186–202.

Li, X.-T., Thomason, L. C., Sawitzke, J. A., Costantino, N., & Court, D. L. (2013). Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. *Nucleic Acids Research*, *41*(22), e204.

Lindberg, U., Claesson, I., Hanson, L. Å., & Jodal, U. (1978). Asymptomatic bacteriuria in schoolgirls. *The Journal of Pediatrics*, *92*(2), 194–199.

Lindberg, U., Claësson, I., Hanson, L. A., & Jodal, U. (1975). Asymptomatic bacteriuria in schoolgirls. I. Clinical and laboratory findings. *Acta Paediatrica Scandinavica*, *64*(3), 425–31.

Lindberg, U., & Winberg, J. (1976). [Asymptomatic bacteriuria]. *Läkartidningen*, *73*(17), 1617–8.

Liu, M. Y., Gui, G., Wei, B., Preston, J. F., Oakford, L., Yüksel, U., … Romeo, T. (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli. *The Journal of Biological Chemistry*, *272*(28), 17502–10.

Liu, M. Y., Yang, H., & Romeo, T. (1995). The product of the pleiotropic Escherichia coli gene csrA modulates glycogen biosynthesis via effects on mRNA stability. *Journal of Bacteriology*, *177*(10), 2663–72.

Loewen, P. C., & Hengge-Aronis, R. (1994). The role of the sigma factor sigma S (KatF) in bacterial global regulation. *Annual Review of Microbiology*, *48*, 53–80.

Loewen, P. C., Hu, B., Strutinsky, J., & Sparling, R. (1998). Regulation in the rpoS regulon of Escherichia coli. *Canadian Journal of Microbiology*, *44*(8), 707–17.

Long, S. S., & Swenson, R. M. (1977). Development of anaerobic fecal flora in healthy newborn infants. *The Journal of Pediatrics*, *91*(2), 298–301.

López, D., Vlamakis, H., & Kolter, R. (2010). Biofilms. *Cold Spring Harbor Perspectives in Biology*, *2*(7), a000398.

Lundberg, J. O., Ehrén, I., Jansson, O., Adolfsson, J., Lundberg, J. M., Weitzberg, E., … Wiklund, N. P. (1996). Elevated nitric oxide in the urinary bladder in infectious and noninfectious cystitis. *Urology*, *48*(5), 700–2.

Mabbett, A. N., Ulett, G. C., Watts, R. E., Tree, J. J., Totsika, M., Ong, C. Y., … Schembri, M. a. (2009). Virulence properties of asymptomatic bacteriuria Escherichia coli. *International Journal of Medical Microbiology : IJMM*, *299*(1), 53–63.

Mahillon, J., & Chandler, M. (1998). Insertion Sequences. *Microbiol. Mol. Biol. Rev.*, *62*(3), 725–774.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., … Spratt, B. G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(6), 3140–5.

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, *351*(6328), 652–4.

Medini, D., Donati, C., Tettelin, H., Masignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, *15*(6), 589–94.

Messer, P. W., & Petrov, D. A. (2013). Frequent adaptation and the McDonald-Kreitman test. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(21), 8615–20.

Mitra, A., Palaniyandi, S., Herren, C. D., Zhu, X., & Mukhopadhyay, S. (2013). Pleiotropic roles of uvrY on biofilm formation, motility and virulence in uropathogenic Escherichia coli CFT073. *PloS One*, *8*(2), e55492.

Mitrophanov, A. Y., & Groisman, E. a. (2008). Signal integration in bacterial two-component regulatory systems. *Genes & Development*, *22*(19), 2601–11.

Mizuno, T. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of Escherichia coli. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, *4*(2), 161–8.

Mobley, H. L., Green, D. M., Trifillis, A. L., Johnson, D. E., Chippendale, G. R., Lockatell, C. V, … Warren, J. W. (1990). Pyelonephritogenic Escherichia coli and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. *Infection and Immunity*, *58*(5), 1281–9.

Mondragón, V., Franco, B., Jonas, K., Suzuki, K., Romeo, T., Melefors, O., & Georgellis, D. (2006). pH-dependent activation of the BarA-UvrY two-component system in Escherichia coli. *Journal of Bacteriology*, *188*(23), 8303–6.

Moolenaar, G. F., van Sluis, C. A., Backendorf, C., & van de Putte, P. (1987). Regulation of the Escherichia coli excision repair gene uvrC. Overlap between the uvrC structural gene and the region coding for a 24 kD protein. *Nucleic Acids Research*, *15*(10), 4273–89.

Mukhopadhyay, S., Audia, J. P., Roy, R. N., & Schellhorn, H. E. (2000). Transcriptional induction of the conserved alternative sigma factor RpoS in Escherichia coli is dependent on BarA, a probable two-component regulator. *Molecular Microbiology*, *37*(2), 371–81.

Nagasawa, S., Tokishita, S., Aiba, H., & Mizuno, T. (1992). A novel sensor-regulator protein that belongs to the homologous family of signal-transduction proteins involved in adaptive responses in Escherichia coii, *6*(6).

Neidhardt, F.C., Ingraham, J.L., and Schaechter, M. (1990) Physiology of the Bacterial Cell. Sunderland, MA: Sinauer Associates.

Nicolle, L. E., Bradley, S., Colgan, R., Rice, J. C., Schaeffer, A., & Hooton, T. M. (2005). Infectious Diseases Society of America guidelines for the diagnosis and treatment of asymptomatic bacteriuria in adults. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *40*(5), 643–54.

Nissle, A. (1918). Die antagonistische Behandlung chronischer Darmstörungen mit Kolibakterien. *Med. Klin*. **2**: 29–30.

Notley-mcrobb, L., King, T., & Ferenci, T. (2002). rpoS Mutations and Loss of General Stress Resistance in Escherichia coli Populations as a Consequence of Conflict between Competing Stress Responses rpoS Mutations and Loss of General Stress Resistance in Escherichia coli Populations as a Consequence of C.

Ochman, H., & Selander, R. K. (1984). Standard reference strains of Escherichia coli from natural populations. *Journal of Bacteriology*, *157*(2), 690–3.

Oelschlaeger, T., & Fünfstück, R. (2006). Recurrent urinary tract infections in women. Virulence of pathogens and host reaction. *Der Urologe. Ausg. A*, *45*, 412, 414–416, 418–420.

Ofran, Y., Mysore, V., & Rost, B. (2007). Prediction of DNA-binding residues from sequence. *Bioinformatics (Oxford, England)*, *23*(13), i347–53.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *27*(1), 29–34.

Olsén, A., Jonsson, A., & Normark, S. (1989). Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli. *Nature*, *338*(6217), 652–5.

Ooka, T., Ogura, Y., Asadulghani, M., Ohnishi, M., Nakayama, K., Terajima, J., … Hayashi, T. (2009). Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in Escherichia coli O157 genomes. *Genome Research*, *19*(10), 1809–16.

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Molecular Systems Biology*, *7*(1), 535.

Oshima, T., Aiba, H., Masuda, Y., Kanaya, S., Sugiura, M., Wanner, B. L., … Mizuno, T. (2002). Transcriptome analysis of all two-component regulatory system mutants of Escherichia coli K-12. *Molecular Microbiology*, *46*(1), 281–91.

Ottemann, K. M., & Miller, J. F. (1997). Roles for motility in bacterial-host interactions. *Molecular Microbiology*, *24*(6), 1109–17.

Palaniyandi, S., Mitra, A., Herren, C. D., Lockatell, C. V., Johnson, D. E., Zhu, X., & Mukhopadhyay, S. (2012). BarA-UvrY two-component system regulates virulence of uropathogenic E. coli CFT073. *PloS One*, *7*(2), e31348.

Parsek, M. R., & Singh, P. K. (2003). Bacterial biofilms: an emerging link to disease pathogenesis. *Annual Review of Microbiology*, *57*, 677–701.

Patrick, J. E., & Kearns, D. B. (2012). Swarming motility and the control of master regulators of flagellar biosynthesis. *Molecular Microbiology*, *83*(1), 14–23.

Patten, C. L., Kirchhof, M. G., Schertzberg, M. R., Morton, R. A., & Schellhorn, H. E. (2004). Microarray analysis of RpoS-mediated gene expression in Escherichia coli K-12. *Molecular Genetics and Genomics : MGG*, *272*(5), 580–91.

Pernestig, A. K., Georgellis, D., Romeo, T., Suzuki, K., Tomenius, H., & Normark, S. (2003). The Escherichia coli BarA-UvrY Two-Component System Is Needed for Efficient Switching between Glycolytic and Gluconeogenic Carbon Sources, *185*(3), 843–853.

Pernestig, A. K., Melefors, O., & Georgellis, D. (2001). Identification of UvrY as the cognate response regulator for the BarA sensor kinase in Escherichia coli. *The Journal of Biological Chemistry*, *276*(1), 225–31.

Petersen, L., Bollback, J. P., Dimmic, M., Hubisz, M., & Nielsen, R. (2007). Genes under positive selection in Escherichia coli. *Genome Research*, *17*(9), 1336–43.

Prigent-Combaret, C., Prensier, G., Le Thi, T. T., Vidal, O., Lejeune, P., & Dorel, C. (2000). Developmental pathway for biofilm formation in curli-producing Escherichia coli strains: role of flagella, curli and colanic acid. *Environmental Microbiology*, *2*(4), 450–464.

Rand, D. M., & Kann, L. M. (1996). Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. *Molecular Biology and Evolution*, *13*(6), 735–48.

Reisner, A., Krogfelt, K. A., Klein, B. M., Zechner, L., Molin, S., & Zechner, E. L. (2006). In Vitro Biofilm Formation of Commensal and Pathogenic Escherichia coli Strains : Impact of Environmental and

Genetic Factors In Vitro Biofilm Formation of Commensal and Pathogenic Escherichia coli Strains : Impact of Environmental and Genetic Factors.

Robinson, V. L., Buckler, D. R., & Stock, a M. (2000). A tale of two components: a novel kinase and a regulatory switch. *Nature Structural Biology*, *7*(8), 626–33.

Romano, a. H., & Conway, T. (1996). Evolution of carbohydrate metabolic pathways. *Research in Microbiology*, *147*(6-7), 448–455.

Romeo, T. (1998). Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Molecular Microbiology*, *29*(6), 1321–30.

Romeo, T., Gong, M., Liu, M. Y., & Brun-Zinkernagel, a M. (1993). Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *Journal of Bacteriology*, *175*(15), 4744–55.

Römling, U. (2002). Molecular biology of cellulose production in bacteria. *Research in Microbiology*, *153*(4), 205–12.

Ronald, A. (2003). The etiology of urinary tract infection: traditional and emerging pathogens. *Disease-a-Month : DM*, *49*(2), 71–82.

Roos, V., Nielsen, E. M., & Klemm, P. (2006). Asymptomatic bacteriuria Escherichia coli strains: adhesins, growth and competition. *FEMS Microbiology Letters*, *262*(1), 22–30.

Roos, V., Schembri, M. A., Ulett, G. C., & Klemm, P. (2006). Asymptomatic bacteriuria Escherichia coli strain 83972 carries mutations in the foc locus and is unable to express F1C fimbriae. *Microbiology (Reading, England)*, *152*(Pt 6), 1799–806.

Roos, V., Ulett, G. C., Schembri, M. A., & Klemm, P. (2006). The Asymptomatic Bacteriuria Escherichia coli Strain 83972 Outcompetes Uropathogenic E . coli Strains in Human Urine, *74*(1), 615–624.

Rossmann, R., Sawers, G., & Böck, A. (1991). Mechanism of regulation of the formate-hydrogenlyase pathway by oxygen, nitrate, and pH: definition of the formate regulon. *Molecular Microbiology*, *5*(11), 2807–2814.

Rozas, J., & Rozas, R. (1995). DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput. Appl. Biosci.*, *11*(6), 621–625.

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology (Clifton, N.J.)*, *132*, 365–86.

Rychlik, W., & Rhoads, R. E. (1989). A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Research*, *17*(21), 8543–51.

Sabnis, N. a, Yang, H., & Romeo, T. (1995). Pleiotropic regulation of central carbohydrate metabolism in Escherichia coli via the gene csrA. *The Journal of Biological Chemistry*, *270*(49), 29096–104.

Sahu, S. N., Acharya, S., Tuminaro, H., Patel, I., Dudley, K., LeClerc, J. E., … Mukhopadhyay, S. (2003). The bacterial adaptive response gene, barA, encodes a novel conserved histidine kinase regulatory switch for adaptation and modulation of metabolism in Escherichia coli. *Molecular and Cellular Biochemistry*, *253*(1-2), 167–77.

Salvador, E., Wagenlehner, F., Köhler, C.-D., Mellmann, A., Hacker, J., Svanborg, C., & Dobrindt, U. (2012). Comparison of asymptomatic bacteriuria Escherichia coli isolates from healthy individuals versus those from hospital patients shows that long-term bladder colonization selects for attenuated virulence phenotypes. *Infection and Immunity*, *80*(2), 668–78.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–7.

Sauer, K. (2003). The genomics and proteomics of biofilm formation. *Genome Biology*, *4*(6), 219.

Sawers, R. G. (2005). Formate and its role in hydrogen production in Escherichia coli. *Biochemical Society Transactions*, *33*(Pt 1), 42–6.

Schägger, H., & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry*, *166*(2), 368–79.

Schellhorn, H. E., Audia, J. P., Wei, L. I., & Chang, L. (1998). Identification of conserved, RpoS-dependent stationary-phase genes of Escherichia coli. *Journal of Bacteriology*, *180*(23), 6283–91.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., … Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, *7*, 3.

Siguier, P., Gourbeyre, E., & Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiology Reviews*, *38*(5), 865–91.

Smith, C. A. (1992). Physiology of the Bacterial Cell. A Molecular Approach. *Biochemical Education*, *20*(2), 124–125.

Snyder, J. A., Haugen, B. J., Buckles, E. L., Lockatell, C. V., Johnson, D. E., Donnenberg, M. S., … Mmun, I. N. I. (2004). Transcriptome of Uropathogenic Escherichia coli during Urinary Tract Infection, *72*(11), 6373–6381.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, *98*(3), 503–517.

Stamey, T. A., & Sexton, C. C. (1975). The role of vaginal colonization with enterobacteriaceae in recurrent urinary infections. *The Journal of Urology*, *113*(2), 214–7.

Steinmetz, M., Le Coq, D., Djemia, H. Ben, & Gay, P. (1983). Analyse génétique de sacB, gène de structure d'une enzyme secrétée, la lévane-saccharase de Bacillus subtilis Marburg. *Molecular and General Genetics MGG*, *191*(1), 138–144.

Stephenson, M., & Stickland, L. H. (1932). Hydrogenlyases: Bacterial enzymes liberating molecular hydrogen. *The Biochemical Journal*, *26*(3), 712–24.

Stock, A. M., Robinson, V. L., & Goudreau, P. N. (2000). Two-component signal transduction. *Annual Review of Biochemistry*, *69*, 183–215.

Sundén, F., Håkansson, L., Ljunggren, E., & Wullt, B. (2006). Bacterial interference--is deliberate colonization with Escherichia coli 83972 an alternative treatment for patients with recurrent urinary tract infection? *International Journal of Antimicrobial Agents*, *28 Suppl 1*, S26–9.

Sundén, F., Håkansson, L., Ljunggren, E., & Wullt, B. (2010). Escherichia coli 83972 bacteriuria protects against recurrent lower urinary tract infections in patients with incomplete bladder emptying. *The Journal of Urology*, *184*(1), 179–85.

Suzuki, K., Babitzke, P., Kushner, S. R., & Romeo, T. (2006). Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E, (Gottesman 2004), 2605–2617.

Suzuki, K., Wang, X., Weilbacher, T., Pernestig, A. K., Georgellis, D., Babitzke, P., & Romeo, T. (2002). Regulatory Circuitry of the CsrA / CsrB and BarA / UvrY Systems of Escherichia coli, *184*(18), 5130–5140.

Tenaillon, O., Skurnik, D., Picard, B., & Denamur, E. (2010a). The population genetics of commensal Escherichia coli. *Nature Reviews. Microbiology*, *8*(3), 207–17.

Tenaillon, O., Skurnik, D., Picard, B., & Denamur, E. (2010b). The population genetics of commensal Escherichia coli. *Nature Reviews. Microbiology*, *8*(3), 207–17.

Teplitski, M., Goodier, R. I., & Ahmer, B. M. M. (2003). Pathways Leading from BarA / SirA to Motility and Virulence Gene Expression in Salmonella, *185*(24), 7257–7265.

Theodorsson-Norheim, E. (1986). Kruskal-Wallis test: BASIC computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples. *Computer Methods and Programs in Biomedicine*, *23*(1), 57–62.

Timmermans, J., & Van Melderen, L. (2010). Post-transcriptional global regulation by CsrA in bacteria. *Cellular and Molecular Life Sciences : CMLS*, *67*(17), 2897–908.

Tomenius, H., Pernestig, A. K., Jonas, K., Georgellis, D., Möllby, R., Normark, S., & Melefors, O. (2006). The Escherichia coli BarA-UvrY two-component system is a virulence determinant in the urinary tract. *BMC Microbiology*, *6*, 27.

Tyler, B. (1978). Regulation of the assimilation of nitrogen compounds. *Annual Review of Biochemistry*, *47*, 1127–62.

Uhlich, G. A., Cooke, P. H., & Solomon, E. B. (2006). Analyses of the red-dry-rough phenotype of an Escherichia coli O157:H7 strain and its role in biofilm formation and resistance to antibacterial agents. *Applied and Environmental Microbiology*, *72*(4), 2564–72.

Valdar, W. S. J. (2002). Scoring residue conservation. *Proteins*, *48*(2), 227–41.

Vejborg, R. M., Friis, C., Hancock, V., Schembri, M. a, & Klemm, P. (2010). A virulent parent with probiotic progeny: comparative genomics of Escherichia coli strains CFT073, Nissle 1917 and ABU 83972. *Molecular Genetics and Genomics : MGG*, *283*(5), 469–84.

Verstraeten, N., Braeken, K., Debkumari, B., Fauvart, M., Fransaer, J., Vermant, J., & Michiels, J. (2008). Living on a surface: swarming and biofilm formation. *Trends in Microbiology*, *16*(10), 496–506.

Villadsen, I. S., & Michelsen, O. (1977). Regulation of PRPP and nucleoside tri and tetraphosphate pools in Escherichia coli under conditions of nitrogen starvation. *Journal of Bacteriology*, *130*(1), 136–43.

Vincze, T., Posfai, J., & Roberts, R. J. (2003). NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Research*, *31*(13), 3688–91.

Wang, X., Dubey, A. K., Suzuki, K., Baker, C. S., Babitzke, P., & Romeo, T. (2005). CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of Escherichia coli. *Molecular Microbiology*, *56*(6), 1648–63.

Wang, X., Preston, J. F., & Romeo, T. (2004). The pgaABCD locus of Escherichia coli promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *Journal of Bacteriology*, *186*(9), 2724–34.

Warming, S., Costantino, N., Court, D. L., Jenkins, N. A., & Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Research*, *33*(4), e36.

Warren, J. W., Tenney, J. H., Hoopes, J. M., Muncie, H. L., & Anthony, W. C. (1982). A prospective microbiologic study of bacteriuria in patients with chronic indwelling urethral catheters. *The Journal of Infectious Diseases*, *146*(6), 719–23.

Weber, H., Polen, T., Heuveling, J., Wendisch, V. F., & Hengge, R. (2005). Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. *Journal of Bacteriology*, *187*(5), 1591–603.

Wei, B. L., Brun-Zinkernagel, a M., Simecka, J. W., Prüss, B. M., Babitzke, P., & Romeo, T. (2001). Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of Escherichia coli. *Molecular Microbiology*, *40*(1), 245–56.

Weilbacher, T., Suzuki, K., Dubey, A. K., Wang, X., Gudapaty, S., Morozov, I., … Romeo, T. (2003). A novel sRNA component of the carbon storage regulatory system of Escherichia coli. *Molecular Microbiology*, *48*(3), 657–70.

West, A. H., & Stock, A. M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in Biochemical Sciences*, *26*(6), 369–376.

White, R. H. (1987). Management of urinary tract infection. *Archives of Disease in Childhood*, *62*(4), 421–7.

WHO. (2002). WHO | Prevention of hospital-acquired infections: A practical guide. 2nd edition.

Wilson, K. (2001). Preparation of genomic DNA from bacteria. *Current Protocols in Molecular Biology /* Edited by Frederick M. Ausubel ... *[et al.]*, *Chapter 2*, Unit 2.4.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6), 80-83.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., … Achtman, M. (2006). Sex and virulence in Escherichia coli: an evolutionary perspective. *Molecular Microbiology*, *60*(5), 1136–51.

Yakhnin, H., Yakhnin, A. V, Baker, C. S., Sineva, E., Berezin, I., Romeo, T., & Babitzke, P. (2011). Complex regulation of the global regulatory gene csrA: CsrA-mediated translational repression, transcription from five promoters by $E\sigma^{70}$ and Eσ(S), and indirect transcriptional activation by CsrA. *Molecular Microbiology*, *81*(3), 689–704.

Yamamoto, K., Hirao, K., Oshima, T., Aiba, H., Utsumi, R., & Ishihama, A. (2005). Functional characterization in vitro of all two-component signal transduction systems from Escherichia coli. *The Journal of Biological Chemistry*, *280*(2), 1448–56.

Yan, F., & Polk, D. B. (2004). Commensal bacteria in the gut: learning who our friends are. *Current Opinion in Gastroenterology*, *20*(6), 565–71.

Zdziarski, J., Brzuszkiewicz, E., Wullt, B., Liesegang, H., Biran, D., Voigt, B., … Dobrindt, U. (2010). Host imprints on bacterial genomes--rapid, divergent evolution in individual patients. *PLoS Pathogens*, *6*(8), e1001078.

Zdziarski, J., Svanborg, C., Wullt, B., Hacker, J., & Dobrindt, U. (2008). Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? *Infection and Immunity*, *76*(2), 695–703.

Zhang, J. P., & Normark, S. (1996). Induction of gene expression in Escherichia coli after pilus-mediated adherence. *Science (New York, N.Y.)*, *273*(5279), 1234–6.

Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., … Kustu, S. (2000). Nitrogen regulatory protein C-controlled genes of Escherichia coli: scavenging as a defense against nitrogen limitation. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(26), 14674–9.

7.3 Publications

Ewert, B., Acquisti, A., Fürst, R., Berger, M., Dobrindt, U. (2015). When less is more: non-functional BarA-UvrY Two-component system may increase *Escherichia coli* fitness in urinary tract infections. (in preparation).

Trost, E., **Ewert, B**., Schiller, R., Ron, E., Oswald, E., Karch, H., Schouler, C., Dobrindt, U. (2015): Phenotypic and genomic characterization of chicken commensal *Escherichia coli* isolates. (in preparation).

Drey, F., Choi, Y.-H., Neef, K., **Ewert, B.**, Tenbrock, A., Treskes, P., … Wahlers, T. (2013). Noninvasive *in vivo* tracking of mesenchymal stem cells and evaluation of cell therapeutic effects in a murine model using a clinical 3.0 T MRI. *Cell Transplantation*, *22*(11), 1971–80.

Werner, G., Fleige, C., **Ewert, B.**, Laverde-Gomez, J. A., Klare, I., & Witte, W. (2010). High-level ciprofloxacin resistance among hospital-adapted Enterococcus faecium (CC17). *International Journal of Antimicrobial Agents*, *35*(2), 119–25.

7.4 Presentations

Ewert, B., Köhler, C.D., Dobrindt, U. (2011): Evaluation of the distribution of mutations in the BarA/UvrY two-component system of Extraintestinal pathogenic *Escherichia coli*. Jahrestagung der Deutschen Gesellschaft für Hygiene und Mikrobiologie (DGHM), Essen. Poster presentation

Ewert, B., Acquisti, C., Dobrindt, U. (2012): The BarA/UvrY two-component system determinants: mutational hotspots and its role for urovirulence in uropathogenic *Escherichia coli*. Jahrestagung der Deutschen Gesellschaft für Hygiene und Mikrobiologie (DGHM), Hamburg. Poster presentation

Ewert, B., Köhler, C.D., Dobrindt, U. (2013): Evaluation of gene regulation *in vivo* by means of mutation analysis of the BarA/UvrY two-component system of extraintestinal pathogenic *Escherichia coli*. Joint Status Seminar 2012 of the ERA-NET PathoGenoMics, Tenerife, Canary Islands. Poster presentation

Ewert, B., Acquisti, C., Dobrindt, U. (2013): The BarA/UvrY two-component system determinants: mutational hotspots and its role for urovirulence in uropathogenic *Escherichia coli.* Annual MGSE (Münster Graduate School of evolution) symposium, Münster. Poster presentation.

Ewert, B., Acquisti, C., Dobrindt, U. (2013): The *Escherichia coli* BarA/UvrY two-component system: a key-factor for the adaptation to growth in the urinary tract? Jahrestagung der Deutschen Gesellschaft für Hygiene und Mikrobiologie (DGHM), Rostock. Poster presentation.

Ewert, B., Acquisti, C., Dobrindt, U. (2014): The *Escherichia coli* BarA/UvrY two-component system: a key-factor for the adaptation to growth in the urinary tract? Molecular UTI Conference, Malmö/Lund, Sweden. Poster presentation

Ewert, B., Acquisti, C., Dobrindt, U. (2014): The *Escherichia coli* BarA/UvrY two-component system: a key-factor for the adaptation to growth in the urinary tract? 1st International Symposium on Autoinflammation Breaks Barriers, Münster, Poster presentation

7.5 Curriculum vitae