



Genomic peculiarity of coding sequences and metabolic potential of probiotic *Escherichia coli* strain Nissle 1917 inferred from raw genome data

Jibin Sun^{a,1}, Florian Gunzer^{b,1}, Astrid M. Westendorf^c, Jan Buer^{b,c},
Maren Scharfe^d, Michael Jarek^d, Frank Gößling^d,
Helmut Blöcker^{d,*}, An-Ping Zeng^{a,*}

^a GBF – German Research Centre for Biotechnology, Experimental Bioinformatics, Mascheroder Weg 1, D-38124 Braunschweig, Germany

^b Medical School Hannover, Department of Medical Microbiology and Hospital Epidemiology,
Carl Neuberg Strasse 1, D-30625 Hannover, Germany

^c GBF – German Research Centre for Biotechnology, Mucosal Immunity, Mascheroder Weg 1, D-38124 Braunschweig, Germany

^d GBF – German Research Centre for Biotechnology, Department of Genome Analysis, Mascheroder Weg 1,
D-38124 Braunschweig, Germany

Received 8 October 2004; received in revised form 16 December 2004; accepted 7 January 2005

Abstract

Probiotic *Escherichia coli* strain Nissle 1917 (O6:K5:H1) is a commensal *E. coli* isolate that has a long tradition in medicine for the treatment of various intestinal disorders in humans. To elucidate the molecular basis of its probiotic nature, we started sequencing the genome of this organism with a whole-genome shotgun approach. A 7.8-fold coverage of the genomic sequence has been generated and is now in the finishing stage. To exploit the genome data as early as possible and to generate hypotheses for functional studies, the unfinished sequencing data were analyzed in this work using a new method [Sun, J., Zeng, A.P., 2004. IdentiCS—identification of coding sequence and in silico reconstruction of the metabolic network directly from unannotated low-coverage bacterial genome sequence. *BMC Bioinformatics* 5, 112] which is particularly suitable for the prediction of coding sequences (CDSs) from unannotated genome sequence. The CDSs predicted for *E. coli* Nissle 1917 were compared with those of all five other sequenced *E. coli* strains (*E. coli* K-12 MG1655, *E. coli* K-12 W3110, *E. coli* CFT073, EHEC O157:H7 EDL933 and EHEC O157:H7 Sakai) published to date. Five thousand one hundred and ninety-two CDSs were predicted for *E. coli* Nissle 1917, of which 1065 were assigned with enzyme EC numbers. The comparison of all predicted CDSs of *E. coli* Nissle 1917 to the other *E. coli* strains revealed 108 CDSs specific for this isolate. They are organized as four big genome islands and many other smaller gene clusters. Based on CDSs with EC numbers for enzymes, the potential metabolic network of Nissle 1917 was reconstructed and compared to those of the other five *E. coli* strains. Overall, the comparative genomic analysis sheds light on

* Corresponding authors.

E-mail addresses: bloecker@gbf.de (H. Blöcker), anping.zeng@gbf.de (A.-P. Zeng).

¹ Both authors contributed equally to this work.

the genomic peculiarity of the probiotic *E. coli* strain Nissle 1917 and is helpful for designing further functional studies long before the sequencing project is completely finished.

© 2005 Published by Elsevier B.V.

Keywords: Genome sequence; Comparative genomics; Coding sequence; Metabolic network; Probiotics; *E. coli* Nissle 1917

1. Introduction

E. coli strain Nissle 1917 (DSM 6601) is a non-pathogenic fecal isolate (Nissle, 1916, 1925) of serotype O6:K5:H1 which is sold as a drug under the name “Mutaflor®”. It has been used as a probiotic agent in the preantibiotic era since the early 1920s to treat infectious diseases of the intestine (Geisse, 1919; Nissle, 1919). Recently, clinical studies have shown that administration of the strain is equivalent to standard medication in remission maintenance of ulcerative colitis (Rembacken et al., 1999; Kruis et al., 2004) and may be an alternative regimen for the treatment of colonic Crohn’s disease (Malchow, 1997). *E. coli* Nissle 1917 successfully colonizes the gut of humans (Lodinova-Zadnikova and Sonnenborn, 1997) and animals (Gunzer et al., 2002; Waidmann et al., 2003). It does not cause colitis, even when gnotobiotic piglets (Gunzer et al., 2002) or interleukin-2-deficient mice (Waidmann et al., 2003) mono-associated with this strain produce high bacterial counts in the feces. Recently we could show, that recombinant intestinal *E. coli* Nissle 1917 had no effect on migration, clonal expansion and activation status of specific CD4⁺ T cells, neither in healthy mice nor in animals with acute colitis (Westendorf et al., 2005). Thus, *E. coli* Nissle 1917 meets all requirements for a biotherapeutic agent generally recognized as a safe organism for human use. Despite the successful therapeutic applications of *E. coli* Nissle 1917, only limited information is available about the beneficial traits contributing to the strain’s probiotic character. Several features such as expression of two microcins (Patzner et al., 2003), presence of six different iron uptake systems (Grozdanov et al., 2004), a semi-rough lipopolysaccharide phenotype and serum sensitivity or absence of known protein toxins (Blum et al., 1995; Grozdanov et al., 2002) have been described that might be advantageous for *E. coli* Nissle 1917 in competing with other intestinal bacteria or adapting to the intestinal situation. However, the mechanisms underlying the probiotic na-

ture, especially at the molecular level, yet have to be elucidated.

The rapid progress in genome sequencing techniques and new tools for functional genomic studies now offer a great chance to reveal the underlying genetic make-up and molecular mechanisms responsible for the probiotic qualities of *E. coli* Nissle 1917. The most detailed analysis on the genome structure of this strain was published recently by Grozdanov et al. (2004). The authors used three genomic approaches, namely sequence context screening of tRNA genes as a potential indication of chromosomal integration of horizontally acquired DNA, sequence analysis of 280 kb of genomic islands (GEIs) and DNA–DNA hybridisation using whole-genome microarrays for comparison of *E. coli* Nissle 1917 genome content with that of other *E. coli* strains. In particular, the analysis of four partially sequenced GEIs identified many known fitness factor determinants such as microcins, different iron uptake systems, adhesins and proteases. It is suggested that the lack of several defined virulence markers such as *E. coli* alpha-hemolysin and P-fimbrial adhesins combined with the expression of the above mentioned fitness factors most likely contributes to the probiotic character of *E. coli* Nissle 1917. However, to have a broader and deeper insight into the genomic content and thus the genetic basis for the probiotic behaviour of this strain, a complete sequencing of its entire genome is necessary. We set out to sequence the whole genome of *E. coli* Nissle 1917 using high throughput automated sequence analysis of shotgun libraries. In order to explore the genomic data for functional experimental studies as soon as possible, we started to analyse an early version of the raw sequence with the program IdentiCS that was specially developed to predict bacterial protein-coding sequences (CDSs) from unannotated genome data (Sun and Zeng, 2004). The CDSs predicted for *E. coli* Nissle 1917 were compared to those of five other sequenced *E. coli* strains, *E. coli* K-12 MG1655 (Blattner et al., 1997), *E. coli* K-12 W3110 (Yamamoto et al., 1997), uropathogenic *E. coli*

(UPEC) CFT073 (Welch et al., 2002), enterohemorrhagic *E. coli* (EHEC) O157:H7 EDL933 (Perna et al., 2001) and Sakai (Hayashi et al., 2001). Efforts were also made to reconstruct the whole metabolic network of *E. coli* Nissle 1917 and to compare its metabolic potential with other *E. coli* strains.

2. Materials and methods

2.1. Preparation of genomic DNA for sequencing

Pure cultures of *E. coli* Nissle 1917 (serotype O6:K5:H1; DSM 6601) were isolated from a tablet of Mutaflor[®] dissolved in PBS and streaked onto MacConkey agar plates (Oxoid, Wesel, Germany). A single colony was inoculated into LB broth (Invitrogen, Karlsruhe, Germany) and grown overnight. Genomic DNA was then isolated from liquid cultures using Qiagen-tip 100 columns and the Genomic DNA Buffer Set from Qiagen (Hilden, Germany), according to manufacturers' protocols.

2.2. Shotgun libraries

Chromosomal DNA was disrupted mechanically with a Hydroshear[®] instrument (GeneMachines, Ann Arbor, USA) into fragments of 1.5–3.0 kb in size. They were cloned into SmaI linearized high copy vectors pUC18 (Norlander et al., 1983) and pTZ18r (Mead et al., 1986). Recombinant plasmids were transformed into electroporation competent *E. coli* DH10B (Invitrogen). Seven shotgun libraries of 69,888 clones were organized and transferred into 384-well plates by a Q-Bot picking robot (Genetix, Munich, Germany) and stored at -70°C .

2.3. BAC library

Genomic DNA fragments of about 20 kb in size were generated by enzymatic digest with Sau3A I and cloned into BamHI linearized PCC 1-Bac vector (EPICENTRE, Madison, USA). The BAC library was transformed into *E. coli* strain TransforMax EPI300-T1 (EPICENTRE) and stored at -70°C .

2.4. Automated sequencing

Plasmid DNA from the shotgun clones was isolated using MultiScreen[™] 96-well columns (Milli-

pore, Schwalbach, Germany) and subjected to Taq cycle sequencing amplification (PTC-225, MJ Research, Waltham, USA) using primers CTGCAGGTCGACTCTAGAGGATC and CTATAGGGAATTCGAGCTCGGTAC in an automated DNA purification and amplification setup. Automated sequencing with dye terminator chemistry was performed on MegaBACE 1000 and 4000 (Molecular Dynamics, Albertville, USA), ABI 377 (Applied Biosystems, Darmstadt, Germany) and Long Reader 4200 (LI-COR, Bad Homburg, Germany) sequencers.

2.5. Coding sequence prediction and function assignment

The raw sequence reads were assembled using Phrap (<http://www.phrap.com>). This resulted in 95 contigs (>2 kb) corresponding to a 7.8-fold coverage (Phred20) of the genome. All contigs were analyzed by using the program IdentiCS (Sun and Zeng, 2004) (<http://genome.gbf.de/bioinformatics/index.html>). This program that uses a reversed query process was developed to identify coding sequences (CDS) directly from unannotated raw genome sequence and at the same time to reconstruct, visualize and compare the metabolic network. This large-scale comparative-genome based method is able to minimize the negative effects of sequencing errors, and therefore, to obtain a better prediction and a simultaneous function assignment of CDSs. Blast search was done through a local Blast program (Altschul et al., 1997) or over the webserver of NCBI (<http://www.ncbi.nlm.nih.gov/blast>).

2.6. Metabolic network reconstruction, visualization and comparison

The metabolic network of *E. coli* Nissle 1917 was reconstructed based on the enzymes (with EC numbers) predicted from the genome sequence and compared to those of the five other *E. coli* strains with IdentiCS. If an enzyme is present in a strain (suppose strain A) but absent in another (suppose strain B), then the absence is confirmed by searching again in the genome B for the homolog of the enzyme in strain A. In this way, mistakes or insufficiencies in the annotation can be avoided. A Microsoft Excel file is then generated for the visualization and comparison of metabolic networks based on the metabolic maps of KEGG (Kanehisa et

al., 2004). The map-by-map comparison serves as a user friendly platform to explore the metabolic potential in detail.

3. Results and discussion

3.1. Overall comparison of protein-coding sequences in different *E. coli* strains

Using the software program IdentiCS (Sun and Zeng, 2004) 5192 CDSs are predicted from the 7.8-fold genomic sequence of *E. coli* Nissle 1917, of which 1065 are assigned with enzyme EC numbers (Table 1). The coding sequences of each *E. coli* isolate were compared to those of the other *E. coli* strains. The differences in the protein-coding sequences of these strains are summarized in Table 2. In consistence with Grozdanov et al. (2004), at genome level *E. coli* Nissle 1917 is found to be more similar to *E. coli* CFT073 than to any other *E. coli* strains compared. *E. coli* MG1655 is more similar to *E. coli* W3110, and EHEC O157:H7 Sakai more similar to EHEC EDL933. Around one fifth of the coding sequences of *E. coli* Nissle 1917 or CFT073 are different from those of the other four strains. An even bigger difference exists when comparing the O157:H7 strain

to others, indicating dramatic changes (i.e. genomic island obtainment, DNA rearrangement and DNA deletion events) among these three groups of *E. coli* strains during their evolution.

Although *E. coli* CFT073 is more similar to *E. coli* Nissle 1917 than to others, significant difference still exists between these two strains. *E. coli* Nissle 1917 possesses 166 CDSs absent in *E. coli* CFT073, whereas *E. coli* CFT073 has 350 CDSs that are absent in *E. coli* Nissle 1917. The *E. coli* CFT073 specific CDSs are mainly organized as big genomic islands including three prophage islands (containing 48, 75 and 36 *E. coli* CFT073-specific CDSs, respectively), two pathogenicity islands (containing 21 and 42 *E. coli* CFT073-specific CDSs, respectively), and another big genomic island (containing 60 *E. coli* CFT073-specific CDSs). The comparison of all predicted CDSs for *E. coli* Nissle 1917 to other five *E. coli* strains reveals 108 CDSs specific for *E. coli* Nissle 1917 (Table 3). They are organized as four big genome islands, including a ferric dicitrate transport island, a capsule synthesis and type II secretion island, a phage island and an island with unknown function, and many other smaller gene clusters. Some of these *E. coli* Nissle 1917 specific CDSs are illustrated in more detail below.

Table 1
Comparison of genome size, protein-coding sequences and enzymes in the six different *E. coli* strains

Abbreviation	Strain	Serotype	Genome size	Number of proteins	Proteins with EC number	Unique EC numbers
ecn	<i>E. coli</i> Nissle 1917	O6:K5:H1	5061544 (167 contigs)	5192	1065	668
ecc	<i>E. coli</i> CFT073	O6:K2:H1	5231428	5379	1077	647
eco	<i>E. coli</i> K-12 MG1655	OR:K ⁻ :H48	4639221	4289	1199	716
ecj	<i>E. coli</i> K-12 W3110	OR:K ⁻ :H48	4641433	4390	1209	720
ecs	EHEC Sakai	O157:H7	5498450	5361	1168	701
ece	EHEC EDL933	O157:H7	5528445	5349	1173	700

Table 2
Number of different coding sequences among six *E. coli* strains

Specific for	In comparison to						All others
	ecn	ecc	eco	ecj	ece	ecs	
ecn	0	166	965	965	882	874	108
ecc	350	0	1243	1243	1047	1039	222
eco	532	508	0	11	384	377	11
ecj	555	540	2	0	413	406	2
ece	1343	1141	1316	1316	0	15	11
ecs	1459	1241	1432	1431	70	0	66

Table 3
Protein-coding sequences specific in *E. coli* Nissle 1917

CDS number	Size (bp)	Contig	Function	Homolog	Similarity <i>e</i> -value/identity (%)	Remarks
ecn0400	1446	107	Hypothetical protein YPO0387	Q8ZIV1	2E–45/28	Island GI1
ecn0401	717	107	Putative glycosyl transferase (Hypothetical)	Q8ZJC6	4E–35/35	Island GI1
ecn0119	1218	83	Hypothetical protein YPO0388	Q8ZIV0	2E–83/41	Island GI1
ecn0118	1732	83	Restriction endonuclease	Q8RNY7	0/99	Restriction-modification system type II
ecn0117	606	83	Site-specific DNA-methyltransferase	Q8YUQ9	2E–25/35	Restriction-modification system type II
ecn0116	1458	83	M6 adenine DNA methyltransferase [EC:2.1.1.73]	Q8RNY5	0/99	Restriction-modification system type II
ecn0120	531	83	GTPase subunit of restriction endonuclease	Q8ZIV1	2E–44/51	Restriction-modification system type II
ecn3541	210	160	M6 adenine DNA methyltransferase	Q8RNY5	1E–23/77	Restriction-modification system type II
ecn3542	345	160	Putative cytoplasmic protein	Q8ZQE1	6E–38/67	Island GI1
ecn3558	915	160	Hypothetical 36.2 kDa protein	Q93F05	1E–178/100	Island GI1
ecn3559	213	160	Hypothetical 7.9 kDa protein	Q93F06	4E–39/100	Island GI1
ecn3560	248	160	Hypothetical 9.5 kDa protein	Q93F07	5E–39/96	Island GI1
ecn3572	1356	160	Hypothetical 49.1 kDa protein	Q93F16	0/100	Island GI1
ecn3588	570	160	Putative lysogenic conversion protein	Q858R9	8E–79/70	Island GI1
ecn3589	528	160	Hypothetical protein	Q87W25	3E–11/26	Island GI1
ecn3592	840	160	Hypothetical protein XCC0340	Q8PDJ6	2E–63/45	Island GI1
ecn0737	174	119	Hypothetical 6.5 kDa protein	Q8W609	7E–15/60	Island GI2
ecn0738	426	119	Putative peptidase [EC:3.4.21.–]	Q8ZNJ5	1E–67/85	UV protection and mutation
ecn0739	210	119	Homolog of MsgA, SsrB-regulated factor	Q8ZNJ4	3E–35/98	Affects survival in macrophages.
ecn0744	3895	119	Tail fiber	Q9MCR7	0/66	Bacteriophage
ecn0745	609	119	Probable phage HK022 GP20-related protein	Q8XYR2	2E–36/38	Bacteriophage
ecn0746	705	119	Gp19	Q9MCU3	5E–64/46	Bacteriophage
ecn0747	726	119	Gp18	O64332	1E–70/51	Bacteriophage
ecn0748	330	119	Putative phage protein (putative phage tail protein)	Q8ZEP0	3E–23/44	Bacteriophage
ecn0753	399	119	Gifsy-2 prophage probable minor tail protein	Q8ZQ88	4E–15/33	Bacteriophage
ecn0755	369	119	Hypothetical protein	Q8FI66	5E–17/46	Bacteriophage
ecn0756	399	119	Gifsy-2 prophage probable minor tail protein	Q8ZQ90	7E–45/60	Bacteriophage
ecn0757	570	119	Gifsy-1 prophage: similar to minor tail protein Z	Q8ZN01	3E–57/59	Bacteriophage
ecn0758	273	119	Hypothetical 106 kDa protein	Q8VNN4	2E–24/57	Bacteriophage
ecn0761	210	119	Hypothetical protein ECs0826	Q8X3E1	9E–26/72	Bacteriophage
ecn0763	474	119	Hypothetical protein Z0963	Q8X876	2E–48/60	Bacteriophage
ecn1527	1308	136	Putative integrase	Q8W658	0/89	Bacteriophage
ecn1528	234	136	Putative excisionase	Q8W657	6E–41/89	Bacteriophage
ecn1529	567	136	Sb30	Q8HAA8	2E–85/78	Bacteriophage

Table 3 (Continued)

CDS number	Size (bp)	Contig	Function	Homolog	Similarity <i>e</i> -value/identity (%)	Remarks
ecn1530	237	136	Hypothetical bacteriophage protein	AAP17294	2E–17/51	Bacteriophage
ecn1531	384	136	Hypothetical bacteriophage protein STY1027	Q8Z7X5	4E–46/70	Bacteriophage
ecn1532	180	136	Ead protein	VEAD_BPP22	6E–16/65	Bacteriophage
ecn1533	237	136	Hypothetical protein ECs2628	Q8X2I3	2E–14/42	Bacteriophage
ecn1534	546	136	Hypothetical protein YfdR	Q8FET1	5E–64/62	Bacteriophage
ecn1535	903	136	Recombination associated protein RdgC	RDGC_SALTY	1E–121/69	Bacteriophage
ecn1536	393	136	Fels-1 prophage, predicted transcriptional regulator	Q8ZQI4	2E–26/45	Bacteriophage
ecn1540	384	136	Sb45	Q8HA93	5E–63/85	Island GI2
ecn1541	624	136	Phage regulatory protein (putative phage antirepressor)	Q8ZEQ5	7E–43/46	Island GI2
ecn1542	987	136	Hypothetical protein YdfU	Q8FEU5	1E–144/69	Island GI2
ecn1543	267	136	Hypothetical protein	Q8KTV8	3E–14/41	Island GI2
ecn1544	669	136	Antitermination protein Q	GQ_BP82	2E–28/30	Bacteriophage
ecn1545	870	136	Periplasmic serine proteases (ClpP class)	Q8NPI9	5E–75/46	Island GI2
ecn1546	276	136	Hypothetical 12.1 kDa protein (hypothetical lipoprotein)	Q8XC83	8E–26/65	Island GI2
ecn1547	1041	136	Hypothetical 40.3 kDa protein	Q8SBE2	1E–165/78	Island GI2
ecn1548	360	136	Hypothetical protein	Q8KTU9	1E–19/35	Island GI2
ecn1551	469	136	Putative secreted protein	Q8Z7W1	2E–43/60	Island GI2
ecn1552	189	136	Gp56	O64364	1E–12/55	Bacteriophage
ecn0089	405	78	KfoB hypothetical protein	Q8L0V3	1E–33/51	Capsule synthesis
ecn0091	270	78	Hypothetical ORF protein	Q47331	2E–37/78	Capsule synthesis
ecn0092	222	78	Hypothetical protein	AAP42476	7E–20/57	Capsule synthesis
ecn0093	1557	78	Putative glycosyltransferase	Q47330	0/97	Capsule synthesis
ecn3705	714	161	KfiA protein	Q47332	1E–139/100	Capsule synthesis
ecn3711	561	161	Hypothetical type II secretion protein GspJ	Q8VRM4	1E–102/97	Type II secretion
ecn3712	369	161	Hypothetical type II secretion protein	Q8VPC3	8E–61/94	Type II secretion
ecn3713	528	161	Hypothetical type II secretion protein GspH	Q8VRM6	3E–92/93	Type II secretion
ecn3715	1221	161	Hypothetical type II secretion protein	Q8VPC6	0/83	Type II secretion
ecn3716	1491	161	Hypothetical type II secretion protein GspE	Q8VRM9	0/98	Type II secretion
ecn3717	1848	161	Hypothetical type II secretion protein GspD	Q8VRN0	0/98	Type II secretion
ecn4407	603	165	Hypothetical protein STY4039	Q8Z2I7	2E–78/71	Island GI4
ecn4408	1287	165	Putative ATP binding protein SugR	Q9Z628	0/86	Island GI4
ecn4412	453	165	Putative radC-like protein YfjY	YFJY_ECOLI	9E–44/54	Toxin–antitoxin
ecn4413	375	165	Putative radC-like protein YkfG	YKFG_ECOLI	3E–13/34	Toxin–antitoxin
ecn4416	909	165	Orf, conserved hypothetical protein	Q83JP5	5E–15/25	Island GI4
ecn4417	585	165	Hypothetical protein	Q8GC57	1E–34/43	Island GI4

Table 3 (Continued)

CDS number	Size (bp)	Contig	Function	Homolog	Similarity <i>e</i> -value/identity (%)	Remarks
ecn4418	177	165	Hypothetical protein	Q8GA57	2E–22/72	Island GI4
ecn4419	861	165	Hypothetical protein sap	AAP18287	5E–61/44	Island GI4
ecn4420	675	165	Hypothetical protein	Q8GA21	4E–54/48	Island GI4
ecn4421	987	165	Conserved hypothetical protein	Q89Z46	2E–26/26	Island GI4
ecn4422	432	165	Hypothetical protein Z0327	Q8X7K4	6E–13/26	Island GI4
ecn4423	1311	165	Conserved hypothetical protein	Q89Z44	1E–20/25	Island GI4
ecn0094	426	79	Hypothetical 16.6 kDa protein	Q93FI4	3E–75/90	Plasmid pMUT1
ecn0095	426	79	Hypothetical 16.2 kDa protein	Q93FI5	2E–79/99	Plasmid pMUT1
ecn0287	1583	102	Putative mobilisation protein	Q9RLF1	0/94	Plasmid pMUT2
ecn0288	534	102	Hypothetical 31.5 kDa protein	O52996	6E–29/40	Plasmid pMUT2
ecn0289	306	102	MobC protein	Q47384	1E–19/48	Plasmid pMUT2
ecn0290	330	102	Conserved hypothetical protein	Q888D5	7E–13/32	Plasmid pMUT2
ecn0291	867	102	Putative replicon protein	Q9RLE8	1E–147/84	Plasmid pMUT2
ecn0292	318	102	Putative RelE protein	Q8GMN9	8E–35/63	Plasmid pMUT2
ecn1754	300	140	Hypothetical protein (putative cytoplasmic protein)	Q8XFJ7	3E–44/84	
ecn1755	495	140	Probable restriction system protein Mrr	D94741	2E–19/41	Restriction-modification system type I
ecn1756	1578	140	HsdM	O34139	0/93	Restriction-modification system type I
ecn1757	405	140	Restriction-modification system, type I [EC:3.1.21.3]	Q82XA3	4E–18/37	Restriction-modification system type I
ecn1758	3261	140	Sty SBLI	P72420	0/92	Restriction-modification system type I
ecn1790	438	140	Hypothetical 19.6 kDa protein	Q9L910	3E–79/97	
ecn0529	258	112	Hypothetical 19.6 kDa protein	Q9L910	9E–46/100	
ecn3228	516	157	Hypothetical 19.6 kDa protein	Q9L910	3E–98/100	
ecn2273	1527	148	IS3 family transposase	Q8UDL7	1E–154/51	Transposon
ecn2293	636	148	L-Fucose phosphate aldolase (L-fucose-1-phosphate aldolase) [EC:4.1.2.17]	FUCA_ECOLI	5E–70/58	L-Fucose metabolism
ecn2294	1080	148	Methylthioribose-1-phosphate isomerase; MTR-1-P isomerase [EC:5.3.1.23]	MTNA_BACSU	1E–126/62	Methionine salvage pathway
ecn2295	1236	148	5-Methylthioribose kinase; MTR kinase [EC:2.7.1.100]	G1EAA23684	3E–91/44	Methionine salvage pathway
ecn0020	687	34	Replication protein RepA	REP5_ECOLI	1E–118/93	Essential for origin function
ecn0021	1233	35	MtaE	Q9RFK7	2E–98/49	Complex polyketide formation
ecn0022	423	37	MtaF	Q9RFK6	3E–62/73	Complex polyketide formation
ecn0024	1178	39	MtaF	Q9RFK6	5E–63/53	Complex polyketide formation
ecn0025	630	39	Mucin 1 precursor (MUC-1)	MUC1_HUMAN	1E–17/32	Major component of the outer cell wall
ecn0060	351	70	LF3	Q8UZE1	7E–11/35	Major component of the outer cell wall

Table 3 (Continued)

CDS number	Size (bp)	Contig	Function	Homolog	Similarity <i>e</i> -value/identity (%)	Remarks
ecn0023	1170	38	LF3	Q8UZE1	3E–14/25	Biosynthesis of thiamine
ecn0026	519	40	Membrane-associated lipoprotein involved in thiamine biosynthesis	Q8R7D3	3E–32/39	
ecn0030	390	48	Probable isoleucyl-tRNA synthetase [EC:6.1.1.5]	D75273	1E–39/61	Protein synthesis
ecn0031	177	52	Recombinase FLP protein (protein able)	FLP.YEAST	4E–29/100	DNA recombination
ecn4546	1158	166	Prophage CP4-57 integrase	INTA.ECOLI	1E–119/54	Bacteriophage
ecn4560	486	166	Gp1	O64316	2E–38/46	Bacteriophage
ecn0040	372	57	Conserved hypothetical protein	Q8FJ24	4E–14/43	

3.2. Genomic peculiarity of coding sequences in *E. coli* Nissle 1917

3.2.1. Capsule biosynthesis

Genes responsible for the biosynthesis of K5 capsule are known to be organized in gene clusters (Pazzani et al., 1993; Whitfield and Roberts, 1999). All of the components for the capsule biosynthesis, including *kfiA–D* required for the polymerization, *kpsC–F*, *M*, *T*, *S*, *U* for the translocation of polysaccharide across the plasma membrane, are identified in the genomic sequence of *E. coli* Nissle 1917 (Fig. 1). No homologs to *kfiA*, *kfiB*, *kfiC* and *kfiD* are found in the other five *E. coli* strains whereas homologs to *kps* genes are also found in *E. coli* CFT073 but not in the other four *E. coli* strains, indicating the diversity of biosynthesis of K antigens in *E. coli* strains. Genes responsible for the translocation of group 1 capsule across the outer membrane, *wza*, *wzb* and *wzc*, are also found in *E. coli* Nissle 1917, clustered with many genes for biosynthesis of colanic acid, an exopolysaccharide (results not shown). Although the average central cavity diameter (2.28 nm) of the ring-like structure formed by Wza multimers is relatively small compared to the large molecular weight of capsular polysaccharides (CPS)

(>100,000 Da), CPSs are considered to be able to be “threaded” through the channel as linear strand because of the flexible random coil structure in solution (Nesper et al., 2003). Till now, experimental proof for genes responsible for the outer membrane translocation is still missing.

3.2.2. Type II secretion apparatus

Interestingly, just next to the cluster of the capsule biosynthesis, another large gene cluster (*gspC–M*) coding for a type II secretion pathway is identified in *E. coli* Nissle 1917 (Fig. 1). The type II secretion pathway, also named as the main terminal branch of the general secretion pathway, is known to be responsible for the extracellular secretion of toxins and hydrolytic enzymes, many of which contribute to pathogenesis of microorganisms (Sandkvist, 2001). The complex secretion apparatus is composed of over 10 proteins. From the genome of *E. coli* Nissle 1917, two different type II secretion clusters are identified. One is nearly identical (95–100%) to the only type II secretion system of the completely sequenced *E. coli* K-12 strains and the uropathogenic CFT073 strain. Its function is believed to secrete chitinase in *E. coli* K-12 (Francetic et al., 2000). In the other type II secretion

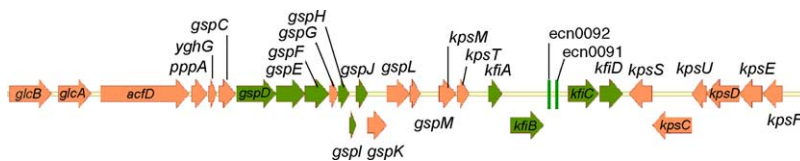


Fig. 1. Genes involved in biosynthesis of capsule and clusters of translocation and type II secretion. Genes marked with green color are specific for *E. coli* Nissle 1917.

cluster, the components GspD, E, F are merely 45–59% identical to the corresponding ones of other *E. coli* strains while GspH, I, J have only weak homologies (identity < 30%, *e*-value > 1E–10) to the big plasmid of EHEC O157:H7 strains. Searching the protein database Swissprot and TrEMBL revealed that all the components of this cluster are 91–99% identical to the protein secretion pathway for the secretion of heat-labile enterotoxin by the enterotoxigenic *E. coli* strain H10407 of serotype O78:K80:H11 (Tauschek et al., 2002). The hits to the corresponding components of GspH, I and J from *Vibrio cholerae* come to the second position. The genes coding for heat-labile enterotoxin cannot be found in the genome of *E. coli* Nissle 1917. This may reflect the incompleteness of the current genomic sequence or indicate another function of this system. It was reported that mutations in the genes of a type II secretion cluster interfere with polysaccharide production and biofilm formation in *V. cholerae* (Ali et al., 2000). Based on the genomic context of the second type II secretion cluster, it may be hypothesized that this system could also interact with or even be involved in the capsule formation process.

3.2.3. Methionine salvage

Methionine is normally activated by *S*-adenosylmethionine synthetase (coded by the gene *metK*) as *S*-adenosylmethionine (SAM) before taking part in metabolic reactions (Fig. 2). SAM can be involved in SAM-dependent methyl transference, producing *S*-adenosyl homocysteine (SAH), or in SAM-dependent aminopropyl-group transference, producing methylthioadenosine (MTA) as by-product. Both SAH and MTA should be recycled to generate methionine to utilize the sulfur more economically, and therefore, gain a better survival in certain ecological niches. All six *E. coli* strains are found to use Pfs and LuxS for the complete cycle of SAH. In contrast, the genes to recycle MTA to methionine are incomplete (Fig. 2). All strains can also activate methionine to SAM, generate spermidine by two successive reactions encoded by the genes *speD* and *speE*, and also convert MTA to methylthioribose (MTR). However, only *E. coli* Nissle 1917 can further convert MTR to methylthioribulose-1-phosphate (MTRu-1-P) by two enzymes MTR kinase (coded by *mtnK*, *ecn2295*) and methylthioribose-1-phosphate

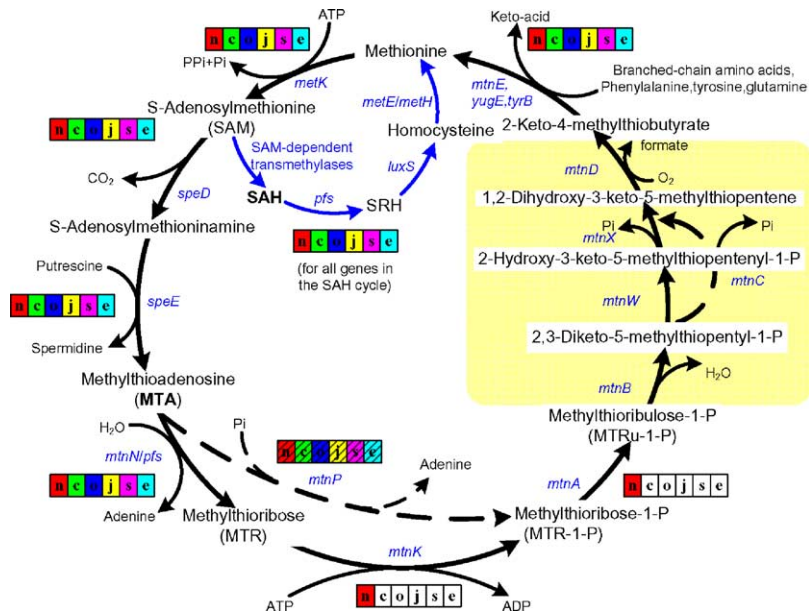


Fig. 2. Genes involved in methionine cycle in different *E. coli* strains. n, c, o, j, s, e represent *E. coli* Nissle 1917, CFT073, MG1655, W3110, EHEC O157:H7 Sakai and EDL933, respectively. Colored backgrounds for these strains represent the existence of corresponding genes. Striped backgrounds of these strains represent the potential existence of the genes. Blank background means absence. The yellow shadow on the metabolic pathway means that genes for these reactions are found in none of the six *E. coli* strains (modified based on Sekowska et al., 2004).

(MTR-1-P) isomerase (coded by *mtnA*, ecn2294). Purine nucleoside phosphorylase, a homolog to MTA phosphorylase (*mtnP*), is found in all six *E. coli* strains, that probably could catalyze the direct phosphorylation of MTA to MTR-1-P. The subsequent pathway for further converting MTRu-1-P through 2,3-diketo-5-methylthiopentyl-1-P to methionine, and thus, closing the methionine cycle, which is present in many other organisms such as *Bacillus subtilis*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Leptospira interrogans*, *Thermoanaerobacter tengcongensis* and *Xylella fastidiosa* (Sekowska et al., 2004), cannot be found in all six *E. coli* strains. Homologs to the gene *mtnE* encoding for 2-keto-4-methylthiobutyrate aminotransferase for the last step of this methionine cycling pathway are found in all six *E. coli* isolates.

In many bacteria (including *E. coli*), MTR was found to accumulate in external culture medium (Schroeder et al., 1973). This is consistent to the incompleteness of the methionine cycle discussed above. No observation for *E. coli* Nissle 1917 regarding MTR accumulation has been reported so far. However, *E. coli* Nissle 1917 may behave differently since it clearly has the two enzymes to further metabolize MTR. This may benefit to its fitness under certain milieu conditions. The missing link in the methionine cycle is still to be defined. Interestingly, some enzymes in the MTA cycle are both diverse and unusual (Sekowska et al., 2004). For example, *P. aeruginosa* uses MTA phosphorylase (coded by *mtnP*) instead of Pfs or MtnK to phosphorylate MTA while *K. pneumoniae* uses an enolase-phosphatase coded by *mtnC* instead of an enolase and a phosphatase coded by *mtnX* and *mtnW*, respectively, for the conversion of 2,3-diketo-5-methylthiopentyl-1-P to 1,2-dihydroxy-3-keto-5-methylthiopentene. The sequence of MTR-1-P isomerase encoded by *mtnA* has a strong similarity to that of the eukaryotic initiation factor eIF-2B (alpha subunit) whereas the sequence of the enolase encoded by *mtnW* in *B. subtilis* is highly similar to RuBisCO. The incompleteness of the MTA cycle in *E. coli* Nissle 1917 may also indicate the existence of an alternative novel pathway. In this context, it is noticed that just adjacent to *mtnA*, there is another *E. coli* Nissle 1917-specific gene coding for L-fucose phosphate aldolase (ecn2293) involved in fucose metabolism. However, since there is another gene (ecn0461) responsible for this function that clusters with other genes for fucose metabolism, the real

function of this paralog remains to be elucidated. It is worthwhile to experimentally test whether this gene is involved in the MTA recycle and whether the genomic peculiarity of *E. coli* Nissle 1917 with respect to the methionine cycle generally offers any advantages or specific properties to this strain in comparison to other *E. coli* isolates.

3.2.4. Mobile genetic elements

Most of the identified *E. coli* Nissle 1917 specific CDSs are associated with mobile elements, in particular with four genomic islands, indicating that *E. coli* Nissle 1917 acquired those genes probably by horizontal gene transfer. These four genomic islands, termed here as GI1, GI2, GI3 and GI4 for convenience, are different from the four islands GEI I–IV analyzed by Grozdanov et al. (2004), except that GI3 is partly covered by GEI II. All the four genomic islands are combined with transposases. GI1 which spans three contigs of the raw assembly contains three transposons at one end, and three transposons and genes for phage life cycle such as lysogenic conversion protein, resolvase and integrase. GI1 includes genes coding for a type II restriction-modification system (Bitinaite et al., 2002) (coded by ecn0116–ecn0120). These genes are not found in the other five *E. coli* strains compared, but found, for example, in the *E. coli* RFL31 strain (Bitinaite et al., 1992). Thirty genes (ecn3555–3584) in this island resemble genes from the pathogenicity island of *Shigella flexneri* 2a YSH6000 (Luck et al., 2001). They encode a ferric dicitrate transport system (Fec). Some of these genes are not found in the other five *E. coli* strains, indicating probably gene deletion events after horizontal gene transfer.

The genomic island GI2 is a bacteriophage characterized by integrase, excisionase and lysozyme and many bacteria phage related proteins. GI2 covers over 50 coding sequences (ecn0737–0764, ecn1527–1552). Twenty-nine of them have no homologues in the other five *E. coli* strains.

A part of GI3 has been discussed in Sections 3.2.1 and 3.2.2. The capsule biosynthesis cluster was also reported in a recent study as GEI II (Grozdanov et al., 2004) while the type II secretion cluster was not reported there. This island was thought to be horizontally acquired and inserted to the tRNA gene *serX*.

Similarly, the genomic island GI4 is inserted downstream of another tRNA gene—*selC*. Many

pathogenicity islands were demonstrated to be integrated at the *selC* locus, as shown for the SHI-2 island of *S. flexneri*, the SPI-3 pathogenicity island of *Salmonella enterica*, the PAI-I island of UPEC strains and the *stx* island of a subgroup of Shiga toxin (*stx*)-producing *E. coli* strains (Schmidt et al., 2001). The GI4 island, flanking with an integrase, encloses at least 17 genes (ecn4407–4423), 12 of which are *E. coli* Nissle 1917 specific. The gene ecn4408, 86% identical to Q9Z628 of *Salmonella typhimurium*, contains an imperfect nucleotide-binding Walker A motif. The genes ecn4409 and 4410 are 59% and 78% identical to P77692 and Q8FG62 of *E. coli*, respectively, which encode for putative toxin–antitoxin pairs (Brown and Shaw, 2003). All other genes in this island are similar to hypothetical proteins of the database so that the function of this island remains unclear.

3.3. Virulence factors

Three thousand two hundred and thirty-nine virulence factors (proteins) from Swissprot-TrEMBL database were queried in the genomes of the six *E. coli* strains to examine their existence. These virulence factors, covering 170 signatures defined by the PRINTS protein fingerprint database, are classified into eight categories: adherence/colonization factors, invasins, cell surface factors, exotoxins, trans-

porters, siderophores, miscellaneous and non-bacterial virulence factors (<http://www.jenner.ac.uk/bacbix3/ppprints.htm>). Two hundred and fifty virulence factors (e -value $\leq 1E-20$) covering 97 signatures were found to be present in at least one *E. coli* strain (see Supplementary Table 1). About 70% of them were found to be common in all the six *E. coli* strains. The existence of a high number of common virulence factors in these strains and the fact that some of the strains are nonpathogenic suggest that these proteins cannot be considered as virulence factors per se. Many of them actually function for general fitness or adaptation purpose rather than as a unique virulence factor.

Surprisingly, the comparison of the known “virulence factors” between the probiotic *E. coli* Nissle 1917 strain and the UPEC isolate CFT073 reveals very little difference (Table 4). Beyond the about 130 common virulence factors, *E. coli* Nissle 1917 has only three specific “virulence determinants”: a prepilin peptidase A and two proteins from a type II secretion system as mentioned above. *E. coli* Nissle 1917 does not possess 12 “virulence factors” that are present in the CFT073 strain, including proteins associated with the P fimbriae (*pili*), PapE and PapF, and proteins associated with alpha-hemolysin, HlyA, C and D. UPEC can attach with the help of P fimbriae to digalactoside receptors that are expressed on the kidney epithelium and then damage the renal epithelium by secreted

Table 4
Specific “virulence factors” in *E. coli* Nissle 1917 and CFT073 identified from the Swissprot-TrEMBL virulence factor protein database

“Virulence factors”	Functions	Virulence categories
Specific for <i>E. coli</i> Nissle 1917		
ecn:ECn3720	Putative prepilin peptidase A	Adherence/colonization factors
ecn:ECn3718	Hypothetical type II secretion protein GspC	Transporters
ecn:ECn3717	Hypothetical type II secretion protein GspD	Transporters
Specific for <i>E. coli</i> CFT073		
ecc:c3585	PapE	Adherence/colonization factors
ecc:c3584	PapF	Adherence/colonization factors
ecc:c3153	Putative outer membrane protein of prophage	Cell surface factors
ecc:c0363	Putative RTX family exoprotein A gene	Cell surface factors
ecc:c2560	UDP-glucose 4-epimerase [EC:5.1.3.2]	Cell surface factors
ecc:c0821	Pal; peptidoglycan-associated lipoprotein precursor	Cell surface factors
ecc:c3570	HlyA; hemolysin A	Exotoxins and transporters
ecc:c3569	HlyC; hemolysin C	Transporters
ecc:c3574	HlyD; hemolysin D	Transporters
ecc:c4786	Sec-independent protein translocase protein TatB	Transporters
ecc:c2294	Flagellar biosynthetic protein FlhB	Transporters
ecc:c4957	HupA; DNA-binding protein HU-alpha	Miscellaneous

Table 5

Enzymes that are absent in at least one *E. coli* strain

EC	ecn	ecc	eco	ecj	ecs	ece	Enzyme	Metabolism
2.7.1.4	n ^a	n	n	n	y ^b	y	D-Fructokinase	Fructose and mannose metabolism
3.2.1.26	n	n	n	n	y	y	Beta-fructofuranosidase (invertase)	Galactose, starch and sucrose metabolism
3.5.1.5	n	n	n	n	y	y	Urease	Purine metabolism
1.1.1.133	n	n	y	y	n	n	Phosphoribosylglycinamide formyltransferase	Nucleotide sugars metabolism
1.1.1.157	n	n	y	y	n	n	3-Hydroxybutyryl-CoA dehydrogenase	Butanoate metabolism
1.2.1.39	n	n	y	y	n	n	Phenylacetaldehyde dehydrogenase	Phenylalanine metabolism
1.4.3.6	n	n	y	y	n	n	Amine oxidase (copper-containing)	Arginine, proline, phenylalanine and tryptophan metabolism
2.1.1.10	n	n	y	y	n	n	Homocysteine S-methyltransferase	Methionine metabolism
3.2.1.37	n	n	y	y	n	n	Xylan 1,4-beta-xylosidase	Nucleotide sugars, starch and sucrose metabolism
5.1.3.13	n	n	y	y	n	n	dTDP-4-dehydrorhamnose 3,5-epimerase	Nucleotide sugars metabolism
6.2.1.30	n	n	y	y	n	n	Phenylacetate-CoA ligase	Phenylalanine metabolism
1.1.1.93	n	n	y	y	y	y	Tartrate dehydrogenase	Glyoxylate and dicarboxylate metabolism
1.13.11.–	n	n	y	y	y	y	2,3-Dihydroxyphenylpropionate 1,2-dioxygenase	Tryptophan metabolism
1.18.1.3	n	n	y	y	y	y	Ferredoxin-NAD ⁺ reductase	Fatty acid metabolism
1.2.1.3	n	n	y	y	y	y	aldH; putative aldehyde dehydrogenase	Glycolysis/gluconeogenesis/pyruvate/propanoate/butanoate/ascorbate and aldarate/fatty acid/Try, Arg, Pro metabolism and Val, Leu and Ile degradation
1.3.1.–	n	n	y	y	y	y	2,3-Dihydroxy-2,3-dihydroxyphenylpropionate dehydrogenase	Propanoate metabolism and Fatty acid biosynthesis
1.3.1.–	n	n	y	y	y	y	2,3-Dihydroxy-2,3-dihydroxyphenylpropionate dehydrogenase	Tryptophan metabolism
3.7.1.–	n	n	y	y	y	y	2-Hydroxy-6-keptonona-2,4-dienedioic acid hydrolase	Ascorbate, aldarate and phenylalanine metabolism
1.2.1.10	n	n	y	y	y	y	Acetaldehyde dehydrogenase	Pyruvate/butanoate metabolism
5.4.99.2	p ^c	p	y	y	y	y	Methylmalonyl-CoA mutase	Propanoate metabolism and Val, Leu and Ile degradation
4.1.2.21	f ^d	f	y	y	n	n	2-Dehydro-3-deoxy-6-phosphogalactonate aldolase	Galactose metabolism
4.2.1.6	f	f	y	y	n	n	Galactonate dehydratase	Galactose metabolism and pentose and glucuronate interconversions
2.3.1.12	y	f	y	y	y	y	Dihydrolipoamide acetyltransferase component (E2) of pyruvate dehydrogenase complex	Glycolysis, gluconeogenesis and pyruvate metabolism
2.7.1.51	y	f	y	y	y	y	FucK; L-fuculokinase	Fructose and mannose metabolism
2.8.3.1	y	y	n	n	n	n	Propionate CoA-transferase	Pyruvate and propanoate metabolism
3.5.3.6	y	y	n	n	n	n	Arginine deiminase	Arginine and proline metabolism
1.1.1.31	y	y	y	y	n	n	3-Hydroxyisobutyrate dehydrogenase	Val, Leu and Ile degradation
1.1.3.15	y	y	y	y	n	n	(S)-2-Hydroxy-acid oxidase	Glyoxylate and dicarboxylate metabolism
2.7.1.53	y	y	y	y	n	n	L-Xylulokinase	Pentose and glucuronate interconversions
2.7.1.58	y	y	y	y	n	n	2-Dehydro-3-deoxygalactonokinase	Galactose metabolism
2.8.3.8	y	y	y	y	n	n	Acetate CoA-transferase	Propanoate and butanoate metabolism
3.2.1.14	y	y	y	y	n	n	Chitinase	Aminosugars metabolism

^a n: does not exist.^b y: exists.^c p: partial gene.^d f: frame shift in the gene.

hemolysin (Kaper et al., 2004). Although *E. coli* Nissle 1917 shares most of the “virulence determinants” with the UPEC strain CFT073, the absence of pathogenicity factors essential for renal pathogenesis, will obviously exclude the possibility that *E. coli* Nissle 1917 becomes a uropathogen as its close relative *E. coli* CFT073 does. The absence of intact *pap* and *hly* was also reported by Grozdanov et al. (2004). This could contribute to this strain’s probiotic nature.

3.4. Metabolic network reconstruction, visualization and comparison

The reconstruction and analysis of metabolic network shows that the metabolic capacity of *E. coli* Nissle 1917 is close to that of the *E. coli* K-12 and O157:H7 strains (Refer to Table 5 and the Supplementary data). They all have the fully functional pathways for glycolysis, TCA cycle, pentose phosphate cycle, sulfur reduction and fixation, etc. They all have the ability to de novo synthesize fatty acids, lipopolysaccharide, all essential nucleotide, all essential amino acids, many vitamins such as riboflavin, nicotinate, pantothenate and CoA, biotin, folate, sideroheme, ubiquinone and so on.

Merely 33 enzymes from the KEGG metabolic maps were found to have different availability in the six *E. coli* strains compared (Table 5). The metabolic network of *E. coli* K-12 strain MG1655 is completely the same as the one of K-12 isolate W3110 while the one of EHEC O157:H7 Sakai is completely the same as that of EHEC O157:H7 EDL933. The metabolic potential of *E. coli* Nissle 1917 resembles mostly that of *E. coli* CFT073. The two *E. coli* K-12 strains have most of those 33 enzymes, indicating these two strains should have a higher metabolic capacity than the other strains compared. Of the enzymes that are absent in the *E. coli* K-12 strains, D-fructokinase (EC 2.7.1.4), beta-fructofuranosidase (EC 3.2.1.26) and urease (EC 3.5.1.5) are merely found in the O157:H7 strains whereas propionate CoA-transferase (EC 2.8.3.1, for pyruvate and propanoate metabolism) and arginine deiminase merely in *E. coli* Nissle 1917 and CFT073. Arginine deiminase (ADI, EC 3.5.3.6) is the first step of the arginine degradation pathway via arginine deiminase, which hydrolyzes arginine to generate energy in many parasitic microorganisms (Zuniga et al., 2002). ADI was reported to have potent anticancer activities

and to be able to halt growth of solid tumors (Park et al., 2003).

4. Concluding remarks

The analysis of the raw genome data of *E. coli* Nissle 1917 as presented above clearly showed a number of genomic peculiarities and significant differences in protein-coding sequences, metabolic potential and possession of possible virulence factors in comparison with other five sequenced *E. coli* strains. Among the 108 *E. coli* Nissle 1917 specific coding sequences identified, we discussed in some more detail about those CDSs that are involved in capsule biosynthesis, secretion apparatus, methionine cycle and mobile genetic elements. To obtain possible links of these genomic peculiarities to the probiotic nature of this strain the function and physiological roles of these and other *E. coli* Nissle 1917 specific CDSs need to be experimentally studied. This in silico analysis of raw genome data has generated useful hypotheses for downstream experiments before sequencing and annotation of the genome under investigation have been completely finished.

Acknowledgements

J.S. and A.-P.Z. were supported by the Ministry for Education and Research (BMBF) through the Braunschweig Bioinformatic Competence Center project “Intergenomics” (Grant No. 031U110A). F.G. and J.B. were supported by the Deutsche Forschungsgemeinschaft, SFB 621.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jbiotec.2005.01.008.

References

- Ali, A., Johnson, J.A., Franco, A.A., Metzger, D.J., Connell, T.D., Morris Jr., J.G., Sozhamannan, S., 2000. Mutations in the ex-

- tracellular protein secretion pathway genes (*eps*) interfere with rugose polysaccharide production in and motility of *Vibrio cholerae*. *Infect. Immun.* 68, 1967–1974.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Bitinaite, J., Maneliene, Z., Menkevicius, S., Klimasauskas, S., Butkus, V., Janulaitis, A., 1992. Alw26I, Eco31I and Esp3I—type II methyltransferases modifying cytosine and adenine in complementary strands of the target DNA. *Nucleic Acids Res.* 20, 4981–4985.
- Bitinaite, J., Mitkaite, G., Dauksaite, V., Jakubauskas, A., Timinskas, A., Vaisvila, R., Lubys, A., Janulaitis, A., 2002. Evolutionary relationship of Alw26I, Eco31I and Esp3I, restriction endonucleases that recognise overlapping sequences. *Mol. Genet. Genomics* 267, 664–672.
- Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1474.
- Blum, G., Marre, R., Hacker, J., 1995. Properties of *Escherichia coli* strains of serotype O6. *Infection* 23, 234–236.
- Brown, J.M., Shaw, K.J., 2003. A novel family of *Escherichia coli* toxin–antitoxin gene pairs. *J. Bacteriol.* 185, 6600–6608.
- Francetic, O., Belin, D., Badaut, C., Pugsley, A.P., 2000. Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion. *EMBO J.* 19, 6697–6703.
- Geisse, A., 1919. Behandlung infektiöser Darmerkrankungen mit “Mutaflo”. *Ther. d. Gegenwart* 60, 90–93.
- Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., Dobrindt, U., 1917. Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle. *J. Bacteriol.* 186, 5432–5441.
- Grozdanov, L., Zahringer, U., Blum-Oehler, G., Brade, L., Henne, A., Knirel, Y.A., Schombel, U., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., Rietschel, E.T., Dobrindt, U., 1917. A single nucleotide exchange in the *wzy* gene is responsible for the semirough O6 lipopolysaccharide phenotype and serum sensitivity of *Escherichia coli* strain Nissle. *J. Bacteriol.* 184, 5912–5925.
- Gunzer, F., Hennig-Pauka, I., Waldmann, K.H., Sandhoff, R., Grone, H.J., Kreipe, H.H., Matussek, A., Mengel, M., 2002. Gnotobiotic piglets develop thrombotic microangiopathy after oral infection with enterohemorrhagic *Escherichia coli*. *Am. J. Clin. Pathol.* 118, 364–375.
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., Shinagawa, H., 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8, 11–22.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., Hattori, M., 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280 (database issue).
- Kaper, J.B., Nataro, J.P., Mobley, H.L., 2004. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140.
- Kruis, W., Fric, P., Pokrotnieks, J., Lukas, M., Fixa, B., Kascak, M., Kamm, M.A., Weismueller, J., Beglinger, C., Stolte, M., Wolff, C., Schulze, J., 2004. Maintaining remission of ulcerative colitis with the probiotic *Escherichia coli* Nissle 1917 is as effective as with standard mesalazine. *Gut* 53, 1617–1623.
- Lodinova-Zadnikova, R.U., Sonnenborn, U., 1997. Effect of preventive administration of a nonpathogenic *Escherichia coli* strain on the colonization of the intestine with microbial pathogens in newborn infants. *Biol. Neonate* 71, 224–232.
- Luck, S.N., Turner, S.A., Rajakumar, K., Sakellaris, H., Adler, B., 2001. Ferric dicitrate transport system (Fec) of *Shigella flexneri* 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes. *Infect. Immun.* 69, 6012–6021.
- Malchow, H.A., 1997. Crohn’s disease and *Escherichia coli*. A new approach in therapy to maintain remission of colonic Crohn’s disease? *J. Clin. Gastroenterol.* 25, 653–658.
- Mead, D.A., Szczesna-Skorupa, E., Kemper, B., 1986. Single-stranded DNA ‘blue’ T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Protein Eng.* 1, 67–74.
- Nesper, J., Hill, C.M., Paiment, A., Harauz, G., Beis, K., Naismith, J.H., Whitfield, C., 2003. Translocation of group 1 capsular polysaccharide in *Escherichia coli* serotype K30. Structural and functional analysis of the outer membrane lipoprotein Wza. *J. Biol. Chem.* 278, 49763–49772.
- Nissle, A., 1916. Über die Grundlagen einer neuen ursächlichen Bekämpfung der pathologischen Darmflora. *Dtsch. Med. Wochenschr.* 42, 1181–1184.
- Nissle, A., 1919. Weiteres über die Mutafloerbehandlung unter besonderer Berücksichtigung der chronischen Ruhr. *Münch. Med. Wschr.* 25, 678–681.
- Nissle, A., 1925. Weiteres über die Grundlagen und Praxis der Mutafloerbehandlung. *Dtsch. Med. Wochenschr.* 44, 1809–1813.
- Norrandner, J., Kempe, T., Messing, J., 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26, 101–106.
- Park, I.S., Kang, S.W., Shin, Y.J., Chae, K.Y., Park, M.O., Kim, M.Y., Wheatley, D.N., Min, B.H., 2003. Arginine deiminase: a potential inhibitor of angiogenesis and tumour growth. *Br. J. Cancer* 89, 907–914.
- Patzner, S.I., Baquero, M.R., Bravo, D., Moreno, F., Hantke, K., 2003. The colicin G, H and X determinants encode microcins M and H47, which might utilize the catechol siderophore receptors FepA, Cir, Fiu and IroN. *Microbiology* 149, 2557–2570.
- Pazzani, C., Rosenow, C., Boulnois, G.J., Bronner, D., Jann, K., Roberts, I.S., 1993. Molecular analysis of region 1 of the *Escherichia coli* K5 antigen gene cluster: a region encoding proteins involved in cell surface expression of capsular polysaccharide. *J. Bacteriol.* 175, 5978–5983.
- Perna, N.T., Plunkett III, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamouisis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen,

- G., Schwartz, D.C., Welch, R.A., Blattner, F.R., 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529–533.
- Rembacken, B.J., Snelling, A.M., Hawkey, P.M., Chalmers, D.M., Axon, A.T., 1999. Non-pathogenic *Escherichia coli* versus mesalazine for the treatment of ulcerative colitis: a randomised trial. *Lancet* 354, 635–639.
- Sandkvist, M., 2001. Type II secretion and pathogenesis. *Infect. Immun.* 69, 3523–3535.
- Schmidt, H., Zhang, W.L., Hemmrich, U., Jelacic, S., Brunder, W., Tarr, P.I., Dobrindt, U., Hacker, J., Karch, H., 2001. Identification and characterization of a novel genomic island integrated at *selC* in locus of enterocyte effacement-negative, Shiga toxin-producing *Escherichia coli*. *Infect. Immun.* 69, 6863–6873.
- Schroeder, H.R., Barnes, C.J., Bohinski, R.C., Mallette, M.F., 1973. Biological production of 5-methylthioribose. *Can. J. Microbiol.* 19, 1347–1354.
- Sekowska, A., Denervaud, V., Ashida, H., Michoud, K., Haas, D., Yokota, A., Danchin, A., 2004. Bacterial variations on the methionine salvage pathway. *BMC Microbiol.* 4, 9.
- Sun, J., Zeng, A.P., 2004. IdentiCS—identification of coding sequence and in silico reconstruction of the metabolic network directly from unannotated low-coverage bacterial genome sequence. *BMC Bioinformatics* 5, 112.
- Tauschek, M., Gorrell, R.J., Strugnell, R.A., Robins-Browne, R.M., 2002. Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7066–7071.
- Waidmann, M., Bechtold, O., Frick, J.S., Lehr, H.A., Schubert, S., Dobrindt, U., Loeffler, J., Bohn, E., Autenrieth, I.B., 2003. *Bacteroides vulgatus* protects against *Escherichia coli*-induced colitis in gnotobiotic interleukin-2-deficient mice. *Gastroenterology* 125, 162–177.
- Welch, R.A., Burland, V., Plunkett III, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L., Donnenberg, M.S., Blattner, F.R., 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17020–17024.
- Westendorf, A.M., Gunzer, F., Deppenmeier, S., Hunger, J.K., Schmidt, M.A., Buer, J., Bruder, D., 2005. Intestinal immunity of *E. coli* NISSLE 1917: a safe carrier for therapeutic molecules. *FEMS Immunol. Med. Microbiol.* 43, 373–384.
- Whitfield, C., Roberts, I.S., 1999. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol. Microbiol.* 31, 1307–1319.
- Yamamoto, Y., Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kimura, S., Kitagawa, M., Makino, K., Miki, T., Mitsuhashi, N., Mizobuchi, K., Mori, H., Nakade, S., Nakamura, Y., Nashimoto, H., Oshima, T., Oyama, S., Saito, N., Sampei, G., Satoh, Y., Sivasundaram, S., Tagami, H., Horiuchi, T., 1997. Construction of a contiguous 874-kb sequence of the *Escherichia coli*-K12 genome corresponding to 50.0–68.8 min on the linkage map and analysis of its sequence features. *DNA Res.* 4, 91–113.
- Zuniga, M., Perez, G., Gonzalez-Candelas, F., 2002. Evolution of arginine deiminase (ADI) pathway genes. *Mol. Phylogenet. E* 25, 429–444.