Comparative genomic hybridization and physiological characterization of environmental isolates indicate that significant (eco-)physiological properties are highly conserved in the species *Escherichia coli*

Julian Ihssen,^{1,2} Elena Grasselli,^{3,4} Claudio Bassin,¹ Patrice François,⁵ Jean-Claude Piffaretti,³ Wolfgang Köster,^{1,6} Jacques Schrenzel⁵ and Thomas Egli¹

Correspondence
Thomas Egli
egli@eawag.ch

¹Swiss Federal Institute for Environmental Science and Technology, PO Box 611, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland

²Empa, Swiss Federal Institute for Materials Testing and Technology, Lerchenfeldstrasse 5, CH-9014 St Gallen, Switzerland

³Istituto Cantonale di Microbiologia, Via Mirasole 22A, CH-6500 Bellinzona, Switzerland

⁴Department of Biology, University of Genova, Corso Europa 26 V piano, 16132 Genova, Italy

⁵Genomic Research Laboratory, University Hospitals of Geneva, rue Micheli-du-Crest 24, CH-1211 Geneva 14, Switzerland

⁶VIDO – Vaccine & Infectious Diseases Organization, University of Saskatchewan, 120 Veterinary Road, Saskatoon, SK S7N 5E3, Canada

Escherichia coli, the common inhabitant of the mammalian intestine, exhibits considerable intraspecies genomic variation, which has been suggested to reflect adaptation to different ecological niches. Also, regulatory trade-offs, e.g. between catabolic versatility and stress protection, are thought to result in significant physiological differences between strains. For these reasons, the relevance of experimental observations made for 'domesticated' *E. coli* strains with regard to the behaviour of this species in its natural environments is often questioned and doubts are frequently raised on the status of *E. coli* as a defined species. The variability of important (eco-)physiological functions, such as carbon substrate uptake and breakdown capabilities, as well as stress defence mechanisms, in the genomes of commensal and pathogenic *E. coli* strains were therefore investigated. Furthermore, (eco-)physiological properties of environmental strains were compared to standard laboratory strain K-12 MG1655. Catabolic, stress protection, and carbon- and energy source transport operons showed a very low intraspecies variability in 57 commensal and pathogenic *E. coli*. Environmental isolates adapted to glucose-limited growth in a similar way as *E. coli* MG1655, namely by increasing their catabolic flexibility and by inducing high-affinity substrate uptake systems. The results obtained indicate that significant (eco-)physiological properties are highly conserved in the natural population of *E. coli*. This questions the proposed dominant role of horizontal gene transfer for niche adaptation.

Received 31 August 2006
Revised 26 March 2007
Accepted 3 April 2007

Abbreviations: CGH, comparative genomic hybridization; HGT, horizontal gene transfer; RSW, raw source water.

Microarray data, design and oligonucleotide sequences GEO DataSets accession nos are GSE6407, GSE6486 and GPL4618.

The raw DNA microarray data of this and other studies and the genomic DNA labelling protocol are available with the online version of this paper.

INTRODUCTION

In recent years, due to the discovery of widespread horizontal gene transfer (HGT) in the domain eubacteria, the applicability of the classical species concept and the extent of linear evolution have been questioned for this group of organisms (Doolittle, 1999; Lan & Reeves, 2000; Ochman *et al.*, 2000). In fact, natural isolates of *Escherichia coli* vary in their chromosome length between 4500 and

5500 kb (Bergthorsson & Ochman, 1998), which is thought to mainly result from incorporation of foreign DNA and gene deletion events (Dobrindt *et al.*, 2003). The standard laboratory *E. coli* K-12 strain MG1655 is estimated to contain up to 18% of horizontally acquired DNA (Lawrence & Ochman, 1998) and only 70% of all predicted MG1655 proteins are shared with both of the fully sequenced pathogenic *E. coli* strains, CFT073 and EDL933 (Welch *et al.*, 2002). Phenotypic variations in LPS structure, fimbriae, capsular antigens, utilization of certain sugars, antibiotic resistance and more are common to natural isolates of *E. coli* (Selander *et al.*, 1996) and are suggested to reflect adaptations to specific hosts or even to different ecological niches (Reeves, 1992; Dobrindt *et al.*, 2004). In the case of various *E. coli* pathotypes, horizontally acquired genomic regions (pathogenicity islands) carry the crucial traits for leading a parasitic life-style (Hacker & Kaper, 2000; Perna *et al.*, 2001; Welch *et al.*, 2002).

This contrasts with numerous physiological characteristics not involved in pathogenicity, which are highly conserved in all or at least in an overwhelming majority of *E. coli* strains, e.g. lactose, maltose and galactose utilization, β -glucuronidase activity, resistance to bile salts, ability to grow at 44 °C, mixed acid fermentation, etc. (Bettelheim, 1992). Indeed, these common characteristics allow a reliable identification of *E. coli* by classical means (Holt, 1994). The large number of shared phenotypic traits is consistent with a large amount of shared DNA sequences constituting the so-called core (or backbone) genome of *E. coli* (Welch *et al.*, 2002; Anjum *et al.*, 2003; Dobrindt *et al.*, 2003).

The question of how conserved or how variable physiological properties and other characteristics of *E. coli* are is further complicated by the rapid emergence of sequence polymorphisms in laboratory evolution experiments. These genetic changes can lead to drastically enhanced fitness (K_s) under glucose-limited growth conditions (Senn *et al.*, 1994; Notley-McRobb & Ferenci, 1999a, b; Wick *et al.*, 2001), but also to a decrease in stress resistance (Chen *et al.*, 2004; Notley-McRobb *et al.*, 2002; Zinser & Kolter, 1999) or to an inactivation of catabolic functions (Cooper & Lenski, 2000).

The normal life cycle of *E. coli* consists of two phases: (i) growth in the primary habitat, the mammalian large intestine, and (ii) survival in the secondary habitat, water and soil, until a new host can be colonized (Savageau, 1983). Both habitats are predominantly carbon- and energy-limited and specific growth rates fluctuate between a small fraction of $\mu_{\max}$ to negative (die-off) for most of the time (Freter *et al.*, 1983a; Savageau, 1983; Macfarlane & Macfarlane, 1997). *E. coli* MG1655 adapts to carbon and energy limitation by complex physiological changes that confer a high degree of catabolic flexibility, the capacity for mixed substrate growth and efficient carbon substrate uptake abilities to the cells (Wick *et al.*, 2001; Ihssen & Egli, 2005; Liu *et al.*, 2005). At the same time, slow-growing or

growth-arrested *E. coli* protect themselves against oxidative cellular deterioration and environmental stresses by expressing various protective proteins, many of which are regulated by the alternative sigma factor RpoS (reviewed by Hengge-Aronis, 1996; Matin, 2004; Nyström, 2004).

If mixed-substrate growth, high-affinity transport and stress defence functions are crucial fitness determinants for the species *E. coli* in its primary and secondary habitats, they should not be subject to variation among strains, i.e. they should be part of the core genome. Therefore, we analysed four sets of comparative genomic hybridization (CGH) data for the presence and absence of known MG1655: (i) catabolic operons, (ii) operons encoding binding protein-dependent ATP-driven transporters or phosphotransferase systems for carbon substrates, and (iii) stress defence operons. In addition, intraspecies variability of the physiological adaptation to carbon- and energy-limited growth and starvation was determined by comparing growth kinetic parameters, RpoS phenotype, derepression of catabolic pathways and expression of high-affinity binding proteins of six environmental *E. coli* isolates with MG1655. Both CGH and physiological results indicate that numerous ecologically relevant properties of *E. coli* are highly conserved and may result from a long-term adaptation of this species to its current niche.

METHODS

Isolation and phenotypic characterization of environmental and faecal *E. coli* strains. The origin, ECOR group and antibiotic resistance of *E. coli* strains analysed in this study are listed in Table 1. Six strains originated from raw source water (RSW) of drinking water catchments in northern Switzerland (WK1–WK8). Based on phenotypic differences (Table 2), they were selected from a total of 45 strains isolated during routine screening. Standard *E. coli* diagnostic agar plates were used for screening and isolation. The 19 faecal strains (Table 1) were isolated from stool samples of healthy adult humans and domestic animals in southern and south-western Switzerland.

The ECOR group of the environmental and faecal isolates was determined by triplex PCR using primers for DNA fragments of *chuA*, *yjaA* and TSPE4.C2 (Clermont *et al.*, 2000). Antibiotic resistance was tested on LB agar plates amended with antibiotics and confirmed for WK1–WK8 in liquid culture. Faecal strains were isolated from healthy human and animal hosts from Geneva and Ticino (south western and south eastern Switzerland), respectively. Antibiotic resistance determination for these strains was performed by the disk diffusion method. Transfers of isolates on laboratory growth media were generally kept as low as possible (maximally three transfers on complex medium agar before storage at –80 °C). Siderophore excretion of the RSW strains was evaluated according to the size and intensity of orange haloes formed around the bacterial colonies grown on blue agar plates containing chrome azurol S (CAS plates) after incubation for 12 h at 37 °C (Schwyn & Neilands, 1987). For analysis of colicin production, the strains were streaked onto LB plates. After incubation for 12 h at 37 °C, the cells were killed by treatment with chloroform vapour for 30 min and then exposed to air for 1 h. The plates were then overlaid with LB soft agar (0.5% agar; Difco) containing a sensitive *E. coli* K-12 strain at 10^6 c.f.u. ml^{–1} (Ölschläger *et al.*, 1984). After incubation for 12 h at 37 °C, growth inhibition zones in the lawn of the indicator strain were analysed. For investigation of haemolysin production, strains were streaked onto

Table 1. Origin, ECOR group and antibiotic resistance of *E. coli* isolates analysed in this study

Amp, ampicillin; Str, streptomycin; Tet, tetracycline; Gnt, gentamycin; Sfa, sulphonamide.

Subgroup	Strain	ECOR group	Origin	Swiss region	Antibiotic resistance
Environmental	WK1	A	Raw source water	Northern Jura	Tet
	WK2	A	Raw source water	Upper Lake Zurich	None
	WK3	A	Raw source water	Linth valley (Glarus)	Amp, Str, Tet
	WK4	A	Raw source water	Linth valley (Glarus)	None
	WK5	A	Raw source water	Linth valley (Glarus)	None
	WK8	A	Raw source water	Limmat valley (Aargau)	Str, Tet
Faecal	520	B2	Human	Ticino	None
	521	D	Human	Ticino	None
	H01	A	Human	Ticino	None
	H02	A	Human	Ticino	None
	H03	B2	Human	Ticino	None
	H05	A	Human	Ticino	Amp, Str, Tet, Gnt, Sfa
	H04	A	Human	Geneva	None
	H12	A	Human	Geneva	None
	H23	D	Human	Unknown	Str, Tet, Sfa
	H27	A	Human	Ticino	Str (weak)
	F01	B2	Cat	Ticino	Sfa (weak)
	F02	B2	Cat	Ticino	None
	C01	B2	Dog	Unknown	None
	C02	B2	Dog	Unknown	None
	B02	D	Cattle	Ticino	None
	B05	A	Cattle	Ticino	None
	S04	A	Swine	Ticino	Str, Sfa
	S05	D	Swine	Ticino	None
	S13	A	Swine	Ticino	Str, Tet, Sfa

sheep blood agar plates and grown for 24–48 h at 37 °C as described in Beutin *et al.* (1989). Clearing zones around colonies were indicative of haemolysin production.

DNA microarrays. The genome composition of the RSW isolates was analysed by CGH using commercially available *E. coli* K-12 V2 microarrays (MWG Biotech). The arrays contained all ORFs of *E. coli* MG1655 (Blattner *et al.*, 1997). Microarrays were hybridized according to the manufacturer's instructions with fluorescently labelled genomic DNA of the reference strain MG1655 and of environmental strains. Total genomic DNA was extracted from stationary phase LB cultures with a genomic DNA purification kit (Fermentas), followed by random hexamer-primed labelling with Cy3-dCTP or Cy5-dCTP (Amersham Biosciences). In each reaction, 2 µg unshredded genomic DNA was labelled with the BioPrime labelling kit (Invitrogen) according to the protocol of DeRisi (Anjum

et al., 2003). After labelling, Cy3- and Cy5-cDNA were combined and co-purified on MinElute PCR purification columns (Qiagen). For each strain two arrays were hybridized with individually labelled DNA derived from two separate overnight LB cultures. The processed slides were scanned with an Affimetrix 428 array scanner. Fluorescent spots and local background intensities were quantified with Affimetrix Jaguar software version 2 and analysed further in Microsoft Excel. Very few spots had a Cy3 signal for the reference strain which was lower than the background signal intensity plus two standard deviations of background intensity. They were excluded from further analysis as suggested by Anjum *et al.* (2003). For each array, Cy5 intensities of all spots were normalized to the median of Cy3 intensities. This procedure resulted in Cy3: Cy5 ratios of approximately one for the majority of spots, which means that labelled target DNA from both the reference and the environmental strain hybridized. Similarly to what has been reported for signal intensity ratios of known gene deletions in laboratory strains (Anjum *et al.*, 2003), certain spots were clear positive outliers. Negative outliers (which would indicate missing genes in our MG1655 strain) were not detected. A cut-off value of 1.6 for logarithmic Cy3: Cy5 ratios (fivefold difference in signal intensity) was chosen as threshold for missing genes, because this value was more than two times the positive range of background variation (Anjum *et al.*, 2003, used a cut-off value of 2.0). Two arrays hybridized with genomic DNA of the same strain yielded similar patterns of missing ORFs, with only a few contradictory results; those were excluded from further analysis. Microarray data for the RSW isolates are available in the GEO online database (www.ncbi.nlm.nih.gov/projects/geo/index.cgi) with series accession no. GSE6407.

The 19 faecal strains were analysed using an in-house-designed oligoarray (Charbonnier *et al.*, 2005). A set of 2700 oligonucleotide

Table 2. Phenotypic differences of *E. coli* strains isolated from raw source water in Switzerland

Strain	Siderophore	Colicin	Haemolysin
WK1	++	+	–
WK2	+	–	(+)
WK3	+	+++	–
WK4	+	–	–
WK5	(+)	–	+
WK8	++++	+++	–

probes covering 63 % of the ORFs of the *E. coli* K-12 chromosome was synthesized using the SurePrint technology (Agilent Technologies). Further details of array design and oligonucleotide sequences can be found in the GEO DataSets under the accession no. GPL4618. In CGH experiments, Cy3-dCTP-labelled K-12 genomic DNA and Cy5-dCTP-labelled DNA of a test strain were heated to 95 °C for 2 min, and then hybridized for 17 h at 60 °C with rotation in a dedicated hybridization oven (Robbins Scientific). Stringent washes were then performed according to the manufacturer's instructions. Slides were dried under nitrogen flow and scanned using 100 % photon multiplier tube power for both wavelengths with the Agilent scanner. Fluorescence intensities were extracted using the Feature extraction software (version 6.1.1; Agilent). Local background-subtracted signals were corrected for unequal dye incorporation or unequal load of the labelled product. The algorithm consisted of a rank consistency filter and a curve fit using the default LOWESS (locally weighted linear regression) method. The background noise of each experiment was evaluated by computing the standard deviation of negative-control intensities. Features whose intensities were smaller than the standard deviation value of the negative controls were considered as inefficient hybridization and discarded from further analysis. For each spot, the logarithm of the ratio between the test channel and the control channel (log ratio) was computed. Computed log ratio values were further sorted into 150 bin categories and fitted with a Gaussian distribution curve, using the Levenberg–Marquardt algorithm. The presence probability of each oligonucleotide probe (EPP) was determined, as described by Kim's algorithm (Kim *et al.*, 2002). Thus, the probability of presence was estimated for each ORF, allowing assessment of the presence or the divergence of genes in relation to K-12 genes. Genes that were either absent or sufficiently different to be hindered from hybridizing to the related oligonucleotide probes were defined as divergent. The most divergent genes (458), which failed to be detected in 25 % of the tested strains, were considered here as missing. Microarray data for the faecal strains are available in the GEO DataSets with series accession no. GSE6486.

Analysis of CGH datasets. Four sets of CGH datasets where total genomic DNA of *E. coli* strains was hybridized to *E. coli* MG1655 arrays were analysed in this study: (i) 5 isolates from raw source water (RSW, Tables 1 and 2), (ii) 19 faecal isolates from human and animal hosts (Faecal, Table 1), (iii) 6 commensal strains plus the laboratory strain *E. coli* B (Dobbrindt *et al.*, 2003) and (iv) 26 pathogenic strains (Anjum *et al.*, 2003). The lists of missing and present MG1655 ORFs were analysed for genes known to encode catabolic, transport, stress protection and global regulatory functions. Substrate-specific catabolic and transport genes and stress defence genes were compiled from Lin (1996), McFall & Newman (1996), Berlyn (1998), Wick & Egli (2004) and the online databases KEGG (Kanehisa & Goto, 2000) and EcoCyc (Keseler *et al.*, 2005). The presence/absence CGH datasets are available as supplementary material with the online version of this paper.

Determination of growth kinetic parameters, growth substrate range and catabolic flexibility in glucose-limited chemostat cultures. Glucose-limited mineral media and cultivation conditions were described previously (Ihssen & Egli, 2004). Maximum specific growth rates in glucose mineral medium ($\mu_{\max, \text{Glc}}$) were determined in triplicate in magnetically stirred flasks at 37 °C. Transfers on LB and mineral medium were kept as few as possible. Glucose affinities ($K_{s, \text{Glc}}$) and growth yields for glucose were determined in carbon-limited chemostat cultures after 20 h of cultivation (for details see Ihssen & Egli, 2004). The growth substrate spectrum of environmental *E. coli* strains and K-12 MG1655 was analysed with Biolog AN MicroPlates, which contain 95 different carbon and energy sources. Plates were inoculated in triplicate with cells grown overnight on TSA agar plates and resuspended in carbon-source-free mineral medium to an OD₅₄₆ of 0.1. Microplates were incubated for 30 h at 37 °C

before A₄₆₀ was measured with a microplate reader. Wells with an A₄₆₀ above 0.4 were counted as growth-positive (background A₄₆₀ of control wells was between 0.13 and 0.17). Expression of alternative catabolic pathways in glucose batch and glucose-limited chemostat cultures was determined with a previously described Biolog respiration assay using chloramphenicol-inhibited cells (Ihssen & Egli, 2005). Wild-type *E. coli* K-12 MG 1655 used as reference laboratory strain was the same as described previously (Wick *et al.*, 2001).

Hydroperoxidase assay, analysis of periplasmic proteins and RpoS-immunoblot. RpoS-dependent hydroperoxidase II specific activities were determined spectrophotometrically with crude cell extracts (Ihssen & Egli, 2004; Visick & Clarke, 1997). Periplasmic proteins were extracted by chloroform shock treatment (Ames *et al.*, 1984) and analysed by SDS-PAGE. For details of sample preparation, electrophoresis and protein identification see Ihssen & Egli (2005). RpoS protein levels were determined by immunoblot analysis (Ihssen & Egli, 2004) after extraction of total cellular proteins by TCA precipitation (Lange & Hengge-Aronis, 1994). Immunoblot loading volumes were normalized to the respective average band intensity of total proteins on Coomassie-stained control SDS-PAGE gels. Relative band intensities on SDS-PAGE gels and immunoblots were quantified with the publicly available software ImageJ (<http://rsb.info.nih.gov/ij/download.html>).

RESULTS

CGH analysis of catabolic and transport functions for carbon substrates

Four CGH datasets, encompassing in total 57 commensal and pathogenic *E. coli* strains, were analysed for operons encoding known catabolic and transport functions for 55 carbon substrates. None of the genes analysed is on the current list of essential genes for *E. coli* (<http://shigen.lab.nig.ac.jp/ecoli/pec/index.jsp>; Hashimoto *et al.*, 2005), i.e. they can hypothetically be subject to deletions or truncations by genome rearrangements and horizontal gene transfer (HGT). However, the incidence of missing MG1655 catabolic genes was very low both in commensal and pathogenic strains (Tables 3 and 4). Only the *atoCDAEB* operon, required for the degradation of C₄ fatty acids, and the glutamate-catabolism-related hypothetical regulatory protein *gluT* were missing consistently in a high percentage of strains. There were no systematic differences in the genome content of MG1655 catabolic and stress defence genes between environmental, faecal and pathogenic strains, and we also found no systematic differences between strains of different ECOR groups. In general, commensal strains predominantly belong to ECOR groups A and B1, while pathogenic strains are found mostly in ECOR groups B2 and D (Clermont *et al.*, 2000; Dobbrindt *et al.*, 2003). The strains isolated from RSW were all classified as ECOR group A (Table 1) and therefore are most likely commensals. The majority of faecal isolates from humans also belonged to group A (Table 1). In contrast, 67 % of isolates from cats, dogs, cattle and swine fell into the pathogenic groups B2 and D (Table 1).

In Biolog growth assays, 36 substrates were used by MG1655 and all RSW strains as sole source of carbon and

Table 3. Percentage of missing transport and catabolic operons in four sets of CGH data

The listed carbon substrates supported growth of all RSW isolates and *E. coli* K-12 MG1655. All genes analysed are non-essential. RSW, raw source water isolates; Faecal, strains isolated from humans, dogs, cattle, rabbits and swine (Table 1); Commensal, CGH data for six commensal *E. coli* strains and the laboratory strain *E. coli* B from Dobrindt *et al.* (2003); PEC, CGH data for 26 pathogenic *E. coli* strains from Anjum *et al.* (2003).

Growth substrate	MG1655 transport and/ or dissimilatory genes	Percentage of strains in which missing			
		RSW <i>n</i> =5*	Faecal <i>n</i> =19	Commensal <i>n</i> =7	PEC <i>n</i> =26
Sugars					
α -D-Glucose	<i>ptsG, ptsH, crr†</i>	0	0	0	0
Glucose-1 (or 6)- phosphate	<i>uhpTCBA</i>	0	0‡	0	0
Dextrin	<i>malS, amyA</i>	0	0	0	0
D-Fructose	<i>fruAKB†</i>	0	0	0	0
L-Fucose	<i>fucOAPIKUR</i>	0	5.3 <i>fucPI</i>	0	3.8 <i>fucP</i>
D-Galactose	<i>mgICABgalS\$, galMKTE, galR</i>	0	0‡	42.9 <i>mgIA</i>	0
α -D-Lactose, methyl β -D-galactoside	<i>lacAYZI</i>	0	0	0	2
D-Mannose	<i>manA, manXYZ†</i>	0	0	0	0
Maltose, maltotriose	<i>malS, malGFE\$, malKmalM</i>	0	0	0	0
	<i>malQPT</i>	0	10.5 <i>malQ</i>	71.4 <i>malP</i>	0
	<i>lamB</i>	0	0	14.3	0
D-Melibiose	<i>melRAV</i>	0	21.1 <i>melB</i>	0	0
D-Rhamnose	<i>rhaDABSRT</i>	0	5.3 <i>rhaD</i>	0	7.6
D-Trehalose	<i>treCBR†, treA, treF</i>	0	0	0	0
N-Acetyl-D-glucosamine	<i>nagDCABE†</i>	0	0	0	0
Sugar alcohols					
Glycerol	<i>glpXKF</i>	0	0	0	0
	<i>glpAB</i>	0	5.3 <i>glpA</i>	0	0
D,L- α -Glycerolphosphate	<i>glpQTABC</i>	0	5.3 <i>glpQA</i>	0	0
	<i>glpRGED</i>	0	0‡	14.3 <i>glpE</i>	3.8 <i>glpGED</i>
	<i>ugpCEAB\$</i>	0	0‡	0	0
D-Mannitol	<i>mtlADR†</i>	0	5.3 <i>mtlDR</i>	0	0
D-Sorbitol	<i>srlAEBDgutMsrlRgutQ†</i>	0	5.3 <i>srlAE</i>	14.3 <i>srlB</i>	7.6
Amino acids and dipeptides					
L-Alanine	<i>cycA, dadAX</i>	0	0	0	0
L-Alanyl-L-glutamine	<i>dppFDCBA\$</i>	0	0‡	0	0
+ L-Alanyl-L-threonine	<i>dcp</i>	0	15.8	0	0
Glycyl-L-aspartic acid	<i>aspA</i>	0	0	0	0
Organic acids					
Acetic acid	<i>acs, ackApta</i>	0	0	0	3.8
	<i>aceBA</i>	0	5.3 <i>aceA</i>	85.7 <i>aceB</i>	0
Fumaric acid	<i>dctA, dctB, cbt, fumA</i>	0	0 <i>dctB</i>	0	0
D-Galacturonic acid	<i>exuTR, kdgT, uxaAC, uxuABR</i>	0	0	0	3.8 <i>kdgT</i>
D-Gluconic acid	<i>eda edd</i>	0	0‡	0	0
	<i>gntT, gntKR, gntP</i>	0	5.3 <i>gntR</i>	0	0
L-Malic acid	<i>mdh</i>	0	0	0	0
DL-Lactic acid	<i>lldPRD, dld, ldhA</i>	0	5.3 <i>lldP</i>	0	3.8 <i>ldhA</i>
Pyruvic acid	<i>pdhR, aceEFlpdA, poxB</i>	0	0	0	0
Succinic acid	<i>sdhCDAB</i>	0	0	14.3 <i>sdhDB</i>	0
Nucleotides and others					
Inosine	<i>gsk</i>	0	5.3	0	0
2'-Deoxy adenosine	<i>tsx, nupG</i>	0	0‡	85.7 <i>nupG</i>	0

Table 3. cont.

Growth substrate	MG1655 transport and/or dissimilatory genes	Percentage of strains in which missing			
		RSW <i>n</i> =5*	Faecal <i>n</i> =19	Commensal <i>n</i> =7	PEC <i>n</i> =26
Thymidine	<i>add deoCABD</i>	0	0	0	0
Uridine	<i>nupC</i>	0	0	0	0
	<i>udp, tktA, tktB</i>	0	0	0	0

*Strain WK2 was not included in CGH analysis due to low signal quality in duplicate array hybridizations.

†Phosphotransferase systems.

‡Not all genes of operon present on array.

§Binding-protein-dependent ATP-cassette (ABC) transport systems.

energy (Table 3). The corresponding substrate-specific genes were present in all RSW strains. In the other datasets included in the CGH analysis, a limited number of these genes were recorded as absent in a low proportion of strains (Table 3).

Table 4 lists catabolic operons for substrates that do not, or only in a subset of *E. coli* strains, support growth in mineral medium. Surprisingly, our CGH data analysis indicated that the respective catabolic operons are nevertheless highly conserved in the *E. coli* population (Table 4). For example, in spite of the fact that dulcitol supported growth of only 50 % of the RSW isolates, the *gatRDCBAZY* operon was present in all of these strains (Table 4). Similarly, although *E. coli* generally lacks the ability to utilize cellobiose (Lin, 1996; compare also second column of Table 4), the *celFDCBA* operon, which is also present on the MG1655 genome, was detected in more than 95 % of the strains (Table 4).

Efficient substrate uptake most likely is a key fitness determinant in bacteria. Phosphotransferase systems are the most efficient uptake systems of *E. coli* under anoxic conditions (Manché *et al.*, 1999), whereas binding-protein-dependent, ATP-driven (ABC) transporters allow *E. coli* to scavenge trace amounts of carbon substrates during aerobic growth (Ferenci, 2001). Operons encoding either type of transport system were only in rare cases (mostly partially) missing from the strains analysed (Tables 3 and 4).

Complex global gene regulation mechanisms have evolved in bacteria which respond rapidly to environmental stimuli. In *E. coli*, most catabolic operons are part of the cAMP-CRP global regulatory network (e.g. the operons for maltose, ribose and arabinose utilization). The genes encoding the key regulatory proteins of the cAMP-CRP network, the *cya* (adenylate cyclase) and *crp* (cAMP receptor protein) genes were present in all RSW, Faecal, Commensal and PEC strains (data not shown). In accordance with known physiological characteristics of *E. coli*, operons needed for mixed acid fermentation and fumarate respiration (*adhE*, *adhP*, *hycIHGFEDCBA*, *fdhF*, *frdDCBA*, *pta*, *ackA*, *poxB*, *ldhA*, *mdh*, *fumCA*, *fumB*, *tdcEDCBAR*) and the respective global regulators (*fur*,

arcA, *arcB*) exhibited a very low frequency of gene deletions (data not shown).

CGH analysis of stress defence functions

Protection against desiccation, acid stress, oxidative stress, starvation and other adverse conditions is a prerequisite for the cycling of *E. coli* between hosts. Therefore, we screened the four CGH datasets for missing or variable stress defence functions. The frequency of missing genes known to encode protective functions was very low in the 57 strains analysed (Table 5). The majority of genes shown in Table 5 are positively regulated by the general stress response sigma factor RpoS (Ishihama, 2000; Loewen *et al.*, 1998). The *rpoS* gene itself, as well as the genes for the heat shock and envelope stress sigma factors (*rpoH*, *rpoE*), showed no variability at all on MG1655 microarrays. One gene for the synthesis of the reserve material glycogen (*glgS*) was variably absent and present; however, the other glycogen synthetase (*glgA*) was detected in all strains (Table 5).

Variable genomic regions

In spite of the large number of shared catabolic and stress defence operons, the RSW strains significantly differed from MG1655 and from one another in their genome composition (Fig. 1). On average, 5.7 % of all MG1655 translatable ORFs spotted on the MWG arrays were absent in the RSW strains (range 4.5–6.9 %). In the 19 faecal isolates, 1.3–17.5 % of MG1655 ORFs were missing (average 7.8 %, median 6.5 %). This corresponds well with an average of 5.8 % MG1655 ORFs missing in 23 pathogenic and commensal *E. coli* strains (Dobrindt *et al.*, 2003).

In most cases not single genes, but stretches of 2 up to 27 ORFs were missing in the RSW strains. The major missing or variable regions (deletion hot spots) were non-randomly distributed and consistent with missing or variable regions in pathogenic *E. coli* (Fig. 1) (Anjum *et al.*, 2003). Most of the missing MG1655 ORFs encoded genes of hypothetical or unknown function. Only three operons with catabolic or transport functions were part of the deletion hot spots, a hypothetical D-allose phosphotransferase system (*yjcVWX*),

Table 4. Percentage of missing transport and catabolic operons in four sets of CGH data

The listed carbon substrates exhibited variable phenotypes in Biolog growth tests with RSW strains or did not support growth at all. Several catabolic operons for substrates not present on Biolog AN microplates were also included in the analysis. All genes analysed are non-essential. RSW, raw source water isolates; Faecal, strains isolated from humans, dogs, cattle, rabbits and swine (Table 1); Commensal, CGH data for six commensal *E. coli* strains and *E. coli* B from Dobrindt *et al.* (2003); PEC, CGH data for 26 pathogenic *E. coli* strains from Anjum *et al.* (2003); NA, not analysed.

Growth substrates	Strains growth-positive (% of 5 RSW strains + MG1655)	MG1655 transport and/or dissimilatory genes	Percentage of strains in which missing			
			RSW <i>n</i> =5*	Faecal <i>n</i> =19	Commensal <i>n</i> =7	PEC <i>n</i> =26
Sugars and sugar alcohols						
L-Arabinose	NA	<i>araDABC</i>	0	0	85.7 <i>araA</i>	0
		<i>araHGF</i> †	0	5.3 <i>araG</i>	0	0
Arbutin	NA	<i>glvBC</i> ‡	0	5.3 <i>glvB</i>	0	0
Methyl β-D-glucoside	66.7	<i>bglA</i>	0	5.3	0	0
		<i>bglBFG</i> ‡	0	15.8/10.5	14.3 <i>bglF</i>	26.9 <i>bglF</i>
		<i>bglJ</i> , <i>bglX</i>	0	0§	0	0
Cellobiose	0	<i>celFDCBA</i> ‡	0	10.5 <i>celF</i>	0	0
Dulcitol	50	<i>gatRDCBAZY</i> ‡	0	NA	42.9 <i>gatBA</i>	7.6
N-Acetyl-D-galactosamine	66.7	<i>agaRZVWASYBCDI</i> ‡	0	5.3 <i>agaS,B</i>	0	0
N-Acetyl-D-mannosamine	16.7	<i>nanTA</i> ‡	0	NA	0	0
D-Ribose	NA	<i>rbsDABCK</i> †, <i>rbsR</i>	0	0	0	0
Xylose	NA	<i>xylBAFGHR</i> †	0	5.3 <i>xylF</i> 10.5 <i>xylH</i>	0	7.6
Amino acids and di/oligopeptides						
L-Arginine	NA	<i>artJMQIP</i> †	0	0	85.7 <i>artI</i>	0
L-Asparagine	66.7	<i>ansAB</i>	0	0	0	0
L-Aspartic acid	0	<i>aspA</i>	0	0	0	0
L-Glutamate	0	<i>aspA</i> , <i>aspC</i>	0	0	0	0
		<i>gltLKJ</i> †	0	15.3 <i>gltJ</i>	0	0
L-Glutamine	66.7	<i>glnQPH</i> †	0	0§	0	0
		<i>gltBDF</i>	100 <i>gltF</i>	0§	71.4 <i>gltF</i>	92.3 <i>gltF</i>
L-Histidine	NA	<i>hisPMQJargT</i> †	0	0§	0	0
L-Proline	NA	<i>proXWV</i> †	0	5.3 <i>proW</i>	0	0
L-Serine	83.3	<i>sdaA</i> , <i>sdaCB</i>	0	5.3 <i>sdaA,B</i>	0	0
L-Threonine	0	<i>tdh</i> <i>kbl</i>	0	42.1 <i>tdh</i>	0	0
		<i>tdcEDCBA</i>	0	0	0	0
L-Tryptophan	NA	<i>tnaAB</i>	0	0	0	0
Glycyl-L-proline	16.7	<i>putAP</i>	0	0	0	0
L-Valine	0	<i>livFGMHKyhKliv</i> †	0	5.3 <i>livF</i>	0	0
Oligopeptides	NA	<i>oppABCD</i> F†	0	0	0	0
Organic acids and others						
C ₄ fatty acids	NA	<i>atoCDAEB</i>	66.7 <i>atoDE</i>	21.1	28.6 <i>atoB</i>	84.6
Fatty acids	NA	<i>fadAB</i>	0	5.3	0	0
		<i>fadR</i> , <i>fadD</i> , <i>fadL</i>	0	0	0	0
Propionic acid	33.3	<i>prpRBCDE</i>	0	5.3 10.5 <i>prpE</i>	14.3 <i>prpC</i>	0
Putrescine	NA	<i>potFGHI</i> †	0	0	0	0
Spermidine	NA	<i>potDCBA</i> †	0	10.5 <i>potC</i>	0	0

*Strain WK2 was not included in CGH analysis due to low signal quality in duplicate array hybridizations.

†Binding-protein-dependent ATP-cassette (ABC) transport systems.

‡Phosphotransferase systems.

§Not all genes of operon present on array.

Table 5. Percentage of missing stress defence operons in four sets of CGH data

All genes listed are non-essential. RSW, raw source water isolates; Faecal, strains isolated from humans, dogs, cattle, rabbits and swine (Table 1); Commensal, CGH data for six commensal *E. coli* strains and *E. coli* B from Dobrindt *et al.* (2003); PEC, CGH data for pathogenic *E. coli* strains from Anjum *et al.* (2003); NA, not analysed.

Function	MG1655 genes	Percentage of strains in which missing			
		RSW <i>n</i> =5*	Faecal <i>n</i> =19	Commensal <i>n</i> =7	PEC <i>n</i> =26
Global stress regulators	<i>rpoS</i>	0	NA	0	0
	<i>rpoH</i>	0	NA	0	0
	<i>rpoE</i>	0	0	0	0
Acid stress defence	<i>gadA</i> †	0	NA	0	0
	<i>gadBC</i> †	0	0	0	0
	<i>gabP</i> †	0	5.3	0	0
	<i>poxB</i> †	0	NA	0	0
Toxic compound stress defence	<i>ada</i> †	0	10.5	0	0
	<i>adhE</i> †	0	0	0	0
	<i>aidB</i> †	0	0	0	0
	<i>aldB</i> †	0	NA	0	0
Oxidative stress defence	<i>dps</i> †	0	NA	14.3	0
	<i>frdA</i> †	0	0	0	0
	<i>gor</i> †	0	0	0	0
	<i>katE</i> †	0	0	0	7.7
	<i>katG</i>	0	0	0	0
	<i>oxyR</i> †	0	0	0	0
	<i>sodB</i>	0	0	0	0
	<i>sodC</i> †	0	0	0	3.8
	<i>soxS</i> †	0	NA	0	0
	<i>bolA</i> †	0	0	0	0
Desiccation/osmotic stress defence	<i>osmB,C,E,Y</i> †	0	0‡	0	0
	<i>otsAB</i> †	0	NA	0	0
	<i>treA</i> †	0	0	0	0
	<i>proP</i> †	0	0	0	0
	<i>cfa</i> †	0	0	0	0
General stress protection/chaperones	<i>dank</i> †	0	0	0	0
	<i>dnaJ</i>	0	0	0	0
	<i>uspA</i> †	0	0	0	0
	<i>xthA</i> †	0	0	0	0
DNA protection/repair	<i>glgA</i>	0	0	0	0
Reserve material synthesis	<i>glgS</i> †	0	NA	42.9	3.8
High temperature stress defence	<i>clpB</i>	0	0	0	0
	<i>htrE</i> †	20	15.3	0	0
	<i>htpG</i>	0	0	0	0
Cold shock	<i>cspA</i>	0	0	14.3	0

*Strain WK2 was not included in CGH analysis due to low signal quality in duplicate array hybridizations.

†RpoS-dependent.

‡Not all genes of operon present on array.

the catabolic operon for C₄-fatty acids (*atoCDAEB*) and an Fe(III)-dicitrate uptake system including a TonB-dependent outer membrane receptor and an ABC transporter (*fecIRABCDE*). Also found on frequently missing DNA stretches were the *mcr* and *hsr* operons involved in DNA

restriction, the LPS synthesis operons *wbb* and *rfb*, the *hof* operon encoding a type II secretion system and the *idn* operon involved in sugar acid metabolism. LPS composition is known to vary strongly between *E. coli* strains and has been associated with HGT.

Fig. 1. Distribution of missing MG1655 genes (black bars) over the genomes of RSW isolates (WK1, 3, 4, 5 and 8) and pathogenic strains (PEC; data from Anjum *et al.*, 2003) of *E. coli*. From left to right: genome position in MG1655 (Blattner no. B0001 to Blattner no. B4403).

Physiological characteristics of environmental isolates in comparison to *E. coli* MG1655

Intraspecies genomic variation and a proposed trade-off between self-protection and nutritional competence (King *et al.*, 2004) can theoretically result in a substantial variation in nutritional versatility, in the RpoS phenotype and in growth kinetic parameters of environmental *E. coli*. However, the RSW isolates varied little from one another and from MG1655 in these physiological characteristics, in spite of clear phenotypic differences in antibiotic resistance, siderophore excretion and haemolysin or colicin production (Tables 1 and 2). The latter functions are predominantly encoded on mobile genetic elements such as plasmids, transposons and integrons.

The substrate spectrum of source water isolates analysed with Biolog AN MicroPlates exhibited some variation, but did not differ significantly from that of *E. coli* MG1655 (data not shown). The total number of growth-supporting

substrates was rather similar and showed no obvious negative correlation with the expression of stress defence functions (Table 6). Variation in the raffinose, dulcitol and sucrose phenotypes (Table 3) is consistent with published data for *E. coli* isolates (Durso *et al.*, 2004; Holt, 1994; Miller & Hartl, 1986). Raffinose and sucrose are two of the few catabolic genes that are easily transferred between *E. coli* strains (Smith & Parsell, 1975). For other substrates, mutational silencing may have led to variation between strains. However, some of the phenotype variability, especially with regard to amino acids, may be an artefact, because these substrates supported in general only weak growth; also with the 'negative' strains, there was usually some colour formation observed, although it did not pass the (arbitrary) threshold of 0.4 A_{460} for recording growth as positive. MG1655 cells grown in complex medium dissimilate 'growth negative' substrates such as L-serine at high rates (Ihssen & Egli, 2005). In the natural habitats of *E. coli*, amino acids are always a part of substrate mixtures.

Table 6. Growth kinetic parameters, hydroperoxidase II (HPII) specific activity, relative RpoS levels and number of growth positive substrates on Biolog microplates of six raw source water isolates (WK1–5 and WK8) and *E. coli* K-12 MG1655

Growth yield [$Y_{x/s}$, g dry wt (g glucose) $^{-1}$], glucose affinity ($K_{s,Glc}$, $\mu\text{g l}^{-1}$), HPII specific activity [$\mu\text{mol H}_2\text{O}_2$ decomposed min^{-1} (mg total protein) $^{-1}$] and RpoS level (percentage of K-12 level) were determined in glucose-limited chemostat cultures 20 h after inoculation ($D=0.3 \text{ h}^{-1}$). Values for μ_{max} , $Y_{x/s}$, K_s and HPII represent means of a minimum of three replicate determinations. Standard deviations for $\mu_{\text{max,Glc}}$ and $Y_{x/s}$ were below $\pm 2\%$ of the mean value.

Strain	$\mu_{\text{max, Glc}}$	$Y_{x/s}$	$K_{s,Glc}$	RpoS level	HPII	Growth-supporting substrates
K-12	0.63*	0.44	539 ± 12	100	28 ± 4	47
WK1	0.92	0.47	468 ± 19	49	49 ± 4	48
WK2	0.85	0.37	1199 ± 160	80	26 ± 1	46
WK3	0.92	0.37	365 ± 20	25	20 ± 4	52
WK4	0.97	0.41	1498 ± 26	54	58 ± 4	52
WK5	0.95	0.45	1061 ± 10	67	55 ± 13	52
WK8	0.92	0.44	1257 ± 26	149	26 ± 5	46

*Data from Wick *et al.* (2002).

All environmental (RSW) isolates expressed detectable amounts of RpoS in glucose-limited chemostat cultures (Table 6). At a fixed specific growth rate of 0.3 h^{-1} , intracellular RpoS concentrations varied by a factor of approximately six (Table 6) and the RpoS level of MG1655 was in the upper third of the observed range (Table 6). For unknown reasons, hydroperoxidase II (HPII) specific activity in the different strains did not correlate well with relative RpoS levels (Table 6), although HPII is a good indicator for RpoS levels in MG1655 when growing at different specific growth rates (Ihssen & Egli, 2004). Similarly to RpoS levels, HPII specific activity of MG1655 was present at a level comparable to that of environmental strains (Table 6). All RSW isolates exhibited both heat-stable HPII (KatE) and heat-labile HPI (KatG) catalase activity (data not shown), which is consistent with CGH data for the corresponding genes (Table 5).

Glucose affinities of the RSW isolates differed by a factor of four at the most. Surprisingly, some of the environmental strains exhibited a considerably lower affinity for glucose ($K_{s,\text{Glc}}$) than wild-type *E. coli* MG1655 (Table 5). Judged from the capacity of MG1655 to dramatically improve its $K_{s,\text{Glc}}$ more than tenfold down to $50 \mu\text{g l}^{-1}$ within 200 generations in glucose-limited chemostat cultures (Wick *et al.*, 2002), neither laboratory nor natural growth environments seem to select for genotypes with high glucose affinity.

RpoS is known to have a negative effect on the expression of high-affinity glucose uptake proteins and thus on $K_{s,\text{Glc}}$ (Notley-McRobb *et al.*, 2002; Wick *et al.*, 2002). Whereas some strains confirmed this correlation, others (WK4 and K12) did not (Table 6). A negative effect of positive antibiotic resistance, siderophore, colicin or haemolysin phenotypes on glucose affinity was not detected (Tables 1, 2 and 6). Thus, carriage of the corresponding genetic elements seems not to decrease competitiveness for carbon substrates.

Glucose growth yields of the RSW strains and MG1655 were rather similar (Table 6), which is in agreement with a

narrow range of growth yields [$0.42\text{--}0.47 \text{ g dry wt (g glucose)}^{-1}$, mean $0.45 \text{ g dry wt (g glucose)}^{-1}$] observed in seven natural isolates and one laboratory strain by Mikkola & Kurland (1992). Maximum specific growth rates in glucose mineral medium also showed remarkably little variation, but were in general approximately 20 % higher than the $\mu_{\text{max,Glc}}$ of MG1655 (Table 6). The reduced $\mu_{\text{max,Glc}}$ of MG1655 is caused by the *rph1* mutation that impairs pyrimidine biosynthesis (Soupeine *et al.*, 2003). Other wild-type laboratory strains exhibit $\mu_{\text{max,Glc}}$ values close to the RSW strains, e.g. 0.78 h^{-1} and 0.87 h^{-1} for strains ML 30G and 017, respectively (Mikkola & Kurland, 1992; Shehata & Marr, 1971). In contrast to our findings, Mikkola & Kurland (1992) reported a broad range of (initial) $\mu_{\text{max,Glc}}$ values from 0.34 to 0.99 h^{-1} (mean 0.64 h^{-1}) for seven natural *E. coli* isolates. Except for differences in experimental procedures, we have no explanation for this contradiction. Durso *et al.* (2004) also reported a low $\mu_{\text{max,Glc}}$ variability for 41 commensal and 81 pathogenic *E. coli* strains, but the mean $\mu_{\text{max,Glc}}$ was significantly lower (0.42 h^{-1}). However, the lower values may have resulted from oxygen limitation in the Biolog microplate wells the strains were cultivated in.

Physiological response of environmental *E. coli* to glucose limitation

E. coli MG1655 adapts to low concentrations of carbon and energy sources in chemostat cultures by the derepression of diverse catabolic functions and high-affinity binding proteins (Ihssen & Egli, 2005). A very similar physiological response to glucose limitation was observed in environmental isolates. The pattern of derepressed catabolic functions resembled that of MG1655 (Fig. 2). Close similarity was also observed for the pattern of high-affinity binding proteins upregulated in response to glucose limitation (Fig. 3). The upregulation included, as judged from the position of the bands on the gel, DppA (A), MalE (B), MglB (C), RbsB (D) and GlnH (E) in most strains (bands were identified in a previous study with MG1655; Ihssen & Egli, 2005).

Fig. 2. Derepression of catabolic functions in five *E. coli* RSW isolates and *E. coli* K-12 MG1655 in response to glucose limitation. Substrate-dependent respiration rates of chloramphenicol-inhibited cells were determined with Biolog AN MicroPlates. White bars, excess glucose batch cultures; black bars, aerobic, glucose-limited chemostat cultures ($D=0.3 \text{ h}^{-1}$). Data for K-12 from Ihssen & Egli (2005).

Fig. 3. Pattern of periplasmic proteins expressed by RSW isolates and *E. coli* K-12 MG1655 under glucose excess and glucose-limited conditions. Ba, exponential phase in glucose batch culture; Ch, glucose-limited chemostat culture ($D=0.3\text{ h}^{-1}$). A, dipeptide-binding protein DppA; B, maltose-binding protein MalE; C, galactose/glucose-binding protein MglB; D, ribose-binding protein MglB; E, glutamine-binding protein GlnH (see also Ihssen & Egli, 2005).

DISCUSSION

Conserved adaptation of *E. coli* to the primary habitat

The primary habitat of *E. coli*, the colon, is a complex ecosystem with a high diversity of bacterial species, which can coexist presumably due to the diversity of growth substrates entering it (Mason & Richardson, 1981). However, still very little is known about the nutritional preferences of gut bacteria (Peekhaus & Conway, 1998). Experimental evidence suggests that *E. coli* successfully competes for a variety of carbon and energy sources in the colon because it can establish itself in multi-species anaerobic bioreactors simulating the digestive tract (Freter *et al.*, 1983a, b; Macfarlane *et al.*, 1998).

It has been suggested that the ecologically relevant traits, i.e. the traits crucial for successful competition against other bacterial species, are those found in all strains and not those that are variable (Mason & Richardson, 1981). In this study, a very low level of intraspecies genomic variation in phosphotransferase systems and dissimilatory enzymes for sugars and sugar alcohols was found (Tables 3 and 4). Therefore, we hypothesize that *E. coli* is specialized on a particular set of fermentable substrates in the gut, mainly monomeric and dimeric sugars, sugar acids and sugar alcohols. The monomeric substrates can be derived either from secreted mucus, e.g. *N*-acetylglucosamine, galactose, glucuronate (Peekhaus & Conway, 1998), or from polymers degraded by extracellular enzymes of other gut bacteria, e.g. arabinose from pectin, gluconate from muscle tissue and glucose from cellulose. The large arsenal of sugar, sugar acid and sugar alcohol catabolic operons presumably allows *E. coli* to engage in kinetically favourable mixed-substrate growth (Egli, 1995) and to flexibly react to transient availability of these substrates. In our view, the genomic variation observed in few catabolic functions can hardly reflect the specialization of different *E. coli* lineages to specific nutritional niches in the colon. Thus, the 'catabolome' of the species *E. coli* seems to be highly conserved.

Conserved adaptation of *E. coli* to the secondary habitat

When leaving the alimentary tract of the host, *E. coli* is faced with (i) very low concentrations of carbon and energy substrates, (ii) oxic conditions, (iii) low temperatures, (iv) UV AB and oxidative stress, and (v) osmotic and desiccation stress. In addition, cells must survive severe acid stress in the stomach before being able to colonize a new host. Both CGH and physiological data suggest that the evolutionary adaptation to these challenges is highly conserved in *E. coli*.

Due to the poor ATP yield per utilized sugar molecule in anaerobically growing *E. coli*, ATP-consuming binding-protein-dependent ABC transporters presumably are of limited use in the anoxic gut. However, in the predominantly oxic secondary habitat (water and soil), the ability to scavenge trace amounts of carbon substrates with ABC transporters may be one of the most important fitness factors. The presence of all known ABC-transport systems for amino acids and sugars in practically all strains and the very low variability in catabolic operons for organic acids and nucleotides (Tables 3 and 4) suggest that *E. coli* takes up and metabolizes variable mixtures of minute concentrations of these substrates for maintenance purposes in the secondary habitat.

A drawback of the microarray approach is the neglect of differences caused by more subtle genetic changes (e.g. single nucleotide exchanges, frame shift mutations) in the catabolic gene itself or in regulatory proteins. Thus, strains might differ significantly in catabolic abilities and stress resistance in spite of a 'similar' genome composition. At least regarding the six RSW strains and *E. coli* K-12 MG1655, physiological analysis indicates that expression of the non-variable catabolic and stress defence genes does not vary to a great extent (Table 6, Figs 2 and 3). Particularly, the remarkably consistent pattern of derepressed catabolic pathways and high-affinity binding proteins under glucose-limited growth conditions (Figs 2 and 3) suggests that a carbon source foraging strategy (Liu

et al., 2005; Ihssen & Egli, 2005) is a highly conserved (eco-) physiological adaptation of *E. coli*. Analysis of the proteome under well-defined conditions might give a more detailed picture of differences in gene expression among *E. coli* isolates, but this is far beyond the scope of this study.

The CGH analysis also suggests that *E. coli* cannot tolerate much genomic variation in stress defence functions. Especially the highly conserved RpoS-dependent genes (Table 5), which play a key role in the survival of starvation and other stressful conditions (Matin, 1991; Munro *et al.*, 1995; Eisenstark *et al.*, 1996; Booth *et al.*, 2002), might be indispensable for the passage of *E. coli* through non-host environments to a new intestinal habitat.

Our data do not support the proposed trade-off between catabolic versatility and RpoS-dependent self-protection (King *et al.*, 2004) in natural *E. coli* populations (Table 6). The trade-off hypothesis is based on observations for one pathogenic and one laboratory strain, which showed very high RpoS/RpoD ratios. The catabolic spectrum of the four other strains used in the latter study (including K-12 MG1655 and two EcoR reference strains) was not improved by an RpoS knockout mutation and these strains had quite similar RpoS/RpoD ratios (King *et al.*, 2004). Presumably, only strains with an extremely strong RpoS phenotype are negatively affected in their catabolic versatility, but such strains are not likely to be widespread in the total *E. coli* population because they would have a competitive disadvantage in many growth situations. One may speculate that the frequency of strong RpoS phenotypes might be higher among pathogenic *E. coli* because they may require a more frequent transfer between hosts and do not compete with the intestinal flora for carbon and energy sources. However, the fact that 17.2% of 58 shiga-like-toxin-producing *E. coli* were RpoS-attenuated (Waterman & Small, 1996) speaks against a generally strong RpoS phenotype in pathogenic strains. To decide how variable the RpoS phenotype and how broad the range of stress resistance truly is in the natural *E. coli* population, and how it affects nutritional abilities, more data for commensal and pathogenic isolates are needed.

How important are variable (HGT-related) traits relative to the core genome?

The relative contribution of horizontal gene transfer (HGT) and vertical inheritance to the evolution of bacteria remains fiercely debated (for recent reviews with opposing views see Dagan & William, 2006; Kurland, 2005 and Philippe & Douady, 2003). The numerous prophages found in *E. coli* genomes (Perna *et al.*, 2001) indicate the widespread occurrence of HGT also in this species. On the one hand, HGT-driven evolution of pathogenic strains (Blum *et al.*, 1994; Hacker & Kaper, 2000) speaks in favour of an adaptation of *E. coli* to novel niches by horizontal acquisition of foreign genes. On the other hand, evidence for HGT-driven niche specialization of commensal *E. coli*, which by far outnumber pathotypes in terms of total

populations size, is scarce. If one assumes that ecology is metabolism for (non-pathogenic) bacteria (Lawrence, 2001), the high degree of similarity in the genome content of catabolic pathways found in our study indicates that the adaptation of *E. coli* to its current niche has been dominated by vertical inheritance. Whether the *lac* operon, which once was proposed as a paradigmatic example of niche adaptation by HGT (Ochman *et al.*, 2000), counts as counter evidence is questionable, because strong doubts have been raised whether the *lac* operon ever was transferred laterally into *E. coli* (Stoebel, 2004).

Horizontally transferred properties like the ability to produce colicins, O-antigen composition, fimbriae and capsules may nevertheless contribute to strain substitutions and differential persistence in the colon (Gordon *et al.*, 1998; Nowrouzian *et al.*, 2006). In an *in vivo* experimental study though, no positive effect of fimbriae or adhesin plasmids on the persistence of particular strains in the human digestive tract was found (Smith & Huggins, 1978). Further ecological studies are needed to clarify the role of HGT-acquired genes in the competition between different commensal *E. coli* strains.

It is striking that the strain-specific genomic regions of *E. coli* MG1655 (K-islands; Perna *et al.*, 2001) are dominated so much by phage-related genes (Anjum *et al.*, 2003; Dobrindt *et al.*, 2003). A high proportion of phage-related genes among variably absent or present ORFs in *E. coli* has also been found in a detailed CGH study on O55:H7 and O157:H7 strains (Wick *et al.*, 2005). These genes are most likely parasitic in nature rather than contributing to the fitness of the host strain (Kurland, 2005). The fact that the majority of horizontally acquired DNA is maintained in the *E. coli* genome only for – in evolutionary timescales – short periods of time (Lawrence & Ochman, 1998) also speaks against routine fitness gains by HGT.

A last point we would like to make is that the term ‘housekeeping genes’, commonly used for the content of the conserved *E. coli* core genome (e.g. Anjum *et al.*, 2003; Dobrindt *et al.*, 2003), is misleading, because it has the connotation ‘non-niche-specific traits’, i.e. genes needed for basic energy generation, biosynthesis and cell division. With the core genome of *E. coli* apparently including so many genes needed for growth and survival in its primary and secondary habitats, the terms ‘life cycle’- or ‘habitat’-specific genes might be more appropriate.

ACKNOWLEDGEMENTS

We thank Beat Birkhofer for the initial phenotypic characterization of raw source water isolates. Thanks are also due to Karin Kovar from the University of Applied Sciences, Wädenswil (HSW), Switzerland, for supervising the diploma thesis of Claudio Bassin and enabling him to do his thesis work at EAWAG. The authors are indebted to Annette Rust for supplying *E. coli* raw source water isolates and to Alexander J. B. Zehnder for support throughout this study. We appreciate the CGH data made available to us by Ulrich Dobrindt and the freely accessible online data of M. F. Anjum and co-workers. Thanks are

also due to Teresa Colangelo for confirmatory antibiotic resistance tests, Alessandro Franchini for help with microarray experiments and Manuela Bento for help with the ECOR group determination. We thank Lukas Wick for helpful methodological suggestions. The financial support by the Swiss Federal Institute of Environmental Science and Technology (EAWAG) is gratefully acknowledged.

REFERENCES

- Ames, G. F., Prody, C. & Kustu, S. (1984). Simple, rapid, and quantitative release of periplasmic proteins by chloroform. *J Bacteriol* **160**, 1181–1183.
- Anjum, M. F., Lucchini, S., Thompson, A., Hinton, J. C. & Woodward, M. J. (2003). Comparative genomic indexing reveals the phylogenomics of *Escherichia coli* pathogens. *Infect Immun* **71**, 4674–4683.
- Bergthorsson, U. & Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Mol Biol Evol* **15**, 6–16.
- Berlyn, M. K. (1998). Linkage map of *Escherichia coli* K-12, edition 10: the traditional map. *Microbiol Mol Biol Rev* **62**, 814–984.
- Bettelheim, K. A. (1992). The genus *Escherichia*. In *The Prokaryotes*, pp. 2696–2736. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K.-H. Schleifer. New York: Springer Verlag.
- Beutin, L., Montenegro, M. A., Orskov, I., Orskov, F., Prada, J., Zimmermann, S. & Stephan, R. (1989). Close association of verotoxin (Shiga-like toxin) production with enterohemolysin production in strains of *Escherichia coli*. *J Clin Microbiol* **27**, 2559–2564.
- Blattner, F. R., Plunkett, G., III, Bloch, C., Perna, N., Burland, V., Riley, M., Collado-Vides, J., Glasner, J., Rode, C. & other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462.
- Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschape, H. & Hacker, J. (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* **62**, 606–614.
- Booth, I. R., Cash, P. & O'Byrne, C. (2002). Sensing and adapting to acid stress. *Antonie Van Leeuwenhoek* **81**, 33–42.
- Charbonnier, Y., Gettler, B., Patrice François, P., Bento, M., Renzoni, A., Vaudaux, P., Schlegel, W. & Schrenzel, J. (2005). A generic approach for the design of whole-genome oligoarrays, validated for genotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus*. *BMC Genomics* **6**, 95.
- Chen, G., Patten, C. L. & Schellhorn, H. E. (2004). Positive selection for loss of RpoS function in *Escherichia coli*. *Mutat Res* **554**, 193–203.
- Clermont, O., Bonacorsi, S. & Bingen, E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66**, 4555–4558.
- Cooper, V. S. & Lenski, R. E. (2000). The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407**, 736–739.
- Dagan, T. & William, M. (2006). The tree of one percent. *Genome Biol* **7**, 118.
- Dobrindt, U., Agerer, F., Michaelis, K., Janka, A., Buchrieser, C., Samuelson, M., Svanborg, C., Gottschalk, G., Karch, H. & Hacker, J. (2003). Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J Bacteriol* **185**, 1831–1840.
- Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**, 414–424.
- Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* **284**, 2124–2129.
- Durso, L. M., Smith, D. & Hutkins, R. W. (2004). Measurements of fitness and competition in commensal *Escherichia coli* and *E. coli* O157:H7 strains. *Appl Environ Microbiol* **70**, 6466–6472.
- Egli, T. (1995). The ecological and physiological significance of the growth of heterotrophic microorganisms with mixtures of substrates. *Adv Microb Ecol* **14**, 305–386.
- Eisenstark, A., Calcutt, M. J., Becker-Hapak, M. & Ivanova, A. (1996). Role of *Escherichia coli* rpoS and associated genes in defence against oxidative damage. *Free Radic Biol Med* **21**, 975–993.
- Ferenci, T. (2001). Hungry bacteria – definition and properties of a nutritional state. *Environ Microbiol* **3**, 605–611.
- Freter, R., Brickner, H., Fekete, J., Vickerman, M. M. & Carey, K. E. (1983a). Survival and implantation of *Escherichia coli* in the intestinal tract. *Infect Immun* **39**, 686–703.
- Freter, R., Brickner, H., Botney, M., Clevon, D. & Aranki, A. (1983b). Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infect Immun* **39**, 676–685.
- Gordon, D. M., Riley, M. A. & Pinou, T. (1998). Temporal changes in the frequency of colicinogeny in *Escherichia coli* from house mice. *Microbiology* **144**, 2233–2240.
- Hacker, J. & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641–679.
- Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., Ote, T., Yamakawa, T., Yamazaki, Y. & other authors (2005). Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol Microbiol* **55**, 137–149.
- Hengge-Aronis, R. (1996). Regulation of gene expression during entry into stationary phase. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, pp. 1497–1512. Edited by F. C. Neidhardt and others. Washington, DC: ASM Press.
- Holt, J. G. (1994). *Bergey's Manual of Determinative Bacteriology*. Baltimore, USA: Williams & Wilkins.
- Ihssen, J. & Egli, T. (2004). Specific growth rate and not cell density controls the general stress response in *Escherichia coli*. *Microbiology* **150**, 1637–1648.
- 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. *Environ Microbiol* **7**, 1568–1581.
- Ishihama, A. (2000). Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* **54**, 499–518.
- Kanehisa, M. & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27–30.
- Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., Peralta-Gil, M. & Karp, P. D. (2005). EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res* **33**, D334–D337.
- Kim, C. C., Joyce, E. A., Chan, K. & Falkow, S. (2002). Improved analytical methods for microarray-based genome-composition analysis. *Genome Biol* **3**. doi:10.1093/nar/gki108
- King, T., Ishihama, A., Kori, A. & Ferenci, T. (2004). A regulatory trade-off as a source of strain variation in the species *Escherichia coli*. *J Bacteriol* **186**, 5614–5620.
- Kurland, C. G. (2005). What tangled web: barriers to rampant horizontal gene transfer. *Bioessays* **27**, 741–747.
- Lan, R. & Reeves, P. R. (2000). Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol* **8**, 396–401.

- Lange, R. & Hengge-Aronis, R. (1994). The cellular concentration of the σ^S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes Dev* **8**, 1600–1612.
- Lawrence, J. G. (2001). Catalyzing bacterial speciation: correlating lateral transfer with genetic headroom. *Syst Biol* **50**, 479–496.
- Lawrence, J. G. & Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* **95**, 9413–9417.
- Lin, E. C. C. (1996). Dissimilatory pathways for sugars, polyols and carboxylates. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, pp. 307–342. Edited by F. C. Neidhardt and others. Washington, DC: ASM Press.
- Liu, M., Durfee, T., Cabrera, J. E., Zhao, K., Jin, D. J. & Blattner, F. R. (2005). Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. *J Biol Chem* **280**, 15921–15927.
- Loewen, P. C., Hu, B., Strutinsky, J. & Sparling, R. (1998). Regulation in the *rpoS* regulon of *Escherichia coli*. *Can J Microbiol* **44**, 707–717.
- Macfarlane, G. T. & Macfarlane, S. (1997). Human colonic microbiota: ecology, physiology and metabolic potential of intestinal bacteria. *Scand J Gastroenterol Suppl* **222**, 3–9.
- Macfarlane, G. T., Macfarlane, S. & Gibson, G. R. (1998). Validation of a three-stage compound continuous culture system for investigating the effect of retention time on the ecology and metabolism of bacteria in the human colon. *Microb Ecol* **35**, 180–187.
- Manché, K., Notley-McRobb, L. & Ferenci, T. (1999). Mutational adaptation of *Escherichia coli* to glucose limitation involves distinct evolutionary pathways in aerobic and oxygen-limited environments. *Genetics* **153**, 5–12.
- Mason, T. G. & Richardson, G. (1981). *Escherichia coli* and the human gut: some ecological considerations. *J Appl Bacteriol* **51**, 1–16.
- Matin, A. (1991). The molecular basis of carbon-starvation-induced general resistance in *Escherichia coli*. *Mol Microbiol* **5**, 3–10.
- Matin, A. C. (2004). Starvation, bacterial. In *Desk Encyclopedia of Microbiology*, pp. 951–959. Edited by M. Schaechter. Amsterdam: Elsevier Academic Press.
- McFall, E. & Newman, E. B. (1996). Amino acids as carbon sources. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, pp. 307–342. Edited by F. C. Neidhardt and others. Washington, DC: ASM Press.
- Mikkola, R. & Kurland, C. G. (1992). Selection of laboratory wild-type phenotype from natural isolates of *Escherichia coli* in chemostats. *Mol Biol Evol* **9**, 394–402.
- Miller, R. D. & Hartl, D. L. (1986). Biotyping confirms a nearly clonal population structure in *Escherichia coli*. *Evolution* **40**, 1–12.
- Munro, P. M., Flatau, G. N., Clement, R. L. & Gauthier, M. J. (1995). Influence of the RpoS (KatF) sigma factor on maintenance of viability and culturability of *Escherichia coli* and *Salmonella typhimurium* in seawater. *Appl Environ Microbiol* **61**, 1853–1858.
- Notley-McRobb, L. & Ferenci, T. (1999a). Adaptive *mgl*-regulatory mutations and genetic diversity evolving in glucose-limited *Escherichia coli* populations. *Environ Microbiol* **1**, 33–43.
- Notley-McRobb, L. & Ferenci, T. (1999b). The generation of multiple coexisting *mal*-regulatory mutations through polygenic evolution in glucose-limited populations of *Escherichia coli*. *Environ Microbiol* **1**, 45–52.
- 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. *J Bacteriol* **184**, 806–811.
- Nowrouzian, F. L., Adlerberth, I. & Wold, A. (2006). Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. *Microbes Infect* **8**, 834–840.
- Nyström, T. (2004). Stationary-phase physiology. *Annu Rev Microbiol* **58**, 161–181.
- Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
- Ölschläger, T., Schramm, E. & Braun, V. (1984). Cloning and expression of the activity and immunity genes of colicins B and M on ColBM plasmids. *Mol Gen Genet* **196**, 482–487.
- Peekhaus, N. & Conway, T. (1998). What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *J Bacteriol* **180**, 3495–3502.
- Perna, N. T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J. & other authors (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533.
- Philippe, H. & Douady, C. J. (2003). Horizontal gene transfer and phylogenetics. *Curr Opin Microbiol* **6**, 498–505.
- Reeves, P. R. (1992). Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiol Lett* **79**, 509–516.
- Savageau, M. A. (1983). *Escherichia coli* habitats, cell types and molecular mechanisms of gene control. *Am Nat* **122**, 732–744.
- Schwyn, B. & Neilands, J. B. (1987). Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* **160**, 47–56.
- Selander, R. K., Caugant, D. A. & Whittam, T. S. (1996). Genetic structure and variation in natural populations of *Escherichia coli*. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 1625–1648. Edited by F. C. Neidhardt and others. Washington, DC: ASM Press.
- Senn, H., Lendenmann, U., Snozzi, M., Hamer, G. & Egli, T. (1994). The growth of *Escherichia coli* in glucose-limited chemostat cultures: a re-examination of the kinetics. *Biochim Biophys Acta* **1201**, 424–436.
- Shehata, T. E. & Marr, A. G. (1971). Effect of nutrient concentration on the growth of *Escherichia coli*. *J Bacteriol* **107**, 210–216.
- Smith, H. W. & Huggins, M. B. (1978). The effect of plasmid-determined and other characteristics on the survival of *Escherichia coli* in the alimentary tract of two human beings. *J Gen Microbiol* **109**, 375–379.
- Smith, H. W. & Parsell, Z. (1975). Transmissible substrate-utilizing ability in enterobacteria. *J Gen Microbiol* **87**, 129–140.
- Soupene, E., van Heeswijk, W. C., Plumbridge, J., Stewart, V., Bertenthal, D., Lee, H., Prasad, G., Paliy, O., Charernnoppakul, P. & Kustu, S. (2003). Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression. *J Bacteriol* **185**, 5611–5626.
- Stoebel, D. M. (2004). Lack of evidence for horizontal transfer of the *lac* operon into *Escherichia coli*. *Mol Biol Evol* **22**, 683–690.
- Visick, J. E. & Clarke, S. (1997). RpoS- and OxyR-independent induction of HPI catalase at stationary phase in *Escherichia coli* and identification of *rpoS* mutations in common laboratory strains. *J Bacteriol* **179**, 4158–4163.
- Waterman, S. R. & Small, P. L. (1996). Characterization of the acid resistance phenotype and *rpoS* alleles of shiga-like toxin-producing *Escherichia coli*. *Infect Immun* **64**, 2808–2811.
- Welch, R. A., Burland, V., Plunkett, G., III, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A. & other authors (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**, 17020–17024.
- Wick, L. M. & Egli, T. (2004). Molecular components of physiological stress responses in *Escherichia coli*. *Adv Biochem Eng Biotechnol* **89**, 1–45.

Wick, L. M., Quadroni, M. & Egli, T. (2001). Short- and long-term changes in proteome composition and kinetic properties in a culture of *Escherichia coli* during transition from glucose-excess to glucose-limited growth conditions in continuous culture and *vice versa*. *Environ Microbiol* **3**, 588–599.

Wick, L. M., Weilenmann, H. & Egli, T. (2002). The apparent clock-like evolution of *Escherichia coli* in glucose-limited chemostats is reproducible at large but not at small population sizes and can be explained with Monod kinetics. *Microbiology* **148**, 2889–2902.

Wick, L. M., Qi, W., Lacher, D. W. & Whittam, T. S. (2005). Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. *J Bacteriol* **187**, 1783–1791.

Zinser, E. R. & Kolter, R. (1999). Mutations enhancing amino acid catabolism confer a growth advantage in stationary phase. *J Bacteriol* **181**, 5800–5807.

Edited by: W. Margolin