Molecular Microbiology (2013) 87(6), 1200–1222 ■
doi:10.1111/mmi.12161
First published online 6 February 2013

Phylogenetic group-associated differences in regulation of the common colonization factor Mat fimbria in *Escherichia coli*

Timo A. Lehti,[1†] Philippe Bauchart,[2‡] Maini Kukkonen,[1§] Ulrich Dobrindt,[2,3¶] Timo K. Korhonen[1] and Benita Westerlund-Wikström[1*]

[1]*Division of General Microbiology, Department of Biosciences, FI-00014 University of Helsinki, Helsinki, Finland.*

[2]*Institute for Molecular Biology of Infectious Diseases, Julius-Maximilians-University Würzburg, D-97080 Würzburg, Germany.*

[3]*Institute of Hygiene, University of Münster, D-48149 Münster, Germany.*

Summary

Heterogeneity of cell population is a key component behind the evolutionary success of *Escherichia coli*. The heterogeneity supports species adaptation and mainly results from lateral gene transfer. Adaptation may also involve genomic alterations that affect regulation of conserved genes. Here we analysed regulation of the *mat* (or *ecp*) genes that encode a conserved fimbrial adhesin of *E. coli*. We found that the differential and temperature-sensitive expression control of the *mat* operon is dependent on *mat* promoter polymorphism and closely linked to phylogenetic grouping of *E. coli*. In the *mat* promoter lineage favouring fimbriae expression, the *mat* operon-encoded regulator MatA forms a positive feedback loop that overcomes the repression by H-NS and stabilizes the fimbrillin mRNA under low growth temperature, acidic pH or elevated levels of acetate. The study exemplifies phylogenetic group-associated expression of a highly common surface organelle in *E. coli*.

Accepted 17 January, 2013. *For correspondence. E-mail benita.westerlund@helsinki.fi; Tel. (+358) 9 19159251; Fax (+358) 9 19159262. Present addresses: †Division of Biochemistry and Biotechnology, Department of Biosciences, University of Helsinki, Helsinki, Finland; ‡Concordia University, Montreal, Canada; §National Supervisory Authority for Welfare and Health, Helsinki, Finland; ¶Institute of Hygiene, University of Münster, Münster, Germany.

Introduction

Escherichia coli is an important commensal and pathogen of humans and animals. Isolates of the species are associated with newborn meningitis and several forms of urinary tract infections or diarrhoeal diseases in children and adults. Currently, two schemes are used to classify *E. coli* isolates, i.e. a pathovar grouping based on the symptoms and mode of transmission of the disease, and a phylogenetic grouping based on genetic properties. The pathovars are characterized by fitness or virulence factors that are frequently encoded by genes present on mobile genetic elements and enable the bacteria to colonize and invade niches not colonized by commensal *E. coli* (Kaper *et al.*, 2004; Croxen and Finlay, 2010). *E. coli* pathovars are not discrete units but have generally arisen on multiple occasions as a result of lateral acquisition of genes (Chaudhuri and Thomas, 2007). Genetic methods separate *E. coli* into four main phylogenetic groups: A, B1, B2 and D, and two minor groups E and F (Herzer *et al.*, 1990; Chaudhuri and Thomas, 2007; Tenaillon *et al.*, 2010). The phylogenetic characteristics include genome size, conserved restriction fragmentation, sequences of housekeeping genes, and possession of genetic elements and specific genes. Diarrhoeagenic isolates are found in various phylogenetic groups, indicating that the two classification systems are not mutually exclusive. On the other hand, urinary and newborn meningitis isolates are genetically more uniform and mainly classify into the B2 group (Picard *et al.*, 1999; Chaudhuri and Thomas, 2007; Touchon *et al.*, 2009) which together with strains in group A is well adapted for colonization of the human host (Obata-Yasuoka *et al.*, 2002; Zhang *et al.*, 2002; Watt *et al.*, 2003; Nowrouzian *et al.*, 2005; 2006; Moreno *et al.*, 2009; Bailey *et al.*, 2010; Tenaillon *et al.*, 2010; de Muinck *et al.*, 2011). In non-human hosts and in environmental samples, strains of group B1 are predominant (Walk *et al.*, 2007; Tenaillon *et al.*, 2010). Genetic analyses indicate that the B2 group is the ancestral phylogroup of *E. coli* (Lecointre *et al.*, 1998; Escobar-Páramo *et al.*, 2004; Sims and Kim, 2011), and their genetic background was proposed to favour expression and maintenance of laterally acquired genes (Escobar-Páramo *et al.*, 2004;

Chaudhuri and Thomas, 2007), mechanistic explanations however remain lacking. The distribution of specific virulence and fitness genes in *E. coli* pathovars and phylogenetic groups have been extensively studied (Dobrindt *et al*., 2003; Moulin-Schouleur *et al*., 2006; 2007; Ewers *et al*., 2007; Bauchart *et al*., 2010), but very little is known about possible differences in gene regulation between or within *E. coli* groups.

Newborn meningitis-causing *E. coli* (NMEC) is the most common cause of Gram-negative newborn meningitis and responsible for a third of the cases of the disease worldwide (Kaper *et al*., 2004; Gaschignard *et al*., 2011). NMECs associate with high rates of mortality and permanent neurological sequelae in survivors. NMEC strains mainly belong to the phylogenetic group B2 (Bingen *et al*., 1998; Johnson *et al*., 2008), and genetically closely similar strains possessing the K1 capsule antigen, especially the serotype O18:K1:H7 strains, are dominant (80–88%) among the NMEC isolates from different parts of the world (Robbins *et al*., 1974; Korhonen *et al*., 1985; Johnson *et al*., 2002; Gaschignard *et al*., 2011). *E. coli* K1 strains also occur in microbiota in healthy individuals (Sarff *et al*., 1975; Obata-Yasuoka *et al*., 2002), and neonates usually acquire the *E. coli* K1 infection through contact with the birth canal of a mother who is a carrier (Sarff *et al*., 1975). A number of potential NMEC-associated virulence or colonization factors have been reported, these include the K1 capsule, the sialyl galactoside-binding S-fimbria and the biofilm-associated Mat fimbria, a variant of the OmpA surface protein, as well as several invasion proteins and iron-acquisition systems (Bonacorsi and Bingen, 2005; Kim, 2008; Lehti *et al*., 2010).

The Mat (meningitis-associated and temperature-regulated) fimbria was identified in O18:K1:H7 NMEC isolates grown at 20°C (Pouttu *et al*., 2001). Later, expression of a structurally related fimbria was observed in pathogenic and non-pathogenic *E. coli* strains grown in the eukaryotic cell culture medium DMEM, and the name *E. coli* common pilus (ECP) was introduced (Rendón *et al*., 2007). The chromosomal 7 kb *mat* operon, also called *yag* or *ecp*, consists of six *matA–F* genes (Pouttu *et al*., 2001; Lehti *et al*., 2010), where *matA* encodes a putative transcription factor and *matB* the major fimbrillin subunit (Pouttu *et al*., 2001; Garnett *et al*., 2012; Lehti *et al*., 2012a,b). The MatB fimbrillin represents a very common and highly conserved surface structure in *E. coli* (Pouttu *et al*., 2001; Rendón *et al*., 2007). The *matB* gene has a prevalence of 91% in *c.* 600 *E. coli* isolates (Pouttu *et al*., 2001; Rendón *et al*., 2007; Blackburn *et al*., 2009; Saldaña *et al*., 2009; Avelino *et al*., 2010; Scaletsky *et al*., 2010; Hernandes *et al*., 2011), and the *mat* gene cluster has an identical chromosomal location as well as an overall 98% DNA sequence identity in the 45 fully sequenced genomes representing different pathovars and phylogenetic groups of *E. coli*. The predicted amino acid sequence of MatB is over 98% identical in the same genomes, and the gene cluster belongs to the so-called persistent genes (Fang *et al*., 2005; Touchon *et al*., 2009) that are present in nearly all isolates of a species. The *mat* cluster is not present in the sequence of currently available *Shigella* genomes or in the type strain of the closest *E. coli* relative, *Escherichia fergusonii*, which suggests that the cluster has been acquired after or during divergence of *E. coli* as a species. An intriguing property of the MatB fimbrillin is that its predicted structure bears only little homology to other fimbrillin structures of *E. coli* (Pouttu *et al*., 2001; Garnett *et al*., 2012). On the basis of the amino acid sequence of the putative usher protein MatD, the Mat fimbria belongs to the alternative chaperone-usher family (Nuccio and Bäumler, 2007), and recently EcpD/MatE was shown to be a tip-associated adhesin for epithelial cells (Garnett *et al*., 2012).

Considerable amount of evidence suggests that the Mat fimbria is an important colonization factor of *E. coli*. The fimbriae mediate attachment of intestinal *E. coli* to cultured human epithelial cells (Rendón *et al*., 2007; Lasaro *et al*., 2009; Saldaña *et al*., 2009; Avelino *et al*., 2010), are needed for biofilm formation by NMEC and uropathogenic *E. coli* (UPEC) (Lehti *et al*., 2010; Garnett *et al*., 2012), and are essential in colonization of infant mouse by the probiotic isolate *E. coli* Nissle 1917 (Lasaro *et al*., 2009). A critical role in biofilm formation is supported by that MatA represses the expression of the flagellar master operon *flhDC* and consequently prevents flagella-driven motility (Lehti *et al*., 2012b). Further, transcription of *mat* is regulated by the Rcs control system which affects several genes encoding surface structures, such as flagella, type-1-fimbria and capsule, that affect biofilm formation in enteric bacteria (Pratt and Kolter, 1998; Majdalani and Gottesman, 2005; Lehti *et al*., 2012a).

Current data suggest that the expression of Mat fimbria in *E. coli* isolates differs (Pouttu *et al*., 2001; Rendón *et al*., 2007). Analyses of Mat fimbria production in *c.* 500 human *E. coli* isolates have detected surface expression in 60% of pathogenic and commensal *E. coli* isolates (Pouttu *et al*., 2001; Rendón *et al*., 2007; Blackburn *et al*., 2009; Saldaña *et al*., 2009; Avelino *et al*., 2010; Hernandes *et al*., 2011). Two environmental cues inducing Mat expression have been reported: low growth temperature for O18:K1:H7 NMEC and DMEM with 5% CO_2 in the case of diarrhoeagenic *E. coli* (Pouttu *et al*., 2001; Rendón *et al*., 2007). In order to better understand the molecular basis behind the differential *mat* expression, we analysed *mat* regulatory DNA in *E. coli* phylogenetic groups and, in more details, in two well-characterized *E. coli* strains, the prototypical O18ac:K1:H7 serotype NMEC strain IHE 3034 (Achtman *et al*., 1983; Korhonen *et al*., 1985; Lehti *et al*., 2010; Moriel *et al*., 2010) and the non-pathogenic laboratory K-12 strain MG1655 (Blattner *et al*., 1997).

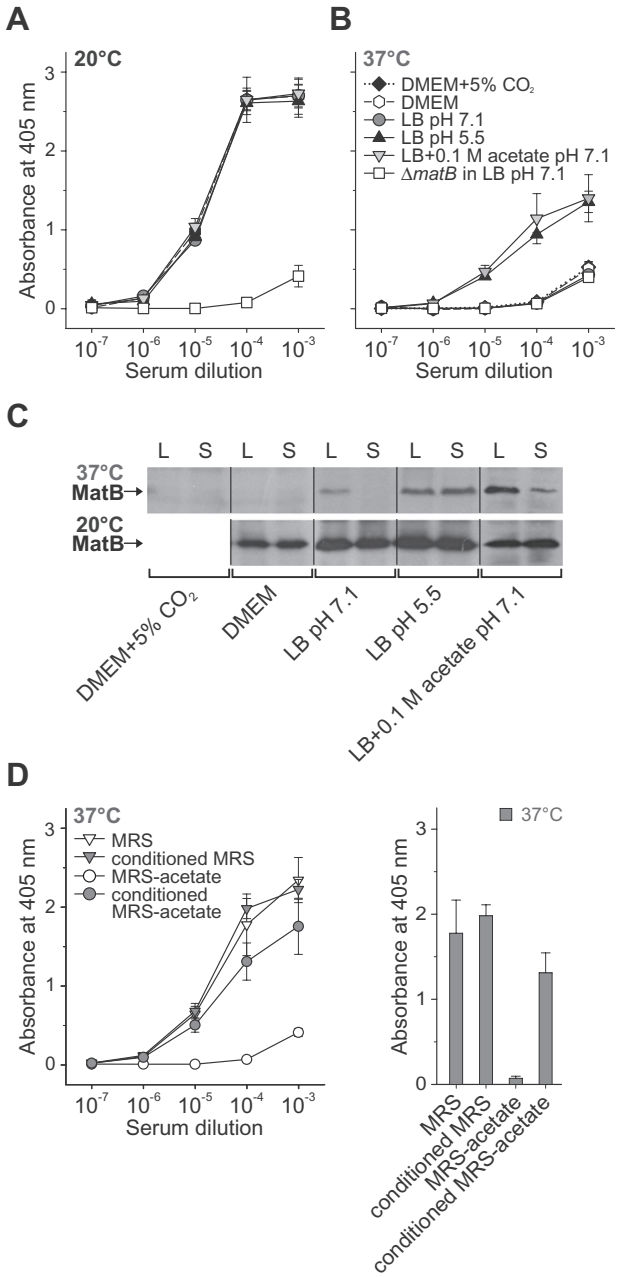

Fig. 1. Acidic pH, exogenous acetate and *Bifidobacterium longum*-conditioned medium, but not DMEM, promote Mat fimbriation in IHE 3034 at 37°C.
A and B. Surface expression of Mat fimbriae in stationary-phase cultures of IHE 3034-Rif was measured by whole-cell ELISA with anti-Mat fimbria antiserum as primary antibodies. The Mat fimbriae-deficient IHE 3034 *matB* deletion mutant was used as a negative control. The bacteria were grown to stationary phase in LB (pH 7.1), in LB with 0.1 M MES (pH 5.5) or with 0.1 M acetate (pH 7.1), and in DMEM (pH 7.2) with and without 5% CO_2 at 20°C (A) and 37°C (B). Cell densities were prior to analysis normalized at OD_{600}. Shown are mean and SD values of three independent assays.
C. Cellular levels of MatB fimbrillin in IHE 3034 at 20°C and 37°C. HCl-treated whole-cell protein samples were prepared from OD_{600} normalized mid-logarithmic (L) and stationary (S) phase cultures and subjected to Western blot analysis using anti-Mat fimbria antiserum as primary antibodies.
D. Effect of *B. longum*-conditioned culture medium on surface expression of Mat fimbriae in IHE 3034 at 37°C as measured by whole-cell ELISA. The strain IHE 3034 was grown to stationary phase in either fresh or conditioned MRS broth or acetate-depleted MRS broth. For clarity, data from an antiserum dilution of 1:10 000 is shown in the right panel. The data represent mean and SD values of two independent assays.

Results

Low pH and high acetate concentration trigger Mat fimbriae expression at 37°C in the NMEC isolate IHE 3034

To analyse the effect of environmental cues on expression of Mat fimbria, we selected the strain IHE 3034 of O18ac:K1:H7 serotype, isolated from a case of fatal late onset neonatal meningitis in Finland in 1976 (Korhonen *et al*., 1985; Selander *et al*., 1986). The strain is one of the best-characterized NMEC strains, belongs to the virulence-associated phylogenetic group B2, and expresses Mat fimbriae at 20°C in Luria–Bertani (LB) broth and M63 medium (Pouttu *et al*., 2001; Lehti *et al*., 2010; Moriel *et al*., 2010). For comparison, we chose the A-phylogroup strain M1655 that carries the *mat* gene cluster at the same chromosomal site as IHE3034 (Blattner *et al*., 1997) but does not express Mat under the same growth conditions (Pouttu *et al*., 2001; Lehti *et al*., 2010).

Earlier work had shown that IHE 3034 expresses Mat fimbriae in LB broth at 20°C but not at 37°C and that iron, osmolarity or human blood do not affect the expression (Pouttu *et al*., 2001). On the other hand, the Mat homologue of an enterohaemorrhagic (EHEC) O157:H7 strain EDL933 was not expressed in LB broth but in the eukaryotic-cell culture medium DMEM, and the production was stimulated by 5% CO_2. In particular, these conditions allowed expression also at 37°C, and the same culture conditions favoured fimbria expression in other diarrhoeagenic pathovars as well (Rendón *et al*., 2007; Blackburn *et al*., 2009; Saldaña *et al*., 2009; Avelino *et al*., 2010). In order to analyse the effect of DMEM medium and presence of 5% CO_2 on the expression of Mat fimbria by IHE 3034, we quantified Mat fimbria on the bacterial surface by whole-cell ELISA using anti-Mat fimbria antibodies (Fig. 1A and B).The assay showed that growth in DMEM or in DMEM + 5% CO_2 did not enhance production of the Mat fimbriae at 20°C (Fig. 1A) or 37°C (Fig. 1B) by IHE 3034. The results obtained by examination of the total cellular content of MatB fimbrillin by Western blotting (Fig. 1C) were consistent with the ELISA results.

We used whole-cell ELISA to identify conditions that trigger Mat fimbria expression on IHE 3034 at 37°C. The strain was cultivated with different carbohydrates, amino

acids and short-chain fatty acids as a carbon source, and at various pH values. Low pH (5.5) and high acetate concentration (50–100 mM, at pH 7.1), supplemented in M9 minimal medium (not shown) or LB broth, stimulated Mat fimbriation at 37°C (Fig. 1A–C), whereas addition of lactate did not affect Mat expression (not shown). The combined effect of high acetate and low pH on Mat fimbria production in IHE 3034 could not be reliably analysed due to the poor bacterial growth. When individually administered, high acetate at pH 7.1 and low pH did not significantly decrease the growth rate.

Transmission of NMEC infection involves colonization of the female genital tract and the intestine, which have a high acetate concentration (Cummings *et al*., 1987; Mirmonsef *et al*., 2011) arising from bacterial fermentation of dietary carbohydrates. To mimic the *in vivo* situation *in vitro*, we tested Mat-fimbriae production by IHE 3034 after growth at 37°C in acetate-depleted MRS broth conditioned by *Bifidobacterium longum* cells. For comparison, fimbriation was assessed after growth in sterile acetate-depleted MRS medium, and in sterile MRS medium that contains 60 mM acetate. In the *B. longum* conditioned medium and in the sterile medium with acetate, IHE 3034 expressed Mat fimbriae, whereas no Mat fimbriae were seen on cells grown in modified MRS medium containing no acetate (Fig. 1D).

Expression of Mat fimbriae correlates with the mat upstream DNA sequence

We next analysed whether the Mat expression response to low pH and high acetate is common in *E. coli* isolates. The strains were selected because their genomic sequences were available from databases or their genetic content has been analysed by comparative genomic hybridization (Grozdanov *et al*., 2004; Zdziarski *et al*., 2008; Bauchart *et al*., 2010). In LB broth, the 11 B2 isolates produced Mat fimbriae at low temperature but not at 37°C, whereas the B1 and A strains did not express the fimbria (Fig. 2A). The levels of Mat fimbria production varied within the B2 group. In 5/11 B2 strains, addition of 0.1 M acetate to the pH 7.1 culture medium increased Mat expression at 37°C, and in 7/11 B2 strains, low pH also stimulated Mat fimbriation. MG1655 did not produce Mat fimbria on the bacterial surface at any of the growth conditions (Fig. 2A) or when grown at 37°C in DMEM or DMEM + 5% CO_2 medium, neither did we detect the MatB fimbrillin in MG1655 cells (not shown).

We compared the *mat* upstream region (−608 bp to −1 bp relative to the translational start of *matA*), including the putative *mat* promoter, in the 48 *E. coli* isolates, from which 41 were publicly available and seven were sequenced in this study. Sequence logo analysis of these nucleotide sequences showed a high conservation across isolates and no clear clusters of mismatches (Fig. S1). However, two distinct lineages, which share 91.3–91.8% sequence identity, were identified (Fig. 2B, left panel). The first lineage contained the B2, the D and the E strains, and the other the A and 8/9 of the B1 strains. The sequences show slight variability within the B2 group; those in the O18:K1:H7 isolates cluster together and also the O157:H7 EHEC strains are in a single cluster. In the A/B1 lineage the bulk of nucleotide polymorphisms, which differentiate the lineage from the B2/D/E lineage, are fixed within the two subgroups, and only a few singletons are present (not shown). The prevalence of sequence polymorphisms in the *mat* upstream region is three times higher than the average genetic diversity in the *matA–F* coding regions (Table S1; for a scheme of the gene cluster, see Fig. 3A). The nucleotide sequence from *matA* to *matF* is highly conserved and gives a less divergent tree topology than what the *mat* upstream region does (Fig. 2B, right panel). The majority of polymorphic changes in the *matA–F* region are shared within A and B1 strains, and partially with E strains. Ratio of non-synonymous (dN, amino acid replacement) to synonymous (dS, silent) substitutions in the *mat* genes indicated that the *matA–F* genes are under purifying selection (dN/dS < 1) (Table S1). Thus, the Mat fimbriae expression and the *mat* upstream regions differ in the B2/D/E compared with the B1/A strains, whereas the genes encoding fimbrial subunit proteins and components of its biosynthetic machinery are well preserved. In particular, IHE 3034 and MG1655 represent extreme cases in regard to Mat expression (Fig. 2A) and primary structure of the *mat* upstream region that contains the putative *mat* promoters (for a sequence alignment, see Fig. S1).

Mat expression in IHE 3034

Subsequently, we characterized the *mat* transcripts in IHE 3034 cells grown to logarithmic and stationary phase at 20°C and 37°C, by Northern blot analysis using probes complementary to each *mat* gene (for a scheme of the gene cluster, see Fig. 3A). Only one transcript, 0.67 kb in length, in the RNA harvested from IHE 3034-Rif cells grown at 20°C, specifically hybridized with a *matB* probe (Fig. 3B). The *matB* transcription was detected at early logarithmic phase and remained high throughout logarithmic growth until transition to the stationary growth phase. On the basis of 5′ end mapping of the *matB* transcript by primer extension analysis (Fig. 3C) and prediction of the position of transcription terminators, the length of the monocistronic *matB* message was calculated to be 677 bp. Probes complementary to *matA* (Fig. 3B), *matC*, *matD*, *matE* or *matF* (data not shown) failed to detect specific mRNAs, indicating that these transcripts are unstable and/or of very low abundance.

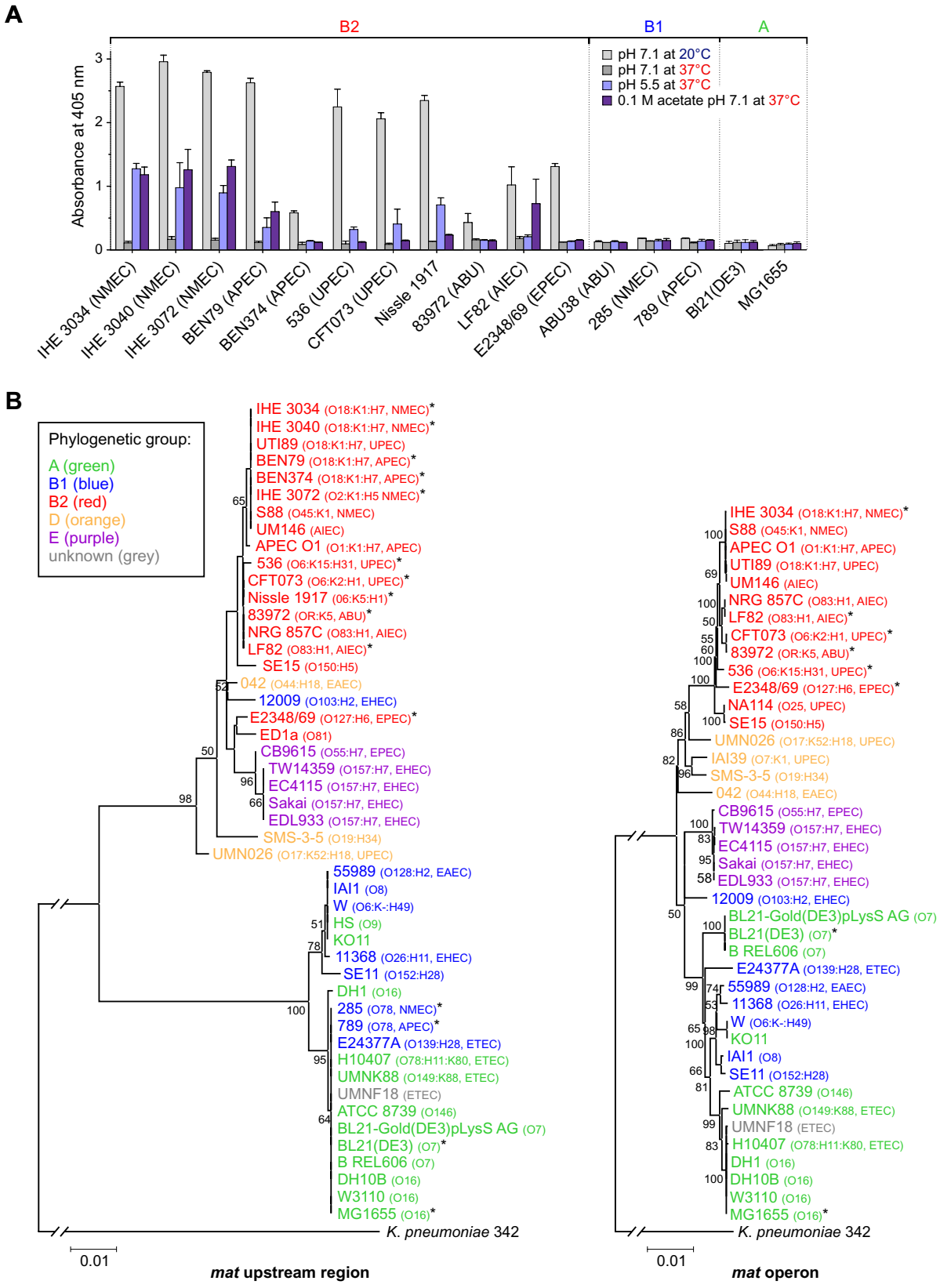
A
B2
B1
A
pH 7.1 at 20°C
pH 7.1 at 37°C
pH 5.5 at 37°C
0.1 M acetate pH 7.1 at 37°C
Absorbance at 405 nm
IHE 3034 (NMEC)
IHE 3040 (NMEC)
IHE 3072 (NMEC)
BEN79 (APEC)
BEN374 (APEC)
536 (UPEC)
CFT073 (UPEC)
Nissle 1917
83972 (ABU)
LF82 (AIEC)
E2348/69 (EPEC)
ABU38 (ABU)
285 (NMEC)
789 (APEC)
BI21(DE3)
MG1655
B
Phylogenetic group:
A (green)
B1 (blue)
B2 (red)
D (orange)
E (purple)
unknown (grey)
IHE 3034 (O18:K1:H7, NMEC)*
IHE 3040 (O18:K1:H7, NMEC)*
UTI89 (O18:K1:H7, UPEC)
BEN79 (O18:K1:H7, APEC)*
BEN374 (O18:K1:H7, APEC)*
IHE 3072 (O2:K1:H5 NMEC)*
S88 (O45:K1, NMEC)
UM146 (AIEC)
APEC O1 (O1:K1:H7, APEC)
536 (O6:K15:H31, UPEC)*
CFT073 (O6:K2:H1, UPEC)*
Nissle 1917 (06:K5:H1)*
83972 (OR:K5, ABU)*
NRG 857C (O83:H1, AIEC)
LF82 (O83:H1, AIEC)*
SE15 (O150:H5)
042 (O44:H18, EAEC)
12009 (O103:H2, EHEC)
E2348/69 (O127:H6, EPEC)*
ED1a (O81)
CB9615 (O55:H7, EPEC)
TW14359 (O157:H7, EHEC)
EC4115 (O157:H7, EHEC)
Sakai (O157:H7, EHEC)
EDL933 (O157:H7, EHEC)
SMS-3-5 (O19:H34)
UMN026 (O17:K52:H18, UPEC)
55989 (O128:H2, EAEC)
IAI1 (O8)
W (O6:K-:H49)
HS (O9)
KO11
11368 (O26:H11, EHEC)
SE11 (O152:H28)
DH1 (O16)
285 (O78, NMEC)*
789 (O78, APEC)*
E24377A (O139:H28, ETEC)
H10407 (O78:H11:K80, ETEC)
UMNK88 (O149:K88, ETEC)
UMNF18 (ETEC)
ATCC 8739 (O146)
BL21-Gold(DE3)pLysS AG (O7)
BL21(DE3) (O7)*
B REL606 (O7)
DH10B (O16)
W3110 (O16)
MG1655 (O16)*
K. pneumoniae 342
0.01
mat upstream region
IHE 3034 (O18:K1:H7, NMEC)*
S88 (O45:K1, NMEC)
APEC O1 (O1:K1:H7, APEC)
UTI89 (O18:K1:H7, UPEC)
UM146 (AIEC)
NRG 857C (O83:H1, AIEC)
LF82 (O83:H1, AIEC)*
CFT073 (O6:K2:H1, UPEC)*
83972 (OR:K5, ABU)*
536 (O6:K15:H31, UPEC)*
E2348/69 (O127:H6, EPEC)*
NA114 (O25, UPEC)
SE15 (O150:H5)
UMN026 (O17:K52:H18, UPEC)
IAI39 (O7:K1, UPEC)
SMS-3-5 (O19:H34)
042 (O44:H18, EAEC)
CB9615 (O55:H7, EPEC)
TW14359 (O157:H7, EHEC)
EC4115 (O157:H7, EHEC)
Sakai (O157:H7, EHEC)
EDL933 (O157:H7, EHEC)
12009 (O103:H2, EHEC)
BL21-Gold(DE3)pLysS AG (O7)
BL21(DE3) (O7)*
B REL606 (O7)
E24377A (O139:H28, ETEC)
55989 (O128:H2, EAEC)
11368 (O26:H11, EHEC)
W (O6:K-:H49)
KO11
IAI1 (O8)
SE11 (O152:H28)
ATCC 8739 (O146)
UMNK88 (O149:K88, ETEC)
UMNF18 (ETEC)
H10407 (O78:H11:K80, ETEC)
DH1 (O16)
DH10B (O16)
W3110 (O16)
MG1655 (O16)*
K. pneumoniae 342
0.01
mat operon

Fig. 2. Comparison of Mat fimbriae expression, *mat* upstream sequences and *mat* operon sequences in *E. coli* isolates.
A. Surface expression of Mat fimbriae in *E. coli* isolates ($n = 16$) was measured by whole-cell ELISA with anti-Mat fimbria antiserum (a dilution of 1:10 000). Bacteria were grown to stationary phase in LB (pH 7.1) at 20°C and 37°C, LB with 0.1 M MES (pH 5.5), and in LB with 0.1 M acetate (pH 7.1) at 37°C. The phylogenetic main groups are indicated at the top, the strain pathotype is given in parentheses, and when no pathotype is indicated the strain is a commensal.
B. Neighbour-joining trees of putative *mat* promoter sequences (–608 bp to –1 bp from the start codon GTG of *matA*) ($n = 48$) in the left panel and of *mat* operon sequences ($n = 41$) in the right panel of selected *E. coli* isolates. Construction of the neighbour-joining trees was based on Tamura-Nei model. The *mat* upstream sequence and the *mat* operon sequence of *Klebsiella pneumoniae* were used as outgroups. The phylogenetic tree topology was evaluated by bootstrap analysis (1000 replicates), and confidence values greater than 50% are shown. The scale bar represents 0.01 base substitutions per site. The phylogenetic main groups for *E. coli* are coded by colours; serotype of each strain (if known) and the pathotype (for pathogenic isolates) is given in parentheses. Asterisks indicate strains used in the ELISA assay in (A). ABU, asymptomatic bacteriuria; AIEC, adherent-invasive *E. coli*; APEC, avian pathogenic *E. coli*; EAEC, enteroaggregative *E. coli*; EHEC, enterohaemorrhagic *E. coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*; NMEC, neonatal meningitis *E. coli*; UPEC, uropathogenic *E. coli*.

We used reverse transcription-PCR (RT-PCR) to analyse possible co-transcription of the *mat* genes. The primers were designed to amplify intergenic junctions of co-transcribed genes (Fig. 3A). Amplicons of the expected size and extending from *matA* to *matC, matC* to *matD*, *matD* to *matE* and from *matE* to *matF* were detected from the cDNA template (Fig. 3D). An amplicon of expected length was also observed for the housekeeping gene *gapA* used as a reference, whereas no DNA was visible when RT-PCR was performed without a reverse transcription.

Quantitative real-time RT-PCR (qRT-PCR) analysis of the transcription of *mat* genes in logarithmic-phase IHE 3034-Rif cells with *frr* and *gapA* as references, showed that all six *mat* genes were transcribed at 20°C, but at varying efficiency. Consistent with RT-PCR results, the transcription levels of *matA* and *matB* were highest, whereas the efficacy of *matC*, *matD* and *matE* transcription was significantly lower and *matF* was transcribed only weakly (Fig. 3E). The level of *matA* transcription was 38-fold higher than that of *frr* and approximately one fourth of the *gapA* transcription level. Taken together, the results indicate that the *mat* determinant is transcribed as a single polycistronic transcript and that the proper stoichiometric expression of different translational units in the polycistronic mRNA is facilitated through partial termination of transcription, ribonucleolytic processing of precursor transcript and/or differential stability of mRNA segments. Instability and/or rapid degradation of the 5′ end of the *matA* transcript are potential reasons for the observed discrepancy between the Northern blot (Fig. 3B), in which a DNA probe complementary to the entire *matA* coding region was used, and the RT-PCR (Fig. 3D) as well as the qRT-PCR (Fig. 3E) which detect only the 3′ half of *matA* (Fig. 3A).

The mat promoter activity depends on growth temperature

To identify promoter regions within the *mat* operon, 0.7 kb fragments containing 5′ upstream region and ~ 0.1 kb of coding sequence of each *mat* gene were fused upstream of the promoterless *lacZ* gene in vector pRS551 for transcriptional reporter analysis. The promoter activities were determined in IHE 3034-Rif cells grown to logarithmic phase at 20°C or 37°C. The transcriptional fusion *pA–lacZ* (–608 to +81 bp relative to *matA* start codon) exhibited a high level of β-galactosidase activity (Fig. 4). The reporter activity was stronger in cells grown at 20°C than at 37°C. The expression of the fusions containing upstream regions from the other *mat* genes (*pB–lacZ*, *pC–lacZ*, *pD–lacZ*, *pE–lacZ* and *pF–lacZ*) was at the level of the vector control (Fig. S2). The pRS551 vector plasmid harbours a weak transcriptional terminator that might interfere with the assays (Liang *et al.*, 1998; Repoila and Gottesman, 2001). We deleted *trpt* from the *lacZ* leader sequence in pRS551. The *pA* promoter exhibited significant (P 0.0013) temperature regulation also in the absence of *trpt* (Fig. S2), confirming that the temperature effects indeed depend on the inserted *mat* promoter element.

We mapped the 5′ end(s) of the polycistronic *mat* mRNA by 5′ rapid amplification of cDNA ends (RACE) and identified three different 5′-termini upstream of *matA*. The first transcriptional start site (T1) is located at –56 bp, the second (T2) at –122 bp and the third (T3) at –344 bp upstream of the *matA* translation start site, and each of them is preceded by potential –35 and –10 promoter motifs (Fig. 4). The –35 and –10 elements upstream of T2 in the *pP2* promoter region most closely resemble the σ^{70} consensus sequences TTGACA and TATAAT (Harley and Reynolds, 1987), containing 10 of the 12 canonical bases of the hexamers. In the proximal *pP1* promoter, eight out of 12 base pairs fit the consensus sequences, while the elements in the distal *pP3* promoter with five mismatches have the weakest similarity to the promoter motifs.

To characterize the active site in the *mat* upstream region (*pA*) in more detail, we constructed *lacZ* fusions of each one of the three putative promoters (encoded in reporter plasmids *pP1–pP3*); a procedure that allowed monitoring of their relative strength. High transcriptional activity at 20°C was primarily associated with the second transcriptional start point T2, which supported 208- and 21-fold higher levels of expression than those observed

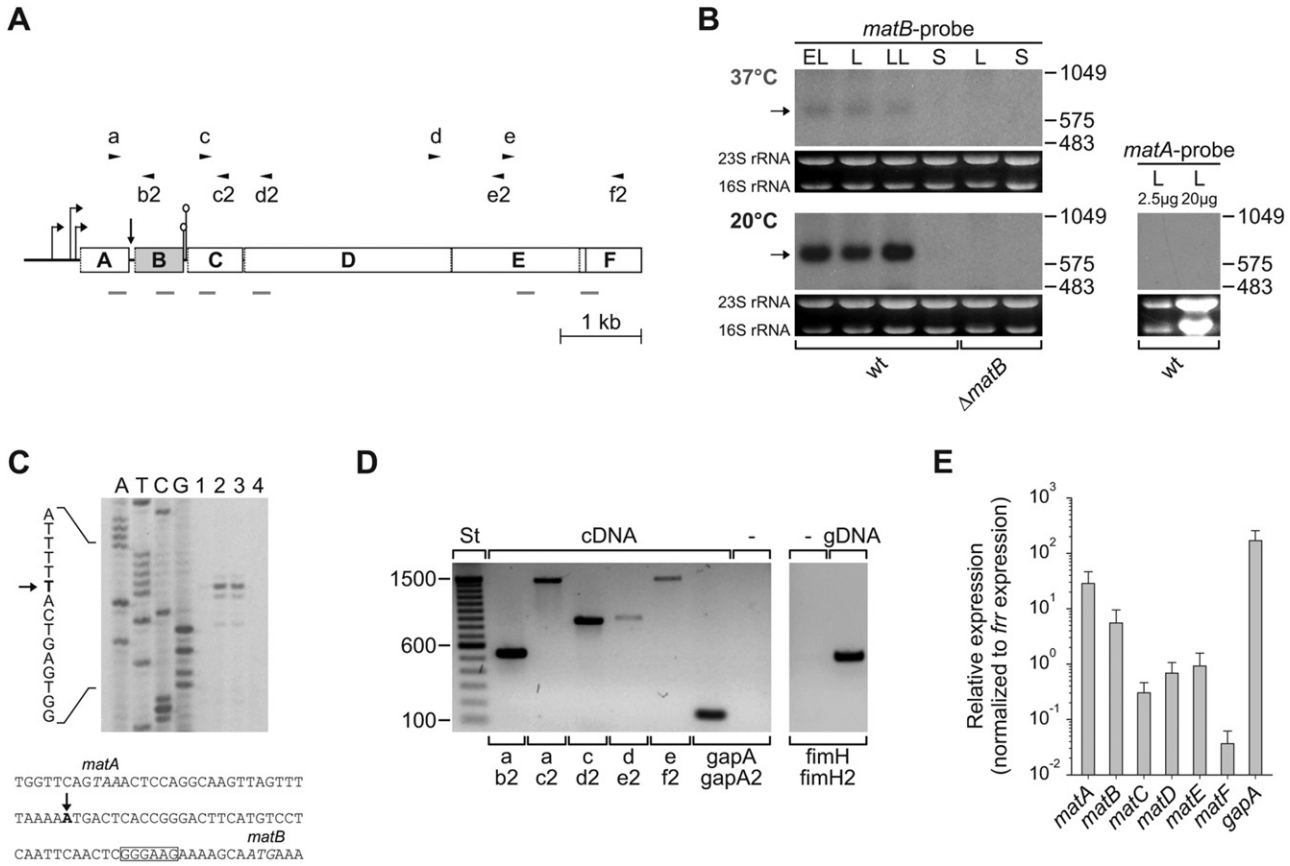

Fig. 3. The major mRNA encoding the MatB fimbrillin is processed from a precursor transcript *matABCDEF*.
A. The *mat* locus of *E. coli*. The *matB* gene encoding the major fimbrillin (Pouttu *et al*., 2001; Garnett *et al*., 2012) is shown in grey. Right-angled arrows indicate the three transcription start sites identified by 5′ RACE in this study. The vertical arrow shows the 5′ end of the monocistronic *matB* mRNA, and stem-loop symbols show putative transcriptional terminators downstream of *matB*. The primers employed in RT-PCR are indicated by horizontal arrowheads and the amplicons of the qRT-PCR are indicated below the *mat* locus.
B. Northern blot analysis of *matB* and *matA* mRNA in total RNA samples derived from IHE 3034-Rif cells grown in LB to early-logarithmic (EL), mid-logarithmic (L), late-logarithmic (LL) and stationary (S) phases at 20°C or 37°C. The amount of total RNA used was 2.5 µg, the *matA* Northern blot was in addition performed using 20 µg of RNA. The position of the *matB* transcript is indicated with arrows, and the length (bp) of DIG-labelled RNA markers are shown on the right. The *matB* deletion derivative of IHE 3034-Rif served as a negative control and ethidium bromide-stained 23S and 16S rRNA as loading controls.
C. Primer extension analysis for determination of the 5′ end of the *matB* transcript was carried out on total RNA (lane 1: 1 µg of RNA, lanes 2–3: 10 µg of RNA and lane 4: no RNA) extracted from IHE 3034-Rif cells. The primary 5′ end is marked as a black arrow and a bold letter. The nucleotide sequence of the *matA*–*matB* intergenic region is shown, with italic letters indicating the stop codon of *matA* and the start codon of *matB*. The putative ribosomal binding site is boxed.
D. RT-PCR analysis of the transcriptional organization of the *mat* operon in IHE 3034. In the cDNA lanes, DNase-digested RNA was subjected to reverse transcriptase synthesis and cDNA amplification. For controls, DNase-treated RNA without a reverse transcription step (– lane) or genomic IHE 3034 DNA (gDNA lane) were used as template in the PCR reaction. Primer pairs used are shown below the lanes, and the primer positions are indicated in (A). Lane St, DNA standard (bp).
E. Transcript levels of *mat* genes relative to the *frr* gene in IHE 3034-Rif grown to mid-logarithmic phase at 20°C as quantified by qRT-PCR using equal amounts of RNA samples. Shown are the mean and SD values of three independent assays.

with *pP1* and *pP3* promoters (Fig. 4). The short promoter *pP1* could not drive expression alone and the activity level resembled that of the vector control. The *pP3* promoter was weakly active at 37°C. Pairing of the promoter *pP2* with *pP1* or *pP3* or including all three transcriptional start sites in *pP123* increased reporter expression at 20°C only modestly above that of the *pP2*. We concluded that *pP2* represents the major promoter for *mat* expression.

Swapping the chromosomal mat promoter regions induces Mat expression in E. coli *MG1655 and silences expression in IHE 3034*

To analyse whether the lineage-restricted polymorphisms in the *mat* regulatory region is a primary determinant of the phylogroup-specific *mat* expression, we chose the phylogroup A strain MG1655, which carries functional

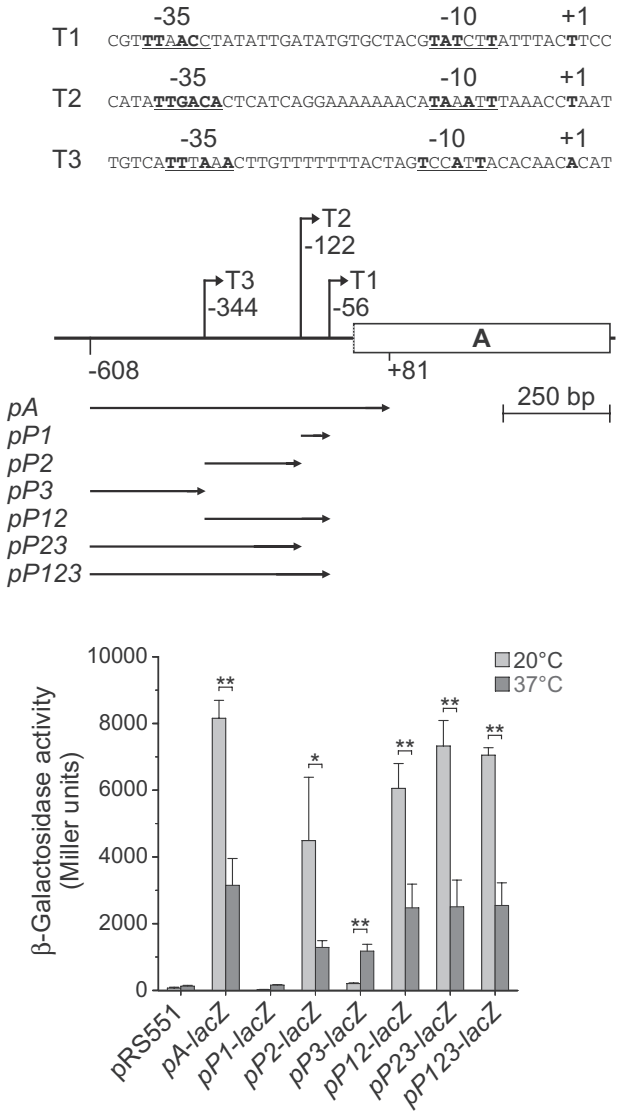

Fig. 4. The promoter of the *mat* operon is subject to thermoregulation in IHE 3034. The three transcriptional start sites (T1, T2, T3; marked as +1 and in bold letters) identified by 5′ RACE are shown. The predicted –10 and –35 promoter sequences are underlined and the bases matching the canonical hexamers of the σ^{70}-type promoters, TTGACA and TATAAT, are indicated in bold. In the scheme below, T1 to T3 are indicated with right-angled arrows and numbers correspond to the distances from the GTG start codon of *matA*. The *mat* upstream DNA fragments used to construct transcriptional *lacZ* fusions are indicated at the bottom of the scheme. The strength of individual and combined promoters in the *mat* upstream region was measured by β-galactosidase activity. The analyses were performed using logarithmic-phase cultures of IHE 3034-Rif harbouring pRS551-based reporter plasmids and grown in LB broth at 20°C and 37°C. The data represent mean and SD values of at least three independent experiments. Significance is indicated as follows: one asterisk, $P < 0.05$; two asterisks, $P < 0.005$; as calculated by Student's *t*-test.

matBCDEF genes but is incapable of Mat fimbriae expression (Lehti *et al.*, 2010), for detailed comparison with IHE 3034. The sequence –608 to +10 bp upstream of *matA* of IHE 3034 and MG1655 share an overall 91.1% sequence identity. Comparison of the putative promoter regions (Figs 5A and S1) revealed that in MG1655, the spacing between the –10 and –35 elements of the major *pP2* promoter is suboptimal as it contains an extra adenine and the T2 site starts with cytosine that is less frequent as an initiation nucleotide than the thymidine present in IHE 3034 (Mendoza-Vargas *et al.*, 2009). In *pP3* and *pP1*, only a few substitutions were found. We tested whether the upstream DNA sequence of MG1655 can direct *lacZ* expression. The level of β-galactosidase activity at 20°C was low in IHE 3034-Rif cells harbouring the pA_{MG1655}-containing reporter plasmid as compared with cells containing the $pA_{IHE\ 3034}$ fusion (Fig. 5B). The induction driven by the promoter variants were similar at 37°C. These findings provide evidence that the promoter sequence variation is a reason for the inability of MG1655 to produce Mat fimbriae.

To further characterize the differential *mat* expression between IHE 3034 and MG1655, we next performed a series of promoter swapping tests where the specific parts of chromosomal *mat* upstream region from one strain were replaced *in cis* with the corresponding regions from the other strain, and the expression levels of Mat fimbriae were analysed. Our promoter-swapping strategy allows precise engineering of promoter elements using conventional *sacB*-based counter-selection system for unmarked allelic exchange while retaining the *mat* genes as single copies at their normal chromosomal location. The regions to be swapped were selected on the basis of the location of transcription start sites and the recognized promoter activities in the *mat* upstream region of IHE 3034, as well as the levels of nucleotide sequence identity between the strains (Figs 4 and 5A).

First, we determined the effect of MG1655-derived sequences in the IHE 3034 background (Fig. 5C). The $S1_{MG1655}$ and the $S2_{MG1655}$ regions immediately upstream of the *matA* gene did not support fimbriation, whereas the replacement of the S3 region, next to T3 in IHE 3034, or the S4 region or their combination (S34) caused only minor reduction on the Mat expression. These results support the notion that the mat_{MG1655} promoter region contains mutations that keep the *mat* operon silent.

For a gain-of-function mutagenesis, the S1 and S2 regions of MG1655 were swapped in combination (S12) to the corresponding IHE 3034 sequence. This change did not convert the strain MG1655 into a Mat producer (Fig. 5C). Also, introducing the S123 fragment from IHE 3034 into MG1655 did not support Mat fimbriation at a detectable level. When the S1234 (–608 to +10 bp) was changed into IHE 3034 sequence, the strain MG1655

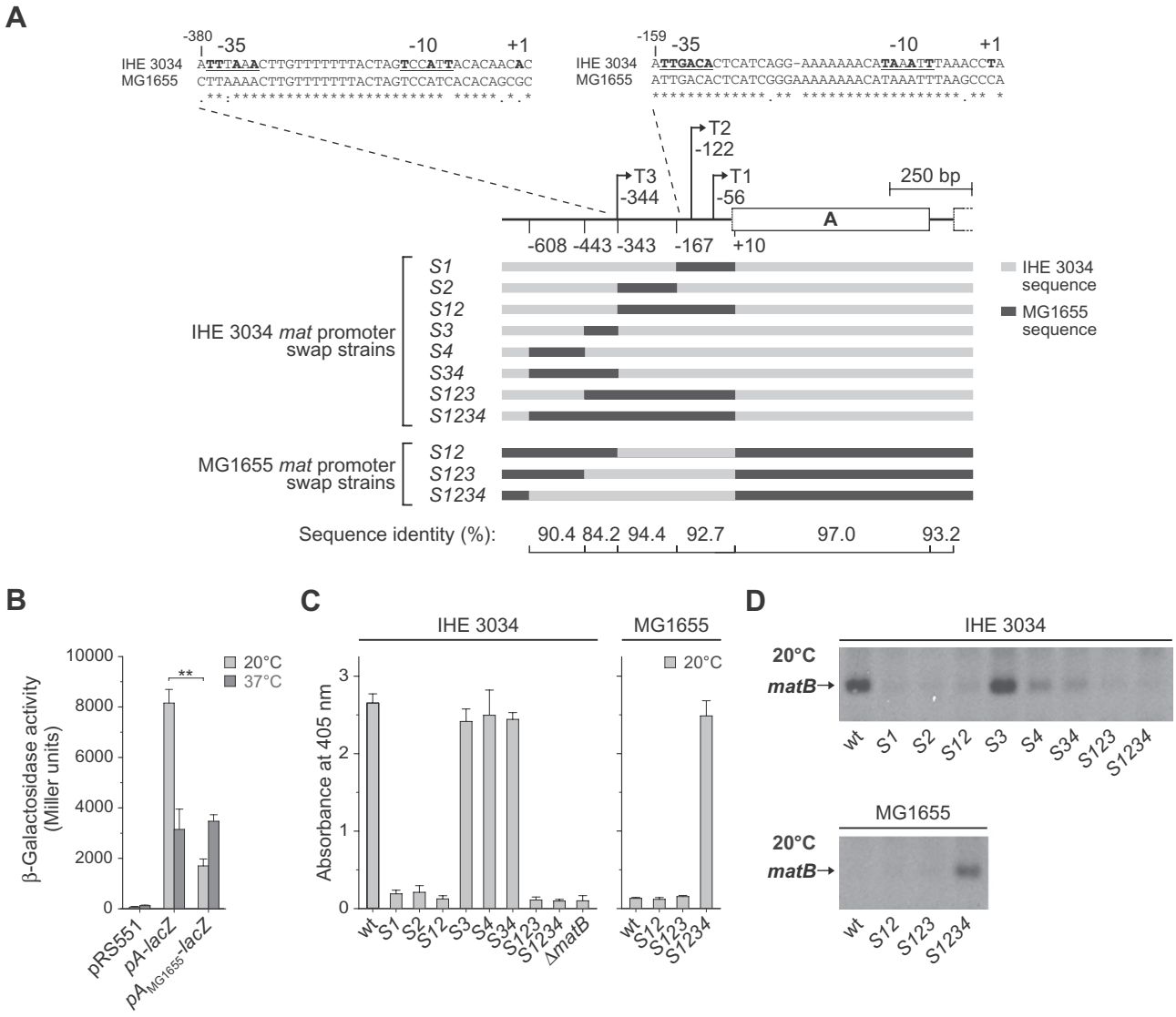

Fig. 5. The *mat* promoter architecture prevents expression of Mat fimbriae in MG1655.

A. DNA sequences upstream of the *mat* transcriptional start sites T2 and T3 in IHE 3034 and MG1655, and a scheme presenting the DNA fragments used to construct promoter swap strains. The predicted −10 and −35 promoter sequences are underlined and the bases matching the canonical hexamers of the σ^{70}-type promoters, TTGACA and TATAAT, are indicated in bold. In the scheme, the transcriptional start sites (T1, T2, T3) are indicated with right-angled arrows and numbers correspond to the distances from the translational start codon of *matA*. The regions for the promoter swapping between IHE 3034 and MG1655 genomes are shown as black or grey boxes, and the sequence identities (%) of the regions between the two strains are indicated below the scheme.

B. Comparison of the expression levels of the *mat* promoters from IHE 3034 (*pA*) and MG1655 (pA_{MG1655}) as measured by β-galactosidase activity. The analyses were performed using logarithmic-phase cultures of IHE 3034-Rif harbouring pRS551-based reporter plasmids and grown in LB at 20°C and 37°C. Shown are mean and SD values of at least three independent assays. The two asterisks mark *P*-value of < 0.005 as calculated by Student's *t*-test.

C. Surface expression of Mat fimbriae by promoter swap strains. The IHE 3034 derivatives including a *matB* deletion mutant as a negative control (left panel) and the MG1655 derivatives (right panel) were analysed by whole-cell ELISA with anti-Mat fimbria antiserum as primary antibodies (a dilution of 1:10 000) after growth to stationary phase in LB medium at 20°C. Shown are mean and SD values of at least two independent assays.

D. Transcription of fimbrillin-encoding *matB* by IHE 3034 and MG1655 promoter swap strains was analysed by Northern blotting with a *matB*-specific probe. Total RNA was isolated from cells grown in LB broth to mid-logarithmic phase at 20°C. The position of the monocistronic *matB* message is indicated.

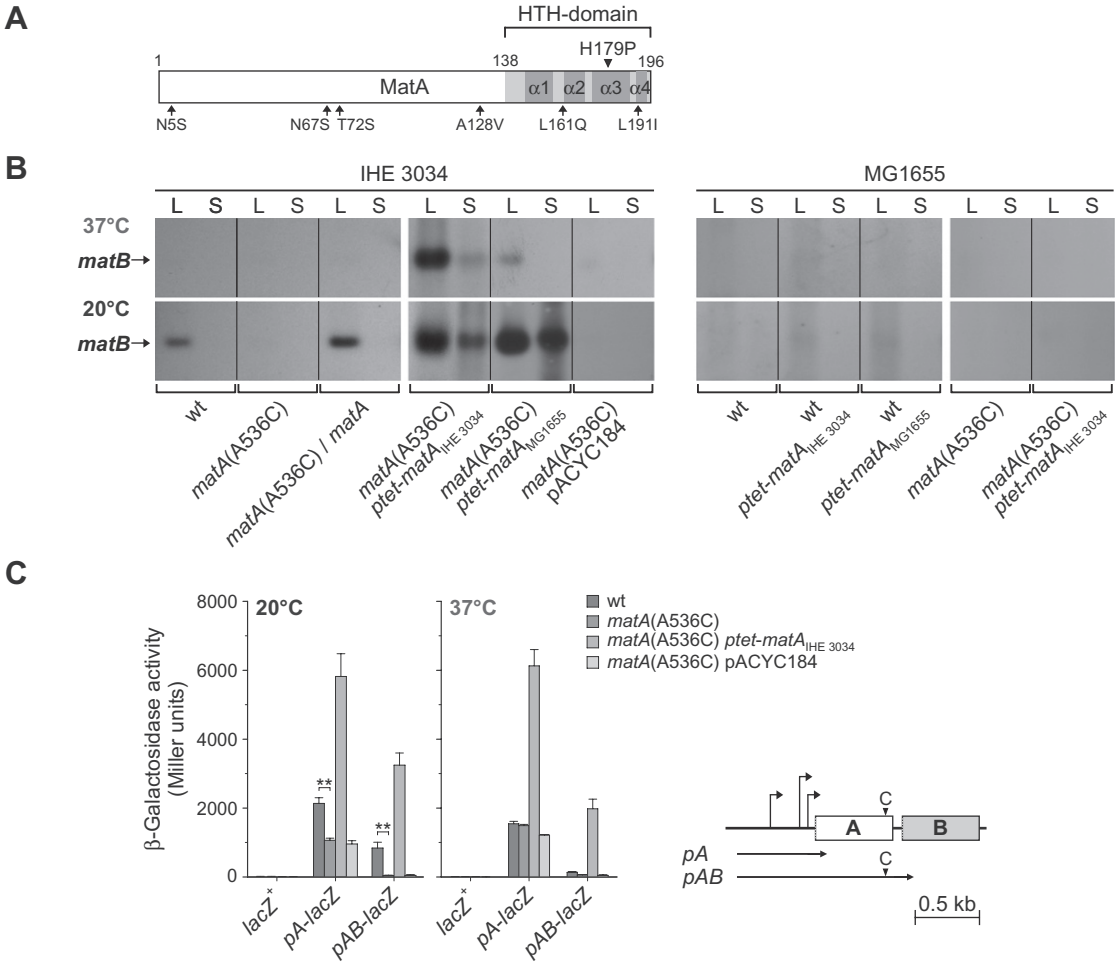

Fig. 6. Influence on *mat* promoter activity by MatA.
A. Schematic representation of the MatA protein of IHE 3034 and MG1655. The putative LuxR-type four-helical HTH DNA-binding motif is shaded in light grey, and alpha helices of HTH are shown in dark grey. The location of H179P (resulting from A536C nucleotide substitution) substitution is indicated by an arrowhead. The arrows indicate amino acid substitutions observed in MG1655.
B. Transcription of *matB* by IHE 3034 and MG1655 derivatives carrying missense mutation within *matA* was analysed by Northern blotting with a *matB*-specific probe. The *matA*(A536C) substitution in IHE 3034-Rif was complemented *in cis* with the wild-type allele; the complementation was performed *in trans* with the *matA* gene from IHE 3034 or MG1655 in IHE 3034-Rif and MG1655-Rif derivatives. Total RNA was isolated from cells grown in LB broth to mid-logarithmic (L) and stationary (S) phase at 20°C and 37°C. The position of the monocistronic *matB* message is indicated with arrows.
C. Effect of *matA*(A536C) mutation and $MatA_{IHE\ 3034}$ overexpression on the activities of *pA* and *pAB* promoter regions in IHE 3034-Rif derivatives. Parent strain and reporter strains containing a *pA*– or *pAB*–*lacZ* fusion inserted at the *lac* operon were grown to mid-logarithmic phase in LB broth at 20°C and 37°C, and assayed for β-galactosidase activity. The outline of the IHE 3034 DNA fragments used to construct single-copy transcriptional *lacZ* fusions is shown on the right. The location of introduced A536C missense mutation in chromosomal *matA* and *matA* in the *pAB* fragment is indicated by a C letter and a black triangle in the scheme. Shown are the mean and SD values of at least three independent assays. The two asterisks mark *P*-value of < 0.005 as calculated by Student's *t*-test.

expressed Mat fimbriae similarly to IHE 3034. Northern blot analysis verified the results (Fig. 5D). Taken together, the *mat* coding region is intact in MG1655 whereas the *mat* upstream region, especially in the S12 region, carries nucleotides which disrupt fimbrial expression under the test conditions.

MatA activates matB expression

MatA was inactivated by the chromosomal substitution *matA*(A536C), which encodes MatA H179P and disrupts the C-terminal helix–turn–helix (HTH) DNA-binding motif (Barlow and Thornton, 1988) (Fig. 6A). The substitution totally abolished transcription of *matB* in IHE 3034 (Fig. 6B) and expression of Mat fimbria (not shown). Insertion of the wild-type *matA* allele back into the native chromosomal location completely restored the *matB* transcription in IHE 3034 (Fig. 6B). The results show that the HTH domain of MatA has a central role in stimulating the expression of the *mat* operon.

To examine the effect of MatA overexpression on *matB* transcription, the *matA* gene from strain IHE 3034 was

inserted into pACYC184 and expressed in the *matA* mutant strain. The *matB* mRNA levels were strongly elevated in the complemented strain, and *matB* expression was also detected at 37°C (Fig. 6B). Thus, overexpression of *matA* abolished temperature control in Mat fimbrillin expression, which suggests that MatA of IHE 3034 is competing with repressor(s) that prevent(s) transcription at 37°C.

We next analysed whether MatA binds to the *mat* upstream regulatory region. Several attempts to purify MatA as His_6- or GST-tagged protein failed to yield soluble fusion protein in quantities enough for functional studies. A small quantity of soluble MBP-$MatA_{IHE\ 3034}$ could however be produced (Fig. S3A). Electrophoretic mobility shift assay (EMSA) revealed that MBP-MatA bound to the *pmatA* fragment, encompassing the transcriptional start sites T1 and T2, but poorly to the *pmatB* fragment that contains the *matA–matB* intergenic region (Fig. S3B and C). Consistent with the Northern blot results (Fig. 6B), a H179P substitution in MatA reduced the binding to the *pmatA* fragment.

MatA of MG1655 is functional

The *matA* coding region in the genome of the *E. coli* K-12 strain MG1655 is 97.0% identical in sequence to that in the IHE 3034 genome and includes six non-synonymous substitutions, from which two are situated within the four-helical HTH domain (see Fig. 6A). We determined whether the crypticity of the *mat* gene cluster in MG1655 might also involve functional differences in its *matA* allele. We cloned *matA* from MG1655 in pACYC184, and analysed *matB* transcription in IHE 3034 *matA*(A536C) complemented with the plasmid carrying $matA_{IHE\ 3034}$ or $matA_{MG1655}$. In the IHE 3034 background, both *matA* alleles enhanced *matB* transcription at 20°C, whereas at 37°C expression of $matA_{IHE\ 3034}$ elicited considerably higher activation than did $matA_{MG1655}$ (Fig. 6B, left panel). In wild-type MG1655 and in MG1655 *matA*(A536C), however, no *matB* transcription was observed after the complementations (Fig. 6B, right panel). These results thus indicate that the MatA protein of MG1655, although not as efficient as MatA from IHE 3034, is able to activate *matB* transcription in the presence of functional *mat* promoter.

The mat promoter activation involves a MatA-mediated feedback loop in IHE 3034 and is affected by acidic pH as well as excess acetate

We integrated a single-copy *pA–lacZ* reporter fusion to the *lac* locus of IHE 3034-Rif and of IHE 3034 *matA*(A536C). Integration of a promoterless *lacZ* reporter into the *lac* locus of IHE 3034-Rif gave the basal level of 4–5 Miller units (MU) in cells grown at 20°C or 37°C. The *pA* region directed high levels of β-galactosidase activity in wild-type background at both temperatures, over 2000 MU at 20°C and 1500 MU Miller units at 37°C (Fig. 6C), whereas the promoter expression was impaired by 50% at 20°C in cells lacking a functional MatA. At 37°C, the *matA*(A536C) mutation had no effect on the promoter activity. Overexpression of *matA* in IHE 3034 *matA*(A536C) yielded strongly enhanced activation of the *pA* promoter at both temperatures.

The result in Fig. 6B showed that inactivation of *matA* resulted in loss of *matB* transcript. However, the *pA* promoter was active in IHE 3034 cells with defective MatA (Fig. 6C). The observed discrepancy between transcription initiation and *matB* mRNA level in the *matA*(A536C) mutant suggests that an additional regulatory element might contribute to the transcription. We constructed a *pAB–lacZ* promoter fusion, which carries the *pA* promoter region together with the *matA* gene and a 73-nucleotide *matA–matB* intergenic region up to position +81 relative to the translational start of *matB*. To silence a functional *matA* allele, we introduced the A536C substitution in *matA* of the *pAB–lacZ* fusion (see Fig. 6C).

In IHE 3034-Rif, the β-galactosidase activity directed by the *pAB–lacZ* fusion was 39% of the activity obtained with the shorter *pA* promoter (Fig. 6C). At 37°C, the promoter activity of *pAB* was low (133 MU) and only 9% compared with *pA*–directed activity at 37°C, indicating that temperature has profound influence on the promoter activation. Moreover, inactivation of MatA reduced the *pAB* expression to negligible levels. As in the case of the *pA–lacZ* fusion, addition of the *matA* plasmid caused upregulation of promoter expression (Fig. 6C). Interestingly, the promoter activity was still partially thermoregulated as evidenced by a reduction in activity to 61% at 37°C of the activity observed at 20°C (Fig. 6C). Taken together, our results demonstrate that MatA autoregulates its own transcription.

We measured the influence of low pH and high acetate on *mat* promoter activity using the single-copy *pAB–lacZ* reporter fusion (Fig. 7A). In IHE 3034-Rif, low pH activated the reporter in particular at 20°C, whereas at 37°C the promoter activation remained lower and was more effective with exogenous acetate. Levels of *matB* mRNA in IHE 3034 corroborated the measured promoter fusion activities (Fig. 7B), whereas in the case of strain MG1655 the stimulations did not lead to detectable *matB* transcription (data not shown). Further, the promoter intensities obtained in IHE 3034 *matA*(A536C) were near to basal level (< 65 MU; Fig. 7A), indicating that both pH- and acetate-driven activation depends on the MatA activator. To determine whether the responses to acidic pH and elevated levels of acetate is attributed mostly to the 0.6 kb *mat* promoter, rather than to strain-specific transcription

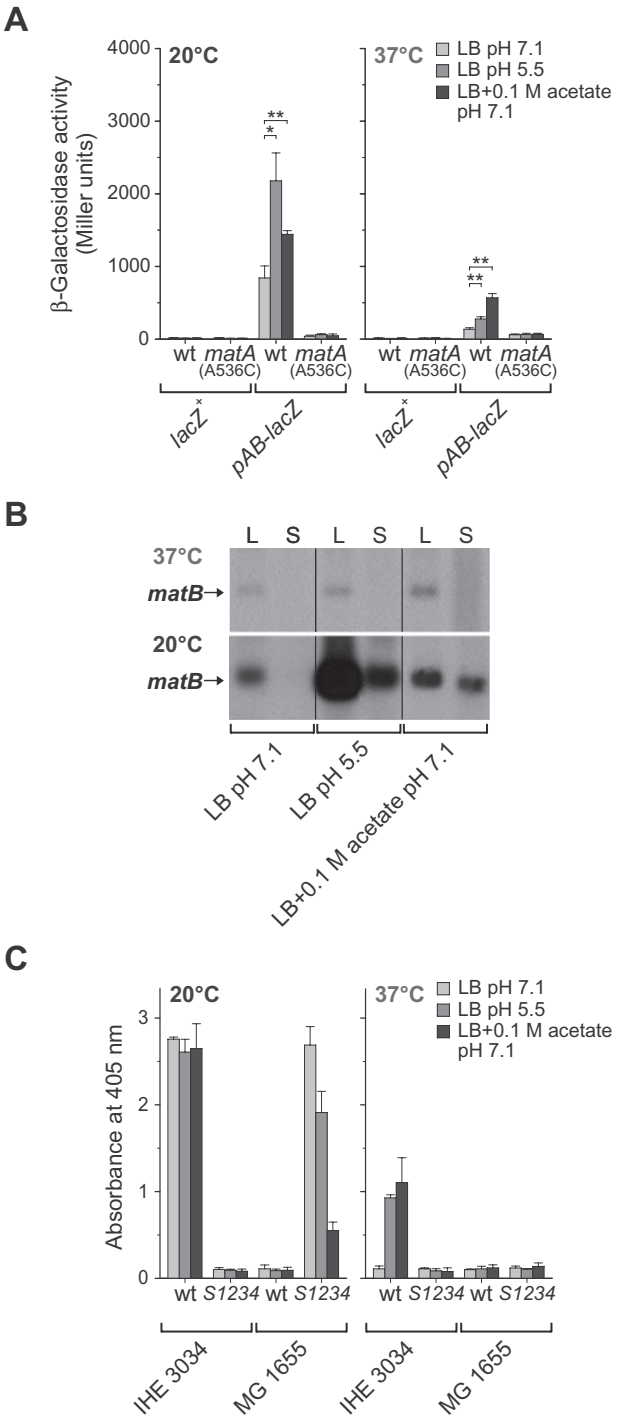

Fig. 7. Influence on *mat* promoter activity by acidic pH and excess acetate.
A. Effect of growth conditions on the activity of *pAB* promoter region in IHE 3034-Rif and *matA*(A535C) mutant. Bacteria were grown to mid-logarithmic phase in LB (pH 7.1), LB with 0.1 M MES (pH 5.5), and LB with 0.1 M acetate (pH 7.1) at 20°C and 37°C. Shown are mean and SD values of at least three independent assays. Significance is indicated as follows: one asterisk, $P < 0.05$; two asterisks, $P < 0.005$; as calculated by Student's *t*-test
B. Influence of different growth conditions on transcription of fimbrillin-encoding *matB* in IHE 3034. Total RNA was isolated from cells grown to mid-logarithmic (L) and stationary (S) phase at 20°C and 37°C. Northern blots were performed with *matB* probe. The position of the monocistronic *matB* message is indicated with arrows.
C. Comparison of the surface expression levels of Mat fimbriae in IHE 3034 and MG1655 *S1234* promoter swap strains. The bacteria were analysed by whole-cell ELISA with anti-Mat fimbria antiserum as primary antibodies (a dilution of 1:10 000) after growth to stationary phase in different growth media at 20°C and 37°C. The *S1234* promoter swap strain is described in Fig. 5A. Shown are mean and SD values of at least two independent assays.

factors or additional elements of the operon, we analysed by whole-cell ELISA the capacity of the *S1234* (–608 to +10 bp relative to *matA* start codon) promoter swap strains to produce Mat fimbriae at different environmental conditions (Fig. 7C). The presence of MG1655-derived sequences resulted in a complete loss of fimbriation in IHE 3034 *S1234* under all tested conditions, a phenotype that corresponds with MG1655. In contrast, the stimulatory effect of low pH or high acetate concentration seen for IHE 3034 was not present in MG1655 *S1234*. The replacement of the promoter region of MG1655 with the corresponding IHE 3034 sequence did not affect Mat expression at 37°C, and the high-level expression of Mat fimbriae of the strain in LB at 20°C was reduced by low pH or high acetate concentration. These results strongly indicate that divergence between *E. coli* strains during evolution has resulted in strain-specific modifications in the regulatory mechanisms of the *mat* operon beyond the *mat* promoter.

Stability of matB mRNA is affected by MatA and culture conditions

To test whether the accumulation of a *matB*-encoding mRNA at low temperature or in the presence of *matA* overexpression involves post-transcriptional regulation, we followed the decay of *matB* message after inhibition of transcription with rifampicin in IHE 3034 grown to exponential phase at 20°C and 37°C (Fig. S4A). The percentage of *matB* transcript that remains at each time point after transcriptional arrest was determined from Northern blots, and half-lives were calculated by using linear fits to time-series data. The *matB* mRNA half-life was estimated to be 38.7 min at 20°C and 6.7 min at 37°C (Figs 8A and S4A). Overexpression of *matA* significantly enhanced the stability of *matB* transcript at both temperatures, in particular, at 37°C the enhancement was 4.2-fold compared with vector control (Figs 8A and S4B). Acidic pH and high acetate had a minor, but statistically significant effect, on *matB* mRNA stability in IHE 3034 grown at 37°C, whereas

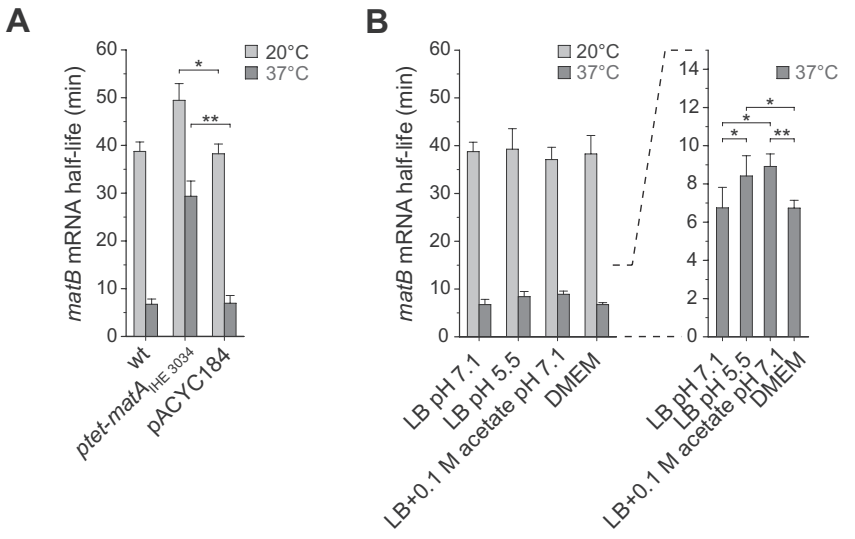

Fig. 8. MatA stabilizes *matB* mRNA and the stability is influenced by acidic pH as well as excess acetate.
A. Effect of $matA_{IHE\ 3034}$ overexpression on *matB* mRNA half-life in IHE 3034 at 20°C and 37°C.
B. Effect of growth conditions on *matB* mRNA half-life in IHE 3034 at 37°C and 20°C.
Bacteria were grown to mid-logarithmic phase in LB (pH 7.1), LB with 0.1 M MES (pH 5.5), LB with 0.1 M acetate (pH 7.1), and in DMEM (pH 7.2) at 20°C and 37°C.
Shown are mean and SD values of at least three independent assays. Significance is indicated as follows: one asterisk, $P < 0.05$; two asterisks, $P < 0.005$; as calculated by Student's *t*-test.

DMEM did not (Fig. 8B). These results suggest that MatA participates in the regulatory process of *matB* expression also at post-transcriptional level by enhancing transcript stability and that low pH as well as excess acetate have a discrete positive effect on *matB* transcript stability.

H-NS is a repressor of the mat operon and determines the crypticity of mat in MG1655

We constructed a mini-Tn*5* library in IHE 3034-Sm and enriched Mat-fimbriated mutants from the library grown at 37°C using immunomagnetic particles. Seven Tn*5* mutants from a library of 12 200 colonies expressed the fimbriae at 37°C. Direct DNA sequencing revealed that the mini-Tn*5* transposon insertions were all located in the *hns* open reading frame at base pair +340, relative to that of the start codon, at a region encoding the DNA-binding domain of H-NS protein (Ueguchi *et al*., 1996). Since the IHE 3034 *hns* mutants displayed similar capacity to produce Mat fimbriae, only one was chosen for further characterization.

The mutation in H-NS dramatically upregulated *matB* transcription both at 20°C and at 37°C in IHE 3034-Sm (Fig. 9A) and the mutant showed *matB* mRNA levels equal to those obtained with overexpression of *matA* (compare to Fig. 6B). In MG1655, inactivation of *hns* by *cat* insertion released repression of *matB* transcription in cells from 20°C, whereas only a poor accumulation of *matB* transcript occurred in cells grown at 37°C. Complementation of the *hns* mutation restored the repression in both strains (Fig. 9A). The results thus suggest that the *matB* expression is subject to negative control by the H-NS protein. We analysed the possible effect of H-NS deletion on transcription of *mat* genes by qRT-PCR analysis of the cells grown at 20°C and 37°C (Fig. 9B). The results revealed an upregulation of the six *mat* genes at both temperatures in IHE 3034.

Activation of H-NS-silenced genes is achieved by a variety of mechanisms including antagonism by other DNA-binding proteins or change in local promoter architecture (Stoebel *et al*., 2008). The temperature-blind induction of *matB* transcription in the absence of H-NS (Fig. 9B) closely resembled transcription in the presence of multicopy *matA*, and we evaluated the effect of *matA* inactivation on *matB* expression in the *hns* mutant background. The inactivation of MatA had no drastic effect on the accumulation of *matB* transcript in *hns*-deficient cells growing at 20°C (Fig. 9A). The results thus demonstrate that in the absence of H-NS, MatA is not compulsory for *matB* expression. This supports the idea that MatA functions as an H-NS antagonist. In contrast, at 37°C only a low level of *matB* transcript was detected in *hns*-deficient cells carrying the *matA*(A536C) allele. This finding suggests that the involvement of MatA in the regulation of the Mat fimbriae expression is more complex than merely antagonizing a transcriptional blockade exerted by H-NS.

The variable effect of H-NS deficiency on *matB* expression in IHE 3034 and MG1655 may result from the reduced capacity of the $MatA_{MG1655}$ variant to enhance the level of *matB* mRNA at 37°C. We expressed $MatA_{MG1655}$ and $MatA_{IHE\ 3034}$ in the MG1655 *hns::cat* mutant, and compared their effects on *matB* transcription. The $matA_{IHE\ 3034}$ allele supported a strong *matB* transcription, whereas the quantity of *matB* transcript was positively affected by $matA_{MG1655}$ only at 20°C (Fig. 9A). These results demonstrated that the global repressor H-NS is a key component in silencing of the cryptic but functional *mat* operon of MG1655 and further support the finding that the two MatA variants differ in their capacity to enhance *matB* transcription.

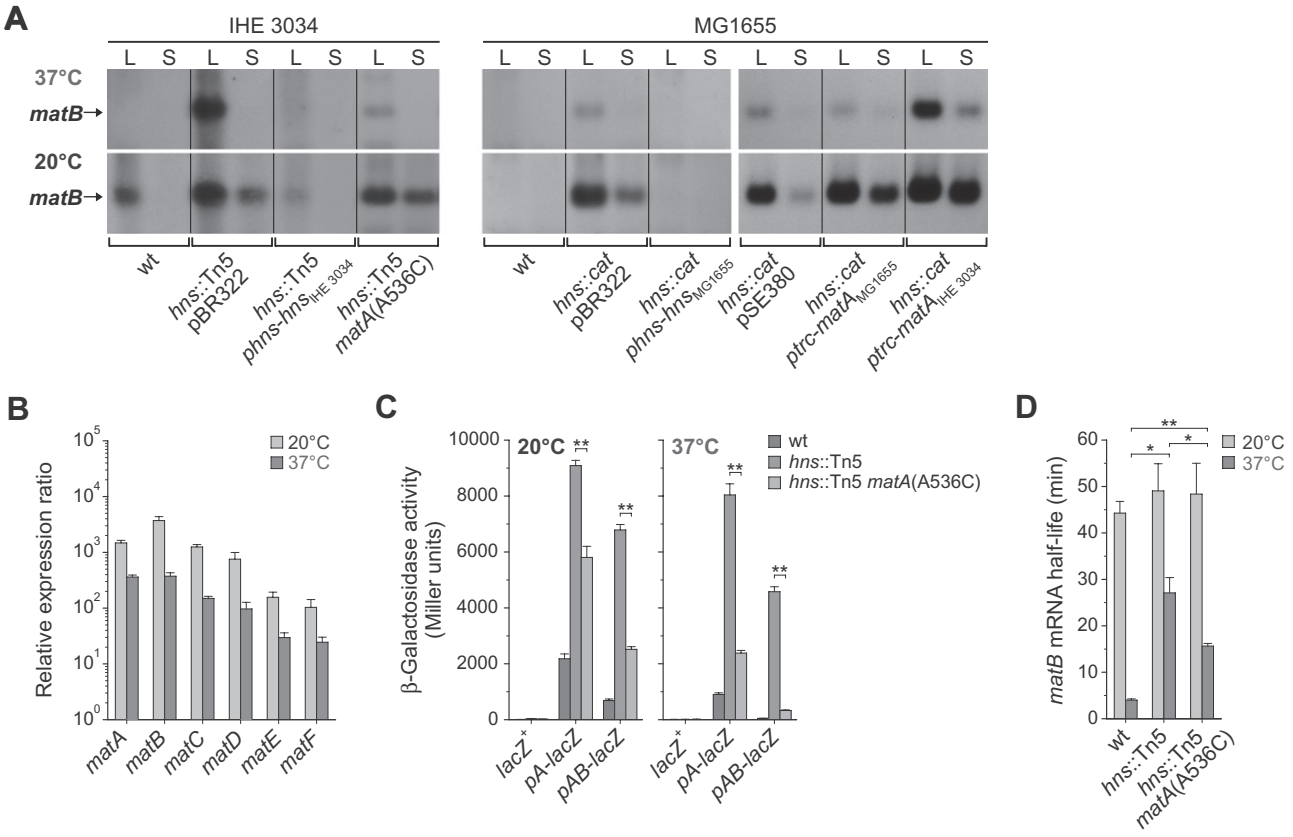

Fig. 9. H-NS acts as a repressor of the *mat* operon.
A. Transcription of *matB* by *hns* and *hns matA*(A536C) mutant strains was analysed by Northern blot with a *matB*-specific probe. For complementation, pBR322 plasmids carrying the promoter-containing *hns* gene from IHE 3034 or from MG1655, and pSE380 plasmids carrying the *matA* gene from IHE 3034 or from MG1655 under the inducible *ptrc* promoter, were used. Total RNA was isolated from cells grown in LB broth to mid-logarithmic (L) and stationary (S) phases at 20°C and 37°C. The position of the monocistronic *matB* message is indicated with arrows.
B. Increase in transcription levels of *mat* genes in IHE 3034 *hns* mutant at 20°C and 37°C was quantified by qRT-PCR. The levels are shown as relative to those seen with IHE 3034-Sm. Shown are mean and SD values of three independent assays with a P-value < 0.05.
C. Induction of *mat* promoter in IHE 3034 *hns* and *hns matA*(A536C) mutants was assessed by measuring β-galactosidase activities derived from *pA*– or *pAB*–*lacZ* fusions. Bacteria were grown to mid-logarithmic phase in LB broth at 20°C and 37°C. Shown are mean and SD values of at least three independent assays with a P-value < 0.005.
D. The effect of *hns* and *hns matA*(A536C) mutations on *matB* mRNA half-life. IHE 3034 derivatives were grown to mid-logarithmic phase in LB broth at 20°C and 37°C. Shown are mean and SD values of at least three independent assays. Significance is indicated as follows: one asterisk, $P < 0.05$; two asterisks, $P < 0.005$; as calculated by Student's t-test.

H-NS decreases mat promoter activity and matB mRNA stability

Inactivation of *hns* strongly increased the activity of the *pA–lacZ* and *pAB–lacZ* promoter fusions in IHE 3034-Sm cells grown at 20°C or 37°C (Fig. 9C). The mutation had a particularly marked effect on the strength of *pAB–lacZ* fusion at 37°C, raising the β-galactosidase level more than 98-fold. In agreement with Northern blot data (Fig. 9A), the inactivation of *matA* significantly decreased promoter activities in *hns*-deficient cells (Fig. 9C). EMSA showed that purified His_6-H-NS protein bound efficiently to the *pmatA* fragment derived from the upstream of IHE 3034 *matA* (Fig. S3D), whereas approximately a sixfold higher concentration of H-NS was required for a detectable binding to the *pmatB* fragment.

The *matB* half-life profile in the *hns* mutant was measured (Fig. 9D) in order to test the hypothesis that accumulation of mature *matB* mRNA in the *hns* mutant also involves MatA-dependent mRNA stabilization. Inactivation of MatA in the *hns* background had virtually no influence on the *matB* half-life at 20°C (Fig. 9D). At 37°C the stability was decreased almost by half, but the message was still 3.9-fold more stable than in the parental wt strain. These findings illustrate that the high stability of the *matB* transcript is achieved regardless of MatA when the action of H-NS is inhibited.

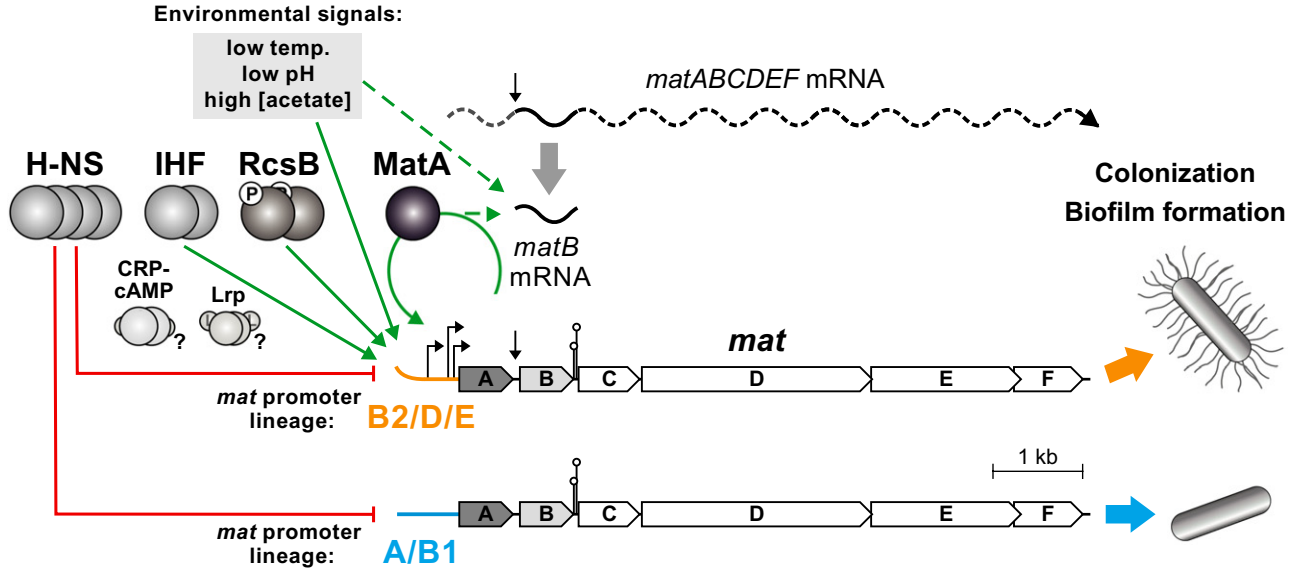

Fig. 10. Overview of environmental and genetic regulation of *mat* gene cluster in *E. coli*. The expression of Mat fimbriae differs in clonal groups of *E. coli*. The two main promoter lineages vary in response to signals from the host/environment. In the meningitis isolate IHE 3034, a representative of the B2/D/E lineage of *mat* promoters, low temperature, acidic pH and elevated concentration of acetate trigger expression of Mat fimbria by activating the *mat* promoter. The *mat* determinant is transcribed as an unstable polycistronic *matABCDEF* transcript, from which the fimbrillin *matB* mRNA is processed and stabilized preferentially at low temperature. The MatA regulatory protein binds to the *mat* upstream region, counteracts the repression mediated by H-NS, activates transcription from the *mat* promoter, and indirectly stabilizes *matB* mRNA. IHF is also required to overcome the repressive action of H-NS (Martínez-Santos *et al*., 2012). H-NS represses expression of the six *mat* genes in a strain-independent manner by binding to the *mat* upstream region. In addition, the DNA-binding response regulator RcsB of the Rcs phosphorelay system and an RcsAB box in the *mat* regulatory region are essential for initiation of *mat* transcription (Lehti *et al*., 2012a). Also, binding sites for the leucine-responsive protein Lrp (Cho *et al*., 2008), and for the stress-related cAMP-receptor protein CRP (Grainger *et al*., 2005) have been identified in the *mat* promoter region. MG1655, a representative of the A/B1 lineage, does not express Mat fimbria upon exposure to low temperature, acidic pH and elevated concentration of acetate. In MG1655, MatA is functional and the crypticity of *mat* is determined by H-NS. Green arrows indicate stimulatory effects, whereas red lines with blunt ends indicate repressing effects. Direct regulation is shown as continuous lines, and indirect effects are represented as dotted lines. Right-angled arrows indicate the three transcription start sites identified by 5′ RACE. The black vertical arrow shows the 5′ end of *matB* mRNA, and stem-loop symbols show putative transcriptional terminators downstream of *matB*.

Discussion

The mat genes are under purifying selection but differ in promoter architecture

We present here an example of alterations in a regulatory DNA element that controls a common but differentially expressed fimbrial adhesin of *E. coli*, and our findings on *mat* regulation are summarized in Fig. 10. The low degree of strain-specific polymorphisms in the coding sequence of the *mat* operon indicates that the *matABCDEF* genes have preserved their ancestral sequences and evolved slowly as an intact genetic unit. We identified two discrete promoter lineages associated with distinct regulatory properties, which indicates that the regulation of the *mat* operon rather than the fimbrial structure has been under divergent evolution. The shared positions of nucleotide substitutions within the lineages suggest that nearly all mutations in the *mat* upstream region have been transmitted vertically, rather than having been acquired on multiple independent occasions.

Comparative sequence analysis of the *mat* promoter in context of *E. coli* phylogeny (Touchon *et al*., 2009; Sims and Kim, 2011) suggests that the B2/D/E-type represents a parental promoter line. Promoter swapping analysis indicated involvement of a 0.6 kb *mat* regulatory region upstream of the start codon of *matA* in the control of *mat* transcription in MG1655 and in temperature regulation in IHE 3034. The proximal ~350 base pairs of the *mat* upstream region was the critical element in expression assays as measured by loss or gain of function, and is almost completely conserved in isolates of the B2/D/E lineage. We identified the region between the nucleotides –343 and –122 as the main *pP2* promoter responsive to thermoregulation and responsible for the expression of the polycistronic *mat* mRNA, whereas the two putative promoters showed low (*pP3*) or no activity (*pP1*). Differing transcriptional start sites influence regulation of several genes whose expression is adapted to different conditions (Ko *et al*., 2006; Sayed and Foster, 2009), and this could be the case also for the *mat* operon.

Multi- as well as single-copy *lacZ* reporter analyses showed that the activity of the *mat* promoter was higher at 20°C than at 37°C. This temperature-responsive transcription initiation differs from the expression of most *E.*

coli fimbrial operons which have a maximal transcription level at 37°C, but resembles the transcription of curli (Olsén *et al*., 1989). The nucleotides −608 to +10 were involved in the *mat* promoter responses to low pH and high acetate, which appear to require additional elements or factors beyond the *mat* operon as shown by analysis of Mat fimbriae expression in the promoter swap strain MG1655 carrying the IHE 3034 promoter element *S1234*. This finding was further confirmed by the fact that also the strains IHE 3034 and BEN374 (or CFT073 and LF82) reacted differently to these signals, although they share an identical DNA sequence spanning the entire *matA* promoter to the end of *matA* coding region. This suggests existence of strain-specific differences in the repertoire of transcription factors acting directly or indirectly in Mat expression. Such factors could include stress-related sigma factors, intergenic regions that contain sRNAs, and genes encoding regulatory proteins, as recently observed in acid responses in EHEC O157:H7 Sakai (King *et al*., 2010).

Differential expression of a surface antigen in *E. coli* has so far been described in a few cases. Sequence variations in the regulatory regions of P- and type 1 fimbrial operons in UPEC facilitate specific fine-tuning in regulation of these determinants (Leathart and Gally, 1998; Totsika *et al*., 2008). Also the variation in OmpT-mediated resistance against α-helical antimicrobial peptides in an EHEC and an EPEC strain is due to differences in the *ompT* promoter region (Thomassin *et al*., 2012). Another mechanism involves differential allocation of RNA polymerase to promoters. The variability of the general stress resistance in *E. coli* is related to strain-specific expression of the stationary-phase sigma factor RpoS and ppGpp, which directly bind to RNA polymerase to alter its activity (Ferenci *et al*., 2011). Our study is the first demonstration of phylogenetic group-associated promoter lineages and fimbriae expression in *E. coli*.

Diverse induction of mat expression under host-mimicking conditions

We found that two physiological cues related to the human gastrointestinal tract and the birth canal, namely low pH and high acetate concentration, control the *mat* promoter in IHE 3034 when grown at 37°C. The 11 strains of the phylogenetic group B2 expressed Mat fimbria when cultured at 20°C, which may potentiate survival also outside a host. The majority of these strains also responded at 37°C to low pH, acetate or both, which may be relevant to bacterial colonization of mammals. Transcriptomic responses of an EHEC isolate to acid stress (pH 3) involved upregulation of structural *mat* genes (House *et al*., 2009), which supports our findings that growth at low pH induces Mat fimbriae expression. The sterile neonatal gut is first colonized by *E. coli* and other facultative anaerobes secreting short-chain fatty acids (SCFAs), primarily acetate, propionate, and butyrate. As the gut progressively becomes more reduced and acidified, the predominant microbiota rapidly changes into fermentative *Bifidobacterium* and *Bacteroides* that excrete acetate and other SCFAs. High SCFA concentration is linked with expression of adhesive Iha protein of Shiga-toxigenic *E. coli* (Herold *et al*., 2009). We found that SCFA-containing, conditioned medium of *B. longum* promotes Mat expression in IHE 3034 at 37°C. This suggests that high concentration of acetate also serves as a signal for *mat* gene expression in the gut, which in turn may enhance colonization and biofilm formation by *E. coli*. We also assessed expression of Mat fimbria in the eukaryotic cell culture medium DMEM in the absence and presence of 5% CO_2. These conditions have previously been shown to stimulate induction of expression of the Mat homologue of diarrhoeagenic *E. coli* pathovars (Rendón *et al*., 2007; Blackburn *et al*., 2009; Saldaña *et al*., 2009; Avelino *et al*., 2010), but no induction of Mat expression was observed in IHE 3034. The results thus indicate that regulatory cues for Mat fimbriae expression are different in diarrhoeagenic and meningitic *E. coli* strains, which may reflect adaptation to ecological niches. The conditions that favour expression of Mat fimbria in IHE 3034 are prevalent in the intestine and the birth canal, where the bacterium colonizes during transmission from mother to the child, and our hypothesis is that Mat-mediated cell-adhesion and biofilm formation have a role in the colonization and transmission.

The MatA activator and the HNS repressor counteract in Mat expression

The MatA activator protein has dual role in the control of *mat* expression in IHE 3034, it forms a positive autoregulatory circuit via the *mat* promoter and enhances the stability of the processed *matB* mRNA. Significant reduction in activities of *pA*– and *pAB–lacZ* promoter fusions in the *hns matA* double mutant suggest that MatA has also an additive effect in the activation of transcription initiation through its counteraction on H-NS. Interestingly, the strictly repressed cryptic *mat* operon of MG1655 was totally insensitive to MatA stimulation in the presence of H-NS, while in the *hns* background the expression was activated by MatA. The strain-specific MatA variants showed slightly variable efficiency in stimulation in *matB* transcription in MG1655 *hns* but the promoter swap experiments demonstrated that change of the entire 0.6 kb *mat* upstream region in MG1655 to the IHE 3034-based sequence led to proficient Mat fimbriation by MG1655-encoded MatA. Thus, these results support the idea that the core function of MatA in the control of *mat*

transcription is maintained across *E. coli* isolates. Consistent with our studies and during the revision of this article, Martínez-Santos *et al.* (2012) independently reported that MatA and H-NS regulates the expression of the *mat* (*ecp*) promoter in attaching and effacing *E. coli*.

The inactivation of *hns* elicited a temperature-independent accumulation of the *matB* mRNA in IHE 3034, and also led to derepression of the cryptic *mat* operon of MG1655. The majority of thermoregulated genes in *E. coli*, such as curli genes, are under the control of H-NS (White-Ziegler and Davis, 2009). A recent study investigated seven out of the 11 chaperone-usher operons in *E. coli* K-12, and found that these mostly cryptic fimbrial operons are all strongly repressed by H-NS (Korea *et al.*, 2010). DNA macroarray analyses of *hns* mutations have shown that in the uropathogenic *E. coli* strain 536, a close relative of IHE 3034 (Moulin-Schouleur *et al.*, 2007), more than 500 genes, including all six *mat* genes, are upregulated (Müller *et al.*, 2006), and in the K-12 strain FB8, *matBCD* genes were among the 250 differently expressed genes (Hommais *et al.*, 2001). In agreement with these studies, our real-time PCR analysis showed that all *mat* genes in IHE 3034 are under the repressive action of H-NS as their expression were strongly upregulated in the *hns* mutant. We demonstrated that the H-NS protein binds to the *mat* promoter. This is consistent with a ChIP-Chip monitoring of the genomic distribution of H-NS and RNA polymerase (RNAP)-associated sequences in *E. coli* K-12 cells, which revealed that the upstream region of the *mat* operon and the *matA*-coding sequence are covered by H-NS. Also RNAP is associated with the promoter, but is apparently repressed by H-NS (Oshima *et al.*, 2006). The existence of a downstream regulatory element (DRE) downstream of *mat* promoter is supported by the observed differences between response profiles of single-copy *pA*– and *pAB*–*lacZ* promoter fusions. In IHE 3034, this element seems to exert a repressive effect that is notably stronger at 37°C, while in the absence of H-NS nearly temperature-blind transcription initiation was detected.

The MatA-dependent *matB* mRNA stabilization seems indirect. The mature *matB* mRNA was preferentially stabilized at low temperature, and overexpression of MatA dramatically increased *matB* mRNA stability. At 37°C the transcript became susceptible to degradation and had a half-life in the range of the global *E. coli* mRNA (5.2–7.5 min; Mohanty and Kushner, 1999; Bernstein *et al.*, 2002; Selinger *et al.*, 2003). The variation in stability may be due to temperature-dependent conformation of *matB* transcript as, e.g. in mRNA of major cold-shock protein CspA (Giuliodori *et al.*, 2010), or to differences in how the RNA degradosome assembles at low and high temperature (Prud'homme-Généreux *et al.*, 2004). We compared the *matB* stability patterns in *hns* and *hns matA* mutants. The high *matB* mRNA half-life in the *hns* mutant closely resembled the turnovers seen in the wild-type after transformation with a MatA-overexpressing plasmid, which can be explained by the increased expression of *matA* in the absence of H-NS. In the *hns* mutant however, the loss of functional MatA had no impact on *matB* turnover at 20°C, and at 37°C only a partial decrease in the stability was detected. These results suggest that MatA acts beyond its local influence on the *mat* operon, as observed in MatA-mediated repression of the flagellar master operon *flhDC* (Lehti *et al.*, 2012b).

The B2 strains of *E. coli* often carry virulence genes and associate with extra-intestinal infections and with long-term persistence in the colon, and it was proposed that their genetic background is suitable for acquisition and expression of fitness or virulence genes (Escobar-Páramo *et al.*, 2004; Nowrouzian *et al.*, 2005; 2006). The genomic islands or pathogenicity islands present in these strains may also encode additional H-NS homologues or other DNA-binding proteins, which may affect global gene expression (Müller *et al.*, 2006). Here we show that the B2 strains, in contrast to A and B1 phylogenetic groups of *E. coli*, possess a *mat* promoter architecture that favours expression of Mat fimbriae at conditions mimicking those in the intestine and the vagina. We suggest that this adaptive divergence contributes to the evolutionary success of the B2 strains by increasing their colonization and transmission capacity. We found that the representative member of the A/B1 lineage, *E. coli* MG1655, harbours functional *mat* genes (Lehti *et al.*, 2010; this study). This implies that the members of A/B1 lineage have retained the capability to express the fimbriae at as-yet unidentified conditions that loosen the tight repression mediated by H-NS.

Experimental procedures

Bacterial strains, plasmids and growth conditions

Escherichia coli strains and plasmids used in this study are listed in Table S2 and Table S3 respectively. Strains were grown to mid-logarithmic or stationary phase [corresponding to an optical density at 600 nm (OD_{600}) of around 0.5 and 2.2 respectively] with vigorous shaking (200 r.p.m.) at 20 or 37°C in LB medium, unless otherwise stated. For surface expression analysis of Mat fimbriae, bacteria were grown statically. When necessary, media were supplemented with ampicillin (Ap, 100 μg ml^{-1}), chloramphenicol (Cm, 25 μg ml^{-1}), kanamycin (Km, 25 μg ml^{-1}), rifampicin (Rif, 75 μg ml^{-1}), streptomycin (Sm, 100 μg ml^{-1}) or tetracycline (Tc, 12.5 μg ml^{-1}). The *mat* derivatives were expressed in the inducible pSE380 vector with 5 μM isopropyl-β-D-thiogalactopyranoside (IPTG) as described earlier (Pouttu *et al.*, 2001; Lehti *et al.*, 2010).

The effect of culture conditions on bacterial growth and expression of Mat fimbria was tested by cultivating the bacteria at 20 or 37°C in static Dulbecco's modified Eagle's medium (DMEM; Gibco) in the presence or absence of 5% CO_2, and in LB or M9 minimal medium (Miller, 1972) in the presence of

0.2% D-galactose, glycerol, D-lactose, DL-lactate, D-mannose or casamino acids; 10 mM L-alanine, L-isoleucine, L-leucine, L-lysine, L-serine, D-serine or ferric citrate; 1% or 5% skim milk; 0.015% or 0.05% bile salt; 100 mM NH_4Cl, formic acid, D-glucuronic acid, sodium pyruvate or sodium citrate; and 50 mM or 100 mM sodium acetate. The influence of pH was tested by cultivating the bacteria at 20 or 37°C in LB or M9 medium buffered with either 100 mM MES [2-(*N*-Morpholino)ethanesulphonic acid] pH 5.5, or BTP (1,3-Bis[tris(hydroxymethyl)methylamino]propane) pH 8.0 or pH 9.0. The bacteria were passaged two times under these conditions before they were tested for Mat fimbria expression.

Conditioned medium of *B. longum* was prepared by culturing *B. longum* ATCC 15707 in MRS (1% Bacto proteose peptone No. 3, 1% Bacto beef extract, 0.5% Bacto yeast extract, 2% glucose, 0.1% Tween-80, 0.2% ammonium citrate, 0.5% sodium acetate, 0.01% $MgSO_4$, 0.005% $MnSO_4$ and 0.2% K_2HPO_4) or acetate-depleted MRS for 48 h at 37°C under anaerobic conditions. The cells were pelleted, and the supernatant was then adjusted to pH 7.0 and filtered through 0.2 μm syringe filter. The conditioned media were used for *in vitro* growth of *E. coli* IHE 3034 at 37°C in the analysis of the effect of medium on Mat expression.

Genetic techniques

Standard recombinant DNA techniques were used (Sambrook and Russell, 2001). Missense mutations, gene complementation, promoter swaps, and single-copy promoter–*lacZ* fusions in the bacterial chromosome were created by site-specific mutagenesis using recombinant PCR and the *pir*-dependent suicide vector pCVD442 essentially as described by Mobley *et al*. (1993). Oligonucleotides used in this study are listed in Table S4. The created mutations were verified by DNA sequencing. To enhance the frequency of excision of integrated suicide vector from the chromosome (from 10^{-3} to almost 10^{0}), an additional sucrose selection step was included into the mutagenesis protocol. The transconjugants grown for 18 h in non-selective LB broth were diluted 1:200 in LB supplemented with 10% sucrose and cultured for additional 5 h before plating on NaCl-free LB agar plates containing 10% sucrose.

For construction of suicide plasmids for single-copy promoter–*lacZ* fusions, two fragments were PCR-amplified and fused with each other by recombinant PCR: the *lacI* gene with the upstream region of IHE 3034 genomic DNA (–420 to +1101 bp from the start codon GTG of *lacI*) and a fragment of pRS551 (–756 to +2066 bp from the EcoRI site) containing a transcription termination region of four *rrnB* T1 terminators, a promoter cloning site and a partial promoterless *lacZ* gene. The fusion product was inserted at the XbaI–SacI restriction site of pUC19 to create pUC19-placZ. Promoter fragments from the IHE 3034 genome were then cloned individually into the BamHI restriction site of the fusion insert in pUC19-placZ, and the XbaI–SacI fragments containing the promoter–*lacZ* fusions were finally subcloned into suicide vector pCVD442.

Mutants able to express Mat fimbriae in LB at 37°C were identified by construction of a mini-Tn*5* library and separation of the fimbriated mutants using magnetic particles coated with anti-Mat fimbria IgG, a method similar to that described by Nuccio *et al*. (2007). Briefly, a pUTmini-Tn*5* Cm vector was conjugated from strain Sm10 λpir into IHE 3034-Sm, and exconjugants were selected on LB agar containing Sm and Cm. Approximately 12 200 mini-Tn*5* transposon mutants were collected from the agar plates into 1 ml of PBS. After cultivation overnight at 37°C statically in LB supplemented with Sm and Cm, approximately 5×10^7 bacteria in 0.1 M sodium-phosphate buffer at pH 7.4 were added to 2×10^7 magnetic particles (Dynabeads M-280 Tosylactivated; Dynal Biotech) coated with anti-Mat fimbria IgG available from previous work (Pouttu *et al*., 2001) according to the manufacturer's instructions. The suspension was incubated for 1 h at 22°C with rotation, the particles were washed three times with PBS containing 0.05% Tween 20 (hereafter PBS-T), and collected using a magnetic particle concentrator (Dynal MPC-6). The collected particles were resuspended in 0.5 ml of PBS, and 10-fold dilutions were spread on LB agar plates containing Cm. 500 randomly selected colonies were picked onto fresh agar plates, cultivated for 18 h at 37°C and assessed for surface expression of Mat fimbriae by colony immunoblotting (as described below).

Plasmid-encoded transcriptional promoter fusions with promoterless *lacZ* were constructed by cloning PCR-amplified *mat* operon regions from IHE 3034 or MG1655 into the BamHI (*mat* upstream region derivatives) or EcoRI–BamHI (*pB*, *pC*, *pD*, *pE* and *pF*) restriction sites of pRS551.

A *matA* overexpression plasmid was constructed by fusion of the *matA* gene from IHE 3034 or MG1655 chromosomal DNA with the constitutive promoter of the *tet* gene from pACYC184 using recombinant PCR, and cloning into the XbaI–SalI restriction sites of pACYC184. Plasmids pMAT19 and pMAT21 were constructed by cloning the *matA* coding region from IHE 3034 or MG1655 chromosomal DNA into the EcoRI–SalI restriction sites of pSE380 under the inducible *ptrc* promoter as described earlier (Pouttu *et al*., 2001).

Plasmids pHNS1 and pHNS2 were constructed for complementation of the *hns* mutations. The *hns* coding region and additional flanking regions (660 bp or 661 bp upstream and 143 bp downstream) were PCR-amplified from IHE 3034 and MG1655 chromosomal DNA and cloned into the BamHI–HindIII restriction sites of pBR322.

RNA techniques

RNA was isolated from bacterial cells using RNeasy Mini kit (Qiagen) as described previously (Lehti *et al*., 2012a). For RT-PCR and qPCR analysis, the bacterial cultures were mixed with RNAprotect (Qiagen) and vortexed prior to RNA isolation. Traces of chromosomal DNA were removed after RNA isolation by treating with RNase-free DNase I (Roche Applied Science) for 2 h at 37°C, and then purifying with RNeasy (Qiagen) clean-up protocol as recommended by the manufacturer. DNase I treatment of RNA samples was also made prior to primer extension with RQ1 RNase-Free DNase (Promega) according to manufacturer's instructions.

For measurements of mRNA decay, total RNA was isolated by the hot phenol method essentially as described by von Gabain *et al*. (1983). The RNA concentration was determined by measuring the optical density at 260 nm.

For quantification of *mat* mRNA by Northern blot analysis, digoxigenin (DIG)-labelled, single-stranded DNA probes of

500–700 bp in length and complementary to the full-length genes (*matA*/*B*/*C*) or the 3′ end of genes (*matD*/*E*/*F*) were used. The probes were prepared by linear PCR amplification using PCR DIG Labelling Mix (Roche Applied Science), appropriate reverse primer, and PCR-amplified DNA fragment as a template. Total RNA samples of 2.5 μg, unless otherwise stated, were subjected to Northern blot analysis as previously described (Lehti *et al*., 2012a). Quantification of hybridized probe was performed using the Tina (v2.0) image analysis program (Raytest Isotopenmessgeräte Gmbh), and the half-life of an mRNA was determined from a least-squares linear fit to a semilog plot of mRNA abundance versus time.

For primer extension, the primer 057 (20 pmol) was 5′ end-labelled using 20 pmol of [γ-^{33}P]-ATP (3000 Ci mmol^{-1}; PerkinElmer) with T4 polynucleotide kinase (Fermentas) at 37°C for 30 min and then at 90°C for 2 min. Labelled primer (1.0 pmol) was annealed with total RNA (1 or 10 μg) at 55°C for 10 min, cooled to 22°C, and extended with Transcriptor reverse transcriptase (Roche Applied Science), in the presence of Transcriptor RT reaction buffer and dNTPs, at 55°C for 30 min and then at 85°C for 5 min. The sample was EtOH precipitated and resuspended in sequence stop solution (95% formamide containing 10 mM NaOH, 0.05% bromophenol blue and 0.05% xylene cyanol). Sequence ladders were generated with the same primer using *fmol* DNA Cycle Sequencing System (Promega) and a PCR-amplified DNA fragment as a template. After denaturation at 95°C for 2 min, the length of the reaction products was determined on a 6% acrylamide–7 M urea sequencing gel. The gel was blotted onto positively charged nylon membrane and autoradiographed on X-ray film.

To analyse co-transcription of *mat* genes by RT-PCR, DNase-treated total RNA was transcribed into cDNA as described previously (Bauchart *et al*., 2010). The RT-PCR primers were selected with the FastPCR software version 3.6.28 (Ruslan Kalendar, Institute for Biotechnology, University of Helsinki, Finland). cDNA samples were diluted 1:3 in water and used for PCR amplification with Red *Taq* polymerase ready mix (Sigma). DNase-treated RNA without a reverse transcription step as a control for DNA contaminations, and genomic DNA as a positive control, was used as template in a PCR reaction with primers binding within the coding sequence of the type 1-fimbrial adhesin gene *fimH* or the housekeeping gene *gapA*. After amplification, PCR samples were directly loaded and analysed on a 2% agarose gel and stained with ethidium bromide.

qRT-PCR was performed as described before (Bauchart *et al*., 2010) to analyse the relative expression levels of *mat* genes in IHE 3034 using *frr* and *gapA* as reference genes and the relative expression levels of *mat* genes in IHE 3034-Sm compared with IHE3034 *hns*.

To determine the transcription starting points upstream of *matA* gene, the 5′ RACE kit from Roche Applied Science was used. Two micrograms of total RNA was added in the first-strand cDNA synthesis step, and DAp GoldStar DNA polymerase (Eurogentec) was used for the nested PCR amplification. Prior to subcloning, the PCR products were cut out from agarose gel and purified using QIAquick Gel Extraction Kit (Qiagen). The products were subcloned into pGEM-T Easy vector followed by sequencing with the promoter primers T7 or SP6.

Measurement of β-galactosidase activity

To measure β-galactosidase activities, cell pellets from 1 ml of mid-exponential cultures were resuspended in 1 ml of ice-cold Z-buffer (60 mM Na_2HPO_4–40 mM NaH_2PO_4–10 mM KCl–1 mM $MgSO_4$) with 50 mM β-mercaptoethanol. β-Galactosidase activity from cell suspensions was then assayed using o-nitrophenyl-β-D-galactopyranoside as a substrate as described by Miller (1972).

Immunological methods

Detection of Mat fimbriae surface expression by whole-cell ELISA was performed basically according to de Ree *et al*. (1986), following the modified procedure described earlier (Pouttu *et al*., 2001; Lehti *et al*., 2010). The polyclonal rabbit anti-Mat fimbria antiserum available from previous work (Pouttu *et al*., 2001) had been applied in whole-cell ELISA before (Pouttu *et al*., 2001; Lehti *et al*., 2010).

For colony immunoblotting, bacteria were grown overnight on LB agar plates and transferred to Protran nitrocellulose membranes (Whatman) by placing the membranes over the plates for 10 s. The membranes were blocked in 2% (w/v) BSA in PBS at RT for 1 h, incubated in anti-Mat antiserum (1:500) in 1% BSA in PBS for 1 h, and washed three times in PBS-T. The membranes were incubated in alkaline phosphatase-conjugated swine anti-rabbit immunoglobulins (1:2000; Dako) in 1% BSA in PBS for 1 h, washed three times with PBS-T, and finally phosphatase substrate was added to visualize Mat fimbriated colonies.

Expression levels of Mat fimbriae by Western blotting were assessed from OD_{600} normalized and HCl-treated whole-cell samples, with anti-Mat fimbria serum (1:750) as primary antibodies, as described previously (Pouttu *et al*., 2001; Lehti *et al*., 2010). Similarly electrophoresed samples were stained with Coomassie Blue for verification of the cell concentration.

Bacterial agglutination in anti-Mat antiserum was done by mixing equal volumes of a suspension of bacteria and a 100-fold dilution of antiserum in PBS on glass slides as described (Rhen *et al*., 1983).

DNA sequencing and computational sequence analysis

Mini-Tn*5* insertion sites were determined by sequencing using isolated genomic DNA of each mutant as a template, transposon-specific primer derived from the I end of mini-Tn*5* Cm and with BigDye terminator enzyme mix v3.1 (Applied Biosystems) according to the manufacturer's instructions. The reactions were analysed on an ABI PRISM 310 or ABI PRISM 3100 genetic analyser (Applied Biosystems).

For sequencing of the *mat* upstream region including a putative *mat* promoter from selected bacterial strains, genomic DNA was isolated, and sequencing templates were amplified by PCR with Phusion High-Fidelity DNA Polymerase (Finnzymes). The *E. coli* strain ABU38 was excluded from the analysis due to the lack of successful PCR production of DNA template. The amplified PCR products were sequenced with BigDye terminator enzyme mix v3.1 (Applied Biosystems). The primers used are specified in Table S4. The reactions were analysed on an ABI PRISM 3100 genetic

analyser (Applied Biosystems; at the Sequencing Core Facility, Haartman-Institute, University of Helsinki, Finland). The sequences were reviewed and edited by visual inspection using Chromas Lite v2.01 software (Technelysium Pty), and the sequence data have been deposited in the GenBank database under Accession No. JN377374–JN377380. These sequences were combined with those from 45 published *E. coli* and one *Klebsiella pneumoniae* genomes for comparison, and aligned using CLUSTALW within MEGA4 software (Tamura *et al.*, 2007). Neighbour-joining trees were constructed using the Tamura-Nei model of nucleotide substitution with default parameters using the MEGA4 software, and the node reliabilities were estimated with 1000 bootstrap replications.

Pairwise alignment of *mat* upstream and coding sequences from IHE 3034 and MG1655 was done with the needle program from the EMBOSS package (Rice *et al.*, 2000). A motif search for HTH domain was performed using the online servers Motif Scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan) and PSIPRED v3.3 (http://bioinf.cs.ucl.ac.uk/psipred).

GenBank accession numbers of nucleotide sequences used in this study are as follows: *E. coli* 042, FN554766; 11368, AP010953; 12009, AP010958; 285, JN377379; 55989, CU928145.2; 789, JN377380; 83972, CP001671; APEC O1, CP000468; ATCC 8739, CP000946; B REL606, CP000819; BEN374, JN377377; BEN79, JN377376; BL21(DE3), AM946981.2; BL21-Gold(DE3)pLysS AG, CP001665; CB9615, CP001846; DH1, CP001637; DH10B, CP000948; E2348/69, FM180568; E24377A, CP000800; EC4115, CP001164; ED1a; CU928162.2; EDL933, AE005174.2; H10407, FN649414; HS, CP000802; IAI1, CU928160.2; IAI39, CU928164.2; IHE 3034, HM102365 and CP001969; IHE 3040, JN377374; IHE 3072, JN377375; KO11, CP002516; LF82, CU651637; MG1655, U00096; NA114, CP002797; Nissle 1917, JN377378; NRG 857C, CP001855; S88, CU928161.2; Sakai, BA000007.2; SE11, AP009240; SE15, AP009378; SMS-3-5, CP000970; TW14359, CP001368; UMN026, CU928163.2; UMN146, CP002167; UMNF18, CP002890; UMNK88, CP002729; UTI89, CP000243; W, CP002185; W3110, AP009048; and *Klebsiella pneumoniae* 342, CP000964.

Statistical analysis

Mean and standard deviation (SD) values were calculated from at least three biologically and technically independent experiments, unless otherwise stated. Differences between mean values were tested for significance by performing an unpaired, two-sided Student's *t*-test. *P*-values of less than 0.05 were considered as statistically significant. The levels of significance of the resulting *P*-values are indicated as follows: one asterisk (*), $P < 0.05$ and two asterisks (**), $P < 0.005$.

Acknowledgements

We thank Arlette Darfeuille-Michaud, Maryvonne Moulin-Schouleur, Eliora Z. Ron, Anja Siitonen, Annika Sjöström and Bernt Eric Uhlin for providing the strains, and Vanessa Sperandio for providing the vector pRS551. We would also like to thank Elina Kajasto and Raili Lameranta for experimental help. This work was supported by University of Helsinki, Viikki Graduate School in Biosciences, Academy of Finland (in the frame of the ERA-NET PathoGenoMics Grant No. 118982 and 130202, and General research Grant 123900). P. Bauchart received a Research Fellowship from the Bavarian Research Foundation. U. Dobrindt was supported by the German Research Foundation (Do 789/4-1). This work was carried out in the frame of the European Virtual Institute for Functional Genomics of Bacterial Pathogens (CEE LSHB-CT-2005-512061) and the ERA-NET PathoGenoMics project 'Deciphering the intersection of commensal and extraintestinal pathogenic *E. coli*' (Grant No. 0313937A).

References

Achtman, M., Mercer, A., Kusecek, B., Pohl, A., Heuzenroeder, M., Aaronson, W., *et al.* (1983) Six widespread bacterial clones among *Escherichia coli* K1 isolates. *Infect Immun* **39:** 315–335.

Avelino, F., Saldaña, Z., Islam, S., Monteiro-Neto, V., Dall'Agnol, M., Eslava, C.A., and Girón, J.A. (2010) The majority of enteroaggregative *Escherichia coli* strains produce the *E. coli* common pilus when adhering to cultured epithelial cells. *Int J Med Microbiol* **300:** 440–448.

Bailey, J.K., Pinyon, J.L., Anantham, S., and Hall, R.M. (2010) Distribution of human commensal *Escherichia coli* phylogenetic groups. *J Clin Microbiol* **48:** 3455–3456.

Barlow, D.J., and Thornton, J.M. (1988) Helix geometry in proteins. *J Mol Biol* **201:** 601–619.

Bauchart, P., Germon, P., Brée, A., Oswald, E., Hacker, J., and Dobrindt, U. (2010) Pathogenomic comparison of human extraintestinal and avian pathogenic *Escherichia coli* – search for factors involved in host specificity or zoonotic potential. *Microb Pathog* **49:** 105–115.

Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., and Cohen, S.N. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* **99:** 9697–9702.

Bingen, E., Picard, B., Brahimi, N., Mathy, S., Desjardins, P., Elion, J., and Denamur, E. (1998) Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. *J Infect Dis* **177:** 642–650.

Blackburn, D., Husband, A., Saldaña, Z., Nada, R.A., Klena, J., Qadri, F., and Girón, J.A. (2009) Distribution of the *Escherichia coli* common pilus among diverse strains of human enterotoxigenic *E. coli*. *J Clin Microbiol* **47:** 1781–1784.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277:** 1453–1462.

Bonacorsi, S., and Bingen, E. (2005) Molecular epidemiology of *Escherichia coli* causing neonatal meningitis. *Int J Med Microbiol* **295:** 373–381.

Chaudhuri, R.R., and Thomas, G.V. (2007) Understanding the model and the menace: a postgenomic view of *Escherichia coli*. In *Bacterial Pathogenomics*. Pallen, M.J.,

Nelson, K.E., and Preston, G.M. (eds). Washington DC: ASM Press, pp. 21–48.
Cho, B.K., Barrett, C.L., Knight, E.M., Park, Y.S., and Palsson, B. (2008) Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. *Proc Natl Acad Sci USA* **105:** 19462–19467.
Croxen, M.A., and Finlay, B.B. (2010) Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* **8:** 26–38.
Cummings, J.H., Pomare, E.W., Branch, W.J., Naylor, C.P., and Macfarlane, G.T. (1987) Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28:** 1221–1227.
Dobrindt, U., Agerer, F., Michaelis, K., Janka, A., Buchrieser, C., Samuelson, M., *et al*. (2003) Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J Bacteriol* **185:** 1831–1840.
Escobar-Páramo, P., Clermont, O., Blanc-Potard, A.B., Bui, H., Le Bouguénec, C., and Denamur, E. (2004) A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol Biol Evol* **21:** 1085–1094.
Ewers, C., Li, G., Wilking, H., Kieβling, S., Alt, K., Antáo, E.M., *et al*. (2007) Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: how closely related are they? *Int J Med Microbiol* **297:** 163–176.
Fang, G., Rocha, E., and Danchin, A. (2005) How essential are nonessential genes? *Mol Biol Evol* **22:** 2147–2156.
Ferenci, T., Galbiati, H.F., Betteridge, T., Phan, K., and Spira, B. (2011) The constancy of global regulation across a species: the concentrations of ppGpp and RpoS are strain-specific in *Escherichia coli*. *BMC Microbiol* **11:** 62.
von Gabain, A., Belasco, J.G., Schottel, J.L., Chang, A.C., and Cohen, S.N. (1983) Decay of mRNA in *Escherichia coli*: investigation of the fate of specific segments of transcripts. *Proc Natl Acad Sci USA* **80:** 653–657.
Garnett, J.A., Martínez-Santos, V.I., Saldaña, Z., Pape, T., Hawthorne, W., Chan, J., *et al*. (2012) Structural insights into the biogenesis and biofilm formation by the *Escherichia coli* common pilus. *Proc Natl Acad Sci USA* **109:** 3950–3955.
Gaschignard, J., Levy, C., Romain, O., Cohen, R., Bingen, E., Aujard, Y., and Boileau, P. (2011) Neonatal bacterial meningitis: 444 cases in 7 years. *Pediatr Infect Dis J* **30:** 212–217.
Giuliodori, A.M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., *et al*. (2010) The *cspA* mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol Cell* **37:** 21–33.
Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J., and Busby, S.J. (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci USA* **102:** 17693–17698.
Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., and Dobrindt, U. (2004) Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J Bacteriol* **186:** 5432–5441.
Harley, C.B., and Reynolds, R.P. (1987) Analysis of *E. coli* promoter sequences. *Nucleic Acids Res* **15:** 2343–2361.
Hernandes, R.T., Velsko, I., Sampaio, S.C., Elias, W.P., Robins-Browne, R.M., Gomes, T.A., and Girón, J.A. (2011) Fimbrial adhesins produced by atypical enteropathogenic *Escherichia coli* strains. *Appl Environ Microbiol* **77:** 8391–8399.
Herold, S., Paton, J.C., Srimanote, P., and Paton, A.W. (2009) Differential effects of short-chain fatty acids and iron on expression of *iha* in Shiga-toxigenic *Escherichia coli*. *Microbiology* **155:** 3554–3563.
Herzer, P.J., Inouye, S., Inouye, M., and Whittam, T.S. (1990) Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol* **172:** 6175–6181.
Hommais, F., Krin, E., Laurent-Winter, C., Soutourina, O., Malpertuy, A., Le Caer, J.P., *et al*. (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol Microbiol* **40:** 20–36.
House, B., Kus, J.V., Prayitno, N., Mair, R., Que, L., Chingcuanco, F., *et al*. (2009) Acid-stress-induced changes in enterohaemorrhagic *Escherichia coli* O157:H7 virulence. *Microbiology* **155:** 2907–2918.
Johnson, J.R., Oswald, E., O'Bryan, T.T., Kuskowski, M.A., and Spanjaard, L. (2002) Phylogenetic distribution of virulence-associated genes among *Escherichia coli* isolates associated with neonatal bacterial meningitis in the Netherlands. *J Infect Dis* **185:** 774–784.
Johnson, T.J., Wannemuehler, Y., Johnson, S.J., Stell, A.L., Doetkott, C., Johnson, J.R., *et al*. (2008) Comparison of extraintestinal pathogenic *Escherichia coli* strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. *Appl Environ Microbiol* **74:** 7043–7050.
Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2:** 123–140.
Kim, K.S. (2008) Mechanisms of microbial traversal of the blood-brain barrier. *Nat Rev Microbiol* **6:** 625–634.
King, T., Lucchini, S., Hinton, J.C., and Gobius, K. (2010) Transcriptomic analysis of *Escherichia coli* O157:H7 and K-12 cultures exposed to inorganic and organic acids in stationary phase reveals acidulant- and strain-specific acid tolerance responses. *Appl Environ Microbiol* **76:** 6514–6528.
Ko, J.H., Lee, S.J., Cho, B., and Lee, Y. (2006) Differential promoter usage of *infA* in response to cold shock in *Escherichia coli*. *FEBS Lett* **580:** 539–544.
Korea, C.G., Badouraly, R., Prevost, M.C., Ghigo, J.M., and Beloin, C. (2010) *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environ Microbiol* **12:** 1957–1977.
Korhonen, T.K., Valtonen, M.V., Parkkinen, J., Väisänen-Rhen, V., Finne, J., Ørskov, F., *et al*. (1985) Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. *Infect Immun* **48:** 486–491.
Lasaro, M.A., Salinger, N., Zhang, J., Wang, Y., Zhong, Z., Goulian, M., and Zhu, J. (2009) F1C fimbriae play an important role in biofilm formation and intestinal colonization by the *Escherichia coli* commensal strain Nissle 1917. *Appl Environ Microbiol* **75:** 246–251.

Leathart, J.B., and Gally, D.L. (1998) Regulation of type 1 fimbrial expression in uropathogenic *Escherichia coli*: heterogeneity of expression through sequence changes in the *fim* switch region. *Mol Microbiol* **28:** 371–381.

Lecointre, G., Rachdi, L., Darlu, P., and Denamur, E. (1998) *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol Biol Evol* **15:** 1685–1695.

Lehti, T.A., Bauchart, P., Heikkinen, J., Hacker, J., Korhonen, T.K., Dobrindt, U., and Westerlund-Wikström, B. (2010) Mat fimbriae promote biofilm formation by meningitis-associated *Escherichia coli*. *Microbiology* **156:** 2408–2417.

Lehti, T.A., Heikkinen, J., Korhonen, T.K., and Westerlund-Wikström, B. (2012a) The response regulator RcsB activates expression of Mat fimbriae in meningitic *Escherichia coli*. *J Bacteriol* **194:** 3475–3485.

Lehti, T.A., Bauchart, P., Dobrindt, U., Korhonen, T.K., and Westerlund-Wikström, B. (2012b) The fimbriae activator MatA switches off motility in *Escherichia coli* by repression of the flagellar master operon *flhDC*. *Microbiology* **158:** 1444–1455.

Liang, S.T., Dennis, P.P., and Bremer, H. (1998) Expression of *lacZ* from the promoter of the *Escherichia coli spc* operon cloned into vectors carrying the W205 *trp–lac* fusion. *J Bacteriol* **180:** 6090–6100.

Majdalani, N., and Gottesman, S. (2005) The Rcs phosphorelay: a complex signal transduction system. *Annu Rev Microbiol* **59:** 379–405.

Martínez-Santos, V.I., Medrano-López, A., Saldaña, Z., Girón, J.A., and Puente, J.L. (2012) Transcriptional regulation of the *ecp* operon by EcpR, IHF, and H-NS in attaching and effacing *Escherichia coli*. *J Bacteriol* **194:** 5020–5033.

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., *et al*. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE* **4:** e7526.

Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Mirmonsef, P., Gilbert, D., Zariffard, M.R., Hamaker, B.R., Kaur, A., Landay, A.L., and Spear, G.T. (2011) The effects of commensal bacteria on innate immune responses in the female genital tract. *Am J Reprod Immunol* **65:** 190–195.

Mobley, H.L., Jarvis, K.G., Elwood, J.P., Whittle, D.I., Lockatell, C.V., Russell, R.G., *et al*. (1993) Isogenic P-fimbrial deletion mutants of pyelonephritogenic *Escherichia coli*: the role of αGal(1-4)βGal binding in virulence of a wild-type strain. *Mol Microbiol* **10:** 143–155.

Mohanty, B.K., and Kushner, S.R. (1999) Analysis of the function of *Escherichia coli* poly(A) polymerase I in RNA metabolism. *Mol Microbiol* **34:** 1094–1108.

Moreno, E., Johnson, J.R., Pérez, T., Prats, G., Kuskowski, M.A., and Andreu, A. (2009) Structure and urovirulence characteristics of the fecal *Escherichia coli* population among healthy women. *Microbes Infect* **11:** 274–280.

Moriel, D.G., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., *et al*. (2010) Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* **107:** 9072–9077.

Moulin-Schouleur, M., Schouler, C., Tailliez, P., Kao, M.R., Brée, A., Germon, P., *et al*. (2006) Common virulence factors and genetic relationships between O18:K1:H7 *Escherichia coli* isolates of human and avian origin. *J Clin Microbiol* **44:** 3484–3492.

Moulin-Schouleur, M., Répérant, M., Laurent, S., Brée, A., Mignon-Grasteau, S., Germon, P., *et al*. (2007) Extraintestinal pathogenic *Escherichia coli* strains of avian and human origin: link between phylogenetic relationships and common virulence patterns. *J Clin Microbiol* **45:** 3366–3376.

de Muinck, E.J., Øien, T., Storrø, O., Johnsen, R., Stenseth, N.C., Rønningen, K.S., and Rudi, K. (2011) Diversity, transmission and persistence of *Escherichia coli* in a cohort of mothers and their infants. *Environ Microbiol Rep* **3:** 352–359.

Müller, C.M., Dobrindt, U., Nagy, G., Emödy, L., Uhlin, B.E., and Hacker, J. (2006) Role of histone-like proteins H-NS and StpA in expression of virulence determinants of uropathogenic *Escherichia coli*. *J Bacteriol* **188:** 5428–5438.

Nowrouzian, F.L., Wold, A.E., and Adlerberth, I. (2005) *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J Infect Dis* **191:** 1078–1083.

Nowrouzian, F.L., Adlerberth, I., and Wold, A.E. (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.

Nuccio, S.P., and Bäumler, A.J. (2007) Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev* **71:** 551–575.

Nuccio, S.P., Chessa, D., Weening, E.H., Raffatellu, M., Clegg, S., and Bäumler, A.J. (2007) SIMPLE approach for isolating mutants expressing fimbriae. *Appl Environ Microbiol* **73:** 4455–4462.

Obata-Yasuoka, M., Ba-Thein, W., Tsukamoto, T., Yoshikawa, H., and Hayashi, H. (2002) Vaginal *Escherichia coli* share common virulence factor profiles, serotypes and phylogeny with other extraintestinal *E. coli*. *Microbiology* **148:** 2745–2752.

Olsén, A., Jonsson, A., and Normark, S. (1989) Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*. *Nature* **338:** 652–655.

Oshima, T., Ishikawa, S., Kurokawa, K., Aiba, H., and Ogasawara, N. (2006) *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res* **13:** 141–153.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., *et al*. (1999) The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* **67:** 546–553.

Pouttu, R., Westerlund-Wikström, B., Lång, H., Alsti, K., Virkola, R., Saarela, U., *et al*. (2001) *matB*, a common fimbrillin gene of *Escherichia coli*, expressed in a genetically conserved, virulent clonal group. *J Bacteriol* **183:** 4727–4736.

Pratt, L.A., and Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol Microbiol* **30:** 285–293.

Prud'homme-Généreux, A., Beran, R.K., Iost, I., Ramey, C.S., Mackie, G.A., and Simons, R.W. (2004) Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'. *Mol Microbiol* **54:** 1409–1421.

de Ree, J.M., Schwillens, P., and van den Bosch, J.F. (1986) Monoclonal antibodies for serotyping the P fimbriae of uropathogenic *Escherichia coli*. *J Clin Microbiol* **24:** 121–125.

Rendón, M.A., Saldaña, Z., Erdem, A.L., Monteiro-Neto, V., Vázquez, A., Kaper, J.B., *et al.* (2007) Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization. *Proc Natl Acad Sci USA* **104:** 10637–10642.

Repoila, F., and Gottesman, S. (2001) Signal transduction cascade for regulation of RpoS: temperature regulation of DsrA. *J Bacteriol* **183:** 4012–4023.

Rhen, M., Knowles, J., Penttilä, M.E., Sarvas, M., and Korhonen, T.K. (1983) P fimbriae of *Escherichia coli*: molecular cloning of DNA fragments containing the structural genes. *FEMS Microbiol Lett* **19:** 119–123.

Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16:** 276–277.

Robbins, J.B., McCracken, G.H., Jr, Gotschlich, E.C., Ørskov, F., Ørskov, I., and Hanson, L.A. (1974) *Escherichia coli* K1 capsular polysaccharide associated with neonatal meningitis. *N Engl J Med* **290:** 1216–1220.

Saldaña, Z., Erdem, A.L., Schüller, S., Okeke, I.N., Lucas, M., Sivananthan, A., *et al.* (2009) The *Escherichia coli* common pilus and the bundle-forming pilus act in concert during the formation of localized adherence by enteropathogenic *E. coli*. *J Bacteriol* **191:** 3451–3461.

Sambrook, J., and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sarff, L.D., McCracken, G.H., Schiffer, M.S., Glode, M.P., Robbins, J.B., Ørskov, I., and Ørskov, F. (1975) Epidemiology of *Escherichia coli* K1 in healthy and diseased newborns. *Lancet* **1:** 1099–1104.

Sayed, A.K., and Foster, J.W. (2009) A 750 bp sensory integration region directs global control of the *Escherichia coli* GadE acid resistance regulator. *Mol Microbiol* **71:** 1435–1450.

Scaletsky, I.C., Aranda, K.R., Souza, T.B., and Silva, N.P. (2010) Adherence factors in atypical enteropathogenic *Escherichia coli* strains expressing the localized adherence-like pattern in HEp-2 cells. *J Clin Microbiol* **48:** 302–306.

Selander, R.K., Korhonen, T.K., Väisanen-Rhen, V., Williams, P.H., Pattison, P.E., and Caugant, D.A. (1986) Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. *Infect Immun* **52:** 213–222.

Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M., and Rosenow, C. (2003) Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res* **13:** 216–223.

Sims, G.E., and Kim, S.H. (2011) Whole-genome phylogeny of *Escherichia coli/Shigella* group by feature frequency profiles (FFPs). *Proc Natl Acad Sci USA* **108:** 8329–8334.

Stoebel, D.M., Free, A., and Dorman, C.J. (2008) Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology* **154:** 2533–2545.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24:** 1596–1599.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010) The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* **8:** 207–217.

Thomassin, J.L., Brannon, J.R., Gibbs, B.F., Gruenheid, S., and Le Moual, H. (2012) OmpT outer membrane proteases of enterohemorrhagic and enteropathogenic *Escherichia coli* contribute differently to the degradation of human LL-37. *Infect Immun* **80:** 483–492.

Totsika, M., Beatson, S.A., Holden, N., and Gally, D.L. (2008) Regulatory interplay between *pap* operons in uropathogenic *Escherichia coli*. *Mol Microbiol* **67:** 996–1011.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., *et al.* (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* **5:** e1000344.

Ueguchi, C., Suzuki, T., Yoshida, T., Tanaka, K., and Mizuno, T. (1996) Systematic mutational analysis revealing the functional domain organization of *Escherichia coli* nucleoid protein H-NS. *J Mol Biol* **263:** 149–162.

Walk, S.T., Alm, E.W., Calhoun, L.M., Mladonicky, J.M., and Whittam, T.S. (2007) Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ Microbiol* **9:** 2274–2288.

Watt, S., Lanotte, P., Mereghetti, L., Moulin-Schouleur, M., Picard, B., and Quentin, R. (2003) *Escherichia coli* strains from pregnant women and neonates: intraspecies genetic distribution and prevalence of virulence factors. *J Clin Microbiol* **41:** 1929–1935.

White-Ziegler, C.A., and Davis, T.R. (2009) Genome-wide identification of H-NS-controlled, temperature-regulated genes in *Escherichia coli* K-12. *J Bacteriol* **191:** 1106–1110.

Zdziarski, J., Svanborg, C., Wullt, B., Hacker, J., and Dobrindt, U. (2008) Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? *Infect Immun* **76:** 695–703.

Zhang, L., Foxman, B., and Marrs, C. (2002) Both urinary and rectal *Escherichia coli* isolates are dominated by strains of phylogenetic group B2. *J Clin Microbiol* **40:** 3951–3955.

Supporting information

Additional supporting information may be found in the online version of this article.