MICROBIAL GENOMICS

RESEARCH ARTICLE

Abd El Ghany *et al.*, *Microbial Genomics* 2021;7:000554

DOI 10.1099/mgen.0.000554

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Functional analysis of colonization factor antigen I positive enterotoxigenic *Escherichia coli* identifies genes implicated in survival in water and host colonization

Moataz Abd El Ghany[1,2,3,4,*], Lars Barquist[5,6], Simon Clare[7], Cordelia Brandt[7], Matthew Mayho[7], Enrique Joffre´[8], Åsa Sjöling[8], A. Keith Turner[7,9], John D. Klena[10], Robert A. Kingsley[7,9], Grant A. Hill-Cawthorne[11], Gordon Dougan[7,12] and Derek Pickard[7,12,*]

Abstract

Enterotoxigenic *Escherichia coli* (ETEC) expressing the colonization pili CFA/I are common causes of diarrhoeal infections in humans. Here, we use a combination of transposon mutagenesis and transcriptomic analysis to identify genes and pathways that contribute to ETEC persistence in water environments and colonization of a mammalian host. ETEC persisting in water exhibit a distinct RNA expression profile from those growing in richer media. Multiple pathways were identified that contribute to water survival, including lipopolysaccharide biosynthesis and stress response regulons. The analysis also indicated that ETEC growing *in vivo* in mice encounter a bottleneck driving down the diversity of colonizing ETEC populations.

DATA SUMMARY

The fastq files containing the TraDIS sequencing data generated in this study have been submitted to the European Nucleotide Archive (ENA) under study accession number PRJEB38958. The RNA-seq data have been deposited in the ArrayExpress database at EMBL (accession number E-MTAB-9364).

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

INTRODUCTION

Diarrhoeal diseases remain a serious health problem in many parts of the world, particularly impacting young children in resource-poor settings [1, 2]. Enterotoxigenic *Escherichia coli* (ETEC) are one of the main diarrhoeal pathogens, affecting children and adults, as well as travellers to endemic regions of the world [3, 4]. ETEC are commonly transmitted via the faecal/oral route, with survival in the environment, on food or in water as a key factor influencing this transmission. They colonize the intestine of humans, persisting on the epithelial surface of the lumen where they can secrete heat-labile (LT) [5] and/or heat-stable (ST) classes of toxins [6]. Colonization is facilitated by the surface-associated expression by

Received 14 December 2020; Accepted 09 March 2021; Published 10 June 2021

Author affiliations: [1]The Westmead Institute for Medical Research, University of Sydney, Sydney, Australia; [2]Marie Bashir Institute for Infectious Diseases and Biosecurity, University of Sydney, Sydney, Australia; [3]The Westmead Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia; [4]King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia; [5]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany; [6]Faculty of Medicine, University of Würzburg, Würzburg, Germany; [7]The Wellcome Trust Sanger Institute (WTSI), the Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; [8]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden; [9]Quadram Institute Bioscience, Norwich Research Park, Norwich, UK; [10]Centers for Disease Control and Prevention, Atlanta, Georgia, USA; [11]School of Public Health, the University of Sydney, Sydney, Australia; [12]Department of Medicine, University of Cambridge, Cambridge, UK.

***Correspondence:** Moataz Abd El Ghany, moataz.mohamed@sydney.edu.au; moataz.mohamed19@outlook.com; Derek Pickard, djjp2@medschl.cam.ac.uk

Keywords: CFA/I; ETEC; H10407; RNAseq; TraDIS; water survival and colonization.

Abbreviations: cDNA, complementary DNA; CF, colonization factor; CFA/I, colonization factor antigen I; cfu, colony forming unit; EAEC, enteroaggregative *Escherichia coli*; EMBL, European Molecular Biology Laboratory; ETEC, enterotoxigenic *Escherichia coli*; FUT2, fucosyltransferase 2; gDNA, genomic DNA; KEGG, Kyoto Encyclopedia of Genes and Genomes; Le^A, Lewis A glycans; LLR, Log2 likelihood ratio; LPS, lipopolysaccharide; LT, heat-labile enterotoxin; RPKM, reads per kilobase per million; ST, heat-stable enterotoxin; TPM, transcripts per million; TraDIS, transposon directed insertion-site sequencing; UI, unique insertion; VBNC, viable but not cultivable.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Two supplementary figures and eight supplementary tables are available with the online version of this article.

ETEC of fimbrial structures, known as colonization factors (CFs), that target receptors on epithelial cells at the luminal surface of the intestine [7, 8]. Human ETEC can express different combinations of antigenically and genetically distinct fimbriae, which target diverse receptors. At least 30 such colonization fimbriae have been identified to date in human ETEC and this number is expected to increase as more novel CFs are identified [9, 10]. Colonization factor antigen I (CFA/I) is one of the most commonly encountered human ETEC fimbriae [9].

Both the toxin and CF genetic determinants of ETEC are frequently encoded by plasmids, although an interplay between plasmid and chromosomal genes is probably required for successful transmission and virulence. Genomic analysis has shown that the global ETEC population is dominated by a limited number of highly successful ETEC clades that have spread to different parts of the world [11]. The reasons why these clades are successful remains unknown, although clades expressing CFA/I are of this type.

CFA/I is a rigid rod-like structure of the class 5a fimbriae, also including CS1, CS2, CS4, CS17 and CS19 [12]. The assembly of CFA/I involves the expression of plasmid-encoded genes clustered in the *cfaABCE* operon, which is mediated via the alternate chaperone pathway that requires CfaA to export the pilus across the periplasm. The major subunit CfaB promotes binding to various glycosphingolipids, including asialo-GM1 [13] and Lewis A (Le^A) glycans present in the mucosa of the small intestine of young children and adults carrying non-functional alleles of *FUT2* [14–16]. The tip protein, CfaE, can promote binding to other receptors on erythrocytes and intestinal epithelial cells [17, 18].

Although both the toxins and CF expressed by human ETEC have been extensively studied, relatively little is known about other genes that can contribute to the ability of ETEC to transmit between hosts and cause disease. The expression of both toxins and CFs is regulated (e.g. by temperature in the case of CFs), but it is likely that other genes are involved. In this study we have used a combination of transposon mutagenesis, transcriptomics, microbial genetics and various selection models to characterize ETEC expressing CFA/I and identify a number of novel survival and virulence mechanisms.

Impact Statement

Enterotoxigenic *Escherichia coli* (ETEC) are characterized by their ability to survive in water and, from there, infect mammalian hosts. However, relatively little is known about the mechanisms employed by the bacteria to survive in these distinct ecological niches. Here, we use a combination of high-throughput mutagenesis and transcriptomics to identify multiple novel survival mechanisms employed by ETEC H10407 to persist in water and within a mammalian host (mice). This approach provides novel insight into the molecular basis of these survival behaviours as well as the pathogenesis of ETEC H10407. The data have potential applications in the control of disease transmission and the treatment of associated infections.

METHODS

Bacteria

CFA/I-positive ETEC H10407 was obtained from the global collection deposited at Gothenburg University, Sweden. This isolate was used for the construction of a transposon library and the mutants developed in this study. The CFs and enterotoxin profiles of H10407 were confirmed using previously described multiplex PCR [19]. The expression of CFA/I was also evaluated using a previously described colony blot assay [20] and inhibitory ELISA [21].

Construction of transposon mutant library in the CFA/I-positive ETEC H10407

A derivative of the transposon EZ-Tn5 (Epicentre) was phosphorylated with polynucleotide kinase (New England BioLabs) and incubated with EZ-Tn5 transposases (Epicentre) at 37 °C for 1 h to prepare the transposome that was stored at −20 °C until use. An overnight culture of H10407 was diluted 1:100 in LB broth and the bacteria were allowed to grow to an OD_{600} of 0.3–0.4. The bacterial cells were then harvested, washed three times in 0.5 volume of 10% chilled glycerol and re-suspended in 1:1000 of the volume of the initial culture. Then, 40 µl of freshly prepared H10407 electrocompetent cells was mixed with 0.3–0.5 µl of transposomes and electrotransformed in a 0.2 cm cuvette (Bio-Rad Laboratories) using bacteria pre-setting parameters (2.5 kV, 25 µF and 200 Ω) in the Gene Pulser Xcell Electroporation System (Bio-Rad Laboratories). Cells were immediately re-suspended in 1 ml of SOC medium (Invitrogen) and incubated at 37 °C for 2 h before being spread on LB agar supplemented with kanamycin (20 µg ml^{-1}). Individual colonies from 20 transformation batches were counted, scraped from plates using LB, and pooled to construct an initial transposon library consisting of ~1.1 million mutants of H10407. The library was aliquoted and kept in 25–30% glycerol stock at −80 °C.

Preparation of TraDIS sequencing libraries

Genomic DNA (gDNA) of the original (input) and screened (output) mutant pools was extracted using a Qiagen Genomic-tip 100/G (Qiagen) and used to prepare TraDIS sequencing libraries as described previously [22]. Briefly, ~2 µg of gDNA was fragmented to an insert size of ~300 bp using a Covaris Focused-ultrasonicator (Covaris). The fragmented DNA was end repaired, A-tailed and adapter-ligated using an Illumina DNA fragment library preparation kit (New England BioLabs) according to the manufacturer's instructions. The libraries were then enriched by 10–20 cycles of PCR using a transposon-specific forward primer and a reverse primer that included Illumina flow cell binding sites [22]. The enriched

libraries were purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified on an Agilent DNA1000 chip (Agilent Technologies) following the manufacturer's instructions. The peak size of the libraries was checked on a Bioanalyzer (Agilent Technologies) and the concentration of the libraries was quantified by quantitative PCR (qPCR). The libraries were sequenced on a HiSeq2000 Illumina platform as single-end reads using a specialized recipe. The sequencing was conducted at the KAUST Bioscience Core Laboratory.

Bioinformatic and statistical analysis

The reads were processed using the Bio::Tradis toolkit [23]. Briefly, reads were filtered for 10 bases matching the expected transposon tag sequence. Filtered reads were then mapped against H10407 reference chromosomal (accession number FN649414) and accessory genomes (accession numbers: FN649415, FN649416, FN649417 and FN649418) using SMALT version 0.7.2, and transposon insertion sites and insertion indices (obtained by dividing the number of unique insertion sites by the gene length) were determined for every gene. The essentiality of every gene in the chromosome and accessory genomes of H10407 was determined as described previously [24]. Gamma distributions were fitted to the two modes observed in the insertion index distribution, and Log2 likelihood ratios (LLRs) were calculated between the fitted distributions to classify genes as essential or non-essential, with cutoffs set at LLRs of less than −2 or greater than 2, respectively. For the calculation of fold change, Log2 fold changes and significance of changes in mutant abundance before (input) and after selection (output; biological replicates of mutant pool screened under the prescribed condition) were calculated using edgeR [25] after normalization using the TMM method [26].

RNA isolation and preparation of RNAseq data

For global transcriptomic analysis, the prototype CFA/I-positive strain H10407 was grown under different *in vitro* conditions to reach the early exponential growth phase (OD_{600} of 0.4). Briefly, H10407 was cultured aerobically in regular LB media (pH 7) (or adjusted to pH of 5 or 9 using appropriate buffers) at 37 °C with shaking. For anaerobic culturing, H10407 was grown overnight in 150 ml of regular LB media at 37 °C in a 200 ml conical flask fitted with a permeable filter in the flask lid (to allow gaseous exchange) and shaken in an anaerobic chamber (Whitley A35 workstation). One millilitre of this culture was added to 100 ml of LB media in a 200 ml conical flask and shaken at 37 °C in the anaerobic cabinet.

For water survival experiments, H10407 that was grown aerobically in regular LB media was washed once before being suspended in water to a final concentration of 1×10^6 c.f.u. ml^{-1} and incubated at ambient temperature (22–25 °C) without shaking for 48 h. For each culturing and survival condition, duplicate independent biological replicates were used to extract bacterial RNA using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. rRNA was depleted using Invitrogen's RiboMinus Prokaryotic Kit (Invitrogen) and contaminating DNA was removed using RNase-free DNase (Invitrogen). The cDNA libraries were prepared using an Illumina TruSeq RNA sample preparation kit, pooled in equimolar concentrations and sequenced on an Illumina HiSeq2000 platform at the KAUST Bioscience Core Laboratory.

Transcriptomic analysis

For each sample, the reads were mapped against H10407 reference chromosomal (accession number FN649414) and accessory genomes (accession numbers: FN649415, FN649416, FN649417 and FN649418) using BWA v0.7.12 [27] to produce a BAM file using default parameters but with the quality threshold for read trimming set to 15 (q=15) and maximum insert size of 300 bp. Gene expression values were computed from the read alignments to the coding sequencing to generate the number of reads mapping and RPKM (reads per kilobase per million) [28]. Only reads with a mapping quality score of 10 were included in the count.

For the purpose of clustering, read counts were converted into transcripts per million (TPM) [29], and averaged across replicates. Genes that were highly expressed in at least one condition (TPM>100) were considered for further analysis, along with a small selection of biologically relevant genes that did not meet this stringent threshold. Hierarchical clustering was performed on log-transformed TPM values using the base R function hclust with default parameters and visualized using heatmap 2.

The Bio-RNASeq pipeline, which makes use of the DESeq Bioconductor R package [30], was used to perform the differential gene expression analysis between pairs of conditions (https://github.com/sanger-pathogens/Bio-RNASeq). The differential gene expression was visualized using heatmapper [31].

Preparation of defined knock-outs in H10407

H10407 mutant derivatives harbouring targeted independent deletions in the chromosomal (*atpB*, *atpC*, *atpD*, *galU*, *gmhD*, *waaC*, *waaF* and *waaO*) and plasmid genes (*aatC*, *cexE* and *cfaE*) were constructed by replacing the wild type gene with a DNA cassette encoding a kanamycin gene using the λ–Red recombination method [32].

Selection for survival in mice and competitive assay

To understand the basis of ETEC persistence in the gut of mice, the transposon input library was used to orally inoculate a group of six streptomycin-treated C57BL/6 mice (5–6 weeks old). Briefly, the mice were orally inoculated with 20 mg of streptomycin 24 h before being infected with the mutant library at a dose of 10^{10} c.f.u. per mouse. The caeca of the infected mice were harvested 48 h post-infection, washed with PBS, homogenized and plated on LB agar supplemented with appropriate antibiotics.

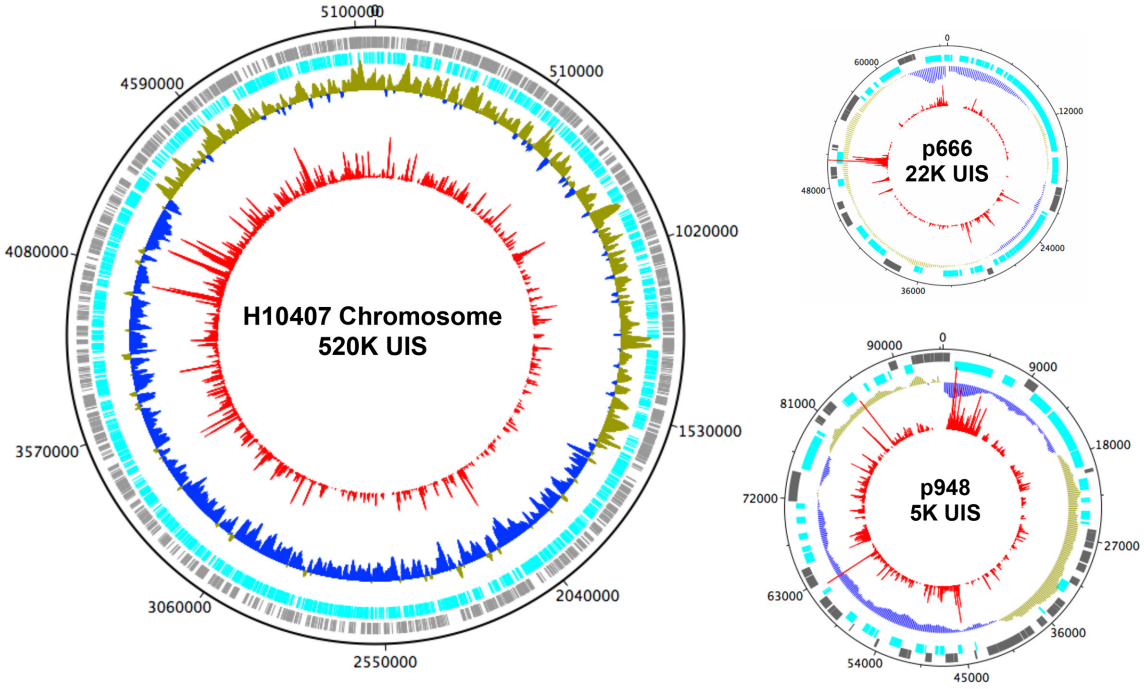

Fig. 1. Frequency and distribution of transposons across the genomes of the ETEC H10407 chromosome and associated virulence plasmids. The unique insertion sites (UIS) across the H10407 genome and associated plasmids p666 and p948 are shown as red lines on the innermost circle. The line length correlates with the frequency of insertions at a given site. The blue/sycamore track refers to the GC skew. The grey and cyan tracks refer to the genes on the forward and reverse strands, respectively.

Competitive infection assays were used to examine the impaired capability of null mutant derivatives in *aatC*, *cexE* and *cfaE* to colonize the caecum of streptomycin-treated mice. For each mutant, a group of five C57BL/6 mice (5–6 weeks old) were infected by oral gavage with an approximately 1:1 mixture of mutants and isogenic parent strains at a dose of approximately 1×10^{10} c.f.u. per mouse. The caeca of the infected mice were harvested 48 h post-infection, washed with sterile PBS and homogenized. Faecal pellets were also collected and homogenized in sterile PBS. Serial dilutions of homogenized caecum, caecal contents (wash) and faecal pellets were plated on LB agar supplemented with appropriate antibiotics and enumerated.

RESULTS

Construction of a transposon library in the CFA/I-positive ETEC H10407

The CFA/I-positive ETEC H10407 was transformed with DNA encoding Tn5 transposomes and transformants were selected on LB plates supplemented with kanamycin. Individual colonies from different transformation batches were pooled to construct an initial transposon library of approximately 1.1 million mutants. DNA from two independent samples was prepared and these were sequenced from the transposon ends using the TraDIS method (Fig. S1, available in the online version of this article). High-quality sequencing reads that contained the 10 base transposon tag sequence were then mapped onto the ETEC H10407 genome (FN649414; 5153435 bp) including the associated plasmids p52 (FN649415; 5175 bp), p58 (FN649416; 5800 bp), p666 (FN649417; 66681 bp) and p948 (FN649418; 94797 bp) in order to simultaneously identify all of the transposon insertion sites.

The transposon unique insertion (UI) sites were densely and randomly distributed throughout the chromosome of H10407 and associated plasmids (Fig. 1). Approximately 520 000, 5000 and 22 000 UIs were identified in the chromosome of H10407 and plasmids p948 and p666, respectively. This resulted in an average of one insertion every 9.9, 18 and 3 bp of H10407 chromosome, plasmids p948 and p666, respectively, which confirms the generation of a highly saturated transposon library (Table S1). To identify the essential genes of CFA/I--positive ETEC H10407, an insertion index was calculated for each gene, normalizing the number of insertions to gene length. Statistical analysis of the bimodal distribution of insertion indices (fitted to a gamma distribution) identified 425 chromosomal genes as candidates essential for H10407 (Table S2). Of these, 247 genes have previously been predicted by TraDIS to be essential in both *E. coli* K12 BW25113 [33] and the uropathogenic *E. coli* EC958 [34]. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis demonstrated that all of the genes that were deemed essential for H10407 have previously been found to be involved in key pathways and vital cellular processes (Table S2).

The RNA expression profiles of H10407 cultured under different conditions were next determined, including in water, a common source of ETEC infection and persistence in the environment. For each gene, TPM values across duplicates were averaged; log TPM values were calculated and these were used to perform an unsupervised hierarchical clustering (Fig. 2). A relatively conserved gene expression profile was detected when *E. coli* H10407 was cultured under different *in vitro* conditions including aerobic [neutral (pH 7), alkaline (pH 9) and acidic media (pH 5)] and anaerobic growth. In contrast, many of these genes were significantly downregulated when the bacteria were placed in water for 48 h (see Methods for conditions) (Fig. 2). These included genes associated with cell division (*n*=16), translation (protein biosynthesis), production of cellular energy/ATP (*n*=16) and ribosomal subunits (*n*=56). This analysis indicates that H10407 uses different genes for water survival compared to growth in richer medium.

TraDIS selection for long-term survival in water

In order to identify H10407 genes that facilitate the survival of H10407 in water, the transposon library consisting of 10^6 mutants was suspended in sterile water at a density of approximately 10^{10} c.f.u. ml^{-1} (input) and incubated at ambient temperature (22–25 °C) without shaking for 5 weeks. At this time point, duplicates of water samples were collected and centrifuged to recover bacterial cells (output). gDNA was isolated from the input and output samples and subjected to TraDIS analysis. In this regard, a comparison was made between input and 5 week screened libraries to identify the genes required for long-term survival in water.

A total of 56 genes with at least a two-fold decrease in the number of transposon insertions per gene in output versus input pools (log_2 equal to or lower than −2, $P<0.05$ and false discovery rate of 5%) were identified as candidates for the long-term survival of H10407 in water (Table S3). The majority of genes identified as essential candidates for the long-term survival of H10407 in water had 1.3–7 log_2 decrease in the unique transposon insertions identified per gene in 5 week screened libraries as compared to 48 h screened libraries. This highlighted the essentiality of these candidate genes for the survival and persistence of H10407 in water.

KEGG analysis demonstrated that many of the candidate water survival genes fell into distinct pathways, including lipopolysaccharide biosynthesis (*n*=12; *waaC*, *waaF*, *waaG*, *waaL*, *waaO*, *waaP*, *waaR*, *waaV*, *gmhA*, *gmhC*, *gmhB* and *gmhD*), energy metabolism, oxidative phosphorylation (*n*=8; *atpC*, *atpD*, *atpG*, *atpA*, *atpH*, *atpF*, *atpE* and *atpB*), biosynthesis of antibiotics (*n*=5; *galU*, *rfbA*, *rfbB*, *pgi* and *fbp*), microbial metabolism in diverse environments (*n*=2; *pgi* and *fbp*), biofilm formation (*n*=2; *rpoS* and *dksA*), and peptidoglycan biosynthesis and degradation (*n*=2; *ldcA* and *lpp*) (Table S4).

The screen also identified four genes implicated in transcription machinery (*rpoS*, *rseA*, *rseB* and *dksA*). *rpoS* codes for the major alternative sigma factor 38 (σ^{38}) and regulates stationary phase transcription under stress conditions [35–37]. Additionally, *rseA* and *rseB* that code for σ^E-negative regulator proteins and *dksA* encoding a suppressor protein (transcription factor), which have been shown to be key determinants of the *E. coli* stringent response (i.e. sensing and responding to nutritional status), were identified [38, 39].

The majority of these candidate water survival genes (*n*=41) were significantly downregulated under water stress compared with aerobic growth in LB (Tables S3 and S5). These include the genes associated with lipopolysaccharide biosynthesis, energy metabolism (oxidative phosphorylation), transcription machinery and biofilm formation. The expression analysis also revealed differential expression of genes associated with flagellar assembly, chemotaxis, motility and fimbrial formation. The majority of fimbrial (*n*=26), phage (*n*=94), chemotaxis (*n*=6; *cheZ*, *cheB*, *cheR*, *tap*, *tar* and *motA*), curli assembly (*n*=5; *csgA*, *csgB*, *csgD*, *csgF* and *csgG*) and two CRISPR genes (ETEC_2945 and ETEC_2951) were relatively upregulated under water stress compared with aerobic growth (Table S6). In contrast, genes associated with flagellar assembly including those coding for flagellar hook (*flgD* and *flgE*), distal (*flgG*) and proximal rods (*flgB*, *flgC* and *flgF*) were significantly downregulated (Table S5). The analyses also identified differential expression in the chemotaxis genes *cheZ* and *cheY*, which interact with the *E. coli* flagellar motor switch complex and thus control the direction of flagella rotation, (impacting smooth swimming – tumbling) [40–43]. Both *cheY* and *cheZ* were upregulated under anaerobic growth while only *cheZ* was upregulated under water stress, compared to aerobic growth (Tables S6 and S7). Previous studies showed that motile behaviour of *E. coli* is a random motion of straight swimming (anticlockwise rotation of flagella) interrupted by tumbling (clockwise rotation of flagella induced by phosphorylated CheY) [40, 41]. Recently, an elevated level of CheZ (which increases the dephosphorylation of CheY-P) has been associated with low tumbling bias of swarming *E. coli* [44, 45]. This suggests that under water survival H10407 can perform high-speed swimming through altering the run tumble bias.

Confirmation of genes required for survival of H10407 in water

In order to confirm if genes identified by the TraDIS screen influenced the survival of H10407 in water, we constructed null mutant derivatives in eight genes and tested their viability in water. The genes *gmhD*, *waaF*, *waaC*, *waaO* (lipopolysaccharide biosynthesis), *atpC*, *atpD*, *atpB* (F_0F_1 ATPase involved in oxidative phosphorylation) and *galU* were inactivated in H10407 with a DNA cassette encoding the *kan* gene, conferring kanamycin resistance. At day 2 post-water incubation, the numbers of c.f.u. recovered from the mutants were slightly lower than that recovered from the wild type parent strain. All the mutants examined exhibited an impaired ability to survive in water following incubation for 5 weeks (Fig. 3). In each case, the c.f.u. recovered from mutant screening were significantly lower (1–2.5 log_{10} decrease) than those recovered

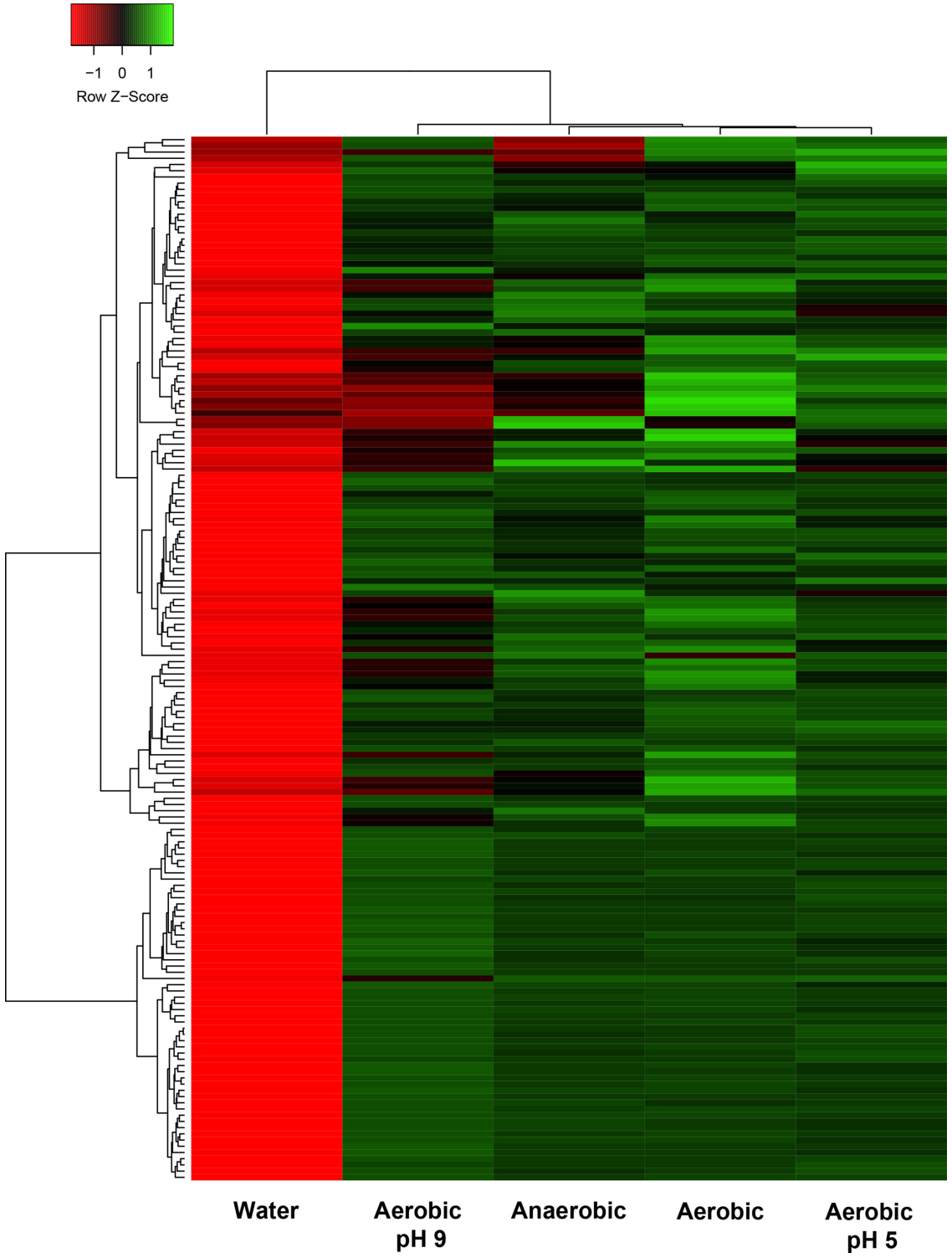

Fig. 2. The expression profiles of H10407 genes. Heat map showing the expression profiles of the essential genes of H10407 under different screening conditions, including water stress and *in vitro* growth. Each column represents the clustering related to the screening conditions. The values of transcript per million (TPM) across replicates were averaged; log TPM values were calculated and used to perform the hierarchical clustering shown. Little differential gene expression (green; higher TPM values) was detected among these genes under *in vitro* aerobic [neutral (pH 7), alkaline (pH 9) and acidic media (pH 5)] and anaerobic growth. In contrast, many of these genes were significantly downregulated (red; lower TPM values) under water stress screening, which provides further evidence for the essentiality of these genes for bacterial growth.

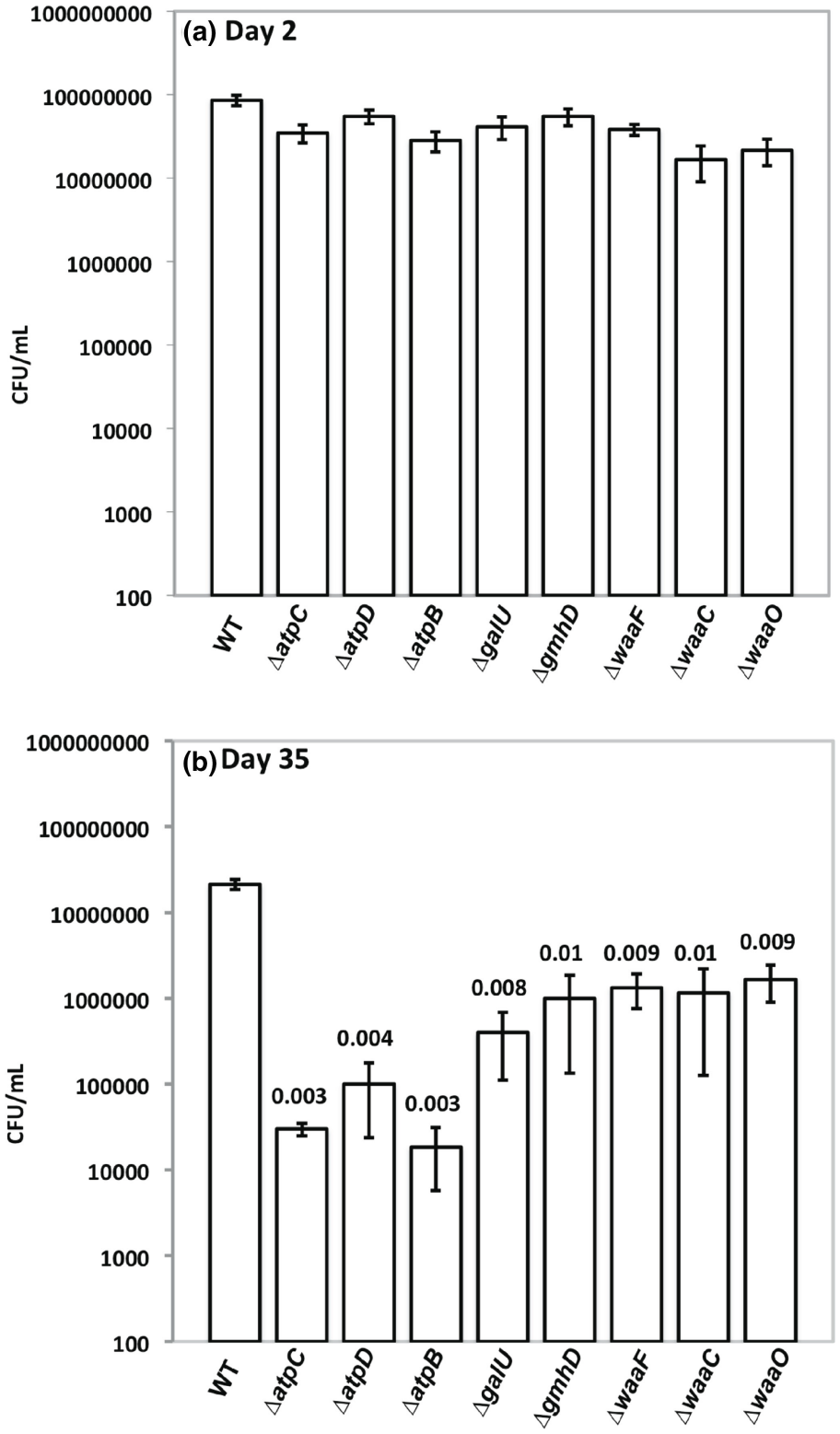

Fig. 3. Confirmation of candidate genes associated with long-term survival of ETEC H10407 in water. The mean of the c.f.u. recovered from the mutant derivatives and isogenic parent strain (wild type) at day 2 and after 5 weeks of incubation in water are plotted with the standard deviation of six biological replicates. At day 35 post-incubation, the c.f.u. recovered from the mutants were significantly lower than those recovered from the wild type parent equivalent derivative. The *P*-values (Student's *t*-test) are given for the comparison of each mutant derivative with the isogenic parent H10407.

from the wild type parent equivalent derivative, confirming the TraDIS data (Fig. 3).

Expression patterns of plasmid genes under different growth conditions

Interestingly, distinct expression patterns were identified between the genes encoded by plasmids p948 (that encodes ETEC-associated virulence genes, including the *cfaABCE* operon and heat-stable enterotoxin *st*) and p666 (a self-transmissible plasmid that encodes the heat-labile enterotoxin *elt*) [46] under anaerobic growth and water stress compared with aerobic growth. The majority of p666 genes, including those associated with conjugal transfer, were significantly upregulated under water stress, whereas the majority of p948 genes were significantly upregulated under anaerobic growth (Fig. 4). Anaerobic growth relatively upregulated the expression of many plasmid-encoded, virulence-associated genes, including the *cfaABCE* operon and *eatA*, encoding a mucin-degrading serine protease. This is in addition to genes associated with plasmid stability and maintenance (replication, partitioning and segregation) (Fig. 4). Differential transcriptional profiles associated with the anaerobic growth of H10407 are provided in Tables S7 and S8.

Selection for survival in mice

We were interested to see if genes associated with survival in water were involved in colonization of and persistence in the mammalian host. Consequently a streptomycin pretreatment model was used to explore the ability of *E. coli* H10407 and associated mutants to colonize mice. To this end, 5–6-week-old C57BL/6 mice were orally inoculated with 20 mg of streptomycin 24 h before infection to disrupt the resident microbiota. These mice were then challenged, by oral gavage, with wild type H10407 at a dose of ~10^{10} c.f.u. per mouse. On average, 2×10^{6} and 10^{5} c.f.u. g^{-1} were recovered from the faecal pellets collected from infected mice 2 and 4 days post-challenge, respectively. The colonizing capability of H10407 to colonize the mouse caecum, colon and small intestine was evaluated on days 1, 2 and 4 post-challenge. At all time points, high c.f.u. values were recovered from the caecum followed by colon and small intestine (Fig. S2). All mice survived the infection and, as expected, showed limited signs of pathology and other infection-associated phenotypes.

The H10407 mutant library was used to orally inoculate mice, and the output pools of bacteria that were capable of survival in the caeca of mice for 48 h post-infection were recovered from homogenized caeca. DNA from six independent output pools was prepared, pooled and subjected to massively parallel sequencing of transposon flanking regions. The bacterial load in mice challenged with the H10407 mutant pool was comparable to that in mice infected with the wild type H10407 challenge (data not shown). However, there was significant loss of the diversity of the mutant pool (complexity of H10407 mutant output pool) (~36 000, 189 and 2458 UIs were identified in the chromosome of H10407 and plasmids p948 and p666, respectively), compromising unbiased TraDIS analysis. An average of a unique insertion was identified every 153, 890 and 29 bp of the H10407 chromosome, and plasmids p948 and p666, respectively (Table S1). This indicated that the loss of pool genetic diversity was random in nature, which precluded TraDIS analysis of H10407 persitence in the mouse gut.

Evaluation of the impaired colonizing phenotype of H10407 in mice

A selected gene targeting approach was used to determine if introducing specific mutations associated with ETEC pathogenicity would compromise mouse colonization. To this end, *E. coli* H10407 null mutant derivatives in plasmid genes *aatC*, *cexE* (coding for non-fimbrial proteins whose implication for pathogenesis are unknown) and *cfaE* (coding for CFA/I minor pillin) that were upregulated under anaerobic growth were constructed and used to orally inoculate a group of streptomycin-treated C57BL/6 mice. Competitive infection assays were used to examine the ability of these mutants to colonize the caeca of mice for 48 h post-infection. These assays demonstrated that all mutants tested were significantly attenuated, characterized by an impaired capability to colonize the caecum of streptomycin-treated mice (Fig. 5). For all mutants examined, the output ratio (mutant:isogenic parent) recovered from the caeca (Fig. 5a), caecum contents (Fig. 5b) and faecal pellets (Fig. 5c) was significantly lower than that present in the inocula used for infection, confirming a role for these genes for H10407 colonization. Interestingly, *cfaE* mutants of H10407 were attenuated in mice. This is consistent with published data that have suggested that the minor pilin CfaE is one of the major virulence factors that mediate colonization of the host intestine by CFA/I-positive ETEC [47, 48].

DISCUSSION

Here we have used a combination of transposon mutagenesis and transcriptomic analysis to investigate the role of chromosomal and plasmid genes encoded by CFA/I-positive ETEC H10407 in water survival and colonization of the murine intestine. These conditions mimic two key stages of the infection/transmission cycle of this pathogen. The analyses demonstrated that both chromosomal- and plasmid-encoded genes are involved in the successful transit and persistence of H10407 within these different niches. The analyses identified particular functional categories playing key roles in the survival of H10407 under these distinct screening conditions. For instance, a number of genes involved in LPS and other carbohydrate modification pathways were found to be essential for the survival of H10407 in water. LPS is known to be important for virulence but its role in environmental survival is less well defined. The gene encoding HldE, which plays key roles in ETEC virulence through protein glycosylation and the correct configuration of LPS, was also identified in the screen. Importantly, *hldE* has previously been shown to be associated with reduced expression of CFA/I and flagellae, therefore impairing its ability to bind Caco2-cells [49]. Interestingly, genes involved in polysaccharide biosynthesis were also recently identified in a screen for survival in water

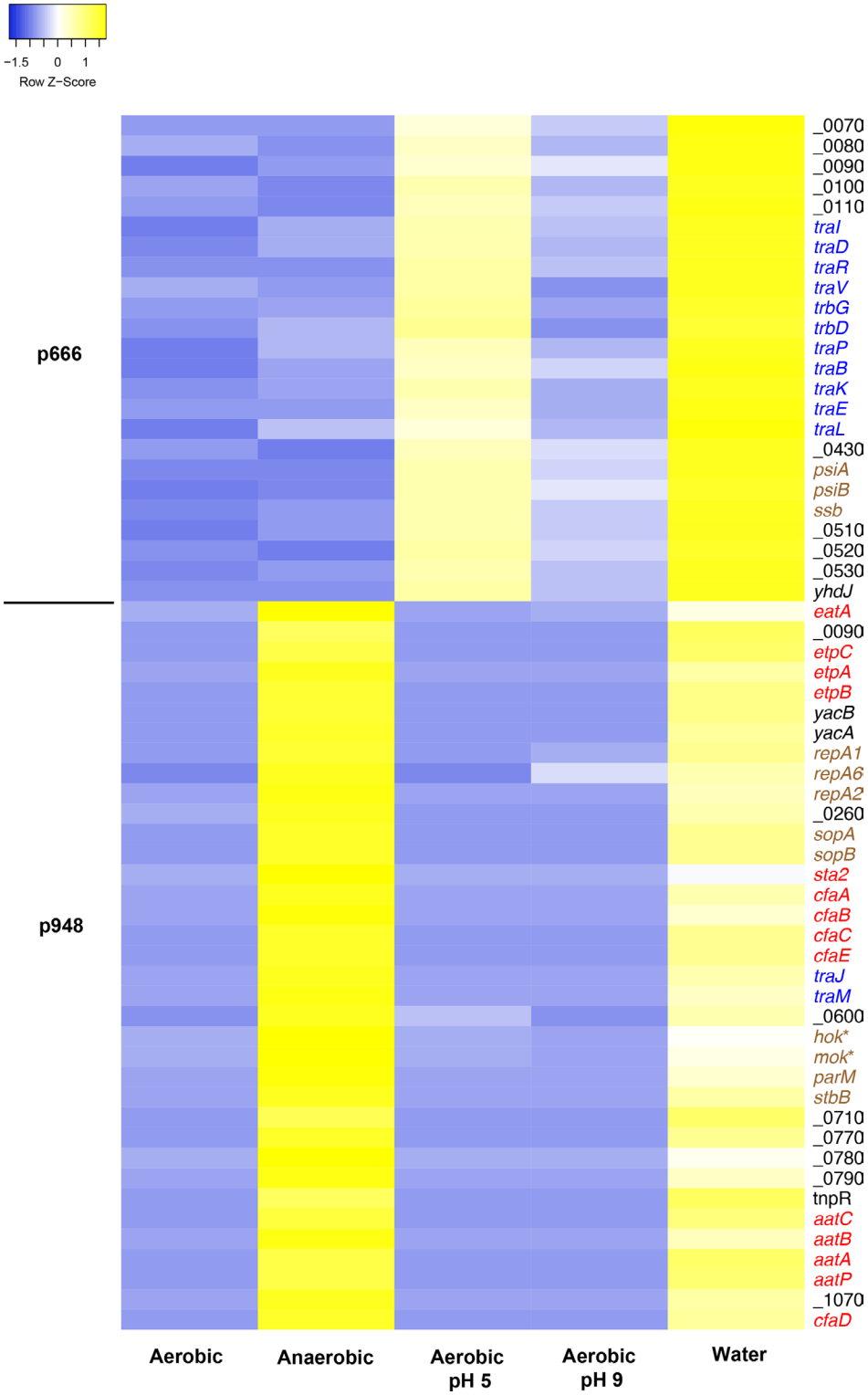

Fig. 4. Differential expression of plasmid genes under different environments. Heat map showing the expression profiles of H10407 plasmid genes (p948 and p666) under different screening conditions, including water stress and *in vitro* growth (log TPM values are plotted). The majority of p666 genes, contrary to those carried on p948, were significantly upregulated only under water stress, whereas the majority of p948 were significantly upregulated under anaerobic growth. The annotation was added to the right of the map and is highlighted based on the gene function. Blue, conjugal transport; red, colonization and virulence-associated; brown, plasmid stability and maintenance (replication, partitioning and segregation); and black, conserved of unknown function. *Toxin–antitoxin system.

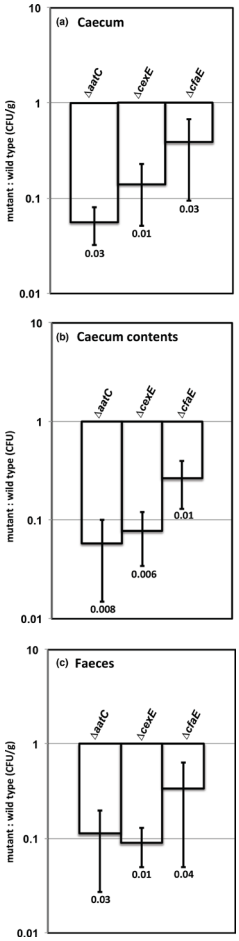

Fig. 5. Competitive infection assay confirming the impaired capabilities of H10407 mutants to colonize mouse intestines. Competitive infection assays were used to examine the impaired capability of mutants Δ*aatC*, Δ*cexE* and Δ*cfaE* to colonize the caecum of streptomycin-treated mice. For each mutant, a group of five C57BL/6 mice (5–6-week-old females) were treated orally with 20 mg streptomycin 24 h before infection. Mice were infected by oral gavage with an approximately 1:1 mixture of mutants and isogenic parent strains at a dose of approximately 1×10^{10} c. f. u. per mouse. The caeca of the infected mice were harvested 48 h post-infection, washed with sterile PBS and homogenized. Faecal pellets were also collected and homogenized in sterile PBS. Serial dilutions of homogenized caecum, caecal contents (wash) and faecal pellets were plated on LB agar supplemented with appropriate antibiotics. The bacteria recovered from the caecum (a), caecal contents (b) and faecal pellets (c) of infected mice were enumerated. The ratio of mutant : wild type strains is given as the mean±SD of the mean. The *P*-values (Mann–Whitney U test) are given for the comparison of each mutant derivative with the isogenic parent strain H10407. For all of the mutants tested, the output ratio was significantly lower than that present in the inoculums used in infection, which confirms the essentiality of these genes for H10407 to colonize the mouse intestine.

of another human-adapted enteric pathogen, *Salmonella enterica* serovar Typhi [50].

A number of different transcriptional regulators were identified in the screen. One key regulator identified, *rpoS* (sigma factor 38), in *E. coli* is known to regulate the transcription of genes associated with entry into the stationary phase and survival under starvation and stress conditions [51–53]. Thus, a role in water survival would be consistent with the function of this gene. Other transcriptional regulators (e.g. *dksA*, *rseA* and *rseB*) were identified as being required for efficient colonization of the mammalian intestines. For example, a previous study has found that *dksA* is required for *Salmonella* colonization of the chicken intestine and for virulence in newly hatched chicks [54].

We identified a number of key metabolic pathways and survival mechanisms implicated in the persistence of H10407 in water. These determinants included genes associated with the ability of bacterial cells to oxidize nutrients to produce energy (ATP synthesis) through oxidative phosphorylation complex V, which is a major energy provider of prokaryotic cells under aerobic conditions. The survival data on defined mutants were consistent with the TraDIS prediction. Many of the genes implicated were distinct from those that were observed when *S.* Typhi was maintained in water, with only limited overlap seen between the genes involved [50]. Interestingly, genes involved in similar metabolic processes were identified in both screens, including those involved in lipopolysaccharide and peptidoglycan biosynthesis. Collectively, these data suggest that ETEC and *S.* Typhi are characterized by somewhat distinct modes of survival in water. Previous studies [55, 56] have demonstrated the ability of ETEC to survive in fresh and marine water for up to 3 months by entering 'viable but not cultivable' (VBNC) status, with bacteria maintaining the enterotoxin gene while apparently not producing/secreting toxins. *S.* Typhi is notoriously difficult to culture from the environment [50]. A VBNC state may be mediated by transcriptional switching, potentially involving alternative sigma factors. RpoA directs the transcriptional activity of *E. coli* during starvation and under stress conditions [57]. In water, ETEC H10407 suppresses most vital cellular processes and virulence-associated mechanisms, including growth, replication, motility and assembly of virulence factors (with the majority of genes associated with these processes being downregulated). Thus, these regulatory pathways may play a key role in survival.

It has been well-established that ETEC colonize the human intestine through a number of adhesins and CFs that are mainly carried on plasmids. Our data demonstrate a key role of the minor pillin CfaE in mediating the colonization of mouse intestine by H10407. Consistently, prophylactic anti-CfaE antibodies [58] and prototype vaccines promoting anti-CfaE-based anti-adhesin responses have been shown to confer protection against H10407 challenge in mice [59], non-human primates [59, 60] and human volunteers [58].

Importantly, our data suggest the involvement of a number of plasmid genes (e.g. *cexE*, *aat* and *eatA*) that code for non-fimbrial proteins in the capability of H10407 to survive in the gut of mice. *cexE*, a well-conserved gene among ETEC strains, codes for a putative dispersin (secretory hydrophilic protein), which coats the bacterial cell membrane and enhances the dispersal of bacterial cells by preventing self-aggregation

mediated by hydrophobic fimbrial adhesion. This process is controlled by an Aat type I secretion system, encoded by the *aat* operon, that seems to play a key role in the expression of CexE and similar dispersins through the outer membrane of H10407 and enteroaggregative *E. coli* (EAEC), respectively [61, 62]. The *aat* operon, which is conserved among ETEC isolates, has been found to be associated with *cexE* in H10407 [46].

The *eatA* gene probably plays a role in mediating CFA/I--dependent colonization through the degradation of the mucoid layer that covers CF receptors expressed on intestinal epithelium cells. Interestingly, a recent study has demonstrated that mucin can induce the expression of different fimbrial CFs, including CFA/I, CS1 and CS3 [63]. *eatA* has been shown to be conserved among diverse phylogenetic groups of ETEC [64], and contributes to pathogenesis in the rabbit ileal loop model [65] and elicits a protective response against ETEC infection in mice [64]. Moreover, SepA, a homologue of EatA, has been shown to promote the ability of *Shigella flexneri* to disrupt the integrity of the host intestinal epithelium [66].

The majority of p948 plasmid genes were upregulated under anaerobic growth on LB media, which is the environment likely to mimic at least some aspects of the *in vivo* conditions. A recent human challenge study found that the expression patterns of ETEC virulence genes in faecal samples from volunteers challenged with H10407 were replicating those that have been identified in *in vitro* anaerobic growth on CFA media [67]. However, the variation in expression profiles can probably be attributed to the impact of the growth media used [68].

Human ETEC, including those that express CFA/I, are human-restricted and largely do not cause serious disease in animal models. Nevertheless, mice can be used to explore more generic adaptations to the mammalian host and here we were successful in identifying such genes. It would require human challenge studies to demonstrate a role for such genes in human infections. In this regard, a recent study has demonstrated that human challenge with H10407 stimulates a robust immune response, not only to classical virulence antigens, but also to a number of conserved *E. coli* molecules (e.g. YghJ and flagellin) and pathovar-specific secretory proteins (e.g. EtpA and EatA) [69].

The data presented here have potential application in the control of ETEC transmission during natural infections and also will help in the selection of new targets (e.g. EatA and CexE) to be included in future ETEC vaccine candidates and for the development of new antimicrobial therapeutics.

Funding information
This work was supported by The Wellcome Trust (G.D.) and Marie Bashir Institute and Sydney Medical School foundation (M.A. and G.A.H.). The work was also supported by an ESPRC Vaccine-Hub grant and the NIMR Cambridge BRC Antibiotic Resistance theme (G.D.) and the Swedish Research Council (Å.S.). R.A.K. was funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent project BBS/E/F/000PR10348. The funders had no role in study design, data collection and interpretation or the decision to submit the work for publication.

Acknowledgement
The authors would like to thank KAUST Bioscience Core Laboratory for the support with generating sequencing data. The authors wish to thank the WTSI Sequencing team for their help with generating TraDIS sequencing data and the Pathogen Informatics Team for their assistance with bioinformatics analysis of TraDIS and RNAseq data.

Author contributions
M.A. and G.D. conceived the research. M.A., D.P., A.K.T., R.A.K. and G.D. designed the experiments. M.A. performed the experiments and generated the data (developed transposon libraries, performed screening experiments, prepared TraDIS and RNA sequencing libraries, developed knock-outs and performed validation experiments). E.J. and Å.S. performed confirmatory phenotypic characterization of the reference strain. M.A. and M.M. contributed to the generation of TraDIS sequencing data. S.C., C.B. and M.A. performed animal experiments. L.B. and M.A. performed bioinformatics and statistical analyses. M.A., D.P., J.D.K., G.A.H. and G.D. analysed and interpreted the data. M.A. wrote the first version of the manuscript. All authors reviewed and edited the manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
All experiments involving mice were performed under a UK animal licence that has been through local ethical review before being approved by the Home Office and performed according to the regulations of the UK Home Office Scientific Procedures Act (1986).

References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. *Lancet* 2015;385:117–171.
2. Liu L, Johnson HL, Cousens S, Perin J, Scott S *et al*. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* 2012;379:2151–2161.
3. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH *et al*. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the global enteric multicenter study, GEMs): a prospective, case-control study. *Lancet* 2013;382:209–222.
4. Steffen R, Hill DR, DuPont HL. Traveler's diarrhea: a clinical review. *JAMA* 2015;313:71–80.
5. Joffre E, von Mentzer A, Abd El Ghany M, Oezguen N, Savidge T *et al*. Allele variants of enterotoxigenic *Escherichia coli* heat-labile toxin are globally transmitted and associated with colonization factors. *J Bacteriol* 2015;197:392–403.
6. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2004;2:123–140.
7. Isidean SD, Riddle MS, Savarino SJ, Porter CK. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine* 2011;29:6167–6178.
8. Svennerholm AM, Lundgren A. Recent progress toward an enterotoxigenic *Escherichia coli* vaccine. *Expert Rev Vaccines* 2012;11:495–507.
9. Qadri F, Svennerholm AM, Faruque AS, Sack RB. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin Microbiol Rev* 2005;18:465–483.
10. von Mentzer A, Tobias J, Wiklund G, Nordqvist S, Aslett M *et al*. Identification and characterization of the novel colonization factor CS30 based on whole genome sequencing in enterotoxigenic *Escherichia coli* (ETEC). *Sci Rep* 2017;7:12514.

11. von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A *et al*. Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat Genet* 2014;46:1321–1326.
12. Anantha RP, McVeigh AL, Lee LH, Agnew MK, Cassels FJ *et al*. Evolutionary and functional relationships of colonization factor antigen i and other class 5 adhesive fimbriae of enterotoxigenic *Escherichia coli*. *Infect Immun* 2004;72:7190–7201.
13. Jansson L, Tobias J, Lebens M, Svennerholm AM, Teneberg S. The major subunit, CfaB, of colonization factor antigen i from enterotoxigenic *Escherichia coli* is a glycosphingolipid binding protein. *Infect Immun* 2006;74:3488–3497.
14. Mottram L, Liu J, Chavan S, Tobias J, Svennerholm AM *et al*. Glycoengineered cell line and computational docking studies reveals enterotoxigenic *Escherichia coli* CFA/I fimbriae bind to Lewis a glycans. *Sci Rep* 2018;8:11250.
15. Mottram L, Wiklund G, Larson G, Qadri F, Svennerholm AM. FUT2 non-secretor status is associated with altered susceptibility to symptomatic enterotoxigenic *Escherichia coli* infection in Bangladeshis. *Sci Rep* 2017;7:10649.
16. Ahmed T, Lundgren A, Arifuzzaman M, Qadri F, Teneberg S *et al*. Children with the Le(a+b-) blood group have increased susceptibility to diarrhea caused by enterotoxigenic *Escherichia coli* expressing colonization factor I group fimbriae. *Infect Immun* 2009;77:2059–2064.
17. Li Y-F, Poole S, Rasulova F, McVeigh AL, Savarino SJ *et al*. A receptor-binding site as revealed by the crystal structure of CfaE, the colonization factor antigen I fimbrial adhesin of enterotoxigenic *Escherichia coli*. *J Biol Chem* 2007;282:23970–23980.
18. Poole ST, McVeigh AL, Anantha RP, Lee LH, Akay YM *et al*. Donor strand complementation governs intersubunit interaction of fimbriae of the alternate chaperone pathway. *Mol Microbiol* 2007;63:1372–1384.
19. Rodas C, Iniguez V, Qadri F, Wiklund G, Svennerholm AM *et al*. Development of multiplex PCR assays for detection of enterotoxigenic *Escherichia coli* colonization factors and toxins. *J Clin Microbiol* 2009;47:1218–1220.
20. Sjoling A, Wiklund G, Savarino SJ, Cohen DI, Svennerholm AM. Comparative analyses of phenotypic and genotypic methods for detection of enterotoxigenic *Escherichia coli* toxins and colonization factors. *J Clin Microbiol* 2007;45:3295–3301.
21. Viboud GI, Binsztein N, Svennerholm AM. Characterization of monoclonal antibodies against putative colonization factors of enterotoxigenic *Escherichia coli* and their use in an epidemiological study. *J Clin Microbiol* 1993;31:558–564.
22. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L *et al*. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res* 2009;19:2308–2316.
23. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ *et al*. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* 2016;32:1109–1111.
24. Barquist L, Langridge GC, Turner DJ, Phan MD, Turner AK *et al*. A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res* 2013;41:4549–4564.
25. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–140.
26. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-Seq data. *Genome Biol* 2010;11:R25.
27. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–595.
28. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* 2008;5:621–628.
29. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:323.
30. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;11:R106.
31. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR *et al*. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res* 2016;44:W147–W153.
32. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 2000;97:6640–6645.
33. Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA *et al*. The Essential Genome of *Escherichia coli* K-12. *mBio* 2018;9:e02096–17.
34. Phan MD, Peters KM, Sarkar S, Lukowski SW, Allsopp LP *et al*. The serum resistome of a globally disseminated multidrug resistant uropathogenic *Escherichia coli* clone. *PLoS Genet* 2013;9:e1003834.
35. Lange R, Hengge-Aronis R. Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. *Mol Microbiol* 1991;5:49–59.
36. Hengge-Aronis R, Fischer D. Identification and molecular analysis of *glgS*, a novel growth-phase-regulated and *rpoS*-dependent gene involved in glycogen synthesis in *Escherichia coli*. *Mol Microbiol* 1992;6:1877–1886.
37. Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R. Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol* 2005;187:1591–1603.
38. Lyzen R, Maitra A, Milewska K, Kochanowska-Lyzen M, Hernandez VJ *et al*. The dual role of DksA protein in the regulation of *Escherichia coli pArgX* promoter. *Nucleic Acids Res* 2016;44:10316–10325.
39. Parshin A, Shiver AL, Lee J, Ozerova M, Schneidman-Duhovny D *et al*. DksA regulates RNA polymerase in *Escherichia coli* through a network of interactions in the secondary channel that includes Sequence Insertion 1. *Proc Natl Acad Sci U S A* 2015;112:E6862–E6871.
40. Fukuoka H, Inoue Y, Ishijima A. Coordinated regulation of multiple flagellar motors by the *Escherichia coli* chemotaxis system. *Biophysics* 2012;8:59–66.
41. Berg HC. The flagellar motor adapts, optimizing bacterial behavior. *Protein Sci* 2017;26:1249–1251.
42. Shukla D, Zhu XY, Matsumura P. Flagellar motor-switch binding face of CheY and the biochemical basis of suppression by CheY mutants that compensate for motor-switch defects in *Escherichia coli*. *J Biol Chem* 1998;273:23993–23999.
43. Parkinson JS, Parker SR, Talbert PB, Houts SE. Interactions between chemotaxis genes and flagellar genes in *Escherichia coli*. *J Bacteriol* 1983;155:265–274.
44. Partridge JD, Nhu NTQ, Dufour YS, Harshey RM. *Escherichia coli* remodels the chemotaxis pathway for swarming. *mBio* 2019;10:e00316–00319.
45. Partridge JD, Nhu NTQ, Dufour YS, Harshey RM. Tumble suppression is a conserved feature of swarming motility. *mBio* 2020;11:e01189–01120.
46. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M *et al*. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *J Bacteriol* 2010;192:5822–5831.
47. Madhavan TPV, Riches JD, Scanlon MJ, Ulett GC, Sakellaris H. Binding of CFA/I pili of enterotoxigenic *Escherichia coli* to asialo-GM1 is mediated by the minor pilin CfaE. *Infect Immun* 2016;84:1642–1649.
48. Luiz WB, Rodrigues JF, Crabb JH, Savarino SJ, Ferreira LC. Maternal vaccination with a fimbrial tip adhesin and passive protection of neonatal mice against lethal human enterotoxigenic *Escherichia coli* challenge. *Infect Immun* 2015;83:4555–4564.
49. Maigaard Hermansen GM, Boysen A, Krogh TJ, Nawrocki A, Jelsbak L *et al*. HldE Is important for virulence phenotypes in Enterotoxigenic *Escherichia coli*. *Front Cell Infect Microbiol* 2018;8:253.

50. Kingsley RA, Langridge G, Smith SE, Makendi C, Fookes M *et al*. Functional analysis of *Salmonella* Typhi adaptation to survival in water. *Environ Microbiol* 2018;20:4079–4090.
51. Cho BK, Kim D, Knight EM, Zengler K, Palsson BO. Genome-scale reconstruction of the sigma factor network in *Escherichia coli*: topology and functional states. *BMC Biol* 2014;12:4.
52. Gruber TM, Gross CA. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* 2003;57:441–466.
53. Maeda H, Fujita N, Ishihama A. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res* 2000;28:3497–3503.
54. Turner AK, Lovell MA, Hulme SD, Zhang-Barber L, Barrow PA. Identification of *Salmonella* typhimurium genes required for colonization of the chicken alimentary tract and for virulence in newly hatched chicks. *Infect Immun* 1998;66:2099–2106.
55. Lothigius A, Sjoling A, Svennerholm AM, Bolin I. Survival and gene expression of enterotoxigenic *Escherichia coli* during long-term incubation in sea water and freshwater. *J Appl Microbiol* 2010;108:1441–1449.
56. Hernroth B, Lothigius A, Bolin I. Factors influencing survival of enterotoxigenic *Escherichia coli*, *Salmonella enterica* (serovar Typhimurium) and *Vibrio parahaemolyticus* in marine environments. *FEMS Microbiol Ecol* 2010;71:272–280.
57. Sharma UK, Chatterji D. Transcriptional switching in *Escherichia coli* during stress and starvation by modulation of sigma activity. *FEMS Microbiol Rev* 2010;34:646–657.
58. Savarino SJ, McKenzie R, Tribble DR, Porter CK, O'Dowd A *et al*. Prophylactic efficacy of hyperimmune bovine colostral antiadhesin antibodies against enterotoxigenic *Escherichia coli* diarrhea: a randomized, double-blind, placebo-controlled, phase 1 trial. *J Infect Dis* 2017;216:7–13.
59. Sincock SA, Hall ER, Woods CM, O'Dowd A, Poole ST *et al*. Immunogenicity of a prototype enterotoxigenic *Escherichia coli* adhesin vaccine in mice and nonhuman primates. *Vaccine* 2016;34:284–291.
60. Rollenhagen JE, Jones F, Hall E, Maves R, Nunez G *et al*. Establishment, validation and application of a New World Primate model of ETEC *Escherichia coli* Disease for vaccine development. *Infect Immun* 2019;87:e00634–18.
61. Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V *et al*. The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. *J Biol Chem* 2003;278:45680–45689.
62. Sheikh J, Czeczulin JR, Harrington S, Hicks S, Henderson IR *et al*. A novel dispersin protein in enteroaggregative *Escherichia coli*. *J Clin Invest* 2002;110:1329–1337.
63. Haines S, Gautheron S, Nasser W, Renauld-Mongenie G. Identification of novel components influencing colonization factor antigen I expression in enterotoxigenic *Escherichia coli*. *PLoS One* 2015;10:e0141469.
64. Luo Q, Qadri F, Kansal R, Rasko DA, Sheikh A *et al*. Conservation and immunogenicity of novel antigens in diverse isolates of enterotoxigenic *Escherichia coli*. *PLoS Negl Trop Dis* 2015;9:e0003446.
65. Patel SK, Dotson J, Allen KP, Fleckenstein JM. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic *Escherichia coli*. *Infect Immun* 2004;72:1786–1794.
66. Maldonado-Contreras A, Birtley JR, Boll E, Zhao Y, Mumy KL *et al*. *Shigella* depends on SepA to destabilize the intestinal epithelial integrity via cofilin activation. *Gut Microbes* 2017;8:544–560.
67. Crofts AA, Giovanetti SM, Rubin EJ, Poly FM, Gutierrez RL *et al*. Enterotoxigenic *E. coli* virulence gene regulation in human infections. *Proc Natl Acad Sci U S A* 2018;115:E8968–E8976.
68. Hazen TH, Michalski J, Luo Q, Shetty AC, Daugherty SC *et al*. Comparative genomics and transcriptomics of *Escherichia coli* isolates carrying virulence factors of both enteropathogenic and enterotoxigenic *E. coli*. *Sci Rep* 2017;7:3513.
69. Chakraborty S, Randall A, Vickers TJ, Molina D, Harro CD *et al*. Human experimental challenge with enterotoxigenic *Escherichia coli* elicits immune responses to canonical and novel antigens relevant to vaccine development. *J Infect Dis* 2018;218:1436–1446.