The ISME Journal (2012) 6, 1916–1925

www.nature.com/ismej

ORIGINAL ARTICLE

Functional metagenomics reveals novel salt tolerance loci from the human gut microbiome

Eamonn P Culligan[1,2], Roy D Sleator[1,3], Julian R Marchesi[1,4] and Colin Hill[1,2]
[1]*Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland;* [2]*Department of Microbiology, University College Cork, Cork, Ireland;* [3]*Department of Biological Sciences, Cork Institute of Technology, Bishopstown, Cork, Ireland and* [4]*Cardiff School of Biosciences, Cardiff University, Cardiff, UK*

Metagenomics is a powerful tool that allows for the culture-independent analysis of complex microbial communities. One of the most complex and dense microbial ecosystems known is that of the human distal colon, with cell densities reaching up to 10^{12} per gram of faeces. With the majority of species as yet uncultured, there are an enormous number of novel genes awaiting discovery. In the current study, we conducted a functional screen of a metagenomic library of the human gut microbiota for potential salt-tolerant clones. Using transposon mutagenesis, three genes were identified from a single clone exhibiting high levels of identity to a species from the genus *Collinsella* (closest relative being *Collinsella aerofaciens*) (COLAER_01955, COLAER_01957 and COLAER_01981), a high G + C, Gram-positive member of the Actinobacteria commonly found in the human gut. The encoded proteins exhibit a strong similarity to GalE, MurB and MazG. Furthermore, pyrosequencing and bioinformatic analysis of two additional fosmid clones revealed the presence of an additional *galE* and *mazG* gene, with the highest level of genetic identity to *Akkermansia muciniphila* and *Eggerthella* sp. YY7918, respectively. Cloning and heterologous expression of the genes in the osmosensitive strain, *Escherichia coli* MKH13, resulted in increased salt tolerance of the transformed cells. It is hoped that the identification of atypical salt tolerance genes will help to further elucidate novel salt tolerance mechanisms, and will assist our increased understanding how resident bacteria cope with the osmolarity of the gastrointestinal tract.
The ISME Journal (2012) **6,** 1916–1925; doi:10.1038/ismej.2012.38; published online 26 April 2012
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: *Akkermansia*; *Collinsella*; *Eggerthella*; human gut microbiome; metagenomics; salt tolerance

Introduction

The ability to respond and adapt to changes in external osmolarity is a key determinant for bacterial survival and proliferation in various environmental niches (Sleator and Hill, 2002). Microorganisms are continually exposed to fluctuations and perturbations in osmolarity in their environment caused by rainwater, drought, salinity and changing solute concentrations. Both transient and symbiotic microorganisms that colonize the gastrointestinal tract are particularly susceptible to water loss due to osmotic stress (Gralla and Huo, 2008). It has been demonstrated that free water is not evenly distributed along the gut, but exists as pockets, meaning changes in osmolarity can be rapid (Schiller *et al.*, 2005). The elevated osmolarity of the upper small intestine (the equivalent of 0.3 M NaCl (Chowdhury *et al.*, 1996)), represents an initial challenge to ingested microorganisms and the osmolarity is likely to be higher in the distal colon following further water absorption in the final stages of the digestive process. *Bacteroides fragilis* isolates from human stool samples have shown increased resistance to both NaCl and bile stress compared with isolates from blood or abscesses, indicating the potential importance of such stress tolerance mechanisms in the gastrointestinal tract (Pumbwe *et al.*, 2007).

In general, bacteria respond to hyper-osmotic stress in a phased manner. Firstly, during the primary response, which is activated within seconds of osmotic upshift, potassium (K^+) is rapidly transported into the cell (for a review see Epstein, 2003). However, a more sophisticated osmotic stress response system is required once the cellular osmolality has been initially stabilized. The next phase, or secondary response, involves the uptake and/or synthesis of compatible solutes (also termed osmolytes or osmoprotectants). As their name suggests, these diverse compounds (Kempf and Bremer, 1998) are for the most part compatible with

Correspondence: C Hill or JR Marchesi, Department of Microbiology or Alimentary Pharmabiotic Centre, University College Cork, Western Road, Cork, Ireland.
E-mail: c.hill@ucc.ie (CH) or marchesijr@cardiff.ac.uk (JRM)
Received 9 August 2011; revised 1 March 2012; accepted 16 March 2012; published online 26 April 2012

vital cellular functions and can restore cell volume and turgor without adversely affecting metabolism (Kunte, 2006). Although the primary and secondary responses constitute the classical response to hyperosmotic stress, the overall osmoadaptation strategy is much more complex, and involves a diverse range of cellular systems and processes seemingly unrelated to the primary and secondary responses. Identifying such diverse genes/proteins will provide us with a broader and more complete view of salt tolerance and the cellular response to salt-induced osmotic stress in bacteria. Indeed, many such ancillary mechanisms have been identified and range from two-component systems and proteolytic systems to numerous general stress and heat stress proteins and can also involve changes to the cell membrane (Kilstrup *et al*., 1997; Sakamoto and Murata, 2002; Gardan *et al*., 2003; Piuri *et al*., 2003; Wonderling *et al*., 2004; Sleator and Hill, 2005; Lopez *et al*., 2006).

Since their inception, metagenomic strategies have led to the identification of numerous novel and diverse genes, enzymes and proteins from many diverse environments through sequence-based and/or functional approaches (Beja *et al*., 2000; Gillespie *et al*., 2002; Lee *et al*., 2007; Banik and Brady, 2008; Heath *et al*., 2009; Meilleur *et al*., 2009). In principle, metagenomics can provide access to all of the genetic resources in a given environmental niche and as such is an extremely powerful tool to access the genomes of difficult-to-culture or unculturable microorganisms (Sleator *et al*., 2008). A recent study identified novel genes from a pond water metagenome that increased resistance to salt-induced osmotic stress when expressed in *E. coli* (Kapardar *et al*., 2010a, b). Although some studies have used functional metagenomics to screen for various phenotypes from the human gut environment (Kazimierczak *et al*., 2008; Lakhdari *et al*., 2010; Gloux *et al*., 2011), to our knowledge this study is the first to identify genes that confer salt tolerance from the human gut microbiota.

Materials and methods

Bacterial strains and growth conditions

Bacterial strains and plasmids used in this study are listed in Table 1. Primers (Eurofins, MWG Operon, Ebersberg, Germany) used in this study are listed in Table 2. *E. coli* EPI300::pCC1FOS (Epicentre Biotechnologies, Madison, WI, USA) was grown in Luria–Bertani (LB) medium (Maniatis *et al*., 1982) containing 12.5 μg ml^{-1} chloramphenicol (Cm) and in 12.5 μg ml^{-1} chloramphenicol plus 50 μg ml^{-1}

Table 2 Primers used in this study

Primer	*Sequence (5′–3′)*[a]
pCI372 FP	CGGGAAGCTAGAGTAAGTAG
pCI372 RP	CCTCTCGGTTATGAGTTAG
mazG(3) FP	AAAACTGCAGGCCCGTCGTTCCCGCAGTCTTAC
mazG(3) RP	GCTCTAGAATCTACGAGGGCGGCGCGTTC
murB(3) FP	AAAACTGCAGCCACCTCCTGGGCGATCTGCTTGAG
murB(3) RP	GCTCTAGACGACACACCGGACTGGGTTATCTGA
galE(3) FP	AAAACTGCAGATGGGTGTGCAGTCCGCCTC
galE(3) RP	GCTCTAGAGTCCCAACGATTTCCACGAACG
mazG(5) FP	AAAACTGCAGCTAAACAGGAGGCGAAGCTC
mazG(5) RP	GCTCTAGAGCAGAAGGCGTCAACGATA
galE(25) FP	GCTCTAGACCGGCTTAACAGCATTGATA
galE(25) RP	AAAACTGCAGGCTGCGTTGTCTTTCCAGTT
EZTn-FP-1	GCCAACGACTACGCACTAGCCAAC
EZTn-RP-1	GAGCCAATATGCGAGAACACCCGAGAA
T7	TAATACGACTCACTATAGGG
M13 R	CAGGAAACAGCTATGACC

Abbreviations: FP, forward primer; RP, reverse primer.
[a]Restriction enzyme recognition sequences are underlined.

Table 1 Bacterial strains, plasmids and transposon used in this study

Strain, plasmid or transposon	*Genotype or characteristic(s)*	*Source or reference*
Strains		
E. coli EPI300	F$^-$ *mcrA* Δ(*mrr-hsd*RMS-*mcrBC*) Φ80d*lacZ*ΔM15 Δ*lac*X74 *recA*1 *endA*1 *araD*139 Δ(*ara, leu*)7697 *galU galK* λ$^-$ *rpsL nupG trfA dhfr*; high-transformation efficiency of large DNA	Epicentre Biotechnologies
E. coli MKH13	MC4100*Δ*(*putPA*)101D(*proP*)2D(*proU*)	(Haardt *et al*., 1995)
E. coli MKH13::pCI372-*murB*(3)	MKH13 containing pCI372 with *murB* gene from SMG 3 (similar to *Collinsella aerofaciens* ATCC 25986)	This study
E. coli MKH13::pCI372-*mazG*(3)	MKH13 containing pCI372 with *mazG* gene from SMG 3 (similar to *Collinsella aerofaciens* ATCC 25986)	This study
E. coli MKH13::pCI372-*galE*(3)	MKH13 containing pCI372 with *galE* gene from SMG 3 (similar to *Collinsella aerofaciens* ATCC 25986)	This study
E. coli MKH13::pCI372-*mazG*(5)	MKH13 containing pCI372 with *mazG* gene from SMG 5 (similar to *Eggerthella* sp. YY7918	This study
E. coli MKH13::pCI372-*galE*(25)	MKH13 containing pCI372 with *galE* gene from SMG 25 (similar to *Akkermansia muciniphila* ATCC BAA-835)	This study
Plasmids		
pCI372	Shuttle vector between *E. coli* and *L. lactis*, Cm^R	(Hayes *et al*., 1990)
pCC1FOS	Fosmid cloning vector, Cm^R	Epicentre Biotechnologies
Transposon		
EZ-Tn*5* <*ori*V/ KAN-2>	Hyperactive Tn*5* transposon, Kan^R, inducible high copy origin of replication—*ori*V	Epicentre Biotechnologies

Kan (kanamycin) following EZ-Tn*5* transposon mutagenesis reactions. *E. coli* MKH13 was grown in LB medium and in LB medium supplemented with 20 μg ml^{-1} Cm for strains transformed with the plasmid pCI372. LB media was supplemented with 1.5% agar when required. All overnight cultures were grown at 37 °C with shaking.

Metagenomic library construction
A metagenomic library that had been constructed previously (Jones and Marchesi, 2007) was used to screen for salt-tolerant clones. Briefly, metagenomic DNA was isolated from a human faecal sample and from this a fosmid library (average insert size of approximately 40 kb) was created using a Copy Control fosmid library production kit (Epicentre Biotechnologies, Madison, WI, USA) according to the manufacturer's instructions. A Genetix (Berkshire, UK) QPix 2 XT colony-picking robot was used to transfer fosmid clones to 384-well micro-titre plates, which were stored at − 80 °C until needed.

Screening the metagenomic library
A total of 23 040 clones from the library (average insert size of ∼40 kb) were screened on LB agar supplemented with 6.5% NaCl for clones showing increased tolerance to NaCl compared with the *E. coli* EPI300 host strain (containing an empty pCC1FOS cloning vector). A Genetix QPix 2 XT colony picking/gridding machine was used to plate the clones onto Q-Trays (Genetix) containing LB agar supplemented with 12.5 μg ml^{-1} Cm and 6.5% NaCl. Q-Trays were incubated at 37 °C for 72 h. Following incubation, likely salt-tolerant clones were replica plated onto LB containing 12.5 μg ml^{-1} Cm and 6.5% NaCl and onto LB containing 12.5 μg ml^{-1} but without NaCl (which served as a positive control).

Growth experiments
Cultures were grown overnight with shaking in LB broth. Subsequently, cells were harvested, washed in one quarter strength sterile Ringer's solution and resuspended in fresh LB broth. A 2% inoculum was sub-cultured in fresh LB broth containing the appropriate stress (that is, 0–10% (w/v) NaCl, 0–8% (w/v) KCl, 0–90% (w/v) sucrose or 0–80% (v/v) glycerol as required) and 200 μl was transferred to a sterile 96-well micro-titre plate (Starstedt Inc. Newton, MA, USA). Plates were incubated at 37 °C for 48 h in an automated spectrophotometer (Tecan Genios, Mannedorf, Switzerland), which recorded the optical density 595 nm every hour. After 48 h, the data was retrieved and analysed using the Magellan 3 software program. Results are presented as the average of triplicate experiments, with error bars being representative of the s.e.m.

DNA manipulations
Extraction of fosmids containing metagenomic DNA: 5 ml of bacterial culture was grown overnight with 12.5 μg ml^{-1} Cm. To inoculate 4 ml of fresh LB broth, 1 ml of culture was used. To this, 5 μl of 1000 × Copy Control Induction solution (Epicentre Biotechnologies) and 12.5 μg ml^{-1} Cm were added. The mixture was incubated at 37 °C for 5 h with vigorous shaking (200–250 r.p.m.) to ensure maximum aeration. Cells were harvested from the whole 5 ml of induced culture by centrifuging at 2100 × *g* for 12 mins. Qiagen (West Sussex, UK) QIAprep Spin mini-prep kit was used to extract fosmids as per manufacturer's instructions. PCR products were purified with a Qiagen PCR purification kit and digested with *XbaI* and *PstI* (Roche Applied Science, West Sussex, UK), followed by ligation (T4 DNA ligase (Roche Applied Science) for *mazG*(3) and *murB*(3); FastLink DNA ligase for *galE*(3), *galE*(25) *and mazG*(5) (Epicentre Biotechnologies) as per manufacturer's instructions) to similarly digested plasmid pCI372. Electrocompetent *E. coli* MKH13 were transformed with the ligation mixture and plated on LB agar plates containing 20 μg ml^{-1} Cm for selection. Colony PCR was performed on all resistant transformants using primers across the multiple cloning site (MCS) of pCI372 to confirm the presence and size of the insert.

Transposon mutagenesis
Transposon mutagenesis was carried out on SMG 3 according to the manufacturer's instructions using the EZTn-5 <*oriV*/ KAN-2> *in vitro* transposition kit (Epicentre Biotechnologies). *E. coli* EPI300 cells were transformed with the transposon reaction mixture and selected on plates containing Cm and Kan (12.5 and 50 μg ml^{-1}, respectively). Subsequently, the transposon insertion clones were replica plated onto LB with and without 6.5% NaCl. Clones which grew on LB but not on LB + 6.5% NaCl indicates a likely insertion event in a gene involved in salt tolerance. Presumptive salt-tolerant knockouts were grown overnight and a fosmid DNA extraction was performed. The extracted fosmid containing metagenomic DNA was subjected to sequencing from the ends of the transposon using the primers EZTn-FP-1 and EZTn-RP-1 (Table 2). All sequencing was performed by GATC Biotech (Konstanz, Germany).

Sequencing and bioinformatic analysis
Fosmids SMG 1, 3, 5, 6 and 25 were fully sequenced and assembled by GATC Biotech using the GS FLX (Roche Applied Science) platform. Putative open reading frames were predicted using Softberry FGENESB bacterial operon and gene prediction software (Mavromatis *et al.*, 2007). Retrieved nucleotide and translated amino-acid sequences

were functionally annotated by homology searches using the Basic Local Alignment and Search Tool (BLAST), to identify homologous sequences from the National Centre for Biotechnology Information website: http://www.ncbi.nlm.nih.gov/blast/Blast.cgi.

End sequencing was performed on all fosmid clones (SMG1-53) using T7 or M13 R primers and the taxonomic distribution of end sequences were assigned using the BLAST program (Supplementary Table S1).

Nucleotide sequences identified in this study were compared against the healthy individuals in the MetaHit data set (Qin *et al*., 2010) using the BLAST program to identify homologous sequences (>50% identity; *e*-value $< 1 \times 10^{-5}$). This data was used to determine the relative abundance of the identified genes in the data set and estimate the amount of DNA that would need to be screened (Mb) to return a hit to one of the genes (Supplementary Figures S2A–C).

In addition, BLASTP analysis was undertaken to identify the protein sequences in GenBank, which showed the highest identity to genes isolated here and which were shown to be responsible for the salt tolerance phenotype. The closest hits were aligned using ClustalW (Thompson *et al*., 1994) in Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). The aligned proteins were analysed using MEGA 5 (Tamura *et al*., 2011) and the evolutionary history was inferred using the Neighbour-Joining method (Saitou and Nei, 1987) from the evolutionary distances, which were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino-acid substitutions per site. The bootstrap method (500 replicates) was used to test the percentage of replicate trees in which the associated taxa clustered together (Felsenstein, 1985) (Supplementary Figure S3).

Results

Screening the metagenomic library

In this study, a metagenomic fosmid library constructed from DNA isolated from human gut microbiota was screened for clones, which conferred increased salt tolerance using a combined functional metagenomic, transposon mutagenesis and bioinformatic analysis approach. The cloning host is incapable of growth on LB supplemented with 6.5% NaCl, providing a positive selection for clones able to cope with elevated osmolarity. In total, 53 salt-tolerant clones were identified, which were annotated SMG (salt metagenome) 1–53. Six clones (SMG 1–6) grew within 24 h; while a further 47 CFUs appeared within 36 h. Physiological growth experiments were conducted on host strain EPI300::pCC1FOS and SMG 1–6. All six clones showed increased salt tolerance relative to the control at 7% NaCl (*E. coli* EPI300::pCC1FOS host strain) (Figure 1). The clones were subjected to further study through transposon mutagenesis.

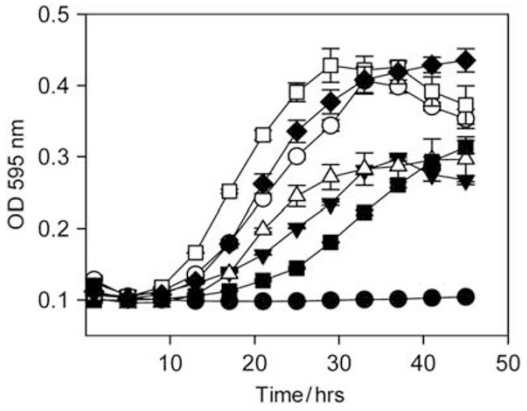

Figure 1 Growth in LB broth supplemented with 7% NaCl. Growth of *E. coli* EPI300::pCC1FOS host strain (●), SMG 1 (○), SMG 2 (▼), SMG 3 (△), SMG 4 (■), SMG 5 (□) and SMG 6 (♦).

Transposon mutagenesis

Transposon mutagenesis was carried out using the EZ-Tn*5* <oriV/ KAN-2> insertion kit (Epicentre Biotechnologies). Transposon mutagenesis proved unsuccessful for SMG 1, SMG 5 and SMG 6, possibly owing to difficulty in obtaining sufficiently high yields of fosmid DNA for *in vitro* transposon mutagenesis, so SMG 3 was chosen for further investigation. Transposon mutagenesis of SMG 3-yielded nine clones incapable of growth on 6.5% NaCl. Sequencing revealed transposon insertions in three distinct genes with high genetic identity (95–98%) to genes encoding hypothetical proteins from a member of the genus *Collinsella* (closest relative being *C. aerofaciens*). The genes; *galE*(3) (COLAER_01955), *murB*(3) (COLAER_01957) and *mazG*(3) (COLAER_01981), encode hypothetical proteins similar to UDP-glucose 4-epimerase (GalE), UDP-*N*-acetylenolpyruvoylglucosamine reductase (MurB) and nucleoside triphosphate pyrophosphohydrolase (MazG family protein). Although GalE has been tentatively linked to salt tolerance in previous studies (Hengge-Aronis *et al*., 1991; Bohringer *et al*., 1995; Nguyen *et al*., 2004) we are unaware of any previous link between MurB or MazG and bacterial salt tolerance.

Sequencing and bioinformatic analysis

End sequencing was carried out on all 53 SMG clones, using T7 or M13 R primers (GATC Biotech). In all, 49 clones were successfully end sequenced and taxonomically assigned based on BLAST hits. Of these, the *Bacteroidetes* dominated, representing 57.14% of the sequences, followed by the *Actinobacteria* and *Proteobacteria* at 14.29% and 12.24%, respectively. The *Verrucomicrobia* made up 8.16% of the sequences, whereas the *Firmicutes* were under-represented at 4.08%. The remainder of the sequences were unidentified and one sequence came from a member of the *Heterokontophyta*

(*Blastocystis hominis*), a eukaryote (2.04% each). (Supplementary Table S1, Supplementary Material).

Full fosmid sequencing was performed on clones SMG 1, 3, 5, 6 and 25, revealing insert sizes of 36, 39, 42, 36 and 44 kb, respectively. The sequences have been submitted to GenBank; with following accession numbers: SMG 1 = JQ269596; SMG 3 = JQ269597; SMG 5 = JQ269598; SMG 6 = JQ269599; SMG 25 = JQ269600; SMG 1 and SMG 6 were found to have the same metagenomic insert, with the highest genetic identity to *Bacteroides thetaiotaomicron*. SMG 5 shared highest genetic identity to *Eggerthella* sp. YY7918 (a member of the high G + C Gram-positive *Actinobacteria*). However, it should be noted that BlastP analysis of the predicted proteins encoded by the SMG 5 genes revealed many variations in the associated species. While all but one corresponded to members of the phylum *Actinobacteria*, the genera were represented by various species of *Eggerthella, Slackia, Cryptobacterium* and *Gordonibacter*. This indicates that the SMG 5 insert represents DNA from a novel species from one of these genera. SMG 25 shared highest genetic identity to *Akkermansia muciniphila* (a member of the Gram-negative *Verrucomicrobia*). Putative open reading frames were identified using FGENESB bacterial operon and gene prediction software from Softberry (Mavromatis *et al*., 2007). BLAST analysis revealed the presence of a hypothetical protein (EGYY_03530) and a UDP glucose 4-epimerase encoded by putative *mazG* and *galE* genes, respectively. These genes were present on SMG 5 and SMG 25 and were designated *mazG*(5) and *galE*(25), respectively. Having previously cloned and expressed homologous genes (*galE*(3) and *mazG*(3)) from SMG 3 in *E. coli* MKH13, resulting in an increased salt tolerance phenotype, it was decided to clone these bioinformatically identified *mazG*(5) and *galE*(25) genes into *E. coli* MKH13 also.

The nucleotide sequences of the genes identified were compared with the MetaHit data set from the healthy individuals (Qin *et al*., 2010) to determine the relative abundance of the genes among the gut metagenomes of the subjects. Homologues of all of the genes identified in this study were found to be present in all MetaHit samples (>50% identity; *e*-value $<1\times10^{-5}$). The *galE* genes (*galE*(3) and *galE*(25)) were found to be much more abundant than both *mazG*(3) and *murB*(3), with a hit rate of approximately 1 per Mb of DNA screened, compared with approximately one per 4–6 Mb DNA for *mazG* and *murB* (Supplementary Figures S2A–C).

Growth experiments

Three genes were identified by transposon mutagenesis from SMG 3, namely *galE*(3), *murB*(3) and *mazG*(3) and two further genes were identified through bioinformatic analysis; *mazG*(5) from SMG 5 and *galE*(25) from SMG 25. The genes were cloned with some flanking DNA into the shuttle plasmid pCI372 and transformed into electrocompetent *E. coli* MKH13. Growth experiments in LB and LB supplemented with NaCl or KCl (ionic osmotic stress) were conducted on MKH13 :: pCI372-*mazG*(3),

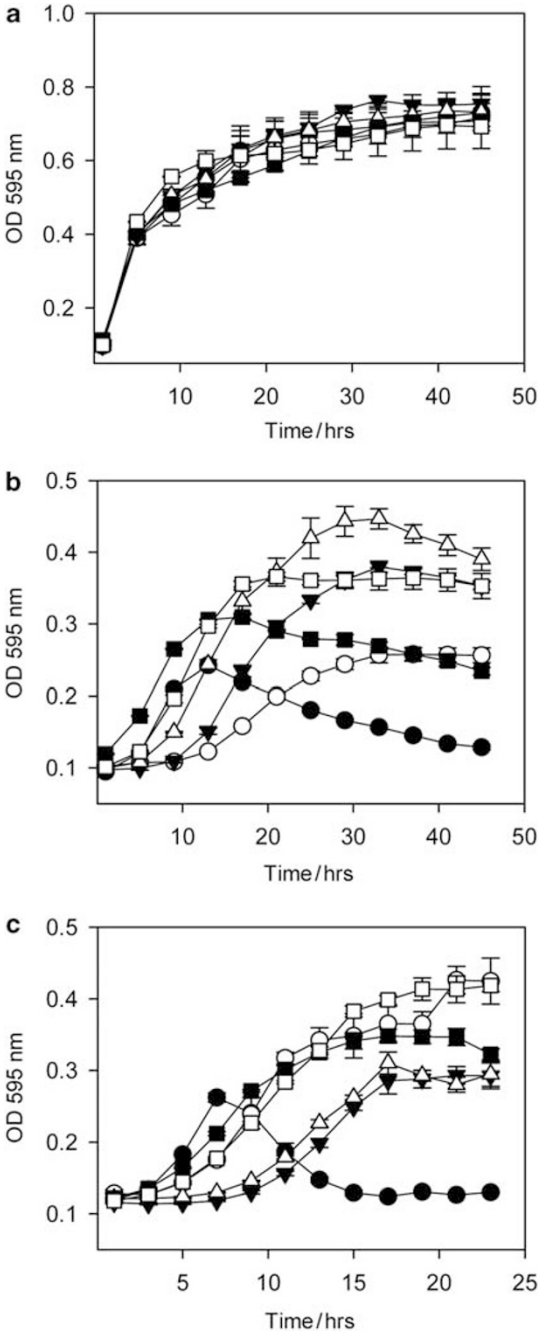

Figure 2 Growth in LB and LB supplemented with 3% NaCl or 4% KCl. The first *P* value in parentheses represents growth in NaCl, while the second P value represents growth in KCl. Growth of WT MKH13 (●), MKH13::pCI372-*galE*(3) ($P<0.006$; $P=0.0004$) (○), MKH13::pCI372-*murB*(3) ($P<0.0001$; $P<0.0001$) (△), MKH13::pCI372-*mazG*(3) ($P<0.0001$; $P<0.0017$) (▼), MKH13::pCI372-*galE*(25) ($P=0.0002$; $P=0.0012$) (■) and MKH13::pCI372-*mazG*(5) ($P=0.0003$; $P<0.0001$) (□) in (**a**) LB broth, (**b**) LB broth supplemented with 3% NaCl and (**c**) LB broth supplemented with 4% KCl.

MKH13::pCI372-*murB*(3), MKH13::pCI372-*galE*(3), MKH13::pCI372-*mazG*(5) and MKH13::pCI372-*galE* (25). Each of the five transformed clones showed a statistically significant increase in salt tolerance (to both NaCl and KCl) compared with wild-type MKH13 (Figures 2b and c), whereas no difference in growth was observed in LB alone (Figure 2a). Each of the five transformed clones was also tested in their ability to grow under conditions of non-ionic osmotic stress (that is, sucrose or glycerol). No increased growth phenotype was observed under these conditions for any of the clones (Supplementary Figure S1), indicating these genes may confer a salt-specific protective effect.

Discussion

Metagenomics has the potential to allow us to advance, or for the most part to begin, the study of the genetic complement of uncultured microbes. In the current study, we used a combined functional metagenomic, transposon mutagenesis and bioinformatic strategy to screen a metagenomic library from the human gut microbiota for potential salt-tolerant clones and identified five genes (Table 3), namely *galE*(3) and *galE*(25), *mazG*(3) and *mazG*(5) and *murB*(3), involved in salt tolerance and likely to be important for life in the gut.

In the most comprehensive analysis to date of the genetic complement of the human gut microbiome, a cohort of over 1200 genes (termed range clusters) were identified, which encode functions important for life in the gut (Qin *et al.*, 2010). These included genes, which encoded proteins similar to those that we have identified in this study, namely, UDP glucose 4-epimerase, NTP pyrophosphohydrolase (for which MazG is a functional homologue), as well as a protein containing a tetrapyrrole methyltransferase domain and a MazG-like domain and UDP-*N*-acetylmuramate dehydrogenase (which is another name for MurB; UDP-*N*-acetylenolpyruvylglucosamine reductase) (Supplementary Table S10 from Qin *et al.*, 2010). The presence of the genes encoding these proteins in the enriched gene set leads us to conclude that they are important for survival in the gut and their putative role in adapting to fluctuating levels of osmotic stress in the gut. Also, homologues of the identified sequences from this study were found to be abundant in each of the individual metagenomes from the MetaHit data set (Supplementary Figures S2A–C). Furthermore, an analysis of human gut genomic and metagenomic data sets identified uncharacterized and novel protein families, which are over-represented in the human gut (Ellrott *et al.*, 2010). Among these, and second on the list of over-represented protein families, are coiled-coil osmosensory transporters. The best characterized of the coiled-coil transporters is ProP of *E. coli*, which is responsible for the uptake of osmoprotectants, such as proline and betaine during osmotic stress (Culham *et al.*, 1993). Such representative proteins were not found in any of the genomes analysed from microbes not associated with the human gut microbiome, indicating their importance to bacteria residing in the human gut. This observation also reinforces the idea that the human gut microbiota may employ novel strategies and possess novel mechanisms for osmoadaptation.

Transposon mutagenesis resulted in the identification of a gene (*galE*) from SMG 3 (*galE*(3)), encoding a UDP glucose 4-epimerase (GalE) (also termed UDP galactose 4-epimerase), whereas another *galE* gene was bioinformatically identified from SMG 25 (*galE*(25)). This enzyme catalyses the direct inter-conversion of UDP glucose and UDP galactose (Leloir, 1951), which are precursors involved in the synthesis of capsular polysaccharide and the compatible solute trehalose as well as the synthesis of lipopolysaccharide and membrane-derived oligosaccharides in Gram-negative bacteria

Table 3 Genes identified in this study

SMG clone	*Putative gene*	*Method of identification*	*Top BLAST hit*	*% Identity (amino acid)*	e-*value*
SMG 3–3	*mazG*(3) COLAER_01981 (861 bp)	EZ-Tn5 transposon mutagenesis	Hypothetical protein; *Collinsella aerofaciens,* Similar to *mazG* encoding a MazG family protein (nucleoside triphosphate pyrophosphohydrolase) (286 aa)	95	2*e*- 94
SMG 3–11	*murB*(3) COLAER_01957 (963 bp)	EZ-Tn5 transposon mutagenesis	Hypothetical protein; *Collinsella aerofaciens,* Similar to *murB* encoding UDP-*N*-acetylenolpyruvoylglucosamine reductase (320 aa)	98	5*e*- 83
SMG 3-17	*galE*(3) COLAER_01955 (1062 bp)	EZ-Tn5 transposon mutagenesis	Hypothetical protein; *Collinsella aerofaciens,* Similar to *galE* encoding UDP glucose 4-epimerase (353 aa)	98	3*e*- 175
SMG 5	*mazG*(5) EGYY_03530 (876 bp)	Bioinformatic analysis	Hypothetical protein; *Eggerthella* sp. YY7918, Similar to *mazG* encoding a MazG family protein (nucleoside triphosphate pyrophosphohydrolase) (291 aa)	61	1*e*-117
SMG 25	*galE*(25) Amuc_1125 (990 bp)	Bioinformatic analysis	Hypothetical protein; *Akkermansia muciniphila,* Similar to *galE* encoding a UDP glucose 4-epimerase (329 aa)	96	0.0

and lipoteichoic acids in Gram-positive bacteria (Fukasawa *et al.*, 1962; Markovitz, 1977; Schulman and Kennedy, 1977; Giaever *et al.*, 1988; Seltman and Holst, 2002; Grundling and Schneewind, 2007).

The *galE* gene is often found in an operon with the *galT* and *galK* genes for galactose metabolism. In SMG 3 (highest genetic identity to *C. aerofaciens*), the *galE* gene (COLAER_01955) is found not on a *gal* operon, but between two genes encoding hypothetical proteins similar to a peptidase and a 4-alpha-glucanotransferase (COLAER_01956 and COLAER_01953, respectively). COLAER_01957 (*murB*) is located 226 nucleotides downstream of COLAER_01956. *C. aerofaciens*' *galE* may still be involved in galactose metabolism, as the organism can utilize galactose (Kageyama *et al.*, 1999). The *galE* gene from SMG 25 is not found on the galactose operon and also exhibits a different genomic arrangement to SMG 3. This gene, *galE*(25), which shares 96% identity to *A. muciniphila* (Amuc_1125), is located between a UDP-galactopyranose mutase homologue from *Chthoniobacter flavus* (52% identity) and a gene encoding a hypothetical protein similar to Amuc_1123 from *A. muciniphila*. Furthermore, the fosmid insert of SMG 25 has a vastly different genomic arrangement to that of *A. muciniphila*. This lack of synteny indicates that the fosmid insert is not from *A. muciniphila*, but from a related but as yet undiscovered species. Indeed, sequences representing at least eight uncharacterized species of the genus *Akkermansia* have been identified from different gut metagenomic libraries (van Passel *et al.*, 2011). Interestingly, a UDP glucose 4-epimerase homologue has been identified as a putative gene involved in drought tolerance in rice by modulating the ability of rice roots to penetrate deeper into the substratum when exposed to drought conditions (Nguyen *et al.*, 2004).

A transposon insertion in the *galE* could potentially affect the cell in different ways; (1) by causing compositional changes in the lipopolysaccharide layer or in lipoteichoic acids of Gram-negative and Gram-positive bacteria, respectively, thus reducing the cell's ability to sense, respond to or resist various environmental stress conditions and (2) by disrupting the inter-conversion of UDP-glucose and UDP-galactose. UDP glucose may be a key molecule, as it can be converted to the osmoprotectant trehalose (Giaever *et al.*, 1988), in addition to potentially modulating the expression of RpoS (which increases expression of a number of genes at high salt concentrations (Hengge-Aronis *et al.*, 1991; Bohringer *et al.*, 1995).

Expressing this *galE* gene in *E. coli* may result in increased UDP glucose levels, which could be converted to trehalose resulting in an osmoprotective effect. Padilla *et al.* (2004) increased the UDP glucose supply and consequently the levels of accumulated trehalose by expressing the *E. coli galU* gene in *Corynebacterium glutamicum*. In our study, the expression of both the *galE*(3) and *galE*(25) genes in the osmosensitive strain *E. coli* MKH13 resulted in a statistically significant increased salt (NaCl and KCl) tolerance phenotype (Figures 2b and c). Of the five genes identified and cloned, *galE*(3) had a lesser effect than the other four genes. MKH13::pCI372-*galE*(3) also had a prolonged lag phase. It has been demonstrated that cellular stress may be caused by imbalances and accumulation of certain intermediary metabolites, particularly in the case of amphibolic pathways (for example, D-galactose pathway), as demonstrated in a *galE* mutant (Lee *et al.*, 2009). In the current study, supplying MKH13 with an additional plasmid-encoded copy of *galE* may also cause an imbalance in an intermediary metabolite, leading to stress and ultimately a prolonged lag phase, which ends when the imbalance is corrected and homoeostasis is restored.

Peptidoglycan is a major component of the bacterial cell wall and has an important role in withstanding osmotic stress (van Heijenoort, 1996). MurB is essential for cell wall biosynthesis and is involved in a two-step process with UDP-*N*-acetylglucosamine enolpyruvyl tranferase (MurA) to form UDP-*N*-acetylmuramate, which is a building block for peptidoglycan (Sylvester *et al.*, 2001). The *murB* gene has been shown to be essential for normal growth in *Bacillus subtilis*, *E. coli* and *Staphylococcus aureus* (Miyakawa *et al.*, 1972; Matsuo *et al.*, 2003; Real and Henriques, 2006). In *S. aureus*, a mutation within the *murB* gene resulted in thermosensitive mutants, which had thinner cell walls (Matsuo *et al.*, 2003). The *murB*(3) gene conferred an increased growth phenotype to *E. coli* MKH13 during both NaCl and KCl stress (Figures 2b and c). Disruption or deletion of the *murB* gene could make cells acutely sensitive to osmotic stress owing to a reduction in cell wall integrity as well as causing a reduction in turgor pressure, which is a driving force for cellular growth and division. Bacteria remodel the structure of their peptidoglycan in response to changes in environmental conditions (Vijaranakul *et al.*, 1995; Quintela *et al.*, 1997), which could be important in the gut by allowing for varying levels of rigidity or elasticity depending on the conditions in the immediate environment.

We have also identified two putative MazG family proteins (encoded by the *mazG* gene from SMG 3 and SMG 5 (*mazG*(3) and *mazG*(5), respectively). These are highly conserved proteins in bacteria and to date there has been no tangible link demonstrated between MazG and salt tolerance. MazG is a nucleoside triphosphate pyrophosphohydrolase (NTPase), which can hydrolyse (deoxy)ribonucleoside triphosphates ((d)NTPs) to their corresponding (deoxy)ribonucleoside monophosphates ((d)NMPs) and pyrophosphate (PPi) (Zhang and Inouye, 2002). It has been proposed that MazG has a role in cellular 'house-cleaning' by removing abnormal (d)NTP's from nascent DNA strands (Galperin *et al.*, 2006), in addition to regulating programmed cell death in

E. coli (Gross *et al.*, 2006). It also regulates intracellular levels of (p)ppGpp, the main nutritional stress signal molecule involved in the stringent response, which has been shown to be an important response in *Campylobacter jejuni* during intestinal colonization (Stintzi *et al.*, 2005). By reducing (p)ppGpp levels, MazG has a central role in maintaining cell viability during nutritional stress, which could be important in the gut during periods of intermittent availability of certain nutrients, whereas (p)ppGpp itself is required for growth during osmotic stress in *L. monocytogenes* (Okada *et al.*, 2002). Moreover, recent work has identified a role for MazG in the mycobacterial oxidative stress response and virulence (Sassetti and Rubin, 2003; Lu *et al.*, 2010). As outlined above MazG has a number of different roles in different bacteria, but it is clear that stress response is a common theme. This study provides evidence for a novel role of MazG in salt tolerance. In the context of the gut environment, MazG may be important in removal of mutagenic nucleotides from growing DNA strands. Damage to bacterial DNA is likely due to exposure to genotoxic compounds, such as nitrosamines and heterocyclic amines (Kurokawa *et al.*, 2007). In addition, MazG could provide energy to cell during stress through the hydrolysis of ATP. Because of the numerous roles of MazG in different stress conditions, it may function as general stress response protein in the bacterial cell. More research and discovery of novel MazG proteins will help identify new physiological roles, substrates and precise mechanisms of action for these proteins in the myriad of stress conditions imposed on microorganisms.

In conclusion, we have identified five genetic loci involved in salt tolerance from within the human gut microbiome using a functional metagenomic approach. The genes represent two different homologues of *galE* and two of *mazG*, as well as a *murB* homologue, from three different species of the genus *Collinsella, Akkermansia* and *Eggerthella*. The identification of three genes within approximately 40 kb of metagenomic DNA from SMG 3 (highest genetic identity to *C. aerofaciens*) is relevant in that functionally related proteins are often co-located on the chromosome of prokaryotic genomes (Sleator and Walsh, 2010; Sleator, 2011). In addition to expanding our knowledge of salt tolerance mechanisms, this study may also facilitate the development of novel drug targets and related approaches to control resident gut microbiota (Sleator, 2010a, b). Ultimately some of the salt tolerance mechanisms identified might be used as part of a 'pathobiotechnology' (Sleator, and Hill, 2006) or 'metabiotechnology' (Culligan *et al.*, 2009) strategy for the design of improved probiotic cultures with greater resistance to process induced stresses (such as spray and freeze drying), as well as improved gut colonization (Sheehan *et al.*, 2006; Sheehan *et al.*, 2007; Watson *et al.*, 2008 for such examples). This will hopefully result in a broader and more comprehensive representation of salt tolerance mechanisms in this unique environment.

Acknowledgements

EPC is funded by Science Foundation Ireland under the CSET Uplift Grant. We acknowledge the continued financial assistance of the Alimentary Pharmabiotic Centre, funded by Science Foundation Ireland. JRM acknowledges funding from The Royal Society, which supports the bioinformatic cluster (Hive) at Cardiff University, School of Biosciences. RDS is an ESCMID Research Fellow.

References

Banik JJ, Brady SF. (2008). Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proc Natl Acad Sci USA* **105**: 17273–17277.

Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP *et al.* (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.

Bohringer J, Fischer D, Mosler G, Hengge-Aronis R. (1995). UDP-glucose is a potential intracellular signal molecule in the control of expression of sigma S and sigma S-dependent genes in *Escherichia coli*. *J Bacteriol* **177**: 413–422.

Chowdhury R, Sahu G, Das J. (1996). Stress response in pathogenic bacteria. *J Biosci* **21**: 149–160.

Culham DE, Lasby B, Marangoni AG, Milner JL, Steer BA, van Nues RW *et al.* (1993). Isolation and sequencing of *Escherichia coli* gene *proP* reveals unusual structural features of the osmoregulatory proline/betaine transporter, ProP. *J Mol Biol* **229**: 268–276.

Culligan EP, Hill C, Sleator RD. (2009). Probiotics and gastrointestinal disease: successes, problems and future prospects. *Gut Pathog* **1**: 19.

Ellrott K, Jaroszewski L, Li W, Wooley JC, Godzik A. (2010). Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput Biol* **6**: e1000798.

Epstein W. (2003). The roles and regulation of potassium in bacteria. *Prog Nucleic Acid Res Mol Biol* **75**: 293–320.

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.

Fukasawa T, Jokura K, Kurahashi K. (1962). A new enzymic defect of galactose metabolism in *Escherichia coli* K-12 mutants. *Biochem Biophys Res Commun* **7**: 121–125.

Galperin MY, Moroz OV, Wilson KS, Murzin AG. (2006). House cleaning, a part of good housekeeping. *Mol Microbiol* **59**: 5–19.

Gardan R, Duche O, Leroy-Setrin S, Labadie J. (2003). Role of *ctc* from *Listeria monocytogenes* in osmotolerance. *Appl Environ Microbiol* **69**: 154–161.

Giaever HM, Styrvold OB, Kaasen I, Strom AR. (1988). Biochemical and genetic characterization of osmoregulatory trehalose synthesis in *Escherichia coli* *J Bacteriol* **170**: 2841–2849.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR *et al.* (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* **68**: 4301–4306.

Gloux K, Berteau O, El Oumami H, Beguet F, Leclerc M, Dore J. (2011). A metagenomic beta-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proc Natl Acad Sci USA* **108** (Suppl 1): 4539–4546.

Gralla JD, Huo YX. (2008). Remodeling and activation of Escherichia coli RNA polymerase by osmolytes. *Biochemistry* **47**: 13189–13196.

Gross M, Marianovsky I, Glaser G. (2006). MazG—a regulator of programmed cell death in *Escherichia coli*. *Mol Microbiol* **59**: 590–601.

Grundling A, Schneewind O. (2007). Genes required for glycolipid synthesis and lipoteichoic acid anchoring in *Staphylococcus aureus*. *J Bacteriol* **189**: 2521–2530.

Haardt M, Kempf B, Faatz E, Bremer E. (1995). The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system ProU of *Escherichia coli* K-12. *Mol Gen Genet* **246**: 783–786.

Hayes F, Daly C, Fitzgerald GF. (1990). Identification of the minimal replicon of *Lactococcus lactis* subsp. lactis UC317 plasmid pCI305. *Appl Environ Microbiol* **56**: 202–209.

Heath C, Hu XP, Cary SC, Cowan D. (2009). Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil. *Appl Environ Microbiol* **75**: 4657–4659.

Hengge-Aronis R, Klein W, Lange R, Rimmele M, Boos W. (1991). Trehalose synthesis genes are controlled by the putative sigma factor encoded by *rpoS* and are involved in stationary-phase thermotolerance in *Escherichia coli*. *J Bacteriol* **173**: 7918–7924.

Jones BV, Marchesi JR. (2007). Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat Methods* **4**: 55–61.

Kageyama A, Benno Y, Nakase T. (1999). Phylogenetic and phenotypic evidence for the transfer of *Eubacterium aerofaciens* to the genus *Collinsella* as *Collinsella aerofaciens* gen. nov., comb. nov. *Int J Syst Bacteriol* **49** (Pt 2): 557–565.

Kapardar R, Ranjan R, Puri M, Sharma R. (2010b). Sequence analysis of a salt tolerant metagenomic clone. *Indian J Microbiol* **50**: 212–215.

Kapardar RK, Ranjan R, Grover A, Puri M, Sharma R. (2010a). Identification and characterization of genes conferring salt tolerance to *Escherichia coli* from pond water metagenome. *Bioresour Technol* **101**: 3917–3924.

Kazimierczak KA, Rincon MT, Patterson AJ, Martin JC, Young P, Flint HJ *et al.* (2008). A new tetracycline efflux gene, tet(40), is located in tandem with tet (O/32/O) in a human gut firmicute bacterium and in metagenomic library clones. *Antimicrob Agents Chemother* **52**: 4001–4009.

Kempf B, Bremer E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch Microbiol* **170**: 319–330.

Kilstrup M, Jacobsen S, Hammer K, Vogensen FK. (1997). Induction of heat shock proteins DnaK, GroEL, and GroES by salt stress in *Lactococcus lactis*. *Appl Environ Microbiol* **63**: 1826–1837.

Kunte HJ. (2006). Osmoregulation in bacteria: compatible solute accumulation and osmosensing. *Environ Chem* **3**: 94–99.

Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A *et al.* (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**: 169–181.

Lakhdari O, Cultrone A, Tap J, Gloux K, Bernard F, Ehrlich SD *et al.* (2010). Functional metagenomics: a high throughput screening method to decipher microbiota-driven NF-kappaB modulation in the human gut. *PLoS One* **5**: e13092.

Lee DG, Jeon JH, Jang MK, Kim NY, Lee JH, Kim SJ *et al.* (2007). Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. *Biotechnol Lett* **29**: 465–472.

Lee SJ, Trostel A, Le P, Harinarayanan R, Fitzgerald PC, Adhya S. (2009). Cellular stress created by intermediary metabolite imbalances. *Proc Natl Acad Sci USA* **106**: 19515–19520.

Leloir LF. (1951). The enzymatic transformation of uridine diphosphate glucose into a galactose derivative. *Arch Biochem Biophys* **33**: 186–190.

Lopez CS, Alice AF, Heras H, Rivas EA, Sanchez-Rivas C. (2006). Role of anionic phospholipids in the adaptation of *Bacillus subtilis* to high salinity. *Microbiology* **152**: 605–616.

Lu LD, Sun Q, Fan XY, Zhong Y, Yao YF, Zhao GP. (2010). Mycobacterial MazG is a novel NTP pyrophosphohydrolase involved in oxidative stress response. *J Biol Chem* **285**: 28076–28085.

Maniatis T, Fritsch EF, Sambrook J. (1982). *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory: Cold Spring Harbor, New York.

Markovitz A. (1977). Genetics and regulation of bacterial capsular polysaccharide synthesis and radiation sensitivity. In: Sutherland WI (ed.). *Surface Carbohydrates of the Prokaryotic Cell*. Academic Press, Inc.: London, pp 415–462.

Matsuo M, Kurokawa K, Nishida S, Li Y, Takimura H, Kaito C *et al.* (2003). Isolation and mutation site determination of the temperature-sensitive *murB* mutants of *Staphylococcus aureus*. *FEMS Microbiol Lett* **222**: 107–113.

Mavromatis K, Ivanova N, Barry K, Shapiro H, Goltsman E, McHardy AC *et al.* (2007). Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nat Methods* **4**: 495–500.

Meilleur C, Hupe JF, Juteau P, Shareck F. (2009). Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. *J Ind Microbiol Biotechnol* **36**: 853–861.

Miyakawa T, Matsuzawa H, Matsuhashi M, Sugino Y. (1972). Cell wall peptidoglycan mutants of *Escherichia coli* K-12: existence of two clusters of genes, mra and mrb, for cell wall peptidoglycan biosynthesis. *J Bacteriol* **112**: 950–958.

Nguyen TT, Klueva N, Chamareck V, Aarti A, Magpantay G, Millena AC *et al.* (2004). Saturation mapping of QTL regions and identification of putative candidate genes for drought tolerance in rice. *Mol Genet Genomics* **272**: 35–46.

Okada Y, Makino S, Tobe T, Okada N, Yamazaki S. (2002). Cloning of *rel* from *Listeria monocytogenes* as an osmotolerance involvement gene. *Appl Environ Microbiol* **68**: 1541–1547.

Padilla L, Morbach S, Kramer R, Agosin E. (2004). Impact of heterologous expression of *Escherichia coli* UDP-glucose pyrophosphorylase on trehalose and glycogen synthesis in *Corynebacterium glutamicum*. *Appl Environ Microbiol* **70**: 3845–3854.

Piuri M, Sanchez-Rivas C, Ruzal SM. (2003). Adaptation to high salt in *Lactobacillus*: role of peptides and proteolytic enzymes. *J Appl Microbiol* **95**: 372–379.

Pumbwe L, Wareham DW, Aduse-Opoku J, Brazier JS, Wexler HM. (2007). Genetic analysis of mechanisms of multidrug resistance in a clinical isolate of *Bacteroides fragilis*. *Clin Microbiol Infect* **13**: 183–189.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**: 59–65.

Quintela JC, de Pedro MA, Zollner P, Allmaier G, Garcia-del Portillo F. (1997). Peptidoglycan structure of *Salmonella typhimurium* growing within cultured mammalian cells. *Mol Microbiol* **23**: 693–704.

Real G, Henriques AO. (2006). Localization of the *Bacillus subtilis murB* gene within the *dcw* cluster is important for growth and sporulation. *J Bacteriol* **188**: 1721–1732.

Saitou N, Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Sakamoto T, Murata N. (2002). Regulation of the desaturation of fatty acids and its role in tolerance to cold and salt stress. *Curr Opin Microbiol* **5**: 208–210.

Sassetti CM, Rubin EJ. (2003). Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci USA* **100**: 12989–12994.

Schiller C, Frohlich CP, Giessmann T, Siegmund W, Monnikes H, Hosten N *et al.* (2005). Intestinal fluid volumes and transit of dosage forms as assessed by magnetic resonance imaging. *Aliment Pharmacol Ther* **22**: 971–979.

Schulman H, Kennedy EP. (1977). Identification of UDP-glucose as an intermediate in the biosynthesis of the membrane-derived oligosaccharides of *Escherichia coli*. *J Biol Chem* **252**: 6299–6303.

Seltmann G, Holst O. (2002). Further cell wall components of Gram-positive bacteria. In: Seltmann and Holst (eds). *Bacterial Cell Wall*. Springer: Germany, pp 131–161.

Sheehan VM, Sleator RD, Fitzgerald GF, Hill C. (2006). Heterologous expression of BetL, a betaine uptake system, enhances the stress tolerance of *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* **72**: 2170–2177.

Sheehan VM, Sleator RD, Hill C, Fitzgerald GF. (2007). Improving gastric transit, gastrointestinal persistence and therapeutic efficacy of the probiotic strain *Bifidobacterium breve* UCC2003. *Microbiology* **153**: 3563–3571.

Sleator RD. (2010a). Probiotic therapy—recruiting old friends to fight new foes. *Gut Pathog* **2**: 5.

Sleator RD. (2010b). Probiotics—a viable therapeutic alternative for enteric infections especially in the developing world. *Discov Med* **10**: 119–124.

Sleator RD. (2011). Phylogenetics. *Arch Microbiol* **193**: 235–239.

Sleator RD, Hill C. (2002). Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol Rev* **26**: 49–71.

Sleator RD, Hill C. (2005). A novel role for the LisRK two-component regulatory system in listerial osmotolerance. *Clin Microbiol Infect* **11**: 599–601.

Sleator RD, Hill C. (2006). Patho-biotechnology: using bad bugs to do good things. *Curr Opin Biotechnol* **17**: 211–216.

Sleator RD, Shortall C, Hill C. (2008). Metagenomics. *Lett Appl Microbiol* **47**: 361–366.

Sleator RD, Walsh P. (2010). An overview of in silico protein function prediction. *Arch Microbiol* **192**: 151–155.

Stintzi A, Marlow D, Palyada K, Naikare H, Panciera R, Whitworth L *et al.* (2005). Use of genome-wide expression profiling and mutagenesis to study the intestinal lifestyle of *Campylobacter jejuni*. *Infect Immun* **73**: 1797–1810.

Sylvester DR, Alvarez E, Patel A, Ratnam K, Kallender H, Wallis NG. (2001). Identification and characterization of UDP-N-acetylenolpyruvylglucosamine reductase (MurB) from the Gram-positive pathogen *Streptococcus pneumoniae*. *Biochem J* **355**: 431–435.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

van Heijenoort J. (1996). Murein synthesis. In: Neidhardt FC, Curtis R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaecter M, Umbarger HE (eds). *Escherichia coli and Salmonella: cellular and molecular biology*. American Society for Microbiology: Washington, DC, pp 1025–1034.

van Passel MW, Kant R, Zoetendal EG, Plugge CM, Derrien M, Malfatti SA *et al.* (2011). The genome of *Akkermansia muciniphila*, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *PLoS One* **6**: e16876.

Vijaranakul U, Nadakavukaren MJ, de Jonge BL, Wilkinson BJ, Jayaswal RK. (1995). Increased cell size and shortened peptidoglycan interpeptide bridge of NaCl-stressed *Staphylococcus aureus* and their reversal by glycine betaine. *J Bacteriol* **177**: 5116–5121.

Watson D, Sleator RD, Hill C, Gahan CG. (2008). Enhancing bile tolerance improves survival and persistence of *Bifidobacterium* and *Lactococcus* in the murine gastrointestinal tract. *BMC Microbiol* **8**: 176.

Wonderling LD, Wilkinson BJ, Bayles DO. (2004). The *htrA (degP)* gene of *Listeria monocytogenes* 10403S is essential for optimal growth under stress conditions. *Appl Environ Microbiol* **70**: 1935–1943.

Zhang J, Inouye M. (2002). MazG, a nucleoside triphosphate pyrophosphohydrolase, interacts with Era, an essential GTPase in *Escherichia coli*. *J Bacteriol* **184**: 5323–5329.

Zuckerkandl E, Pauling L. (1965). Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds). *Evolving Genes and Proteins*. Academic Press: New York, pp 97–166.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)